77926

Constrained Gazelles: High Potentials in West Africa's Informal Economy♣

Michael Grimm,[a,b*] Peter Knorringa,[a] Jann Lay[c,d]

March 2012

Abstract

The informal sector is typically characterised as being very heterogeneous and possibly composed of two clearly distinct segments, sometimes called the lower and upper tier. However, empirical evidence shows that even among lower tier entrepreneurs profitability can be quite high. We combine these findings and develop an innovative approach to identify what we call 'constrained gazelles', next to the well-known survivalists in the lower tier and growth-oriented top-performers in the upper tier. Our sample of informal entrepreneurs in seven West-African countries allows to link the relative size of these three groups to the structural and macroeconomic environment in these countries.

Keywords: Informality, Entrepreneurship, Capital, Managerial Ability, Targeting, West-Africa.

JEL codes: D22, O17.

[a] *International Institute of Social Studies, Erasmus University Rotterdam, The Hague, The Netherlands*

[b] *Institute for the Study of Labor (IZA), Bonn, Germany.*

[c] *GIGA German Institute of Global and Area Studies, Hamburg, Germany*

[d] *University of Göttingen, Germany*

[*] Corresponding author: Michael Grimm, International Institute of Social Studies, Erasmus University Rotterdam, Kortenaerkade 12, 2518 AX The Hague, The Netherlands, Phone: +31-70-4260694, Fax: +31-70-4260799, E-mail: grimm@iss.nl.

Acknowledgements

Disclaimer

1. INTRODUCTION

Informal sector heterogeneity has been recognized for decades (Mead and Morrisson 1996). From an initial perception of a dichotomy, coined by Hart (1972), the academic debate moved towards seeing the myriad variation in types of firms, entrepreneurs and workers as a continuum, within which authors identified various types of segmentation, at least partly depending on their thematic interest, disciplinary background, and adherence to particular strands in development theory (Lubbel, 1991; Chen *et al.* 2005). Also context specificity is an important underlying source of heterogeneity. For example, the debate on the informal sector in Latin America and in economically more dynamic regions of South East Asia focuses on whether most informal entrepreneurs would prefer to have a rather secure job or are constrained genuine entrepreneurs (Maloney, 2004; De Mel *et al.*, 2010). For economically less dynamic regions in Africa with a glaring missing middle of small and medium-sized enterprises (Fafchamps, 1996; Tybout, 2000), the debate has focused on the often visible bifurcation between a rather small group of successful entrepreneurs and a much larger group of entrepreneurs that apparently struggle to survive (House, 1984; Fields, 1990; Rogerson, 1996; Mead and Liedholm, 1998; ILO, 2002; World Bank, 2007; Nichter and Goldmark, 2009). A focus on this qualitative bifurcation has proven useful to avoid one-size-fits-all thinking, but it also has inhibited progress in identifying possible other segments of informal entrepreneurs.[i] Moreover, progress in producing more rigorous and generalizable findings on informal entrepreneurs has been hampered because representative samples hardly exist.

To push the debate forward, we apply an innovative empirical strategy to a unique data set based on a *representative* sample of informal entrepreneurs in seven capital cities in francophone West Africa. Instead of two, we identify three groups of entrepreneurs: First, the well-known success stories or top performers; second, a group of constrained gazelles[ii] who share some characteristics with these top performers such as education, language skills, sector choice and some basic management abilities, but who are not (yet) successful; and third, a group of survival entrepreneurs with fundamentally different characteristics. While constrained gazelles and top performers have some similarities, the constrained gazelles – similar to survivalists – possess very low levels of capital. Our data show that constrained gazelles earn a very high marginal return to capital, which underscores their potential to become top performers. Moreover, our empirical strategy enables us to arrive at some clear criteria for separating out constrained gazelles, so that they can benefit from targeted interventions to unlock their potential. Given that our findings are based on a representative sample, we also discuss the extent to which these criteria could be applied more generically to identify potentially successful but constrained entrepreneurs in the urban informal sector in other regions of the developing world. This could significantly improve targeting and thus the effectiveness of informal sector interventions. Finally, given the multi-sectoral and cross-country dimensions of our data, we can link the relative size of the three groups of entrepreneurs to the sectoral composition and the macroeconomic environment in these seven countries.

The remainder of the paper is organised as follows. In Section 2 we combine key findings from the two main analytical views on why distinct groups of informal entrepreneurs co-exist. Moreover, we position constrained gazelles in the intersection between these views. Section 3 discusses the context and data. Section 4 derives and characterises three distinct groups of entrepreneurs. Section 5 analyzes differences in productivity across these three groups. Section 6 discusses policy implications and Section 7 summarises our findings.

2. COMBINING VIEWS ON HETEREGONEITY

Different analytical views exist on why heterogeneity persists among informal entrepreneurs with differing policy implications. Given our focus on seven capital cities in francophone West Africa, an economically less dynamic region with a wide gap between its formal and informal economies, we focus on the following two main strands in the literature. The first multi-disciplinary strand posits the existence of two qualitatively distinct subgroups of entrepreneurs. On the one hand, relatively better-off growth-oriented entrepreneurs who can afford to specialise and can focus on accumulation. On the other hand, survival entrepreneurs who need to be risk averse, and thus diversify their economic activities in search of stability. The main policy implication is that these two distinct groups require distinct policy approaches: a targeted business development approach for growth-oriented entrepreneurs and a more generic approach to assist survival entrepreneurs in enhancing their employability and alleviating their poverty (Berner *et al.*, 2012; Cotter, 1996; Mead and Liedholm, 1998).

A second strand - including a study based on the same data used in this paper - shows that entrepreneurs who operate with low levels of invested capital can earn high returns to capital (De Mel, McKenzie and Woodruff, 2008; McKenzie and Woodruff, 2006; Kremer, Lee and Robinson 2010; Grimm, Krüger and Lay, 2011). In other words, in these firms the return to investing in an additional hammer or a basket to carry goods to the market is remarkably high. Hence, this literature contradicts the validity of the poverty trap models that often assume that some fixed costs have to be incurred before above subsistence returns can be earned (Banerjee and Newman, 1993) and that generally associate low productivity and product quality to informal firms as separate from formal firms (La Porta and Shleifer 2011). The high returns to capital literature also casts serious doubt on the overriding assumption in much of the first strand of informal sector literature that none of the smallest firms can generate significant value added. Moreover, among practitioners it is commonplace to say that the smallest entrepreneurs are not entrepreneurial and would be better off finding a job. While this may well be true for real survival-oriented entrepreneurs, we will argue that this does not hold for a significant third segment of informal entrepreneurs, which we call constrained gazelles (Boston and Boston, 2007).[iii]

The two strands in the literature mentioned above usually rely on different (types of) data sets and have not systematically tackled their distinct assumptions. In this study we will demonstrate the added value of combining the qualitative characteristics of informal entrepreneurs with their returns to capital in one representative data set. We will show that combining these analytical views allows us to identify a third group of potentially successful entrepreneurs that so far remained 'hidden' inside the amorphous survival segment. Moreover, our empirical strategy allows isolating these constrained gazelles from top performers and survivalists, a key precondition for effective targeting.

A final point to consider in this section is the empirical importance of constrained gazelles in the informal sector. As we will show in some detail below, constrained gazelles are a significant group in the informal sector of urban West Africa, their share ranging from between 20% to 35%, with outliers of even almost 60%, depending on the country and the method of calculation. Country comparisons suggest that it is more likely to find more constrained gazelles in more dynamic economies. As the countries in our sample arguably possess less dynamic economies compared to many other regions in the world, we believe that the presence of constrained gazelles will also be significant in other regions.

3. DATA SOURCE AND DESCRIPTIVE OVERVIEW

(a) Enquêtes 1-2-3

For our empirical analysis we use a set of survey data stemming from the so-called 1-2-3 surveys or in its French synonym 'Enquêtes 1-2-3'. The data have been collected in 2001 and 2002 in seven urban centers in West Africa: Abidjan, Bamako, Cotonou, Dakar, Niamey, Lomé and Ouagadougou.[iv] A 1-2-3 survey is a multi-layer survey organized in three phases and specifically designed to study the informal sector (see Brilleau, Ouedraogo and Roubaud (2005) for a presentation of data and methodology). Phase 1 is a representative labor force survey collecting detailed information about individual socio-demographic characteristics and employment. Phase 2 is a survey which interviews a sub-sample of informal production units identified in Phase 1. The focus of this phase is on the characteristics of the entrepreneurs and their production unit, including the characteristics of employed workers. It also contains detailed information on costs, input use, investment, sales and the unit's forward and backward linkages. Phase 3 is a household expenditure survey interviewing (again) a representative sub-sample of Phase 1 and hence part of the Phase 2 households. The data of all three phases are organized such that it can be linked. Hence, for a (representative) sub-sample of informal entrepreneurs we have information from Phase 1 and Phase 2 (n=6,580) and, again for a subsample, information from all three phases. Phase 3 is not available for Abidjan because of the onset of violence in 2001 before it could be implemented.

Besides the detail of information, a major advantage of the 1-2-3 survey is that Phase 1 ensures that Phase 2 delivers a representative picture of the informal sector, because being sampled does for instance not depend on whether the entrepreneur has a fixed location or simply operates a business at home or in a fully mobile way. Thus the survey also includes entrepreneurs who are likely to be overseen in one-stage surveys where the sample population is produced from enterprise sampling frames. The 1-2-3 surveys define informal enterprises as small production units that (a) do not have written formal accounts or (b) are not registered with the tax administration or both. Part (b) of this definition varies slightly between countries, as registration may not always refer to registration with tax authorities. The 1-2-3 surveys do not apply a size criterion. The detail of information, the representativeness and the cross-country dimension make this data particularly suitable for our purpose.

(b) Heterogeneity of sampled firms

Table 1 shows some basic characteristics of the sampled firms and their owners for the four quartiles and the top decile of the capital distribution. Capital is measured as the used physical capital stock evaluated at replacement costs. It includes buildings, machines, furniture, vehicles and tools used for the business. Price differences between countries are adjusted via the use of PPPs. This is explained along with definitions for all other used variables in Table A1 in the Appendix. As expected the variation across capital quartiles is large in many of the shown characteristics. The most striking is probably gender. Whereas in the bottom quartile 65% of the firm managers are women, it is only 30% in the top quartile. 37% of the entrepreneurs in the bottom quartile speak French compared to 64% in the top quartile. In the bottom quartile 74% are without any formal education while this percentage is 'only' 56% in the top quartile. In this quartile 8.2% have at least some secondary education, which is the case for only 4.1% in the bottom quartile. However, maybe surprisingly, firm age does almost not vary across the capital distribution. It also seems as if the household's activity portfolio had only limited impact on capital accumulation of informal firms. We cannot see more capital in firms that belong to households with income from a formal activity, possibly

because the dominance of formal earnings implies that entrepreneurial activities constitute a secondary activity only complementing the main source of income. Regarding the sectoral distribution, we tend to find more top quartile firms in the transport, hotel and restaurant and repair services sectors.

[Table 1 about here]

4. IDENTIFICATION OF CONSTRAINED GAZELLES

Our empirical strategy to identify three categories of informal entrepreneurs is as follows: First we define a category of 'top performers' based on used physical capital and generated value added. Then, we identify a set of owner and firm characteristics that are correlated with these performance measures. Using these correlations we predict the empirical probability of being a 'top-performer'. Based on the actual status of being a top performer and the predicted probability of being one (although the entrepreneur is not), we then classify entrepreneurs into three groups: 'top performers', 'constrained gazelles' and 'survivalists'. 'Constrained gazelles' are those entrepreneurs who have a high empirical probability of being a 'top-performer' given their *observable* characteristics. Based on these classifications, we discuss the differences in the informal sector composition across countries and relate these differences to the country-specific structural and macroeconomic environment. Finally, we identify and discuss the structural and behavioural differences between these groups. A particular focus will be on factors that are related to the way entrepreneurs manage their firms.

(a) Defining a top performer

To define 'top-performers' we combine a size and a productivity criterion. The size reflects past performance and the capacity to grow. Productivity measures current performance. We proceed in two steps. First, we select those entrepreneurs in our sample that are in the top 25% of the capital distribution of their respective country. From these firms, we then select the 40% with the highest capital profitability, i.e. the highest value added per unit of physical capital. This means that the group of top-performers comprises by definition 10% of all sampled entrepreneurs in every country. Value added is measured as turnover minus the costs for intermediate inputs; hence it includes all paid wages, the implied income of the owner and unpaid family members or other unpaid workers, such as trainees. Alternatively, one could just use the income of the owner. However, given that family labour is widespread with a number of entrepreneurs reporting wages for those family members, we prefer value added over the owner's income. Moreover, we do not use a performance criterion based on employed staff. Informal firms typically do not conclude labour contracts, are often engaged in seasonal activities and can hardly be separated from the household. This is why current employment is probably not a very reliable size indicator. We test, however, the robustness of our approach using two alternative definitions; one where performance is just determined by total value added and another using value added per worker. It turns out that the correlation between all three definitions is relatively high (close to 60% and more) and that there are no substantial differences in terms of the shares of top-performers across economic sectors. The alternative definitions yield slightly more top-performers in the very capital intensive sectors transport and construction as well as in the less capital intensive sectors of retail and petty trade. The results hence do not seem to be biased by using capital as a size criterion. In addition, capital accumulation is central to economic growth.

(b) Partitioning into top-performers, constrained gazelles and survivalists

The set of variables that are correlated with being part of the group of top-performers should be largely given or predetermined at the start of the enterprise. 'Predetermined' thus simply means determined prior in time. These factors may still be endogenous in an econometric sense, because both the entrepreneur's success and the correlated variables may depend on the same third variables, for instance family background. We include in the set of predetermined factors: age, age squared, gender, formal education, knowledge of French, and the motivation of the entrepreneur to set up the enterprise. The motivation of the entrepreneur to set up the enterprise is measured via a dummy variable that takes the value 'one' if the entrepreneur created the enterprise because he or she could not find a job as a wage worker (either in a small or larger company). Alternative answers to this question are 'to receive higher earnings', 'prefer to be independent' and 'enterprise created because of a family tradition'. If the respondent chooses one of these alternative answers, we take this are a sign of a minimum 'entrepreneurial spirit'. In the literature, this variable is sometimes used as a measure of the entrepreneur's ability (see e.g. McKenzie and Woodruff, 2006). Amin (2009) uses the same variable to separate 'opportunity' from 'necessity' firms. In addition to the predetermined factors, we include sector and country dummies and control for the firm's age to avoid discriminating against younger firms with a high potential. We think the virtue of our approach is that at the start we limit the classification to a fairly limited set of variables, which are not much prone to endogeneity issues, but can then show below that based on these variables we identify a group of entrepreneurs that have business skills which resemble in many respects those of the top-performers but are clearly different of those characterizing survivalists. We take this then as an indication that this group has a potential that can be unlocked. Once we have analyzed structural and behavioral differences between these three groups, we can further refine the classification of these entrepreneurs into three groups. This will be done at the end of this section.

Hence, to start with, we regress the binary variable 'being a top-performer', Y^{Perf}, on the list of variables just discussed assuming that there is a latent unobserved probability of being a 'top performer' underlying the observed binary variable being or not being part of that group. Hence, the appropriate econometric model is a binary response model. We estimate a probit model of the following form:

$$\Pr(Y_i^{Perf} = 1) = \theta(\beta_0 + X_i'\beta_1 + \omega_i),$$ (1)

where X_i stands for the vector of the above owner and firm characteristics. β is the vector of coefficients that describes how these characteristics relate to the probability of being a top-performer. θ stands for the cumulative standard normal distribution function, i.e. the underlying probability distribution in a probit model. Table 2 shows the results of this regression. We show the estimated coefficients and the implied marginal effects, i.e. the change in the underlying probability if the explaining variable is changed by one unit (from zero to one for binary variables) while all other variables are kept at the sample means. Means of the independent variables can be found in Table 1.

[Table 2 about here]

We see that the probability of being a 'top-performer' is increasing in age, but at a decreasing rate (the quadratic term is negative). Firms managed by men are more likely to be in the group of top-performers than firms managed by women. The marginal effect indicates a higher chance by almost 9 percentage points for men being in that group. Given that overall only 10% of all entrepreneurs are classified as top-performers, this effect points to some important

gender issues, such as access to household resources and the need of women to combine their professional activity with housework. The effect of schooling is positive and significant, but there is no significant difference between primary, secondary and any other type of higher than primary schooling, including technical schooling. Speaking French also increases the probability of being a 'top performer' (by 2.5 percentage points). The age of the firm has a small positive effect; at the margin an increase in the firm's age by one year increases its probability of being a top-performer by just 0.1 percentage point). The sector effects indicate that firms in the hotel and restaurant or in the transport sector have a higher probability of being in the group of top-performers relative to the firms that are in textile manufacturing (clothing and apparel). Almost all country effects (not shown) are insignificant. They should be insignificant, as we simply define 10% of the firms as top-performers in every country. The small differences we still see are just driven by the small differences in sample sizes across countries and possibly by the non-linearity of the choice model.

Based on this regression, we now predict for every entrepreneur in the sample the statistical probability of being a top-performer – conditional on the estimated parameters β_0 and β_1 and on the vector of observed variables X, i.e. we compute

$$\hat{\mathrm{Pr}}(Y_i^{Perf}=1) = \theta(\hat{\beta}_0 + X_i^{'}\hat{\beta}_1), \tag{2}$$

where 'hats' indicate estimated parameters. We use these predicted probabilities to partition in each country the sample of informal entrepreneurs into three groups: top-performers, constrained gazelles and survivalists. The 'top-performers' are those entrepreneurs that are actually 'top-performers', i.e. the top 40% entrepreneurs in terms of capital productivity from the top 25% of the distribution of capital. The shares of the two other groups are determined as follows: In each country, we choose the cut-off point for $\hat{\mathrm{Pr}}(Y_i^{Perf}=1)$ between constrained gazelles and survivalists such that the mean of $\hat{\mathrm{Pr}}(Y_i^{Perf}=1)$ is the same in the group of top-performers and constrained gazelles, i.e. constrained gazelles should on average be equally likely to be a top-performer than the actual performers. This should imply that the distribution of the observable variables X is the same in both groups. This is confirmed by a regression (results not presented) of Equation (1) on the pooled sample of top-performers and constrained gazelles; all regression coefficients are insignificant, meaning that there can only be minor (and no systematic) differences in the observables between both groups. This will be further discussed below. As survivalists we define all entrepreneurs that have $\hat{\mathrm{Pr}}(Y_i^{Perf}=1)$ below the threshold. We also check graphically whether these partitions satisfy two basic criteria. First, the distribution of predicted probabilities, $\hat{\mathrm{Pr}}(Y_i^{Perf}=1)$, of top-performers and constrained gazelles should more or less overlap while the distribution of survivalists should be clearly distinct. As illustrated in Figure 1 (Kernel densities), this is the case. Second, the distribution of capital should show similarly low levels of capital for survivalists and constrained gazelles, while top-performer's higher capital stock should be clearly visible. Indeed, Figure 2 shows this pattern. Although the distribution of capital of constrained gazelles is a bit more on the right than the one of survivalists, it is far closer to the latter than to the one by top-performers, i.e. constrained gazelles have only slightly more capital than survivalists. Hence, constrained gazelles might have a high potential to increase their capital stock.

[insert Figures 1 and 2]

In Table 3, we present the partition into the three informal sector segments for each of the seven countries and relate the differences in this partition across countries to the country-specific structural and macroeconomic environment. Obviously, one has to be careful in drawing conclusions based on seven data points, but some correlations clearly stand out. First, the share of survivalists is strongly negatively correlated with the share of private formal and public sector employment and strongly positively correlated with the urbanisation rate and the share of agriculture in GDP. We also consider various indicators that measure in one way or another the quality of the 'business climate', namely the number of days it takes to enforce a contract – drawn from the World Bank's 'doing business surveys' - and two measures drawn from the business freedom data base hosted by the Heritage Foundation (business freedom, investment freedom). All three measures suggest that a better business climate is associated with a lower share of survivalists. Both data bases provide alternative indicators, but the three selected ones are in our view those that are the most relevant for informal small enterprises. In addition, these are available for 2001 or, in case of doing business, at least for 2003. To summarize, all these correlations are plausible and point to important bottlenecks that keep many people in survival-oriented self-employment with little chance of developing a dynamic business. It is interesting to see that GNI is not correlated with the share of survivalists, suggesting that structural features and the quality of the business climate, as measured here, seem to matter more than the income level per se. The general level of formal education does also only play a minor role.

[Table 3 about here]

(c) Socio-demographic heterogeneity and sector choice

We now study the heterogeneity in the informal sector across these three groups. The upper panel of Table 4 (basic characteristics) shows the means of the variables that were used to estimate Equation (1) and to predict $\hat{\Pr}(Y_i^{Perf} = 1)$ (only country dummies are not shown). By construction, of course, we expect similar means for the group of top-performers and constrained gazelles, while the group of survivalists should differ in those characteristics – as shown above. The most striking contrast is obtained for the variable gender; only 13% of survivalist firms are led by men. The dominant sector among survivalists is clearly 'petty trade' with some activity in 'other manufacturing and food processing' and 'other services'. In contrast, the activities of the two other groups are more equally distributed across sectors with most entrepreneurs engaged in 'other manufacturing and food processing', 'clothing and apparel' and 'construction'. The transport sector is the fourth most important sector of top-performers. The fact that 'other manufacturing and food processing' is important for all groups, illustrates this sector's heterogeneity in terms of firm performance. Comparing survivalists and constrained gazelles it turns out that survivalists are a bit younger (34.7 vs. 37.7 years), have lower education, speak French less often, show less 'entrepreneurial spirit' and, on average, have firms that are younger (6.4 vs. 9.4 years). In Table 4 we also show the mean capital and mean labour used by these firms. Top-performers employ on average 2.7 workers including the owner, while constrained gazelles employ 1.8 and survivalists only 1.3 workers.

[insert Table 4 about here]

(d) Structural heterogeneity

We now focus on structural variables that have not been used in determining group membership. These variables come from the economic, institutional as well as the social

domain. Among the economic factors we consider whether the enterprise was set up by the entrepreneur alone or with another person. We also take into account whether other household members of the entrepreneur (or the entrepreneur him/her-self) are employed in the public or formal private sector. Again, setting up an enterprise with others or having access to formal employment and regular earnings may lift credit market constraints and allow sharing business related risks. For the same reason, we include an asset index based on the ownership of various non-business assets in the household of the entrepreneur. This index is constructed using principal component analysis (Filmer and Pritchett, 2001; Sahn and Stifel, 2000). We also include the type of location used by the entrepreneur. An additional economic factor is whether an entrepreneur has accomplished an apprenticeship in a larger firm or brings some experience from such a firm. Institutional factors focus on access to public services, such as access to water, electricity and telecommunication. The social factors include a possible entrepreneurial tradition of the household in this particular activity. Social networks and connectedness that may improve access to output and factor markets are proxied by whether the entrepreneur belongs to the first, second, third or less important ethnic group in the country (in terms of the proportion), by the share of the population in the entrepreneur's neighbourhood (cluster) that belong to entrepreneur's ethnic group and by migrant status (distinguishing non-migrants, internal migrants, (international) immigrants and return migrants). The second panel of Table 4 (structural factors) presents the means of all variables for the three groups of entrepreneurs. Obviously, none of these variables can be considered unambiguously exogenous, but the objective is not to identify causal relationships, but to identify factors that correlate with entrepreneurial success and to separate these from those that seem rather unrelated. Identifying causal relationships needs a different type of analysis and is of course the next step towards the design of specific policies that would allow lifting constrained gazelles to the group of top-performers.

A first interesting finding is the correlation between the way the firm was set up and the classification across the three groups of entrepreneurs. Both top-performers and constrained gazelles have done this as a 'joint venture' in almost in one out of ten cases. This happens less frequently among survivalists. However, the household activity portfolio does not significantly differ across the three groups. Hence, entrepreneurs in households with a formal income source are not systematically more successful than entrepreneurs without such links. This does of course not imply that such regular income is not a potential advantage. The lack of correlation might reflect that some households with formal-wage-earning household members pursue the informal entrepreneurial activity as a side activity, maybe undertaken by the spouse, which is not intended to grow in size or to be very important in general. This interpretation is supported by the fact that wealth, in contrast, shows a quite sizeable correlation with success. Top-performers clearly come from wealthier households. This may be a sign of a lack of access to credit that holds back the group of constrained gazelles. Wealthier people may also have other (non-observed) traits that give them better opportunities to set up and develop their business. Of course, it is also possible that successful entrepreneurs simply accumulate more wealth. Four other correlations stand out. First, access to electricity and telecommunication is the highest among top-performers and, with the exception of water, still higher among constrained gazelles than among survivalists. Second, top-performers operate more often from fixed locations (shops and restaurants) than from mobile selling points. This applies also to constrained gazelles, relative to survivalist, but to a lesser extent. Third, pursuing an activity because of a family tradition is associated with a significantly lower probability of being a top-performer). Fourth, internal and return migrants (from abroad) are less likely than non-migrants to belong to the top-performers. In contrast, international migrants, i.e. migrants coming from neighbouring countries or even further away, are not different from non-migrants in that respect. To summarize, this part of Table 4

suggests that constrained gazelles differ from top-performers mainly in terms of wealth, access to utilities and the type of location they can use. This, again, may be an indication that constrained gazelles are not mainly restricted by internal constraints, such as skills, but rather by external constraints, such as access to capital. With respect to prior experience in a larger firm, which should be a major determinant of business relevant skills, both groups are in fact very similar and clearly different from survivalists.

(e) Behavioural heterogeneity

Next, we examine to what extent survivalists, constrained gazelles and top-performers differ in terms of their behaviour in managing and controlling their firm. Hence, we include variables for book-keeping and an index of financial literacy based on questions regarding the knowledge of credit and other financial products. We also construct dummy variables from entrepreneurs' answers to the following questions: Whether the entrepreneur (1) adjusts supply to actual or expected demand instead of just producing at the level of capacity, looks actively for new clients, (2) has problems to manage the firm, (3) would consider in case of a decline in demand to reduce profits, to diversify the activity or to improve the quality of the products, (4) is member of a business association,[v] and (5) would consider formalizing his or her firm. Moreover, we include log monthly working hours of the entrepreneur as a measure of individual effort. Finally we include the ratio of employed non-paid workers (so mainly family members) to total staff (including the owner), a control for self-employment and a dummy for whether the entrepreneur would consider hiring family members irrespective of their qualification rather than ex-workers from small or larger enterprises. We also add a very simple measure of risk aversion on the basis of a question that asks for the motivation for taking up the specific business activity. Here, one possible motivation is "Assurance of more stable receipts than in other products", which we take as a sign of risk aversion. For about 10% of all entrepreneurs the corresponding 'chosen because of stable-profits' dummy is set to 1. We believe that this list covers the important dimensions of how entrepreneurs are capable of and actually managing their firms.

The means of these variables are again shown in Table 4. The first striking feature is that top-performers do better than the two other groups almost along all variables we consider here. Recall that our initial classification is just based on past capital accumulation and capital productivity, hence these findings validate our approach along many dimensions, although to some extent the causal effect might be reverse, i.e. entrepreneurial success leads to the adoption of better management practices. The second striking feature is that the constrained gazelles are in many cases very close to the top-performers and in almost all cases clearly different from the survivalists. Constrained gazelles clearly work more hours than survivalists, they keep more often books, have a much higher financial literacy, react better to demand shocks and search actively for new clients; the latter as often as the top-performers. Moreover, similar to top-performers, the constrained gazelles have a much lower preference for hiring family members and a much higher preference for hiring workers from the market. Constrained gazelles are also more often than survivalists member of a business association and consider more often to formally register their business. Finally, constrained gazelles show the same level of risk aversion - as imperfect as this measure might be - as the top-performers. Here as well, the significant difference in the management behaviour of constrained gazelles compared to survivalists and their closeness to the top-performers can be seen as an ex-post validation of our approach. Although survivalists and constrained gazelles have similar levels of capital stock, constrained gazelles show much higher management skills and seem to be more entrepreneurial than survivalists. Again, all the skills we consider here were not used as a criterion in the initial classification; hence it is not obvious to find these patterns.

Taken together the results in Table 4 suggest that constrained gazelles are entrepreneurs who show a clear potential to become top-performers. They are mainly constrained by their external environment, i.e. access to finance, to a fixed business location and to basic infrastructure. However, internal constraints do not seem to be binding or at least to be the main problem. We think it is this feature, which makes this group so interesting for possible policy interventions. Survivalists in contrast do not only face serious external constraints; they would also need policy interventions that improve their education and skills, i.e. interventions with long-term targets that are likely to be much more costly.

(f) Extended partitioning

Based on the behavioural differences introduced above we have repeated the partition procedure, i.e. based on the probit model specified in Equation (1) augmented by behavioural factors. This revision allows for a finer classification of firms and for isolating a narrower group of constrained gazelles. Such an augmented model implies a larger variance in the distribution of predicted probabilities. We can also expect on average higher predicted probabilities of being a top-performer among top-performers since we better account for their strengths. This in turn will lead – given our partition criterion requiring identical average predicted probabilities of top-performers and constrained gazelles – to a smaller share of constrained gazelles and, consequently, a larger share of survivalists. Both partitions together can be interpreted as an upper and lower bound estimate for the share of constrained gazelles. The modified partition is shown in the Appendix (Table A2); between 20% and 30% of all informal entrepreneurs can be identified as constrained gazelles. However, for the remainder of our paper, we further use the initial partition performed in Section 4(b), but get back to the extended partition in our conclusion.

5. RETURNS TO CAPITAL

In this section we estimate marginal returns to capital for all firms together and for each of the three segments of firms separately. We are particularly interested in whether such estimates confirm an untapped growth potential of constrained gazelles relative to survivalists. The regressions explain value added by capital and labour and a number of characteristics of the entrepreneur including proxies for management skills. Estimating such functions will mainly answer two important questions: First, to what extent do the above described differences between the three groups result in higher value added, i.e. how important is education, knowledge of French and a certain managerial attitude? Second, do the returns to capital differ between the three groups under consideration? This is important as capital is generally seen as a crucial input and bottleneck targeted by many policy interventions (e.g. micro-credit programs).

We estimate the value added function separately for each group using our initial partition. However, we need to make one modification to this partition. Given the definition of top-performers (high capital stock and high value-added to capital ratio) the group of constrained gazelles exhibits a capital-value added scatter plot, in which the upper-left corner is virtually empty, since the firms located there are classified as top-performers. This may lead to a biased estimate of the return to capital as it is difficult to adjust a regression line to such a selected sample. Hence, in order to perform these regressions, we split the group of top-performers and constrained gazelles differently. For each country we compute the median capital stock

among the group of top-performers. We then attribute all firms with a capital stock above that limit to the group of top-performers independent of their capital productivity and all firms below that limit to the group of constrained gazelles. This procedure reclassifies 540 constrained gazelles to the group of top performers and 322 top-performers to the group of constrained gazelles. This manipulation does not change significantly the structure of the group of constrained gazelles. The changes are slightly larger for the group of top-performers, but also far from being substantial.

The value added function we estimate reads as follows:

$$\ln va_{ij} = \pi_{0j} + \pi_{1j} \ln[\max(K_{ij}, 1 \times NOK_{ij})] + \pi_{2j} NOK_{ij} + \pi_{3j} \ln L_{ij} + X_{ij}^{'} \pi_{4j} + M_{ij}^{'} \pi_{5j}$$
$$+ S_{ij}^{'} \pi_{6j} + E_{ij}^{'} \pi_{7j} + \xi_{ij}, \tag{5}$$

where $\ln va_{ij}$ stands for the log of value added of firm i in segment j. The segments refer to our groups of survivalists, constrained gazelles and top-performers, respectively. Ln K_{ij} stands for the log of physical capital. Given that many firms do not use any capital, we set ln K_{ij} to zero for those firms and introduce an additional dummy variable, NOK_{ij}, that takes the value one for the same. This ensures that we obtain unbiased estimates of the returns to capital in the presence of zero observations of capital (Battese, 1997). Ln L_{ij} stands for the monthly working hours including those of the firm owner and unpaid family members or other unpaid staff, such as trainees. The vector X_{ij} includes standard characteristics of the owner including age, age squared, education, knowledge of French, gender and the age of the firm. The vector M_{ij} includes the above discussed management skills. Finally, we control for sector (S_{ij}) and country effects (E_{ij}). For all three groups we first estimate Equation (5) with and then without M_{ij}.[vi]

The results are shown in Table 5. The first two columns show the coefficients we obtain if we estimate the value added function for all three groups together. The signs of all coefficients are as expected. The estimated capital elasticity is roughly 0.15. We note a clear positive association between education and value added. Most of the business skills are also significantly associated with value added. If we estimate the models for the three groups of entrepreneurs separately, a first striking finding is that the production elasticities of capital, $\pi_{1}j$, are high in all three segments, but continuously increasing if we go from the group of survivalists to the group of top-performers. The latter is significantly different for top-performers compared to constrained gazelles and for survivalists. However, the more relevant measure is the monthly marginal return to capital, i.e. the production elasticity times the actual capital productivity, $\pi_{1j} \cdot va_j/K_j$. To compute the monthly marginal return to capital, we use the average value added and capital stock observed in each segment. The marginal returns are shown at the bottom of Table 5. The marginal return is indeed highest among the constrained gazelles. If we increased the capital stock of a constrained gazelle by 10 Intl. $, this firm would increase its monthly value added by about 2 Intl. $, or 20%. The corresponding marginal increase of survivalists would only be 0.2 to 3 Intl. $, a fairly low return. The finding of high returns at very low levels of capital for the constrained gazelles is at odds with the typical assumption of a standard poverty trap model *à la* Banerjee and Newman (1993). It is important to highlight that this result is quite robust with respect to the functional form. For instance, we find very similar results if we estimate a more flexible specification in polynomials with capital and capital squared instead of the double-log specification. Finally, similar estimates are obtained using a semi-parametric approach that explores in even more detail the pattern of the apparent heterogeneity in returns. Moreover, the parametric estimates of the controls (in this semi-parametric specification) are very close

to those of the fully parametric specification, which again shows that the estimated parameters are not overly sensitive to changes in the specification. Finally, these results also hold if we specify and estimate a profit equation with value added minus all costs including the imputed value of the owner's time on the left-hand-side.[vii]

It is important to note that although returns of the top-performers are considerably lower than those of constrained gazelles, they are reasonably high with about 6% to 7%% per month. Put differently, the finding of lower returns at the top compared to the returns experienced by constrained gazelles should not be taken as an indication that investments do not pay off in the long run. The high returns for the constrained gazelles are rather the immediate consequence of the constraints these entrepreneurs face. Returns to labour are also high, but differ less across the three segments. Education effects are not very pronounced due to the fact that by construction each segment is relatively homogenous in these characteristics. They are in turn highly significant in columns (1) and (2), when the pooled sample is used. Moreover, value added increases with the firm's age (recall that firms can change ownership; hence the current manger's age and the firm's age are two different dimensions). Older firms are more productive - even when controlling for the accumulated capital stock and labour used. Whether these and the other effects are significantly different across groups can best be seen in Table 6 (see discussion below).

Gender effects again show a very interesting pattern. While women within the survivalists and to some extent within the constrained gazelles generate lower value added, top-performer women generate higher value added, ceteris paribus, than their male counterparts. About 20% of the top-performers are women and they are mainly in the manufacturing sectors and in the restaurants/hotels sector. Studying the driving forces of these gender effects is beyond the scope of this paper. Yet, recent studies that have examined these forces in more detail do, for instance, not find evidence for a lower inherent demand for finance. It rather seems that performance differences can be explained by gender gaps related to the use of financial services and education. In addition, female business is often only a secondary activity next to housework (Aterido *et al.*, 2011). Similarly, the obligation to combine housework with the market activity seems to force many women to operate their business from home, which often means to be far from input markets and clients. This latter aspect is also confirmed by our data. Whereas only 17% of the male entrepreneurs operate from hone, among women this share is 35%. All this may explain why women among the survivalists and constrained gazelles generate a lower value added. For Ghana, Fafchamps *et al.* (2011) also find female entrepreneurs to have a higher propensity to channel resources to household expenses and transfers instead of investment. Their findings also suggest that women who operate with larger capital stocks are quite successful and do take advantage of business opportunities. This is consistent with our results.

[insert Table 5 here]

The effects associated with the indicators we use to account for the management abilities are in most cases as expected. Practicing some form of accounting for instance increases value added by 30% to 50%. Again, we have to be careful with the estimated orders of magnitude as we do not attempt to control here for reverse causality. In addition, there are unobservable variables that can introduce a bias. However, the large set of introduced measures for skills and management attitude should dampen the typical unobserved ability bias. The capital elasticities are indeed slightly lower once we control for skills and management attitude.[viii] Being member of a business association, adjusting quickly to demand shocks and having an active approach to find new clients is typically associated with higher value added. Financial

literacy comes out as significant for the survivalists and the constrained gazelles, with a larger effect for the former. The coefficients associated with the ratio of non-paid staff to total staff are very interesting. While it has a positive sign for the group of survivalists, it is negative for constrained gazelles and top-performers, although for the latter statistically not significant. This suggests that non-paid staff – i.e. primarily family members – are less productive than hired staff in growth-oriented firms. This may partly be driven by behaviour (e.g. more difficult to monitor) and partly by skill differences. It also explains why a large share of top-performers and somewhat lower share of constrained gazelles reports to prefer to hire non-family members (cf. Table 4). Our measure of risk aversion is insignificant in all three regressions. The sector-specific level effects vary across the three groups. Survivalists achieve the highest value added in construction and restaurants/catering. Constrained gazelles are most productive in construction, as survivalists, and in transport. Among the top-performers, it is again construction, transport, retail shops and, surprisingly, petty trading.

Finally, to further analyze the sources of the differences in value added between the three groups, we perform a Blinder-Oaxaca decomposition (Blinder, 1973, Oaxaca, 1973). This decomposition allows attributing differences in outcomes to endowment and price effects. In our case this exercise decomposes group differences in value added into differences in factor use including capital, labour, skills and other characteristics, on the one hand, and, differences in the returns to these factors, on the other. The relative importance of so-called endowment versus price effects is important for the design of policies, as policies can either address endowments per se or the effectiveness of using them. Regarding the comparison between constrained gazelles and top performers, we expect differences in value added to be mainly explained by differences in capital, and possibly labour. In contrast, we expect that differences between survivalists and constrained gazelles can mainly be attributed to all endowments other than capital (and labour), such as education and management skills, as well as differences in the returns to these endowments.

The decomposition formula reads as follows:

$$
\begin{aligned}
\Delta \ln va = & [E(K_j, L_j, X_j, M_j, S_j, E_j) - E(K_k, L_k, X_k, M_k, S_k, E_k)]^{'} \pi_k \\
& + E(K_k, L_k, X_k, M_k, S_k, E_k)^{'} (\pi_j - \pi_k) \\
& + [E(K_j, L_j, X_j, M_j, S_j, E_j) - E(K_k, L_k, X_k, M_k, S_k, E_k)]^{'} (\pi_j - \pi_k),
\end{aligned}
\tag{6}
$$

where the first component is the endowment effect, the second component is the coefficient effect, and the third component is an interaction term. The latter accounts for the fact that differences in endowments and coefficients exist simultaneously. Note that the decomposition is formulated from the viewpoint of group k. If k stands for constrained gazelles it means that we compute the endowment effect by evaluating the contribution of endowments using the constrained gazelles' returns. Conversely, coefficient effects are computed alternatively the other groups' returns to evaluate the constrained gazelle's endowments.[ix]

[insert Table 6 here]

Table 6 shows the decomposition results of the comparison between constrained gazelles and survivalists (column 1) and between constrained gazelles and top-performers (column 2), respectively. On top figures the log difference in value added for these two group-comparisons. For the comparison between survivalists and constrained gazelles, we hence aim to explain the sources of the observed 1.021 difference in the average log annual value added between constrained gazelles and survivalists (the value added of constrained gazelles is

about 2.7 times higher). This difference can be attributed to endowment and coefficient components, and within these components, to single factors. The difference in log value added between constrained gazelles and survivalists can mainly be attributed to endowment effects (84%), while coefficient or price effects explain 36%. There is also a residual that contributes negatively to the observed difference. The endowment effect is mainly driven by the gender difference (constrained gazelles are more often operated by men), labour hours (constrained gazelles work more hours per month) and sector choice (constrained gazelles chose sectors with higher potential returns). The coefficient effect is mainly driven by a difference in returns to labour and to education. The 'returns to capital effect' is in favour of constrained gazelles, but is not statistically significant.

The decomposition for the comparison 'top-performers versus constrained gazelles' renders completely different results: Here, the entire difference in value added is explained by the endowment effect, in particular capital and labour. In contrast, the coefficient effect is largely driven by the return to capital and sector choice in favour of top-performers. This is partly offset by gender effects in favour of constrained gazelles. It should be noted that a significant effect implies that the effect of the corresponding variable is significantly different with respect to its effect on value added between the two groups of entrepreneurs considered.[x]

6. Policy Implications

Informal small-scale economic activities account for an important share of production in many developing countries. The contribution of these activities to employment is even larger, in some cases exceeding 80% of total employment. Early views on structural change and economic growth often posited that a modern sector consisting of larger firms would absorb those employed in a traditional or informal sector. This has happened to some extent in some developing economies, but not to the extent implied by earlier development theories. Moreover, especially in somewhat more dynamic economies of Latin America and (South) East Asia, many informal entrepreneurs choose to stay informal, and should not be perceived as waiting for a job in the formal sector (Maloney 2004). In contrast, in less dynamic economies with a less pronounced missing middle in particular in Africa (Fafchamps 1994, Tybout 2000), much larger shares of involuntary survivalist entrepreneurs are likely to be found, who would prefer a relatively secure formal sector job if only such jobs were available. In addition, and very urgent from a policy perspective, millions of young people will enter the labour market in the developing world in the next two decades, in particular in Africa, the Middle East and South Asia, and many of them will be absorbed in the informal sector. This scenario makes it necessary to rethink and challenge some common assumptions and views on informal entrepreneurs and to improve the integration of informal sector with labour market policies.

Our main message to policy makers is that the typical informal entrepreneur, also in non-dynamic economies in Africa, should not too easily be labelled a survivalist waiting for a job opportunity, without entrepreneurial capacities or growth potential. We argue and show that among those entrepreneurs typically considered survivalists – mainly because they operate with very little capital and generate low profits in absolute terms – there is a substantial share of entrepreneurs with business skills and an entrepreneurial behaviour that resembles skills and behaviour of upper tier entrepreneurs – or top-performers as we call them. Our data clearly show that very high marginal returns to capital can be earned by many of the very small informal businesses. In our basic partition, the share of the constrained gazelles ranges from 28% (Benin) to 58% (Cote d'Ivoire). In our more refined partition, that explicitly

16

includes the distribution of managerial abilities of entrepreneurs, this share ranges from 19% (Togo) to 34% (Cote d'Ivoire). The two partitions can be seen as an upper and lower bound of the gazelles' segment's size. It would be naïve to think that an exact threshold could be defined. We rather want to emphasise that a substantial share of firms with very low capital stocks shows strong entrepreneurial dynamism.

These constrained gazelles are mainly constrained by their business environment, i.e. external constraints such as the lack of access to capital, insurance and productive infrastructure. Individual or internal constraints such as education, specific business skills are not the key restriction to these businesses. They already show a minimum of financial literacy, keep books and react flexibly to change in market conditions. The true survivalists in turn, lack all these capacities. Hence, our findings may be taken as an argument for providing these constrained gazelles with credit, savings devices, and insurance; the latter not necessarily to insure business-related risks, but rather to cover important household related risks, such as health shocks and death. Obviously, each of these policies needs to be examined carefully in its specific context, but such an assessment goes beyond the scope of this paper.

For survivalist in turn the policy menu needs to be broader as *both* business environment and individual constraints are binding. In this segment, we find in particular female entrepreneurs with very low education levels who routinely have to combine their entrepreneurial activity with housework. Amine and Staub (2009) rightly emphasize that the socio-cultural environment and prevailing norms place a particular burden on women who desire to become entrepreneurs or to expand an entrepreneurial business. This implies that this very comprehensive set of constraints needs to be taken into account when policies towards female survivalists are designed. Many male survivalist entrepreneurs would prefer and be better off with a job as a wage worker. Some of these jobs could be provided by constrained gazelles if their potential can be unlocked.

In terms of targeting, our analysis suggests that a relatively small set of commonly observable characteristics, such as education, age, language skills, sector choice, management practices, are probably sufficient to keep targeting errors within a reasonable range, in particular given the likely costs of more precise targeting mechanisms. In such an approach, we would expect that between a third and a half of all micro and small entrepreneurs would be able to make good use of credit and savings devices. Given that at least in the countries we focus on almost none of these services are available for this informal segment of the economy and given the significant numbers of new labour market entrants, we feel it is worthwhile to start taking steps in this direction.

7. Conclusion

The literature on urban informal entrepreneurs emphasises at least two types of heterogeneity. First, a more anthropological and sociological strand posits the existence of two qualitatively distinct and persisting groups of entrepreneurs in the informal sector, an upper tier with a growth-orientation versus a lower tier of entrepreneurs focused on survival. Various recent studies in a second more economic strand find heterogeneity in capital productivity among informal entrepreneurs, where entrepreneurs working with (very) low levels of capital stock earn very high marginal returns to capital. In this paper we combine these observations, using a representative sample of urban informal entrepreneurs from seven capital cities in West Africa.

We empirically identify a third segment of entrepreneurs who are neither top performers nor survivalists. This group – which we label 'constrained gazelles' – shares many characteristics with top-performers, they even show similar managerial abilities in running their firm, but they operate with substantially lower capital stocks. Their stock of capital is almost as low as that of most survivalists. However, they are much more productive and can thus earn much higher returns to capital than survivalists.

From a policy perspective, being able to isolate constrained gazelles from the other two segments of informal entrepreneurs is a potentially important step towards a more differentiated informal sector policy menu. Moreover, our data shows that identification of constrained gazelles could be straightforward, which makes effective targeting possible for this attractive additional target group for dedicated policy interventions.

Informal sector heterogeneity continues to defy easy generalizations (Mead and Morrisson 1996). Nevertheless, based on our representative sample from seven capital cities in francophone West Africa, we show that the relative size of the group of constrained gazelles relative to survivalists is related to the structural characteristics of the respective economies, such as the urbanisation rate, the weight of agriculture in the economy and the size of the public and formal private sectors. Generally, we find a higher share of constrained gazelles in the more dynamic economies. Arguably, the seven capital cities in West Africa belong to the economically less dynamic capitals in the developing world. Therefore, we expect to find at least similar shares of constrained gazelles in the informal sector of economically more dynamic developing countries.[xi]

Appendix

Definition of all variables used

[insert Table A1]

Modified partition of the informal sector into groups, using a larger set of explanatory variables

[insert Table A2]

References

Amin, M. (2009). Labor Productivity in the Informal Sector: Necessity vs. Opportunity Firms. Mimeo, World Bank, Washington D.C.

Amine, L.S. & Staub, K.M. (2009). Women entrepreneurs in sub-Saharan Africa: An institutional theory analysis from a social marketing point of view. *Entrepreneurship and Regional Development*, 21 (2), 183-211.

Aterido, R., Beck, T. & Iacovone, L. (2011). Gender and Finance in Sub-Saharan Africa. Are Women Disadvantaged? World Bank Policy Research Working Paper #5571, World Bank, Washington D.C.

Banerjee, A.V. & Newman, A.F. (1993). Occupational Choice and the Process of Development. *Journal of Political Economy*, 101(2), 274-298.

Battese, G.E. (1997). A note on the estimation of Cobb-Douglas production functions when some explanatory variables have zero values. *Journal of Agricultural Economics*, *48*, 250–252.

Belsley, D.A., Kuh, E. & Welsch, R. E. (1980). *Regression Diagnostics: Identifying influential data and sources of collinearity*. John Wiley, New York.

Berner, E., Gomez, G. & Knorringa, P. (2012). Helping a Large Number of People Become a Little Less Poor: The Logic of Survival Entrepreneurs. *European Journal of Development Research*, in press.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8, 436-455.

Brilleau A., Ouedraogo, E., & Roubaud, F. (2005). L'Enquête 1-2-3 dans les Principales Agglomérations de l'UEMOA: la Consolidation d'une Méthode. *Stateco*, 99, 15-19.

Brilleau A., Roubaud, F. & Torelli, C. (2005). L'emploi, le chômage et les conditions d'activité, Enquêtes 1-2-3, phase 1. *Stateco*, 99, 54-84.

Boston, T.D. & Boston, L.R. (2007). Secrets of Gazelles: The Differences Between High-Growth and Low-Growth Business Owned by African-American Entrepreneurs. *The Annals of the American Academy of Political and Social Science*, 613 (3), 227-236.

Chen, M., Vanek, J., Lund, F., Heintz, J., Jhabvala, R. & Bonner, C. (2005). Progress of the World's Women 2005: Women, Work and Poverty. New York: United Nations Development Fund for Women.

Cotter, J. (1996). Distinguishing between poverty alleviation and business growth. *Small Enterprise Development*, 7 (2), 49–52.

Cunningham, W.V. & Maloney, W.F. (2001) Heterogeneity Among Mexico's Micro-Enterprises: An Application of Factor and Cluster Analysis. *Economic Development and Cultural Change*, 50 (1), 131-156.

De Mel, S., McKenzie, D. & Woodruff, C. (2008). Returns to Capital in Microenterprises: Evidence from a Field Experiment. *Quarterly Journal of Economics*, 123(4), 1329-1372.

De Mel, S., D. McKenzie & C. Woodruff (2010). Who are the Microenterprise Owners? Evidence from Sri Lanka on Tokman versus De Soto. In Lerner, J. & A. Schoar (Eds.), *International Differences in Entrepreneurship* (pp. 63-87). Chicago: University of Chicago Press.

Djankov, S., La Porta, R., López-de-Silanes, F., & Shleifer, A. (2002). The Regulation of Entry. *Quarterly Journal of Economics*, 117, 1-37.

Fafchamps, M. (1994). Industrial structure and microenterprises in Africa. *The Journal of Developing Areas*, 29: 1-30.

Fafchamps, M., McKenzie, D., Quinn, S. & Woodruff, C. (2011). When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana. Mimeo, University of Oxford.

Fields, G.S. (1990). Labor Market Modelling and the Urban Informal Sector: Theory and Evidence. In D. Turnham, B. Salomé & A. Schwarz (Eds.), *The Informal Sector Revisited* (pp. 44-69). OECD, Paris.

Filmer, D. & Pritchett, L. (2001). Estimating Wealth Effects Without Expenditure Data –Or Tears: An Application to Educational Enrolments in States of India. *Demography,* 38 (1), 15-132.

Grimm, M., Krüger, J. & Lay, J. (2011). Barriers to entry and returns to capital in informal activities: Evidence from Sub-Saharan Africa. *Review of Income and Wealth,* 57, S27-S53.

Hart, K. (1972). Employment, income and inequality: A strategy for increasing productive employment in Kenya. Geneva: International Labour Organization.

House, W.J. (1984). Nairobi's informal sector: dynamic entrepreneurs or surplus labour? *Economic Development and Cultural Change*, 32 (2), 277–302.

International Labour Organisation (2002). Decent Work and the Informal Economy. Geneva: International Labour Organisation.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *Stata Journal*, 8, 453-479.

Kremer, M., Lee, J.N., & Robinson, J.M. (2010). The Return to Capital for Small Retailers in Kenya: Evidence from Inventories. Mimeo, Harvard University.

La Porta, R. & Shleifer, A. (2011). The unofficial economy in Africa. NBER Working Paper #16821, NBER, Boston, MA.

Lubbel, H. (1991). *The informal sector in the 1980s and 1990s.* OECD Development Centre, Paris: OECD.

Maloney, W. F. (2004). Informality Revisited, *World Development*, 32 (7), 1159-1178.

McKenzie, D.J., & Woodruff, C. (2006). Do Entry Costs Provide an Empirical Basis for Poverty Traps? Evidence from Mexican Microenterprises. *Economic Development and Cultural Change*, 55 (1), 3-42.

McKenzie, D. & Woodruff, C. (2008). Experimental Evidence on Returns to Capital and Access to Finance in Mexico. *World Bank Economic Review*, 22 (3), 457-482.

Mead, D.C., & Liedholm, C. (1998). The dynamics of micro and small enterprises in Developing Countries. *World Development*, 26 (1): 61–74.

Mead, D.C. & Morrisson, C. (1996). The Informal Sector Elephant, *World Development*, 24 (10), 1611–1619.

Nichter, S. & Goldmark, L. (2009). Small firm growth in Developing Countries. *World Development*, 37 (9), 1453–1464.

Oaxaca, R.L. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14, 693-709.

Rogerson, C.M. (1996). Urban poverty and the informal economy in South Africa's economic heartland. *Environment and Urbanization*, 8 (1), 167–179.

Sahn, D.E. & Stifel, D.C. (2000). Poverty Comparisons Over Time and Across Countries in Africa. *World Development*, 28 (12), 2123-2155.

Tybout, J.R. (2000). Manufacturing firms in developing countries: How well do they do, and why? *Journal of Economic Literature*, 38 (1), 11-44.

World Bank (2007). *Informality: Exit and Exclusion: Latin American and Caribbean Studies*. Washington: World Bank.

World Bank (2009), World Development Indicators 2009. CD Rom, World Bank, Washington D.C.

Tables and Figures

Table 1: Descriptive statistics

	All	1st (bottom) quartile	2nd quartile	3rd quartile	4th (top) quartile	Top Decile
Owner characteristics						
Male	0.500	0.353	0.369	0.582	0.697	0.779
Age owner	36.3	35.0	36.5	36.4	37.3	38.3
Speaks French	0.476	0.373	0.360	0.532	0.637	0.701
No diploma	0.689	0.740	0.785	0.673	0.559	0.499
Primary completed	0.200	0.161	0.151	0.220	0.268	0.283
Some secondary	0.053	0.041	0.035	0.053	0.082	0.094
Other post primary	0.058	0.057	0.029	0.054	0.091	0.124
Ethnic group 1	0.410	0.427	0.410	0.412	0.392	0.405
Ethnic group 2	0.203	0.162	0.191	0.198	0.262	0.275
Ethnic group 3	0.098	0.103	0.089	0.095	0.106	0.098
Household Characteristics						
Household size	6.883	7.350	7.027	6.838	6.311	6.232
Only informal firm	0.796	0.792	0.802	0.781	0.810	0.841
Public wage earner	0.103	0.113	0.094	0.113	0.093	0.067
Private formal wage earner	0.095	0.089	0.102	0.102	0.086	0.080
Other combination	0.007	0.007	0.002	0.005	0.013	0.014
Firm characteristics						
Age of firm	8.1	7.3	8.4	8.6	8.0	8.3
Ann. VA in intl.$ PPP	5370.3	2712.4	3052.1	5585.9	10098.2	16363.2
Monthly hours owner	201.5	184.8	191.4	203.3	226.6	235.7
Total monthly hours	337.6	203.5	244.7	351.6	550.1	645.6
Total staff incl. owner	1.7	1.1	1.3	1.8	2.4	2.8
Hired paid staff	0.205	0.036	0.067	0.214	0.503	0.707
No capital	0.138	0.541	0	0	0	0
Physical. cap. in intl. $ PPP	1004.0	4.2	39.3	207.2	3768.9	7923.4
Sector						
Clothing and apparel	0.108	0.072	0.041	0.172	0.146	0.123
Other manufact. and food	0.158	0.113	0.186	0.173	0.159	0.152
Construction	0.075	0.024	0.104	0.116	0.059	0.053
Wholesale/retail shops	0.101	0.111	0.090	0.089	0.113	0.123
Petty trading	0.274	0.447	0.371	0.162	0.115	0.073
Hotels and restaurants	0.063	0.014	0.069	0.099	0.069	0.058
Repair services	0.057	0.022	0.033	0.084	0.089	0.081
Transport	0.046	0.025	0.008	0.018	0.134	0.216
Other services	0.119	0.172	0.098	0.088	0.116	0.122
Country						
Benin	0.142	0.060	0.110	0.129	0.269	0.315
Burkina Faso	0.148	0.170	0.126	0.146	0.149	0.172
Cote d'Ivoire	0.151	0.086	0.158	0.201	0.160	0.142
Mali	0.149	0.231	0.138	0.127	0.097	0.085
Niger	0.113	0.131	0.151	0.104	0.067	0.054
Senegal	0.153	0.170	0.161	0.152	0.128	0.121
Togo	0.145	0.152	0.155	0.141	0.131	0.110
N	6558	1662	1606	1645	1645	656

Source: 123-survey, WAEMU 2001; own computations.

Table 2: Probability of being a top-performer (Probit model)

	Coefficients	Marginal effects
Age owner	0.049***	0.007
	(0.012)	
Age owner squared	-0.055***	-0.008
	(0.015)	
Male (=1)	0.600***	0.087
	(0.058)	
No diploma	Ref.	
Primary completed	0.210***	0.033
	(0.064)	
Some secondary	0.188*	0.030
	(0.101)	
Other post primary	0.158	0.025
	(0.100)	
Speaks French (=1)	0.165***	0.024
	(0.061)	
No other opportunity	-0.185***	-0.024
	(0.058)	
Age of firm	0.008***	0.001
	(0.003)	
Clothing and apparel	Ref.	
Other manufact. and food	-0.216**	-0.028
	(0.085)	
Construction	-0.287***	-0.034
	(0.100)	
Wholesale/retail shops	-0.282***	-0.034
	(0.096)	
Petty trading	-0.496***	-0.061
	(0.085)	
Hotels and restaurants	0.300***	0.051
	(0.105)	
Repair services	-0.193*	-0.024
	(0.108)	
Transport	0.108	0.017
	(0.108)	
Other services	-0.395***	-0.045
	(0.094)	
Country effects	yes	
Pseudo-R2	0.094	
N	6503	

Notes: * $p<0.10$, ** $p<0.05$, *** $p<0.01$, standard errors (in parentheses) are clustered at neighborhood level.
Source: 123-survey, WAEMU 2001; own estimations.

Figure 1: Distribution of predicted probability of being a top-performer
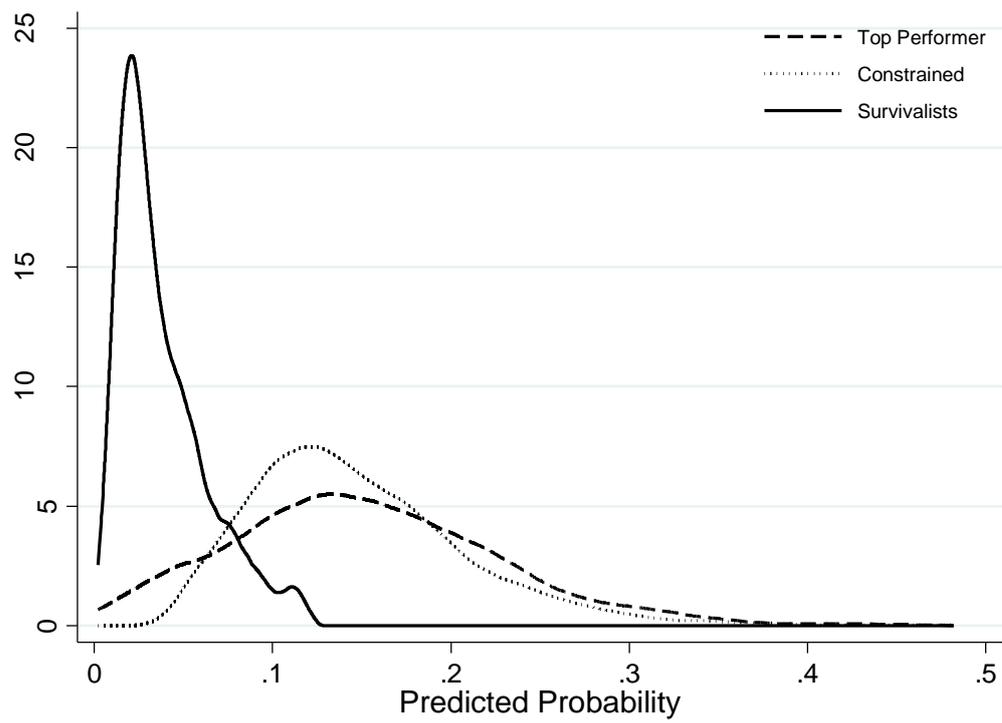
Source: 123-survey, WAEMU 2001; own estimations.

Figure 2: Distribution of capital
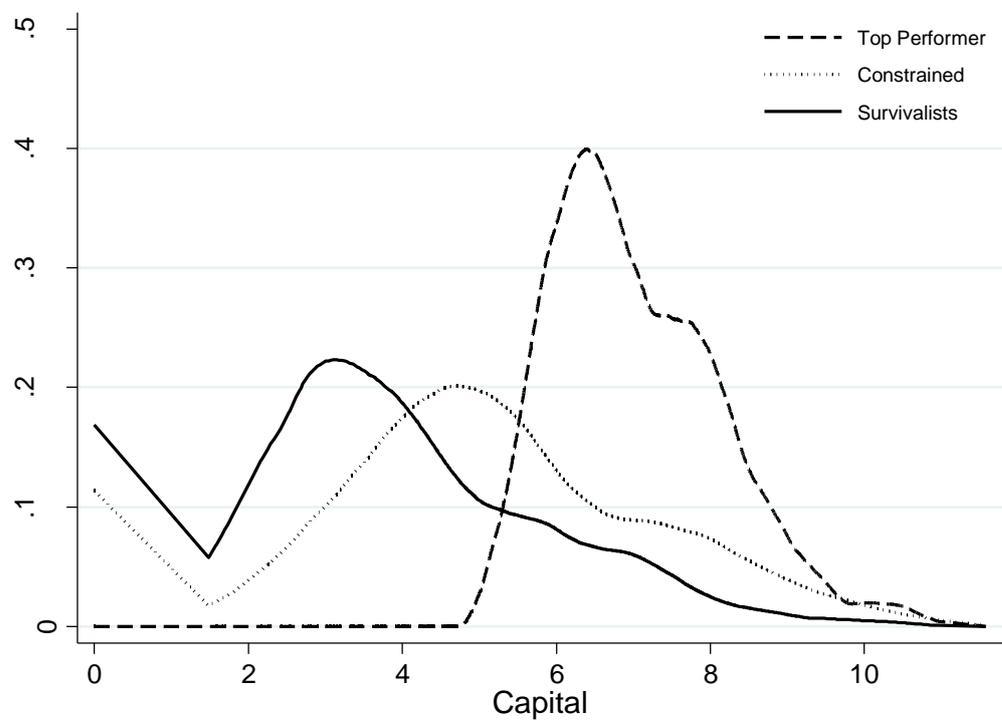
Source: 123-survey, WAEMU 2001; own estimations.

Table 3: Partition of informal sector and structural characteristics of regional labour markets

	Benin	Burkina Faso	Cote d'Ivoire	Mali	Niger	Senegal	Togo
Partition informal sector (shares)							
Top-performers	0.099	0.097	0.096	0.099	0.097	0.099	0.096
Constrained gazelles	0.280	0.434	0.587	0.539	0.581	0.439	0.348
Survivalists	0.620	0.469	0.317	0.362	0.322	0.462	0.556
Structural characteristics (2001)							
Geography	coastal	Sahel	coastal	Sahel	Sahel	coastal	coastal
Dominant religion	Christian	Muslim	Christian	Muslim	Muslim	Muslim	Christian
GNI per capita, PPP (current Intl. $)[a]	1170	850	1450	760	530	1330	680
Private/public sector employment, %[b]	19.7	26.6	25.3	22.5	29.9	23.6	19.0
Adult literacy rate, %[a]	35	22	49	24	9	39	53
Share w/t any schooling (10 y. and older), %[b]	20.2	33.7	28.5	40.8	36.3	31.4	16.1
Urbanisation rate, %[a]	38	17	44	28	16	41	37
Agriculture, value added % of GDP[a]	36	37	25	38	40	19	38
Time to enforce a contract (days)[c]	720	446	525	860	360	780	535
Business freedom[d]	70	55	55	70	55	55	40
Investment freedom[d]	50	70	50	70	30	50	30

Sources: [a] World Development Indicators, World Bank (2009), [b] 1-2-3 Survey, phase 1, Brilleau, Roubaud and Torelli (2005), [c] World Bank, Doing business survey, Djankov *et al.* (2002) and [d] Business Freedom Data Base/Heritage Foundation, Score between 0 (worst) – 100 (best) (http://www.heritage.org/index/explore) The partition of the informal sector is based on the 123-survey, WAEMU 2001 and own estimations.

Table 4: Group comparisonTable notes see, next page.

	(1)	(2)	(3)	(4)		(5)	
				t-Test		t-Test	
		Constr.	Top			(2) vs.	
	Survivalists	Gazelles	Perform.	(1) vs. (2)		(3)	
Basic characteristics							
Age owner	34.7	37.7	37.5	0	***	0.807	
Male	0.129	0.812	0.760	0	***	0.003	***
No diploma	0.807	0.601	0.564	0	***	0.084	*
Primary completed	0.142	0.240	0.277	0	***	0.049	**
Some secondary	0.030	0.070	0.074	0	***	0.725	
Other post primary	0.021	0.089	0.085	0	***	0.756	
Speaks French	0.337	0.583	0.620	0	***	0.080	*
No other opportunity	0.243	0.200	0.184	0	***	0.345	
Age of firm	6.4	9.4	9.5	0	***	0.758	
Sector							
Clothing and apparel	0.071	0.136	0.146	0	***	0.501	
Other manufact. and food	0.153	0.161	0.165	0.413		0.789	
Construction	0.015	0.129	0.109	0	***	0.168	
Wholesale/retail shops	0.094	0.109	0.096	0.062		0.363	
Petty trading	0.491	0.095	0.106	0	***	0.413	
Hotels and restaurants	0.035	0.084	0.089	0	***	0.731	
Repair services	0.013	0.094	0.089	0	***	0.661	
Transport	0.003	0.077	0.101	0	***	0.044	**
Other services	0.126	0.115	0.100	0.174		0.260	
Structural factors							
Firm set up with other person	0.032	0.065	0.083	0	***	0.097	
Public wage worker in hh	0.110	0.098	0.096	0.135		0.859	
Private formal wage worker in hh	0.112	0.081	0.079	0	***	0.884	
Household wealth index	-0.182	-0.056	0.342	0.010	***	0	***
Hands-on exper. larger firm	0.032	0.099	0.127	0	***	0.039	**
Water connection	0.095	0.074	0.138	0.004	***	0	***
Electricity connection	0.141	0.254	0.465	0	***	0	***
Telec. (land line)	0.038	0.097	0.209	0	***	0	***
Mobile location on the street	0.128	0.081	0.066	0	***	0.216	
Fixed location on the street	0.110	0.107	0.111	0.662		0.734	
Vehicle serves as location	0.002	0.033	0.040	0	***	0.397	
Firm located at home	0.304	0.206	0.141	0	***	0	***
Fixed location on market	0.181	0.183	0.210	0.838		0.108	
Shop/restaurant	0.052	0.147	0.278	0	***	0	***
Other location	0.113	0.285	0.372	0	***	0	***
By family tradition	0.068	0.079	0.046	0.129		0.004	***
Share same ethnic group in neighb.	0.402	0.367	0.383	0	***	0.181	
Largest ethnic group in country	0.424	0.402	0.386	0.089	*	0.461	
Second ethnic group in country	0.217	0.183	0.233	0	***	0.003	***
Third ethnic group in country	0.098	0.097	0.104	0.935		0.636	
Internal migrant	0.469	0.481	0.437	0.352		0.038	**
Immigrant	0.053	0.097	0.076	0	***	0.095	*
Return migrant	0.079	0.068	0.067	0.116		0.884	

Table continues next page.

Table 4 *... continued*

	(1)	(2)	(3)	(4)		(5)	
				t-Test		t-Test	
		Constr.	Top			(2) vs.	
	Survivalists	Gazelles	Perform.	(1) vs. (2)		(3)	
Behavioural factors							
Labour hours of owner	188.6	207.5	232.1	0	***	0	***
Book keeping	0.253	0.337	0.467	0	***	0	***
Financial literacy index	-0.156	0.074	0.397	0	***	0	***
Adjustment of supply to demand	0.649	0.684	0.745	0.004		0.002	***
New clients by publicity	0.255	0.243	0.253	0.296		0.572	
Active search for new clients	0.146	0.172	0.172	0.006		0.977	
Active reaction to demand shocks	0.683	0.686	0.762	0.761		0	***
Self-employment	0.830	0.629	0.352	0	***	0	***
Would hire family members	0.531	0.375	0.298	0	***	0	***
Would hire 'profess. from the market	0.120	0.304	0.355	0	***	0.011	***
Member of business association	0.026	0.054	0.141	0	***	0	***
Would register firm	0.222	0.442	0.611	0	***	0	***
Risk aversion	0.113	0.070	0.074	0	***	0.748	
N	2957	2965	643				

Notes: * p<0.10, ** p<0.05, *** p<0.01.
Source: 123-survey, WAEMU 2001; own computations.

Table 5: Drivers of performance

	All		Survivalists		Constrained perf.		Top performers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of capital	0.148***	0.137***	0.105***	0.089***	0.133***	0.126***	0.352***	0.331***
	(0.009)	(0.010)	(0.016)	(0.016)	(0.019)	(0.019)	(0.043)	(0.043)
No capital (=1)	0.404***	0.346***	0.113	0.073	0.482***	0.440***		
	(0.060)	(0.060)	(0.083)	(0.083)	(0.109)	(0.109)		
log total hours of labour	0.482***	0.458***	0.367***	0.329***	0.512***	0.512***	0.552***	0.538***
	(0.018)	(0.020)	(0.031)	(0.033)	(0.026)	(0.029)	(0.048)	(0.051)
Age owner	0.038***	0.029***	0.034***	0.023***	0.029**	0.019	0.002	0.003
	(0.007)	(0.006)	(0.009)	(0.009)	(0.012)	(0.012)	(0.024)	(0.023)
Age owner squared	-0.046***	-0.035***	-0.045***	-0.030***	-0.034**	-0.019	-0.009	-0.007
	(0.008)	(0.008)	(0.011)	(0.011)	(0.015)	(0.015)	(0.028)	(0.028)
Male (=1)	0.364***	0.337***	0.338***	0.295***	0.083	0.027	-0.570***	-0.566***
	(0.035)	(0.035)	(0.075)	(0.075)	(0.069)	(0.069)	(0.117)	(0.115)
No diploma	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Primary completed	0.075*	0.047	0.126*	0.041	-0.018	-0.050	-0.252**	-0.241**
	(0.044)	(0.044)	(0.076)	(0.077)	(0.062)	(0.062)	(0.110)	(0.109)
Some secondary	0.263***	0.153**	0.381**	0.174	0.122	0.049	0.171	0.011
	(0.072)	(0.072)	(0.148)	(0.151)	(0.094)	(0.094)	(0.159)	(0.158)
Other post primary	0.370***	0.197***	0.215	-0.075	0.284***	0.220**	0.049	-0.127
	(0.073)	(0.073)	(0.185)	(0.188)	(0.097)	(0.099)	(0.138)	(0.141)
Speaks French (=1)	0.122***	0.077**	0.037	0.011	0.124**	0.064	0.212**	0.189*
	(0.038)	(0.038)	(0.059)	(0.059)	(0.057)	(0.057)	(0.103)	(0.101)
Age of firm	0.016***	0.014***	0.019***	0.018***	0.012***	0.010***	0.008	0.008
	(0.002)	(0.002)	(0.004)	(0.004)	(0.003)	(0.003)	(0.006)	(0.005)
Book keeping		0.434***		0.479***		0.465***		0.279***
		(0.042)		(0.080)		(0.058)		(0.088)
Financial literacy index		0.040***		0.087***		0.017		0.006
		(0.010)		(0.016)		(0.014)		(0.022)
Adjustment of supply to demand		0.120***		0.141***		0.148***		-0.026
		(0.031)		(0.046)		(0.047)		(0.085)

Table continues next page.

Table 5 ... *continued*

	All		Survivalists		Constrained perf.		Top performers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New clients by publicity		0.078**		0.044		0.042		0.213**
		(0.035)		(0.052)		(0.053)		(0.097)
Active search for new clients		0.253***		0.301***		0.171***		0.151
		(0.040)		(0.064)		(0.061)		(0.102)
Active reaction to demand shocks		0.035		0.027		0.016		0.240***
		(0.031)		(0.047)		(0.046)		(0.082)
Faces management problems		-0.054		-0.087		-0.050		-0.036
		(0.089)		(0.136)		(0.151)		(0.191)
Share non-paid staff (excl. owner)		-0.034		0.135*		-0.174***		-0.077
		(0.041)		(0.074)		(0.058)		(0.090)
Member of business association		0.436***		0.493***		0.418***		0.187
		(0.069)		(0.141)		(0.100)		(0.116)
Risk aversion		-0.045		-0.021		-0.003		-0.047
		(0.050)		(0.069)		(0.087)		(0.134)
Clothing and apparel	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Other manufact. and food	0.390***	0.377***	0.313***	0.359***	0.470***	0.522***	0.597***	0.557***
	(0.056)	(0.056)	(0.101)	(0.100)	(0.078)	(0.077)	(0.141)	(0.139)
Construction	1.039***	1.026***	1.288***	1.216***	1.003***	1.024***	1.661***	1.686***
	(0.070)	(0.069)	(0.225)	(0.223)	(0.082)	(0.083)	(0.185)	(0.184)
Wholesale/retail shops	0.408***	0.396***	0.206*	0.269**	0.552***	0.535***	0.591***	0.561***
	(0.063)	(0.063)	(0.110)	(0.111)	(0.090)	(0.091)	(0.147)	(0.147)
Petty trading	0.290***	0.250***	0.295***	0.293***	0.433***	0.431***	1.147***	1.117***
	(0.052)	(0.053)	(0.090)	(0.090)	(0.090)	(0.091)	(0.188)	(0.187)
Hotels and restaurants	0.743***	0.722***	1.125***	1.097***	0.340***	0.308***	0.520***	0.435**
	(0.072)	(0.072)	(0.150)	(0.148)	(0.097)	(0.098)	(0.177)	(0.177)
Repair services	0.109	0.133*	-0.066	-0.072	0.125	0.204**	0.480***	0.449***
	(0.074)	(0.074)	(0.223)	(0.227)	(0.090)	(0.090)	(0.165)	(0.163)
Transport	0.745***	0.711***	0.582	0.695	0.732***	0.694***	0.807***	0.699***
	(0.081)	(0.081)	(0.774)	(0.765)	(0.115)	(0.115)	(0.146)	(0.151)

Table continues next page.

Table 5 ... *continued*

	All		Survivalists		Constrained perf.		Top performers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Other services	0.416***	0.420***	0.511***	0.528***	0.357***	0.408***	0.624***	0.544***
	(0.060)	(0.060)	(0.103)	(0.102)	(0.088)	(0.087)	(0.167)	(0.165)
Country effects	yes	yes	yes	yes	yes	yes	yes	yes
Test of equality of capital elasticity col. (6) vs. col. (4)						0.433		
Test of equality of capital elasticity col. (6) vs. col. (8)						0.000***		
R-squared	0.421	0.448	0.216	0.244	0.325	0.345	0.391	0.406
N	5757	5676	2580	2540	2384	2346	786	780
Mean yearly value added (Intl. $ 2001)	4083	4202	1673	1646	4675	5027	11620	11458
Mean capital stock (Intl. $ 2001)	919	924	351	350	253	253	5204	5160
MRK at mean (monthly), %	0.05	0.05	0.04	0.03	0.20	0.21	0.07	0.06

Notes: * p<0.10, ** p<0.05, *** p<0.01, standard errors (in parentheses) are clustered at neighborhood level. MRK stands for Marginal Return to Capital.
Source: 123-survey, WAEMU 2001; own estimations.

Table 6: Oaxaca Decomposition of differences in mean value added between groups

	Constr. vs. Survival		Top-perf. vs. Constr.	
Log annual value added				
Constr. / Top-perf.	7.719		8.580	
Survival / Constr.	6.698		7.708	
Difference	1.021	***	0.872	***
Endowment effects				
Capital	8%	***	41%	***
Labour (hours)	11%	***	33%	***
Exp./Age, Educ., Language	4%	**	3%	***
Male	20%	***	0%	
Age of firm	5%	***	0%	
Business attitude	7%	***	6%	***
Sector	14%	***	0%	
Country	16%	***	-5%	***
Total	84%	***	77%	***
Coefficients				
Capital	15%		74%	***
Labour (hours)	70%	***	-3%	
Exp./Age, Educ., Language	8%		-44%	
Male	-2%	*	-34%	***
Age of firm	-4%		-5%	
Business attitude	-7%		5%	
Sector	0%		24%	**
Country	-1%		-16%	
Constant	-44%		-55%	
Total	36%	***	-53%	***
Interaction	-17%	**	72%	***

Notes: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.
Source: 123-survey, WAEMU 2001; own estimations.

Table A1: Definition of all variables used

Owner characteristics

Male	Takes the value one if the entrepreneur is a man
Age owner	Age of entrepreneur in years
Speaks French	"Yes" answered to the question "Do you know to read and write in French?"
Education	Measured via highest diploma achieved in formal school system
Internal migrant	Entrepreneur migrated from another city or the country side to the capital.
Immigrant	Entrepreneur immigrated from abroad.
Return migrant	Entrepreneur was abroad, and returned to his/her country of origin.
Ethnicity	To ensure comparability across countries, we classify entrepreneurs whether they belong to the most frequent group in their country, the second most frequent etc.
Share same ethnic group in neighbourhood	Share of the population in the neighbourhood that belongs to the same ethnic group
No other opportunity	Entrepreneur started business because he/she could not find a job as a wage worker
By family tradition	Entrepreneur started business because it is a family tradition
Hands-on experience in larger firm	Entrepreneur worked previously in a larger farm as a trainee (larger than current firm)
Book-keeping	Entrepreneur keeps books of his/her activity
Financial literacy index	Index based on knowledge of credit and other financial products
Adjustment of supply to demand	Entrepreneur reports to adjust his/her supply to demand
New clients by publicity	Entrepreneur reports to undertake efforts to make publicity in his/her neighbourhood and family
Active search for new clients	Entrepreneur reports to actively look for new clients.
Active reaction to demand shocks	In response to a strong decline in demand, entrepreneur reduces his/her profit, diversifies and/or increases the quality of products
Would hire family members	If entrepreneur expanded he/she would hire family members
Would hire 'professionals' from the market	If entrepreneur expanded he/she would hire workers on a competitive basis
Business organization	Entrepreneur is member of a business organization
Register	Entrepreneur is in principle ready and willing to formally register his/her firm
Risk aversion	Dummy variable indicating the motivation for taking up the specific activity. Answer: "Assurance of more stable receipts than in other products" (no multiple answers possible)

Household characteristics

Household size	The number of all household members that stayed in the household for more than six months eat their meals together cooked in the same kitchen and accept the authority of the same household head
Household wealth index	First principal component of a principal components analysis, included variables: Housing characteristics (floor, number of rooms, electric light, tap water, sanitation facilities), TV, Hifi, Video, gas cooker
Only informal firm	Household does only have one or several informal firms, but no other sources of market income
Public wage earner	At least one household member is a wage earner in the public sector
Private formal wage earner	At least one household member is a wage earner in the private formal sector
Other combination	Any other combination, e.g. household has in addition a formal firm

Table continues next page.

Firm characteristics

Age of firm	Years since firm was established irrespective whether current owner was already involved
Firm set up with other person	Firm was established jointly with another person
Annual. VA in intl.$ PPP	Annual value added evaluated at 2001 international $. Value added is measured as turnover minus the costs for intermediate inputs; hence it includes all paid wages, the implied income of the owner and unpaid family members or other unpaid workers, such as trainees.
Monthly hours owner	Self-reported number of hours worked in firm last month
Total monthly hours	Self-reported total number of hours used in firm last month
Self-employment	Owner operates without any additional staff.
Hired paid staff	Number of hired staff that is paid
Share of non-paid staff (excl. owner)	Share of non-paid staff (excl. owner) in firm
Share of non-paid staff	Share of workers that are not paid
No capital	Firm operates without any physical capital
Capital in intl. $ PPP	Replacement value of all business-related assets, that have been used for the operation of the enterprise in the last year, including the business establishment, machines, furniture, vehicles and utilities
Water connection	Water connection available in firm
Electricity connection	Electricity connection available in firm
Telecommunication (land line)	Phone connection (land line) available in firm
Mobile location on the street	No fixed location for activity
Fixed location on the street	Fixed location on the street
Vehicle serves as location	Activity is operated from vehicle
Firm located at home	Firm is located at home.
Fixed location on market	Fixed location on a market
Shop/restaurant	Firm is a shop or a restaurant
Other location	Any other type of location, not listed above
Sector dummies	Take value 1 if firm is in respective sector (exclusive), 0 otherwise
Country dummies	Take value 1 if firm is in respective country (exclusive), 0 otherwise

Source: 123-survey, WAEMU 2001; own presentation.

Table A2: Modified partition based on augmented set of explanatory variables

	Benin	Burkina Faso	Cote d'Ivoire	Mali	Niger	Senegal	Togo
Partition informal sector (shares)							
Top-performers	0.099	0.097	0.096	0.099	0.097	0.099	0.096
Constrained gazelles	0.240	0.220	0.342	0.272	0.254	0.311	0.191
Survivalists	0.661	0.683	0.562	0.629	0.649	0.590	0.713

Source: 123-survey, WAEMU 2001; own estimations

[i] A segment is defined as a relatively homogenous group of entrepreneurs that share similar characteristics and are significantly different from other groups (Nichter and Goldmark, 2009).

[ii] The term 'gazelle' originates from the United States literature on high performance small firms (Boston and Boston, 2007).

[iii] It might also not apply to those entrepreneurs that see their activity clearly as a side activity that complements other income sources but is on purpose kept small.

[iv] The surveys were carried out by AFRISTAT and the National Statistical Institutes (INS) with the support of DIAL as part of the Regional Program of Statistical Support for Multilateral Surveillance (PARSTAT).

[v] Business associations mainly provide training and access to input and final product markets but usually they do not provide any finance.

[vi] When running the regression, we always correct standard errors for intra-cluster correlations. In total across the seven countries, there are 558 clusters in Phase 2. Moreover to reduce a bias due to measurement and reporting errors, we trim the data and drop influential outliers from our sample that we identify by the DFITS-statistic. As suggested by Belsley, Kuh, and Welsch (1980), we use a cut-off-value $\left|DFITS\right|_{ihj} > 2\sqrt{k/N}$ with k, the degrees of freedom (plus 1) and N the number of observations. Depending on the estimation, this procedure removes between 50 and 130 observations from our sample.

[vii] All these estimates can be obtained upon request from the authors.

[viii] For various robustness checks of similar estimates, see Grimm *et al.* (2011).

[ix] To perform this decomposition, we used the Stata macro 'oaxaca'. For details see Jann. (2008).

[x] For instance the fact that the labour hours effect in Table 6 makes a significant contribution to the explanation of the difference in value added between survivalists and constrained gazelles implies that the effect of labour on value added is significantly different for both groups (the coefficient in Table 6 is much higher for constrained gazelles compared to survivalists).

[xi] Indeed, applying exactly the same method we find similar evidence for Madagascar (41.4%), Vietnam (40.2%) and Peru (30.6%). For Madagascar and Peru the data is based on exactly the same type of questionnaire.