DEPOCEN

The Dynamics of Vietnam Informal Sector: Which Policy Should Be the Best?

Nguyen Dinh Chuc
Tran Thi Bich
La Hai Anh

The Dynamics of Vietnam Informal Sector: Which Policy Should Be the Best?

Nguyen Dinh Chuc
Tran Thi Bich
La Hai Anh

I. INTRODUCTION

The concept of what constitutes "informal" is vague with a multitude of definitions having been put forward by various authors. The term 'informal sector' was first coined in the 1970s by the International Labour Office (ILO) in their country study of Ghana (International Labour Office 1972) and was used to describe less structured working conditions in a series of studies related to urbanization and poverty. It has since grown to broadly encompass the dualistic economic structure in many developing countries where transactions occur outside the traditional, official or state-regulated channels. The definition is highly heterogeneous and used differently by different authors as shown by "*the wide range of terms: non-observed, irregular, unofficial, second, hidden, shadow, parallel, subterranean, informal, cash economy, black market, unmeasured, unrecorded, untaxed, non-structured, petty production, and unorganized*" (Sindzingre 2006, pp.59).

The informal sector is wide encompassing term referring to the part of the economy that is neither taxed nor monitored by the government (Becker 2004). From the government perspective, informality is where entrepreneurs avoid or subvert authority of the government (Kenyon 2007). Informality is also defined from firms' characteristics perspective. Harding and Jenkins (1989) emphasize on the labour characteristics of firms (undeclared labor, lack of social benefits, sub-minimum wages, poor working conditions) as to define the informality. Tokman (2001) defines the informal sector as comprising firms with limited ownership (i.e. self-employed) that utilize unpaid family members, domestic servants, less educated employees, and have less than five workers (including the owner).

In an effort to give a clear guidance for statistician, ILO (International Labour Office 1993) adopted an international statistical definition of the informal sector and characterised enterprises as informal on the basis of the following criteria: (i) the enterprises are owned by individuals or house-holds which do not have complete accounts that allow a financial separation of the production activities of the enterprise from the other activities of its owner(s), (ii) they produce goods or services produced for sale or barter, (iii) they are not registered under specific forms of legalized businesses or their employment size is below a certain threshold defined by the government, (iv) they are engaged in non-agricultural activities. Due to the complexity and multifaceted nature of the informal sector it is difficult

to pin it with a single definition, instead Gerxhani (2004) proposes that it should be classified by the following criteria: (i) amount of government regulation; (ii) illegal activities; and (ii) national statistics.

By any definition, a large informal sector is a typical feature of transitional and developing countries. The World Bank estimates that at least 30% of output and 70% of labour in developing countries exists outside the formal sector and contrary to previous belief that this would disappear, the share of informality is growing (Kenyon 2007). In Vietnam, it is estimated 11 out of 46 million jobs are generated by the informal sector and 8.4 million household businesses operate informally, contributing to at least 20% to GDP (Cling, Razafindrakoto et al. 2011). The magnitude of the sector would suggest that the challenges and opportunities it represents should be the subject of greater policy focus. From the dualistic view, informal economy provides a safety net for people who lose or are unable to fund a job within the formal sector. Additionally it also provides employment for disadvantaged groups including women, the handicapped, and children (USAid 2005). According to Becker (2004), 84% of female employment in sub-Saharan Africa comes from the informal sector. On the other hand, from structuralist and legalist view, the informal enterprises select themselves into the sector in the industrial world and workers sometimes prefer self-employment to salaried jobs (Maloney 2004; De Mel, McKenzie et al. 2011).

Large informal sector is usually seen as weaknesses in economic management. According to Moser (1984), the appearance of informal sector reflects failure of the macro-economy. Moreover, employment within the informal sector is not as secure as that in the formal sector (USAid 2005). At the country level, high proportion of informal sector results in loss in tax revenues and lower economic growth (see Loayza 1996; Dabla-Norris and Feltenstein 2005). At the enterprise level, formalization increases profits and investment (Fajnzylber, Maloney et al. 2011; Rand and Torm 2011). At the household level, working in formal sector implies having access to social insurance and more secured employment (Rand and Torm 2011).

There is a general consensus that mainstreaming the informal sector into the formal economy is a desired objective but there remains debate on the degree of formalization that is necessary, what formalization entails and how one goes about achieving it. In particular, policy debates run the risk of an overemphasis on regulation, penalties, or frameworks to institute administrative legitimacy. Rather, balanced and purposeful policy would focus on the main benefits of formalization, that is, to build economic security for both firms and individuals. There is no consensus on factors affecting formalization of firms. Research by USAid (2005) argues that formalization of enterprises is limited by the regressive fees (e.g., business registration fees), taxes, minimum capital requirements that effectively penalize smaller firms. While Mead and Morrisson (1996) shows that formalization is affected by firms characteristics, including legality, size, capital intensity, tax burden, labor and other regulations… As a result, finding out incentives for firms to choose formalization based on understandings of firms' characters and business environment give policy implications, enhance the efficiencies of policy interventions.

In Vietnam, the informal sector has predominated in term of job creations. Results from the Vietnam labour force survey in 2007 reveal that this sector generates 23.5 per cent of

employment whereas the public, foreign direct investment (FDI), and private domestic sectors create 10.7 per cent, 2 per cent, and 5.7 per cent of jobs, respectively (Cling, Razafindrakoto et al. 2011). The generation of large numbers of jobs has been one of the key factors behind the country's success in reducing poverty incidence during the past two decades. The existence of informal enterprises is, therefore, a resource for the country to ensure the middle-income status. However, along with the development in the world, the government wants to apply policies to facilitate the formalization process of firms in informal sector. Understanding characteristics and environment variables that affects the transition of firms especially from informal to formal is critical to the success of such policies.

This study is one of the efforts to understand an important sector of Vietnamese economy – the informal sector. The main focus of the study is the formalization process of enteprirses in Vietnam. The study try to answer the following question: What are systematic relations between firm characteristics, behaviours, incentives with formalization process? How are the dynamic transition between formality and informality affected by characteristics, behaviours, incentives of firms? Does performance of firms play any role in the dynamic transition between formality and informality of firms?

The analysis is conducted with several hypotheses, inclusing: the cost of formalization outweighs its benefit from the business perspective so that it discourages businesses in informal sector to move to the formal one. Moreover, firms with higher performance will move to formal sector. Since formalizing requires costs, firms that are more efficient in using inputs to maximizing outputs/profit will overcome this barrier. And legal and institutional environment, especially at the local level, needs to be strengthened to facilitate movement from informal to formal sector.

The paper uses non-parametric approaches in doing quantitative analysis. To examine the technical performance of enterprises, an nonparametric technique will be used. Also two techniques including the multiple factor analysis and cluster analysis are used to build informality index and explore the relations between firms characteristics, behaviours, incentives, performance with formalization process. The understandings, therefore, help to design effective policy interventions.

II. REVIEW OF PAST RESEARCHES

The concept "informal sector" has evolved to encompass various types of cash and non-cash transaction in both developing and industrialized economies–transactions that benefit both the poor and non-poor. The informal economy includes those enterprises and activities that may not comply with standard business practices, taxation regulations, and/or business reporting requirements but are otherwise not engaged in overtly criminal activity. In the Kenya report (ILO, 1972), the popular view of informal sector activities is that they are primarily those of petty traders, street hawkers, shoeshine boys and other groups underemployed on the streets of the big towns. The characteristics of the informal sector include: easy to join, depending on the available resources at local, family ownership of enterprises, small-scale, labor - intensive and adapted technology, without formal training, difficult to control the market and low competitiveness.

The informal sector also may be described in general terms as a set of units producing goods or services principally in order to create jobs and income for the persons concerned. Informal sector production units have the characteristics of unincorporated enterprises as defined in the 4th revision of the SNA (System of National Accounts). The activities of informal sector production units are necessarily carried out with the deliberate intention of avoiding payment of tax or social security contributions, or of infringing labor or other laws or other administrative provisions. The concept of informal sector activities should therefore be set apart from hidden or underground economy activities.

In empirical studies, informal sector has its own operational definition. According to these studies, there are two types of informal enterprise can be distinguished: (i) informal enterprises of persons working for their own account; (ii) enterprises of informal employers.

According to the ILO (International Conference of Labor Statisticians), the informal sector is defined as all unregistered unincorporated enterprises (called informal household businesses; IHBs). As in many countries, farm activities (agriculture, forestry and fisheries) are excluded since their characteristics are very different from non-farm activities (seasonality, labor organization, level of income generated, legislation, etc.).

Until now, there is no consensus on the definition of the informal sector in Vietnam lead to controversy about the number of non-farm household businesses in Vietnam. Two main statistical sources have been used to try to accurately measure these businesses are VHLSS and AHBS. In 2007, a suitable framework was designed and put into practice to measure the informal sector and informal employment in Vietnam. This framework was developed in line with international recommendations and tailored to the Vietnamese context (Razafindrakoto, Roubaud and Le Van Duy, 2008). Accordance with International Labour Organization recommendations (ILO, 2002), the following definition has been adopted: The informal sector is defined as all private unincorporated enterprises that produce at least some of their goods and services for sale or barter, are not registered (no business licence) and are engaged in non-agricultural activities". The results drawn from the Labor force survey shows that the informal economy is predominant in Vietnam. Most employment (82 per cent) in Vietnam can be defined as informal employment. A new improved questionnaire was designed for the Labour Force Survey (LFS) to paints an overall picture of the informal sector in Vietnam. Also in 2007, the informal sector accounted for almost 11 million jobs out of a total of 46 million. This represents nearly a quarter of all main occupations (24 per cent), with nearly half of non-farm work found in the informal sector. The characteristics of this sector in Vietnam is similar to that observed in other developing countries: proportion of the economy, scale size, demographic characteristics of social workers, unstable operation and working conditions, low productivity and income, lack of funds and lack of investment and integration into the economy of other issues.

In developed economies, informality involves tax evasion and undeclared labor rather than a significant share of unregistered businesses. Many formal firms seem to be evading taxes and using undeclared labor. The share of self-employed workers in emerging economies is still relatively high, when compared to more advanced economies. The design of appropriate policies to reduce informal activities and promote formalization is the first require.

Gutmann (1977) and Duncan (1992) argue that formal and informal works are gross substitutes for one another, that is, when people lose jobs in the formal economy, they turn to informal jobs to make ends meet. The degree of dependence between the formal sector and informal is strong and growing. The economic restructuring have strong impact to the transition from the informal sector into the formal sector. This process requires high capital investment and is connected to the formal economy (either through capital transactions, product distribution, or wage setting).

The transition from the informal economy to the formal is the desirable goal in strategy promote transition from informal to formal activities in Africa (ILO, 2008), however, the concept how is formalized and how to achieve it still exist many different perspectives. The development goals are focused on making the labor market becomes more efficient and fair to both workers and businesses. The benefits of converting to consider in terms of enhancing social and economic security, create platforms to attract investment and a long-term vision about their future than maintaining informal activities. In the path towards formalizing, expanding the capacity of the system and organization is necessary, whether the reform should go much further in reinventing the frameworks, instruments and culture of outreach to suit the specific conditions of the informal economy; Promoting two objectives of preserving and expanding the employment, income generation, poverty reduction of the informal economy and extending social protection, rights to the vast majority of those working in the informal economy.

To solve the problem of unofficial economy, macroeconomic policy should be oriented toward achieving growth and poverty reduction, with the macro-economic stability. The goal is how to enhance productivity, facilitate economic transformation, and increase the availability of suitable jobs, and improve the ability to work of employees. ILO (2008) suggests the process of formalization including: promoting a greater awareness of the benefits and protection that come with formalization, creating an enabling policy and regulatory environment that reduces, both at the national and local levels, the barriers to formalization while protecting workers' rights, a particular focus on women entrepreneurs, fostering linkages between enterprises of different sizes in value chains and clusters to improve market access, access to finance and business development services, and encouraging informal enterprises to join together in production conglomerates or cooperatives.

According to Thiam (2007), the transition from informal to formal is a normal process of industrialization and development. Process is triggered through incentives and enabling environment reforms (e.g., access to credit, trade facilitation, formalization of business linkages), making costs of formalization worthwhile. Informality limits enterprise growth and access to markets. High degree of informality outside tax nets limits ability of governments to mobilize resources to provide public goods. Informality represents absence of social protection and benefits. Barriers to formalization include: (i) reluctance of entrepreneurs to expose themselves (especially when revenue collection is characterized by rent-seeking); and (ii) costs that may put formal firms at a competitive disadvantage.

Research about the voluntary formalization in Tazania focus mainly on the process of formalization by observes from the perspective of owner managers of informal enterprises. According to this paper, process of formalization proceeding in two main steps from

informal to semi-formal and then to formal status, however, an operator may graduate directly from informal status to formal and ignore the intermediate steps and an operator may remain at any level of informal or formal status differing periods of time, from one year to 29 years (Nelson and Bruijn, 2005).

In exploring determinants of formalization, using a general-equilibrium model on firm-level dataset collected by WB's World Business Environment survey Dabla-Norris, Gradstein, & Inchauste (2008) find that both regulation burden and legal quality are important determinants of informality. Other researchers shows that informality relates to tax burden (e.g., Cebula, 1997; Giles and Tedds, 2002), entry costs (Auriol and Warlters, 2005); institutional quality and regulatory burden, in particular of labor (Friedman and others, 2000, Johnson and others, 1997, 1998, 2000; Botero and others, 2004); and financial development (Straub, 2005). Andrews, Sánchez, & Johansson (2011) found that, cross-country estimates of informality suffer from large measurement problems and suggest to use household and firm level data to adequately capture the behavioural responses of firms and households to policies.

Review of previous researches shows that there is no concensus on informal sector. At the same time, there are many factors that are potential determinants of informality. Previous studies that use a macroeconomic framework in analysing the informal sector are biased and do not adequately capture the behavioural responses of firms and households. Therefore firm level data is needed in analysing the informal sector and formalization process.

III. METHODOLOGY

There are various definitions of formal enterprises which are subjective to researchers and policy makers. To better capture the formality of each firm, we propose, in the first step of the analysis, to construct a formality index via synthesizing these definitions. This index is set as the score on the first principal axis derived from many firms' characteristics related to the formality definitions (capital, labor, registration status, enterprise type...). These characteristics can be a mixture of quantitative, categorical and frequency data, which could be integrated via Multiple Factor Analysis (MFA). The formal enterprise is defined when the constructed index is positive otherwise the informal enterprise is implied. Using dataset collected from Small and Medium Enterprise (SME) surveys, formality indices are constructed. In the *second* step, the combination of MFA and the cluster analysis is used to group firms sharing similar characteristics within and between groups. These techniques can be applied for cross-sectional datasets (to extract main features of the informal and formal enterprise groups at one time and check the consistency of these features over time) as well as panel datasets (to draw out main firms' characteristics affecting formality transition status in short and long time).

Step 1: Construct the formality index using MFA

The informality of firms has been described in relation to aspects such as compliance with government regulations (via registration, payment of tax and adherence to labour regulations); size of the firm; resource endowment and applied technology (labour or capital intensive), location, the physical place of operation and the characteristics of workforce and

ownership. Tokman (2001) defines the informal enterprises as firms with limited ownership (i.e. self-employed) that utilize unpaid family members, domestic servants, less educated employees, and have less than five workers (including the owner). However, the descriptions of informal firms according to their size are not uniform. Some scholars apply a cut-off point of ten people (e.g. Arimah 2001) and others five employees (e.g. Tokman 1978). Furthermore, some researchers relate the informal firms to micro or small ones since they argue that firms of these sizes are not likely to comply with government regulations (Anderson 1998; Mlinga and Wells 2002). However, there are also large firms that do not follow all government regulation, particularly labour regulations (Tokman 2001; Mlinga and Wells 2002). In reality, there are firms that should be defined as semi-informal since parts of their business are informal but they utilise casual labour contract (Djankov, Lieberman et al. 2003).

Several issues are emerged when studying informal sector in Vietnam. *First*, the formalization of a business involves acquiring business registration certificate (BRC) and tax registration certificate (TRC).[1] However, in fact, some firms were operating with both a BRC and a TRC, while others had a BRC and no TC (Rand and Torm, 2011). *Second*, although legally registered firms[2] improve access to formal credit, unregistered firms are able to use their land use-right certificate as collateral for a loan and therefore do not need a BRC for this purpose. *Third*, normally when firms (with more than 10 employees) formalize, they should be required to register the use of labor. However, in fact, there is no requirement for small and medium firms such as household firms to register the use of labor, and therefore the employment relationship does not automatically become subject to regulation. In addition, workers in unregistered firms are often hired on a casual basis without contracts and therefore not entitled to receiving social benefits. However, many of these firms covered medical expenses for their workers as well as the costs of work-related accidents (Rand and Torm, 2011).

With the broad and controversial definition as well as conflicted facts of informal sector, we try to construct a formality index via synthesizing these definitions. This index is set as the score on the first principal axis derived from many firms' characteristics/variables related to the formality definitions. They include variables representing (i) *compliance with government regulations* (having a BRC, having a TRC, payment of tax, contracted labor, employees with social insurance, employees covered medical and accidental expenses) and (ii) *firm's resource endowment* (number of employees). These variables can be a mixture of quantitative, categorical and frequency data, which could be integrated via the multiple factor analysis (MFA) instead of the principal component analysis (PCA).[3]

Multiple Factor Analysis (MFA)

The multiple factor analysis (MFA)[4] is used to analyze a set of observations described by several groups of variables. The number of variables in each group may differ and the nature of the variables (quantitative, categorical or frequency) can vary from one group to the

---

[1] The Government Decree No. 88/2006/ND-CP dated August 29, 2006.
[2] Under the Vietnamese Enterprise Law, when a firm has more than 10 employees and/or owns more than one business premise, it should register for BRC.
[3] Normally, the PCA is relevant to numerical variables only.
[4] See Escofier and Pages (1990, 1994), Abdi and Valentin (2007)

other but the variables should be of the same nature in a given group. The analysis derives an integrated picture of the observations and of the relationships between the groups of variables.[5]

Multiple Correspondence Analysis (MCA) has reached a high level of development and use (eg. Nishisato, 1980; Gifi, 1981; Greenacre, 1984; Lebart, Morineau, & Warwick, 1984; Richardson & Kuder, 1933; Hirshfeld, 1935; Horst, 1935; Fisher, 1940; Guttman, 1941; Burt, 1950; Hayashi, 1950) under a variety of rationales. In statistics, multiple correspondence analysis (MCA) is a data analysis technique applied for nominal and categorical data in order to (a) extract the most important information from the data table, (b) compress the size of the data set by keeping only important information, (c) simplify the description of the data set, and (d) analyze the structure of the observations and the variables. By representing data as points in a low-dimensional Euclidean space, MCA thus appears to be the counterpart of PCA for categorical data (Le Roux and Rouanet, 2004, Greenacre *et al.* 2006).

MCA is performed by applying the CA algorithm to an indicator matrix (Greenacre, 2007). In the indicator matrix approach, associations between variables are uncovered by calculating the chi-square distance between different categories of the variables and between the individuals (Le Roux and Rouanet, 2004). Oppositions between individuals and variables are then maximized to find out the underlying dimensions which are best able to describe the central oppositions in the data. As PCA, the first axis is the most important dimension in terms of the amount of variance accounted for. The number of axes retained for analysis is determined by modified eigenvalues.

We hence will use both definitions of informality: the classical one and the composition index which is estimated by MCA. Similar to Nguyen *et al.* (2011) study[6], the latter is defined as being informal if positive and formal if negative. This analysis standardizes the highest axial *inertia* of every set of variables to 1 for balancing their importance. Assume I observations are described by J sets of variables: J_q sets of quantitative variables, J_C sets of categorical variables and J_f sets of frequency variables, in which:

$J = J_q + J_C + J_f$

- if j is a quantitative set, the value x_{ikj} of variable k for observation i
- if j is a categorical set, $z_{ikj} = 1$ if i belongs to category k and 0 if not
- if j is a frequency set, proportion f_{ikj}, computed as the ratio between the number of occurrences of event k for observation i and the total over the table that gathers all the J_f frequency tables.

---

[5] MFA is a data reduction method that is similar to the Principle Component Analysis (PCA). The PCA was developed in early 20th century (Pearson 1901, Hotelling 1933) in psychometrics and multivariate statistical analysis. In economics, the method has been applied to the studies of cointegration and spatial convergence, development, panel data, forecasting… (see Kolenikov, S. and G. Angeles (2009) for detail review).

[6] Despite of the same definition, Nguyen *et al.* (2011) do not try to use sets of variables directly establishing informality definition. Source: Nguyen Huu Chi, Tran Kim Thu and Tran Thi Bich (2011), "Informality, business outcomes and heterogeneity among rural non-farm household enterprises in Vietnam: An application of multivariate analysis", *DEPOCEN Working Paper Series No. 2011/18.*

In order to balance the influence of the different sets, theoretically, MFA considers the cloud associated with each set of variables and standardizes the *inertia* of every cloud on the first principal axis to 1. Technically, this property is obtained by dividing the weight of the columns belonging to set j by λ_1^j, the first *eigenvalue* of the separate analysis of set j. Let p_i be the weight assigned to observation i.

The extended MFA is based on a non-standardized weighted PCA performed on the multiple table using:

• $\{p_i = f_{i.}; i = 1,...,I\}$ as row-unit weights

• the initial weights of the columns (belonging to set j) divided by λ_1^j as column weights, that is, $(1/\lambda_1^j)$ in the case of a quantitative set, $(w_{kj}/\lambda_1^j Q_j)$ in the case of a categorical set[7], $(f_{.kj}/\lambda_1^j)$ in the case of a frequency set.

MFA induces a distance between observations corresponding to a weighted sum of the separate distances induced by every set of variables. The square distance between units i and l is computed by:

$$d^2(i,l) = \sum_{j \in I_Q} \frac{1}{\lambda_1^j} \sum_{k \in K_j} \left[\frac{x_{ikj} - x_{lkj}}{s_{kj}} \right]^2 + \sum_{j \in I_c} \frac{1}{\lambda_1^j} \sum_{k \in K_j} \frac{1}{Q_j w_{kj}} \left[z_{ikj} - z_{lkj} \right]^2 +$$

$$\sum_{j \in I_f} \frac{1}{\lambda_1^j} \sum_{k \in K_j} \frac{1}{f_{.kj}} \left[\left(\frac{f_{ikj}}{f_{i.}} - \frac{f_{lkj}}{f_{l.}} \right) - \frac{f_{.kj}}{f_{..j}} \left(\frac{f_{i..}}{f_{i.}} - \frac{f_{l..}}{f_{l.}} \right) \right]^2 \tag{1}$$

The relation that gives the coordinate $F_s(i)$ along axis s corresponding to the coordinates of the columns $\{G_s(kj)\}$ is obtained by applying the general transition formula (Pagès, 2002):

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j \in I_q} \frac{1}{\lambda_1^j} \left[\sum_{k \in K_{jq}} x_{ikj} G_s(kj) \right] + \frac{1}{\sqrt{\lambda_s}} \sum_{j \in I_c} \frac{1}{\lambda_1^j Q_j} \left[\sum_{k \in K_j} z_{ikj} G_s(kj) \right]$$

$$+ \frac{1}{\sqrt{\lambda_s}} \sum_{j \in I_f} \frac{1}{\lambda_1^j} \frac{f_{i..}}{f_{i.}} \left[\sum_{k \in K_j} \frac{f_{ikj}}{f_{i..}} G_s(kj) \right] \tag{2}$$

MFA computes new variables called principal components F_s obtained as linear combinations of the original variables. The *first principal component F_1* is required to have the largest possible variance. The *second component F_2* is computed under the constraint of being orthogonal to the first component and to have the largest possible inertia. The other

---

[7] To integrate categorical sets into MFA, The starting point is to make use of the equivalence between MCA and a non-standardized weighted PCA. The results of MCA can be obtained by performing PCA to the table of $(z_{ikj} - w_{kj})/w_{kj}$ where $w_{kj} = \sum_{i \in I} p_i z_{ikj}$ and $\sum_{k \in K_j} w_{kj} = Q_j$ (Q_j is number of variables belonging to set j), using the weight w_{kj}/Q_j to column k of set j and the weight p_i to row i.

components are computed likewise. The values of these new variables for the observations are called *factor scores*, which can be interpreted geometrically as the projections of the observations onto the principal components.

With *I* firms and *J* sets of variables/characteristics related to the formality definitions (as described above), factor scores from the first principal component F_1 will reflect the formality index. For checking purposes, two definition of formality is used.

(i) *The basic definition*: A firm is formal if they have a TRC (Rand and Torm, 2012).

(ii) *Our definition*: We define the positive index as formality and the negative index as informality. This definition is also used in Nguyen (2011)[8].

From the SMEs 2005, 2007, 2009 and 2011, we have four formality indices. Taking advantages of getting the panel datasets from these four surveys, changing signs between every two indices give information on two main groups: (i) transition groups from informality to formality or from formality to informality; and (ii) unchanged groups. These groups can be examined in short-term (2 years) or longer-term (4-6 years, in the case of panel datasets built between 2005 and 2009, 2007 and 2011, or 2005 and 2011). The latter consideration is especially useful to look at firms' characteristics varying in long time (e.g. technical efficiency index that is estimated using non-parametric model).

Step 2: Cluster firm's homogenous characteristics within and between the four transition groups using the MFA and the cluster analysis

This step will answer the research questions. Firms with homogenous characteristics will be grouped by cluster analysis techniques. The objective of cluster analysis is to partition a set of observations into clusters in such a manner that all observations within a cluster are similar, while observations in different clusters are differential. To discover the clusters in the data, an algorithm is applied using the proximity measure, which is either the degree of distance called 'dissimilarity' or the degree of 'association' named as 'similarity' between clusters. This method is applied to analyze a set of principals factors extracted from MFA and *active* variables, which describe enterprises' characteristics to *supplementary* variables. Using SME surveys, the study analyzes a set of *active* and *supplement* variables as follows:

(i) *Entrepreneurial characteristics*: age, gender, education, professional/skill level, experience.

(ii) *Client networks*: main clients, revenue share from main clients, input providers, social networks.

(iii) *Business environment*: difficulties in business establishment and development process, facing administrative procedures, access to credit.

(iv) *Enterprises' characteristics:* years of operation, business location, applied technology, capital-labor ratio, type of main products, number and characteristics of employees, participating in business associations.

---

[8] Despite of the same definition, Nguyen *et al.* (2011) do not try to use sets of variables directly establishing informality definition.

(v) Enterprises' outcomes: technical efficiency, value-added, profit, initial capital, competitive capacity.

(vi) *Enterprise dynamics*: change in technical efficiency, change in number of employees, change in profits, change in value-added, change in capital.

The *complement* variables include entrepreneurial characteristics and business environment variables, applied technology, technical efficiency, change in technical efficiency. The other variables are in the *active* group. In clustering the enterprises, a procedure is implemented, which consists of five steps presented as follows:

(i) *Selecting variables for analysis*: As the cluster analysis is undertaken as the second step after MFA, the *active* variables selected in MFA are factors which reduce the complexity of the dimensional space of the original variables.

(ii) *Selecting a distance measure*: Cluster analysis is performed based on assessing similarity of *supplement* variables by measuring the distance between them. The closer the observations are, the more similarity they have. The most common measure used in cluster analysis is Euclidean distance (Equation 1) to assess the shortest distance between two points. In the case cluster analysis is performed based on categorical data, the preferred distance measure is chi-square[9].

(iii) *Selecting a clustering procedure*: Clustering procedure is a hierarchical algorithm that produces clustering results. The hierarchical clustering algorithm starts with assigning each firm to its own cluster and joins the two closest clusters together. The Euclidian distances between firms are used as criterion to assess the similarity. The joining process of similar clusters continues until all firms are grouped into clusters. The output of this process is a hierarchical tree.

(iv) *Deciding the number of clusters*: Choosing the number of clusters can be facilitated by examining the histogram of the increasing level of indices. Each bar histogram indicates the loss of inertia obtained from one partition of *s* clusters to the partition of *s - 1* clusters.

(v) *Interpreting the profile of clusters*: the interpretation is based on a set of variables and the associated modalities that can be best used to characterize each cluster. In defining these variables, inferential comparisons of means or percentages of firms in a cluster with the means or percentages on all firms are undertaken. Value tests are performed to determine the modalities significantly correlated with the cluster. Those modalities positively correlated with the analyzed cluster are representable for that cluster.

Therefore, the combination between MFA and the cluster analysis clusters firms sharing similar characteristics within and between groups. These techniques can be applied for cross-sectional datasets (to extract main features of the informal and formal enterprise groups at one time and check the consistency of these features over time) as well as panel datasets (to draw out main firms' characteristics affecting formality transition status in short and long time).

---

[9] See Greenacre (1984) for more details

Non-parametric approach to efficiency analysis

One of contribution of the paper is to explore the relation between formalization of firms and their performance. The performance of firms in this study is analysed from technical efficiency perspective, measuring how close is a firm to production frontier. The technical efficiencies of firms are estimated by using a popular non-parametric approach to efficiency analysis, DEA (Data Envelopmen Analysis). The production frontier in this case is formulation non-parametrically from firm observations. The nonparametric frontier is a frontier of a convex production set, invented by Charnes et al. (1978), who coined the term Data Envelopment Analysis (DEA) for the approach to identifying the convex efficient frontier. The DEA estimation of technical efficiency could be presented as a measurement of ratio between given inputs of a set to the boundary input for the same set of outputs. As presented by Daraio and Simar (2007) we will have production levels of firms that dominate other units in the same industry and create the famous DEA production frontier:

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

Where $\left(X_i, Y_i\right)$ are observations in a convex hull of $\chi = \left\{\left(X_i, Y_i\right), i = 1,...,n\right\}$ covering unit $\left(x,y\right)$.

The above formula allows the variable returns to scale production technology, where outputs under efficient production change by a different proportional to the change in inputs. Other 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 presentation of a constant returns to scale technology as described earlier. While setting $\sum_{i=1}^{n} \gamma_i \geq 1$ or ≤ 1 we allow respectively for non-decreasing or non-increasing returns to scale, respectively.

With variable returns to scale production technology, the input-oriented technical efficiency score for a production unit operating at the level $\left(x_0, y_0\right)$ will be:

$$\hat{\lambda}_{DEA}\left(x_0, y_0\right) = \min\left\{\lambda \left| y_0 \leq \sum_{i=1}^{n} \gamma_i Y_i; \lambda x_0 \geq \sum_{i=1}^{n} \gamma_i X_i; \lambda \geq 0; \right.\right.$$

$$\sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0; i = 1,...,n$$

with the input-oriented technical efficiency score $\hat{\lambda}_{DEA}$, to achieve the output level $\left(y_0\right)$ the projection of $\left(x_0, y_0\right)$ on the efficient boundary is $\hat{\lambda}_{DEA} * x_0$. Therefore the difference

between x_0 and $\hat{\lambda}_{DEA} * x_0$ is the radial distance which measures the efficiency of a production unit in producing a given level of output (y_0).

Similarly, the output-oriented approach to technical efficiency will arrive at the DEA efficiency by solving the optimization problem:

$$\hat{\theta}_{DEA}(x_0, y_0) = \max \left\{ \theta \middle| \theta y_0 \leq \sum_{i=1}^{n} \gamma_i Y_i ; x_0 \geq \sum_{i=1}^{n} \gamma_i X_i ; \theta \geq 0; \right.$$

$$\sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0; i = 1, \dots, n$$

In our study, since firms in the sample is small and medium enterprises are price takers and affected by business environment, the appropriate approach is input-oriented technicall efficiencies. This dimension of technical efficiency as firms' characteristics will be estimated before the two-step approach of analysis as presented above being conducted. Using technical efficiency, the study is possible to analyze the relation between performance and formalization of firms in dynamic transition framework. Particularly, the panel data allow us to apply Malmquist index decomposition method (Thanassoulis 2001; Ray 2004; Cooper, Seiford et al. 2006) to formulate the "catching up" index in term of technical efficiency by comparing production combination of one enterprise to its current and past production frontier. Using the "catching up" index enables us to investigate the relationship between the change over time in production technology and formalization of enterprises in the sample.

In sum, instead of using a particular definition of formality as used in previous researches, this study propose to use MFA to formulate a formality index to measure the formalization propensity of firms. Utilizing MFA and cluster analysis, in addition to technical efficiency dimension of firms and a panel data of 4 years, this study explore the systematic relations between formalization and characteristics, performance, behaviours, and incentives of firms in a dynamics process.

IV. INFORMALITY AND FIRM PERFORMANCE

4.1. Informality

The informal dimensions of organizational life became increasingly important and a commonplace topic for research since the 1950s and 1960s (Gouldner, 1954; Blau and Scott, 1963). However, until the 1970s Keith Hart (1970, 1973) was the first one to bring the term 'informal sector' into the academic literature, which is described as a part of the urban labor force and takes place outside of the formal labor market. In spite of the early work by Hart, the report of the International Labor Office in Kenya (ILO, 1972) is widely considered as the pioneering research on the informal sector, which is mainly characterized by the avoidance of government regulations and taxes. In 1993, informal enterprises defined by the 15th International Conference of Labour Statisticians (ICLS)[10] are defined as those owned by

---

[10] 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.

individuals or households, constituted as non-separate legal entities and without complete accounts. To distinguish informal enterprises, the 15th ICLS recommended to use the following three criteria: non-registration of the enterprise; small size in terms of employment; and non-registration of the employees.

In addition to defining unregistered enterprises as informal, a substantial number of analyzes of informality simply assume that the informal sector corresponds to micro or small enterprises (Anderson, 1998; Mlinga and Wells, 2002) or applies various cut-off points 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), 20 for Sudan (Cohen and House, 1996). Their views are based on the assumption that the majority of firms in these sizes are likely not to comply with government regulations.

According to Duval-Hernández (2006), informality ought to be measured not only according to the employer but also according to the worker's legal status. Theoreticaly, contract status might provide a good discriminator for informal status if the allocation of workers to the informal sector is governed by a free assessment of the costs and benefits of labor contract registration. Practically, however, such a measure has no relevance for the case of self-employed workers since they cannot contract with themselves. Therefore, the alternative 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. Their results show that the informal sectors measured by the latter criteria produce moderately higher rates than those estimated by the former. For instance, 43.2 compared to 30.8 percent in Mexico in 1990 (Marcoullier et al., 1997), 54.8 vs. 53.3 percent in Peru in 1995 (Saavedra and Chong, 1999). Similarly, a recent work by Pisani and Pagán (2004) examines the informal sector in Nicaragua by employer size (5 or less employees) and by social security registration. The informal sector measured by social security registration appears to have grown fast during the 1990s from 73 to 85 percent for male and 67 to 79 percent for female. In contrast, in terms of employer size, the informality rates were observed with a sharply falling trend. This suggests that different measures may behave very diversely.

Like other countries, several issues are emerged when studying informal sector in Vietnam. *First*, the formalization of a business involves acquiring a business registration certificate (BRC) and registering for a tax code (TC)[11]. However, in fact, many firms were operating with both BRC and TC, while others had BRC but no TC[12] (Rand and Torm, 2012). As TC authorities require BRC, it is not possible to have a TC but no BRC. Holding BRC depends on types of businesses. For example, BRC is not compulsory for some specific kinds of household businesses, including those related to (i) agriculture, forestry, fishery, salt production; (ii)

---

[11] The Government Decree No. 88/2006/ND-CP dated August 29, 2006.
[12] This indicated that government officials would come to collect (usually on a monthly basis) a lump-sum tax/fee.

street vendors; (iii) mobile businesses; and (iv) low-income[13] services (Article 49, Decree 88/2006/NĐ-CP). Note that a household business employing more than ten employees must be transformed into an enterprise. An enterprise has to hold Business Registration Certificate (BRC) and a Tax Code (TC)[14] and open a simple accounting book without double entries.

Second, normally when firms (with more than 10 employees) formalize, they should be required to register the use of labor.[15] However, in fact, this registration is not required for small and medium firms (eg. household businesses), and therefore the employment relationship does not automatically become subject to regulation. In addition, workers in unregistered firms are often hired on a casual basis without contracts and therefore not entitled to receiving social benefits. However, some of these firms covered medical expenses for their workers as well as the costs of work-related accidents (Rand and Torm, 2011).

Therefore, beside using the simple and classical definition of informal businesses (those without BRC and/or TR), this paper constructs an informality index via synthesizing a number of criteria. This index is set as the score on the first principal axis derived from many firms' characteristics/variables related to the informality definitions. They include variables representing (i) incompliance with government regulations (no BRC, no TC, employees without social insurance, employees without health insurance, no accounting books, unpaid workers), (ii) firm's resource endowment (number of employees) 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 multiple correspondence analysis (MCA) instead of the principal component analysis (PCA).[16]

4.2 Data

4.2.1 Variable Description

There are totally 2,500 - 2,800 businesses in SME surveys each year. Eight variables are selected: (i) having BRC, (ii) having TC, (iii) paying social insurance, (iv) paying health insurance, (v) having an accounting book, (vi) types of ownership, (vii) number of full-time employees, (viii) number of unpaid employees. Table 1 shows the description of these varialbes. While 60.6 percent of enterprises having BRC, most of them have tax code (96.5 percent). The percent of businesses holding BRC and therefore tax code increase over time up to 73 percent in 2011. However, the social and health insurance benefits go to a much smaller amount of enterprises which explain for 15.7 percent of total businesses in 2005 although this ratio increases over time up to 20 percent in 2011. More than 70 percent of

---

[13] The low income levels are regulated by Provincial People's Committees.

[14] According to Article 8 - Decree 43/2010/NĐ-CP, enterprise code is (i) also enterprise registration code and tax code; (ii) unique for an enterprise; and (iii) exists during the operation of an business.

[15] According to Article 141 (The 1994 Labour Law), it is compulsory for all enterprises and registered HBs to register their permanent employees (with at least a three-month employment contract) with the Vietnam Social Security (VSS). The employer has to pay 15 percent and employees pay 5 percent of salary for social insurance contribution.

[16] Normally, the PCA is relevant to continuous variables only.

registered businesses have not paid any social and/or health insurance. Only 37-42 percent of enterprises have accounting books.

Out of all interviewed enterprises, the majority are household businesses, explaining 68 percent in 2005 and this rate decreases steadily over time up to 64 percent in 2011. There is also an increasing trend of micro and small businesses who employ fewer than five workers, from 37.6 percent in 2005 to 45.4 percent in 2011 while larger businesses (with five or more employees) tend to decrease. Moreover, unpaid workers seem to be popular among SMEs with only 22.7 percent of enterprises without any unpaid workers in 2005, increasing to nearly one third in 2011. These figures imply that if only taking the condition of holding BRC/TC to define the formality then in many cases, other conditions of formality will be violated.

Table 1. Description of variables to define informality index

Percentage (%) of enterprises	2005	2007	2009	2011
Have Tax Code	58.49	61.21	65.63	70.92
Have Business Registration Code	60.58	62.20	67.32	73.00
Pay Social Insurance	15.35	16.09	19.37	20.49
Pay Health Insurance	15.70	16.09	19.33	20.96
Have Accounting Book	36.94	36.39	41.52	39.30
Ownership				
- Household Businesses	*68.27*	*67.97*	*65.21*	*64.26*
- Private company	*10.10*	*7.93*	*8.05*	*7.95*
- Partnership or Collective	*3.72*	*4.14*	*3.12*	*2.78*
- Limited company	*15.74*	*17.42*	*19.97*	*20.85*
- Other ownership	*2.16*	*2.54*	*3.65*	*4.15*
Full-time workers				
- Under 5 workers	*37.61*	*40.72*	*42.16*	*45.42*
- 5 to 10 workers	*30.27*	*28.88*	*28.21*	*27.08*
- Above 10 workers	*32.12*	*30.40*	*29.64*	*27.51*
Unpaid workers				
- No unpaid worker	*22.65*	*26.30*	*29.26*	*30.53*
- 1 unpaid worker	*25.63*	*22.16*	*21.66*	*22.02*
- 2 unpaid workers	*35.94*	*35.29*	*35.43*	*35.70*
- 3 unpaid workers	*8.93*	*8.50*	*8.42*	*7.29*
- 4 unpaid workers	*3.97*	*5.31*	*3.84*	*3.17*
- 5 unpaid workers	*1.84*	*1.52*	*0.79*	*0.82*
- From 6 unpaid workers	*1.03*	*0.91*	*0.60*	*0.47*
Number of enterprises	**2,821**	**2,635**	**2,659**	**2,552**

Source: SME 2005, 2007, 2009, 2011

4.2.2 Criteria interaction

Figure 1 gives the visual view of the interaction among variables using SME 2011. It can be recorganized that only small part of unregistered businesses (those without TC and BRC) hold any other condition of formality, such as paying social and health insurance, having accounting book, non-household businesses, above 10 workers or no unpaid employees. In contrast, a large percent of registered businesses violate any other condition of formality. Some of them still keep their household businesses instead of legally transforming into enterprises because of employing more than 10 workers. Although nearly a half of non-household businesses haven't participated into social security fund, nearly 100 percent of household businesses don't care about that. Employees in household businesses also face much higher risk of being unpaid.

Figure 1. Venn Diagrams showing correlation between criteria
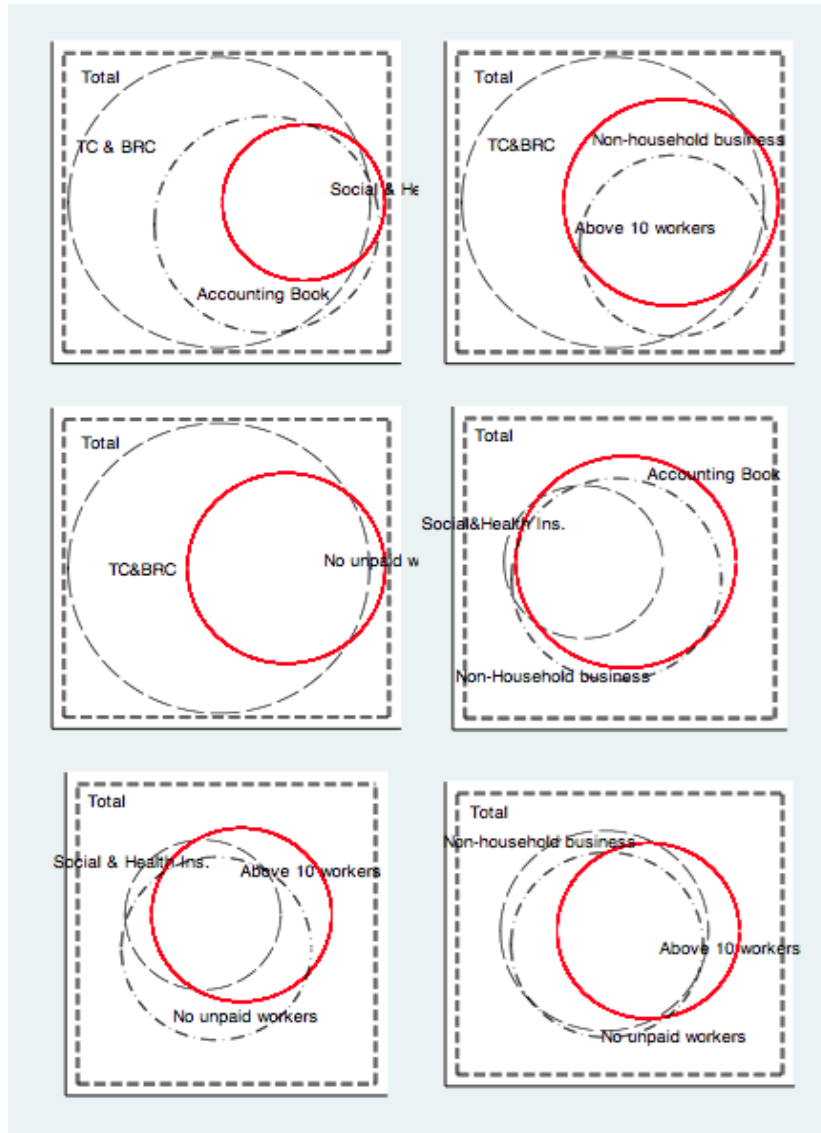
Source: SME 2011

4.3. Multi-dimension of classical informality index

Using the basic definition where the informality is defined as businesses without BRC or TC, almost all informal businesses include all informal characteristics: no social and health insurance, no accounting booking, household businesses, at least one unpaid worker, below 10 workers (Figure 2). However, according to this definition, formal businesses are not defined well. Among them, 75 percent has not paid social and health insurance, 40 percent without accounting book, 50 percent being household businesses and employing less than 10 workers, and 60 percent having at least one unpaid worker (Figure 2). This also suggests the use of MCA to define the informality index.

Figure 2. Multi-dimension of classical informality index
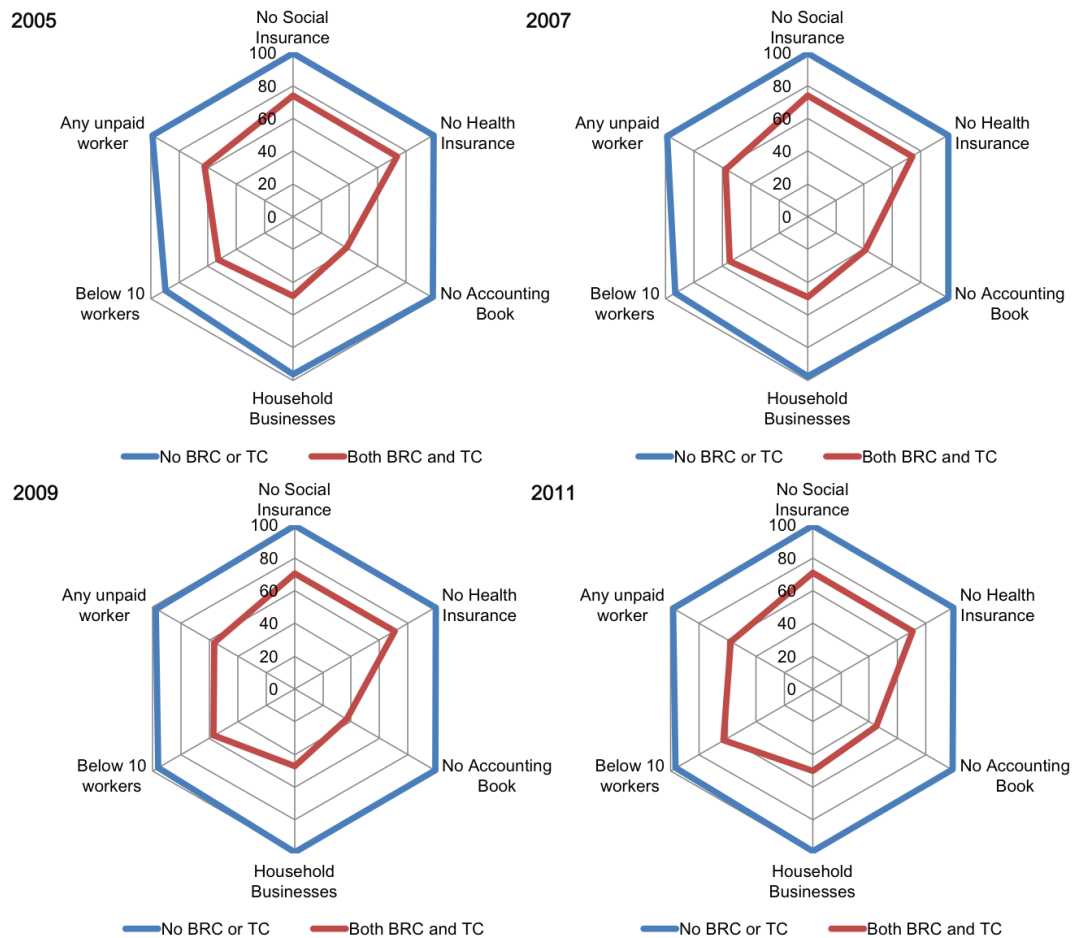
Source: SME 2005, 2007, 2009, 2011

4.4. Multi-dimension of MCA informality index

In order to measure the new informality index, MCA is applied. A set of eight nominal and categorical variables is used including (i) holding BRC, (ii) holding TC, (iii) paid social insurance, (iv) paid health insurance, (v) types of ownership, (vi) number of full-time employees, (vii) number of unpaid workers and (viii) having an accounting book. The variances (eigenvalues) of the first five axes and their inertia rates are given in Table 2. The MCA result[17] shows that the first dimension can explain 91-92 percent of the total variance (Table 2). The one-dimensional solution was adopted because the two-dimensional solution would have only explained about 2 percent of the variance.

Table 2. MCA Calculation

---

[17] Note that the PCA and MCA give the quite similar results which are shown in Appendix D.

	Dim1	Dim2	Dim3	Dim4	Dim5
2005					
Eigenvalues	**0.31**	0.01	0.00	0.00	0.00
% inertia (variance)	**91.63**	1.98	0.06	0.02	0.00
2007					
Eigenvalues	**0.35**	0.01	0.00	0.00	0.00
% inertia (variance)	**91.27**	1.71	0.07	0.01	0.00
2009					
Eigenvalues	**0.34**	0.01	0.00	0.00	0.00
% inertia (variance)	**92.12**	1.77	0.03	0.00	0.00
2011					
Eigenvalues	**0.34**	0.01	0.00	0.00	0.00
% inertia (variance)	**91.16**	2.22	0.03	0.01	0.00

Source: SME 2005, 2007, 2009, 2011

The principal coordinates of nominal and categorical variables in two dimensions are plotted in Figure 3 using the SME 2011. From Figure 3 we see that BRC and TC are located on the right of the first dimension and the top of the second one while no BRC and no TC are on the left of the first dimension and the bottom of the second one. For other variables, the positive values of the first axis represent perfectly all conditions of informality: no social and health insurance, under five employees, at least one unpaid worker, no accounting book, and household businesses. 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. Meanwhile, the values of the second principal axis doesn't interpret well the informality index. This visual view suggests taking positive values of the first principal axis as informality and negative values as formality.[18] The results are consistent and robust for all four SME surveys (see Appendix A).

Figure 3. MCP Coordinate Plot

---

[18] For a complete and correct interpretation of the graphical display in Figure 3, we use additional information presented in Appendix B and Appendix C.

MCA Coordinate Plot, 2011

coordinates in standard normalization
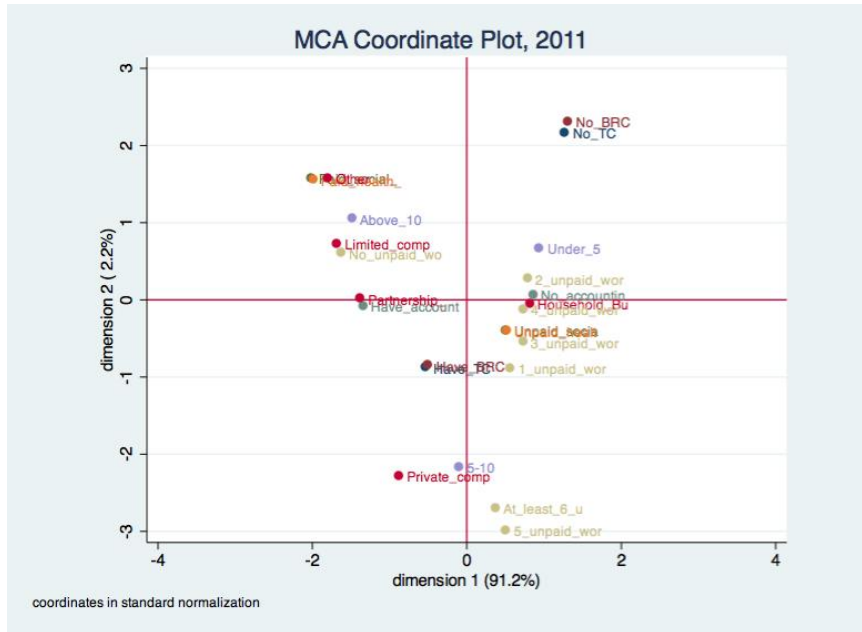
Source: SME 2011

Besides that, we check the MCA index by dividing this index by 5 quintiles. Figure 4 shows that the first three quintiles reflect almost perfectly the informality because of statisfying all informality conditions. Meanwhile, in spite of imperfect description, the fourth quintile is more likely to reflects the informality because of some illegal factors (enterprises with non-household ownership and accounting books but without paying social and health insurance and having unpaid workers). In contrast, the last quintile represents relatively well the formality, especially for SME 2011 where only a small percentage of businesses violate any formality conditions.

Based on the above arguments as well as the difference between the informality index defined by MCA and classical definition with only less than 50 percent matched (see Figure 5), we use the MCA informality index for the rest of our analysis.

Figure 4. Multi-dimension of MCA informality index
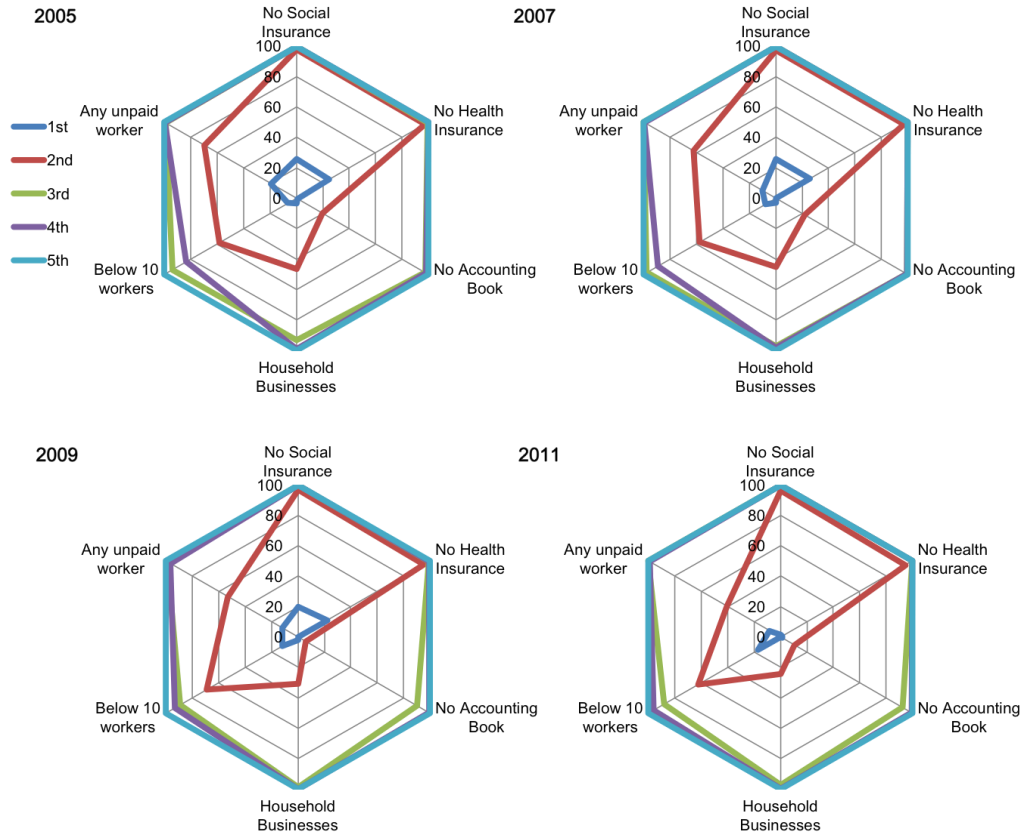
Source: Our calculation from SME 2005, 2007, 2009, 2011

Figure 5. TC/BRC and Informality
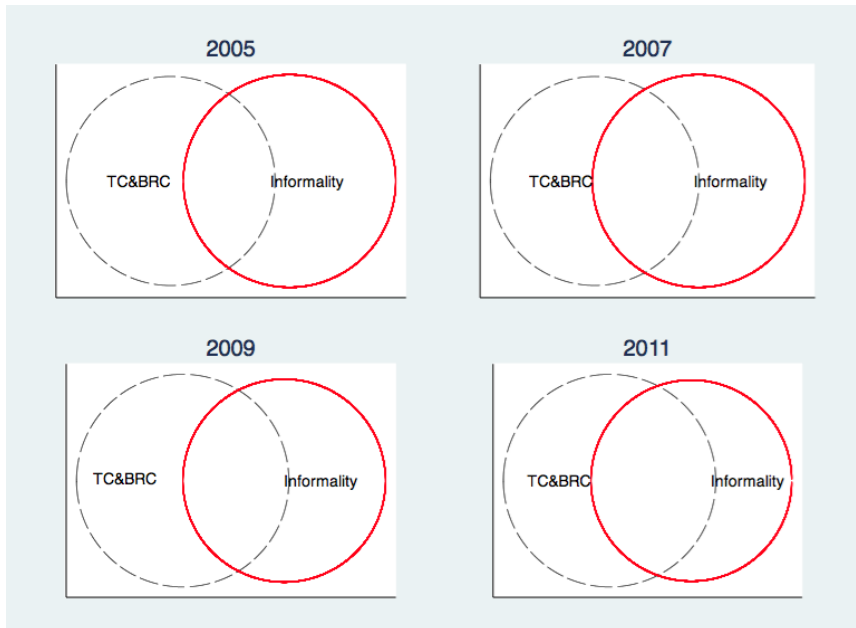
Source: Our calculation from SME 2005, 2007, 2009, 2011
Notes: Informality is defined = 1 if positive MCA index

4.5. Firm performance: output and efficiency

In this research we add one more dimension to the analytical senarior by estimation of technological efficiency of firms in the surveys. The following section shows the performance of the firms in accordance with several definitions of informal section presented above.

Firstly, in accordance with number of full-time employees, **Table 3** shows that in average larger firms have absolute advantages over other firms with less employees. Particularly, in 2009 in average each firm in the group of having more than 10 employees has revenue 31 times higher than a firm with less than 5 employees and 7 times higher than firm with the number of employees between 5 to 10. In 2011 the gap is even larger between firms of different size. In term of productivity, larger firms have higher productivity, both labour and capital productivity. However, larger firm does use the same technology as smaller firm. Technical efficiency of a larger firm is even lower than one of small firm. Larger firm is obviously more efficient in scale, as showed by significantly higher than small firm.

Comparing two time period, the figures in the **Table 3** show that there is improvement over time for firms in all sizes. Firms performed better in 2011 than in 2009 in all indicators, including revenue, value added, size of firms, productivity. However, there is no significant improvement in term production technology used by firms.

Table 3. Output and efficiency of firms by size

Number of employee	<5	5-10	>10
2009			
Total revenue	359,717	1,593,865	11,500,000
Total value added	88,882	374,835	2,924,274
Fixed assets	623,097	2,008,119	9,031,859
Variable capital	44,529	245,007	1,755,973
Total labour	3	7	40
Labour productivity	33,043	52,273	72,055
Capital productivity	0.62	0.64	0.69
Technical efficiency scores	0.62	0.48	0.55
Scale efficiency scores	0.62	0.83	0.87
2011			
Total revenue	502,736	2,203,045	28,700,000
Total value added	141,334	585,002	4,338,657
Fixed assets	1,263,047	3,526,096	14,100,000
Variable capital	73,875	376,838	2,036,616
Total labour	3	7	40
Labour productivity	52,703	82,116	100,613
Capital productivity	0.70	0.59	0.79
Technical efficiency scores	0.59	0.51	0.62
Scale efficiency scores	0.65	0.87	0.87

Sources: Authors' estimation

Secondly, in term of ownership household businesses have much lower perfomance in comparison with other types of ownership. They have significant lower revenue, size in both labour and capital, and value added (**Table 4**). While the gap in productivity and efficiency is much smaller. The data also show improvement over time of firms of different ownership types.

Table 4. Output and efficiency of firms by type of ownership

	Household Businesses	Private company	Partnership or Collective	Limited company	Other
2009					
Total revenue	849,801	7,075,220	4,376,842	10,600,000	18,500,000
Total value added	223,843	1,448,379	1,167,082	2,837,776	4,335,405
Fixed assets	1,126,237	5,166,849	6,472,281	8,937,590	11,400,000
Variable capital	110,864	871,810	710,932	1,815,340	2,467,999
Total labour	5	21	31	35	53
Labour productivity	37,030	63,984	48,475	82,827	80,516
Capital productivity	0.64	0.58	0.76	0.63	0.78
Technical efficiency scores	0.56	0.51	0.48	0.58	0.59
Scale efficiency scores	0.70	0.82	0.88	0.86	0.80
2011					
Total revenue	1,009,209	4,626,644	7,375,983	32,900,000	16,900,000
Total value added	301,074	1,295,759	1,704,436	4,252,657	4,727,583
Fixed assets	2,157,800	5,666,842	15,900,000	12,800,000	11,800,000
Variable capital	135,829	941,954	786,258	1,977,632	2,521,727
Total labour	5	18	24	33	55
Labour productivity	56,019	87,003	74,025	121,759	87,172
Capital productivity	0.62	0.61	1.02	0.87	0.93
Technical efficiency scores	0.56	0.58	0.52	0.64	0.63
Scale efficiency scores	0.72	0.86	0.82	0.88	0.88

Sources: Authors' estimation

Thirdly, in term of the number of unpaid employees **Table 5** show that, firms without unpaid employees perform much more better than the rest. The figures also show that firms with unpaid employees operate at the same level of performance regardless the number of the unpaid employees. An interesting observation from the data is that all firms of different size of unpaid employees have the same efficiency level, both technical and scale efficiency. This fact show that firms in Vietnam are similar in technology level. The technical performance of firms by number of unpaid employees improves over time from 2009 to 2011.

Table 5. Output and efficiency of firms by number of unpaid employees

	Number of unpaid employee						
	0	**1**	**2**	**3**	**4**	**5**	**6**
2009							
Total revenue	11,000,000	1,227,768	1,209,765	1,148,692	976,565	1,067,645	1,422,649
Total value added	2,762,671	300,583	316,674	291,746	254,817	367,150	464,289
Fixed assets	8,801,555	1,543,426	1,325,523	1,232,250	1,227,750	1,569,923	2,159,543
Variable capital	1,752,416	152,361	153,874	121,981	153,036	173,560	185,995
Total labour	35	6	7	6	13	7	13
Labour productivity	76,468	43,722	39,596	34,332	26,801	43,320	32,322
Capital productivity	0.66	0.69	0.66	0.54	0.42	0.41	0.33
Technical efficiency scores	0.56	0.57	0.57	0.55	0.51	0.51	0.42
Scale efficiency scores	0.86	0.64	0.74	0.74	0.73	0.82	0.78
2011							
Total revenue	26,100,000	1,279,939	1,051,252	1,327,186	1,650,542	1,526,248	1,406,142
Total value added	3,863,686	386,008	322,328	341,711	454,599	511,268	460,305
Fixed assets	12,400,000	2,134,477	2,339,192	2,188,743	4,312,314	2,300,877	2,295,227
Variable capital	1,917,525	208,209	144,322	153,593	131,182	181,967	206,583
Total labour	34	6	5	5	7	8	10
Labour productivity	109,009	63,830	56,048	51,986	66,851	64,737	50,810
Capital productivity	0.88	0.60	0.68	0.42	0.40	0.32	0.94
Technical efficiency scores	0.62	0.60	0.55	0.50	0.51	0.47	0.41
Scale efficiency scores	0.87	0.70	0.74	0.73	0.79	0.85	0.81

Sources: Authors' estimation

V. DYNAMICS OF THE INFORMAL SECTOR

5.1. Sources of informality and dynamics

A high level of informality often exists in a society where rules and regulations are applied unevenly (Tenev et al., 2003; Dabla-Norris et al., 2005). Existing theories assume that formality entails better access to public goods as well as imposes taxes and/or costs of complying with regulatory requirements (Marcouiller and Young, 1995; Azuma and Grosman, 2002; Straub, 2005). The literature also provides contradictory hypotheses on firm dynamics that (i) small firms stay as small because of either satisfying behaviours or structural constraints to growth; and (ii) desire to evade of formal regulations versus lack of access to them (Cunningham and Maloney, 2001).

Empirical studies find mixed impacts of tax burden on the size of informality (Friedman et al., 2000; Giles and Tedds, 2002). Other factors such as entry costs, institutional quality and regulatory burden, labour, and financial development are found to have significant impacts on the informality level (Botero et al., 2004; Dabla-Norris et al., 2005; Auriol and Warlters, 2005).

In terms of labour market point of views, formal or informal jobs are considered as a selection process. On the one hand, the informal sector is seen as the shelter for workers who queue for a job in formal enterprises. Contrarily, the sector is chosen by people who prefer self-employment to waged jobs (Hart, 1972). Therefore, performance of enterprises do not matter on the transition from informal to formal businesses and vice versa.

Although opinions on the existence of the informal sector are contradictory, there is a unanimous agreement that informal firms are often at small and medium size, and informality is costly. At the national level, informality means the loss of taxes for social welfare and mis-allocation of resources that hinder the economy development. At firm level, informality impedes firms to access to secure property rights, formal contract mechanism, and financial services (Levenson and Maloney, 1998), as well as trade across the country border(Tenev et al., 2003). At worker level, informal labourers are not provided with social security by their entrepreneurs. Therefore, identifying reasons why firms select the status as they are is very important for policy responses to enhance the movement towards formality.

Cunningham and Maloney (2001) argue that contradictory theories and mixed empirical evidence can be reconciled by accepting the fact that small and medium firms are very heterogeneous. If we can segment firms into different fractions, then theoretical and empirical evidence spectrum are correct for each segment of firms. Inspired by this idea, this section uses the cluster analysis technique to cluster a panel sample SMEs in 2009-11 in Vietnam into different clusters.

5.2. The dynamic of informal sector in Vietnam

5.2.1 Descriptive statistics

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 our informality index calculated in Section 4, Table 6 documents the formal-informal status and dynamics of firms during the period 2009-11.[19]

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)*

---

[19] Since the results for other years are similar, therefore we present here only analysis of the two years 2009 and 2011. This fact does not change our analysis conclusion.

As can be seen from **Table 6**, informal businesses are over-representative 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 stated out in Section 5.1, tax burden, entry cost, and institutional quality are among the determinants of informality level. However, those might not be appropriate in the case of Vietnam. The Law of enterprise revenue tax (2008) and Nghi dinh 56/2009/ND-CP exempts taxe for SMEs and the entry cost is excluded from obstacles listed by surveyed firms. These two factors and institutional quality are, thus, excluded from our analysis because provinces in Vietnam experience a similar pattern of regulations.

In the case of Vietnam, we try to investigate the following hypotheses:

(i) Small firms stay as small to satisfy their behaviours or to cope with structural constraints to growth;
(ii) Firms desire to evade of formal regulations or there is lack of access to formal institutions;
(iii) Informal sector is a shelter for workers who queue for a job in the formal sector or this is a selection process.

Cluster analysis is applied for the panel sample of 1774 firms. As pointed out in Section 2, the first step in CA is to select variables. In our study, we select four groups of variables including characteristics of entrepreneurs and firms, firm performance and dynamics, and policy environment. Table 2 presents summary statistics of variables included in CA.

Table 7. Summary statistics of variables used in cluster analysis

	Stayed as informal		Formal		Moving from informal to formal		Moving from formal to informal		Total
Number	1051		620		31		72		1774
Percentage	59.24		34.95		1.75		4.06		100
1. Characteristics of entrepreneurs									
Sex									
Male (n=1,108)	64.2	[60.8,67.3]	29.5	[26.5,32.7]	2.1	[1.3,3.3]	4.3	[3.1,5.9]	100
Female (n=666)	50	[45.6,54.3]	44.6	[40.3,49.0]	1.3	[0.6,2.7]	4.1	[2.7,6.3]	100
Age (mean)	47.81		44.45		43.77		48.57		46.60
Education									
Not finished primary (n=31)	85.1	[67.5,94.0]	5.5	[1.3,19.9]	0		9.4	[2.9,26.9]	100
Finished secondary (n=650)	88.1	[84.9,90.6]	6.7	[4.8,9.2]	1.9	[1.0,3.6]	3.3	[2.1,5.3]	100
Finished high school (n=1,093)	41.1	[37.8,44.5]	52.5	[49.1,55.9]	1.7	[1.0,2.8]	4.6	[3.3,6.3]	100
Technical skill									
Unskilled (n=121)	78.6	[69.2,85.7]	12.4	[7.4,20.2]	1.7	[0.4,6.6]	7.3	[3.3,15.5]	100
Elementary worker (n=92)	64.9	[53.2,75.0]	29.2	[19.8,40.8]	2.4	[0.6,8.8]	3.5	[1.0,11.0]	100
Technical worker (n=1,121)	75.5	[72.5,78.3]	18	[15.5,20.7]	2	[1.3,3.3]	4.5	[3.3,6.1]	100
College and above (n=440)	12.3	[9.3,16.1]	83.8	[79.5,87.3]	1	[0.4,2.8]	2.9	[1.5,5.6]	100
Have social network (n=1,269)	53.3	[50.1,56.4]	40.9	[37.8,44.0]	2	[1.3,3.0]	3.9	[2.8,5.3]	100
2. Characteristics of firms									
Firm age (mean)	16		11.11		11.97		13.17		14.11
Have electronic access (n=635)	14.9	[12.0,18.4]	78.8	[75.0,82.2]	2.6	[1.6,4.3]	3.7	[2.4,5.7]	100
3. Firm performance and dynamics									
Expansion and innovation in 2009 (n=1,024)	46.9	[43.5,50.5]	47.1	[43.6,50.6]	1.9	[1.2,3.1]	4.1	[2.9,5.7]	100
Expansion and innovation in 2011 (n=1,051)	47.7	[44.3,51.2]	43.6	[40.2,47.1]	2.8	[1.9,4.1]	5.9	[4.4,7.8]	100

Efficiency (mean)									
Technical efficiency in 2009	0.57		0.54		0.44		0.49		0.56
Technical efficiency in 2011	0.56		0.61		0.62		0.49		0.57
Scale efficiency in 2009	0.68		0.86		0.83		0.83		0.75
Scale efficiency in 2011	0.69		0.86		0.86		0.85		0.76
4. Policy environment									
Constraints in 2009									
No constraint (n=300)	77.1	[71.1,82.2]	17.7	[13.2,23.5]	1	[0.3,3.1]	4.1	[2.2,7.7]	100
Capital constraint (n=554)	51.9	[47.1,56.7]	42.9	[38.2,47.6]	2.1	[1.1,3.9]	3.2	[1.9,5.4]	100
Labour constraint (n=46)	47.1	[30.8,63.9]	43.3	[27.9,60.1]	1.5	[0.2,9.9]	8.2	[2.4,24.3]	100
Technical constraint (n=72)	55.5	[42.1,68.1]	39.5	[27.3,53.1]	0.5	[0.1,3.8]	4.5	[1.4,13.7]	100
Market constraint (n=509)	57.8	[52.8,62.7]	35.7	[31.1,40.6]	1.1	[0.5,2.7]	5.4	[3.5,8.2]	100
Outside service constraint (n=101)	64	[52.5,74.1]	27.5	[18.4,38.8]	7	[2.9,16.0]	1.5	[0.4,5.1]	100
Land constraint (n=135)	49.2	[39.9,58.6]	42.5	[33.5,52.0]	2.5	[0.7,8.6]	5.8	[2.6,12.6]	100
Policy constraint (n=45)	70.7	[54.4,82.9]	26.9	[15.2,43.1]	0		2.4	[0.3,15.2]	100
Constraints in 2011									
No constraint (n=303)	82.5	[76.9,86.9]	15.7	[11.4,21.1]	0.6	[0.2,2.0]	1.3	[0.4,3.9]	100
Capital constraint (n=661)	49.2	[44.8,53.6]	43.3	[39.0,47.7]	3.4	[2.2,5.4]	4.1	[2.6,6.3]	100
Labour constraint (n=85)	46.1	[34.7,57.9]	47.2	[35.7,58.9]	1.3	[0.2,8.6]	5.4	[2.0,13.8]	100
Technical constraint (n=63)	55.2	[41.2,68.5]	42.8	[29.7,56.9]	0		2	[0.3,12.3]	100
Market constraint (n=439)	61.1	[55.7,66.2]	33	[28.1,38.2]	0.6	[0.1,2.6]	5.3	[3.3,8.4]	100
Outside service constraint (n=73)	54.7	[41.1,67.6]	29.6	[18.7,43.4]	4.5	[1.1,16.8]	11.2	[5.2,22.5]	100
Land constraint (n=83)	44.3	[32.4,57.0]	47.1	[35.0,59.6]	0		8.5	[3.6,18.7]	100
Policy constraint (n=41)	51.5	[34.3,68.4]	47.2	[30.4,64.5]	0		1.3	[0.2,8.9]	100
Having financial assistance in 2009 (n=532)	56.2	[51.2,61.0]	39	[34.3,43.9]	1.4	[0.7,3.1]	3.4	[1.9,5.9]	100
Having financial assistance in 2011 (n=188)	54.8	[46.8,62.5]	38.3	[30.9,46.2]	2.8	[1.1,6.5]	4.2	[1.8,9.2]	100
Having technical assistance in 2011	23.1	[12.7,38.3]	71.9	[56.5,83.4]	1.8	[0.2,11.6]	3.2	[0.6,16.6]	100

(n=52)									
Having technical assistance in 2011 (n=63)	25.4	[14.9,39.9]	74.2	[59.8,84.8]	0		0.3	[0.0,2.4]	100
% of management time dealing with government regulations in 2009 (mean)	0.68%		1.83%		1.21%		1.21%		1.11%
% of management time dealing with government regulations in 2011 (mean)	1.61%		3.57%		3.02%		2.26%		2.35%

As can be seen from **Table 7**, entrepreneurs of informal firms and those moving from formal to informal status are the oldest ones. 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 level 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).

Younger firms have more tendency of being formal and electronically accessible than the older informal enterprises. Their average age is just above 10 years old while the informal businesses and those moving from formal to informal are aged of 16 and 13, respectively. This might be because the business environment in Vietnam is less constrained to the new-enters.

In terms of firm performance, formal firms are more efficient than the informal ones. It can be found from Table 2 that the formal group has the highest scale efficiency level (86% compared the 75% on average) while informal businesses have the lowest technical and scale efficiency. Firms moving from informal to formal can improve both technical and scale efficiency level (from 44% to 62% for technical efficiency and from 83% to 86% for scale efficiency). Firms moving in the opposite direction can only improve their scale efficiency level. Their technical efficiency is the same at 49% for the period 2009-11.

Informal firms experience less constraint than their formal counterparts. In case of having constraint, policy obstacle is often cited by the informal group (70.7% in 2009 and this constraint is loosen in 2011 at 51.5%). For formal enterprises, capital, labour, and land constraints are often cited as ones of the most obstacles. In line with theoretical and empirical evidence, formal firms suffer more burdens of regulation inspections than the informal ones. Similarly, firms moving to formal status are more inspected while those moving in the opposite direction can avoid the regulation radar (3.02% and 2.26% of management working time compared to 2.35% on average in 2011). Being formal also receives more assistance from the government (**Table 7**).

Overall, descriptive statistics show that formal firms are more efficient and have more advantages than informal businesses. They, however, experience more constraints and inspections. This might be because formal enterprises have more prospect of expansion and need more resources. The picture is, however, a bit ambiguous. Cluster analysis, which segments firms into different clusters, sheds a deep light into policy environment and firm dynamics.

5.2.2 Cluster and multiple correspondence analyses

In this study, we use hierarchical method because the number of clusters is not known prior to the analysis. The Ward's linkage cluster analysis assigns observations into more homogenous groups based on variables selected in Section 6.1.1. 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) index indicates more distinct clustering. Accompanied with Duda_Hart Je(2)/Je(1) indexes are pseudo_T_squared values. Smaller pseudo_T_squared values indicate more distinct clustering (Duda and Hart, 1973). Dendrogram is indicated in Figure 1 while Duda_Hart indexes are presented in **Table 8**.

Figure 6. Dendrogram and Duda_Hart index
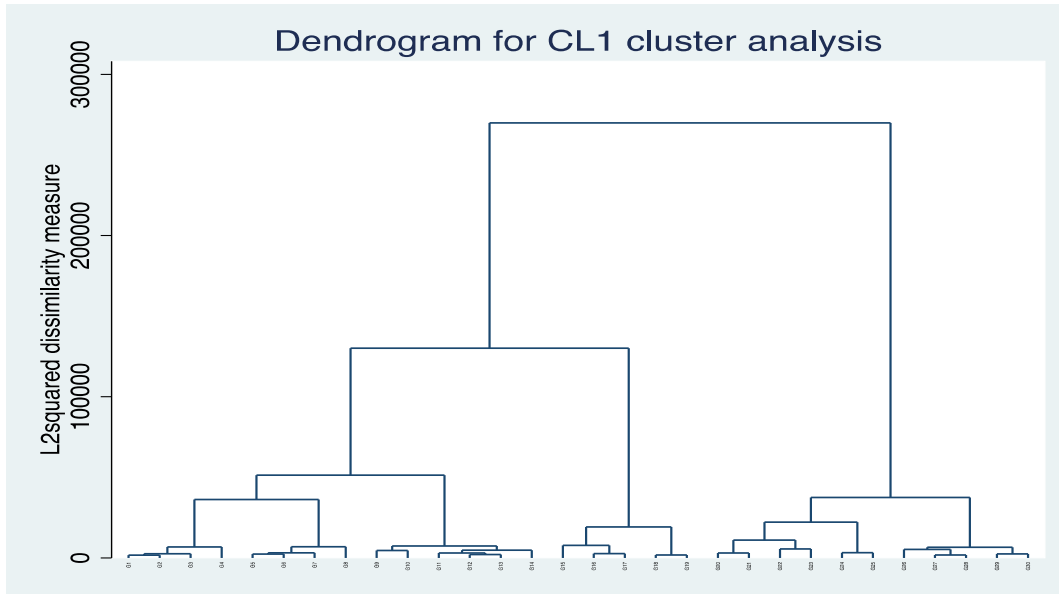
Source: Authors' calculation

Table 8. Duda_Hart indexes for SME clustering

```
+--------------------------------+
Duda/Hart
Number of     pseudo
clusters Je(2)/Je(1) T-squared
-------------+-------------+--------
```

Number 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.

The final step in CA is the interpretation of the clusters. The step starts with the validation of whether clusters are distinct from one another. This can be done by applying oneway ANOVA for quantitative variables and Chi-squared test for categorical variables. Table 4 documents Sidak post-hoc for pair-group comparison after ANOVA for quantitative variables while Table 5 reflects Chi-squared test for categorical variables.

Table 9. ANOVA and Sidak post-hoc for pair-group comparison

	ANOVA	Cluster 1	Cluster 2	Cluster 3	Cluster
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
		***		***	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 10. 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 cluster analysis are meaningful and 4 clusters are distinct from each other. We, then, proceed to the interpretation of clusters. **Table 10** provides streamlined information on cluster characteristics while full table can be found in the Appendix. It should be noted that the number of observations reduces to 1,765 because some missing values are removed from the sample.

Table 11. 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	31%	18%	7%	44%	100%
Formal	15%	28%	4%	53%	100%
Moving from informal to formal	23%	19%	0%	58%	100%
Moving from formal to informal	26%	33%	3%	38%	100%
Efficiency (mean)					
Technical efficiency 2009	0.55	0.56	0.55	0.55	0.55

	Group 1	Group 2	Group 3	Group 4	Total
Number of observation	**441**	**392**	**97**	**835**	**1,765**
Percentage	**25%**	**22%**	**5%**	**47%**	**100%**
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%
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

Based on **Table 11**, characteristics of clusters can be interpreted as follows:

Cluster 1– Informal and moving from formal to informal status. This group is characterised by informal businesses and those moving from formal to informal status. This group of firms run by old entrepreneurs aged of 56 on average with low education, technical skills, and low social networks. 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

regulation interventions (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 and businesses moving from formal to informal status. 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 firms in this cluster are the most interfered by government regulation (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. This group is the oldest one, run by the oldest and low educated people. Having lowest efficiency levels and policy constraint is cited as the most obstacle by firms in this group. The group receives no government assistance as well as interventions.

Cluster 4 – Formal and moving from informal to formal: includes formal firms and those move from informal to formal. This group is at 10 years old on average and run by the youngest entrepreneurs with high level of education and technical skills. 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 growth. 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 intervention because 52% of firms in this group state that they are inspected in 2011 compared to 47% of the sample average.

5.2.3 Policy discussion

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 find out which policies constrain firms' existence and growth and pay attention on technical assistance for firms in this group.

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 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 on those firms. Regulation interventions should be lessened on formal firms to encourage the formality.

Formal firms are often young and stronger enterprises run by capable entrepreneurs. Formality benefits firms on receiving government assistance. To promote the growth of formal businesses, policies should release capital and labour constraints on those firms.

Finally, firms moving from informal to formal status are young and have capacity to growth. Benefits from formality encourage firms moving to formal status but government regulation may be an obstacle of the transition. Therefore, we suggest that to promote the formality of young and strong firms, regulation interventions and policy constraint should be released.

VI. CONCLUSION

By using non-parametric approaches to analyse data collected from surveying more than 2000 SMEs firms in 2005, 2007, 2009, 2011, we define informality of these firms by a combined index. The informality index is built on eight types of information related to the informality of firms. They include (i) holding BRC, (ii) holding TC, (iii) paid social insurance, (iv) paid health insurance, (v) types of ownership, (vi) number of full-time employees, (vii) number of unpaid workers and (viii) having an accounting book. By introducing the new definition of informality of firm, we are able to take into account all available information on informality of firms.

In this research, by looking into the technical performance of the firms we are able to analyse the impact of technology on informal-formal transformation of firms. The panel data for two years 2009 and 2011 enabled us to look into the mechanism and factors influencing the transformation of firms. Particularly, regarding group of *informal and moving from formal to informal firms*, our analysis shows that the most constraint faced by this group is policy constraint, followed by technical and market constraints.

For *formal and moving from formal to informal* firms, this group achieves the highest technical and scale efficiency, while land and market constraints are the most obstacles facing by this group. Also, observation on this group leads to the conclusion that formal firms suffer regulation burdens and the desire to avoid this burden by moving to informal status seem to be correct. Concerning group of *informal and vulnerable firms, t*his group is the smallest groups, run by the oldest and low educated enterpreneurs. This group also has very low efficiency levels, which prevent them from catching up with new technology frontier. The group of f*ormal and moving from informal to formal firms* is characterised by modern enterprises with high electronic access. Their constraint is listed including capital, labour, and technical.

From the analysis, several policy recommendations are made. However, to apply successfully policies for boosting up the efficiency and profitability of firms regarding

their informality, study on impact of policies is crucially needed and should be conducted before application of policies. This is not the focus of this research and need to be dealt with in another study.

References

1. Auriol, E., and M. Warlters, 2005, "Taxation Base in Developing Countries," Journal of Public Economics, Vol. 89, pp. 625–46.
2. Azuma, Y., and H. Grossman, 2002, "A Theory of the Informal Sector," NBER Working Paper No. 8823 (Cambridge, Massachusetts: National Bureau of Economic Research).
3. 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.
4. 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.
5. Dabla-Norris et al., 2005. What Causes Firms to Hide Outputs? The Determinants of Informality. IMF Working Paper, WP/05/160.
6. Duda, R. O., and P. E. Hart. 1973. Pattern Classification and Scene Analysis. New York: Wiley.
7. Friedman, E., S. Johnson, D. Kaufman, and P. Zoido-Lobatón, 2000, "Dodging the Grabbing Hand: The Determinants of Unofficial Activity in 69 Countries," Journal of Public Economics, Vol. 76, pp. 459–92.
8. Giles, D., and L. Tedds, 2002, "Taxes and the Canadian Underground Economy," Canadian Tax Paper No. 106 (Toronto: Canadian Tax Foundation).
9. Hart, K., 1972. Employment, Income and Inequality: a strategy for increasing productive employment in Kenya. Geneva: ILO.
10. 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.
11. 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.
12. 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.
13. Mooi, E., and M. Sarstedt, 2011. A Concise Guide to Market Research. 1st Edition, Springe-Verlag Berlin Heidelberg.
14. Tenev et al., 2003. Informality and The Playing Field in Vietnam's Business Sector. IFC, World Bank, MPDF. Washington, D.C.
15. Straub, S., 2005. Informal sector: the credit market channel. Journal of Development Economics, Vol. 78, Iss. 2, pp. 299-321.
16. Gouldner, A.W., 1954. Patterns of Industrial Bureaucracy, New York, Free Press.
17. Blau, P.M. and Scott, W.R, 1963. Formal Organizations: A Comparative Approach, London, Routledge and Kegan Paul.
18. Hart, K. (1970) Small-scale Entrepreneurs in Ghana and Development Planning. Journal of Development Studies, 6 (4): 104-120.

19. Hart, K., 1973. "Informal Income Opportunities and Urban Employment in Ghana", Journal of Modern African Studies, 11, 1, pp 61-89.
20. ILO (1972). Employment, Income and Equality: A Strategy for Increasing Productive Employment in Kenya. International Labour Office, Geneva.
21. 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.
22. 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.
23. 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.
24. Funkhouser, E. (1996). The urban Informal Sector in Central America: Household Survey Evidence, World Development, 24(11), 1731-1751.
25. Pradhan, M. and van Soest, A. (1995), "Formal and informal sector employment in urban areas of Bolivia", Labor Economics, 2: 275-297.
26. 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.
27. Pradhan, M. and van Soest, A. (1995), "Formal and informal sector employment in urban areas of Bolivia", Labor Economics, 2: 275-297.
28. 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.
29. Livingstone, I. (1991), "A reassessment of Kenya's rural and urban informal sector", World Development, 19(6): 651-670.
30. Arimah, B.C. (2001), "Nature and Determinants of Linkages between Formal and Informal Sector in Nigeria", in African Development Bank, Vol. --, 114-144.
31. 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.
32. Duval-Hernández, Robert. 2006. Informality, segmentation and earnings in urban Mexico. Mimeo. Center for U.S.-Mexican Studies. San Diego, CA, University of California.
33. 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.
34. 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.
35. 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.

36. 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. John Rand and Nina Torm (2012). The benefits of formalization: Evidence from Vietnamese Manufacturing SMEs. World Development, 40(5): 983-998, May 2012.
42. Nishisato, S. (1980). Analysis of categorical data: Dual Scaling and its applications. Toronto: University of Toronto Press;
43. Girl, A. (1981). Nonlinear multivariate analysis. Leiden, The Netherlands: University of Leiden, Afdeling Data-theorie;
44. Greenacre, M. J. (1984). Theory and applications of correspondence analysis. London: Academic Press;
45. Lebart, L., Morineau, A., & Warwick, K. M. (1984). Multivariate descriptive analysis: Correspondence analysis and related techniques for largle matrics. New York: Wiley-Interscience;
46. Richardson, M., & Kuder, G. F. (1933). Making a rating scale that measures. Personnel Journal, 12, 36-40;
47. Hirshfield, H. O. (1935). A connection between correlation and contingency. Cambridge Philosophical Society Proceedings, 31, 520-524;
48. Horst, P. (1935). Measuring complex attitudes. Journal of Social Psychology, 6, 369-374;
49. Fisher, R. A. (1940). The precision of discriminant functions. Annals of Eugenics, 10, 422-429;
50. 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;
51. Burt, C. (1950). The factorial analysis of qualitative data. British Journal of Psychology, 3, 166-185;
52. 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.

53. Le Roux, B. and H. Rouanet (2004). Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis. Dordrecht. Kluwer: p.180.
54. Greenacre, Michael and Blasius, Jörg (editors) (2006). Multiple Correspondence Analysis and Related Methods. London: Chapman & Hall/CRC.
55. Greenacre, Michael (2007). Correspondence Analysis in Practice, Second Edition. London: Chapman & Hall/CRC.
56. Le Roux, B. and H. Rouanet (2004), Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis, Dordrecht. Kluwer: p.179.

APPENDIX 1. CHARACTERISTICS OF CLUSTER

	Group 1	Group 2	Group 3	Group 4	Total
1. Characteristics of entrepreneurs					
Sex					
Male	29%	21%	5%	45%	100%
Female	19%	24%	6%	51%	100%
Total	25%	22%	5%	47%	100%
Age (mean)	56.44	50.88	59.32	37.90	46.59
Education					
Not finished primary	39%	13%	26%	23%	100%
Finished secondary	32%	20%	7%	41%	100%
Finished high school	21%	24%	4%	52%	100%
Total	25%	22%	5%	47%	100%
Technical skill					
Unskilled	29%	18%	2%	51%	100%
Elementary worker	29%	17%	4%	49%	100%
Technical worker	27%	22%	7%	43%	100%
College and above	17%	25%	3%	56%	100%
Total	25%	22%	5%	47%	100%
Social capital					
No network	28%	21%	6%	45%	100%
Have network	24%	23%	5%	48%	100%
Total	25%	22%	5%	47%	100%
2. Characteristics of firms					
Firm age	18.50	9.02	40.84	10.97	14.06
Electronic access					
No electronic access	30%	20%	7%	43%	100%
Electronic access	17%	26%	3%	54%	100%
Total	25%	22%	5%	47%	100%
3. Firm performance and dynamics					
Informal-formal transition					
Stayed as informal	31%	18%	7%	44%	100%
Formal	15%	28%	4%	53%	100%
Moving from informal to formal	23%	19%	0%	58%	100%
Moving from formal to informal	26%	33%	3%	38%	100%
Total	25%	22%	5%	47%	100%
Expansion and innovation in 2009					
No expansion	29%	24%	8%	39%	100%
Expansion	22%	21%	4%	53%	100%

Total	25%	22%	5%	47%	100%
Expansion and innovation in 2011					
No expansion	29%	20%	7%	43%	100%
Expansion	22%	24%	4%	50%	100%
Total	25%	22%	5%	47%	100%
Efficiency					
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%
other	22%	0%	22%	56%	100%
Total	25%	22%	5%	47%	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%
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%
Others	11%	26%	21%	42%	100%
Total	25%	22%	5%	47%	100%
Financial assistance in 2009					
No financial assistance in 2009	25%	23%	5%	46%	100%
Financial assistance in 2009	24%	20%	6%	50%	100%
Total	25%	22%	5%	47%	100%
Financial assistance in 2011					
No financial assistance in 2011	25%	22%	5%	47%	100%
Financial assistance in 2011	21%	24%	6%	48%	100%
Total	25%	22%	5%	47%	100%

Technical assistance in 2009					
No technical assistance in 2009	25%	22%	6%	47%	100%
Technical assistance in 2009	23%	35%	0%	42%	100%
Total	25%	22%	5%	47%	100%
Technical assistance in 2011					
No technical assistance in 2011	25%	22%	6%	47%	100%
Technical assistance in 2011	13%	38%	2%	48%	100%
Total	25%	22%	5%	47%	100%
Being inspected in 2009					
Not inspected	29%	17%	7%	47%	100%
Inspected	22%	26%	5%	47%	100%
Total	25%	22%	5%	47%	100%
Being inspected in 2011					
Not inspected	26%	22%	6%	47%	100%
Inspected	14%	29%	4%	52%	100%
Total	25%	22%	5%	47%	100%
% of management time dealing with government regulations in 2009	0.92%	1.35%	0.92%	1.13%	1.12%
% of management time dealing with government regulations in 2011	1.75%	3.15%	1.82%	2.36%	2.36%
N	441	392	97	835	1,765

APPENDIX 2. MCP COORDINATE PLOTS
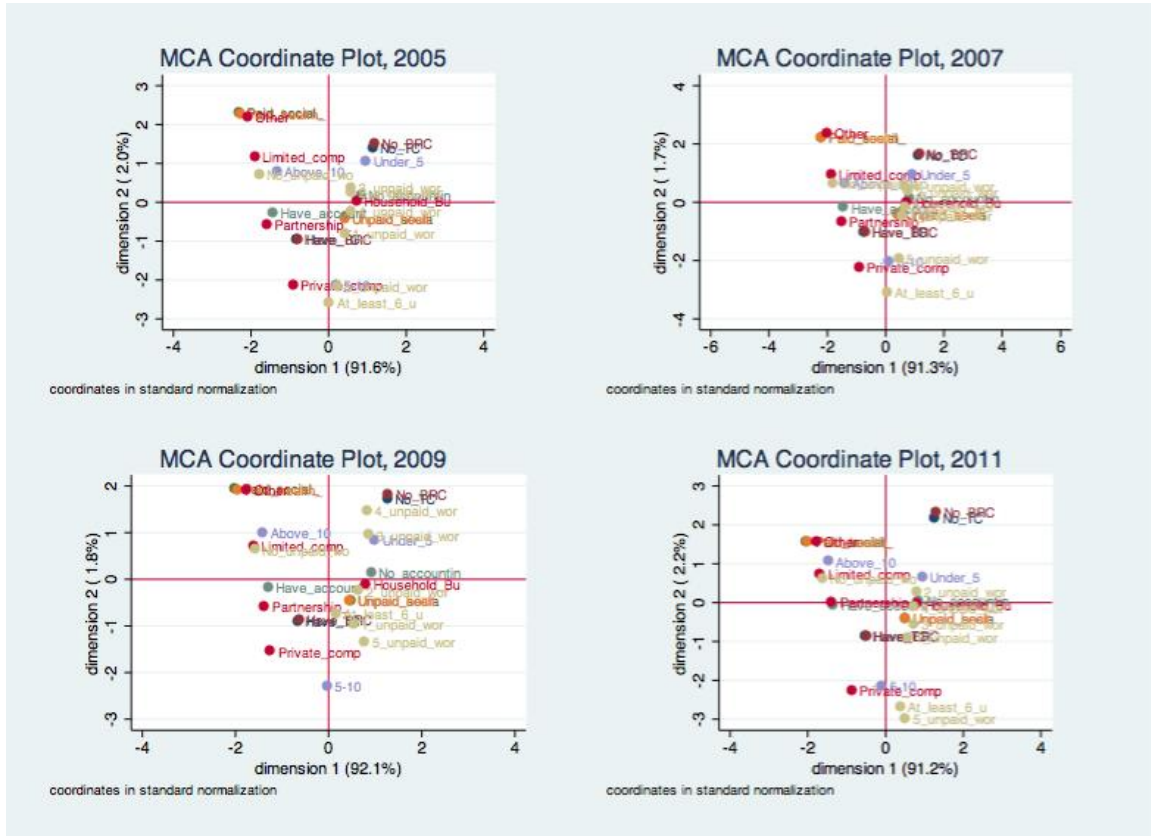
Source: SME 2005, 2007, 2009, 2011

APPENDIX 3. STATISTICS FOR COLUMN CATEGORIES IN STANDARD NORMALIZATION

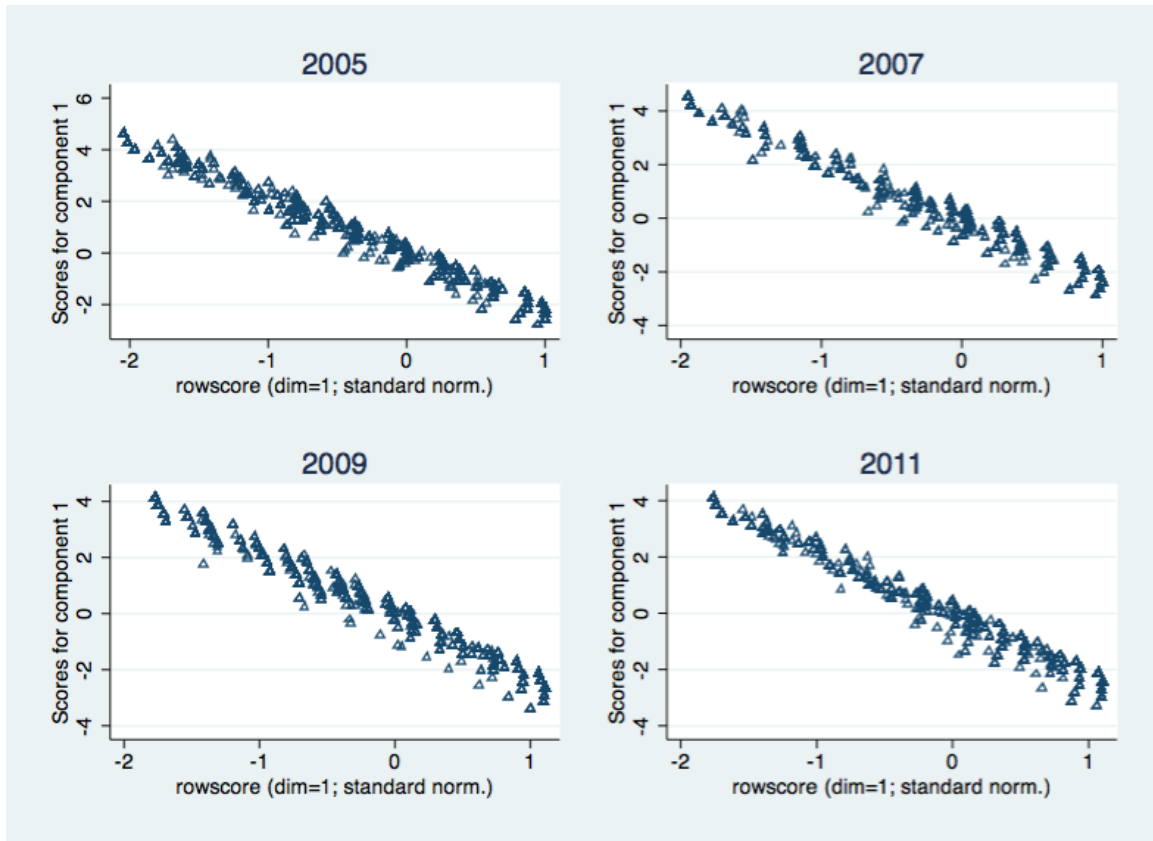|	Overall			Dimension 1			Dimension 2		
Categories	Mass	Quality	% Inertia	Coord-inate	Square corre-lation	Contri-bution	Coord-inate	Square corre-lation	Contri-bution
BRC									
No	0.036	0.877	6.64	1.282	0.820	0.060	2.158	0.057	0.169
Yes	0.089	0.877	2.72	-0.526	0.820	0.024	-0.885	0.057	0.069
TC									
No	0.034	0.873	6.61	1.321	0.813	0.059	2.308	0.060	0.180
Yes	0.091	0.873	2.44	-0.489	0.813	0.022	-0.854	0.060	0.067
Social insurance									
No	0.099	0.930	2.66	0.519	0.917	0.027	-0.404	0.014	0.016
Yes	0.026	0.930	10.33	-2.013	0.917	0.104	1.567	0.014	0.063
Health insurance									
No	0.099	0.930	2.71	0.526	0.917	0.027	-0.410	0.014	0.017
Yes	0.026	0.930	10.23	-1.981	0.917	0.103	1.547	0.014	0.063
Accounting book									
No	0.076	0.967	5.36	0.866	0.967	0.057	0.056	0.000	0.000
Yes	0.049	0.967	8.28	-1.337	0.967	0.088	-0.086	0.000	0.000
Ownership									
HB	0.080	0.947	5.25	0.824	0.947	0.055	-0.054	0.000	0.000
Private	0.010	0.888	0.90	-0.866	0.759	0.007	-2.284	0.129	0.052
Partnersh/Collective	0.003	0.885	0.67	-1.368	0.885	0.007	0.011	0.000	0.000
Limited	0.026	0.937	7.12	-1.672	0.933	0.073	0.723	0.004	0.014
Other	0.005	0.935	1.64	-1.784	0.918	0.017	1.569	0.017	0.013
Number of full-time workers									
< 5	0.057	0.990	4.77	0.950	0.978	0.051	0.661	0.012	0.025
5 - 10	0.034	0.772	0.49	-0.090	0.051	0.000	-2.173	0.721	0.160
> 10	0.034	0.990	7.02	-1.480	0.978	0.075	1.047	0.012	0.038
Unpaid workers									
0	0.038	0.948	9.66	-1.620	0.944	0.100	0.598	0.003	0.014
1	0.028	0.882	1.00	0.577	0.833	0.009	-0.893	0.049	0.022
2	0.045	0.966	2.67	0.796	0.963	0.028	0.268	0.003	0.003
3	0.009	0.947	0.48	0.737	0.934	0.005	-0.550	0.013	0.003
4	0.004	0.929	0.21	0.742	0.928	0.002	-0.131	0.001	0.000
5	0.001	0.600	0.08	0.519	0.331	0.000	-2.994	0.268	0.009
6	0.001	0.365	0.05	0.380	0.163	0.000	-2.700	0.201	0.004

Source: SME 2011

APPENDIX 4. INERTIA 1
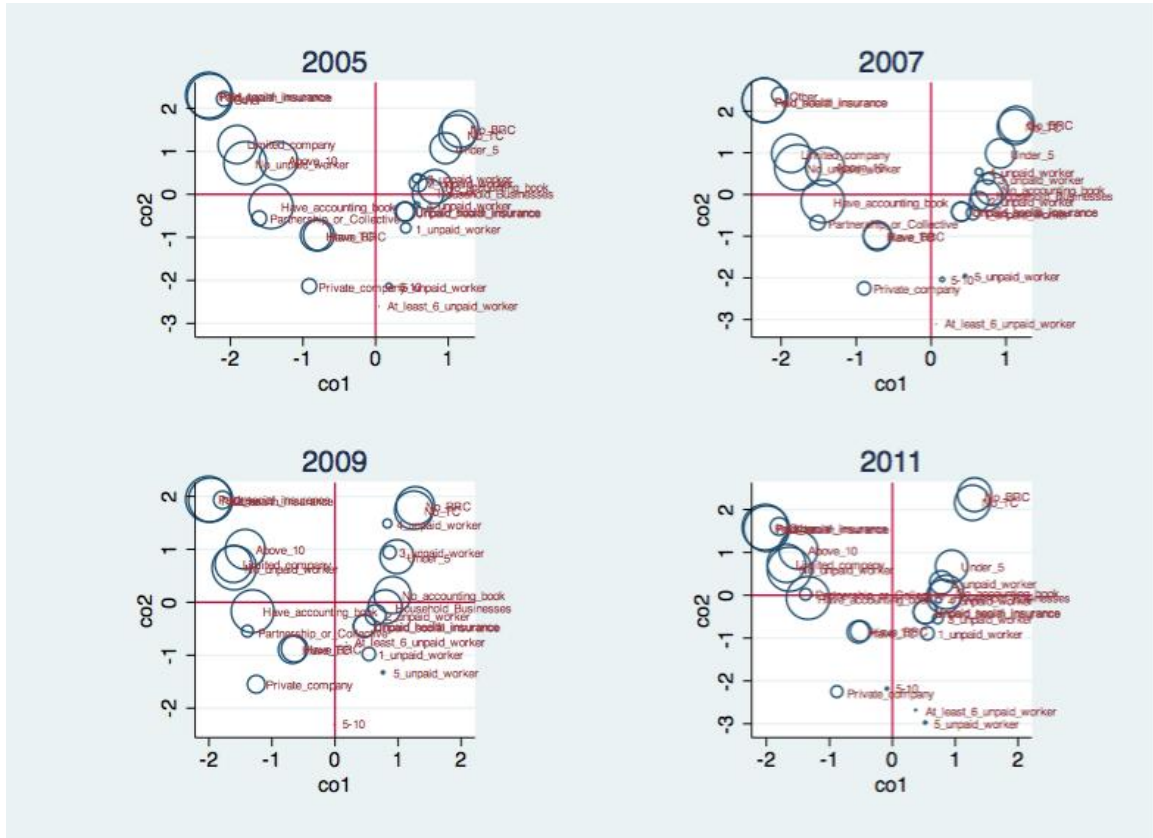
Source: SME 2005, 2007, 2009, 2011

APPENDIX 5. PCA AND MCA

Source: SME 2005, 2007, 2009, 2011