DISCUSSION PAPER SERIES

DP12401

SMALL FIRM DEATH IN DEVELOPING COUNTRIES

David J. McKenzie and Anna Luisa Paffhausen

DEVELOPMENT ECONOMICS

SMALL FIRM DEATH IN DEVELOPING COUNTRIES

David J. McKenzie and Anna Luisa Paffhausen

SMALL FIRM DEATH IN DEVELOPING COUNTRIES

Abstract

Small firms are an important source of income for the poor in developing countries, and the target of many interventions designed to help them grow. But there is no systematic information on the failure or death of such firms. We put together 16 panel surveys from 12 different developing countries to develop stylized facts from over 14,000 firms on how much firm death there is; on which types of these firms are most likely to die; and on why they die, paying careful attention to issues of measurement and attrition. We find small firms die at an average rate of 8.3 percent per year over the first five years of following them, so that half of all firms observed to be operating at a given point in time are dead within 6 years. Death rates are higher for small firms in richer countries, younger firms, retail firms, less productive and less profitable firms, and those whose owners are female and not middle-aged. We propose three theories of why small firms die: firm competition and firm shocks, occupational choice, and non-separability from the household. We find the cause of firm death to be heterogeneous, with different subgroups of firms more likely to die for reasons consistent with each of these theories.

David J. McKenzie - dmckenzie@worldbank.org
World Bank and CEPR

Anna Luisa Paffhausen - apaffhausen@worldbank.org
World Bank

Small Firm Death in Developing Countries[#]

David McKenzie, *Development Research Group, World Bank*
Anna Luisa Paffhausen, *University of Passau and Development Research Group, World Bank*

Abstract

Small firms are an important source of income for the poor in developing countries, and the target of many interventions designed to help them grow. But there is no systematic information on the failure or death of such firms. We put together 16 panel surveys from 12 different developing countries to develop stylized facts from over 14,000 firms on how much firm death there is; on which types of these firms are most likely to die; and on why they die, paying careful attention to issues of measurement and attrition. We find small firms die at an average rate of 8.3 percent per year over the first five years of following them, so that half of all firms observed to be operating at a given point in time are dead within 6 years. Death rates are higher for small firms in richer countries, younger firms, retail firms, less productive and less profitable firms, and those whose owners are female and not middle-aged. We propose three theories of why small firms die: firm competition and firm shocks, occupational choice, and non-separability from the household. We find the cause of firm death to be heterogeneous, with different subgroups of firms more likely to die for reasons consistent with each of these theories.

Keywords: firm death; survival; microenterprise dynamics
JEL codes: O12, O17, D22, L26

---

1. Introduction

Twenty-seven percent of the non-agricultural labor force in developing countries consists of self-employed business owners with no employees (Gindling and Newhouse, 2014), and 99 percent of the firms in many poor countries have 10 workers or fewer (McKenzie, 2017). These firms are an important source of income for the poor, and many policy interventions are designed to help people start and grow such firms. Yet there has been much less attention devoted to the death of such firms, with no systematic evidence available as to the rate of small firm death, which firms are more likely to die, and why they die. Almost two-thirds of published randomized experiments testing policy interventions for small firms in developing countries ignore firm death completely, neither reporting the death rate nor examining it as an outcome.[1]

Understanding the prevalence, characteristics, and causes of firm death is important for poverty, productivity, and policy. Since self-employment is an important source of income for the poor, firms shutting down could mean a large loss in income for firm owners. A growing body of literature (e.g. Hsieh and Klenow, 2009) has noted large misallocation of resources in developing countries. Firm death can improve aggregate productivity if less productive firms die and reallocate resources and customers to more efficient competitors. The optimal policy response to firm death then depends on whether it involves large income losses for the poor, and whether it is efficiently selecting out the least productive firms or not.

This paper provides systematic evidence on firm death in small firms in developing countries by collating data on more than 14,000 small firms from 16 firm panel surveys in 12 countries, enabling estimation of the rate of firm death over horizons as short as 3 months, and as long as 17 years. This overcomes many of the limitations of the existing literature on firm death in developing countries, which have had to rely on a small number of time periods, usually just in a single country, often with small samples of micro firms, and high rates of attrition.[2] Moreover, we included detailed questions in nine of these surveys to measure cause of death, and most of our surveys also continue to track the firm owner after their firm has died to provide data on what they do once their firm dies.

---

[1] Table A.1 provides details from 35 published randomized experiments testing different interventions for small firms in developing countries. Out of these 35 experiments, only 12 report the firm death rate and 13 examine firm death as an outcome.

[2] We discuss how firm death has been measured in the existing literature in the next section, and in Appendix Figure A.2. and Table A.3, compare our estimates of death rates to those in existing studies.

We begin by estimating the prevalence of firm death. Existing estimates of the rate at which small firms die in developing countries range from as low as 3 percent per year (Frazer, 2005; Davies and Kerr, 2015) to over 30 percent annually (Fajnzylber et al, 2006; Nagler and Naudé, 2017). We find the rate of firm death is approximately linear over time horizons up to five years, averaging 8.3 percent per year, but then declines over longer time horizons, so that firm death rates average less than 5 percent per year when looking over intervals of 10 years or longer. The result is that half the stock of firms operating at a given point in time will die within the next six years.

We then examine which firms are more likely to die, through examining the extent to which firm and owner characteristics predict firm death. Younger firms are more likely to die, with the first year particularly failure-prone: a firm in its first year is twice as likely to die in the next year as a five-year old firm. Retail firms and less profitable firms are more likely to die. Death rates are higher for younger and older entrepreneurs than for middle-aged owners, and higher for female-owners than males. Firm death rates are higher in richer developing countries than poorer countries.

Finally, we turn to why firms die. We consider three separate theories of firm death. The first views firm death as arising from firm competition and firm-level shocks causing firms to make losses and exit. Here firm death is involuntary, tends to cull less productive firms, and lowers the income of the owner. A second theory predicts firm death as arising from occupational choice decisions of the owner. Firm death can be voluntary here, with owners choosing to shut down when better outside opportunities arise, with less impact on poverty and less selectivity on productivity. A final theory is that firm death results from non-separability of business and household decisions due to imperfect markets. The result is that illness and shocks in the household can cause the business to have to close, resulting in income loss for the owner and not necessarily the least productive firms closing.

We bring these three theories to our data by examining the extent to which firm death occurs for more, or less, productive firms, measuring the change in total labor earnings of the owner after death, and using cause of death data. On average, firms that die are less productive to begin with, and their owners suffer a fall in labor earnings. This is most consistent with the firm competition theory, and least consistent with firm death being a voluntary occupational choice. However, we find considerable heterogeneity in the cause of death, with subsets of firms

appearing to die in ways consistent with each of these three theories. We characterize which types of firms are most likely to die for each of these three reasons.

The remainder of the paper is structured as follows: Section 2 discusses the different approaches to measuring firm death that have been used in the existing literature; Section 3 outlines our approach to firm death and the sample we have put together; Section 4 examines the prevalence of firm death; Section 5 which firms are more likely to die; Section 6 why firms die; and Section 7 concludes.

2. Measurement of Firm Death in the Existing Literature

At least four different approaches have been used to measure firm death in the existing literature, each with particular strengths and drawbacks.[3] A first approach is seen in some of the earliest empirical work on the topic, undertaken by Carl Liedholm and Donald Mead in the early 1990s in several sub-Saharan African countries (Liedholm and Mead, 1995; Mead and Liedholm, 1998). They carried out cross-sectional "closed firm" surveys by asking a random sample of households about enterprises they previously ran, but no longer operate, in addition to those they currently run. Based on this data, they report an average closure rate of 12.9 percent per year; although using the same data McPherson (1995) reports annual hazard rates of firm death of 3 to 4 percent per year. This approach has the advantage of allowing for large samples, and for representativeness of the areas in which sampling occurs. However, as the authors themselves acknowledge, it is likely to be much less accurate than panel surveys which track businesses over time, since people may forget or not want to talk about businesses that failed, and there is likely to be substantial recall error as to exactly when the business closed. Moreover, it will not capture people who close their businesses and then migrate out of the area.

A second approach has been to use manufacturing censuses or company registers. Examples include Roberts and Tybout (1996) who provide annual firm exit rates of 10.8 percent in Chile, 11.1 percent in Colombia, and 6.0 percent in Morocco; Bartelsman et al. (2009) who graphically show annual exit rates of approximately 4 to 5 percent in 10 Latin American and Eastern European economies; Klapper and Richmond (2011) who use a firm census of formal firms in Côte d'Ivoire and find annual exit rates of around 10 percent; and Shiferaw (2009) who reports

---

[3] A fifth, related, approach has been to use panel labor force surveys to track movements into and out of self-employment status (e.g. Demirgüc–Kunt et al. 2011; Fajnzylber et al, 2006). This does not enable examination of whether a particular firm survives, since individuals can shut down one firm and start another one, whilst remaining self-employed. Furthermore, labor force surveys typically do not contain information on the firm such as firm age, number of employees, or profitability.

annual exit rates of 16 percent for privately-owned Ethiopian manufacturing firms in the census. These censuses offer the advantage of providing population rather than sample data on the dynamics of larger firms. However, they are unsuited for examining the death of small firms for several reasons. First, many of the censuses are restricted to firms with 10 or more workers, and/or those which are formally registered, and they are extremely unlikely to capture non-visible businesses operating within households. Second, Shiferaw (2009) notes that it is not possible to distinguish firms that have closed from those that have switched sector out of manufacturing, or which have fallen below the cut-off point for inclusion. Third, they do not include firms in the retail and service sectors, which comprise the majority of small businesses.

A third approach is to use multi-purpose household living standards panel surveys such as the World Bank's Living Standards Measurement Study (LSMS). This approach was used in Vietnam by Vijverberg and Haughton (2004) who report 61 percent of firms dying over a five-year period; and McCaig and Pavcnik (2016) who find 30 to 35 percent death rates over two year periods. Kraftt (2016) uses the Egypt Labor Market Panel Survey to find 51 percent of firms died over an 8-year period, and 61 percent over a subsequent 6-year period. The advantages of these datasets are that the samples can be large and representative, and they capture micro and small businesses operating within households. However, these authors note several challenges to the use of these household surveys to track enterprises. Since the surveys do not identify specific enterprises by either name or code, authors need to match firms over time based on characteristics such as the age of the firm, identity of which household member runs the firm, and its sector. This can result in measurement errors if firms switch their sector of business or have trouble recalling the age of the firm. Further downsides are that the surveys are often conducted at irregular intervals, limiting the periods over which death rates can be measured, and often contain relatively little information on firm characteristics to enable exploration of which firms die and why.

The final approach is then to use dedicated firm panel surveys to track a sample of firms over time. The main challenge for this approach has been the dearth of firm panel surveys in developing countries, small sample sizes, and the difficulty of tracking firms with relatively low attrition. This approach was first tried by Mead and Liedholm (1998), who report an annual closure rate for micro and small enterprises in the Dominican Republic of 29 percent in 1992 and 22 percent in 1993, based on a sample of unspecified size with an unspecified attrition rate. They note that they also tried the approach in Zimbabwe, but were unable to relocate 42 percent

of firms. Panel surveys of manufacturing firms in several sub-Saharan African countries were collected as part of the Regional Program on Enterprise Development (RPED) by Oxford University and the World Bank, and have been used in several papers on firm dynamics (Frazer 2005; Söderbom et al, 2006; Sandefur, 2010). These surveys typically include 100-200 manufacturing firms per country, and do not include household enterprises. They report firm death rates of 6.3 percent over two years (Frazer, 2005) and 19 to 44 percent over 5 years (Söderbom et al, 2006), although they do not report attrition rates. Davies and Kerr (2015) conducted a 10-year follow-up survey of a random sample of 1000 firms surveyed in the 2003 Ghanaian manufacturing census. They could not find what had happened to 29 percent of the 386 firms with under 10 workers in their sample, while 25 percent had died.

This review of the existing literature shows that there has been relatively little evidence on the rates of firm death in the most prevalent types of enterprises found in developing countries: micro and small household enterprises that include retail and service firms in addition to manufacturing. Few of these existing studies measure attrition rates, nor do any provide bounds on what the rate of firm death will be when this is taken into account. A number of the studies provide descriptive information on what types of firms are more likely to die, and we will compare our findings on these correlates in Section 5. However, the fact that firm death is only a fraction of the sample, and that single country samples have often been relatively small, has meant that studies typically have relatively few deaths with which to examine the characteristics of who dies. For example, Frazer (2005) pools four rounds of firm surveys from Ghana, and still has only 30 deaths versus 479 survival episodes to compare; while Davies and Kerr (2015) have 95 deaths of firms with 0 to 9 workers.

3. Our Approach to Measurement and Our Sample

Our conceptual unit of analysis is the firm, rather than the entrepreneur, and our focus is on non-farm micro and small enterprises in developing countries. The vast majority of such firms have fewer than ten workers, with the modal firm in many developing countries consisting of just the owner with no paid workers (McKenzie, 2017). We define firm death as having occurred if a firm is open at one point in time, and then is reported as having shut down by the owner in a subsequent survey round. By shut down, we mean that the owner of the firm has decided to stop operating the firm, and no one else is operating it. It is not intended to include temporary closures of a few days or weeks that may occur when the owner is ill or away.

3.1 Putting Together a Large Panel of Micro and Small Firms

We draw on information on firms from both multi-purpose household and firm panel surveys in order to construct a large panel with which to examine death rates. Our resulting dataset combines data from 16 surveys in 12 developing countries, and includes a total of 14,426[4] firms at baseline, with a longitudinal dimension that is able to cover the dynamics of these firms in the short, medium and long term. Table 1 lists the surveys underlying the dataset, appendix A.1 discusses more details of each survey and how firm death was measured in each case.

Our inclusion criteria for using multi-purpose household panel surveys was to include nationally representative surveys that included a detailed non-farm business module, had low rates of attrition, a relatively large sample size, and enabled us to measure firm survival and death rates over at least three or more waves. This yielded four surveys. The first two are the Mexican Family Life Survey (MxFLS) and the Indonesian Family Life Survey (IFLS), both of which provide large samples of households. These surveys occur at low frequency, enabling us to measure firm death over horizons of three, four and a half, and eight years in Mexico, and seven, eight, and fifteen years in Indonesia. We examined the set of LSMS surveys to determine the ease at which they could be used to construct household firm panels. This led to the inclusion of the Nigeria General Household Survey (NGGHS), which allows measurement of firm death over 0.5, 1.5, 2, and 2.5 year horizons. Finally, we also include data from the Townsend Thai Project. This is a smaller sample, but resamples households annually, over an extended period of time, enabling firm death to be measured over periods from one to seventeen years.

These household panel surveys were not designed to track businesses over time, and do not link the businesses reported from one survey wave to the next. We therefore consider households that only operate one enterprise, and follow Vijverberg and Haughton (2004), McCaig and Pavcnik (2016) and Kraftt (2016) in using information on firm age, business ownership, and sector to attempt to link firms over time.[5] Since these household panels also contain data on individuals who are not operating firms in the first period, they enable us to also capture information on new firms that households start between one survey round and the next. We then pool together multiple survey rounds to get a larger sample, Appendix A.1. details these procedures in more detail.

---

[4] This number refers to the total number of unique firms we observe.

[5] The majority of households who operate a business operate only one business. In the 2000-round of the IFLS these are 80 percent, in the 2002-round of the MxFLS 93 percent, in the 2010-round of the NGGHS 58 percent, and in the 1997-round of the Townsend Thai Project 83 percent.

The second form of data comes from specialized firm panel surveys that have been collected as part of randomized controlled trials to assess different interventions designed to help micro and small firms in different countries. In each case we only use the control group data, in order to be able to assess firm death in the absence of any intervention. We searched publicly available datasets for impact evaluations targeted at microenterprises, but found most data made available only provided sufficient detail for replicating particular papers, and were not suitable for assessing survival dynamics. We use 11 panel surveys carried out in randomized trials conducted by one of the authors.[6] These include microenterprise surveys of informal firms used for formalization experiments in Benin, Malawi and Sri Lanka; surveys of microenterprises used in business training interventions in Kenya, Sri Lanka, and Togo; microenterprises that were part of cash grant interventions in Ghana and Sri Lanka; microfinance clients that were part of an insurance intervention in Egypt; business plan competition applicants in Nigeria; and microenterprises that were part of an intervention that gave wage subsidies, savings help, and business training in Sri Lanka. In addition, the one additional publicly available dataset which we were able to include was data from a program for female small businesses in Uganda from Blattman et al. (2016).[7]

Many of these specialized firm surveys are representative of urban microenterprises of particular size cutoffs or informality status, having been found through door-to-door listing surveys. Some of the samples also screen on the gender of the owner. The main exceptions are the business plan competition applicants in Nigeria, the microfinance clients in Egypt, and the Uganda study which was targeted at women in marginalized villages in Northern Uganda.

These specialized firm surveys collect data at relatively high frequencies, with a number collecting data at quarterly or semi-annual frequencies. However, few of them follow firms for more than three and a half years. The exceptions are three surveys from Sri Lanka, which tracked firms at 5 to 6 years, and at 10 to 11 years in the case of the Sri Lanka Microenterprise Survey (SLMS). Table A.4. provides an overview of the follow-up periods provided by each survey.

---

[6] We excluded data from surveys with high attrition rates for which death rates could not be accurately calculated, including data from a grants experiment in Mexico and a formalization experiment in Brazil.

[7] The Uganda study targeted women, rather than firms, as the unit of observation. We take the sample of control group participants who started a firm and use this sub-sample to track firm dynamics.

3.2 Dealing with Attrition

A key threat to our ability to measure firm death rates comes from survey attrition, in which owners of businesses could not be found or refused to be re-interviewed in the follow-up survey. The studies we draw on mitigate this threat in two main ways. The first has been to conduct multiple follow-up attempts and track individuals who move, thereby reducing overall levels of attrition. Second, the specialized firm surveys usually record whether the business is still in operation or not, even if the owner refuses to be re-interviewed or is away from the business when survey attempts are made. Table A.5. summarizes the attrition rates in our combined dataset by period of measurement. Attrition rates are relatively low for this combined sample, at 4.8 percent for one year, 10 percent for two years, and 9.5 percent over three years, and are around 20 percent in many periods beyond 5 years, with some exceptions of lower attrition rates.

We then form bounds for death rates to account for attrition. A lower bound for firm death is obtained by assuming that all enterprises which were operating at baseline, and then have missing values for firm death status at follow-up survived. An upper bound for firm death is obtained by assuming that all these firms died.

3.3 Summary Statistics on Firm Characteristics

Table 2 provides baseline summary statistics of our sample of firms in order to provide a description of the types of firms for which we are able to examine firm death rates. The sample consists of small enterprises run by individuals with relatively low levels of education (only 6 percent have any tertiary schooling), and who have an average age of 43 years. 56 percent of the owners are female. The firms have been in business for a mean of 9 and median of 6 years, with just over half (52 percent) in retail, 31 percent in services, and 14 percent in manufacturing. Only 29 percent of firms have any paid workers, with a mean of 0.75 workers per firm. Conditional on having paid workers, the median firm has only 1 worker and only 10 percent have 5 or more workers. Mean (median) monthly profits are US$129 (US$63). The median firm is thus earning approximately $2.65 per day it operates.

4. How Much Firm Death is There?

The first question that we wish to use this data to answer is just what the rate of firm death is for small firms in the economy. That is, starting with a sample of existing firms today, how many of them will have closed a year from now, or five years from now? The answer to this

question is useful for understanding the rate of churn in these small firms, as well as for policymakers and researchers planning interventions to help these firms.

4.1 The Fitted Relationship Between Death Rates and Time

The surveys vary in their frequency, so that we cannot measure firm death for the same interval (e.g. one year) for all the different surveys. We therefore begin by calculating firm death rates for each of the 79 survey-time interval combinations in our data, and report these rates in Appendix Table A.2. along with an upper and lower bound that accounts for survey attrition. We then plot this data in Figure 1, using a point to indicate the death rate for a survey-period combination, and bands around the point to indicate the bounds on this rate once attrition is allowed for. We fit a quadratic to these points, to give the predicted relationship:

Firm death rate = 0.015 + 0.092*Years Elapsed -0.00215*Years Elapsed^2 R^2 =0.789

           [0.024] [0.013]   [0.00079]

where robust standard errors are shown in parentheses. This relationship suggests firms die at the rate of approximately 9 percentage points per year over the first three years of following them, with this rate then declining over time, to around 7 percentage points per year at five years and 4.6 percentage points a year at ten years.

The majority of our data (58 out of 79 survey-time intervals) comes from horizons of up to 5 years. Figure 2 focuses on this sub-sample to better visualize the data within this interval. We cannot reject linearity of the relationship between death rates and time over this range (p=0.837 on the quadratic term), so we fit the following linear relationship for periods of up to five years:

Firm death rate = 0.024 + 0.083*Years Elapsed R^2 = 0.409

           [0.027] [0.016]

That is, firms die at an average of 8.3 percentage points per year over the first five years. Fitting the same slope through the lower bounds of the intervals gives a slope of 6.9 percentage points per year, while fitting it through the upper bounds of the intervals gives a slope of 8.7 percentage points per year.

In appendix Table A.3. we collate together estimates of firm death in developing countries from other literature. Figure A.1 then plots these alongside with our data and fitted quadratic. The firm death rates from other studies are all relatively close to the fitted quadratic, showing that results are generally consistent across a number of contexts. The one exception comes from the ten year death rates of Ghanaian manufacturing firms reported in Davies and Kerr (2015), which are much lower than our data would predict. Their death rates would look more similar if we assume that the more than 20 percent attrition comes from firms that die.

It may also be of interest to some readers to benchmark these death rates against those in the United States. The main source of firm dynamics data in the U.S. comes from data from the Bureau of Labor Statistics and U.S. Census Bureau for firms with at least one paid employee. Among firms with 1 to 4 paid workers, the one year death rate over 1995 to 2016 averaged 19.7 percent. It was highest for younger firms, averaging 28.4 percent for firms aged 1 to 4 years, 20.8 percent for firms aged 5 to 9 years, and 16.6 percent for firms aged 10 or more years. Death rates were lower for firms with more workers, averaging 4.6 percent for firms with 5 to 9 workers and 4.0 percent for firms with 10 to 19 workers[8]. U.S. data also shows high exit rates from self-employment, with 28 percent of the incorporated self-employed exiting over a year (Schweitzer and Shane, 2016). This evidence suggests that firm death rates for small firms are higher in the U.S. than the averages seen in our sample.

4.2 The Half-Life of Firms and Variability

Using these fitted relationships, we can then estimate the half-life of firms that are observed to be in existence at a given point in time. From the quadratic fit, 50 percent of firms are predicted to die within 6.2 years, while from the linear fit, 50 percent are predicted to die within 5.7 years. By way of comparison, Klapper and Richmond (2011) find the median lifespan of formally registered firms in Côte d'Ivoire is five years.

Since the death rates are approximately linear over the first five years, we can annualize them to get a death rate per year. This enables examination of the variability in death rates across the different samples. The mean annualized death rate over the 58 survey-time intervals is 10.4 percent, with a standard deviation of 7.9 percent. The [10,90] range is [4.0%, 22.7%]. So there

---

[8] Source: Bureau of Labor Statistics: https://www.bls.gov/bdm/business-employment-dynamics-data-by-age-and-size.htm

is considerable variability in the death rates found. We explore determinants of this variability in the next section.

4.3 Firm Death Need Not Mean Permanent Exit From Self-Employment

As noted, our focus is on the firm. However, using our panel surveys we can also track what happens to the owner once the firm has closed. Of all owners that we observe closing a firm within a year, 20 percent are already operating a new firm again within the same year. Figure A.2. looks at these re-opening rates for the five surveys for which we have the largest numbers of closed firms. We see re-opening rates increase with time, with approximately 40 percent of owners of closed firms opening a different firm within 3 years, although this rate is lower in the Townsend Thai data than in the specialized firm surveys. We will return to this issue when examining the causes of firm death, but note the importance of being able to track specific firms rather than just self-employment status in order to accurately measure firm lifespans.

5. Which firms are more likely to die?

We next examine how firm death varies with several key firm and owner characteristics. The aim in this section is to provide evidence that helps answer descriptive questions about which types of firms are more likely to die. The next section then provides some theoretical structure to help better understand why certain firms are more likely to die.

We begin with examination of univariate associations, and then move to multivariate analysis. Since the previous section shows that unconditional death rates are approximately linear over a five-year period, we summarize some of the information graphically by calculating standardized annual death rates. Graphically, we focus on time horizons of up to 3 years, since this is the range over which we observe firm death rates in a wide range of surveys and countries. Standardized annual death rates are then calculated by dividing the death rate by the period of measurement, so that a death rate of 18 percent over two years will be shown graphically as an annualized rate of 9 percent.

We then use probit and proportional hazard models to complement this graphical analysis and to examine the multivariate associations of firm and owner characteristics with firm death. We estimate a probit regression specification of the following form:

$$Probit(death_{iT} = 1 \mid X_i) = \theta(X_i'\beta + \varepsilon_i) \qquad ()$$

with $death_{iT}$ a dummy indicating whether firm i dies over observation period T, X_i includes the firm and owner characteristics of interest, as well as the horizon T over which the firm is observed, and survey dummies, and ε_i being the error term. To account for the fact that the same firm enters the regression more than once if it is observed over multiple time horizons, we cluster the standard errors at the firm level. We restrict this analysis to time periods T of less than or equal to five years. This accounts for the majority of our observations and enables us to describe results in terms of annualized death rates since death rates are linear with time over this interval (see Figure 2).

The Cox proportional hazard specification we estimate has the following form:

$$h_i(t \mid X_i) = h_0(t)exp(X_i'\beta) \tag{1}$$

where $h_i(t)$ is the hazard at time t, i.e. the likelihood that firm i dies at time t, $h_0(t)$ is the baseline hazard, and X_i includes firm and owner characteristics and survey dummies. Again, standard errors are clustered at the firm level.

The choice of firm and owner characteristics X_i here was determined by a combination of commonly asked descriptive questions about which firms are more likely to die, insights from the existing literature, and data availability. While the existing literature has examined some of these correlations, the number of firm deaths observed in any single study is typically small, limiting their ability to examine multiple predictors of firm death.

5.1 Univariate Firm-Level Correlates of Firm Death

We begin with an exploration of univariate associations between firm and owner characteristics and firm death, to establish stylized facts. Figure 3 graphically displays the univariate associations between firm characteristics and firm death. Each point is the mean death rate or annualized death rate for all firms with a given characteristic for a survey-year combination. The first five columns of Table 3 then provide the marginal effects from probit estimation, after controlling for survey fixed effects and a quadratic in time.

Are younger firms more likely to die? Perhaps the most common stylized fact about firm death is that younger firms are more likely to die than older firms. Theoretically this can arise from

new entrepreneurs learning that they are inefficient and exiting as in Jovanovic (1982), from starting with a low level of investment and not being able to keep up with the advances of competitors as in Ericson and Pakes (1995), or from taking on high levels of risk and failing early as in Cressy (2006). Empirically younger firms have been found to be more likely to die in the United States (e.g. Bernard and Jensen, 2002; Dunne, Roberts and Samuelson, 1989; Evans, 1987) and in developing countries (e.g. Vijverberg and Haughton, 2004; Frazer, 2005; Shiferaw, 2009) although Söderbom et al. (2006) found no significant relationship between firm age and death rates, and Davies and Kerr (2016) find that firms in the middle age category (5-14 years old) are less likely to die than younger firms (less than 5 years old), while older firms (older than 14 years) do not have higher likelihoods of dying than younger firms.

Figure 3a shows the relationship between annualized death rates and firm age in our data. There is a negative relationship, with older firms less likely to die on average than younger firms. However, we also see considerable heterogeneity in death rates across the different surveys. Among firms that are less than 2 years old, we see annual death rates ranging from less than 10 percent to more than 40 percent. Column 1 of Table 3 shows the marginal effect of age on the likelihood of firm death in our probit estimation when a linear term in age is included. An additional year of firm age is associated with a 0.6 percentage point lower likelihood of dying. Column 2 allows for firm age to have a non-linear effect, adding a quadratic term in firm age, and also a dummy variable for being in the first year of the business. Both these additional terms are significant, and show that an additional year of age matters more when firms are very young. The predicted probability of dying in the next year is then 26.6 percent for firms in their first year, 17.1 percent for firms in their second year, 13.7 percent for firms that are five years old, and 11.5 percent for firms that are ten years old.

Are smaller firms more likely to die? Firm size is closely linked to firm age in many theoretical models, with firms starting small and either growing or dying. Lower death rates for larger firms do appear to be the case in studies which compare medium or large-sized firms to small firms. For example, Davies and Kerr (2015) find Ghanaian firms with 75 or more workers are 5 percentage points more likely to survive over ten years than firms with 0 to 9 workers. Likewise Söderbom et al. (2006) estimate that a firm with 50 workers is 7 percentage points more likely to survive than a firm with 10 workers in their data from Ghana, Kenya and Tanzania, and Shiferaw (2009) finds Ethiopian firms with more than 100 employees are more likely to survive than those with 10 to 30 workers. The relationship has been less strong when

examining smaller firms and household enterprises. Vijverberg and Haughton (2004) find larger firms less likely to die in Vietnam, whereas McPherson (1995) finds no significant relationship with firm size in Swaziland and Botswana, and even finds larger firms being more likely to die in Zimbabwe.

Figures 3b and 3c examine the relationship between firm death and firm size, considering two measures of firm size: employment, and profitability. Since the modal firm in our sample has no paid workers, and 99.4 percent have 10 workers or fewer, we compare firms with no workers to those with 1 to 10 workers in Figure 3b. There is no strong size effect apparent in this figure. Column 3 of Table 3 confirms this lack of employment size effect. Figure 3c shows firms earning more profits are slightly less likely to die than the smallest firms. Column 4 of Table 3 confirms this, with log profits being a significant negative predictor of firm death.

Are firms in certain sectors more likely to die? While many studies have just used data from manufacturing firms, studies which have included other sectors have found death rates to vary with sector. However, they come to different conclusions about which sectors have higher death rates. McPherson (1995) and Mead and Liedholm (1998) find that firms in the retail sector are more likely to close than firms in the services or manufacturing sector; Klapper and Richmond (2011) find manufacturing firms to face a higher risk of closing compared to firms in the services sector and Vijverberg and Haughton (2004) find manufacturing and services firms to be significantly more likely to close than those in the retail sector. Figure 3d shows that in our data, retail firms have the highest death rates, and manufacturing firms the lowest, with services in the middle. However, column 5 of Table 3 shows that these differences are not statistically significant.

5.2 Univariate Individual-Level Correlates of Firm Death

Figure 4 and columns 6 through 8 of Table 3 examine how the characteristics of the owner are correlated with firm death. These have been less studied in much of the literature, since many surveys of firms do not provide personal characteristics of the owner.

Are firm death rates higher for younger or older entrepreneurs? Younger owners have had less time to accumulate skills and capital, and are at an age where job mobility is also higher even in wage work. But older entrepreneurs may close as they near retirement, or see less time to recover from shocks. The result might be that survival probabilities are highest for the middle

aged. Fajnzylber et al. (2006) find this in the context of the likelihood of staying self-employed (not necessarily in the same business) over a year in Mexico, where the maximum is found for ages 36 to 50. Likewise, Vijverberg and Haughton (2004) find firm death rates to be lower for 26 to 45 year olds than for older or younger firm owners. Figure 4a suggests this inverse-U shaped pattern, which is confirmed in column 6 of Table 3. Firm death rates are highest for youth, fall until age 50, and then start rising again. The predicted annualized death rate averages 18.0% for 20 to 24 year olds, compared to 9.6% for 45 to 49 year olds.

Are male or female owners more likely to have their firms die? Female-owned firms are, on average, smaller and less profitable than male-owned firms in developing countries. Both these factors would suggest their firms would be more susceptible to failure. However, the existing literature has found mixed associations between gender and firm death. McPherson (1995) finds female-owned firms are more likely to close in two of the four countries he studies, but finds no significant relationship in the other two. In contrast, Shiferaw (2009) and Vijverberg and Haughton (2004) find female-owned firms to be less likely to die than male-owned firms in Ethiopia and Vietnam respectively. In contrast, female-owned firms are more likely to die in our data. Figure 4b shows that female-owned firms have higher death rates across the range of time intervals we consider. Column 7 of Table 3 shows this difference is statistically significant, with a 2.7 percentage point higher death rate for female-owned firms.

Are firms run by more educated owners less likely to die? To the extent that higher education also reflects higher ability, we might expect firms run by more educated owners to be more likely to survive. However, the existing literature has either found no relation between firm death and education (Vijverberg and Haughton (2004); Krafft (2016) in one time period), or that firms with more educated owners are actually more likely to die (Fajnzylber et al. (2006) for self-employment; Krafft (2016) in a second time period). Figure 4c shows that firm death rates appear higher for the more educated in our data. However, this appears to largely reflect differences across surveys. After controlling for survey-year fixed effects, column 8 of Table 3 shows that the firm death rate for tertiary educated owners is not statistically significant (p=0.366).

5.3 Multivariate analysis of the correlates of firm death

Column 9 of Table 3 then includes all these variables together to examine how they jointly predict firm death with the probit. Column 10 gives odds ratios from the analogous proportional

hazard model. The results are largely similar to the univariate results: firm death is higher for younger firms, especially those in their first year, lower for firms with higher baseline profits, lower for middle-aged owners compared to youth or older owners, and higher for female owners. Business sector, which was not statistically significant in the univariate specification, is significant in both the probit and hazard multivariate models. Manufacturing firms are 2.7 percentage points less likely to die than retail firms, while service firms are 1.7 percentage points less likely to die compared to retail firms. The number of employees also becomes significant at the 10 percent level in the hazard model, suggesting that larger firms are less likely to die. However, the odds ratio of 0.98 shows this effect is small in magnitude.

5.4 Do death rates vary with level of development?

All of the above analysis controls for survey fixed effects, which captures country-level differences across the samples. But it is still of interest to see descriptively whether firm death rates vary with level of development. On one hand, we might expect businesses in less developed countries to face more constraints such as less access to credit, and so be less likely to survive negative shocks. However, there may be fewer other options for employment in poorer economies, suggesting fewer voluntary exits. Figure 5 plots the observed correlation between annualized firm death rates and per-capita GDP. We see a positive relationship (correlation 0.41), with firm death rates higher in richer developing countries. Each log point increase in per capita GDP is associated with a 5.2 percentage point higher annualized firm death rate (p=0.004). This is also consistent with the evidence presented earlier in this paper suggesting that small firm death rates are higher in the U.S. than in our sample.

6.   Why do firms die?

The previous section shows that firm death does not simply occur at random, but is more likely to occur for certain types of firms and firm owners than others. We consider three competing theories of why micro and small firms may close down. These theories differ in the extent to which firm closure is a voluntary choice for the owner, and in terms of the relationship that can be expected between firm productivity and firm death. As a result, they lead to different conclusions about how firm death will affect aggregate productivity and individual welfare. We then relate these theories back to our data on which firms die, and use data on cause of death to help understand why firms die.

6.1 Three theories of why firms die

Firm competition and firm-level shocks: A first view of firm death takes the firm as the decision-making unit, and comes from firm competition and industrial organization models of firm dynamics (e.g. Hopenhayn, 1992; Ericson and Pakes, 1995). Firm exit occurs for two main reasons in these models. The first is the entry of more productive competitors, which take away market share from less productive incumbents, ultimately leading them to close down. Secondly, firms face exogeneous shocks, such as shocks in the demand for the firm's products. Firm closure is thus involuntary in this model, and is expected to lower the welfare and earnings of the firm owner since it arises from deteriorating firm conditions. In the absence of distortions and adjustment frictions, it will be the least productive firms which close down, resulting in an increase in aggregate productivity. However, distortions can allow some unproductive firms to remain in business and cause some more productive firms to exit (Bartelsman et al, 2013), weakening the extent to which firm death selects on productivity.

Occupational choice and positive shocks to the entrepreneur's outside options: the first view completely ignores the role of the firm owner or entrepreneur in decision-making. This may be appropriate for examining the death of larger firms with many owners, but for micro and small firms, the firm is tightly linked to the occupation of the owner, and it is rare for the firm to continue when the owner leaves.[9] This leads to examining firm death through theories of occupational choice (e.g. Evans and Jovanovic, 1989; Banerjee and Newman, 1993). In these theories individuals initially make their choice to be self-employed in a particular firm given their endowments of financial assets and human capital, and the prevailing returns to their skills and assets in alternative occupations. Business closure then occurs when relative returns change so that working in a wage job or starting a different firm becomes more attractive. This could occur through negative shocks to the existing firm, in which case we are back in the first theory, or through positive shocks to outside options such as an attractive wage job offer or idea for starting a different business. Firm closure is then voluntary, and will improve, or at least not lower, welfare and earnings of the firm owner. The extent to which firm death selects on productivity will depend on which types of entrepreneurs are most likely to receive these outside offers. Low productivity entrepreneurs may be close to the margin of preferring wage

---

[9] For instance, in the SLMS, only 1.5 percent of owners who stopped operating their business from round to round had sold it, while only 2.9 percent said the firm was being operated by another family member. In the Lomé Informal Enterprise Survey, 1.2 percent of owners who were not running their business anymore had sold it, and 1.8 percent had passed it to a family member. In the Ghana Microenterprises Survey, 7.5% of the businesses that were not being operated anymore by the original owner were being run by another family member, while no business had been sold.

work, so that small improvements in the wage market may induce them to quit. Conversely, high productivity individuals may receive more outside wage offers and also be more likely to think of ideas for alternative businesses they could run.

Non-separability of business and household decisions: in both of the first two theories, business decisions are made separately from household time allocation and consumption decisions of the business owner. However, the classic agricultural household model (e.g. Singh, Squire and Strauss, 1986; Bardhan and Udry, 1999) notes that production and consumption decisions may not be separable in the presence of missing markets. Adapting this idea to an urban context, with imperfect labor markets, illness of household members may require the firm owner to shut down the business in order to look after sick family members as they cannot hire someone else to do the job. Likewise, illness of the owner may lead to the business shutting down.[10] Firm closure is involuntary, and will be associated with lower earnings and welfare for the owner. The impact on aggregate productivity will depend on which types of businesses are most likely to shut down from household shocks. Gender norms may make women more likely to shut their business to look after other family members, and since female-owned businesses tend to be less profitable and less productive on average, this would lead to less productive businesses being more likely to die. Secondly, less profitable and less productive businesses may be closer to a viability margin below which it no longer is profitable to run the business. As a result, a temporary reduction in sales during a period of illness and the possibility of having to sell business assets to cover household expenses may result in closure for these less profitable firms.

6.2 Which theories are our data on correlates of firm death most consistent with?

The patterns found in section 5 for which firms are most likely to die can generally be consistent with any one of these three theories. For example, consider the finding that younger (and smaller, less profitable) firms are more likely to die. This is a direct prediction of the firm competition and firm-shocks literature, as young firms learn that they are inefficient or are unable to keep up with competition. But if we think of firm owners having heterogeneous likelihoods of receiving positive wage job offers or of experiencing negative health shocks, then death rates will be high for young firms as the highest risk firms exit, and then older firms will be those with lower risks of experiencing these shocks. Thus, the other two theories would also give rise to older firms being less likely to die. Likewise, the higher death rate for female-

---

[10] People with higher likelihoods of illness may also self-select into running household enterprises rather than more brawn-based activities such as farming (Adhvaryu and Nyshadham, 2017).

owned firms is a direct prediction from the non-separability theory, but could also arise as a prediction from firm competition theory since female-owned firms are less productive on average and so may find it harder to compete with new entrants; and could also occur in the occupational choice theory since females may be on average less attached to the labor market, and so may need less of a change in outside options to lead them to shut down their firm.

The U-shaped relationship between firm death and age of the owner, and the finding that firm death rates are higher in richer developing countries are most consistent with the occupational choice model. Young workers are likely to experience a higher arrival rate of alternative job offers than middle-aged workers, while older workers may be choosing between continuing to run the business and retirement or at least working less. Richer countries may offer a wider range of employment opportunities beyond subsistence self-employment, pulling people out of jobs.

The theories differ most in their predictions for the effect of firm exit on individual labor earnings, and in their predictions of the link between productivity and firm death. We therefore estimate the following regression for the outcome (Y) of log total labor earnings (the sum of wage earnings plus profits+1, in last month):

$$Y_{i,t} = a + b * FirmDeath_{i,t} + \sum \theta_{sj} 1(i\epsilon j, t = s) + \varepsilon_{i,t} \tag{3}$$

Where *FirmDeath* takes value one if firm i is dead at time t, and $1(i\epsilon j, t = s)$ are a set of dummy variables for survey j measuring follow-up in period s. The standard errors are clustered at the firm level. The coefficient b then measures the earnings loss or gain associated with the firm dying.

To test whether it is the more or less productive, or profitable, firms who die, we instead use baseline labor productivity (log of sales per worker) and log profitability, and estimate the regression:

$$Y_i = a + b * EverDie_i + \sum \theta_j 1(i\epsilon j) + \varepsilon_i \tag{4}$$

Where *EverDie* takes value one if firm i is dead by the last survey round (within 5 years) of a survey. The coefficient b then measures whether those firms who go on to die were lower or higher productivity and profitability to begin with than those that survive.

Panel A of Table 4 shows the results of estimating equations (3) and (4). In column 1, we see that firm death is associated with a sizeable fall in total labor earnings. The coefficient suggests

an average 84 percent fall in earnings of the owner when the firm dies. Columns 2 and 3 show that the firms which die are, on average, 17 percent less productive and 19 percent less profitable in the baseline period than those that go on to survive.

These results are not in line with the occupational choice theory in which firm owners are voluntarily closing down to pursue better opportunities. They are more in line with the firm competition and firm shocks theory, or non-separability theory. However, Figure 6 shows that this average hides considerable overlap in the productivity distributions of firms that die with those that survive (Figure A.3. shows this same pattern within surveys). There is not a sharp productivity cutoff below which death occurs, but rather many deaths from high productivity firms as well as many low productivity firms surviving. Similarly, while Figure 7 shows considerable mass in the distribution of labor income near zero for owners whose firms die, there are also firm owners who earn relatively high incomes after firm death. These results suggest that there may be some firms that die for reasons related to these alternate theories, even if the average is due to firm competition and firm shocks.

6.3 Cause of Death Data to Examine Heterogeneity in Reasons for Firm Death

To further explore the heterogeneity in reasons why firms die, and to see whether each of the theories seems applicable for different subsets of firms, we use cause of death data. This comes from direct questions asked of firm owners whose businesses have closed, and is asked in 9 of our 16 surveys. Table 5 summarizes what the owners report as the main cause of death of their firm, and what their main activity is after the firm dies. We see that the most common cause of death is in line with the first theory of being driven out by competition and firm shocks: 41 percent say they closed the firm because it was making a loss. The second most common cause is in line with the non-separability theory, with 26 percent closing for sickness or family reasons, while 11 percent give responses in line with occupational mobility theory, closing to pursue better wage opportunities or another idea for a firm.[11] Appendix Table A.6 shows that the cause of death varies by gender. Among male enterprise owners, 49 percent say that they closed their firm because it was making a loss, while among women 35.5 percent give this reason. For women, sickness (16 percent) or family reasons (18 percent) taken together are almost as common as making a loss, and only 9 percent of female owners give reasons related to the occupational mobility theory. We observe a different pattern for male enterprise owners,

---

[11] 22 percent give "other" as cause of death. This includes some firms in Sri Lanka, where the data for what "other" refers to was never entered, as well as reasons like the government shutting down the firm, disasters causing assets to be destroyed, and international migration.

of which 12 percent give reasons in line with the non-separability theory – and among which only 2 percent say they closed their firm because of family reasons. Instead, closing for better opportunities is the second most common reason of closure given by male enterprise owners (16 percent). Appendix Table A.7 strengthens the suggestion that non-separability is especially relevant for female business owners.

One might question whether these self-administered firm autopsies give reliable information on cause of death. A first check is then to see whether differences in cause of death are accompanied by differences in what the owner is doing after the firm dies. Table 5 shows this to be the case. Those whose firms died because they sought better opportunities are most likely to be in wage work or operating new firms; those whose firms died for family reasons or sickness are most likely to be doing housework and family care; while those whose firms died because the firm was making a loss are spread across a range of activities, including having the highest rate of still looking for a job of any of the causes of death.

We can then test whether different causes of death result in differences in selection on productivity and profitability, and in changes in labor earnings, in the way predicted by the three theories. Panel B of Table 4 re-runs the regressions in equations (3) and (4) separately by cause of death, while Figures 8 and 9 show the distributions of initial productivity and follow-up labor earnings in the month preceding a given survey round by cause of death. We see that the mean selection on labor productivity is similar, and negative, across all three causes of death. However, Figure 9 shows the variance is much higher for those closing for reasons of better opportunity – some very high productivity firms close for this reason, as well as some very low productivity firms. In contrast, the different reasons for closure are associated with very different changes in labor earnings. Those closing for illness or family reasons have a large mass earning zero, and an average 97 percent drop in labor earnings. Those closing because the business was making a loss also experience a large earnings fall, averaging 77 percent. In contrast, the earnings change associated with closing for better opportunities is not statistically different from zero. We see in Figure 8 that this group has a bimodal earnings distribution, with many earning higher incomes than survivors, while a larger proportion also have zero income compared to the survivors. This suggests that the better opportunities do not always work out, but that when they do, they do result in higher incomes.

6.4 What determines cause of death?

These results suggest that firms do close for all three of the theoretical reasons described above, but with closing for firm competition and firm shocks the most common. This then leads to the question of which types of firms are more likely to die for each of the theoretical reasons. To examine this, we estimate a multinomial logit model, where the outcome is firm operating status. The omitted (base) case is surviving, and the coefficients are then odds ratios for the likelihood of the firm dying for each of the three reasons (business loss, family and illness, better opportunities) relative to survival. We include the same firm and owner characteristics we used in Table 3 to enable comparability.

Table 6 presents the results. We see that firms are more likely to close because they are making a loss if they have lower baseline profits to start with, are in retail rather than manufacturing or the services sector, and if they are run by women. The chance of closing for this reason decreases with firm age, but at a decreasing rate. Firms are more likely to close for family reasons and illness if they have fewer employees and lower profits, and especially if they are run by female owners, with the chance decreasing with firm age, at a decreasing rate. In contrast, the strongest predictor of closing for reasons of better opportunity is if the owner has tertiary education, and if the firm is in the first year of business, with no significant size effect, nor a significant effect of firm age after the first year. There is no significant impact of gender on closing for this reason, and firms in manufacturing are less likely to close for reasons of better opportunity than retail firms.

Taken together with the results of the previous subsection, this highlights that the firm death experience differs by type of firm. One group of less productive and profitable firms, which are more commonly run by women, tend to die because they are making a loss, and suffer a loss in labor income for the owner when they do. Another group of less productive and profitable firms, with fewer employees, and even more likely to be run by women, tend to die because of the family reasons or illness of the owner, and result in an even larger drop in labor earnings for the owner. In contrast, a smaller group of firms, that tend to be young and run by highly educated owners, die because better opportunities come about for the owner, and there is not a large fall in owner labor earnings as a result.

7. Conclusions

Firm death is extremely common among small firms, with half of the current stock of small firms in a developing country likely to die within the next six years. Yet because panel surveys

of such firms are still relatively rare, often do not track firms for long periods, and often lose firms which die to attrition, systematic data on the rates, correlates, and causes of firm death have not been available. This paper overcomes these issues by pooling together data from 16 panel surveys in 12 countries, in which special efforts were made to limit attrition and ascertain the status of firms which closed down. Using this, we are able to provide stylized facts on the rate of firm death over time, provide descriptive evidence on which types of firms are most likely to die, and test three competing theories for why firms die. Our evidence suggests that the most common reason for firm death is that less profitable and less productive firms end up making losses and closing. However, other small firms, particularly those run by women, close because of illness and family reasons, suggesting non-separability between the household and firm, while a minority of firms, largely run by more educated owners, close because better opportunities arise for the owner.

We see several potential uses for this research. The first is as a benchmarking tool. There is currently little evidence for researchers or policymakers tracking the failure rates of firms in their study or intervention to compare to, to know whether the rates they see in their data are high or low. This paper provides guidance as to what death rate can be expected at different time horizons. The second potential use is for targeting of programs. Programs that want to work actively with businesses for several years may wish to avoid small firms at greatest risk of closing down, while other programs that are designed to help firms at risk of failure to survive may wish to use the information provided here to decide which firms to offer the program to. Finally, the results may be of interest in deciding whether interventions are needed, and what types of intervention. To the extent that firms are exiting because they are low productivity and competition is reallocating production to more efficient firms, policymakers may not want to intervene. Similarly, if firm death arises because of voluntary exit in pursuit of better opportunities, again no intervention may be warranted. In contrast, if firms die because of a lack of separability with household shocks, or because of firm shocks that were not able to be insured, policymakers may want to experiment with policies to insure these types of shocks. Our results suggest that there is a subset of smaller firms, run by women, for whom such policies may be of particular importance.

References

Adhvaryu, Achyuta and Anant Nyshadham (2017) "Health, enterprise, and labor complementarity in the household", *Journal of Development Economics* 126: 91-111.

Banerjee, Abhijit and Andrew Newman (1993) "Occupational choice and the process of development", *Journal of Political Economy* 101(2): 274-98.

Bardhan, Pranab and Christopher Udry (1999) *Development Microeconomics*. Oxford University Press, Oxford.

Bartelsman, Eric, John Haltiwanger and Stefano Scarpetta (2013) "Cross-country differences in productivity: The role of allocation and selection", *American Economic Review* 103(1): 305-34.

Bartelsman, Eric, John Haltiwanger and Stefano Scarpetta (2009) "Measuring and Analyzing Cross-country Differences in Firm Dynamics", pp. 15-76 in Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts (editors) *Producer Dynamics: New Evidence from Micro Data*, NBER, Cambridge, MA.

Benhassine, Najy, David McKenzie, Victor Pouliquen and Massimiliano Santini (2016) "Can enhancing the benefits of formalization induce informal firms to become formal? Experimental evidence from Benin ", World Bank Policy Research Working Paper 7900.

Bernard, Andrew B. and J. Bradford Jensen (2002) "The deaths of manufacturing plants", NBER Working Paper 9026.

Blattman, Christoper, Jeannie Annan, Eric P. Green and Julian Jamison (2016) "The returns to microenterprise support among the Ultrapoor: A field experiment in Postwar Uganda", *American Economic Journal: Applied Economics* 8(2): 35-64.

Campos, Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hillary Johnson, David McKenzie and Mona Mensmann (2017) "Teaching personal initiative beats traditional business training in boosting small business in West Africa", *Science*, forthcoming.

Campos, Francisco, Markus Goldstein and David McKenzie (2015) "Short-term impacts of Formalization Assistance and a Bank Information Session on Business Registration and Access to Finance in Malawi", *World Bank Policy Research Working Paper* no. 7183

Cressy, Robert (2006) "Why do most firms die young?", *Small Business Economics* 26: 103-116.

Davies, Elwyn and Andrew Kerr (2015) "Firm Survival and Change in Ghana, 2003-2013", CSAE Working Paper 2015-06.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2016) "Labor drops: Experimental evidence on the return to additional labor in microenterprises", *World Bank Policy Research Working Paper* 7924.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2014) "Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka", *Journal of Development Economics* 106: 199-210.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2013) "The demand for, and consequences of, formalization among informal firms in Sri Lanka", *American Economic Journal: Applied Economics* 5(2): 122-150.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2012) "One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka", *Science* 335(6071): 962-966.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2010) "Wage subsidies for microenterprises", *The American Economic Review Papers and Proceedings* 100(2): 614-618.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2009) "Are women more credit constrained? Experimental evidence on gender and microenterprise returns", *American Economic Journal: Applied Economics* 1(3): 1-32.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2008) "Returns to capital in microenterprises: Evidence from a field experiment", *The Quarterly Journal of Economics* 123(4): 1329-1372.

Demirgüc-Kunt, Asli, Leora F. Klapper, and Georgios A. Panos (2006) "Entrepreneurship in post-conflict transition. The role of informality and access to finance", *Economics of Transition* 19(1): 27-78.Dunne, Timothy, Mark J. Roberts, and Larry Samuelson (1989) "The growth and failure of US manufacturing plants", *The Quarterly Journal of Economics* 104(4): 671-698.

Ericson, Richard and Ariel Pakes (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62 (January): 53–82.

Evans, David S., and Boyan Jovanovic (1989) ''An Estimated Model of Entrepreneurial Choice under Liquidity Constraints.'' *Journal of Political Economy* 97(4):808–27.

Evans, David S. (1987) "The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries." *The Journal of Industrial Economics 35*(4): 567-581.

Fafchamps, Marcel, David McKenzie, Simon Quinn and Christopher Woodruff (2014) "Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana", *Journal of Development Economics* 106: 211-226.

Fajnzylber, Pablo, William Maloney, and Gabriel Montes Rojas (2006) "Microenterprise Dynamics in Developing Countries: How Similar are They to Those in the Industrialized World? Evidence from Mexico", *World Bank Economic Review* 20(3): 389-419.

Frazer, Garth (2005) "Which Firms Die? A Look at Manufacturing Firm Exit in Ghana", *Economic Development and Cultural Change* 53(3): 585-617.

Gindling, T.H. and David Newhouse (2014) "Self-employment in the Developing World", *World Development* 56: 313-31.

Groh, Matthew and David McKenzie (2016) "Macroinsurance for microenterprises: A randomized experiment in post-revolution Egypt", *Journal of Development Economics* 118: 13-25.

Hopenhayn, Hugo (1992) "Entry, Exit, and Firm Dynamics in Long-Run Equilibrium." *Econometrica* 60:1127–50.

Hsieh, Chang-Tai and Peter Klenow (2009) "Misallocation and Manufacturing TFP in China and India", *Quarterly Journal of Economics* 124(4): 1403-48.

Jovanovic, Boyan (1982) "Selection and the Evolution of Industry", *Econometrica* 50(3): 649-70.

Klapper, Leora and Christine Richmond (2011) "Patterns of business creation, survival and growth: Evidence from Africa", *Labour Economics* 18(1): S32-S44.

Krafft, Caroline (2016) "Understanding the Dynamics of Household Enterprises: Birth, Death, Growth, and Transformation", Economic Research Forum Working Paper Series 983.

Liedholm, Carl and Donald Mead (1995) "GEMINI Action Research Program I: Final Report" – The Dynamic Role of Micro and Small Enterprises in the Development Process, Gemini, Bethesda, MA.

McCaig Brian and Pavcnik Nina (2016) "Out with the old and unproductive, in with the new and similarly unproductive: Microenterprise dynamics in a growing low-income economy", Mimeo.

McKenzie, David (2017) „Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition", *American Economic Review*, 107(8): 2278-2307.

McKenzie, David and Susana Puerto (2017) "Growing markets through business training for female entrepreneurs: A market-level randomized experiment in Kenya" World Bank Policy Research Working Paper 7993.

McPherson, Michael A. (1995) "The hazards of small firms in Southern Africa.", *The Journal of Development Studies* 32(1): 31-54.

Mead, Donald and Carl Liedholm (1998) "The Dynamics of Micro and Small Enterprises in Developing Countries", *World Development* 26(1): 61-74.

Nagler, Paula and Wim Naudé (2017) "Non-farm entrepreneurship in rural sub-Saharan Africa: New empirical evidences", *Food Policy* 67: 175-191.

National Bureau of Statistics of the Federal Republic of Nigeria (2014) "Nigeria General Household Survey – Panel Wave 2 2012/2013", Available at http://microdata.worldbank.org/index.php/catalog/1952.

Roberts, Mark J., and James R. Tybout (1996) "A Preview of the Country Studies" pp. 188-199 in Roberts, Mark J., and James R. Tybout, (editors) *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*, Oxford University Press, New York.

Rubalcava, Luis and Graciela Teruel (2013) "Mexican Family Life Survey, Third Round", Available at: www.ennvih-mxfls.org.

Sandefur, Justin (2010) "On the Evolution of the Firm Size Distribution in an African Economy", CSAE Working Paper no. 2010-05.

Schweitzer, Mark E., and Scott Shane (2016). "The Ins and Outs of Self-Employment: An Estimate of Business Cycle and Trend Effects," *Federal Reserve Bank of Cleveland Working Paper*, no. 16-21.

Shiferaw, Admasu, (2009) "Survival of private sector manufacturing establishments in Africa: The role of productivity and ownership", *World Development* 37(3), 572–584.

Singh, Inderjit J., Lyn Squire, and John Strauss (1986) *Agricultural Household Models: Extensions, Applications and Policy.* John Hopkins University Press, Baltimore.

Söderbom, Måns, Francis Teal, and Alan Harding (2006) "The Determinants of Survival Among African Manufacturing Firms", *Economic Development and Cultural Change* 54(3): 533-55.

Strauss, John, Firman Witoelar, and Bondan Sikoki (2016) "The Fifth Wave of the Indonesia Family Life Survey (IFLS5): Overview and Field Report", WR-1143/1-NIA/NICHD.

Townsend, Robert M. (2016) "Townsend Thai Project Household Annual Resurvey, 2014 (Urban)", doi:10.7910/DVN/UOOUAZ, Harvard Dataverse, V1.

Van Biesebroeck, Johannes (2005) "Firm Size Matters: Growth and Productivity Growth in African Manufacturing", *Economic Development and Cultural Change* 53(3): 545-83.

Vijverberg, Wim and Jonathan Haughton (2004) "Household Enterprises in Vietnam: Survival, Growth, and Living Standards", pp. 95-132 in David Dollar and Paul Glewwe (eds.) *Economic Growth, Poverty, and Household Welfare: Policy Lessons from Vietnam*. The World Bank, Washington D.C.

Figure 1. Firm death rates over different time horizons
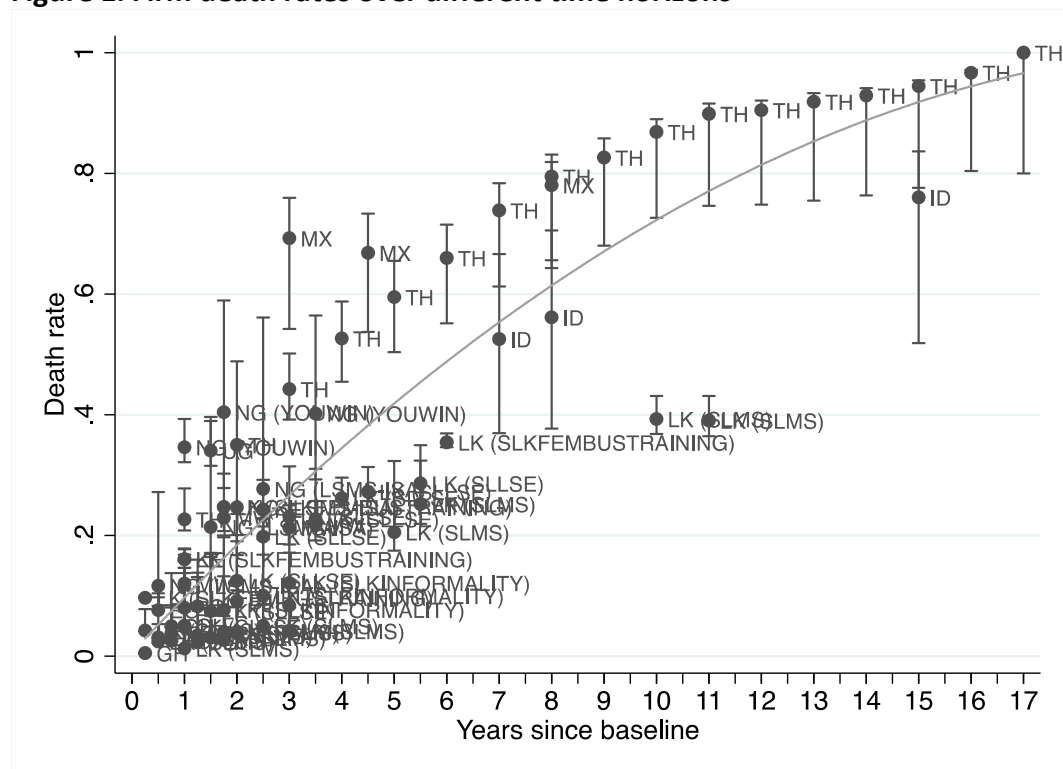
Figure 2. Firm death rates over different time horizons – up to five years
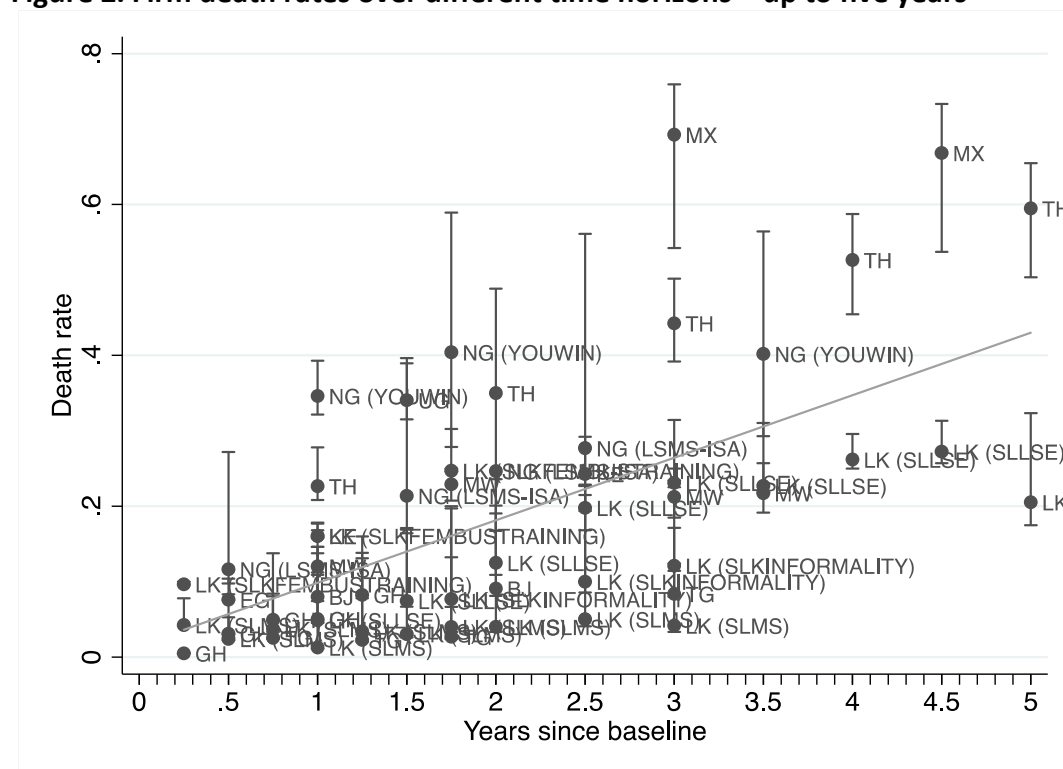
Figure 3. Death rates by firm characteristics

(a) Standardized death rates by firm age
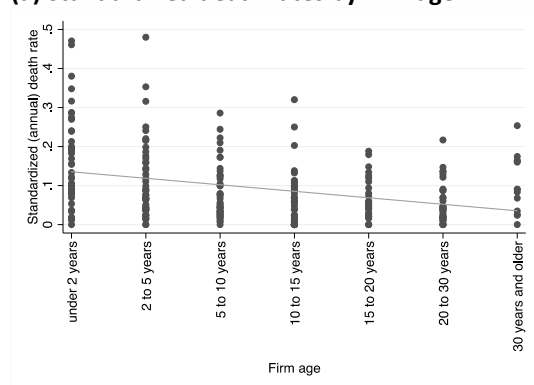
(b) Death rates by firm size (employment)
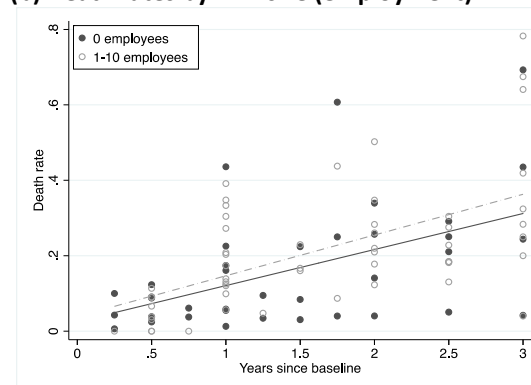
(c) Standardized death rates by daily profits
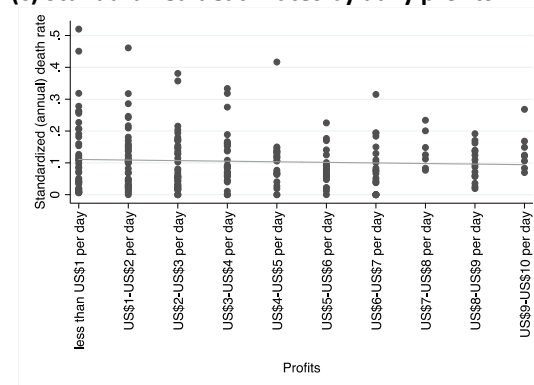
(d) Death rates by firm sector
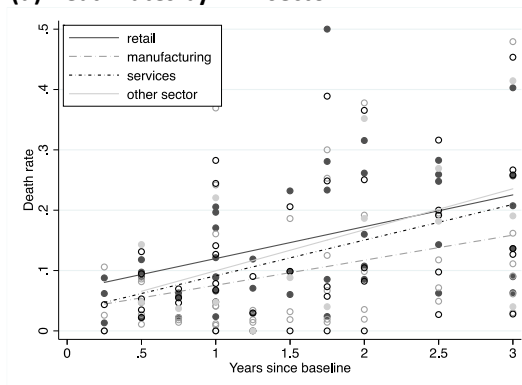
Figure 4. Death rates by owner characteristics

(a) Standardized death rates by age of owner
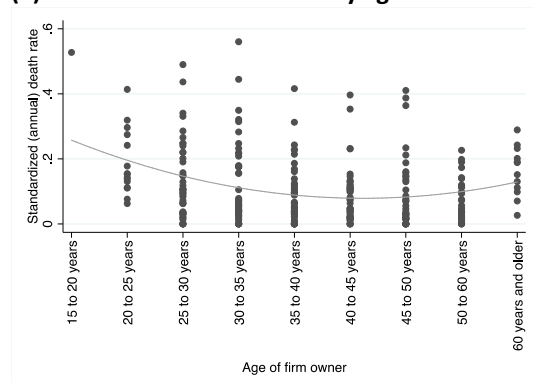
(b) Death rates by gender of owner
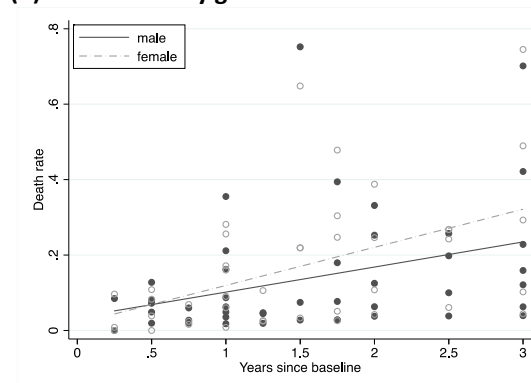
(c) Death rates by education of owner
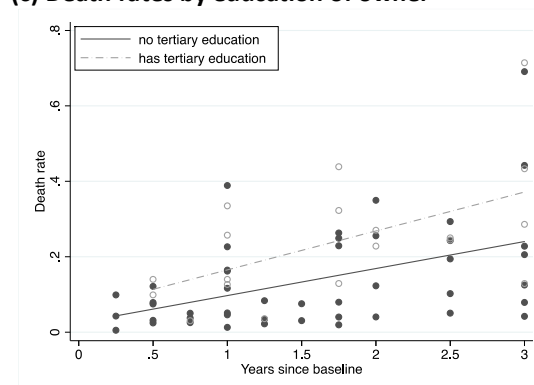
Figure 5. Death rates by per capita GDP
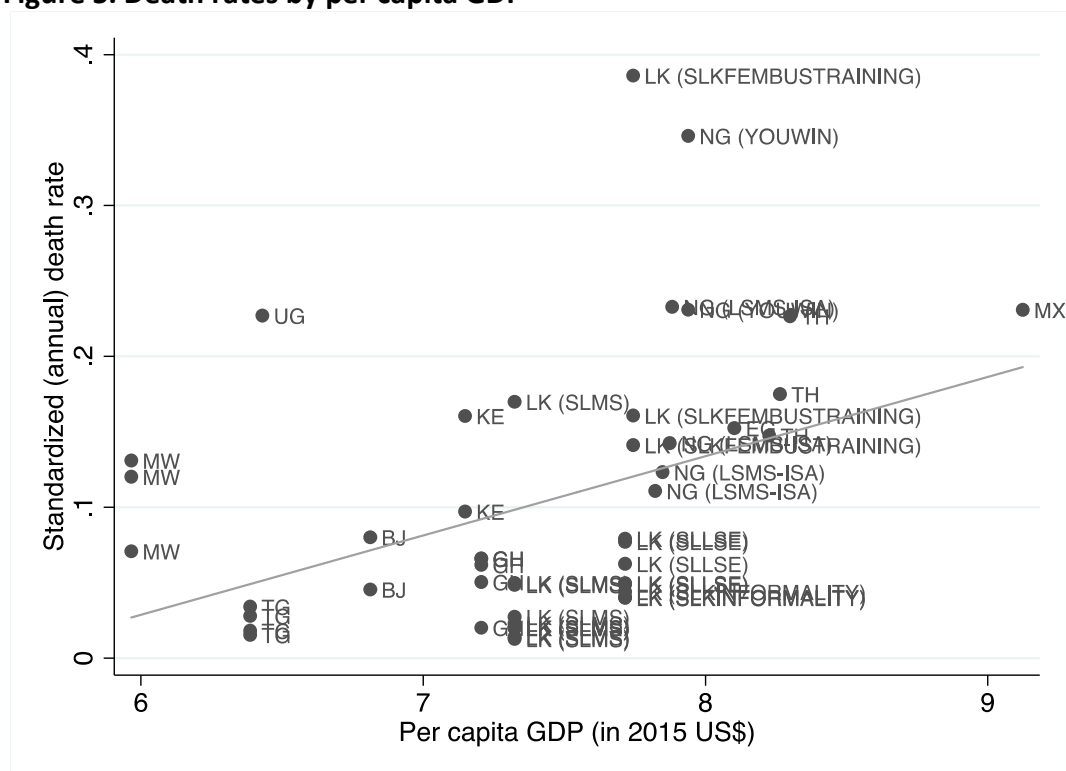
Figure 6. Comparison of labor productivity at baseline for firms that closed vs. survivors
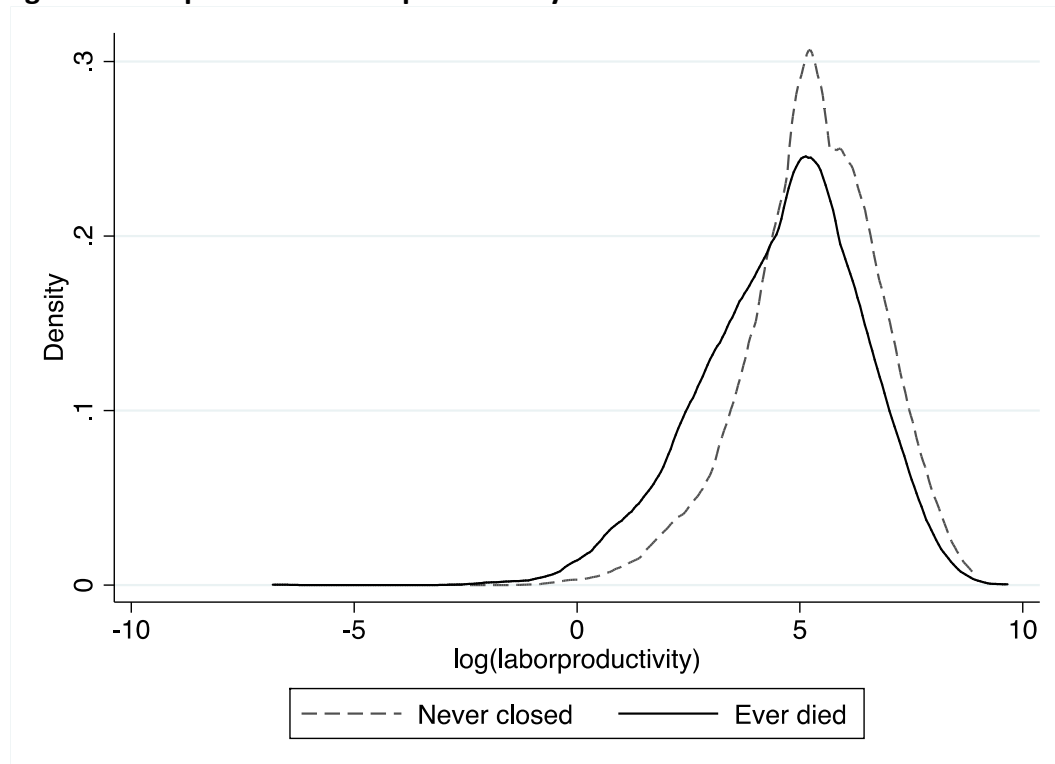
Includes data from BENINFORM, GHANAFLYP, IFLS, KENYAGETAHEAD, MALAWIFORM, MXFLS, NGLSMS-ISA, NGYOUWIN, SLKFEMBUSTRAINING, SLKINFORMALITY, SLLSE, SLMS, TOGOINF, TTHAI.

Figure 7. Comparison of monthly labor income at follow-up of owners of firms that closed vs. survivors
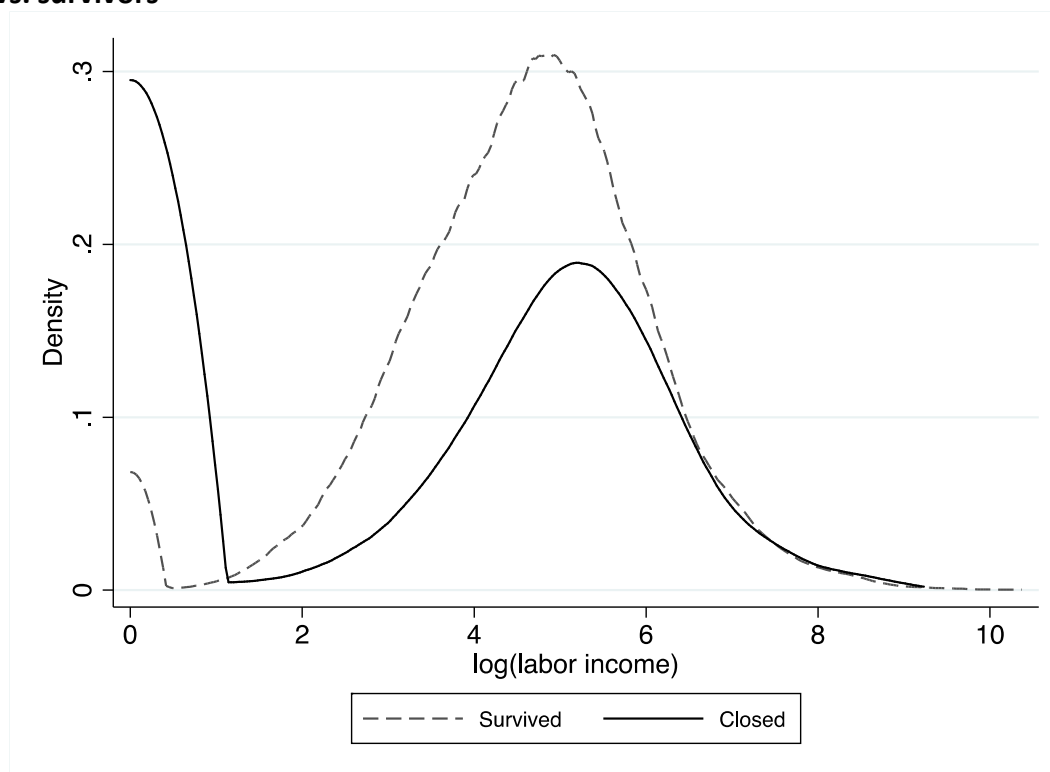
Figure 8. Comparison of labor productivity at baseline for firms that closed vs. survivors, by reason for closure
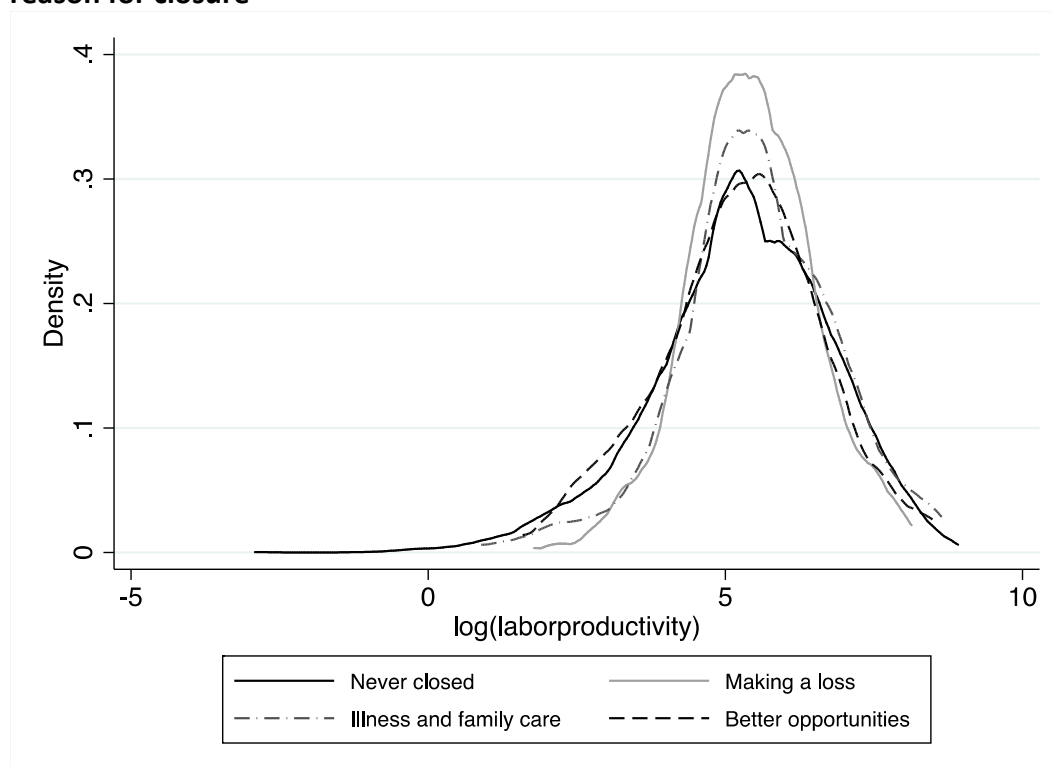
Figure 9. Comparison of monthly labor income for owners of firms that closed vs. survivors, by reason for closure
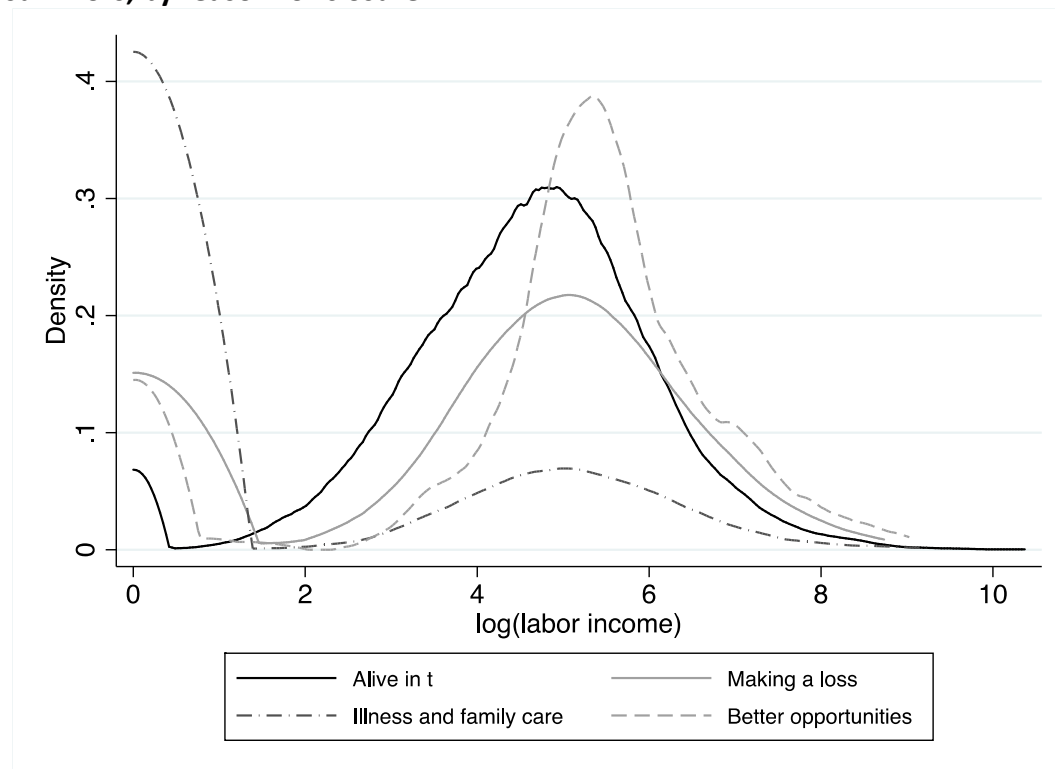
Table 1. Overview of included surveys

Survey	Rounds	Sample Size
A. General Household Surveys:		
Indonesian Family Life Survey (IFLS)	2000,2007/2008,2014/2015	2,462
Mexican Family Life Survey (MxFLS)	2002,2005/2006,2009/2010/2011/2012	723
Nigeria General Household Survey (NGGHS)	2010/2011 - 2012/2013 (post-planting and post-harvest each)	1,804
Townsend Thai Survey (TTHAI)	1997 - 2014 (Annual)	1,122
B. Specialized Firm Impact Evaluation Surveys		
Cotonou Informal Enterprise Survey (BJINFORMALITY)	2014 - 2016	1,197
Egypt Macroinsurance for Microenterprises Survey (EGMACROINSURANCE)	2012-03 - 2012-11	1,441
Ghana Microenterprises Survey (GHMS)	2008 - 2010	397
Kenya GET Ahead Business Training Program Impact Evaluation Survey (KEGETAHEAD)	2013 - 2016	2,365
Malawi Business Registration Impact Evaluation Survey (MWBRIE)	2011/2012 - 2013/2014	757
Nigeria YouWiN! National Business Plan Competition Impact Evaluation Survey (NGYOUWIN)	2011 - 2016	280
Sri Lanka Microenterprise Survey (SLMS)	2000 - 2016	269
Sri Lanka Impact Evaluation Survey of Business Training for Women (SLFEMBUSTRAINING)	2009 - 2015	260
Sri Lanka Informal Enterprises Survey (SLINFORMALITY)	2008 - 2011	105
Sri Lankan Longitudinal Survey of Enterprises (SLLSE)	2008 - 2014	284
Lomé Informal Enterprise Survey (TGINFORMALITY)	2013/2014 - 2015	499
Uganda Women's Income Generating Support Impact Evaluation Survey (UGWINGS)	2009 - 2011	461

Note: Only the control group is used from surveys that are part of randomized experiments.

Table 2. Summary Statistics on Firms

	Number of Observations	Mean	S.D.	10th	Median	90th
Owner Characteristics						
Owner is Female	18,761	0.56	0.50	0	1	1
Age of Owner	17,961	43.41	13.36	27	42	62
Owner has tertiary education	15,294	0.06	0.24	0	0	0
Firm Characteristics						
Age of firm (years)	15,027	9.47	9.50	1	6	22
Monthly sales (USD)	17,159	1053	18175	2	205	1664
Monthly profits (USD)	15,424	129	733	-63	63	417
Any paid worker	15,737	0.29	0.45	0	0	1
Number of paid workers	15,645	0.75	5.02	0	0	2
Retail firm	17,457	0.52	0.50	0	1	1
Manufacturing firm	17,449	0.14	0.34	0	0	1
Services firm	17,449	0.31	0.46	0	0	1

Table 3. Univariate and Multivariate Correlates of Firm Death

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Years since baseline	0.0679***	0.0674***	0.0688***	0.0673***	0.0688***	0.0684***	0.0686***	0.0688***	0.0663***	
	(0.00186)	(0.00186)	(0.00183)	(0.00184)	(0.00183)	(0.00183)	(0.00183)	(0.00183)	(0.00186)	
Age of firm (years)	-0.00575***	-0.0108***							-0.00929***	0.960***
	(0.000601)	(0.00135)							(0.00138)	(0.00632)
Age of firm squared		0.000180***							0.000150***	1.001***
		(3.20e-05)							(3.28e-05)	(0.000152)
Firm in first year of life		0.0774***							0.0666***	1.283***
		(0.0114)							(0.0113)	(0.0620)
Number of employees			-0.00148						-0.00224	0.983*
			(0.00261)						(0.00241)	(0.00889)
Log Business Profits				-0.0363***					-0.0302***	0.892***
Business sector (omitted category retail):				(0.00320)					(0.00314)	(0.0119)
Manufacturing					-0.0183				-0.0269**	0.878**
					(0.0131)				(0.0124)	(0.0544)
Services					-0.00619				-0.0171*	0.928
					(0.0106)				(0.00994)	(0.0436)
Other					-0.00831				-0.00466	1.001
					(0.0259)				(0.0264)	(0.120)
Age of Owner						-0.0159***			-0.0111***	0.950***
						(0.00178)			(0.00179)	(0.00698)
Age of Owner Squared						0.000154***			0.000114***	1.001***
						(1.93e-05)			(1.95e-05)	(7.63e-05)
Owner is Female							0.0273***		0.0197*	1.112**
							(0.0105)		(0.0102)	(0.0521)
Owner has tertiary education								0.0195	0.0166	1.105
								(0.0216)	(0.0203)	(0.0951)
Observations	40,954	40,954	40,957	40,957	40,957	40,957	40,957	40,957	40,954	40,954

Standard errors in parentheses, clustered at the firm level. All specifications include survey fixed effects. Columns 1-9 are marginal effects from a probit of firm death; column 10 are odds ratios from a proportional hazard model. Estimation is over time horizons of five years or less. *, **, and *** denotes significance at the 10, 5, and 1 percent levels respectively.

Table 4. Changes in Productivity, Profitability, and Labor Earnings Associated with Firm Death

	Log(Labor Earnings + 1)	Log(Labor Productivity)	Log(Profitability+1)
Panel A: Full Sample			
Firm Death	-1.8567***	-0.1829***	-0.2095***
	(0.0666)	(0.0313)	(0.0318)
Observations	15,187	7,927	7,261
Panel B: By Cause of Death			
Making a Loss	-1.5123***	-0.1774***	-0.2207***
	(0.1109)	(0.0591)	(0.0579)
Illness or Family Care	-3.3645***	-0.1828**	-0.1990***
	(0.1051)	(0.0769)	(0.0635)
Better Opportunities	-0.1911	-0.1642	-0.0739
	(0.1650)	(0.1232)	(0.1040)
Observations	13,114	6,035	5,661

Note: Robust standard errors, clustered at the firm level, in parentheses. Labor earnings are the sum of wage earnings plus profits in the month preceding a given survey wave. Labor productivity is measured as business sales in the month preceding a given survey wave divided by the number of employees +1. Profitability is measured by business profits in the month preceding a given survey wave.

Table 5. Cause of Firm Death and Main Activity of Owner After Firm Death

Reason for closure	Percent	Frequency	Working for a wage	Looking for a job	Operating a different business	Housework looking after children	Other	Total
Making a loss	40.55	590	30.92	7.96	18.26	17.54	25.32	553
Sickness	13.54	197	8.13	2.50	10.63	46.25	32.50	160
Care for family	12.37	180	5.29	1.76	4.12	71.18	17.65	170
Better wage job	6.53	95	95.56	1.11	2.22	1.11	0.00	90
Better business opportunity	4.88	71	10.77	1.54	75.38	6.15	6.15	65
Other	22.13	322	19.51	6.27	18.12	33.10	23.00	287
Total	100.00	1,455	342	71	228	392	292	1,325

Note: Data for columns (1)-(2) from BENINFORM, KENYAGETAHEAD, MALAWIFORM, NGYOUWIN, SLKFEMBUSTRAINING, SLKINFORMALITY, SLLSE, SLMS, and TOGOINF. The category "Other" includes 3.09% of businesses, which had to close because of a catastrophe (e.g. fire or burglary), 1.51% of business which were shut down by government, a remaining 1.72% of owners who discontinued their businesses because either the business was taken from them, they married, or migrated. The remaining 15.81% of owners who closed their businesses either did not specify this other reason more in detail or closed because of any other miscellaneous reason, which could not be coded into one of the above categories (for instance because of a lack of workers or inputs, or a machine broke down).
Data for columns (3)-(8) from BENINFORM, KENYAGETAHEAD, MALAWIFORM, SLKFEMBUSTRAINING, SLLSE, SLMS, and TOGOINF.

Table 6. Multinomial logit regression results

	Reason for closure (Base outcome: Survived at time t)		
	(1)	(2)	(3)
	Making a loss	Illness and family care	Better opportunities
Years since baseline	1.1824***	1.3050***	1.2972***
	(0.0514)	(0.0838)	(0.1104)
Age of firm (years)	0.8967***	0.9249***	0.9661
	(0.0180)	(0.0195)	(0.0442)
Age of firm squared	1.0016***	1.0012***	0.9997
	(0.0004)	(0.0004)	(0.0014)
Firm in first year of life	1.3569*	1.4249*	2.0584***
	(0.2228)	(0.2610)	(0.5764)
Number of employees	0.9674	0.7591**	0.9270
	(0.0529)	(0.1044)	(0.1973)
Log business profits	0.8272***	0.8608**	0.9343
	(0.0459)	(0.0580)	(0.0803)
Sector (Omitted category = Retail)			
Sector=Manufacturing	0.6488**	0.6803	0.4160**
	(0.1320)	(0.1999)	(0.1441)
Sector=Services	0.7583**	0.8832	1.0270
	(0.1055)	(0.1441)	(0.2254)
Sector=Other	0.4181	0.6776	0.4866
	(0.2410)	(0.4621)	(0.4990)
Age of owner	0.9714	0.9467	1.2123
	(0.0389)	(0.0337)	(0.1486)
Age of owner squared	1.0002	1.0008*	0.9971*
	(0.0005)	(0.0005)	(0.0017)
Owner is female	1.3811**	2.4835***	0.8244
	(0.2107)	(0.6069)	(0.2129)
Owner has tertiary education	1.4118	0.9010	2.1314**
	(0.3170)	(0.2553)	(0.6714)
Constant	0.1851	0.0010***	10799000***
	(0.3269)	(0.0018)	(25950000)
Observations	16,308	16,308	16,308

Note: Robust standard errors in parentheses, clustered at the firm level. The specification includes survey fixed effects and controls for missing values in firm age, number of employees, log business profits, sector=other, age of owner, and owner has tertiary education. Estimation is over time horizons of five years or less. *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix

A.1 Additional details on surveys

We discuss each of the sixteen surveys used in our study, and how firm death is defined and calculated in each. We begin with the multi-purpose household surveys, before turning to specialty firm panel surveys.

Multi-purpose household surveys:

Indonesian Family Life Survey (IFLS)

The Indonesian Family Life Survey (IFLS) is a publicly available multi-purpose panel survey of the Indonesian population. It surveys over 7,000 households, consisting of over 30,000 individuals, and is representative of about 83 percent of the Indonesian population. Further details on study design and data can be obtained in Strauss et al. (2016) and online at http://www.rand.org/labor/FLS/IFLS.html. There are currently five waves of data available and our sample draws from IFLS3, which was fielded in 2000, IFLS4 (fielded in 2007/2008) and IFLS-5 (fielded in 2014/2015). Non-farm business modules of IFLS1 and IFLS2 are less detailed, complicating the identification of the business owner as well as the matching of businesses with later rounds, which is why we decided not to include them.

We used the non-farm business modules of the different survey waves to identify household businesses in each of the three waves we included in our sample. A household was coded as operating a business if there was a householder who had worked in a family-owned, non-farm business like trade/retailing or been self-employed in a non-farm enterprise during the past 12 months preceding the survey, which was entirely owned by the household, and still operating at the time of the survey. Owners were identified as those household members who were indicated in the non-farm business module as being primarily responsible for the reported business. Up to two members could be reported per business and their individual survey identification numbers were given, so that information on owner characteristics could be matched from the other survey modules.

Since the IFLS is not a panel of firms, but of households who are asked in each round if they operate a business, it allows us to capture information on firms that were not operating in IFLS3, but in IFLS4. We can thus use two baselines for the analysis of firm dynamics over time, as each round provides us with a sample of firms operating at a given point in time. Taking IFLS3 as baseline, we observe horizons of 7.5 and 14.5 years to IFLS4 and IFLS-5, respectively. We observe a horizon of seven years from IFLS4 to IFLS-5 with IFLS4 as the baseline. Since horizons using IFLS3 and IFLS4 as baseline differ, we do not pool together data for the same observation period with this approach, but increase the sample with information on different horizons.

The non-farm business modules of the IFLS did not ask whether the business reported in a given round is the same as the business reported in the previous round, or what happened to the one reported earlier. We therefore needed to match reported businesses across rounds in order to be able to determine whether the business survived or not, and in the latter case, whether a new firm was started. To simplify this matching exercise, we only kept firms of households who never operated more than one enterprise in any of the three survey rounds in our sample.

Taking IFLS3 [IFLS4] as our baseline, and IFLS4 [IFLS-5] as follow-up round, we then coded the business reported at baseline as surviving, if a business was reported in the follow-up round, and:

- start dates (i.e. date when the business was started) were identical across the two waves; or
- the field of the business was the same across the two waves, and the start date of the business reported in the follow-up round was a date preceding the interview date of the baseline wave; or
- at least one of the business owners (person(s) indicated as being responsible for the business) was the same across the two waves, and the start date of the business reported in the follow-up round was a date preceding the interview date of the baseline wave.

We coded the business reported at baseline as closed, if:
- no business was reported in a subsequent round.

Otherwise, we coded survival/closure as missing.

We coded a new firm start, if:
- the start date of the business reported in the follow-up round was a date after the interview date of the baseline wave.

We coded the variable *newfirmstart* to zero, if:
- the business reported in the baseline round survived; or
- no business was reported in the follow-up round and the business of the baseline round was coded as closed; or
- the start date of the business reported in the follow-up round was a date preceding the interview date of the baseline wave and no business was reported in the baseline wave.

Otherwise, the variable *newfirmstart* was coded as missing.

We used these three types of business characteristics (start date / age of the business, ownership, and business sector) together, because start dates turned out not to be very accurate, and it is possible that ownership of a household business changes over time. Even so, there remained a number of inconsistencies and cases which could not be matched, so that in a final step, we inspected the characteristics of these businesses in all three consecutive survey waves visually and coded survival on a case-by-case basis, again, taking into account information on firm age, ownership and the business sector reported in the three survey waves. For these remaining observations, we decided to exclude firms from the sample if information on firm age was missing, since this information is important in determining *survival* and *newfirmstart*.

Given that the non-farm business module of the IFLS allowed for more than one owner to be reported, and ownership could change from one round to another, we also had to decide on which of the owners to include in case we identified multiple owners for a business, as we were interested in assessing also owner characteristics in our study. For firms for which we had identified more than one owner, we decided to include the owner, who was reported as the owner in all of the waves in which we observed the firm. If this was the case for both owners, we included the owner,
- who had the lowest household roster number (this chose the household head as the owner if he or she was among the owners), if no other business was reported for at least one of the owners over the whole observation period; or
- who had already been identified as the owner of a different business, observed during one or more different waves over the whole observation period;

and in the remaining cases we chose the owner who was reported as the owner in most of the waves in which we observed the firm, or, if this was the case for both owners, who had the lowest household roster number.

Finally, we excluded firms for which we could not identify at least one owner among the household members who were in the IFLS sample.

Mexican Family Life Survey (MxFLS)
The Mexican Family Life Survey (MxFLS) is a publicly available multi-purpose panel survey conducted in Mexico. It surveys 8,400 households (35,000 individuals) and is representative of the Mexican population at the national, urban, rural, and regional level. Further details on study design and the data can be obtained in Rubalcava and Teruel (2013). Three rounds of data have been collected (2002, 2005-2006, and 2009-2012) and we included all three of them in our sample.

We identified household businesses in the MxFLS using the non-farm business modules of the different survey waves. A household was coded as operating a business if there was a household member who had owned or shared a non-agricultural business during the past 12 months preceding the survey. Owners were identified as those household members who took the most important decisions regarding the business. Given that only their relationship to the respondent was reported, as opposed to their survey identification number, owners could only be uniquely identified, if they were either the respondent or his or her spouse.

Given the household panel structure of the MxFLS, we also extracted information on firms that were not operating in round 1, but in round 2, and use two baselines for the analysis of firm dynamics over time. Taking round 1 as baseline, we observe horizons of three and 7.5 years to rounds 2 and 3, respectively. We observe a horizon of 4.5 years from round 2 to round 3 with round 2 as the baseline. Since horizons using rounds 1 and 2 as baseline differ, we do not pool together data for the same observation period, but increase the sample with information on different horizons with this procedure.

In order to code whether a business observed at baseline survived or not, and in the latter case, whether a new firm was started, we needed to match reported businesses between rounds based on a number of business characteristics, as the MxFLS does not link household businesses from round to round. For simplification, we decided to only include firms of households in our sample who never operated more than one enterprise in any of the three survey rounds.

Taking round 1 [round 2] as our baseline, and round 2 [round 3] as follow-up rounds, we then coded the business reported at baseline as surviving, if a business was reported in the follow-up round, and:
- start dates (i.e. year the business was started) were identical across the two waves; or
- at least one of the business owners (person(s) taking the most important decisions regarding the business) was the same across the two waves, and the business reported in the follow-up round was not started after 2001 [in or after the year in which the baseline survey was conducted].

We coded the business reported at baseline as closed, if:
- no business was reported in a subsequent round.

Otherwise, we coded *survival* as missing.

We coded a new firm start, if:
- the start date of the business reported in the follow-up round was 2002 or later [the year in which the baseline survey was conducted or later].

We coded the variable *newfirmstart* to zero, if:
- the business reported in the baseline round survived; or
- no business was reported in the follow-up round and the business of the baseline round was coded as closed; or
- the start date of the business reported in the follow-up round was a date preceding 2002 [the interview date of the baseline wave] and no business was reported in the baseline wave.

Otherwise, the variable *newfirmstart* was coded as missing.

We coded survival and newfirmstart from round 1 to round 3 as well, based on these same criteria described above.

After this matching and coding, we inspected the characteristics of the businesses for which *survival* and *newfirmstart* remained unclear or inconsistent on a case-by-case basis, taking into account information on firm age and ownership reported in the three survey waves. For this, we decided to exclude firms from the sample if information on firm age was missing, since this information is important in determining *survival* and *newfirmstart*.

For firms for which we had identified more than one owner, we decided to include the owner, who was reported as the owner in all of the waves in which we observed the firm. If this was the case for both owners, we included the owner,
- who had the lowest household roster number (this chose the household head as the owner if he or she was among the owners), if no other business was reported for at least one of the owners over the whole observation period; or
- who had already been identified as the owner of a different business, observed during one or more different waves over the whole observation period;
and in the remaining cases we chose the owner who was reported as the owner in most of the waves in which we observed the firm, or, if this was the case for both owners, who had the lowest household roster number.

Finally, we excluded firms for which we could not identify at least one owner among the household members who were in the MxFLS sample.

Nigeria General Household Survey (NGGHS)
The Nigeria General Household Survey (NGGHS) is a cross-sectional multipurpose household survey of 22,000 households, carried out annually in Nigeria. Under the World Bank's Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) project a panel component was introduced into the multi-purpose NGGHS, which collects longitudinal data for 5,000 households. The sample is representative at the national level and provides reliable estimates of key socio-economic variables for the six zones in the country. Further details on study design and the data can be obtained in National Bureau of Statistics of the Federal Republic of Nigeria (2014).

The panel component was first introduced in 2010/2011 and has since then been implemented every two years. There are currently three waves available, of which we use the two first ones (2010/2011, and 2012/2013), given that the third wave was not yet available at the time of our

data analysis. Each wave consists of two visits, one in the post-planting period and another in the post-harvest period, which we consider separately, resulting in a total of four survey waves from the NGGHS in our sample.

We identified household businesses in the NGGHS panel component using the non-farm business modules of the different survey waves. A household was coded as operating a business if there was a household member who had worked for himself, other than on a farm or raising animals (e.g. operating his or her own business, trade, working as a self-employed professional or craftsman) during the past 12 months preceding the survey. Owners were identified using information on who in the household owned the business, who managed it, and who was the survey respondent for the non-farm business module. Up to two members could be reported as owners and/or managers per business and their individual survey identification numbers were given, so that information on owner characteristics could be matched from the other survey modules. We then identified the business owner as the person who was indicated as the business owner if the household indicated only one and the same owner over the whole observation period. For the remaining cases (i.e. households who reported two owners in the same wave, or different owners in different waves) we identified the business owner(s) in a given round as the owner(s):

- who was also reported to be the manager and was the respondent for the non-farm business module; or
- who was/were also the manager of the business, if the respondent was neither a business owner nor a manager of the business; or
- who was also the respondent for the non-farm business module, if no information on the manager was available or none of the owners were identified as the manager; or
- who was/were the business owner(s), if no information was provided on manager and respondent; or
- who was the household head if no information was given on who was the business owner, manager, and respondent; or
- who was also reported to be the manager if one owner was indicated as manager and the other was the respondent for the business module.

In contrast to the other multi-purpose household surveys we use, the non-farm business modules of the NGGHS panel component to check whether the households surveyed reported any non-farm household enterprise or income generating activity in the previous round. If so, they ask, whether this activity is still operating, closed permanently, temporarily or seasonally. We coded enterprises as surviving, if they were still being operated, and as closed if they were either closed permanently, temporarily, or seasonally. Households were also asked if the firm they reported was a new or an original one. We coded the variable *newfirmstart* to one if the reported firm was a new one, and to zero if it was an original one. For simplification, we only included firms in our sample if the household never operated more than one firm in any of the four rounds we considered. Nevertheless, relying on the matching and information on closure and new firm start provided by the NGGHS was not sufficient. There remained a number of inconsistent and unclear cases, which we then coded on a case by case basis. For this we considered ownership and business sector over the horizon over which we observed these unclear cases for a given household, as well as, if available, information provided on closure and new firm start.

Given that the NGGHS panel component asks households in each survey round, whether they are operating a business, regardless of whether they were operating one in the previous round, we also extracted information on firms that were not operating in round 1, but in round 2, or 3 and use three baselines for the analysis of firm dynamics over time. Taking round 1 as baseline, we observe horizons of six months, 1.5 years, and two years to rounds 2, 3, and 4 respectively.

With round 2 as the baseline, we observe horizons of 1.5 and two years to round 3 and 4, respectively. And with round 3 as baseline, we observe a horizon of six months to round 4. In the analysis, we pool the data for the same horizons with different baselines together to increase the sample size for a given horizon. We also use different baselines when horizons differ to increase the sample with information on different observation periods.

In order to be able to assess owner characteristics in our study, for jointly owned businesses, and businesses for which ownership changed over the observation horizon, we determined one owner as the owner for all waves in which we observed the firm. In the case of jointly owned businesses, we decided to include the owner, who was reported as the owner in all of the waves in which we observed the firm. If this was the case for both owners, we included the owner,

- who had the lowest household roster number (this chose the household head as the owner if he or she was among the owners), if no other business was reported for at least one of the owners over the whole observation period; or
- who had already been identified as the owner of a different business, observed during one or more different waves over the whole observation period;

and in the remaining cases we chose the owner who was reported as the owner in most of the waves in which we observed the firm, or, if this was the case for both owners, who had the lowest household roster number.

Townsend Thai Survey (TTHAI)
The Townsend Thai Project is a data collection project consisting of ongoing annual and monthly multi-purpose household, institutional, and key informant (village leader) panel surveys, in addition to collecting environmental data in Thailand. The baseline was conducted in 1997, collecting data from rural areas in four Thai provinces, two in the Northeast and two in the Central region. In the early and mid-2000s the sample was extended to include data from urban areas as well as from the Southern and Northern region of Thailand. Further details on study design can be obtained in Townsend et al. (2016). We use data from the annual household surveys conducted between 1997 and 2014, resulting in the inclusion of 18 survey waves in our sample.

We identified non-farm household businesses using the household business modules of the different survey waves and coded a household as operating a business if anyone in this household owned a business, which was entirely owned by the household or this person. Since our interest was in non-farm enterprises, we did not consider entrepreneurial activities if they were classified as rice mills, shrimp farms, fish farms, shrimp and fish farms, or any other agricultural activity, as provided by the description of the type of business operated. The household business module did not ask who was the owner or manager of the business reported. We therefore identified business owners using the occupation module, as those household members whose occupation was reported to be *"owner of a business"* in case the household was operating a non-farm household business.

Households in the panel were asked in each survey round if anyone in the household owns a business. Each round provides us hence with a sample of firms operating at a given point in time. Taking advantage of this, we pool together data for the same observation period to increase our sample for a given horizon. For instance, we include baseline information from 17 survey waves (with baselines from 1997 to 2013) to assess survival over a horizon of one year, from 16 survey waves (with baselines from 1997 to 2012) for an observation period of two years, and so forth. For the longest observation horizon, 17 years, we can only use information on baseline firms from one survey round, i.e. from 1997.

The annual Townsend Thai Survey (TTHAI) we use, does not link household businesses from round to round. It does hence not provide information about survival of a given firm that has been identified in one of the survey rounds to later rounds. In order to code whether a business observed in a given round survived or not, and in the latter case, whether a new firm was started, we needed to match reported businesses between rounds based on a number of business characteristics. For simplification, we decided to only include firms of households in our sample who never operated more than one enterprise in any of the three survey rounds.

We then coded the business of a household reported in a given survey round as surviving to a given follow-up round, if a business was reported for the same household in the follow-up round, and:
- firm ages coincided across the two waves; or
- the sector of the business was the same across the two waves, and the age of the business reported in the follow-up round was not less than the time difference between the two survey waves; or
- at least one of the household members identified as the business owners was the same across the two waves, and the age of the business reported in the follow-up round was not less than the time difference between the two survey waves.

We coded the business reported at baseline as closed, if:
- no business was reported in the follow-up round.

Otherwise, we coded survival/closure as missing.

We coded a new firm start, if:
- the age of the business reported in the follow-up round was less than the time difference between the two survey waves.

We coded the variable *newfirmstart* to zero, if:
- the business reported in the baseline round survived; or
- no business was reported in the follow-up round and the business of the baseline round was coded as closed; or
- the age of the business reported in the follow-up round was not less than the time difference between the two survey waves.

Otherwise, the variable *newfirmstart* was coded to missing.

For businesses for which *survival* and *newfirmstart* remained unclear or inconsistent, we inspected the characteristics on a case-by-case basis, taking into account information on firm age, type of activity, and ownership reported in all survey waves in which we observed these unclear cases. In addition, we used information on whether the household decided to stop running any business in the past 12 months, and if so, the type of activity they were no longer running, as well as whether the business reported in a given round had been founded in the past 12 months, which was provided in the household business modules of the follow-up questionnaires.

Finally, we included one owner per firm in our sample, in case we had identified more than one owner for a given firm, either because we had identified more than one owner in one or more waves, or because different owners were reported in different waves. In these cases, we decided to include the owner, who was reported as the owner in all of the waves in which we observed the firm. If this was the case for both owners, we included the owner,

- who had the lowest household roster number (this chose the household head as the owner if he or she was among the owners), if no other business was reported for at least one of the owners over the whole observation period; or
- who had already been identified as the owner of a different business, observed during one or more different waves over the whole observation period;

and in the remaining cases we chose the owner who was reported as the owner in most of the waves in which we observed the firm, or, if this was the case for both owners, who had the lowest household roster number.

Specialized firm impact evaluation surveys:
We use data from the control group only in a number of panel data surveys collected to measure the impacts of firm interventions in randomized experiments.

Cotonou Informal Enterprise Survey (BJINFORMALITY)
The sample surveyed for the Cotonou Informal Enterprise Survey (BJINFORMALITY) consists of 3,596 informal businesses and is representative of informal businesses operating in Cotonou, Benin. The survey was designed to evaluate the effect of a simple free business registration process on formalization, and test the effectiveness of supplementary efforts to enhance the presumed benefits of formalization by facilitating its links to government training programs, support to open bank accounts, and tax mediation services. Details on sampling and the interventions tested can be found in Benhassine et al., (2016). Survey participants were randomly allocated into three treatment groups and one control group, with the control group consisting of 1,197 informal firms. The baseline survey was conducted between March and April 2014. Two follow-up surveys followed, the first was conducted between April and June 2015, and the second between May and June 2016. Attrition rates at first and second follow-up were 11.8 percent and 15.9 respectively and were not correlated with treatment status.

We coded *survival* to one if:
- the owner reported to be still operating in the same sector as in the previous wave; or
- the business was said by a person knowledgeable on this matter to be still operating, if the owner refused to be re-interviewed or could not be interviewed when survey attempts were made.

We coded *survival* to zero if:
- the owner reported not to be operating anymore in the same sector as in the previous wave because of a change in activity; or
- the survey was not completed because the business was found to have been closed definitely; or
- the business was said by a person knowledgeable on this matter not to be operating anymore, if the owner refused to be re-interviewed or could not be interviewed when survey attempts were made.

We coded *newfirmstart* to one if the owner reported not to be operating anymore in the same sector as in the previous wave because of a change in activity. For the endline wave, we additionally coded new firm start to one if the enterprise had closed definitely and the owner was now operating a different enterprises.

Newfirmstart was coded to zero if:
- an enterprise survived; or

- the survey was not completed because the business was found to have been closed definitely (and in the endline: the owner was not operating a different enterprise); or
- the business was said by a person knowledgeable on this matter not to be operating anymore, if the owner refused to be re-interviewed or could not be interviewed when survey attempts were made.

With this procedure, we coded *survival* and *newfirmstart* from round to round. Given that we were interested in following only the baseline businesses, and not businesses started in the midline, we recoded *survival* to be capturing survival of the baseline firm. This implied, recoding *survival* to zero in the endline wave, if the business had been coded as closed and a new firm had been started in the midline wave.

Egypt Macroinsurance for Microenterprises Survey (EGMACROINSURANCE)
The survey was conducted to evaluate the impact of an insurance to protect microenterprise owners against macroeconomic and political shocks in Egypt, described in detail in Groh and McKenzie (2016). The baseline survey was conducted in March 2012 among a sample of clients from Egypt's largest microfinance institution at particular loan sizes, who were deciding about the renewal of their loans. The sample consisted of 2,980 individuals, with equal numbers assigned to treatment and control. In November 2012, a follow-up survey was conducted, which was able to re-interview 98.8 percent of respondents in the control group and 98.9 percent in the treatment group. We use these two survey waves and are hence able to examine survival over a horizon of approximately 6 months. Using only data from the control group yields 1,441 business owners at baseline, which we include in our sample.

We coded *survival* to one, if:
- the respondent said that he or she had not closed or sold the business in which he or she worked at the time of the baseline survey in March 2012; or
- the business was said by a person knowledgeable on this matter, to be still being operated by either the business owner, an immediate family member, or by some other person, if the respondent refused to participate in the survey or could not be located.

We coded *survival* to zero, if:
- the respondent said that he or she had closed or sold the business in which he or she worked at the time of the baseline survey in March 2012; or
- the business was said by a person knowledgeable on this matter, to be closed, if the respondent refused to participate in the survey or could not be located.

If the respondent said that he or she had closed or sold the business in which he or she had worked at the time of the baseline survey in March 2012, we coded *newfirmstart* to one, if the respondent said that he or she was operating a different business after closing or selling the business previously operated, and had opened this business since March 2012. Otherwise, we coded *newfirmstart* to zero.

Ghana Microenterprises Survey (GHMS)
The Ghana Microenterprises Survey (GHMS) was conducted to evaluate the impact of cash and in-kind grants to microentrepreneurs in urban Ghana. Details of the intervention and study design are described in Fafchamps et al. (2014). The baseline survey was conducted in October and November 2009, surveying a representative sample of 907 microenterprise owners in urban Ghana, in particular in the cities of Accra and Tema. A second pre-treatment survey of these firms was conducted in February 2009 to eliminate firms most likely to attrit, yielding a final baseline sample of 793 firms for the experiment, of which 397 were assigned to the control

group. Four additional quarterly follow-up surveys followed in May 2009, August 2009, November 2009, and February 2010. In March 2012 a final, longer-term follow-up survey was conducted. We include the 397 baseline control group enterprises in our sample and use data from all six survey waves.

Using information from round to round, we coded *survival* in each round to one if:
- the owner said that he or she was still owning and operating a business, and the line of the business had not changed; or
- the line of the business had changed, but when asked what happened to the business previously ran, the owner said that he or she was either operating the same business, or that the business was now being run by another family member; or
- the owner or a person knowledgeable of the matter told the enumerator that the owner was still operating the business, if the respondent refused to participate in the survey or could not be located; or
- the owner or a person knowledgeable of the matter told the enumerator that he owner was not operating the business anymore but that someone else was operating it, if the respondent refused to participate in the survey or could not be located.

We coded *survival* to zero if:
- the line of the business had changed, and when asked what happened to the business previously run, the owner said that the business had closed down, had been modified to become the current business, or gave any other answer that implied the closure of the original business; or
- the owner or a person knowledgeable of the matter told the enumerator that the business had closed down, if the respondent refused to participate in the survey or could not be located.

We coded *newfirmstart* to one, if the line of the business had changed from one round to another and the respondent said that his or her main activity was then operating a different business. Otherwise, we coded *newfirmstart* to zero.

To be able to assess survival of firms operating at baseline, we recoded *survival* in a final step, by recoding it to zero for all subsequent rounds, if it had been coded as zero in a given round and a new firm start had been coded in the same round or later.

Kenya GET Ahead Business Training Program Impact Evaluation Survey (KEGETAHEAD)
The survey was part of an evaluation of a business training program for women in four counties of Kenya: Kakamega and Kisii in the Western region, and Embu and Kitui in the Eastern region. In each county a census was taken of all female-owned businesses in market centers deemed as medium-sized or large, and then a screening was done to select a sample in which the business did not have more than 3 employees; the business had profits in the past week between 0 and 4000 Kenyan Shillings (KSH) (1 US Dollar averaged approximately 85 KSH over the survey period); sales in the past week less than or equal to 50,000 KSH; and the individual had at least one year of schooling. The survey is therefore representative of microenterprises of this size run by women in these four counties.

The baseline survey took place between June and November 2013, and included 3,537 individuals. These individuals were then assigned to treatment and control for the business training intervention in a two-stage process, with markets being assigned to treatment or control in the first stage, and then, within markets that had been assigned to treatment, individuals being assigned to treatment or control. We include all control group observations, i.e. individuals in

pure control markets and individuals assigned to the control group in treatment markets, totaling 2,365 firms, in our sample. One year after the baseline survey, between June and October 2014, the first follow-up survey was conducted, which was supplemented by a rapid short survey that was conducted between November 2014 and February 2015, and which we mainly used to obtain information on firms not found in the first follow-up. The second follow-up survey took place between February 2016 and July 2016. Further details on the survey, intervention, and study design can be found in McKenzie and Puerto (2017). We used all four survey waves, joining information from the first follow-up survey and the subsequent short follow-up survey and treating them as one round, hence including three rounds of data from this survey in our sample.

Using information from round to round, we coded *survival* in each round to one if:
- the owner answered that she was still operating the same business as in the previous round; or
- the owner was said by a proxy respondent to still be operating a business, if the respondent refused to participate in the survey or could not be located.

We coded *survival* in each round to zero if:
- the owner answered that she had closed the business and was now operating a different business; or
- the owner answered that she was no longer operating a business; or
- the owner said that she had never operated a business; or
- the owner said or was said by a proxy respondent, not to be in business anymore or to have left the country, if the respondent refused to participate in the survey or could not be located; or
- the owner had died.

We coded *newfirmstart* to one, if the owner answered that she had closed the business and was now operating a different business. If the owner answered either that she was still operating the same business as in the previous round, or that she was no longer operating a business, we coded *newfirmstart* to one. In the remaining cases *newfirmstart* was coded to missing.

To transform the coding into survival to a given round for firms that were operating at baseline, and to deal with some inconsistencies that were due to the combination of data from the first follow-up and the subsequent short follow-up, we recoded *survival* in a final step. If *survival* to round 2, as well as *newfirmstart* to round 2 had been coded to one, we recoded *survival* to round 2 to zero, assuming that these were either cases in which the original business had survived to the first follow-up, but not to the subsequent short follow-up survey, and the owner was operating a new firm by the time of the short follow-up survey, or cases in which the owner had closed and opened new firm in the first follow-up and then the business was coded as surviving in the short subsequent follow-up survey, because this new business continued to be open. If a business had been coded to be closed by round 2, but survived to round 3 and a new firm had been started by round 2, we recoded *survival* to round 3 to zero.

Malawi Business Registration Impact Evaluation Survey (MWBRIE)
The Malawi Business Registration Impact Evaluation Survey (MWBRIE) was designed to evaluate the impact of a randomized experiment designed to increase business formalization in Malawi. The survey was conducted on a sample of 3,002 informal firms in urban Lilongwe and Blantyre, which had been identified through a listing of over 100 business centers (i.e. concentrations of firms including industrial parks, markets, streets with shops, set of workshops, etc.). As a result the sample does not include household-based enterprises.

Moreover, firms in the sample comply with one of the following criteria: (i) they had at least one worker contracted outside of family members and business owners, (ii) they were operating in a fixed location with more than one person working in the business, (iii) they were at the 25 percentile of revenues or above. More details are provided in Campos et al. (2015).

The baseline survey was conducted between December 2011 and April 2012. After that, the sample was randomly assigned to either one of three treatment arms or the control group, which consisted of 757 firms. Details on the intervention and experimental design can be found in Campos et al. (2015). A first follow-up survey took place between November 2012 and March 2013. Attrition was 5.7 percent, and uncorrelated with treatment status. The second follow-up survey was implemented between November 2013 and March 2014. Attrition for this second follow-up survey was 9.4 percent, and also uncorrelated with treatment status. Between November 2014 and April 2015, a third follow-up survey took place, and a fourth follow-up survey was conducted between July 2015 and January 2016. Including data from all five survey rounds into our sample, we are able to observe firms over horizons of up to 3.5 years.

We coded *survival* to one if owners said that they were still operating the business that they had when they were first interviewed, i.e. around December 2011 to April 2012. If they said they were not operating this business anymore, we coded *survival* to zero. *Newfirmstart* was coded to one if the business had closed and the owner had started a new business, and to zero if the business had closed but the owner was engaged in an activity different than operating another business.

Nigeria YouWiN! National Business Plan Competition Impact Evaluation Survey (NGYOUWIN)
The survey sample consists of applicants to a nationwide business plan competition launched in late 2011 and described in McKenzie (2017). Applicants submitted their business plans in January 2012, together with baseline information on firm and owner characteristics collected via a baseline data sheet. Business plans were then narrowed down to 2,312 semi-finalists, of which 1,200 received awards averaging US$50,000 each, and among which 729 winners had been randomly selected from a group of 1,841 semi-finalists. This left 1,112 control group individuals. Three follow-up surveys were taken, approximately at yearly intervals after individuals had applied for the program, with the first follow-up conducted between November 2012 and May 2013, the second between October 2013 and February 2014, and the third between September 2014 and February 2015. A fourth follow-up survey took place between July and November 2016.

Given that at baseline, there were not only already existing firms in the sample, but also individuals who intended to start a new business with the grant, and that, in addition, limited firm data collected at baseline, we decided to include only the four follow-up surveys in our sample, taking businesses that were existing by follow-up 1 (either because they survived from the baseline to this round, or because they had been created in between these two rounds) as the baseline sample. In addition, we only kept firms, which were not in the business of agriculture, forestry or fishing, whose owners never operated more than one business in any of the rounds in which we observed them, and which had been assigned to the control group. This resulted in the inclusion of 280 firms from the Nigeria YouWiN! National Business Plan Competition Impact Evaluation Survey (NGYOUWIN) in our sample at baseline.

We coded *survival* to one, if:

- the respondent was currently operating a business in the follow-up round and start dates of the business were the same across the follow-up round and the baseline round (i.e. follow-up 1); or
- the respondent was currently operating a business in the follow-up round and the start date of the business was missing in the follow-up round[12]; or
- a person knowledgeable on the matter answered that the respondent was currently owning a business in the follow-up round, if the owner could not be interviewed after repeated attempts.

We coded *survival* to zero, if:
- the respondent was not operating a business in the follow-up round; or
- the respondent was currently operating a business in the follow-up round, but start dates of the businesses reported in the follow-up and baseline round (i.e. follow-up 1) did not coincide; or
- a person knowledgeable on the matter answered that the respondent was not owning a business in the follow-up round, if the owner could not be interviewed after repeated attempts.

We coded *newfirmstart* to one, if the respondent was currently operating a business in the follow-up round, but start dates of the businesses reported in the follow-up and baseline round (i.e. follow-up 1) did not coincide. Given that firms were only asked for their start dates in follow-up rounds 2 to 4, if they had been established in the survey year, this should have picked up new firm start.

We coded *newfirmstart* to zero, if the respondent was not operating a business, or *survival* had been coded to one.

In a final step, we then recoded *survival* for firms that had been started in a given round and had then been coded as surviving in a later round, because they had appeared with missing start dates in this round.

Sri Lanka Microenterprise Survey (SLMS)
The Sri Lanka Microenterprise Survey (SLMS) was designed to evaluate an experiment to measure the impact of providing capital grants to microenterprises. It took place in three Southern and South-Western districts of Sri Lanka: Kalutara, Galle and Matara. One-time grants of between US$100 and US$200 were given in cash or in-kind (as business equipment) to male and female-owned enterprises, some of whom had been affected by the December 2004 tsunami. A door-to-door screening survey of households was used to identify enterprises with invested capital of 100,000 rupees (about US$1000) or less, excluding investments in land and buildings, resulting in a baseline sample of 617 enterprises in retail trade and manufacturing, operated by owners 22 to 65 years old, and with no paid employees.

---

[12] Businesses were only asked their start dates in follow-up rounds 2 to 4 if they had been established in the survey year. We hence assumed that all firms which had missing values for start dates in these rounds, had been established earlier. We corrected for potential erroneous coding as a result of this rule in a final step. Initially we had tried to use ISIC codes of the businesses reported in different rounds to code *survival*, coding *survival* to one if a firm reported in a follow-up round was operating in the same sector as in in the baseline (i.e. follow-up 1) round and as closed, if they were in different sectors. This led however to extremely high closure rates, and a visual inspection of the sectors in two consecutive survey waves revealed that sectors of firms coded as closed with this rule, were very similar, although having different ISIC codes. As a result, we decided not to use ISIC codes for the coding of *survival*.

The baseline survey was conducted in April 2005. Firms in the sample were then randomly allocated into one of two treatment groups, which received US $100 or $200 either after the first or the third survey round and either in cash or in-kind. Details on the experimental design and intervention can be found in de Mel et al. (2008, 2009, and 2012). Follow-up surveys were conducted at quarterly intervals, from July 2005 to April 2007, and semi-annually in October 2007 and April 2008. Further follow-up surveys were carried out in June and December 2010, in September 2015, and in March 2016. We included all 15 waves of data collection in our sample, which allows us to observe firms for horizons of up to eleven years. Using firms assigned to the control group, we include 269 firms at baseline from the SLMS in our sample

For rounds 2 to 11, we coded *survival* from round to round to one, if the owner was still working in the same line of business, he or she had been working in, in the previous round., and to zero if the line of business had changed. *Newfirmstart* was coded to one for rounds 2 to 11 if the line of business had changed and the owner stated that his or her main activity was to operate a new business. If the line of business had changed but the main activity was a different one (for instance, working for a wage, looking for work, or housework or looking after children), we coded *newfirmstart* to zero.

In round 12, questions on business continuation referred to the original baseline business. We coded *survival* to one, if the owner was still working in the same line of business, he or she had been working in, during the baseline survey in April 2005, and to zero if the line of business had changed. *Newfirmstart* was coded to one if either the line of the original business had changed or the owner had started a new line of business while continuing to operate the original business, and the owner stated that his or her main activity after closing, selling or changing the business was either to operate a different business or to ooperate a new business along with the original one. If the line of business had changed but the main activity was different to operating another business, we coded *newfirmstart* to zero. For round 13, we coded *survival* and *newfirmstart* based on the same criteria, only that the questions asked here referred to the business being operated in round 12 and *survival* and *newfirmstart* were hence being coded from round to round.

In rounds 14 and 15 we coded *survival* from round to round to one, if:
- the owner stated that he or she was currently owning a business, and had not changed the industry of the business since the previous round or closed down the business he or she had been running and started a new business; or
- a person knowledgeable of the matter answered that either the enterprise owner was still running the business, or the business was being run by an immediate family member, or the business was being run by some other person, if the respondent refused to participate in the survey or could not be located.

We coded *survival* from round to round to zero in these rounds, if:
- the owner stated that he or she was currently not owning a business; or
- the owner said that he or she had changed the industry of the business since the previous round, or that he or she had closed down the business he or she had been running and started a new business; or
- a person knowledgeable of the matter answered that the enterprise was closed or not functioning, if the respondent refused to participate in the survey or could not be located.

Newfirmstart was coded in rounds 14 and 15 to one, if:

- the business being owned in the given round had been started after the date of the previous survey round; or
- the owner stated that he or she had either changed the industry in which the business was operating, or started another business while still operating the original business, or closed down the previously run business and started a new one.

Newfirmstart was coded to zero in these rounds, if:
- the owner stated that he or she was currently not owning a business.

Except for round 12, we then recoded *survival* in a way that it referred to survival of firms operating at baseline by recoding it to zero for all subsequent rounds, if it had been coded as zero in a given round and a new firm start had been coded in the same round or later.

Sri Lanka Impact Evaluation Survey of Business Training for Women (SLFEMBUSTRAINING)
The Sri Lanka Impact Evaluation Survey of Business Training for Women (SLFEMBUSTRAINING) was conducted to evaluate a business training intervention aimed at increasing female labor force participation rates and to raise the income levels of low-earning women already in business. It surveys two different groups of females: (i) a representative sample of 628 current female business owners operating in greater Colombo and greater Kandy areas, who had monthly profits of 5,000 Rs or less (US$43), and (ii) a representative sample of 628 women between 25 and 45 years of age in greater Colombo and greater Kandy areas, who were out of the labor force, but who were likely to start working within the next year, called potential business owners. These two groups were first interviewed in January 2009, after which 400 current and 400 potential business owners were allocated to treatment. Four rounds of follow-up surveys were then conducted in September 2009, January 2010, September 2010, and June 2011. A fifth and long-term follow-up survey was conducted in June 2015. Details on the survey and study design are described in de Mel et al. (2014).

Since our interest lies in existing businesses, we use follow-up 1 to 5, including current business owners who were operating a business by the first follow-up round and potential business owners who had started a business by then, and who had been assigned to the control group, as our baseline sample.[13] This resulted in the inclusion of 260 firms at baseline in our sample.

We coded *survival* from round 1 to round 2 (i.e. from follow-up 1 to follow-up2) to one, if the owner reported that she was still working in the same line of business as the one she was working in in September 2009. If the owner had changed the line of business of the original business, or was not operating a business anymore, we coded *survival* from round 1 to 2 to zero. *Newfirmstart* was coded to one from round 1 to round 2, if the owner answered that she was either operating a different business or operating a new business along with the old one, after closing, changing or selling the business, and to zero, if she was engaged in a different activity then.

In rounds 3 and 4 (the fourth and fifth follow-up) separate questionnaires were given to females owning only one business and those owning more than one business. For single business owners in rounds 3 and 4, we coded *survival* from round to round to one, if the owner reported that she was still working in the same line of business as the one she was working in in the previous wave. If the owner had changed the line of business of the original business, or was not

---

[13] For current business owners we also use round 1 to extract information on time-invariant characteristics that were not asked in round 2.

operating a business anymore, we coded *survival* from round to round to zero. In round 4, we additionally used information from the attrition section to code *survival*, coding a firm of a single business owner as surviving if a person knowledgeable on the matter said that the owner was still running the business, the business was being run by an immediate family member, or the business was being run by some other person, in case the respondent refused to participate in the survey or could not be located. If the proxy respondent answered that the business was closed, we coded *survival* to zero. We coded *newfirmstart* to one for single business owners in rounds 3 and 4, if the owner answered that she was either operating a different business or operating a new business along with the old one after closing, changing or selling the business, and to zero, if she was engaged in an activity different than operating another business, after closing, changing or selling the business.

For owners of multiple businesses in rounds 3 and 4, we coded *survival* from round to round to one, if:
- none of the businesses operated in the previous round had been closed and business lines for all businesses had not changed since the previous round; or
- after closing, changing or selling one of the businesses operated in the previous round, the owner was either operating a new business along with one previous business, or operating only one of the previous businesses.

We coded *survival* from round to round to zero, for owners of multiple businesses in rounds 3 and 4, if, after closing, changing or selling the businesses operated in the previous round, the owner was either working for a wage, looking for work, operating a different business, or doing housework or looking after children.

In round 4, we additionally used information from the attrition section to code *survival*, coding a firm of a multiple business owner as surviving if a proxy respondent answered that:
- the owner was still running all businesses; or
- the owner was not running all businesses, but at least one; or
- the owner had closed down at least one of the businesses, but not all; or
- at least one of the businesses was being run by an immediate family member; or
- at least one of the businesses was being run by some other person,
if the respondent refused to participate in the survey or could not be located, and coding *survival* to zero if the proxy respondent said that all businesses had closed down.

For owners of multiple businesses in rounds 3 and 4, we coded *newfirmstart* to one, if, after closing, changing, or selling at least one of the businesses, or starting another business, the owner was either operating a different business, or operating a new business along with one or both previous businesses, and to zero, if after closing, changing, or selling the business, the owner was either engaged in wage work, looking for work, doing housework or looking after children, or she was operating only one of the previous businesses, or none of the businesses operated in the previous round had been closed and business lines for both businesses had not changed since the previous round.

In the final round, we coded *survival* to one, if:
- the respondent was currently owning a business and, since January 2009, she had not changed the industry in which the business was operating, or closed down a business she was operating and then started a new business; or
- a proxy respondent reported that the owner was still running the business, or the business was being run by an immediate family member, or the business was being run

by some other person, if the respondent refused to participate in the survey or could not be located.

We coded *survival* to zero in the final round, if:
- the respondent was not owning a business; or
- the respondent said that, since January 2009, she had changed the industry in which the business was operating or she had closed down a business she was operating and then started a new business; or
- a proxy respondent said that the business was closed, if the respondent refused to participate in the survey or could not be located.

We coded *newfirmstart* in round 5 (the final round) to one, if the business had been started after the previous survey round, or, since January 2009, the owner had either changed the industry in which the business was operating, started another business while still operating the original one, or closed down the original business and then started a new business. If the respondent was not operating a business in the final round or the business she was operating had been started before the previous survey round, we coded *newfirmstart* to zero.

According to the criteria described above, *survival* had been coded since baseline only in the second and final round. We hence recoded *survival* in rounds 3 and 4, in order to reflect *survival* since baseline, by recoding it to zero for all subsequent rounds, if it had been coded as zero in a given round and a new firm start had been coded in the same round or later.

Sri Lanka Informal Enterprises Survey (SLINFORMALITY)
The Sri Lanka Informal Enterprises Survey (SLINFORMALITY) was part of a field experiment that provided informal firms incentives to formalize. A sample of 520 informal firms owned either exclusively by men or jointly owned by husband and wife, with one to 14 paid employees in the two largest Sri Lankan cities – Colombo and Kandy – were randomly divided into four treatment groups and a control group. Details on sample selection, experimental and survey design can be found in de Mel et al. (2013).

The baseline survey was conducted in December 2008. Treatment was then assigned randomly, with 105 businesses being assigned to the control group. Three follow-up surveys were carried out in August 2010, March 2011 and December 2011, and we include data on the control firms from all four survey rounds in our sample.

We coded *survival* from round to round to one, if, in a given wave, the owner was still engaged in the same line of business as in the previous survey round, and to zero, if the line of business had changed. In the last round, we were able to use information obtained from proxy respondents to code *survival*, if the owner had refused to participate in the survey or could not be located. If the proxy respondent answered that the owner was still running the business, or the business was being run by an immediate family member, or the business was being run by some other person, we coded *survival* to one. We coded *survival* to zero if the proxy respondent said that the business was closed.

For rounds 3 and 4 we then recoded *survival*, so that it reflected survival since baseline, by recoding it to zero for all subsequent rounds, if it had been coded as zero in a given round and a new firm start had been coded in the same round or later.

Sri Lankan Longitudinal Survey of Enterprises (SLLSE)
The Sri Lankan Longitudinal Survey of Enterprises (SLLSE) is a long-term panel survey of microenterprises designed to assess the impact of a wage subsidy, plus a savings and a business training intervention as supplementary interventions (see de Mel et al, 2010, 2016 for details on the study and survey design). The baseline survey was intended to be a representative survey of urban microenterprises with two or fewer paid employees, owned by males aged 20 to 45, and operating in non-agricultural sectors, in the Colombo, Kandy, and Galle-Matara areas of Sri Lanka. It was conducted between April and October 2008. The first phase occurred in April 2008 as part of a larger panel survey and also included other urban areas in Sri Lanka. In October 2008, a booster listing exercise and survey took place in order to attain a larger sample for the intervention. The original sample was also re-interviewed during the October 2008 survey, resulting in a total sample of 1,533 firms. These firms were then randomly allocated into treatment and control groups, the latter of which consisted of 286 enterprises. After these two survey rounds, the sample was re-interviewed in 10 additional follow-up rounds, consisting of six-monthly surveys every April and October from 2009 through 2012, and two additional surveys in April 2013 and April 2014.

We included firms allocated to the control group that were either part of the booster sample, or surveyed in April 2008 and reported no change in the line of business in October 2008, taking the survey conducted in October 2008 as the baseline. Our baseline sample then comprises 284 firms, which we follow for a total of eleven rounds (baseline plus ten follow-ups).

We coded *survival* from round to round to one, if:
- the respondent reported that he was still working in the same line of business as in the previous round; or
- the business was now being operated by another household member; or
- the respondent had sold the business to someone else to run.

Otherwise, *survival* was coded to zero.

Starting in round 6, we additionally used information from the attrition section to code *survival*, coding it to one if the respondent could not be located or refused to participate in the survey, and the proxy respondent said that the owner was either still operating the business, or the business was being run by another family member or by someone else. If the proxy respondent said that the business had closed down, we coded *survival* to zero.

Newfirmstart was coded to one, if the respondent was operating a different business, after closing, changing or selling the business, and to zero otherwise.

In a final step, we recoded *survival* for rounds 3 to 11, so that it reflected survival since baseline, by recoding it to zero for all subsequent rounds if it had been coded as zero in a given round and a new firm start had been coded in the same round or later.

Lomé Informal Enterprise Survey (TGINFORMALITY)
The Lomé Informal Enterprise Survey (TGINFORMALITY) was used for an impact evaluation of two training interventions for informal businesses in Lomé. The sample was drawn from applicants to the government project under which these trainings were conducted and consists of 1,500 microenterprises. Applicants had to be in business for at least 12 months, have fewer than 50 employees, operate outside of agriculture, and not be a formally registered company. Further details on the intervention, sample selection and survey can be found in Campos et al. (2017). The baseline survey was conducted between October 2013 and January 2014. Firms were then randomly assigned into one of two treatment groups and one control group, each of

which included 500 firms. The first follow-up survey took place in September 2014, followed by a second in January 2015, a third in September 2015 and a fourth in September 2016. We included data for control group firms from all five survey rounds and excluded firms that were in the agricultural sector, resulting in the inclusion of 499 firms at baseline into our sample.

In each of the follow-up rounds, we coded *survival* to one, it the owner reported to still be working for the enterprise he or she operated in the previous round, and to zero if the owner said he or she was not working for this enterprise anymore, either because it was closed down or because of another reason. In round five, we additionally used information for owners, who refused to participate in the survey or could not be located, to code *survival*. If the proxy respondent said that the owner was still working for the enterprise he or she operated in the previous round, we coded *survival* to one, and if the proxy respondent said that the owner was not working for this enterprise anymore, we coded *survival* to zero. *Newfirmstart* was coded to one if the owner was operating a business and was not working for the previous business anymore, and to zero, if the owner was engaged in a different activity than operating a business and not working for the previous business anymore. For rounds 3 to 5, in a final step, we then recoded *survival* to zero, if it had been coded as zero in a given round and a new firm start had been coded in the same round or later, so that it reflected survival since baseline.

Uganda Women's Income Generating Support Impact Evaluation Survey (UGWINGS)
The Uganda Women's Income Generating Support Impact Evaluation Survey (UGWINGS) was conducted to evaluate individual business training, a one-time unconditional cash grant of US$150, supervision, and business advising, provided to very poor Ugandans, mostly women. Details of the intervention and experimental design are described in Blattmann et al. (2016). The baseline survey was conducted on a sample of 1,800 individuals from April to June 2009 who had been selected for the intervention. Of these, 896 were randomized to receive either training, or cash and supervision (phase 1 treatment group), and 904 were randomized to waitlist treatments (phase 2). The first follow-up survey was conducted between November 2010 and February 2011. Phase 2 participants then all received business training and cash and in addition supervision and advise, which was randomized among them. They received a short follow-up survey between September and October 2011 (roughly a month after receiving the grant), and a further, more detailed, follow-up survey between June and August 2012.

Individuals were not asked whether they were already operating a small business at baseline, but in the follow-up surveys conducted between November 2010 and February 2011, and between June and August 2012, respondents were asked whether they had started a small business between the last time they had been surveyed and the current survey. In particular they were asked how many small businesses they had started or had tried to start within this period. They were then asked how many of these were still being operated. We used this information to construct the variables *newfirmstart* and *survival*. In particular, we constructed them in the following way:
  ▪ We coded *newfirmstart* to 1 if the individual reported to have started or tried to start at least one new small business since the last time he or she was surveyed.
  ▪ We coded *survival* to 1 if the individual reported to have started or tried to start at least one new small business since the last time he or she was surveyed and at least one of these businesses was still active. If none of these businesses was active anymore, we coded *survival* to zero.

In the endline for phase-2 individuals (June and August 2012), individuals were not asked if they had started a business between the baseline and the midline survey (conducted between November 2010 and February 2011), so it was not possible to follow the businesses they had

started then until the final round. They were only asked whether they started a new small business between the last time they were surveyed and the current

survey. Furthermore, at that time they had already received treatment. We hence included in our sample data from 461 phase-2 businesses which had been started between round 1 and round 2, and observe them until round 2, which is a maximum horizon of 18 months.

Table A.1. Many Published RCTs of Small Firm Interventions do not discuss Firm Death

Author(s)	Type of Intervention	Firm Size	Country	Sample Size	Time Frame (approximate)	Attrition Rate	Firm Death/Survival is an Outcome	Death Rate in Control Group	Note
Angelucci et al. (2015)	microcredit	micro	Mexico	Total: 2,912	20 months	37%	yes	n.a.	a.
Atkin et al. (2017)	export promotion	micro	Egypt	Total: 219 Control: 145	2.5 years (S1) 15 months (S2)	28% (S1) 7% (S2)	no		b.
Beaman et al. (2014)	other	micro	Kenya	Total: 508	n.a. / n.r.	2-3%	no		c.
Berge et al. (2015)	business training and cash grant	micro	Tanzania	Total: 644 Control: 199	3 years	2.6%	no		
Blattman and Dercon (forthcoming)	business training and cash grant	micro	Ethiopia	Total: 643 Control: 358	1 year	14%	no		d.
Blattman et al. (2014)	cash grants	micro	Uganda	Total: 2,677	4 years	18%	no		
Bruhn et al. (forthcoming)	management consulting	micro, small and medium	Mexico	Total: 432 Control: 282	1 year	12%	no	3.3%	e.
Bruhn and Zia (2013)	business training	micro	Bosnia and Herzegovina	Total: 445 Control: 148	1 year	11%	yes	36%	
De Andrade et al. (2016)	formalization incentives	micro	Brazil	Total: 1,348 Control: 689	1 year	36.8%	yes	14-16%	f.
De Giorgi and Rahman (2013)	formalization incentives	small and medium	Bangladesh	Total: 3,000 Control: 1,500	2 years	n.r.	no		
De Mel et al. (2014)	business training and cash grant	micro	Sri Lanka	Total: 624 Control: 224	2.5 years	7.9%	yes	17%	g.
De Mel et al. (2013)	formalization incentives	micro and small	Sri Lanka	Total: 520 Control: 105	31 months	11%	yes	n.r.	
De Mel et al. (2012)	cash and in-kind grants	micro	Sri Lanka	Total: 387	5.25 years 5.75 years	10% (5.25 years) 8% (5.75 years)	yes	29% (males) 26% (females)	h.
De Mel et al. (2009)	cash and in-kind grants	micro	Sri Lanka	Total: 387	3 years	13.8%	no		
De Mel et al. (2008)	cash and in-kind grants	micro	Sri Lanka	Total: 408 Control: 280	2 years	9.6%	no		
Drexler et al. (2014)	business training	micro	Dominican Republic	Total: 1,193 Control: 387	1.5 years	13%	no		

Dupas and Robinson (2013)	formal savings accounts	micro	Kenya	Total: 392 Control: 197	6-18 months	36%	no		
Fafchamps and Quinn (forthcoming)	other	small	Ethiopia Tanzania Zambia	Total: 126 (E), 134 (T), 85 (Z) Control: 40 (E), 44 (T), 22 (Z)	1 year	n.r.	no		i.
Fafchamps and Woodruff (2016)	business training	small	Ghana	Total: 140 Control:70	2 years	0%	yes	n.r.	j.
Fafchamps et al. (2014)	cash and in-kind grants	micro	Ghana	Total: 793 Control: 396	3 years	14%	no		
Field et al. (2013)	microcredit	micro	India	Total: 845 Control: 425	3 years	9%	yes	38.6%	
Field et al. (2010)	business training	micro	India	Total: 636 Control: 212	4 months	5.3%	no		
Glaub et al. (2014)	business training	micro and small	Uganda	Total: 100 Control: 53	1 year	0%	yes	9.4%	k.
Groh and McKenzie (2016)	insurance against macroeconomic uncertainty	micro	Egypt	Total: 2,961 Control: 1,481	20 months	1.2%	no		
Higuchi et al. (2015)	business training	small	Vietnam	Total: 312 Control: 115	2.5 years	12.2%	no		
Karlan et al. (2015)	management consulting services and cash grants	micro	Ghana	Total: 160 Control: 45	2 years	6.9%	no		l.
Karlan and Valdivia (2011)	business training	micro	Peru	Total: 4,591	1-2 years	24%	no		
Karlan and Zinman (2011)	microcredit	micro	Philippines	Total: 1,601 Control: 329	11-22 months	30%	no		
Mano et al. (2012)	business training	micro and small	Ghana	Total: 167 Control: 107	3.5 years	10.8%	yes	9.3%	k.
Martínez et al. (forthcoming)	business training and in-kind transfer	micro	Chile	Total: 1,948 Control: 566	45 months	23%	yes	75.1%	m.
McKenzie (2017)	large cash grants	micro and small	Nigeria	Total: 541 Control: 263	3 years	28.1% (1 year) 12.8% (2 years) 19.5% (3 years)	yes	13% (1 year) 15.6% (2 years) 24.1% (3 years)	n.

Study	Intervention	Size	Country	Sample	Timeframe	Attrition		Death rate	
McKenzie et al. (2017)	matching grants	small and medium	Yemen	Total: 416 Control: 200	15-17 months	51%	no	1%	k.
McKenzie and Woodruff (2008)	cash and in-kind grants	micro	Mexico	Total: 207 Control: 77	1 year	33.8%	no		
Premand et al. (2016)	business training	micro	Tunisia	Total; 1,702 Control: 851	1.5 years	7.2%	no		
Valdivia (2015)	business training and technical assistance	micro	Peru	Total: 1,979 Control: 565	1 year (FU1) 2 years (FU2)	16.2% (FU1) 19.8% (FU2)	yes	15.7% (FU1) 20% (FU2)	o.

Notes:

n.r. denotes not reported. n.a. denotes not applicable.

Where available attrition rate refers to the control group, otherwise attrition rate refers to the total sample.

a. Only part of the full sample has a panel structure. Sample size, time frame, and attrition reported here refer only to the panel sample. One of the outcomes the authors assess is whether the respondent no longer has a business. The timeframe is indefinite and the outcome is measured as a binary variable equal to 1 if the respondent used to have a business but no longer has one.

b. The authors use different samples in the analysis. S1 denotes sample 1 and S2 denotes sample 2. Sample sizes refer to the combined sample, which is the analysis sample.

c. The intervention consisted of different methods to increase the salience of lost sales due to changeouts (running out of change). Firms were surveyed weekly and the experimental design relied on random variation in the introduction into the study. Hence the number of survey waves and length of follow up vary.

d. There was an additional treatment arm, to which 304 study participants were assigned. We report only on the entrepreneurship program treatment arm and the control group.

e. Data comes from two sources: (1) a one-year follow-up and (7) social security data, covering a horizon of up to 7 years. We do not consider the longer term horizon since the data is not able to assess survival. The death rate of the control group is provided in a footnote in the context of attrition.

f. Sample sizes and attrition only refer to the panel sample.

g. The authors study two groups of women: women who were already operating a business at baseline (current business owners) and women out of the labor force but interested in starting a business (potential business owners). Survival is only assessed for current business owners, and so we focus only on this sample here.

h. Death rates are based on the pooled information from the two long-term follow-up surveys.

i. The intervention randomly created social networks among firm owners.

j. The business training was part of a larger business plan competition. Training was randomly assigned among enterprise owners who presented their project bevor the competition panel. We report only on the sample that presented before the panel.

k. We calculated the death rate in the control group, based on information provided in the article.

l. The authors note that no firm permanently exited the tailoring business during the course of the study, but that some closed their businesses temporarily for health or other reasons.

m. Survival is measured as 1 if the individual was self-employed in 2011 and self-employed in 2013.

n. Survival is only measured for firms that already existed at the start of the intervention. Part of the full intervention sample consisted of entrepreneurs interested in starting a business. We report sample sizes only for the existing businesses sample.

o. Death rates are computed asking business owners whether they closed a business in the last two years.

References for Appendix Table A.1.

Angelucci, Manuela, Dean Karlan, and Jonathan Zinman (2015) "Microcredit impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco", *American Economic Journal: Applied Economics* 7(1):151-182.

Atkin, David, Amit K. Khandelwal, and Adam Osman (2017) "Exporting and firm performance: Evidence from a randomized experiment", *The Quarterly Journal of Economics* 132(2): 551-615.

Beaman, Lori, Jeremy Magruder, and Jonathan Robinson (2014) "Minding small change among small firms in Kenya", *Journal of Development Economics* 108: 69-86.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, and Bertil Tungodden (2015) "Human and financial capital for microenterprise development: Evidence from a field and lab experiment", *Management Science* 61(4): 707-722.

Blattman, Christopher and Stefan Dercon (forthcoming) "The impacts of industrial and entrepreneurial work on income and health: Experimental evidence from Ethiopia", *American Economic Journal: Applied Economics.*

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez (2014) "Generating skilled self-employment in developing countries: Experimental evidence from Uganda", *The Quarterly Journal of Economics* 129(2): 697-752.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar (forthcoming) "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico", *Journal of Political Economy.*

Bruhn, Miriam and Bilal Zia (2013) „Stimulating managerial capital in emerging markets: The impact of business training for young entrepreneurs. *Journal of Development Effectiveness*, 5(2): 232-266.

De Andrade, Gustavo Henrique, Miriam Bruhn, and David McKenzie (2016) "A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms", *The World Bank Economic Review* 31(1): 24-54.

De Giorgi, Giacomo and Aminur Rahman (2013) "SME's registration: evidence from an RCT in Bangladesh", *Economics Letters* 120(3): 573-578.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2014) "Business training and female enterprise start-up, growth and dynamics: Experimental evidence from Sri Lanka", *Journal of Development Economics* 106: 199-210.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2013) "The demand for, and consequences of, formalization among informal firms in Sri Lanka", *American Economic Journal: Applied Economics,* 5(2): 122-150.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2012) "One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka", *Science* 335(6071): 962-966.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2009) "Are women more credit constrained? Experimental evidence on gender and microenterprise returns", *American Economic Journal: Applied Economics* 1(3): 1–32.

De Mel, Suresh, David McKenzie, and Christopher Woodruff (2008) "Returns to capital in microenterprises: Evidence from a field experiment", *The Quarterly Journal of Economics*, 123(4): 1329-72.

Drexler, Alejandro, Greg Fischer and Antoinette Schoar (2014) "Keeping it simple: Financial literacy and rules of thumb", *American Economic Journal: Applied Economics,* 6(2): 1-31.

Dupas, Pascaline and Jonathan Robinson (2013) "Savings constraints and microenterprise development: Evidence form a field experiment in Kenya", *American Economic Journal: Applied Economics,* 5(1): 163-192.

Fafchamps, Marcel and Simon Quinn (forthcoming) "Networks and manufacturing firms in Africa: Results from a randomized field experiment", *The World Bank Economic Review.*

Fafchamps, Marcel and Christopher Woodruff (forthcoming) "Identifying gazelles: Expert panels vs. surveys as a means to identify firms with rapid growth potential", *The World Bank Economic Review.*

Fafchamps, Marcel, David McKenzie, Simon Quinn and Christopher Woodruff (2014) "Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana", *Journal of Development Economics* 106: 211-226.

Field, Erica, Rohini Pande, John Papp and Natalia Rigol (2013) "Does the classic microfinance model discourage entrepreneurship among the poor? Experimental evidence from India", *American Economic Review* 103(6): 2196-2226.

Field, Erica, Seema Jayachandran and Rohini Pande (2010) "Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India", *American Economic Review: Papers and Proceedings* 100(2): 125-129.

Glaub, Matthias E., Michael Frese, Sebastian Fischer and Maria Hoppe (2014) "Increasing personal initiative in small business managers or owners leads to entrepreneurial success: A theory-based controlled randomized field intervention for evidence-based management", *Academy of Management Learning & Education* 13(3): 354-379.

Groh, Matthew and David McKenzie (2016) "Macroinsurance for microenterprises: A randomized experiment in post-revolution Egypt", *Journal of Development Economics* 118: 13-25.

Higuchi, Yuki, Vu Hoang Nam and Tetsushi Sonobe (2015) "Sustained impacts of Kaizen training", *Journal of Economic Behavior and Organization* 120: 189-206.

Karlan, Dean, Ryan Knight and Christopher Udry (2015) "Consulting and capital experiments with microenterprise tailors in Ghana", *Journal of Economic Behavior and Organization* 118: 281-302.

Karlan, Dean and Martín Valdivia (2011) "Teaching entrepreneurship: Impact of business training on microfinance clients and institutions", *The Review of Economics and Statistics* 93(2): 510-527.

Karlan, Dean and Jonathan Zinman (2011) "Microcredit in theory and practice: Using randomized credit scoring for impact evaluation", *Science* 332(6035): 1278-1284.

Mano, Yukichi, Alhassan Iddrisu, Yukuta Yoshino and Tetsushi Sonobe (2012) "How can micro and small enterprises in Sub-Saharan Africa become more productive? The impacts of experimental basic managerial training", *World Development* 40(3): 458-468.

Martínez, Claudia, Esteban Puentes and Jaime Ruiz-Tagle (forthcoming) "The effects of micro entrepreneurship programs on labor market performance: Experimental evidence from Chile", *American Economic Journal: Applied Economics.*

McKenzie, David (2017) "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition", *American Economic Review* 107(8): 2278-2307.

McKenzie, David, Nabila Assaf and Ana Paula Cusolito (2017) "The additionality impact of a matching grant programme for small firms: Experimental evidence from Yemen", *Journal of Development Effectiveness* 9(1): 1-14.

McKenzie, David and Christopher Woodruff (2008) "Experimental Evidence on Returns to Capital and Access to Finance in Mexico", *The World Bank Economic Review* 22(3): 457-482.

Premand, Patrick, Stefanie Brodmann, Rita Almeida, Rebekka Grun and Mahdi Barouni (2016) "Entrepreneurship education and entry into self-employment among university graduates", *World Development* 77: 311-327.

Valdivia, Martín (2015) "Business training plus for female entrepreneurship? Short and medium-term experimental evidence from Peru", *Journal of Development Economics* 113: 33-51.

Table A.2. Firm death rates over different time horizons, pooled sample and individual surveys

Years since baseline	Pooled sample	BENINFORM	EGYPTMACROINS	GHANAFLYP	IFLS	KENYAGETAHEAD	MALAWIFORM	MXFLS	NGLSMSISA	NGYOUWIN	SLKFEMBUSTRAINING	SLKINFORMALITY	SILSE	SLMS	TOGOINF	TTHAI	UGWINGS
0.25	4% [4%,5%]			1% [1%,1%]							10% [10%,10%]			4% [4%,8%]			
0.5	9% [8%,18%]		8% [8%,8%]	3% [3%,5%]					12%* [10%,27%]*					2% [2%,10%]			
0.75	4% [3%,10%]			5% [5%,8%]										4% [3%,14%]	3% [2%,8%]		
1	13% [13%,18%]	8% [7%,15%]		5% [5%,15%]		16% [16%,17%]	12% [11%,18%]			35% [32%,39%]	16% [16%,18%]		5% [5%,11%]	1% [1%,14%]		23%* [21%,28%]*	
1.25	5% [4%,14%]			8% [8%,13%]										3% [3%,16%]	2% [2%,14%]		
1.5	20% [17%,34%]								21% [16%,40%]				7% [7%,17%]	3% [3%,17%]			34% [32%,39%]
1.75	17% [15%,28%]						23% [21%,30%]			40% [28%,59%]	25% [25%,25%]	8% [7%,20%]		4% [3%,20%]	3% [2%,13%]		
2	20% [17%,33%]	9% [8%,19%]							25%* [17%,49%]*				13% [11%,24%]	4% [3%,20%]		35%* [32%,40%]*	
2.5	23% [19%,36%]					24% [23%,29%]			28% [17%,56%]			10% [9%,23%]	20% [19%,21%]	5% [4%,23%]			
3	30% [26%,38%]						21% [18%,31%]	69% [54%,76%]				12% [11%,17%]	23% [23%,25%]	4% [3%,23%]	8% [8%,12%]	44%* [39%,50%]*	
3.5	25% [22%,35%]						22% [19%,31%]			40% [29%,56%]			23% [22%,26%]				
4	36% [32%,43%]												26% [25%,30%]			53%* [45%,59%]*	
4.5	51% [44%,58%]							67% [54%,73%]					27% [26%,31%]				
5	41% [35%,49%]													21% [17%,32%]		59%* [50%,65%]*	
5.5	27% [25%,34%]												29% [27%,32%]	25% [22%,35%]			
6	50% [43%,57%]										35% [35%,37%]					66%* [55%,72%]*	
7	58% [42%,69%]				53% [37%,67%]											74%* [61%,78%]*	
8	65% [47%,74%]				56% [38%,71%]			78% [64%,82%]								79%* [66%,83%]*	

Years since baseline	Pooled sample	BENINFORM	EGYPTMACROINS	GHANAFLYP	IFLS	KENYAGETAHEAD	MALAWIFORM	MXFLS	NGLSMSISA	NGYOUWIN	SLKFEMBUSTRAINING	SLKINFORMALITY	SLLSE	SLMS	TOGOINF	TTHAI	UGWINGS
9	83% [67%,86%]															83%* [68%,86%]*	
10	66% [58%,70%]													39% [37%,43%]		87%* [73%,89%]*	
11	65% [57%,70%]													39% [36%,43%]		90%* [75%,92%]*	
12	90% [74%,91%]															90%* [75%,92%]*	
13	91% [75%,93%]															92%* [75%,93%]*	
14	92% [76%,93%]															93%* [76%,94%]*	
15	77% [53%,84%]				76% [52%,84%]											94%* [78%,95%]*	
16	96% [80%,96%]															97%* [80%,97%]*	
17	100% [80%,100%]															100% [80%,100%]	

Notes:

Death rate bounds accounting for missing information on survival/death, including attrition, are given in parentheses.

* indicates that multiple rounds are available for this combination of survey and time horizon, so that death rates, lower and upper bounds shown are averages of all their respective round-specific values available for this particular horizon.

Observed drops in death rates for a given survey are possible, since some firm closure is temporary (in the sample of unique firms, we observe 6.22 percent as closed at one point in time and open in a later round), and as a result of attrition (1.1 percent of unique firms in our sample are either closed or open in t and t+2, and not observed in t+1).

For the MXFLS, the baselines for the horizons over which we observe firm death are different. We use the 2002-round for the 3-year and 8-year horizon and the 2005-round for the 4.5 year horizon.

Table A.3. Estimates of Firm Death Rate in Existing Literature

Study	Country	Type of firm	Sample size	Period	Death rate	Annualized Rate	Attrition rate	Note
Cabal (1995)	Dominican Republic	micro and small enterprises	n.r.	n.r.	n.r.	22%-29%	n.r.	e.
Daniels (1995)	Zimbabwe	micro and small enterprises	n.r.	n.r.	n.r.	11.5%	42%	e.
Davies and Kerr (2015)	Ghana	All Manufacturing	1000	10 years	21%	2.1%	22%	
		0-9 worker manuf.	386	10 years	24.7%	2.5%	28.5%	
Fajnzylber et al. (2006)	Mexico	self-employed individuals (own-account workers or employers)	n.r.	2 years	n.r.	30.1%	n.r.	
Fisscha (1994)	Jamaica	micro and small enterprises	n.r.	n.r.	n.r.	4.1%	n.r.	f.
Frazer (2005)	Ghana	selected manufacturing	200	2 years	6.3%	3.2%	n.r.	a.
Haddad et al. (1996)	Morocco	Manufacturing firms	n.r.	1 year	4.6%-7.1%	4.6%-7.1%	n.r.	
Kilby (1994)	Nigeria	micro and small enterprises	n.r.	n.r.	n.r.	1.3%	n.r.	f.
Klapper and Richmond (2011)	Cote d'Ivoire	Formal firms	5941	20 years	n.r.	17%	n.r.	b.
Krafft (2016)	Egypt	Household non-farm	828	8 years	51.4%	6.4%	n.r.	
		Household non-farm	1584	6 years	60.9%	10.2%	n.r.	
McCaig and Pavcnik (2016)	Vietnam	Household enterprises	n.r.	2 years	30-35%	15-17%	n.r.	
Mead and Liedholm (1998)	Botswana	micro and small enterprises	n.r.	n.r.	n.r.	6.0%	n.r.	
	Kenya	micro and small enterprises	n.r.	n.r.	n.r.	15.9%	n.r.	
	Malawi	micro and small enterprises	n.r.	n.r.	n.r.	15.0%	n.r.	
	Swaziland	micro and small enterprises	n.r.	n.r.	n.r.	10.5%	n.r.	
	Zimbabwe	micro and small enterprises	n.r.	n.r.	n.r.	7.0%	n.r.	

Study	Country	Description	Sample	Period	Rate	Annualized	Rate	Note
Nagler and Naudé (2017)	Uganda	Household non-farm	1849	1 year	30.8%	30.8%	n.r.	
			1829	1 year	26.2%	26.2%	n.r.	
			1707	4 years	30.5%	7.6%	n.r.	
	Nigeria	Household non-farm	4448	0.5 years	13.2%	26.3%	n.r.	
			5367	0.5 years	1.9%	3.8%	n.r.	
			4516	1.5 years	4.5%	3.0%	n.r.	
Roberts (1996)	Colombia	Manufacturing firms	n.r.	1 year	8.3%-12.9%	8.3%-12.9%	n.r.	
Sandefur (2010)	Ghana	formal and informal manuf. firms with less than 10 employees	n.r.	n.r.	n.r.	5.1%	n.r.	
Shiferwaw (2009)	Ethiopia	Manufacturing with 10+ workers	471-686	1 year	16%	16%	n.r.	c.
Söderbom et al. (2006)	Ghana	Manufacturing firms	154	5 years	19%	3.8%	n.r.	
	Kenya	Manufacturing firms	133	5 years	39%	7.8%	n.r.	
	Tanzania	Manufacturing firms	102	5 years	44%	8.8%	n.r.	
Tybout (1996)	Chile	Manufacturing firms	n.r.	1 year	7.1%-18.2%	7.1%-18.2%	n.r.	
Vijverberg and Haughton (2004)	Vietnam	Household enterprises	2484	5 years	61%	12.2%	9.6%	d.
Mean Annualized Rate						11.5%		
Median Annualized Rate						9.5%		

Notes:

n.r. denotes not reported.

Annualized rate is expressed relative to initial stock, and is death rate divided by period over which death is measured.

Sample size reported for experimental studies is number in the control group.

a. Author doesn't report attrition, but reports sample of 800 with only 479 observations on survival status.

b. Death is measured from filings at register of companies, so authors cannot distinguish death from becoming informal/not reporting.

c. Death is measured from manufacturing census, cannot distinguish between dropping below size threshold from death.

d. Attrition rate does not include firms in households that were re-interviewed, but where authors were not able to match the firm over time.

e. Cited in Mead and Liedholm (1998)

f. Cited in Liedholm and Mead (1995). Death rates are based on tracer surveys which are subject to severe selectivity biases that potentially understate these death rates.

Table A.4. Overview of the follow-up periods covered, by survey

Years since baseline	BENINFORM	EGYPTMACROINS	GHANAFLYP	IFLS	KENYAGETAHEAD	MALAWIFORM	MXFLS	NGLSMSISA	NGYOUWIN	SLKFEMBUSTRAINING	SLKINFORMALITY	SLLSE	SLMS	TOGOINF	TTHAI	UGWINGS
0.25			x							x			x			
0.5		x	x					x					x			
0.75			x										x	x		
1	x		x		x	x		x	x			x	x		x	
1.25			x										x	x		
1.5								x				x	x			x
1.75						x			x	x	x		x	x		
2	x							x				x	x		x	
2.5					x			x				x	x			
3						x	x				x	x	x	x	x	
3.5						x			x			x				
4												x			x	
4.5							x					x				
5													x		x	
5.5												x	x			
6									x						x	
7				x											x	
8				x			x								x	
9															x	
10													x		x	
11													x		x	
12															x	
13															x	
14															x	
15				x											x	
16															x	
17															x	

Table A.5. Attrition and missing information on survival by follow-up horizon

Years since baseline	Percent attrited	Percent missing	Percent attritted + missing
0.25	1,1%	0,1%	1,2%
0.5	2,4%	7,1%	9,5%
0.75	6,1%	0,0%	6,1%
1	4,8%	0,0%	4,8%
1.25	9,2%	0,7%	9,9%
1.5	6,3%	0,4%	6,7%
1.75	10,0%	0,6%	10,6%
2	10,0%	12,0%	21,9%
2.5	7,0%	9,5%	16,5%
3	9,5%	3,6%	13,1%
3.5	11,6%	0,0%	11,6%
4	5,7%	0,0%	5,7%
4.5	5,6%	0,0%	5,6%
5	15,4%	0,0%	15,4%
5.5	9,0%	0,0%	9,0%
6	4,1%	0,0%	4,1%
7	20,0%	0,0%	20,0%
8	15,5%	13,8%	29,3%
9	20,0%	0,0%	20,0%
10	7,4%	0,0%	7,4%
11	8,0%	0,0%	8,0%
12	20,0%	0,0%	20,0%
13	20,0%	0,0%	20,0%
14	20,0%	0,0%	20,0%
15	19,3%	12,4%	31,6%
16	20,0%	0,0%	20,0%
17	20,0%	0,0%	20,0%

Table A.6. Reason for closure by gender

Reason for closure	Male	Female	Total
Making a loss	49.34%	35.50%	590
Sickness	9.79%	15.69%	197
Care for family	2.45%	18.07%	180
Better wage job	8.10%	5.63%	95
Better business opportunity	7.91%	3.14%	71
Other	22.41%	21.97%	322
Total	531	924	1,455

Table A.7. Gender of owner by reason of closure

| Reason for closure	Male	Female	Total
Making a loss	44.41%	55.59%	590
Sickness	26.40%	73.60%	197
Care for family	7.22%	92.78%	180
Better wage job	45.26%	54.74%	95
Better business opportunity	59.15%	40.85%	71
Other	36.96%	63.04%	322
Total	36.49%	63.51%	1,455

Figure A.1. Comparison of our Firm Death Rates to Those Reported in Existing Literature
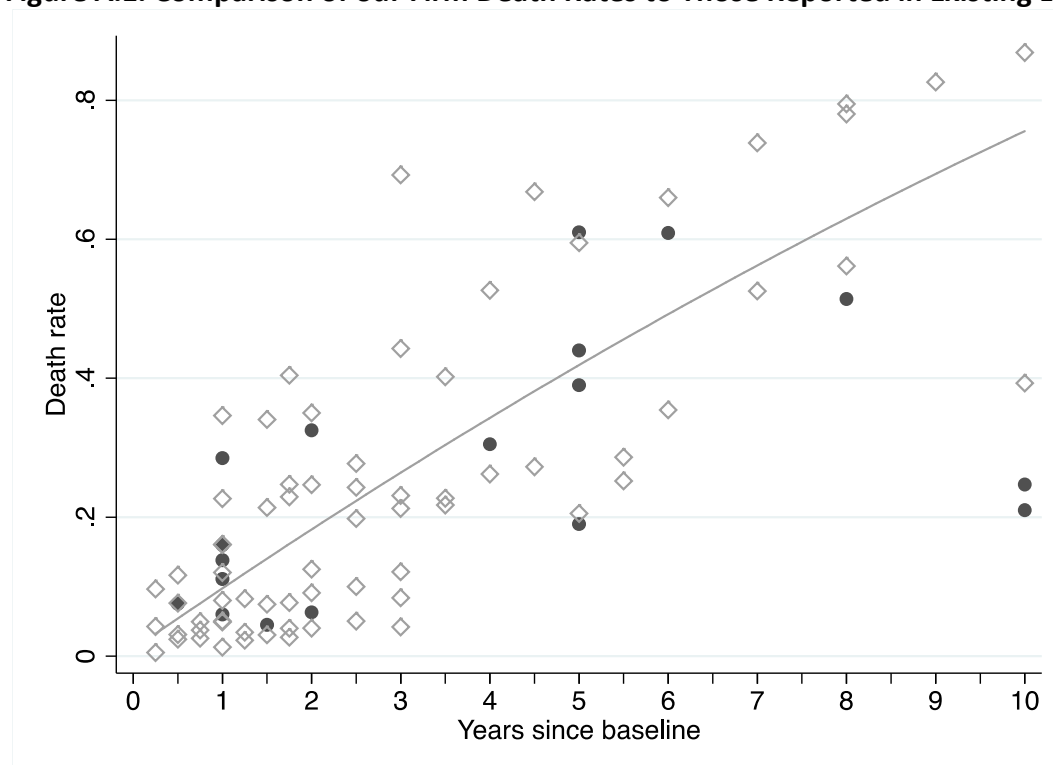
Note: Black circles are death rates from existing literature in Table A.4. Open diamonds are data points in our study. Curve shown is fitted quadratic in Figure 1. The two outliers in the existing literature at 10 years are from Davies and Kerr (2015).

Figure A.2. Reopening rate over different horizons for firms that closed within 1 year
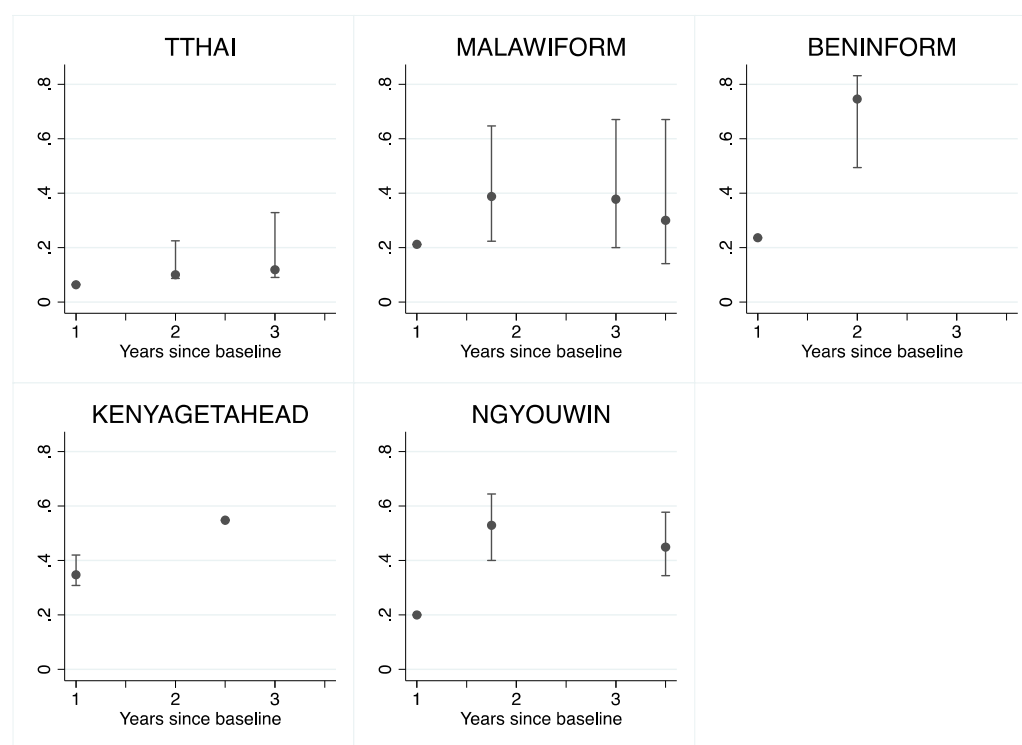
Notes: only surveys in which 85 or more firms died within a year and subsequent re-opening can then be observed are shown.

Figure A.3. Comparison of labor productivity at baseline for firms that closed vs. survivors, by survey