The Distribution of Firm Size in Africa’s Manufacturing Sector and Its Implication for Industrial Policy

Janvier D. Nkurunziza

Abstract

This paper analyzes the distribution of Kenyan firms into four size categories: micro, small, medium and large size. Four questions are investigated. First, do small firms grow faster than large ones or is firm growth independent of size as stipulated by Gibrat’s “Law of Proportionate Effect” Second, what is the steady state size of firms? Third, how long does it take to reach the steady state size? Fourth, does the use of bank credit affect a firm’s growth and the process of firm size convergence? On the basis of ergodic probability distributions, we derive information on firms’ steady state growth and convergence. Using data on Kenya’s manufacturing sector, empirical results suggest that firm growth is associated with overall economic performance. Before the early 1990s, firms had a high growth potential. In equilibrium, 70 percent of surviving firms converged to large size. In contrast, growth and convergence in the 1990s reflected the economic crisis that hit the Kenyan economy: the steady state distribution of firm size was concentrated in the first two quartiles, with suggestive evidence that smaller firms recorded the highest rates of failure. Also, the use of credit increased the growth of surviving firms but appears to have contributed to precipitating firm failure. These results suggest that support policies such as special credit schemes are not a panacea for a firm’s survival and growth.

Keywords  Firm size, Ergodic distribution, Kenya

JEL Classification  D22; D4; 055

Introduction

Small and medium-scale enterprises (SME) represent the largest share of the stock of firms in the world. In Africa, however, the economic landscape is dominated by firms in the lowest
extreme of the size distribution, which are informal firms and microenterprises. Indeed, survey evidence shows that the majority of SMEs in Africa are made of just one person and the bulk of the remainder employ less than 10 workers (Mead and Liedholm, 1998). These micro firms are less productive in terms of output per worker than those which are larger (even those with two persons), suggesting that very small size is a severe handicap to firm performance. In this regard, knowing the growth potential of firms of different size helps to understand the pattern of industrial development in Africa.

Four questions are investigated. First, do small firms grow faster than large ones or is firm growth independent of size as stipulated by Gibrat’s “Law of Proportionate Effect--LPE” (Gibrat, 1931)? Second, what is the steady state size of firms? Third, how long does it take to reach the steady state size? Fourth, does the use of bank credit affect a firm’s growth and the process of firm size convergence? This last question has important policy implications given that easing credit constraint for SMEs has been one of the key elements of programmes of assistance to SMEs both in Africa and elsewhere (Manu, 1998).

Understanding the pattern of firm growth, particularly the growth differences across the size distribution, is important for many other reasons. For example, there have been initiatives taken to encourage the growth of small firms, including the establishment of special credit schemes. Whether or not such initiatives have been successful in helping these firms to grow is often not known. Moreover, the dominance of an industry by small firms has been associated with backwardness, a problem that calls for government intervention (Lucas, 1978). Knowing the growth pattern of firms within an industry can inform such government policy. Furthermore, in many societies, firm size mirrors the pattern of income distribution and social status (Nugent, 1996). It is normally the poor who own small firms as they lack the required level of education to work as salaried employees and do not have enough resources to establish viable businesses. As a result, constraints to firm growth may result in social immobility.1

Studies on firm growth in Africa including Teal (1999), Aguilar and Kimuyu (2002), McCormick et al (1997), Mead and Liedholm (1998), and Nkurunziza (2010) have used regression analysis to identify the key determinants of growth. Despite its popularity, regression analysis is not suited to answer the question of interest in this paper, which is to know how the groups of micro and small firms, which dominate the continent’s productive sector, grow relative to other size groups. This paper addresses this question by deriving equilibrium distributions of firm size using a Markov chain that models the dynamics of size distributions.2

Using firm level data on the Kenyan manufacturing sector, firms are grouped in size quartiles representing micro, small, medium and large firms. The data used were collected in the framework of the World Bank’s Regional Programme on Enterprise Development (RPED) Project. The data were collected for the years 1992, 1993, 1994 and 1999.3 More than 200 firms were surveyed in Kenya in each wave (see details later) covering four regions namely Nairobi (about 60 percent of the sample), Mombasa (about 20 percent), Eldoret (about 10 percent) and Nakuru (about 10 percent). Firms were stratified into four size groups and four industries.4

1 The main assumption underlying the analysis in this paper is that firms grow in size when they are prosperous and downsize or exit when they face difficulties. However, Lucas (1978) rightly argues that some small firms may not be willing to grow if their managers realise expected returns to their level of managerial ability. On the other hand, large firms may not be ‘willing’ to downsize and revert to their optimal size as predicted by the classical Marshall-Viner competitive model.

2 The terms Markov chain and Markov process are used interchangeably.

3 The 1999 wave of data collection was funded by UNIDO and carried out by the Centre for the Study of African Economies, University of Oxford, in collaboration with other academic institutions. The author of this paper participated in this wave, leading the teams that worked in Nairobi and Mombasa. Other teams collected data in Nairobi, Nakuru and Eldoret.

4 For more details on the sampling procedure used to collect this dataset and other associated aspects, see Aguilar and Bigsten (2002).
Micro firms were firms having up to 5 full time workers. Small firms were those with six to twenty-five workers. Medium firms had between twenty-six and one hundred workers while large firms were those with more than 100 full time workers. The four sectors were textile, food, metal and woodwork industries. Sampling drew more or less an equal number of firms in each size group and each industry. In each subsequent survey, the original firms were tracked in order to constitute a panel. Information was collected about the status of firms that were no more at their original addresses; in most cases, they had closed down. This information was used to identify the firms that collapsed and determine the rate of attrition.

The availability of the RPED database has led to the publication of dozens of articles covering a range of issues such as investment [Bigsten, et al. (1999)]; inventory holding as a risk coping mechanism [Fafchamps, et al. (2000)]; firm survival [Harding, et al. (2004); Nkurunziza (2012)]; firm growth and productivity [Teal (1999); Nkurunziza (2010)]; credit constraints [Bigsten, et al. (2003)]; contract flexibility and dispute resolution [Bigsten et al (2000)]; and trade credit [Fafchamps (1997)], among others. This firm-level survey database provides a unique source of information on African manufacturing. RPED data collection efforts remain the best attempt ever made to collect comparable and detailed firm level data on Africa’s manufacturing sector in a multi-country setting.

Growth is measured by the proportions of firms moving across quartiles over a period of time. These proportions, also interpreted as probabilities, are contained in a transition matrix which summarizes the pattern of short-term growth. Assuming that the transition matrix is homogeneous (see section 3 for details), long-term growth is computed on the basis of ergodic probabilities which show the equilibrium distribution of firms in the different size groups. The speed of convergence to steady state is measured by the asymptotic half-life of convergence computed using the second highest Frobenius-Perron’s eigenvalue. Moreover, the Markovian framework allows analysis of firm growth, entry and exit in a unified framework, addressing the problem of self-selection often overlooked in growth regressions (Nkurunziza, 2010).

One limitation of this approach is that it can only allow for a limited number of conditioning variables, particularly given the relatively small sample we use in this paper. As discussed under Section 3.1.2, a bivariate analysis with the four-state transition matrix implies solving a system of 16 linear equations while analysis conditioning firm mobility on two variables would imply solving a system of 64 linear equations. Despite this limitation, the Markovian approach is appropriate for analysing the dynamics of firm size distributions. To the best of our knowledge, this study is the first applying this methodology to model firm size convergence in an African context.

The empirical analysis has two parts. The first focuses on mobility of firms from their creation to 1992, the first year when they were surveyed, noting that firms were created at different times. In the second part, the analysis is restricted to the period from 1992 to 1999, where equal observation periods are used for every surviving firm from the original 1992 sample. Restricting the analysis to the group of firms surveyed in 1992 and observing them throughout the 1990s makes it possible to integrate the impact of firm exit in the analysis.

The paper proceeds as follows. In the second section, the paper briefly reviews the macroeconomic context in Kenya before and during the 1990s and why it mattered for firm performance. The third section develops the methodology and discusses the empirical data used to model convergence of the distribution of firm size. Section 4 presents empirical results for the pre-1992 and the post-1992 period. Section 5 concludes.

5 The 4 groups have been widely used in empirical research using RPED data even though the definition of their boundaries is arbitrary (e.g. Bigsten, et al., 1999; 2000; 2003).
Macroeconomic shocks and manufacturing firms in Kenya

Macroeconomic trends

In the 1960s and 1970s, Kenya had a robust economy. Growing at an average annual rate of 6.5 percent in the first decade following independence in 1963, Kenya was among the fastest growing economies in Africa. In 1974, Kenya was hit by the oil shock that ravaged economies throughout the world. As a result, the cost of Kenyan oil imports quadrupled, while the price of other imports increased by 30 percent. Hence, in 1975, the terms of trade declined by 23 percent. The country also suffered from a drought that substantially decreased agricultural production in 1974 and 1975.

The doubling of coffee prices between 1975 and 1976 provided unexpected revenue to government, which eased the pressure of the oil shock. The coffee boom changed the country’s macroeconomic management from a tight policy in the aftermath of the drought and the oil shock to a lax policy. This change had dramatic consequences on the country’s economy. Kenya adopted expansionary policies and the control of the Treasury over spending broke down (Bevan, et al., 1990). As the coffee boom came to an end in 1978, it became clear that Kenya needed to change its policies. In the 1980s, the economy remained weak in the first half but recovered in the second half. As a whole, however, the performance remained disappointing with an average growth rate of 0.6 percent. The 1990s followed more or less the same pattern. In the early 1990s, GDP growth rates were close to zero and the overall average annual GDP growth for the decade was -0.73 percent. In the 2000s, between 2000 and 2008, the annual rate of GDP growth averaged almost 4 percent (World Bank, 2009).

Kenya’s economic problems were rooted in internal policies pursued by the government starting from the 1970s (Njeru and Randa, 2001). They included the poor management of the coffee windfall in the second half of the 1970s and the widespread economic controls put in place in the 1980s as a way of addressing the mistakes made during the boom years. There were controls on foreign exchange allocation, consumer and producer prices, the setting of interest rates, etc (Bevan, et al., 1990). The economy was also overburdened by loss making state firms that had been kept afloat by large transfers of state funds. It is only in the 1990s that Kenya resolutely implemented liberalization policies in a way that sent a shockwave across the economy. The country’s economic reform programme included liberalisation measures that resulted in high inflation, large devaluations and high interest rates, as figure 1 illustrates.

Figure 1 shows the trends of four key variables, namely the rates of inflation and devaluation, and the lending and deposit interest rates. Inflation increased markedly from the mid-1980s to the mid-1990s, reaching 46 percent in 1993. Over almost four decades from 1970 to 2008, inflation fell below 5 percent in only 5 years and it fell below 8 percent in 10 years. These two thresholds are considered by the World Bank-IMF programmes and Collier and Gunning (1999), respectively, as the rates above which inflation is harmful to growth. Hence, inflation in Kenya remained well above the two benchmarks for most of the period.

Figure 1 also shows that relative to the American dollar, the Kenyan shilling depreciated continuously from the late 1980s. The peak was in 1993 when the currency lost 80 percent of its

---

6 For a detailed analysis of Kenya’s shocks and their impact on the performance of the country’s economy, see Bevan, et al. (1990).
7 On a detailed discussion of the early impact of macroeconomic liberalisation and its episodes, see Adam (1992) and Reinikka (1993).
8 Inflation is defined as the percentage change in consumer prices. The rate of devaluation is the percentage change in the amount of Kenyan shillings required to buy one American dollar.
9 Others such as Bruno and Easterly (1998) and Pollin and Andong (2005) suggest even higher rates.
value in just one year. Between March and May 1993, the Kenyan shilling was devalued three times: by 25 percent on 9 March, 31 percent on 20 April and by 6 percent in May. Exchange rate liberalisation led to instability of the Shilling due to speculation in foreign currency trading. Between the 1980s and 1990s, the average exchange rate increased from 14.25 to 49.50 Kenyan shilling to the dollar, implying 245 percent devaluation. In 2001, the exchange rate had reached 79 Kenyan shilling per dollar but, by 2008, the Kenyan currency had appreciated to 69 shillings per dollar (World Bank, 2009).

One of the measures taken to liberalise the financial sector was the decision by the Capital Market Development Authority (CMDA) to promote secondary trading in government bonds and equity. As a result, interest rates on bonds increased from an average of about 12 percent in the 1980s to 22 percent in the 1990s, translating into an 83 percent increase. In 1993 alone, the interest rate on bonds jumped from 16.5 to 50 percent, a staggering 200 percent increase in one year. Lending rates practiced by commercial banks followed the same pattern. Although the controlled interest rates had been regularly revised upwards starting from the mid-1980s, they were not fully liberalised until July 1991. The lending rate peaked in 1994 at 36 percent. However, the most dramatic change was observed in 1993 when the lending interest rates increased by 42 percent from an already high level of 21 percent in 1992. Lending interest rates remained above 20 percent until 2000. These policy shocks created high macroeconomic uncertainty, with negative consequences on the activities of firms.

Macroeconomic shocks and manufacturing firm performance

The policy environment was an important determinant of SMEs performance in Kenya. Indeed, some of the liberalization policies adopted in the early 1990s, including exchange rate and interest rate policies, were incompatible with the government’s objective to encourage production, including by manufacturing firms. Macroeconomic shocks were transmitted to
the manufacturing sector through different channels. First, as most manufacturing firms produced for the domestic market, overall economic contraction resulted in low domestic demand of manufacturing output, limiting SMEs activities. During our field research, we noted that most firms were using only a part of their total production capacity, corroborating firm managers’ view that low demand was one of the most important constraints they were facing. This led to low investment and low firm growth. On average, the firms in our sample had an average growth rate of 7.0 percent per annum over the period before 1992, this being the year when these firms were interviewed for the first time. During the period between 1992 and 1999, the average annual rate of growth of firms was -1.3 percent.

Second, in a climate of economic instability and uncertainty due to devaluations and high inflation, suppliers not only required cash payment but also quoted prices in dollars instead of Kenyan shillings (Bigsten, 2002). Hence, for firms relying on imports of raw material, the depreciation of the Kenyan shilling increased their production costs, reducing firm profitability. In the same connection, firms financing imports by trade credit were unable to predict how much foreign currency they would need to pay when their letters of credit are called (Ndung’u, 1998). In theory, the depreciation of the Kenyan shilling could have led to more exports from manufacturing firms but there is no evidence that Kenyan firms increased their exports during this period, as most of them were producing for the local market (Bigsten, et al., 1999a).

Third, high interest rate increases meant that rates of returns on projects needed to increase even more for investment projects to be bankable. That was difficult in a context of sluggish economic activity. As a result, bank lending was curtailed. Fourth, the sudden increases in interest rates increased the debt burden of firms which were already in debt. Subsequently, some of them were unable to reimburse their loans so they either downsized or simply collapsed as empirical results in section 4 will show. Fifth, the high interest rates paid on treasury bills induced firms with financial reserves to invest in bonds rather than their production capacity. Therefore, the issuance of bonds by the government to finance its large fiscal deficits diverted economic agents’ attention from investment in the real sector to trade in financial assets.

In addition to the destabilising impact of policy changes, a drought ravaged the country in 1990-91, resulting in increased domestic food prices and high inflation. In one year, food prices more than doubled. Expectedly drought had an impact on firms in the food industry that suffered from an input scarcity. In this light, information collected through interviews of manufacturing firm managers in the 1990s (see Table 1) shows that three out of four key problems they faced are directly related to the developments discussed above.

### Table 1 Managers’ Ranking of Three Most Important Problems.

<table>
<thead>
<tr>
<th>Years</th>
<th>Problem No. 1</th>
<th>Problem No. 2</th>
<th>Problem No. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>Credit</td>
<td>Foreign exchange</td>
<td>Low demand</td>
</tr>
<tr>
<td>1993</td>
<td>Credit</td>
<td>Low demand</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>1994</td>
<td>Credit</td>
<td>Low demand</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>1999</td>
<td>Low demand</td>
<td>Credit</td>
<td>Infrastructure</td>
</tr>
</tbody>
</table>

Source Regional Programme on Enterprise Development (RPED) database.

10 Mead and Liedholm (1998) discuss in detail the link between macroeconomic variables and the performance of small and medium firms in African countries.

11 Many firms surveyed in 1992 and 1999 communicated data on start-up size but the 1993 and 1994 waves did not collect this information. Therefore, the analysis using start-up size is restricted to the 1992 sample and the firms interviewed for the first time in 1999.

12 Table 2 in Bigsten, et al. (1999) shows that in comparison with Cameroon, Ghana and Zimbabwe, Kenya had the highest real interest rate in 1995 at 24 percent.
According to firm managers, credit was the most important problem faced by firms. It came out four times, thrice as the most important problem and once as the second most important problem. Low demand came out four times, once as the most important problem, twice as the second most important problem and once as the third most important problem. A survey of small firms in six Southern African countries notes that business managers cite the same two problems, namely lack of credit and low demand, as the most important factors explaining business closures (Mead and Liedholm, 1998). On the basis of the discussions above, low demand due to the economic crisis and the changes in the credit market may have had a long term effect on firm growth and convergence. Empirical results in section 4 will attempt to confirm this hypothesis.

**Methodology and empirical data**

Transition analysis is used to derive the distribution of firm size into four categories both in the short term and long term. Owing to the focus of the original application of transition analysis to intergenerational mobility, most studies have been concerned with unconditional mobility or the position of the Markov chain given its position in the previous period. For instance, studies look at the social position of a person given that of his parents. By design, these are long-term studies as the time unit of analysis is a generation. This paper analyses a different phenomenon. Firms rather than individuals or households are the primitive unit of observation and the empirical data used covers a short period relative to the studies of intergenerational mobility. Moreover, the objective of this paper is different as it deals with both short term and steady state distributions of firm size. Furthermore, given the importance of credit for firm performance (see Table 1 and Nkurunziza, 2010), part of the empirical analysis conditions convergence on the use of credit. The first subsection develops a detailed empirical methodology for the measurement of convergence. The second subsection discusses the empirical data used.

**Modelling Firm Convergence with a Markov Chain**

Does a Markov process characterise the pattern of firm growth across different size categories? The answer is yes, as long as the following conditional probability holds:

\[ P[X(t+1) = x_{t+1} | (X(1) = x_1) \cap (X(2) = x_2) \cap \ldots \cap (X(t) = x_t)] \]

This formulation simply states that the probability that a variable \( X \) has a value of \( x_{t+1} \) in period \( t+1 \) given the entire path of the variable depends only on the value of the variable at period \( t \). In the literature, the word ‘state’ is used to characterise the ‘position’ of a Markov process. Using the formulation in (1), we have: \( P[X(t+1) = j | X(t) = i] \) as the probability that the process is in state \( j \) at step (or period) \( t+1 \) given that it was in state \( i \) at the previous step (or period). In a discrete Markov chain, this formulation allows indexation of different states by \( 0,1,2, \ldots, N \) where \( N < \infty \). A period can be any time unit, depending on the dynamic behaviour of the process under analysis. For instance, studies of intergenerational income dynamics use a generation as the time unit while studies of cross-country growth usually use five or ten year periods as time units.
Definitions and Analytical Set-up

We adopt the following concepts, assumptions and notations. First, mobility from size group $i$ to size group $j$ between periods $t$ and $t+1$ follows a first-order Markov chain. To characterise a Markov process, two sets of information are needed (Benjamin and Cornell, 1970). First is information about ‘initial conditions’. What is the distribution over the initial states? In probabilistic terms this is given by:

$$q_i = P[X(0) = i], \forall i \in \{A\}$$

(2)

where $\{A\}$ is the state space. The second set of information concerns transition probabilities $p_{ij}$ describing the transition from one state to another. This is:

$$p_{ij} = (t + 1) = P[X(t + 1) = j | X(t) = i]$$

(3)

A Markov matrix also called stochastic or transition matrix, is the matrix of transition probabilities. It is conveniently represented as a $n \times n$ matrix, where

- $n$ is the number of states. Given four size groups, the stochastic matrix $P$ characterizing transition between any two periods is given by:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix}$$

(4)

where $p_{11}$ is the proportion of micro firms in period $t$ that remain micro in the next period; $p_{12}$ is the proportion of micro firms that grow from micro to small firms between the two periods; and $p_{21}$ is the proportion of small firms that become micro firms, and so on. The transition matrix $P$ is nonnegative as each transition probability $p_{ij} \geq 0$ and:

$$0 \leq p_{ij} \leq 1$$

(5)

As probabilities, elements in each row must sum to 1.

$$\sum_{j=1}^{n} p_{ij} = 1$$

(6)

If transition is analysed over more than one period ahead, the transition probability becomes:

$$p_{ij}^{(m)} = P[X(t + m) = j | X(t) = i]$$

(7)

where $m$ is the number of steps. Transition probabilities may or may not be a function of time. When they are not, the process they describe is said to be ‘homogeneous’ in time such that $p_{ij}(t+1) = p_{ij}(t) = p_{ij}$. When this is the case, the $m$-step transition matrix is the product of the matrix itself $m$-times. For non-homogeneous chains, the $m$-step transition probabilities must take into account the changes in transition between the initial and the final period. Therefore, the $m$-step probabilities are:

$$p_{ij}^{t+m}(t) = \sum_{k=1}^{r} p_{ik}^{t+m-1}(t) p_{kj}(t+m); \text{ where } p_{ik} \neq p_{kj}$$

(8)
If all the states communicate among them, or if \( i \leftrightarrow j: \forall i, j; \) then the Markov process is said to be ‘irreducible’. This is an important property as it gives an indication of the sensitivity of the Markov chain to initial conditions (Robert and Casella, 1999).

A state \( i \) that does not allow mobility out of it is called a ‘trapping state’. For instance, the group of firms that close down constitutes a trapping state: failing firms cannot transition to other size groups. A state with a unitary transition probability is called ‘absorbing’ state.

It is clear that Markov chains with trapping or absorbing states are not irreducible. If \( g_i \) is the probability that a process starting from state \( i \) eventually re-enters the state at some time, then state \( i \) is said to be ‘recurrent’ if \( g_i = 1 \); otherwise, if \( g_i < 1 \), the state is called ‘transient’. With a transient state, the number of times that the process will be in state \( i \) is geometrically distributed with a finite mean equal to \( 1/(1 - g_i) \). With a recurrent state, the expected number of times a Markov process will be in state \( i \) is infinite. Such a process is called positive recurrent as the time between two passages in the same state is finite. All states in a finite irreducible Markov process are recurrent (see Ross, 2003). With irreducible processes, recurrence and transience are properties of the process not of particular states (Robert and Casella, 1999). Finally, a chain is ‘periodic’ if it displays a cyclical pattern. In other words, the chain’s visits to a specific state are separated by a number of steps equal to a multiple of some integer greater than one (Luenberger, 1979) otherwise it is ‘aperiodic’.

In order to have meaningful empirical comparisons across time, it is important to study growth and convergence based on the same size groups defined in the initial period (equation 2). For instance, if initial discretization defines \( n \) groups with \( z_{i0} \) boundaries, all subsequent groups must have the same \( z_{i0} = z_{i1} = z_{i2} = \ldots = z_{iT} \) state boundaries (section 4 provides empirical results).

Although the analysis of transition probabilities provides interesting insights into the short-term distribution of firm size, it gives no information on the distribution in steady state. Thanks to the stability of ergodic probability distributions, it is possible to draw conclusions on the steady state characteristics of the distribution derived from the transition matrix. The next subsection focuses on the derivation of ergodic probability distributions.

Convergence, Ergodicity and Steady State Probability Distributions

An important question when analysing mobility is to know the limiting behaviour of a Markov chain or its probabilistic steady state characteristics. These are given by ergodic probabilities. Central to this analysis is the Frobenius-Perron theorem and its associated Basic Limit Theorem for Markov Chains. The theorem states that all positive finite Markov chains have a unitary dominant eigenvalue also called Frobenius-Perron eigenvalue. After a large number of steps, such a chain converges to a unique and stable probability vector that is independent of initial conditions. The theory of linear homogeneous systems shows that the Frobenius-Perron eigenvalue and its associated eigenvector determine the behaviour of a system in steady state.

Convergence of a transition matrix towards this stable long-run probability distribution is called ‘ergodicity’. Hence, ‘ergodic distribution’ refers to the steady state or limiting probability distribution of a dynamic system. Borovkov (1998: ii) defines ergodicity as

13 In a three-state Markov chain where the states are \( i, j, k \), irreducibility requires that \( i \leftrightarrow j \text{ and } j \leftrightarrow k \) as communication between \( i \) and \( j \) guarantees communication between \( i \) and \( k \) (Ross, 2003: 190).
14 Robert and Casella (1999) provide a formal definition. Examples of periodic processes include the sine and cosine trigonometric functions (see Chiang, 1984: 516-517).
15 See Luenberger (1979) for a formal statement and proof of the theorem.
16 The term ‘Ergodicity’ originates from statistical mechanics where it is used in the theory of dynamic systems.
“convergence, in large time limit, of a distribution generated by a stochastic process to some limiting distribution that is stationary and independent of the initial conditions”.

17 In view of the relatively small sample used to analyze mobility and derive ergodic probabilities, a discrete modelling framework is adopted. A Markov process is ergodic if:

\[
\pi_j = \lim_{t \to \infty} p_{ij}; \forall j \geq 0
\]

(9)

where \(\pi_j\) is the invariant marginal probability distribution over the state space. Ross (2003) shows in his Theorem 4.1 and thereafter that the existence and uniqueness of \(\pi_j\) depends on two main conditions. First, the Markov chain must be irreducible ergodic. 18 This condition is satisfied in all of the matrices we analyse below but one. A weak sufficient condition for convergence is that all the elements in the main diagonal of a transition matrix are non-zero (Shorrocks, 1976: 567). Secondly, the matrix \(\pi_j\) must solve the following system of equations:

\[
\begin{align*}
\pi_j &= \sum_{i=0}^{n} \pi_i p_{ij}; \forall j \geq 0 \\
\sum_{j=0}^{n} \pi_j &= 1
\end{align*}
\]

(10)

The limiting probability is also interpreted as the long-run proportion of times that the process will be in state \(j\).

Equation (10) provides a convenient way of deriving ergodic probabilities. The information needed is just a homogeneous stochastic matrix. In order to clarify the relationships that produce ergodic probabilities, the system of equations (10) may be expanded in the form of a simultaneous equations system as follow

\[
\begin{align*}
\pi_1^* &= \pi_1^* p_{11} + \pi_2^* p_{21} + \pi_3^* p_{31} + \pi_4^* p_{41} \\
\pi_2^* &= \pi_1^* p_{12} + \pi_2^* p_{22} + \pi_3^* p_{32} + \pi_4^* p_{42} \\
\pi_3^* &= \pi_1^* p_{13} + \pi_2^* p_{23} + \pi_3^* p_{33} + \pi_4^* p_{43} \\
\pi_4^* &= \pi_1^* p_{14} + \pi_2^* p_{24} + \pi_3^* p_{34} + \pi_4^* p_{44}
\end{align*}
\]

(11a)

and

\[
\pi_1^* + \pi_2^* + \pi_3^* + \pi_4^* = 1
\]

(11b)

Where \(\pi^*\) represents the ergodic probability. Re-writing equation (11a) as a homogeneous system:

\[
\begin{pmatrix}
p_{ij}
\end{pmatrix}_{n \times n} \left( I - \begin{pmatrix}
p_{ij}
\end{pmatrix} \right) \pi_j = 0
\]

(12)

The problem with (12) is that it has multiple solutions. To ensure uniqueness of the solution, one equation in (11a) must be dropped and replaced by the constraint in equation (11b). Dropping the fourth equation in (11a), the system of equations becomes:

17 Note the presence of the terms convergence, limiting distribution, stationary, and independent of initial conditions in the definition. A stationary Markov chain is one where the transition probabilities are independent of the number of past periods (Luenberger, 1979: 225). Transition probabilities are stationary when, given an initial distribution over the states equal to \(\pi\), the probability of being in any state at any time is also \(\pi\), implying that \(X_t \sim \pi\) if \(X_t \sim \pi\).

18 Ergodic states are states that are positive recurrent and aperiodic in a finite Markov chain.
which can be compactly written as: \( P \times \Pi = d \). Finally, provided \( P \) satisfies the non-singularity condition, the vector of ergodic probabilities is obtained by inverting the system in (13).

\[
\begin{pmatrix}
(p_{11} - 1) & p_{21} & p_{31} & p_{41} \\
 p_{12} & (p_{22} - 1) & p_{32} & p_{42} \\
 p_{13} & p_{23} & (p_{33} - 1) & p_{43} \\
 1 & 1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
\pi_1 \\
\pi_2 \\
\pi_3 \\
\pi_4
\end{pmatrix}
= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{13}
\]

It should be clear that the computations become more complex the higher the number of states and the higher the dimension of the analysis. For instance, a bivariate analysis in a four-state stochastic matrix requires solving a system of sixteen linear equations while a three-variable system requires solving 64 linear equations. If size groups are defined as deciles rather than quartiles, the number of equations to deal with increases from 10 for a univariate analysis to 1000 for a three-variable system.

The speed of convergence to the steady state is computed on the basis of the asymptotic half-life of convergence.

\[
h = - \frac{\ln 2}{\ln |\lambda_2|} \tag{15}
\]

where \( \lambda_2 \) is the second highest eigenvalue of the transition probability matrix. The value of \( h \) is interpreted as the number of periods required to halve the norm of the difference between the current and ergodic distributions.

Summary of the Main Assumptions

Before presenting empirical results, it is useful to recapitulate the foregoing discussions to show how the concepts presented relate to the problem we set out to analyse.

**Markov assumption**: Is the Markovian assumption appropriate? The answer is yes. Firm size at any given time reflects its own history. Hence, the use of a first-order Markov chain assumption is theoretically plausible.

**Homogeneity**: It is hard to justify the homogeneity assumption when analysing firm growth. The growth process is dependent on many changing factors such as overall economic performance, changing institutional factors, changes in the population’s education, change in economic openness, technological factors, etc. Homogeneity, therefore, must be considered as a simplifying assumption that enables us to understand the long-term behaviour of size distributions.

**Irreducibility**: There is no theoretical reason to expect that each size group does not have a nonzero probability of moving to other groups even if this process may take a long time. Therefore, communication of states ensures that the Markov process is irreducible.\(^{19}\)

\(^{19}\) A special case where exit is specifically introduced in the transition matrix violates this assumption so the analytical framework has to be adapted to the case.
**Trapping and absorbing states:** The analysis considers first the case where there is no trapping state. Then we define a Markov chain with a trapping state that comprises failing firms. To compute ergodic probabilities in this case, only the sub-matrix which excludes the trapping states is considered. With respect to absorbing states, there is no theoretical reason to expect their existence.

**Recurrence:** As we are dealing with a finite Markov chain, recurrence is automatically implied by the irreducibility assumption, which rules out transience.

**Periodicity:** There is no theoretical basis to suggest that the system we analyse may be periodic. We, therefore, assume that it is aperiodic.

**Ergodicity:** Taken together, the above properties imply that the long-term behaviour of firm size can be characterised by a unique ergodic probability distribution over the state space.

**Empirical data**

The data used were collected in the framework of the Regional Programme on Enterprise Development (RPED) Project. The World Bank launched RPED to collect and analyse firm level data on seven African countries, including Kenya. The data on Kenya were collected for the years 1992, 1993, 1994 and 1999.20 The questionnaire used in the seven countries and different waves was basically the same. In 1999 for example, firms in Kenya were asked a series of questions covering the following aspects: general information on firm structure, location and ownership, entrepreneurs’ profiles, production structure and costs, investment, labour market, government regulations, financial markets, infrastructure, major problems hampering firm activity and expansion, investor confidence, networks and governance. More than 200 firms were surveyed in Kenya in each wave (see earlier discussions) Table 2 shows the general trend in the number of firms surveyed across four waves as well as the rate of attrition per period.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Number of Firms and the Rate of Attrition</th>
</tr>
</thead>
<tbody>
<tr>
<td>From 1992</td>
<td>224</td>
</tr>
<tr>
<td>From 1993</td>
<td>215</td>
</tr>
<tr>
<td>From 1994</td>
<td>219</td>
</tr>
<tr>
<td>From 1999</td>
<td>223</td>
</tr>
<tr>
<td>Total firms</td>
<td>224</td>
</tr>
<tr>
<td>New firms</td>
<td>224</td>
</tr>
<tr>
<td>Attrition</td>
<td>30</td>
</tr>
<tr>
<td>Attrition rate</td>
<td>13.4%</td>
</tr>
</tbody>
</table>


The data in Table 2 suggests that the rate of firm survival in Kenya is relatively low. For instance, Harding and Teal (2000) report that 59 percent of firms in the Tanzanian original 1992 sample were known to be still in existence in 1999 whereas only 33 percent were still operational in Kenya. The data in Table 2 show that the average annual rate of exit in Kenyan
manufacturing is about 12.5 percent for the period 1992 to 1999, almost twice the rate in Tanzania where it is 7 percent. It is relevant to distinguish between firms owned by Kenyans of African origin from those owned by Kenyans of Indian origin. The country’s manufacturing sector has been dominated by Kenyans of Indian origin for a long time. The process leading to this situation is complex and its discussion is beyond the objective of this paper.21

**Empirical Results**

The results are presented in two parts, starting with the findings relating to the period prior to 1992 and followed by those covering the 1990s. These findings should reflect the differences in the macroeconomic environment that characterized these two periods.

**Firms size convergence prior to 1992**

Although the literature on industrial organization in Africa and other developing countries recognizes that there are four distinct size groups (micro, small, medium and large size), there is no consensus on what are the size boundaries separating these groups. Rather than defining these size groups arbitrarily as most studies do, this paper divides the sample into quartiles on the basis of start-up size. Among the 224 firms surveyed in 1992, 208 firms that provided information on their initial size are used in the analysis.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Initial Distribution of Firm Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion</td>
</tr>
<tr>
<td>Micro Firms</td>
<td>0.29</td>
</tr>
<tr>
<td>Small Firms</td>
<td>0.24</td>
</tr>
<tr>
<td>Medium Firms</td>
<td>0.22</td>
</tr>
<tr>
<td>Large Firms</td>
<td>0.25</td>
</tr>
</tbody>
</table>

The values in Table 3 are read as follows. About 29 percent of start-ups were micro firms, which are businesses starting with 3 permanent workers or less.22 A quarter of the sample was made up of large firms or those with more than 17 full-time workers. To determine the level of firm mobility, for example from the micro size group between start-up and 1992, we calculate the proportion of firms which had up to 3 workers at start-up but which were larger than 3 workers in 1992. It should be noted that the analysis of mobility from start-up relates only to surviving firms as those that collapsed were not observed at the time the data were collected in 1992. Applying the methodology developed in section 3 produces the results summarized in Table 4.

---

21 Details on this topic can be found in Bigsten (2002), Phillips et al (2000) and Delf (1963).

22 Although we divided the sample in quartiles, low variability in the data meant that all firms with the same size had to fit in the same size group. For instance, 22.6 percent of start-ups had up to two workers. Moving from two to three workers to have a full quartile increased the proportion to 29 due to the large number of firms with three workers. This explains why we do not have groups with exactly 25 percent of the distribution each.
Table 4 shows several interesting results. First is the noise in the transition matrix. All transition probabilities of surviving firms are nonzero implying that there is mobility in all directions. Hence, the transition matrix is clearly irreducible (See Section 3.1.2). This contrasts with the ‘triple diagonal’ matrix of Kremer, et al. (2001) and other studies of cross-country income mobility where all elements outside the two minor diagonals are zeros.23 Thirty-three percent of surviving micro firms grew during the period, with 4 percent making it to the large size group. Large surviving firms tended to remain in their size group, with only 10 percent of the firms shrinking in size.24 The relatively small proportion of firms leaving the large size group for the smaller size group means that once they become large, firms tend to stay large. This result implies that there was more growth of micro firms than downsizing of large firms, which is an indication of a growing manufacturing sector.

The second interesting result relates to the ergodic probability distribution. Conditional on firm survival, seventy percent of the distribution’s mass is concentrated in the largest size group. In contrast, firms in the micro and small size groups account for only 10 and 9 percent of the distribution suggesting that when they survive, even these firms have an opportunity to grow into medium and large size. Therefore, before 1992, starting small did not seem to be an insurmountable constraint to future growth. The design of industrial policy in these circumstances does not need to focus too much on helping small firms since the economic environment allows them to grow. Indeed, given the importance of the macroeconomic environment on firm performance, this high level of growth was most probably associated with the fact that the transition matrix covered a period during which the Kenyan economy was performing relatively well.

The third result is about the speed of convergence. Using equation (15), it appears that the growth potential of firms was very high. It took 2.6*16 ≈ 42 years to reduce by half the distance between the current and steady state size, where 16 is the average age of the surviving firms. Up and down movements that delayed transition into the large size group in the steady state occurred in the smallest size groups as shown in the transition matrix.

When convergence is conditioned on the use of credit at start-up, the rate of growth is even higher. The results are presented in table 5 which has two panels. The top panel shows the results for firms that used loans at start-up; the second panel shows the findings for firms not using credit.

---

23 This implies we cannot use Kremer, et al. (2001) simplification to calculate ergodic probabilities.

24 Some rows of the transition and ergodic probabilities may not add up exactly to one due to rounding.
The conditioning credit variable is a dummy which takes value 1 when a firm used formal credit at start-up and zero otherwise. Clearly, the transition matrix shows that the firms that used credit grew larger than that which did not. For instance, 80 percent of micro firms that used credit at start-up grew, with 20 percent of them becoming large. The steady state distribution has 82 percent of firms in the large size group, and 93 percent of firms would be either large or of medium size in the steady state. This finding confirms the association of credit use and growth for surviving firms.

Firms not using credit had lower growth than those using it. In the steady state, only two-thirds of the distribution mass is in the large size group and a quarter of firms are either micro or small (relative to only 7 percent of firms using credit). The main finding is that surviving firms that used formal credit to fund their start-up capital recorded higher growth rates. This result is in line with the thesis that credit is important for firm performance, lending support to policies that seek to help SMEs by facilitating their access to financial resources.

Firm size convergence in the 1990s

Following the same methodology applied to the pre-1992 period, firms are grouped into quartiles using 1992 as the base year. Group sizes are different from those in table 3. The thresholds of the different groups are: 4, 18 and 62 permanent workers. These higher thresholds indicate overall growth in the sector in the period prior to 1992. For instance, whereas firms that started with 18 workers or more are ranked as large firms in the pre-1992 period, this is the upper size threshold of small firms in 1992.

By restricting the analysis to the period from 1992 to 1999, equal observation periods are used for every surviving firm. Furthermore, restricting the analysis on the group of firms surveyed in 1992 and observing them throughout the 1990s makes it possible to integrate...
the impact of firm exit in the analysis. We discuss three cases. First, we focus on mobility of surviving firms from the original 224 that were in the sample in 1992. In the second case, exit is specifically integrated into the transition matrix. The third case analyses mobility patterns of entrants, defined here as firms that were five years old or younger in 1992. The transition matrix covering this period is based on equation (8); it captures transitions from 1992 to 1993; 1993 to 1994; and 1994 to 1999.

Case 1: Growth of Surviving Firms

Table 6 summarizes the results of short term growth shown in the transition matrix and the pattern of convergence shown by the ergodic distribution of surviving firms.

<table>
<thead>
<tr>
<th></th>
<th>Size ≤ 4</th>
<th>4 &lt; Size ≤ 18</th>
<th>18 &lt; Size ≤ 62</th>
<th>Size &gt; 62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size ≤ 4</td>
<td>0.73</td>
<td>0.27</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4 &lt; Size ≤ 18</td>
<td>0.23</td>
<td>0.62</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>18 &lt; Size ≤ 62</td>
<td>0.01</td>
<td>0.14</td>
<td>0.81</td>
<td>0.04</td>
</tr>
<tr>
<td>Size &gt; 62</td>
<td>0.00</td>
<td>0.04</td>
<td>0.13</td>
<td>0.83</td>
</tr>
<tr>
<td>Frobenius-Perron eigenvalues</td>
<td>1.00</td>
<td>0.85</td>
<td>0.75</td>
<td>0.39</td>
</tr>
<tr>
<td>Ergodic distribution: 1992 to 1999</td>
<td>0.29</td>
<td>0.33</td>
<td>0.29</td>
<td>0.09</td>
</tr>
</tbody>
</table>

From 1992 to 1999, surviving micro firms recorded the highest level of short term growth: 27 percent compared with 15 percent for small firms and only 4 percent for medium firms. Small firms declined more than any other group followed by large firms. In the steady state the probability mass is concentrated in the small size group (33 percent of the distribution). In contrast with the pre-1992 period where firms tended to concentrate in the large size group, the latter vanishes from the distribution in the post-1992 period. It is likely that firms’ downsizing was the result of the shocks that hit the economy in the 1990s. Convergence, in this case, is from large to smaller size, which is consistent with the finding that, on average, firms experienced a decline rather than growth in the 1990s. On the basis of equation (15), the speed of convergence to this steady state distribution is $4.26 \times 7 = 30$ years.

Could the difference in mobility pattern with the previous case (the period from start-up to 1992) be attributed to the difference in the use of higher size thresholds? That does not seem to be the main reason. If it was the case, the distribution should at least have 25 percent of its mass in the large size group keeping the number of firms that were there in the initial distribution of 1992. It is more plausible that other factors account for this difference. As Figure 1 illustrates, Kenyan firms went through a very difficult period throughout the 1990s, which more likely explains their poor growth performance during this period.

---

25 A fourth case where there are more firms entering than those exiting is not pursued because it is theoretically implausible. More entry eventually drives profit margins to zero implying no entry in equilibrium. In addition, such a hypothesis would require that each period have a different transition matrix, violating the homogeneity assumption.

26 Seven years is the period of analysis (from 1992 to 1999)
Introducing credit changes both the transition matrix and the ergodic probability distribution as shown in table 7.$^{27}$

**Table 7** Effect of credit on Growth of Surviving Firms in the 1990s

(i) Growth of Firms Using Loans

<table>
<thead>
<tr>
<th>Size</th>
<th>$\leq 4$</th>
<th>$4 &lt; Size \leq 18$</th>
<th>$18 &lt; Size \leq 62$</th>
<th>$&gt; 62$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.64</td>
<td>0.29</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.08</td>
<td>0.92</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.03</td>
<td>0.10</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.86</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ergodic distribution</td>
<td>0.00</td>
<td>0.17</td>
<td>0.74</td>
<td>0.09</td>
</tr>
</tbody>
</table>

(ii) Growth of Firms Not Using Loans

<table>
<thead>
<tr>
<th>Size</th>
<th>$\leq 4$</th>
<th>$4 &lt; Size \leq 18$</th>
<th>$18 &lt; Size \leq 62$</th>
<th>$&gt; 62$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80</td>
<td>0.20</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td>0.58</td>
<td>0.10</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.18</td>
<td>0.79</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.11</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>0.87</td>
<td>0.76</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>Ergodic distribution</td>
<td>0.49</td>
<td>0.31</td>
<td>0.17</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The results in Table 7 suggest that surviving micro and small firms using loans had much higher rates of short term and long-term growth than in the unconditional case. The ergodic distribution shows that if surviving firms had access to loans, there would be no micro firms in the steady state. Three-quarters of the firms would be in the medium size category with 9 percent of firms in the large size group. This result confirms previous findings that conditional on survival, credit fosters firm growth. However, growth is less important than in the period before 1992 where 82 percent of surviving firms using credit would become large. The bottom panel of Table 7 shows that the lack of access to credit pushes half of the surviving firms into the micro size group and 31 percent in the small size group. Therefore, if surviving firms do not have access to loans, 80 percent of them would be micro or small in the steady state, which signals a generalised decline of large and medium size firms. Credit does not seem to affect the speed of convergence given that the second highest eigenvalues in the top and bottom panels are almost equal.

$^{27}$ Some probabilities are based on limited observations given that only a quarter of the firms in the sample used credit. Credit use is particularly limited for micro and small firms.
Case 2: Firm Growth Taking Exit into Account

Introducing a new group of failing firms, which is an absorbing state, changes the picture. The results in Table 8 illustrate the effect of exit on the distribution of firm size.

### Table 8  Transition Matrix of Firm Growth Integrating Exit

<table>
<thead>
<tr>
<th>Exit</th>
<th>Size ≤ 4</th>
<th>4 &lt; Size ≤ 18</th>
<th>18 &lt; Size ≤ 62</th>
<th>Size &gt; 62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size ≤ 4</td>
<td>0.29</td>
<td>0.522</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>4 &lt; Size ≤ 18</td>
<td>0.27</td>
<td>0.17</td>
<td>0.45</td>
<td>0.09</td>
</tr>
<tr>
<td>18 &lt; Size ≤ 62</td>
<td>0.23</td>
<td>0.01</td>
<td>0.11</td>
<td>0.63</td>
</tr>
<tr>
<td>Size &gt; 62</td>
<td>0.24</td>
<td>0.00</td>
<td>0.03</td>
<td>0.10</td>
</tr>
</tbody>
</table>

The results in Table 8 show that medium size firms were the most resilient group with an exit probability of 0.23 followed by large firms with a probability of 0.24. Small and micro firms post probabilities of failure of 0.27 and 0.29, respectively. The transition matrix in Table 8 does not have an ergodic distribution because when there are exits but no entry, the steady state distribution is degenerate. These results suggest that the probability of failure monotonically declines with firm size a fact that should be taken into account when designing firm growth strategies. Integrating credit and exit at the same time leads to the following results (see Table 9).

### Table 9  Effect of Loans on Growth Taking Exit into Account

(i) Growth of Firms Using Loans

<table>
<thead>
<tr>
<th>Exit</th>
<th>Size ≤ 4</th>
<th>4 &lt; Size ≤ 18</th>
<th>18 &lt; Size ≤ 62</th>
<th>Size &gt; 62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size ≤ 4</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4 &lt; Size ≤ 18</td>
<td>0.55</td>
<td>0.00</td>
<td>0.27</td>
<td>0.18</td>
</tr>
<tr>
<td>18 &lt; Size ≤ 62</td>
<td>0.39</td>
<td>0.00</td>
<td>0.00</td>
<td>0.61</td>
</tr>
<tr>
<td>Size &gt; 62</td>
<td>0.45</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

(ii) Mobility of Firms Not Using Loans

<table>
<thead>
<tr>
<th>Exit</th>
<th>Size ≤ 4</th>
<th>4 &lt; Size ≤ 18</th>
<th>18 &lt; Size ≤ 62</th>
<th>Size &gt; 62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size ≤ 4</td>
<td>0.39</td>
<td>0.51</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>4 &lt; Size ≤ 18</td>
<td>0.23</td>
<td>0.21</td>
<td>0.46</td>
<td>0.10</td>
</tr>
<tr>
<td>18 &lt; Size ≤ 62</td>
<td>0.24</td>
<td>0.00</td>
<td>0.12</td>
<td>0.61</td>
</tr>
<tr>
<td>Size &gt; 62</td>
<td>0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
</tr>
</tbody>
</table>

The results in Table 9 provide a clear picture of the difference between the effect of loans on the growth of surviving firms and on firm survival. Whereas all the results reviewed so far show that the use of loans helped surviving firms to grow, credit use tended to increase the probability of firm failure. Micro firms are most affected by the use of credit because
all micro firms using loans collapsed. Medium size firms appeared to be the most resilient among the group of firms using credit. It is important to note that large firms either survive or die but do not downsize. Forty-five percent of large firms using credit collapsed rather than downsizing while only 13 percent of large firms not using credit downsized. The implication is that the impact of credit on firm survival may be rather brutal. This is not surprising. Firms that contracted medium term loans in 1992 when the average interest rate was 21 percent were required to pay 30 percent the next year and 36 percent two years later, in a situation of economic downturn. It is possible that many such firms were unable to cope with the increase in interest rates and therefore collapsed (Nkurunziza, 2012).

Case 3: Growth of Entrants

Old and new firms may have a different mobility pattern. Firms that were created during the period of economic growth may have had better chances to grow relative to those created in the late 1980s. To investigate this hypothesis, we analyse mobility of firms that had been in operation for a maximum of five years in 1992; the results are presented in Table 10. Only surviving firms are used to compute the steady state distribution of firm size by adjusting the transition probabilities to ensure that marginal probabilities sum to unity.

<table>
<thead>
<tr>
<th>Exit</th>
<th>Size ≤ 4</th>
<th>4 &lt; Size ≤ 18</th>
<th>18 &lt; Size ≤ 62</th>
<th>Size &gt; 62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size ≤ 4</td>
<td>0.39</td>
<td>0.44</td>
<td>0.17</td>
<td>0.00</td>
</tr>
<tr>
<td>4 &lt; Size ≤ 18</td>
<td>0.18</td>
<td>0.32</td>
<td>0.44</td>
<td>0.06</td>
</tr>
<tr>
<td>18 &lt; Size ≤ 62</td>
<td>0.40</td>
<td>0.00</td>
<td>0.10</td>
<td>0.50</td>
</tr>
<tr>
<td>Size &gt; 62</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>Frobenius-Perron eigenvalues</td>
<td>1.00</td>
<td>0.82</td>
<td>0.56</td>
<td>0.27</td>
</tr>
<tr>
<td>Ergodic distribution</td>
<td>0.50</td>
<td>0.36</td>
<td>0.15</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The transition matrix in Table 10 suggests that among young firms, micro and medium size enterprises had the highest exit rate followed by large and then small firms. These results contrast with those in Table 8 where exit rates decrease monotonically with size. The ergodic probability distribution of the entrants shows that in the steady state half of the firms would be micro and 86 percent of all firms would be either micro or small. Interestingly, there is no large firm in the steady state, an indication that firms created in the late 1980s and early 1990s had a lower growth potential than their older counterparts, probably as a result of difficult economic conditions that prevailed in the 1990s. The speed of convergence to the steady state is 24 years, the lowest half-life of convergence for all the four cases analysed.

---

28 In this case, the stochastic matrix is not irreducible so the transition matrix has no ergodic probability distribution.
29 The choice of five years is arbitrary. The objective is to determine whether the firms created at the beginning of the crisis period have a different mobility pattern than those created before.
Conclusion

Using a Markov chain to model growth and convergence of firm size in Kenyan manufacturing sector, this paper has shown that the extent of firm growth and convergence depends on the overall economic environment and whether or not a firm uses credit. The main findings provide answers to the four questions the paper proposed to address. First, if the analysis is limited to the period of economic growth covering the years prior to 1992, 70 percent of surviving firms would become large in the steady state, on average. Using credit increases the share of large firms to 82 percent of the distribution. This result suggests that for a firm operating in a good economic environment, starting small should not be considered as a binding constraint to its growth. Indeed most micro and small firms eventually grow and become large. The fact that the steady state distribution is concentrated in the large size group contradicts Gibrat’s law of proportionate effect. Also, the results show the positive effect of credit on the growth of surviving firms.

Second, during the 1990s, a period of economic slowdown, the steady state distribution of surviving firms is relatively balanced in the first three size groups, a notable difference with the results over the pre-1992 period. Restricting the analysis to entrants, namely firms that were 5 years old or younger in 1992, the ergodic distribution has most of its mass in the lowest two size groups. The implication is that the growth potential was limited, constraining small firms to remain small. The use of credit helped surviving firms to increase their growth but not to the same extent as that observed in the pre-1992 period. During the 1990s, three-quarters of surviving firms using loans converged in the medium not the large size group.

Third, incorporating firm exits in the analysis changes the results drastically: all micro firms using loans collapse. Hence, bearing in mind the caveat concerning the limited number of observations some probabilities are based on, the general conclusion from the analysis is that the use of credit by surviving firms was beneficial. However, the use of credit amidst a severe economic crisis during which the financial sector witnessed severe shocks to interest rates caused failure for fragile firms, especially those from micro and small size groups.

These conclusions are based on just one case study so they should not be generalized before similar studies on other African countries are carried out to confirm their robustness. If additional studies lead to the same conclusions, it will be appropriate to call for a reassessment of some industrial support policies that have been implemented in several countries to help SMEs to grow. In light of the results in this study, if a country’s main objective is the development of its industrial sector, support policies such as special credit schemes would have to focus on the firms that have shown some level of resilience in order to avoid concentrating limited resources on young and small fragile firms that would not survive the next economic or financial shock. This does not, however, mean that support for start-ups should be abandoned altogether. If a country’s policy objective is to fight against poverty and improve fairness in resource allocation, targeting small and young firms would be appropriate given that it is generally the poor and people at the lower end of income distribution who create micro and small firms as a survival coping strategy. Moreover, if a country intends to develop a specific sector, it would be understandable that it assists new entrants in the sector.

References


