NBER WORKING PAPER SERIES

ENTRY AND EXIT OF INFORMAL FIRMS AND DEVELOPMENT

Brian McCaig
Nina Pavcnik

Working Paper 28986
http://www.nber.org/papers/w28986

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2021

We thank Adam Rivard, Tess Snyder, Justin Han, and Archita Harathi for excellent research assistance, the editor, two anonymous referees, and participants at the Structural Transformation and Economic Growth Theme 0 workshop and Universitat Autonoma de Barcelona for helpful comments. An earlier version of this project circulated under the title "Out with the old and unproductive, in with the new and similarly unproductive: Microenterprise dynamics in a growing low-income economy." That earlier version is an output from a project funded by the UK Department for International Development (DFID) and the Institute for the Study of Labor (IZA) for the benefit of developing countries. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Entry and Exit of Informal Firms and Development
Brian McCaig and Nina Pavcnik
NBER Working Paper No. 28986
July 2021
JEL No. J46,L2,O17,O53

ABSTRACT

Non-farm informal businesses comprise the majority of the firm distribution in developing countries. We document novel stylized facts about entry and exit of informal, non-farm firms using nationally representative panel data over 15 years and across regions with varying levels of local economic development in Vietnam. First, we find that informal businesses exhibit rates of entry and exit around 14-18% annually. Entry and exit rates are similar and highly correlated at a point in time, within industries, and within regions. They both decline over time and across space with economic development. Second, although market selection influences which firms survive, entry and exit has little net effect on aggregate (revenue) productivity or hiring of workers outside the household. This owes to overlapping labor productivity of entering and exiting firms and low subsequent productivity growth and hiring among the surviving entrants. Nonetheless, entry and exit are associated with large changes in individual income. Third, the large overlap in revenue of entering and exiting informal businesses and the high correlation between entry and exit rates are related to the education of owners and their economic activities before and after operating an informal business. Informal business owners are less educated on average than wage workers in the formal sector, but more educated than agricultural workers. The transitions in and out of operating an informal business reflect the underlying structure of economic activities of the working age population, with education gaps also playing a role. The most common transition into non-farm businesses is to and from self-employment in agriculture. The likelihood of this transition declines with economic development, highlighting the role of net entry from agriculture into informal non-farm businesses in structural change.

Brian McCaig
Department of Economics,
Wilfrid Laurier University
75 University Avenue West
Waterloo, ON, N2L 3C5 Canada
bmccaig@wlu.ca

Nina Pavcnik
Department of Economics
6106 Rockefeller Center
Dartmouth College
Hanover, NH 03755
and NBER
nina.pavcnik@dartmouth.edu

1 Introduction

A large literature has emphasized the importance of firm entry and exit in developing countries (see Roberts and Tybout (1996), Tybout (2000)), highlighting their contribution to aggregate productivity and their role in firms' responses to policy changes (such as trade policy) and business cycles. Because of data constraints, most of the literature on facts about firm entry and exit focuses on formal establishments (Li and Rama (2015)).[1] In contrast, we have less systematic evidence on entry and exit of informal firms (Woodruff (2012), Li and Rama (2015), McKenzie and Paffhausen (2019), Ulyssea (2020)).

Why should one focus on entry and exit of informal businesses? These businesses comprise the majority of firms in developing countries (Gollin, 2002, 2008; La Porta and Shleifer, 2008, 2014; Hsieh and Olken, 2014).[2] They influence aggregate productivity because they tend to be less productive than larger, formal firms (La Porta and Shleifer, 2008, 2014; McCaig and Pavcnik, 2018) or face different distortions in labor and capital markets, leading to potential misallocation of resources (Atkin and Khandelwal, 2020). Importantly, informal firms matter for the livelihoods of households in poorer countries. They are a notable source of household income (Benjamin *et al.*, 2017) and account for a large share of employment (Gollin, 2002, 2008; Nataraj, 2011).[3] Their performance and dynamics are therefore key to understanding the livelihoods of poor households, generation of jobs, and aggregate labor productivity in low-income countries.

In this paper, we document several facts about the entry and exit of informal non-farm businesses. We overcome the usual data constraints by using a large sample of almost 130,000 informal non-farm businesses based on nationally representative household surveys over a period of 15 years in Vietnam.[4] Although we focus on a single country, we observe informal

[1] Firm-level datasets usually cover registered firms or registered firms above a certain employment threshold in manufacturing.

[2] For example, in Vietnam in 2010, there were 9.1 million informal businesses compared to 287 thousand formal firms. These figures are based on authors' calculations using the 2010 Vietnam Household Living Standards Survey for the estimate of informal businesses and the 2010 Enterprise Census for the estimate of formal businesses.

[3] For example, they account for 21% of household income in Vietnam (Benjamin *et al.*, 2017). They account for 80% of manufacturing employment in India (Nataraj, 2011). In Vietnam, 58% of non-agricultural workers worked in informal businesses at the time of the 2009 population census (authors' calculation).

[4] We focus on informal firms precisely because they are usually excluded from other datasets and they account for a large share of employment and household income in poor countries. However, we do not focus

businesses in about 60 local economies within Vietnam with levels of economic development ranging from about 1800 to 12,000 USD and differing sectoral compositions of employment.[5] The panel component allows us to follow about 48,000 informal firms over a period of up to four years (something most studies cannot do). We more systematically document prevalence and facts about entry and exit of informal firms over longer time periods, across industries, and with levels and sectoral structure of economic development than is usually possible with smaller, not nationally representative samples. Importantly, we can simultaneously examine the rate of entry and exit of informal businesses and the implications of entry and exit for labor productivity, paid employment and household income. Our results thus complement McKenzie and Paffhausen (2019) which focuses on small firm exit.

Finally, we relate entry and exit patterns of informal businesses to the education and economic activities of their owner before or after operating a business. Most informal businesses have only one worker, the owner, who is the business's key input (Gollin, 2008; La Porta and Shleifer, 2008, 2014; de Mel *et al.*, 2010). This person owns the business, works at the business, and manages its daily operations. We perform these comparisons by linking the longitudinal data on informal businesses to a panel of households and individuals. Importantly, we systematically compare education and economic activities of informal business owners that are starting or shutting down a business to the education and economic activities of the overall working-age population. This enables us to examine selection of individuals into and out of informal businesses in a comprehensive fashion over long periods of time and across locations. Our work complements studies that have largely focused on the largest cities (de Mel *et al.*, 2010) or urban centers in middle-income countries (Maloney, 2004).

We find that informal businesses exhibit rates of entry and exit of around 14-18% annually, which is higher than what is typically observed for formal firms in developing countries. Entry and exit rates are similar and highly correlated at a point in time, within industries, and within regions that vary in economic development. They both decline with economic development over time or across space. Second, although market selection influences which firms survive, informal business entry and exit has little net effect on aggregate (revenue)

on why these firms are informal. See Ulyssea (2018, 2020) for recent contributions to that literature.

[5]The amounts are from 2010 and based on GDP per capita, PPP adjusted in constant 2017 USD. See section 4 for further details on provincial GDP per capita estimates.

labor productivity or hiring of workers outside the household. We find that there is a large overlap in the revenue and total factor productivity distribution of entrants and exiters. Moreover, almost half of the entrants exit within two years and the surviving entrants do not significantly improve their performance over time or begin to hire paid workers. Overall this suggests that market-selection is one factor influencing entry and exit of informal businesses, with a broader range of owner-specific characteristics and shocks (Maloney, 2004; Hurst and Pugsley, 2011; Astebro *et al.*, 2014), alternative employment options (La Porta and Shleifer, 2008, 2014), and household-specific shocks (such as age of the owner, illness in the family) potentially also playing a role. Importantly, business exit and entry is associated with large decreases and increases, respectively, in individual and household income.

Third, the large overlap in revenue of entering and exiting informal businesses and the high correlation between entry and exit rates are related to the education and economic activities of business owners prior to starting, or after closing down, an informal business. Informal business owners are less educated on average than wage workers in the formal sector, which is consistent with Gollin (2008), but have higher levels of education than workers in agriculture. Entering and exiting owners have very similar levels of education. Additionally, they are more likely to transition to and from other activities with similar levels of educational attainment.

The economic activities that owners transition to and from reflect the structure of economic activities of the working age population at a point in time and location.[6] Owners most often transition to and from self-employment in agriculture, accounting for about a third of informal non-farm business entrants and exiters. Net entry from self-employment in agriculture is consistently positive and the rate falls with economic development, both over time and locations. This underscores the role of net entry out of agriculture into informal non-farm businesses in the process of structural change. Transitioning to and from the formal sector, which accounts for about 10% of entrants and exiters, is less common than the overall prevalence of work in the formal sector (in part because formal wage work requires

[6]Economic activities include self-employment in agriculture, self-employment in the informal sector, self-employment in the formal sector, wage work in agriculture, wage work in the informal sector, wage work in the formal sector, and not in the workforce. Our data also enables us to take into account measurement of informal business entry and exit due to changing composition of the household.

higher education).

Overall, these results highlight the importance of simultaneously examining entry and exit of informal businesses with nationally representative data that fully captures the variation of economic activities (and thus outside opportunities for informal business owners) over time and locations.

The paper proceeds as follows. Section 2 summarizes related theoretical and empirical literature. Section 3 describes the legal framework for private businesses in Vietnam and data. Section 4 provides background information on the prevalence of informal businesses over time and across locations. Sections 5 and 6 report patterns of informal business entry and exit and the economic activities of owners before and after operating an informal business.

2 Related Literature

In this section we discuss theoretical and empirical literature related to informal firm entry and exit in developing countries and the relation of our study to this literature.

2.1 Theoretical reasons for informal firm entry and exit

What determines entry and exit of informal firms in low-income countries?

We focus on entry and exit of informal firms because they are usually excluded from datasets of formal registered firms used for studying firm entry and exit in low-income countries. However, they account for a large share of employment and household income in poor countries. Most informal firms are very small, often operated by only the owner. Because our focus on informal firms is motivated by their exclusion from the data (rather than their informal status), we provide an overview of what drives the existence, and entry and exit decisions, of small firms as opposed to the informal status of most small firms.[7]

[7]We do not focus on why these firms are informal rather than formal. Recent papers studying the decision to formalize and discussing conceptual issues around formalization include Maloney (2004), Levy (2008), de Mel *et al.* (2013), La Porta and Shleifer (2014), and Ulyssea (2018). See Ulyssea (2020) for a recent review.

Lucas (1978) and Gollin (2008) tie the existence of small firms to economic development. They posit models where individuals vary in their entrepreneurial ability and this variation leads to endogenous selection between wage work (the least entrepreneurial) and entrepreneurship (the most entrepreneurial). Gollin (2008) features an intermediate case, self-employment without employees, where the self-employed can split their time between being a wage worker and managing their own business, a common occurrence in many low-income countries. The models suggest that with low levels of wealth or aggregate productivity, the average firm size will be small, consistent with the empirical firm size distribution in many low-income countries.[8] As the economy develops, average firm size will increase as marginal entrepreneurs (or the self-employed) find outside wage opportunities to be more attractive. This suggests declines in self-employment and a reduction in the number of the smallest firms with development. This view of the world is consistent with the idea that many individuals in developing countries run an informal business out of necessity and lack of other employment opportunities and that these individuals tend to switch to paid work as the economy develops (La Porta and Shleifer, 2008, 2014).

Other frameworks link the existence of smaller firms to a market imperfection, whereby productive entrepreneurs are precluded from expanding due to imperfections in the capital market, barriers to hiring labor outside the family, costs of registration and formalization, or lack of business training (see Banerjee (2013) and Banerjee *et al.* (2015) for the impacts of microcredit, de Mel *et al.* (2019) and McCasland and Hardy (2020) for hiring barriers, Bruhn and McKenzie (2014) for registration, and McKenzie and Woodruff (2014) for business training). To the extent that policy interventions reduce these distortions, these businesses could better flourish, expand, and formalize.

Beyond entrepreneurial ability or luck and market frictions, several studies have emphasized the role of individual characteristics such as nonpecuniary tastes, risk aversion, and overconfidence to explain entry and persistence of entrepreneurs (Maloney, 2004; Hurst and Pugsley, 2011; Astebro *et al.*, 2014). Importantly, these studies note the large heterogeneity in why individuals choose to run a business.

[8]Note that the small firm size with low levels of development in these models is not due to distortions (Atkin and Khandelwal, 2020).

What determines entry and exit of businesses? Hopenhayn (1992) and Jovanovic (1982) build on Lucas (1978), but propose a model of business entry as a process of learning. In this setting, a business owner needs to operate the business to learn about his or her abilities and profitability of the business, as opposed to knowing his or her productivity as in Lucas (1978). These models predict simultaneous entry and exit of businesses, with the less successful businesses exiting.

This discussion suggests that the decision to start a business could be due to the owner having a good business idea or foreseeing a market opportunity. The decision to exit could be driven by firm-specific shocks (such as low productivity, increased product market competition, increase in costs of inputs). These market-selection related reasons for entry and exit of informal firms are similar to the considerations of formal firms. However, informal businesses often have only one worker, the owner, who is the business's key input. Hence their entry and exit might be closely tied to the owner's characteristics, alternative employment opportunities (Gollin, 2008; La Porta and Shleifer, 2008), or household-specific shocks (such as age of the owner or illness in the family). As noted by Ulyssea (2020), we know little about the dynamics of informal firms. The discussion above suggests that the entry and exit of informal firms might not be related just to market-selection mechanisms, but also to the owner's characteristics, individual and owner specific shocks, and labor market opportunities available to owners.

Why would economic development be related to entry and exit rates? Higher levels of economic development might be associated with more or better outside opportunities for business owners, leading to lower voluntary entry (and thus also) lower exit rates. Models from industrial organization likewise suggest that entry (and thus) exit rates are lower in tougher/more competitive markets, highlighting the positive correlation between entry and exit rates within an industry or a local market. Higher economic development is associated with fewer constraints on informal business operations (such as access to credit), thus higher entry and lower exit rates (McKenzie and Paffhausen, 2019). Overall, this discussion highlights the importance of simultaneously studying the entry and exit of firms.

2.2 Empirical evidence on informal firm entry and exit

Our study relates to several empirical literatures. A large literature examines the effect of a wide range of input-market interventions on the performance of microenterprises.[9] These interventions include microcredit provision, business training, and easing formalization requirements. As noted by McKenzie and Paffhausen (2019) most of the above studies do not focus on firm exit.[10]

The challenges in examining the dynamics of microenterprises and transitions of businesses over long periods are well known (de Mel *et al.*, 2010, 2013; Woodruff, 2012; Pages, 2012). As a result, much of what we know about informal firm entry and exit stems from panels of household businesses over short periods of time.[11] Alternatively, people have used labor force surveys from urban areas in middle-income countries in Latin America (see a survey by Maloney (2004), Mondragón-Vélez and Peña (2010)), or the Townsend Thai Project.[12] We contribute to this literature by systematically examining entry and exit over a long period of time and localities with a broad range of economic development.

Our study is most closely related to McKenzie and Paffhausen (2019), which studies small firm exit. It provides descriptive analysis of why firms exit, characteristics of those that exit (relative to existing business owners), and how exit varies with economic development using samples that cover 14,000 small firms across 12 countries.[13] Our study is complementary, providing simultaneous analysis of entry and exit using nationally representative data. Although our data covers a single country, we are able to follow businesses that operate in local economies that differ in income per capita and the sectoral distribution of employment.

[9]The definition of a microenterprise varies across studies, but will typically cover most, if not all, informal businesses since few informal businesses ever become very large. See surveys by Banerjee (2013) and Banerjee *et al.* (2015) for the impacts of microcredit, Bruhn and McKenzie (2014) for registration, and McKenzie and Woodruff (2014) for business training and Jayachandran (2020) for a comprehensive review of microenterprises in developing countries. Most of the studies find little evidence of large impacts of these interventions on firm growth or job creation.

[10]For example, less than a third of recently published randomized control trials report an exit rate or uses microenterprise exit as an outcome variable (McKenzie and Paffhausen, 2019).

[11]See the extensive review in McKenzie and Paffhausen (2019) and Vijverberg and Haughton (2002) for evidence from Vietnam during an earlier period, a one-period panel between 1993 and 1998.

[12]See, for example, Paulson and Townsend (2004); Paulson *et al.* (2006); Kaboski and Townsend (2011, 2012); Samphantharak and Townsend (2012).

[13]These samples continue to be small and not nationally representative. See McKenzie and Woodruff (2014) for a discussion of issues related to small and selective samples.

We also relate to an older literature on the facts of entry and exit among firms in developed countries as well as formal firms in developing countries. Using data from the U.S. manufacturing census, Dunne *et al.* (1988) reports entry rates over 5-year periods of around 0.52 (an annualized rate of about 0.14) and exit rates over the same periods of 0.45 to 0.50 (annualized rates of 0.11 to 0.13), with higher entry and exit rates among small manufacturing plants. Dunne *et al.* (1989) reports that exit rates decrease with plant age and size as do growth rates, conditional on surviving. Tables A9 through A14 report exit and entry rates across various studies and are organized by microenterprises (developing countries) and formal firms (from developing and developed countries). The annual entry and exit rates among microenterprises are much more varied than among formal firms, ranging from about 8 to 32% annually for entry and from 3 to 31% annually for exit. Entry and exit rates for formal firms in developing countries are almost as varied, ranging from 4 to 26% annually for entry and from 3 to 17% annually for exit.[14] Lastly, among developed countries, entry rates range from 5 to 20% and exit rates vary between 6 and 11%.

Our paper also relates to a literature on firm growth, particularly for small firms. Astebro *et al.* (2014) and Haltiwanger (2015) report that a small share of new firms experience significant growth, while the majority remain small. Similar evidence has been found in India and Mexico by Hsieh and Klenow (2014). They follow cohorts of informal and formal firms and report dramatically lower growth of these firms relative to in the US. McCaig *et al.* (2021) track private, formal firms over 16 years in Vietnam and report similarly low average growth rates, especially in comparison to foreign-invested firms. Among informal firms and microenterprises in developing countries, the consensus is that few of these firms grow (McKenzie, 2017; Fafchamps and Woodruff, 2017; Diao *et al.*, 2018).

Finally, while most of the literature on the consequences of trade policy or export promotion has focused on firms and workers in the formal sector, there is a nascent literature on the link between microenterprises, trade policy (and other market access interventions), and transitions from the informal to the formal sector (Nataraj, 2011; Brambilla *et al.*, 2012; Startz, 2016; Atkin *et al.*, 2017; McCaig and Pavcnik, 2018; Bergquist and Dinerstein, 2020).

[14]Note that the highest entry rates among formal firms in developing countries are found in countries transitioning away from a centrally planned economy, such as Estonia, Slovenia, and Vietnam.

The facts about informal firm entry and exit provide a useful input for understanding how informal firms and their owners are affected by general equilibrium changes brought about by policy changes.

3 Informal firms in Vietnam and data description

3.1 Informal firms in Vietnam

Our analysis focuses on informal businesses, which are referred to as household or individual businesses in Vietnam.[15] In Vietnam, a household or individual business is one that is not registered with the national government under the Enterprise Law. Hence, our definition of an informal business is based on the registration status of the business.[16] All state, foreign and collective businesses are legally required to register as enterprises under Vietnam's Enterprise Law.[17] However, private businesses can legally operate in two different modes, a household business and a private enterprise, depending on their size and number of locations. All private businesses with 10 or more regular workers or more than one location are legally obligated to register as an enterprise.[18] Malesky and Taussig (2009) report that enterprises, relative to household businesses, have easier access to export licenses, customs certificates, opportunities to bid on government contracts, the right to open branches and to operate outside their home district, but running an enterprise, as opposed to a household business, entails the registration cost and more rigorous accounting.[19] Enterprises (formal firms) are also required to contribute to social insurance for their workers McCaig and Pavcnik (2018).

[15] Vietnam transitioned from a low- to lower middle-income country according to the World Bank's classification in 2010.

[16] The definition of an informal business based on registration status is common in the literature (La Porta and Shleifer, 2008, 2014; Ulyssea, 2020).

[17] See the Law on Enterprises passed in 1999.

[18] See Decrees No. 02/2000/ND-CP and No. 109/2004/ND-CP.

[19] Similarly, Taussig and Hang (2004) reports the benefits of being an enterprise, relative to a household business, as greater ability to trade beyond their home district, ability to expand, value added tax receipts, legal ability to establish branch locations, a stamp for making transactions more official, more predictability, law based interactions with government, ability to access equity for limited and joint stock companies, and greater access to government investment incentives while the costs of formalization include registration costs, annual registration fee, certified chief accountant, greater reporting requirements, potential for increased attention from local authorities, and potential for increased taxes with movement from lump sum to standard tax calculations. The information on the costs of registering as a private enterprise in Vietnam is summarized by the World Bank's Doing Business Survey.

3.2 Household survey description

We use the 2004 to 2018 rounds of the Vietnam Household Living Standard Surveys (VHLSS). The surveys are conducted every two years. These household surveys are nationally representative and contain a detailed business module on non-farm businesses owned and run by the households. The surveys capture both formal (i.e., private enterprises) and informal businesses operated by the household. Our data thus enables us to study a representative sample of informal businesses owned by households, which is rare. Most of these businesses are usually not covered in the more widely available firm-level data, which tend to focus only on registered (or formal) firms or formal firms above a certain employment size cut off.

Each survey round contains information on about 20,000 businesses. The business modules consistently collect the following information for all businesses, both formal and informal: the industry of operation, the number of months it operated during the past 12 months, monthly revenue (including the value of self-consumption), the annual wage bill, and other annual expenditures (materials, repairs, water, etc.).[20] For all surveys except 2004, it also asked whether the business was registered as an enterprise (our definition of formal). The 2006 and 2008 surveys additionally contain information on who is the most knowledgeable person (hereafter referred to as the owner), the number of workers, the number of paid workers, and the year the business started. Information on the number of workers, the number of paid workers, and the start year is available in the 2004 survey for only about 1/5th of the businesses. For the surveys other than 2006 and 2008, we have designed an algorithm to predict the owner of the business within the household. Since most households operate only one business and most businesses have only one individual working in the business, predicting the owner is straightforward for most businesses. Appendix B contains further details.

Importantly, the surveys contain panel components that allow us to examine entry and exit using large, nationally representative datasets covering 15 years. Table 1 displays the number of households in each panel.[21] Between each successive survey, with the exception of

[20] According to the Operational Handbook for the 2004 survey, "The enumerator interviews those people who have the most information about each activity [business], who usually coordinate or manage these activities." The handbooks for later surveys provide the same guidance to the enumerators.

[21] The VHLSSs feature a rotating panel by enumeration area. Thus, not all enumerations areas surveyed

between 2008 and 2010, there is a panel of approximately 20,000 households.[22] Additionally, we have 3-survey panels of approximately 9,500 households. These panels start with the 2004, 2010, 2012, and 2014 surveys. Nationally representative longitudinal data on informal businesses owned by households that spans four years is very rare in low-income countries (Woodruff, 2012). Longitudinal data is useful for two reasons. First, it provides for a representative sample of households who could potentially start a business, an important piece of information not available in the usual firm-level data sets. This enables us to measure informal business entry and exit, correlates of informal business formation, and to follow informal businesses over time. The household surveys include data on all businesses owned by a household, regardless of their formal registration status. As a result, we can distinguish between true exit and formalization. Second, the household panel allows the creation of a panel of businesses. We create a panel of businesses operated by the household by taking advantage of the information on the number of businesses a household operates and the information about the business that is not likely to change in a short period.

Table 1: Number of households in each panel

Panel	Number of households
2004-06	21714
2006-08	21005
2010-12	20929
2012-14	20719
2014-16	20766
2016-18	20453
2004-08	9682
2010-14	9732
2012-16	9350
2014-18	9426

For each two-survey panel, Table A3 shows the number of businesses operated by panel

in 2004 were intended to be resurveyed in 2006, for example, but all households in a panel enumeration area were to be surveyed again. This accounts for why the number of panel households is noticeably lower than the total number of households surveyed in each VHLSS, as only about half of the enumeration areas surveyed in 2004 were surveyed again in 2006, for example.

[22]The authors have cleaned and verified the originally suggested household and individual panels. Details of the cleaning can be found on McCaig's website for the 2004 through 2008 VHLSSs while details for the 2010 through 2018 surveys can be found in Appendix B.

households at the beginning of the panel, at the end of the panel, and the number that were matched between surveys (see Appendix B for details on the matching procedure). In this context, we define entry as a business that was operated at the end of the panel, but cannot be matched to a business operating within the household at the beginning of the panel. Similarly, we define exit as a business that was operated at the beginning of the panel, but cannot be matched to a business operating within the household at the end of the panel.[23] Our definition of informal business entry and exit could lead to an overestimation due to individuals joining and leaving households. For example, consider an individual who was a member of a household and was operating a business, but then left the household and took their business with them. In this case, we no longer observe the business and we record it as a business exit when in fact the business continues to be operated, but no longer by a household member. To address these concerns, when we present entry and exit rates we present alternative estimates that remove from the calculations businesses operated by individuals that join or leave the household.

Finally, we link the business module to the information about its key input, the owner, from the survey modules related to demographics, education, and employment. For each working individual, the employment module collects information on the industry, occupation, ownership category, hours worked during the past 12 months, and wages and other payments for both the most and second most time consuming jobs during the past 12 months. Demographic modules contain information on the individual's age, gender, ethnicity, completed level of education, and location. We also use information from the employment modules to measure the number of household members working in the business.[24] Note that the VHLSSs also include a longitudinal dimension at the individual level.[25] The individual panel enables

[23]Appendix B discusses in detail the high degree of consistency between the employment information reported in the employment module by members of the household and the business information reported in the business module. In other words, for a business to be missing from those reported by the household, and thus contributing to false exit or entry, it must be unreported in both modules. Nonetheless, this is a possibility.

[24]This is based on the household member reporting being self-employed in the same industry as the business. We cannot do this for household members for which self-employment in a non-farm business is neither their primary nor secondary job during the past year. Additionally, we cannot do this for businesses that operate in the same industry as another business in the household since we do not know which household member to assign to which business. Overall, we are able to estimate the number of household members working in the business for 91% of informal businesses in our sample.

[25]The surveys do not follow individuals that left a household, although the datasets often provide some

us to examine economic activities of business owners prior to starting and after closing down a business.

Table 2 displays summary statistics from the repeated cross sections of informal businesses. In the appendix, we provide summary statistics from all businesses, including formal businesses, in the repeated cross sections (Table A4).

Table 2: Informal business and manager characteristics, repeated cross sections

	2006	2008	2010	2012	2014	2016	2018
Business characteristics							
Primary industries	.008	.008	.006	.005	.004	.002	.001
Secondary industries	.267	.259	.242	.231	.235	.219	.222
Tertiary industries	.725	.732	.751	.764	.761	.779	.777
Urban	.347	.35	.389	.401	.394	.401	.391
Operated for past 12 months	.593	.631	.64	.67	.676	.706	.733
Annual revenue	86.6	100.4	117.4	135.4	143.6	154.2	169.3
Annual expenses	53	60.6	67.7	78.3	81.8	79.7	89.2
Labor expense share	.03	.032	.038	.039	.041	.044	.045
Positive labor expenses	.09	.097	.108	.115	.11	.115	.116
Household workers[1]	1.33	1.32	1.33	1.34	1.31	1.31	1.3
Workers	1.59	1.62					
Paid workers	.29	.31					
Age	8.183	8.492					
Manager characteristics							
Primary job	.766	.79	.79	.817	.793	.809	.814
Female	.589	.586	.577	.577	.578	.575	.565
Age	41.7	42.8	41.9	42.9	43.9	44.8	45.8
Grade	7.57	7.63	7.78	7.91	7.97	8.05	8.17
Less than primary	.186	.179	.167	.156	.155	.153	.145
Completed primary	.296	.295	.291	.286	.276	.267	.256
Completed lower secondary	.34	.339	.331	.337	.337	.33	.336
Completed upper secondary	.178	.187	.209	.22	.231	.249	.261
Ethnic minority	.086	.09	.093	.085	.09	.094	.099
Number of businesses	20001	20033	19126	17565	17830	17815	17463

All monetary values are reported in January 2018 prices, millions of VND.
1 Data available for 18,042 observations in 2006, 18,110 in 2008, 17,390 in 2010, 16,221 in 2012, 16,224 in 2014, 16,230 in 2016, and 15,838 in 2018.

Before leaving the data section, we briefly discuss concerns about reporting error. As very few informal businesses keep formal accounts, this is a greater concern than among formal firms. de Mel *et al.* (2009) find that microenterprises in Sri Lanka under report

basic information of why they left (i.e., for marriage, for work, etc.) and where they moved to.

revenue by about 30 percent on average. They note this is not necessarily intentional, for example to avoid taxation, but rather due to recall error. They suggest that sales revenue, as compared to implicit revenue from self-consumption or exchanging good and services, is measured with less error. Motivated by this, our main measure of revenue is sales, which account for 95 percent of total revenue among informal businesses, with goods or services exchanged, goods or services consumed by the household, and by-products consumed by the household representing the remaining 5 percent.

Under reporting to avoid taxation is not likely to be a big concern in our context because companies with less than 10 workers are simply not required, given their size, to formally register as an enterprise and operate under the legal framework of the Enterprise Law. 90% of informal businesses have 2 or fewer workers and thus are well away from the legal requirement to formally register as an enterprise. Furthermore, a significant percentage of informal businesses (around 35-37%) report positive taxes and/or fees paid in the annual expense category "Taxes, fees, and other fees considered as taxes."

4 Informal businesses over time and space

In order to provide a broader context for the analysis of informal business entry and exit in Sections 5 and 6, we first document a few facts about the prevalence of informal businesses over time and space. Our data spans 15 years of economic growth, during which GDP per capita nearly doubled and was accompanied by significant structural change (McCaig and Pavcnik, 2013).[26]

Fewer Vietnamese households operate non-farm businesses over time during this period of economic growth. Figure 1 plots the share of households that operate a business and shows

[26] According to estimates in the World Development Indicators database, between 2004 and 2018, Vietnam's GDP per capita PPP adjusted grew from $3,690 to $7,586 in constant 2017 international dollars for an average annual growth rate of 5.3%. For comparison, the business samples included in McKenzie and Paffhausen (2019) range in terms of GDP per capita from about $400 (Malawi) to over $8,000 (Mexico). The data on owner transitions in Maloney (2004) primarily comes from middle-income countries such as Argentina and Mexico.

a noticeable decline from 0.38 in 2004 to about 0.32 in 2018, a 16% reduction.[27,28] This reduction is due to a decline in the share of households operating informal businesses and is only partially offset by an increase in the share of households operating formal businesses. Figure A14 shows that the incidence of operating a formal business is extremely low, about 1% in 2006, and grows to 1.3% by 2018. The drop in the share of households operating a business is consistent with Gollin (2008), which predicts a decline in self-employment businesses as the economy grows. As the share of households operating an informal business has fallen, the informal businesses look marginally more successful, as measured by the share that hire workers, operate as the owner's primary job, and operate as a year-round business (see section A.1).

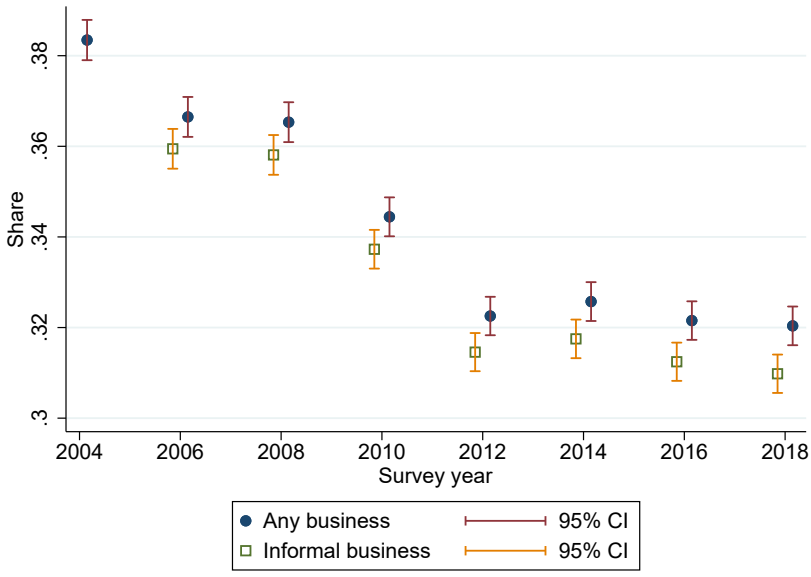


Figure 1: Share of households operating any business and operating an informal business

The share of households operating an informal non-farm business increases with provincial income per capita. Figure 2 shows a strong positive relationship between provincial income per capita in 2006 and the share of households within a province that operate an informal

[27]All figures in this section rely on repeated cross sections and national estimates are obtained using survey sampling weights. We find a very similar pattern without using sampling weights. The decline is occurring in both urban and rural areas (Figure A15), but more so in urban areas, which started off with a greater prevalence of households operating a business.

[28]This reduction is consistent with the results of McCaig and Pavcnik (2015) which reports a reduction in the share of workers in the informal sector between 1999 and 2009 using population census data.

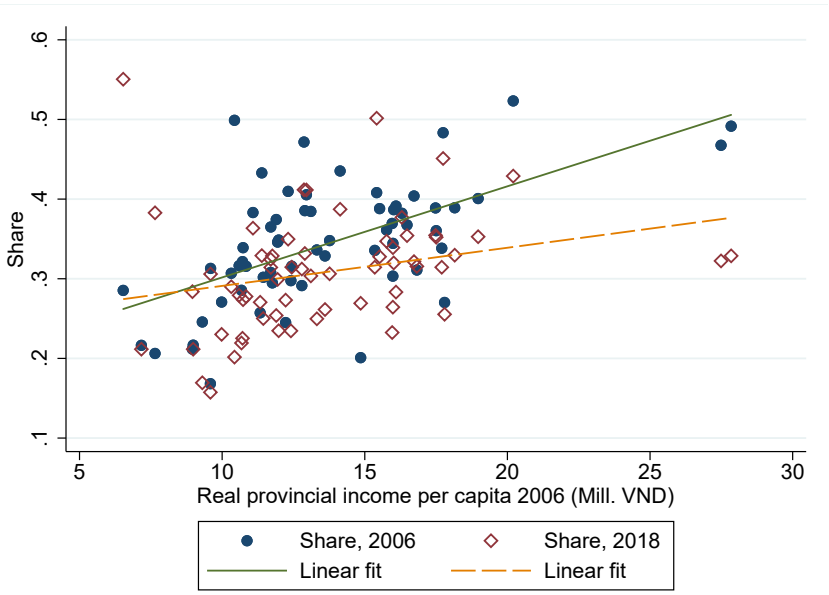


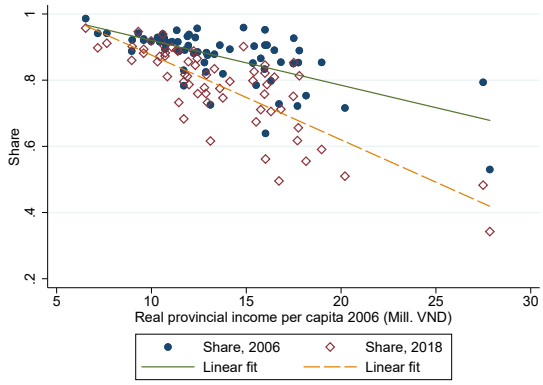
Figure 2: Share of households within a province operating an informal business

business.[29] However, the positive relationship weakens over time. Households in the initially richest provinces have become less likely to operate an informal non-farm business over time. The lines of best fit are statistically different starting at about 14 million VND (see Figure A20). Thus, the aggregate decline in the share of households operating an informal business from Figure 1 is driven, on average, by households in initially richer provinces.

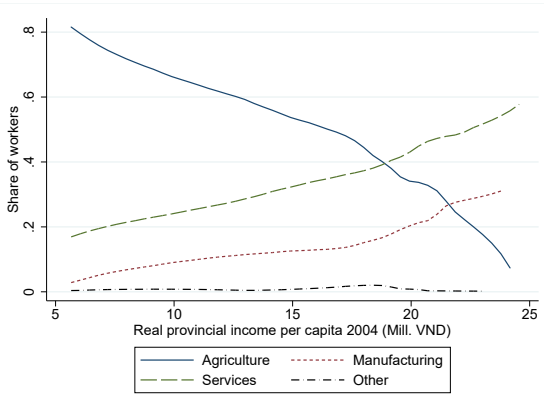
Existing literature documents a negative relationship between the prevalence of self-employment (in agriculture and non-farm businesses) and development across countries (La Porta and Shleifer, 2008) and over time within countries (Gollin, 2008). Our results in Figure 2 seem inconsistent with this pattern. However, when we include both farm and non-farm businesses (as is customary in the cross country literature), we in fact see the same pattern across provinces. In Figure 3a, we find that the prevalence of operating any business, farm or non-farm, declines with development across provinces. The contrast across the two figures reflects the differences in the underlying sectoral employment structure across provinces. Figure 3b shows that the share of employment in agriculture decreases and the share of employment in manufacturing and services increases across provinces with income.

[29]The income per capita estimates are from Benjamin *et al.* (2017). The informal businesses in richer provinces are marginally more successful than in poorer provinces, as measured by the share that hire workers, operate as the owner's primary job, and operate as a year-round business (see Figure A2).

Our data distinguishes between owning an informal farm and a non-farm business. This allows us to focus on the role of non-farm informal businesses in the process of development, taking into account the sectoral structure of employment.



(a) Share of households operating a farm or non-farm business



(b) Structure of employment

Figure 3: Household businesses and sectoral structure of employment

This analysis yields two important takeaways. First, the large variation in the incidence of households operating informal businesses over time and across provinces at different levels of development might be an important consideration for studies that examine the effects of interventions on informal businesses or microenterprises in a specific location. It is useful to interpret the results of those studies keeping in mind the level and sectoral structure of local development. Second, a better understanding of the prevalence of informal non-farm businesses requires more facts about the patterns of entry and exit of informal non-farm businesses over time and space, what informal business owners do before and after they run the business, and how these transitions relate to the level and structure of local development. We turn next to dynamic analysis focusing on entry and exit of informal businesses.

5 Informal business dynamics

We first document several facts about entry and exit rates of informal businesses over time, industries, and locations. In order to better understand the implications of entry and exit for aggregate productivity of the informal sector, we establish several facts about entry and exit

and informal business performance in Section 5.2 and post-entry performance of entrants in Section 5.3. We examine how entry and exit affect the livelihoods of business owners through income in Section 5.4.

5.1 Entry and exit rates are similar and correlated over time, industries, and space

We begin our analysis of the dynamics of informal businesses by examining exit and entry rates. As McKenzie and Paffhausen (2019) note, relatively little is known about the exit of small firms in developing countries, despite the large literature on (formal) firm dynamics in developing countries (Tybout, 2000). Even less is known about entry and overall informal firm dynamics (Ulyssea, 2020). One of the greatest strengths of our data is the existence of multiple large, nationally representative panels of informal businesses that span a period of 15 years, which enable us to simultaneously examine entry and exit of informal businesses. We provide several facts about entry and exit rates over time, within industries, and across space with economic development.

We estimate annual entry and exit rates of informal businesses using each of the six two-survey panels. We define an entrant as a business that did not operate in year t-2, but operated in year t and an exiter as one that did not operate in t+2, but operated in t.[30] These are reported in Figure 4. Several facts emerge.

Exit rates range from about 18% in the 2006-08 panel to about 14% in the 2012-14 panel. In Tables A9 through A11 we report exit rates from existing studies. The exit rates in Figure 4 are slightly higher than the mean exit rates in McKenzie and Paffhausen (2019), but well within their range of reported estimates across surveys and countries as well as for other studies of microenterprise and informal firms in developing countries (see Table A9). These rates are slightly higher than what are typically observed among formal firms in developing countries (Table A10) and firms in developed countries (Table A11).

[30] We calculate the annual exit rate as follows. We calculate the share of firms operating in year t that are still operating in year t+2. We calculate the square root of this to calculate the annual survival rate that is consistent with the two-year survival share. The annual exit rate is then 1 minus the annual survival rate. For the annual entry rate, we first calculate the two-year share of continuing firms and convert this to a one-year survivor share by taking the square root. We then calculate the one year entry rate as 1 minus this rate.

There is substantial simultaneous entry and exit. Entry rates range from about 17.5% to 15.5%, and are relatively similar to the exit rates in a given time period. In Tables A12 through A14 we report entry rates from existing studies. The entry rates in Figure 4 are within the range of reported entry rates for informal and microenterprises in developing countries (Table A12), slightly higher than what is typically observed among non-transition developing countries (Table A13) and higher than commonly observed in developed countries (Table A14).[31]

Importantly, exit and entry rates are similar within a period and decline over time from 17.5-18% in 2006 to around 15% in 2016, as Vietnam experiences economic growth.[32] We observe similar, but slightly lower, rates of exit and entry when we exclude businesses run by individuals joining or leaving the household (Figure A18). We also observe quantitatively similar rates of exit and entry when we include all businesses reported by the household, in case of misclassification of businesses as formal or informal (Figure A19).

While entry and exit rates vary across industries, the differences across industries are not large. Figure A21 plots entry and exit rates for services and manufacturing, followed by entry and exit rates in the most common industries (retail, hotels and restaurants, food and beverages, and land transportation). The rates are relatively similar across industries (with manufacturing experiencing higher entry and exit rates than services). For the most part, industry entry rates tend to be a similar magnitude as industry exit rates in a given time period. This is consistent with models from industrial organization that suggest that entry (and thus) exit rates are both determined by industry-level conditions such as the fixed cost of entry (and thus exit) and the competitiveness of the market (Dunne *et al.*, 1988; Asplund and Nocke, 2006).

[31] Although informal business exit and entry rates are only slightly higher than what is typically observed for formal firms in developing countries, this nonetheless likely means greater employment churning since there is little size variation among informal firms and larger formal firms are less likely to exit than smaller formal firms. Hence, similar rates of entry and exit would be associated with a greater share of employment turnover in the informal sector.

[32] The decline in the exit and entry rates is not simply due to better matching of businesses within a household in the more recent surveys. As we saw in the repeated cross section analysis, the share of households operating a business has fallen over time. Additionally, if we aggregate over all businesses run by a household and examine household-level entry and exit from/to operating any businesses, the rates also decline over time. As expected, the rates are lower with aggregation across businesses. See Figures A16 and A17 in the appendix for households operating any business and for households operating an informal business.

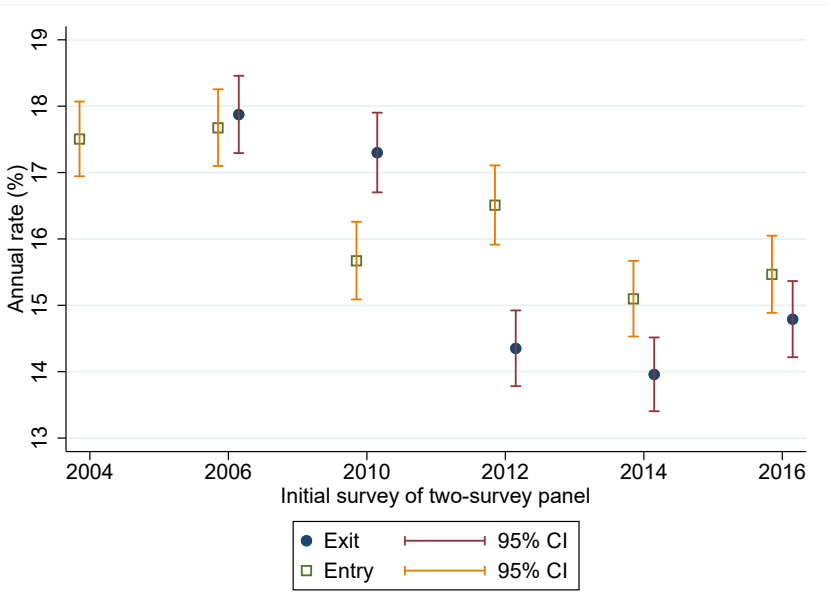


Figure 4: Annual exit and entry rates of informal businesses

The prevalence of exit and entry is negatively correlated with economic development across provinces. Using the 2006-08 panel as an example, in Figure 5 we plot the annual exit and entry rates of informal businesses in a province relative to income per capita in 2006 across provinces. On average, more developed provinces experienced lower rates of entry and exit. The lower exit rates in richer provinces is consistent with the evidence from Mead and Liedholm (1998) that the exit rates of micro and small enterprises (<50 workers) in Botswana, Kenya, Malawi, Swaziland, Zimbabwe, and the Dominican Republic were lower in urban areas than in rural areas, but contrasts higher rates of exit in larger cities in developed countries (Asplund and Nocke, 2006). The pattern is consistent over time and as such we only report it for one panel.[33] We find a very similar pattern when we exclude businesses run by individuals joining or leaving the households (Figure A23) or to major industries, such as retail (available upon request).

Why would economic development be related to entry and exit rates? Higher levels of economic development might be associated with more outside employment opportunities for business owners (such as formal wage work), leading to lower voluntary entry (and thus also) lower exit rates. At the same time, higher economic development is associated with fewer

[33]Figure A22 shows that the 95% confidence intervals for the entry and exit lines of best fit overlap.

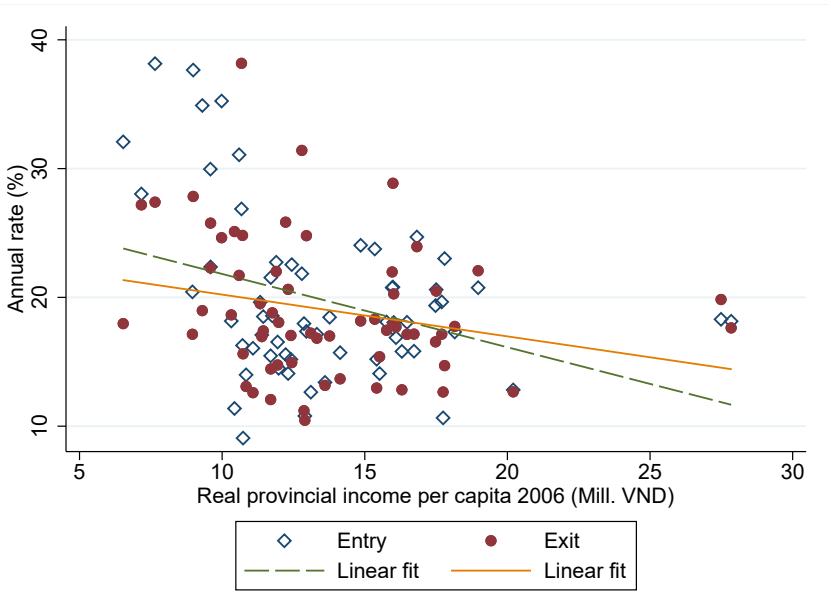


Figure 5: Annual informal entry and exit rates by province, 2006-08 panel

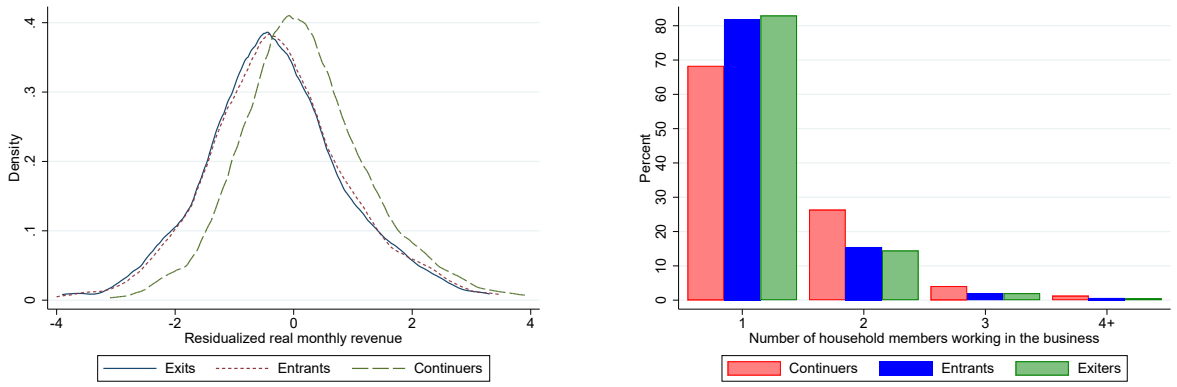
constraints on informal businesses (such as access to credit), thus higher entry and lower exit rates (McKenzie and Paffhausen, 2019). Our findings that entry and exit rates both decline with economic development is consistent with models that suggest a high correlation of the entry and exit rates with industry or market characteristics and the importance of outside employment options for microenterprise owners. Although all else equal, one might expect that exit would be higher in more developed areas with more outside options for a business owner (and hence lead to a positive relationship between economic development and exit (see McKenzie and Paffhausen (2019)), outside opportunities also reduce entry and thus subsequent exit. This highlights the importance of simultaneously examining entry and exit for understanding the dynamics of informal businesses.

5.2 Performance of entrants and exiters overlaps

Entry and exit could contribute to improvements in aggregate productivity of the informal sector if less productive businesses exit and entering businesses are more productive and ultimately grow. In this subsection, we focus on the link between entry and exit and informal business performance.

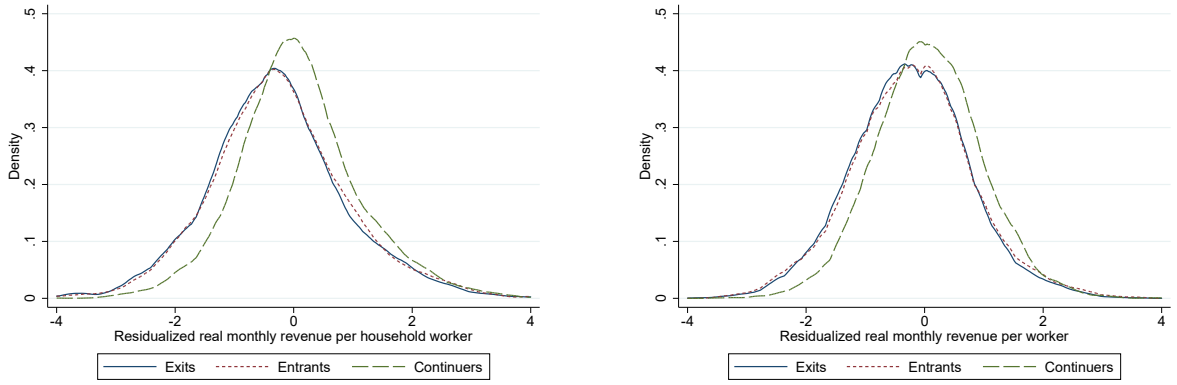
To the extent that exiting informal businesses are less productive than continuing and

entering ones, the informal business dynamics from section 5.1 could contribute to a better allocation of resources and aggregate labor productivity improvements. In Figure 6a, we plot real monthly revenue for entering, exiting, and continuing informal businesses based on pooled 3-survey panels.[34] We plot residualized revenue by removing survey fixed effects. In Figure 6b we also plot the number of household members working in these businesses.



(a) Revenue

(b) Household workers



(c) Revenue per household worker

(d) Revenue per worker

Figure 6: Distribution of revenue and number of household workers among entering, exiting, and continuing informal businesses

Note: We removed survey fixed effects from the revenue distribution by regressing ln monthly revenue on survey fixed effects and then plotting the residuals. In panel (d), revenue per worker is calculated using only the 2004-06-8 panel.

First, market selection is at work: the surviving informal businesses tend to perform better than those that exit. The revenue distribution for the continuing informal businesses is shifted to the right of the distribution of those exiting, although there is substantial

[34]We define an entrant as a business that did not operate in year t-2, but operated in year t and an exiter as one that did not operate in t+2, but operated in t. A continuer operated in all three surveys in the panel.

overlap.[35] The patterns for informal firms are consistent with the role of exit in weeding out inefficient firms in the formal sector (Roberts and Tybout, 1996; Tybout, 2000). A decision to exit could be driven by firm-specific shocks that influence its profitability, such as low productivity, increases in costs of inputs, or increased product market competition. These are common reasons for exit of formal firms and the market selection evidence suggests they also play a role among informal firms.

Importantly, the entering informal businesses tend to perform equally to their exiting counterparts. The two distributions nearly perfectly overlap in Figure 6a and are statistically indistinguishable (see Figure A4 for the same figure with confidence intervals). We find similar results using measures of revenue labor productivity in Figures 6c and 6d. This pattern is at odds with the higher productivity of entrants relative to exiting firms found among formal firms in developing countries (Roberts and Tybout, 1996; Aw *et al.*, 2001), where net entry can account for a sizable share of industry-level productivity improvements. In Appendix section A.3, we plot distributions of estimated total factor productivity based on a residual approach and find similar results.

Many studies of informal firms focus on their employment as a key measure of their size and performance (Hsieh and Klenow, 2014; Haltiwanger, 2015; McKenzie and Paffhausen, 2019). We find similar overlap in the number of household members that entering and exiting informal non-farm businesses employ (Figure 6b). A little over 80 percent of exiting and entering businesses have only 1 household member working in the business as compared to slightly less than 70 percent for continuing businesses. Continuing businesses are about twice as likely to have 2 household members working in the business and very few businesses, continuers, entrants, or exiters, have more then 2 household members working in the business. In Figure A7, we confirm these findings using the total number of workers, both household members and from outside the household, for the 2004-06-08 panel (the only panel for which this information is available). We also find overlap in education and hours worked by the business owner for entrants and exiters (see Appendix section A.2).

We have ruled out several explanations for the large overlap in performance between

[35]Figure A3 excludes businesses run by individuals joining or leaving the household and obtains similar patterns.

entering and exiting informal businesses. The overlap is not primarily driven by creation of informal non-farm business as a coping mechanism to smooth agricultural shocks as in Adhvaryu *et al.* (2019). The overlap does not simply reflect adding and dropping of business activities that are complementary to existing household activities, mismeasurement of businesses within a household, or churning in underlying household composition (see subsection A.2 for details).

Thus, despite market selection through the exit of less efficient informal firms, it is unlikely that point-in-time entry and exit of informal firms contribute much to improvements in aggregate productivity of the informal non-farm sector. We next examine post-entry growth of entrants.

5.3 Modest survival and growth for entrants

The longer-term implications of entering firms on aggregate sector performance depend not only on the initial performance of entrants, but also their post entry growth. Additional analysis of post-entry performance suggests relatively modest performance of entrants two years following entry.

First, almost half of the informal entrants do not survive until the next survey. Figure 7 plots, for each three-survey panel, the share of businesses that enter in the middle survey t (relative to the survey at t-2) and exit by the t+2 survey, spanning a period of 14 years. Around 49% of informal entrants exit within 2 years, about twice the average rate of exit. This large turnover contributes to the high degree of overlap between entrants and exiters in Figure 6a, although substantial overlap persists after the removal of exiting entrants.

Second, the initial revenue of entrants predicts post entry performance, highlighting that market selection is at work. Figure 8 plots the initial monthly revenue distribution as well as the number of household members working in the business for informal entrants at time t for those that survive to t+2 and those that exit by t+2. Consistent with market selection, the initial distribution of surviving entrants is positioned to the right of the distribution of exiting entrants. Surviving entrants also somewhat increase monthly revenue two years after the entry (see Figure A24 for a version with confidence intervals that demonstrates the statistically significant differences in the distributions). However, the growth in revenue

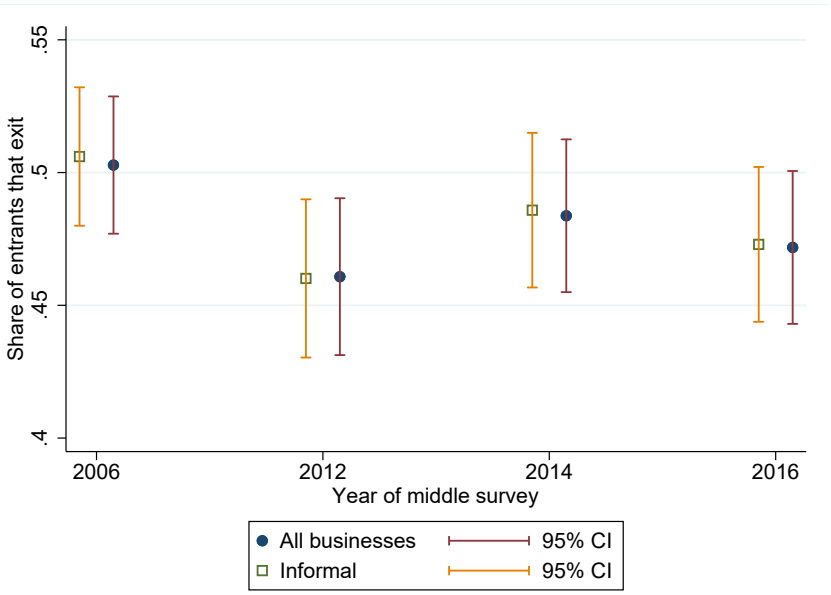
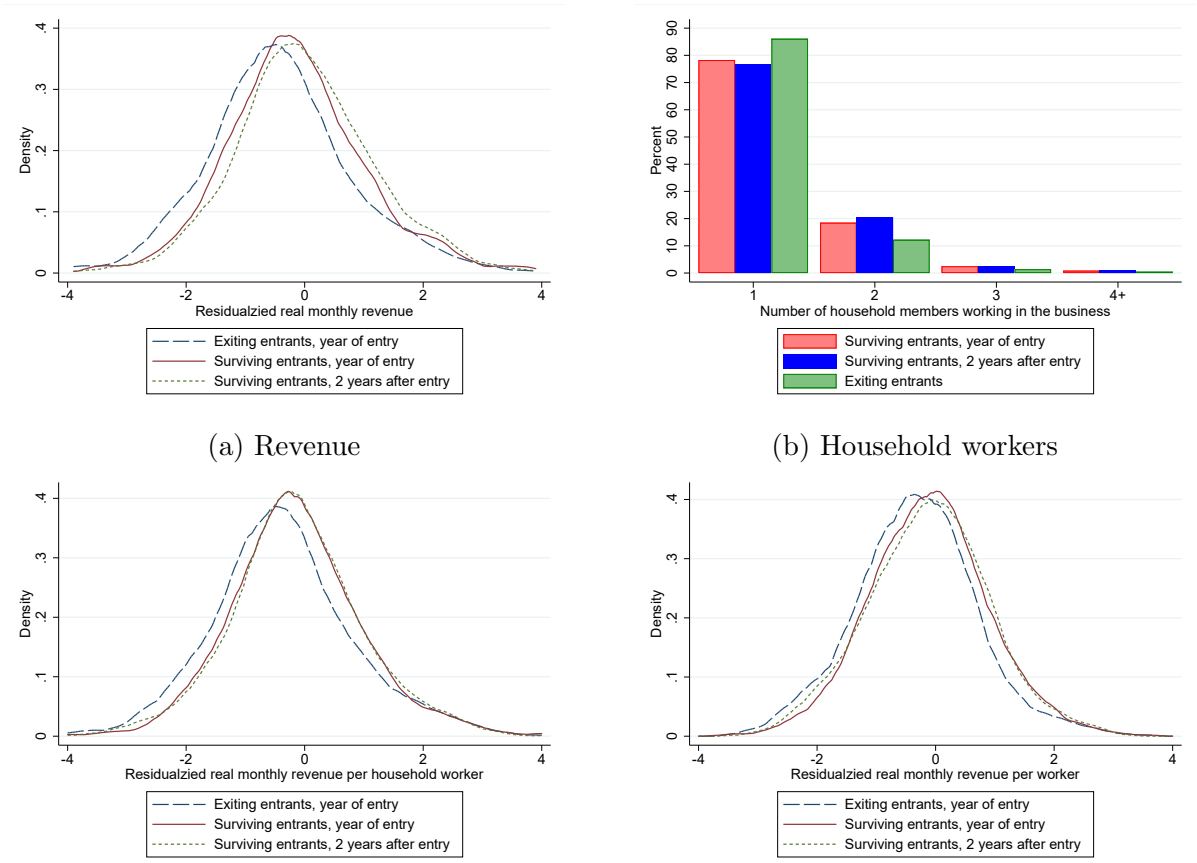


Figure 7: Share of entrants that exit by next survey

among surviving entrants is largely due to growth in the number of workers in the business as Figures 8c and 8d show that the revenue per worker distributions for surviving entrants at entry and two years later overlap to a much greater degree. Thus, it is not an increase in labor productivity that is driving the revenue growth in Figure 8a, it is growth in the number of workers.

High exit rates of entering informal businesses and the survival of initially better performing entrants are consistent with settings where firms learn about their productivity or demand after entry and the lower performing firms exit (Jovanovic, 1982; Foster *et al.*, 2008). Some of this entry and exit might also reflect individual or household specific shocks (such as dealing with a household income shock or an illness). They are also consistent with evidence of exit being the highest among young small firms in studies that examine the link between firm age and survival (Dunne *et al.*, 1989; Haltiwanger *et al.*, 2013), including in developing countries (McKenzie and Paffhausen, 2019). We find evidence of this as well (Panel B of Table A6).

The modest average growth may hide significant growth among a small number of successful surviving entrants if most entrants do not grow. Haltiwanger (2015) reports that most young firms in the U.S. do not grow, but a small fraction experience rapid growth.



(a) Revenue

(b) Household workers

(c) Revenue per household worker

(d) Revenue per worker

Figure 8: Distributions of surviving and exiting informal entrants

Note: We removed panel fixed effects from the revenue distribution by regressing ln monthly revenue on a series of indicators for which 3-survey panel the business belongs to. We plot the residuals. In panel (d), revenue per worker is calculated using only the 2004-06-8 panel.

A similar pattern is found among firms in developing countries (Fafchamps and Woodruff, 2017; McKenzie, 2017). Indeed, when we focus on employment (Figure 8b), we find that the distribution of the number of household members working in surviving entrants two years after the entry was not different from the employment at the time of the entry. Figure A25 shows the same plot using the total number of workers, household and non-household members, for the 2004-06-08 panel and demonstrates marginally higher growth on average. We explore transitions to higher employment using an employment transition matrix. Among surviving entrants, in Table A7 we report the share of businesses based on the number of household members working in the business in their start year and in the subsequent survey. We pool across all 3-survey panels. We find that most businesses remain very small two

years after entry, with only one or two household members working in the business.[36] In Appendix Section A.5, we provide additional evidence that surviving entrants experience modest growth relative to continuing informal firms, in terms of the likelihood of survival and the contribution to the share of firms, revenue and hiring of workers in the informal sector in the two years following the entry.

In the analysis so far, we can only follow entrants for two years after entry. Hsieh and Klenow (2014) and Foster *et al.* (2008) note that it might take time for firms to grow after entry. Although this is an important concern, it is useful to keep in mind that almost 50% of entrants exit within the first two years. We, in part, overcome this concern by additional analysis of performance of continuing firms.

In Appendix Section A.6 we find gradual improvements in the performance of continuing informal businesses with age, relative to 0-2 year old businesses. Businesses that survive 8-12 and 13-17 years post entry are more likely to hire outside workers, have more workers, and higher monthly revenue. Unlike the panel-data analysis above, the observed trends could reflect performance improvements or selection of better firms over time. It is thus important to keep in mind that about 25% of continuing firms survive to age 8-12 and 10% to 13-17 years. Further analysis of continuing firms in Appendix Section A.7 confirms that despite market selection at work, most informal firms to do not contribute much to aggregate employment (beyond the family work). The existing literature has emphasized hiring of paid workers (outside of the family) and formalization as indicators of business expansion (de Mel *et al.*, 2010, 2013). Yet, we find that transitioning to employing paid workers and formalization is rare even for continuing firms.

In summary, although we find some evidence in Sections 5.2 and 5.3 that market selection influences informal business dynamics, the large overlap in performance of entrants and exiters and modest post-entry growth of entrants are consistent with little net effect of entry and exit on aggregate (revenue) productivity or hiring of workers outside the household. Market selection is only one factor influencing entry and exit of informal businesses, with a broader range of owner-specific characteristics and shocks potentially also playing a role

[36] In Table A8 we report the same transitions matrix but focus on the total number of workers, both from within and outside the household. The results are very similar as most businesses do not employ workers from outside the household.

(Maloney, 2004; Hurst and Pugsley, 2011; Astebro *et al.*, 2014). The overlap in performance is consistent with the idea that informal businesses might enter (or exit) the market because their owners do not have alternative means of employment and need to rely on the informal business to generate income (Gollin, 2008; La Porta and Shleifer, 2008, 2014). In these cases, underlying entrepreneurial ability or productivity of the business is only one factor influencing entry and exit decisions. Consistent with Hurst and Pugsley (2011) and Astebro *et al.* (2014), owners list a wide range of exit reasons (Section A.4). The most common is "other activities with more income," accounting for 27% of all exits and which is more common in richer areas, where entry and exit are lower on average (Figure 5). We further explore how entry and exit affect individual income in Section 5.4 and the role of the owner's education and outside employment opportunities that vary with the level and structure of development for entry and exit in Section 6.

5.4 Entry and exit matter for individual income

Entry and exit could still contribute importantly toward the livelihoods of their owners. To the extent that informal business income accounts for a sizable share of household income, entry and exit could expose households to large income fluctuations. Income from informal businesses accounts, on average, for 20% of household income in Vietnam.[37] In this subsection, we examine whether business entry and exit is associated with significant changes in individual income.

To do so, we construct a sample of informal business owners that either (i) operated an informal business in surveys t and t+2, (ii) operated an informal business in survey t, but not t+2 (an exit), or (iii) operated an informal business in survey t+2, but not t (an entrant). We sum profits from the business and any additional wage jobs reported by the owner. We exclude any informal business owners that also report additional self-employment income, such as from household agriculture, due to the difficulty of assigning such income across multiple household members participating in the activity.

[37] Benjamin *et al.* (2017) estimate that formal and informal businesses accounted for 22.4% of household income in the 2008 survey and informal businesses accounted for 91.9% of profits in the cross section of businesses in 2008.

We estimate the following equation using data from each of the 2-survey panels:

$$\Delta Y_{it} = \beta_1 exit_{it} + \beta_2 enter_{it} + \theta_t + \epsilon_{it} \tag{1}$$

where ΔY_{it} is the change in owner i's income during panel t, *exit* is an indicator variable for business exit, *enter* is an indicator variable for business entry, and θ_t are panel fixed effects. Owners that continue to operate an informal business are the reference group. This specification is similar to McKenzie and Paffhausen (2019) except that it also includes informal business entry.

We find that informal business entry and exit indeed significantly contribute to changes in the owner's income, as reported in Table 3. In columns 1 and 2 the dependent variable is the change in owner income, with observations in the top and bottom percentiles windsorized in column 2. Since informal business earnings are measured with potentially large errors (de Mel *et al.*, 2009), we also report a specification based on an indicator for a positive change in household income in column 3. We find that exit is associated with a reduction in an owner's income and entry is associated with an increase in an owner's income. Relative to the overall mean change in owner income of 13, the loss and gain of income associated with closing and starting a business is important. The reduction in income in connection with exit is consistent with results from McKenzie and Paffhausen (2019).[38] In appendix section A.8 we use household income and thus capture all income generating activities in the household, even those attributed to multiple members. We find a similar pattern of results.

Overall, this analysis highlights that even though entry and exit of informal businesses might not contribute significantly to the aggregate share of household income derived from informal businesses, it does have a significant impact on individual and household income. Entry enhances individual and household and income and exit reduces them. Entry and exit of informal businesses are therefore important factors to consider when thinking about income fluctuations for households in low-income countries.

[38]Unfortunately, the VHLSSs do not survey individuals about reasons for shutting down their business to directly examine whether individuals that shut down their business for better economic opportunities observe increases in income. McKenzie and Paffhausen (2019) find that negative exit gets reversed for those the owners that report they are exiting due to better opportunities. However, their data does not enable them to examine entry.

Table 3: Changes in individual income due to informal business entry and exit

	(1) Change	(2) Change	(3) Positive change
Exiter	-27.09	-23.68	-0.18
	(7.26)	(1.38)	(0.02)
Entrant	26.24	23.29	0.18
	(8.39)	(3.03)	(0.01)
Observations	18735	18735	18735

Notes: The dependent variable in columns 1 and 2 is the change in individual income (including wages and profits) and in column 3 is an indicator for a positive change. The sample is all individuals that operated an informal businesses in either or both surveys of each 2-survey panel. We exclude individuals that worked in self-employment in agriculture. Observations in column 2 have been windsorized at the 1st and 99th percentiles. Panel fixed effects are included in each regression. All income are reported in million VND (Jan 2018 prices).

6 What do owners do before and after operating a business?

We further connect entry and exit patterns of informal businesses to the education and economic activities of their owners, who are the key production input in an informal business. Our data allows us to systematically compare education and economic activities of informal business owners that are starting or shutting down a business to the education and economic activities of the overall working-age population. This analysis sheds further light on the high overlap in the performance of entering and exiting informal businesses, the high correlation between entry and exit, and how entry and exit vary across time and space with the level and sectoral structure of economic development.

6.1 Owners have intermediate levels of education

To set the stage, we compare the education of owners of informal businesses and workers economy-wide across economic activities. We focus on an individual's most time consuming job during the past 12 months and distinguish between wage work and self-employment in agriculture and the informal and formal sectors outside of agriculture. Self-employment includes both owners of businesses as well as unpaid family members.

Individuals with relatively low levels of education are self-employed in informal businesses, especially when compared to wage workers in the formal sector (Table 4).[39] With an average of 8.3 years of education, the self-employed in the informal sector are more educated than self-employed and wage workers in agriculture. However, they have substantially lower levels of education than wage workers in the formal sector, who have completed on average 10.5 years of education.

Table 4: Mean highest grade completed by occupation

	All		2006		2018	
Economic activity	Share	Grade	Share	Grade	Share	Grade
Wage worker in agriculture	.039	5.0	.042	4.6	.039	5.5
Wage worker in informal sector	.126	7.9	.112	7.7	.14	8.1
Wage worker in formal sector	.244	10.5	.181	10.4	.316	10.6
Self-employed in agriculture	.397	6.9	.476	6.8	.307	6.9
Self-employed in informal sector	.189	8.3	.184	8.0	.192	8.7
Self-employed in formal sector	.005	10.2	.004	9.3	.006	10.6
All	1	8.1	1	7.7	1	8.5

Columns 1 and 2 are pooled over 2006 through 2018 surveys. Sample restricted to individuals between 15 and 64 years old inclusive.

Even informal business owners that hire workers tend to be less educated than wage workers in the formal sector. In Table 5 we focus specifically on owners of informal businesses and distinguish between owners that hire and those that do not.[40] Owners of informal businesses that hire workers have 9.1 years of education, less than wage workers in the formal sector. Informal business owners that do not hire (which accounts for most of the informal businesses) have similar education levels as wage workers in the informal sector. These patterns persist throughout our 14-year sample. Years of education is clearly an imperfect measure of an individual's ability. However, studies such as de Mel *et al.* (2010) that measure ability based on years of education and more specific tests (such as score on digit span recall test, Raven test, and Cognitive reflection test) reach similar conclusions across these measures.

[39] The mean grade reported is based on highest grade completed up to grade 12.

[40] We include all owners, including those running a business as a job beyond their most time consuming job. In instances in which an owner operates more than one business, we only include the owner once.

Table 5: Mean grade completed by informal business owners

	All	2006	2018
All owners	8.2	7.8	8.5
Owners that hire	9.1	9.0	9.2
Owners that do not hire	8.0	7.7	8.4

Sample restricted to individuals between 15 and 64 years old inclusive.

Our results have several implications. This analysis highlights relatively low levels of education among informal non-farm business owners, especially relative to wage earners in the formal sector. Earlier studies find that microenterprise owners tend to have lower levels of education than owners of businesses that hire workers and comparable education to wage workers (La Porta and Shleifer, 2008; de Mel *et al.*, 2010; Woodruff, 2012). We show that it is important to distinguish comparisons with informal and formal wage work. Even owners of more successful informal businesses that hire workers are on average less educated than formal wage workers. This is consistent with theories that suggest that more educated individuals select into wage work in the formal sector (rather than run an informal business) when formal wage work is available (Lucas, 1978; Gollin, 2008). Other studies have emphasized the role of owner characteristics such as risk aversion, overconfidence, and nonpecuniary tastes to explain entry and persistence of entrepreneurship (Maloney, 2004; Hurst and Pugsley, 2011; Astebro *et al.*, 2014), which our data does not allow us to examine. As the literature review in Section 2 suggests owners could also differ in their exposure to market imperfections, including access to credit. However, the large differences in education might affect individuals' ability to transition in and out of owning an informal business into other activities and thus influence the patterns of entry and exit. We next more directly focus on these transitions and the education of individuals that transition in and out of informal businesses.

6.2 Economic activities of owners prior to starting and after shutting down a business overlap

We systematically examine the economic activities of business owners prior to starting or after closing down an informal business. We compare owner transitions to the rates of transitions between activities among the working age (15 to 64) population. Table 6 presents the results.[41] The activities in which the working age population is engaged are listed in row headings. The share of informal owners entering from and exiting to each activity is reported in columns 1 and 2, respectively. The average grade of a person starting a business after each activity is reported in column 3 and the mean grade of the informal business owner shutting down the business and transitions to each activity is reported in column 4. Several interesting patterns emerge.

Table 6: Before and after operating a business

	Informal mangers				All	
	Before entering (share)	After exiting (share)	Before entering (mean grade)	After exiting (mean grade)	First survey (share)	Second survey (share)
Wage worker in agriculture	.013	.009	5.9	6.1	.037	.036
Wage worker in informal sector	.152	.139	8.2	8.1	.083	.089
Wage worker in formal sector	.103	.098	10.1	9.9	.139	.151
Self-employed in agriculture	.375	.338	6.7	6.9	.352	.336
Self-employed in informal sector	.158	.153	8.2	8.3	.137	.14
Self-employed in formal sector	.004	.005	9.5	9.8	.003	.003
Not in the workforce	.105	.109	8.4	7.8	.192	.135
Not in the household	.09	.149			.058	.11
All	1	1	7.8	7.8	1	1

Sample restricted to individuals between 15 and 64 years old inclusive.

There is a large overlap in economic activities and in levels of education of most entering and exiting informal business owners. In particular, the share of individuals that start or shut down an informal business from a particular activity is very similar, with activities that account for a high share of activities of business owners before the entry also accounting for a

[41] The analysis for the working age population proceeds as follows. We look at all working age individuals in the survey in year t and then track what activity they are doing in year t+2 and report this as second survey. Similarly, we look at all working age individuals in survey t+2 and track what activity they were doing in survey t and report this as first survey.

high share of activities to which the business owners return after closing down the business. Along the same lines, education levels of business owners leaving/returning to a category before and after running a business are similar to each other, as well as similar to the general education levels for that category in Table 4.[42] This overlap in economic activities and levels of education of most entering and exiting informal business owners might contribute to the overlap in performance of entering and exiting informal businesses noted in Subsection 5.2 and Figure 6.

Importantly, the last two rows of the table show the importance of transitions in and out of the household and in and out of the workforce. Jointly, these two categories account for 20-26% of owner transitions. Individuals joining households account for 10% of entry of informal businesses, while individuals departing households account for 16% of informal business exit. These might not reflect true entry or exit as we cannot track the owner before joining or after leaving the household. Such changes in household composition introduce a common measurement issue that we have dealt with throughout the paper and others working on this topic must address as well. Transitions in and out of the workforce, 10-11% of owner transitions, also matter and illustrate the role of individual characteristics or household-specific shocks affecting entry and exit of informal household businesses. Some of these movements are consistent with a number of these businesses being run alongside household responsibilities such as housework (the reason for not working given by around 50% of non-working individuals).[43] Other household or individual specific shocks play a role in business entry and exit. The second most common reason for being out of the workforce is small/at school before entry (16%) and too old/retired, particularly after exiting (25% of business owners that exit to out of the workforce report being too old/retired). Disabled/ill is another notable category, especially after exiting.

[42]The exception is education of formal wage workers that start a business being noticeably higher and education of those moving from the formal sector self-employment being lower than the average education in the formal wage and self-employment sector. This is consistent with more marginal people moving between the two sectors as in McCaig and Pavcnik (2015).

[43]The 2008, 2014, 2016, and 2018 surveys ask respondents who answered that they did not work during the past year to indicate the main reason for not working. We use this data restricted to informal business owners age 15 to 64 prior to opening a business or after closing a business.

6.3 Transitions in and out of informal businesses reflect the structure of employment

The transitions to starting or exiting an informal business in large part reflect the underlying structure of employment/participation of the working age population across economic activities. The most common transition to starting or exiting an informal business in Table 6 is self-employment in agriculture, accounting for 37% of entrants and 34% of exit. Note that education levels of owners in agriculture are the lowest among all groups in Table 4. Transitions from and to wage work or self-employment in the informal sector are also common, jointly accounting for almost 30% of transitions. Fewer individuals are transitioning to and from wage work in the formal sector. About 10% of owners start an informal business after being employed as a wage worker in the formal sector (and 9.7% leave for that sector). Transitions from and to wage work in agriculture and self-employment in the formal sector are very rare.

The right panel of Table 6 confirms the transitions to and from operating an informal business in columns 1 and 2 are relatively similar to the overall distribution of workers across these activities in the economy. The panel reports the share of participation in each activity among the overall working age population. The first such column is based on members of panel households in the end survey of the two-survey panels and reports what they were doing in the start survey (to be compared to activities before business entry in column 1). The second such column is based on members of panel households in the start survey and reports what they were doing in the end survey (to be compared to activities after business exit in column 2). For example, among the overall working age population, 33 to 34% are self-employed in agriculture, which is very close to the rate of transitions into self-employed agriculture to and from operating an informal business.[44]

Nonetheless, there are some education-related deviations from these economy-wide patterns in Table 6. Most notably, relative to the working-age population, informal business owners are more likely to work in informal wage work before or after running a business

[44]The larger share of not in the workforce in the working age population, especially in the first survey, is largely driven by individuals age 15 to 20, likely due to still being in school. If we restrict the sample to individuals ages 20 to 64, the share out of the workforce becomes more similar to transitioning informal business owners.

and they are less likely to work in formal wage work. These deviations might be related to differences in education levels across these categories. Education levels of wage workers and self-employed in the informal sector are comparable, particularly for owners that do not hire (see Tables 4 and 5), as are the average levels of education of informal business owners that enter and exit from these two activities (see columns 3 and 4 of Table 6), making these activities more substitutable for each other. On the other hand, informal business owners are on average less educated than wage workers in the formal sector (see Table 4) and individuals that start or shut down an informal business from and to formal wage work tend to have on average two more years of education than than informal business owners that start or shut down an informal business. More educated individuals with formal wage jobs potentially have better employment opportunities than running an informal business, consistent with (Gollin, 2008).[45]

6.4 Structure of employment matters for transitions over time and space

The importance of the underlying structure of employment for transitions in and out of owning an informal business is further confirmed by examining what informal business owners do before and after operating a business at the beginning and end of our sample, a period during which GDP per capita nearly doubled and significant structural change occurred (see Section 4 for details). Recall that entry and exit rates fell over this period (Figure 4). Appendix Table A5 repeats the analysis of Table 6 but reports the transitions separately for the first two-survey panel, 2006-08, and the last two-survey panel, 2016-18. In general, as the underlying structure of employment changes away from agricultural self-employment toward wage work in the formal sector, so has the nature of transitions for owners of informal businesses. Transitions to and from self-employment in agriculture have become less common

[45]In addition, relative to the working-age population, informal business owners starting a business are less likely to be out of the workforce and somewhat more likely to be joining or leaving a household than the overall working age population. This highlights that these transitions, which tend to be associated with individual or household specific-shocks, need to be considered in studying of entry and exit of informal businesses. Individuals that have a business are more likely not to be part of the household in the previous period than general working-age population. Individual that have a business are more likely to leave the household than the general population.

for informal business owners over time, while transitions to wage work in the formal sector have increased.

Finally, the entry and exit transitions are also influenced by the level of local economic development. In Table 7 we separate the results from Table 6 by provincial income, focusing on the 15 richest and 15 poorest provinces. There are large differences in the pre-entry and post-exit economic activities of business owners across the two sets of provinces. Transitions from and to self-employment in agriculture account for 67 and 59% of transitions in the poorest provinces, while only for 23 and 17% of transitions in the richest provinces. In the richest provinces, it is more common for informal business owners to be transitioning from and to informal wage work and informal self-employment than in the poorer provinces. Likewise, the richest provinces also have a higher share of transitions to and from wage work in the formal sector than in the poorer provinces. This evidence is consistent with the differences in employment opportunities present in these two sets of provinces and consistent with lower entry and exit rates of informal businesses in richer provinces noted in Figure 5 in Subsection 5.1.

Overall, this analysis suggests that transitions mostly reflect the underlying structure of employment in the economy, with education gaps also playing a role. Net entry into informal non-farm businesses from self-employment in agriculture is consistently positive, both across time and locations. Transitioning to and from self-employment in agriculture falls with economic development both over time and across provinces. This underscores the role of net entry into informal non-farm businesses from agriculture in the process of structural change, particularly in poorer provinces where employment opportunities in the formal sector are more limited than in richer provinces. This suggest that informal businesses may serve as a stepping stone out of agriculture. Overall, this evidence is consistent with outside employment opportunities of individuals affecting exit and entry of informal businesses (Gollin, 2008; La Porta and Shleifer, 2008). This highlights the importance of examining the dynamics of informal business entry and exit in the economy-wide context.

Table 7: Before and after operating a business, by initial provincial income per capita

		Informal mangers				All	
		Before entering (share)	After exiting (share)	Before entering (mean grade)	After exiting (mean grade)	First survey (share)	Second survey (share)
15 richest provinces	Wage worker in agriculture	.026	.025	5.1	4.9	.049	.046
	Wage worker in informal sector	.165	.163	7.9	7.4	.092	.096
	Wage worker in formal sector	.141	.112	10.3	9.7	.192	.207
	Self-employed in agriculture	.231	.171	6.6	7.2	.211	.203
	Self-employed in informal sector	.185	.183	7.7	7.8	.169	.171
	Self-employed in formal sector	.006	.004	8.4	10.0	.004	.004
	Not in the workforce	.155	.167	8.4	8.3	.221	.168
	Not in the household	.09	.174			.061	.104
	All	1	1	7.9	7.9	1	1
15 poorest provinces	Wage worker in agriculture	.011	.011	8.5	5.5	.015	.017
	Wage worker in informal sector	.057	.059	8.9	9.0	.043	.053
	Wage worker in formal sector	.043	.079	10.6	10.6	.092	.1
	Self-employed in agriculture	.684	.599	5.4	6.0	.586	.555
	Self-employed in informal sector	.112	.104	9.0	9.2	.075	.079
	Self-employed in formal sector	0	.005		10.6	.001	.002
	Not in the workforce	.042	.058	9.4	9.5	.132	.084
	Not in the household	.051	.085			.055	.11
	All	1	1	6.5	7.2	1	1

Sample restricted to individuals between 15 and 64 years old inclusive.

7 Conclusion

Our study highlights several advantages of studying informal business entry and exit with nationally representative panel data that matches owners and businesses in a fast-growing, low-income country, spanning a period of 15 years. We can follow about 48,000 informal firms over a period of up to four years (most studies cannot), including the activities of their owners before and after operating an informal business. The analysis improves our understanding of the economics of informal firm entry and exit in several dimensions.

For example, it highlights the importance of simultaneously examining entry and exit for understanding the dynamics of informal businesses. Entry and exit rates are similar (14-18%) and highly correlated at a point in time. Aggregate entry and exit decline with economic development over time, while previous work found a positive relationship between

economic development and exit across countries (McKenzie and Paffhausen, 2019). Despite a greater prevalence of non-farm informal businesses in richer provinces, entry and exit rates are lower in these provinces relative to poorer provinces. Although everything else equal, one might expect that exit would be higher in more developed areas with better outside options for a business owner (and hence lead to a positive relationship between economic development and exit), better outside opportunities are also associated with reduced entry.

Second, despite a significant amount of churning, net entry and market selection has little effect on aggregate productivity and hiring of workers outside the family in informal firms. Although less productive informal businesses are more likely to exit, the productivity of new entrants perfectly overlaps that of exiting businesses. This is in contrast to what has been shown for formal firms in developing countries (Aw *et al.*, 2001; Roberts and Tybout, 1996) and to our knowledge this has not been studied for informal firms. Moreover, almost half of the entrants exit within two years and the surviving entrants demonstrate modest growth in revenue and employment, not contributing much to improvements in aggregate productivity in the sector nor hiring of workers outside the family. This is consistent with Hsieh and Klenow (2014) who follow cohorts of firms over time, but don't observe gross exit and entry. Despite churning mattering little for aggregate performance, it matters for individual households: business exit and entry are associated with large decreases and increases in household income, respectively. Moreover, the entry and exit might be closely tied to the owner's characteristics (Maloney, 2004; Hurst and Pugsley, 2011; Astebro *et al.*, 2014), alternative employment opportunities (La Porta and Shleifer, 2008; Gollin, 2008), or individual or household-specific shocks (such as age of the owner, illness in the family).

Third, the large overlap in revenue of entering and exiting informal businesses and the high correlation between entry and exit rates are related to the education and outside employment opportunities of business owners. The individuals that start or shut down a business have very similar levels of education. They are less educated on average than wage workers in the formal sector, which is consistent with Lucas (1978); Gollin (2008), but they are better educated than workers in agriculture.

Fourth, our analysis highlights the importance of examining economic activities of owners before and after operating a business with nationally representative data, which is rare. We

show that the share of individuals that start or shut down a business from a particular economic activity reflects the underlying structure of economic activities of the working age population, with education also playing a role. The most common transition into non-farm businesses is to and from self-employment in agriculture within a location. The likelihood of this transition declines with economic development (both across provinces and over time), highlighting the role of net entry into informal non-farm businesses in structural change out of agriculture. This suggest that informal businesses may serve as a stepping stone out of agriculture and complements existing work focused on informal businesses in major cities or urban areas (de Mel *et al.*, 2010; Maloney, 2004). Our analysis suggests that it is important to keep in mind labor market linkages across agriculture and non-farm informal businesses. As the formal sector has grown, we see more transitions to and from wage work in the formal sector.

Our analysis focuses on informal businesses and shows that their dynamics are influenced by the underlying local economic structure. This demonstrates the importance of studying more broadly how the informal sector interacts with both agriculture and the formal sector. Our current analysis suggests that informal businesses in Vietnam do not often formalize, but individuals–particularly younger ones– do transition from employment in the informal to the formal sector (McCaig and Pavcnik, 2015, 2018). Recent literature from other countries demonstrates important interactions between the informal and formal sectors (see, for example, evidence from Bangladesh (Gutierrez *et al.*, 2019), Brazil (Ulyssea, 2018), India (Allen *et al.*, 2018) and Mexico (Koelle, 2019)). This highlights the importance of examining both the informal and formal sector to understand the sources of economy-wide growth and the allocation of resources between the informal and formal sectors.

References

Adhvaryu, A., Kala, N. and Nyshadham, A. (2019) Booms, busts, and household enterprise: Evidence from coffee farmers in Tanzania, *The World Bank Economic Review*.

Allen, J., Nataraj, S. and Schipper, T. C. (2018) Strict duality and overlapping productivity distributions between formal and informal firms, *Journal of Development Economics*, **135**, 534–554.

Asplund, M. and Nocke, V. (2006) Firm turnover in imperfectly competitive markets, *The Review of Economic Studies*, **73**, 295–327.

Astebro, T., Herz, H., Nanda, R. and Weber, R. A. (2014) Seeking the roots of entrepreneurship: Insights from behavioral economics, *Journal of Economic Perspectives*, **28**, 49–70.

Atkin, D. and Khandelwal, A. K. (2020) How distortions alter the impacts of international trade in developing countries, *Annual Review of Economics*, **12**, 213–238.

Atkin, D., Khandelwal, A. K. and Osman, A. (2017) Exporting and firm performance: Evidence from a randomized experiment, *The Quarterly Journal of Economics*, **132**, 551–615.

Aw, B. Y., Chen, X. and Roberts, M. J. (2001) Firm-level evidence on productivity differentials and turnover in Taiwanese manufacturing, *Journal of Development Economics*, **66**, 51–86.

Banerjee, A., Karlan, D. and Zinman, J. (2015) Six randomized evaluations of microcredit: Introduction and further steps, *American Economic Journal: Applied Economics*, **7**, 1–21.

Banerjee, A. V. (2013) Microcredit under the microscope: What have we learned in the past two decades, and what do we need to know?", *Annual Review of Economics*, **5**, 487–519.

Benjamin, D., Brandt, L. and McCaig, B. (2017) Growth with equity: Income inequality in Vietnam, 2002–14, *The Journal of Economic Inequality*, **15**, 25–46.

Bergquist, L. F. and Dinerstein, M. (2020) Competition and entry in agricultural markets: Experimental evidence from Kenya, *American Economic Review*, **110**, 3705–47.

Brambilla, I., Porto, G. and Tarozzi, A. (2012) Adjusting to trade policy: Evidence from U.S. antidumping duties on Vietnamese catfish, *The Review of Economics and Statistics*, **94**, 304–319.

Bruhn, M. and McKenzie, D. (2014) Entry regulation and the formalization of microenterprises in developing countries, *The World Bank Research Observer*, **29**, 186–201.

de Mel, S., McKenzie, D. and Woodruff, C. (2010) Who are the microenterprise owners? Evidence from Sri Lanka on Tokman versus De Soto, in *International Differences in Entrepreneurship* (Eds.) J. Lerner and A. Schoar, University of Chicago Press, pp. 63–87.

de Mel, S., McKenzie, D. and Woodruff, C. (2013) The demand for, and consequences of, formalization among informal firms in Sri Lanka, *American Economic Journal: Applied Economics*, **5**, 122–50.

de Mel, S., McKenzie, D. and Woodruff, C. (2019) Labor drops: Experimental evidence on the return to additional labor in microenterprises, *American Economic Journal: Applied Economics*, **11**.

de Mel, S., McKenzie, D. J. and Woodruff, C. (2009) Measuring microenterprise profits: Must we ask how the sausage is made?, *Journal of Development Economics*, **88**, 19–31.

Diao, X., Kweka, J. and McMillan, M. (2018) Small firms, structural change and labor productivity growth in Africa: Evidence from Tanzania, *World Development*, **105**, 400–415.

Dunne, T., Roberts, M. J. and Samuelson, L. (1988) Patterns of firm entry and exit in us manufacturing industries, *The RAND Journal of Economics*, pp. 495–515.

Dunne, T., Roberts, M. J. and Samuelson, L. (1989) The growth and failure of US manufacturing plants, *The Quarterly Journal of Economics*, **104**, 671–698.

Fafchamps, M. and Woodruff, C. (2017) Identifying gazelles: Expert panels vs. surveys as a means to identify firms with rapid growth potential, *The World Bank Economic Review*, **31**, 670–686.

Foster, L., Haltiwanger, J. and Syverson, C. (2008) Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?, *American Economic Review*, **98**, 394–425.

Gollin, D. (2002) Getting income shares right, *Journal of Political Economy*, **110**, 458–474.

Gollin, D. (2008) Nobody's business but my own: Self--employment and small enterprise in economic development, *Journal of Monetary Economics*, **55**, 219–233.

Gutierrez, I. A., Kumar, K. B., Mahmud, M., Munshi, F. and Nataraj, S. (2019) Transitions between informal and formal employment: results from a worker survey in Bangladesh, *IZA Journal of Development and Migration*, **9**, 1–27.

Haltiwanger, J. (2015) Job creation, job destruction, and productivity growth: The role of young businesses, *Annuel Review of Economics*, **7**, 341–358.

Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2013) Who creates jobs? Small versus large versus young, *Review of Economics and Statistics*, **95**, 347–361.

Hopenhayn, H. A. (1992) Entry, exit, and firm dynamics in long run equilibrium, *Econometrica*, **60**, 1127–1150.

Hsieh, C.-T. and Klenow, P. J. (2014) The life cycle of plants in India and Mexico, *The Quarterly Journal of Economics*, **129**, 1035–1084.

Hsieh, C.-T. and Olken, B. A. (2014) The missing "missing middle", *Journal of Economic Perspectives*, **28**, 89–108.

Hurst, E. and Pugsley, B. W. (2011) What do small businesses do?, *Brookings Papers on Economic Activity*, **2011**, 73–118.

Jayachandran, S. (2020) Microentrepreneurship in developing countries, Working Paper 26661, National Bureau of Economic Research.

Jovanovic, B. (1982) Selection and the evolution of industry, *Econometrica*, **50**, 649–670.

Kaboski, J. P. and Townsend, R. M. (2011) A structural evaluation of a large-scale quasi-experimental microfinance initiative, *Econometrica*, **79**, 1357–1406.

Kaboski, J. P. and Townsend, R. M. (2012) The impacts of credit on village economies, *American Economic Journal: Applied Economics*, **4**, 98–133.

Koelle, M. (2019) Microenterprises and the lure of wage work: Theory and evidence from mexican export manufacturing, Working Paper 2019-08, CSAE.

La Porta, R. and Shleifer, A. (2008) The unofficial economy and economic development, *Brookings Papers on Economic Activity*.

La Porta, R. and Shleifer, A. (2014) Informality and development, *Journal of Economic Perspectives*, **28**, 109–26.

Levy, S. (2008) *Good intentions, bad outcomes: Social policy, informality, and economic growth in Mexico*, Brookings Institution Press.

Li, Y. and Rama, M. (2015) Firm dynamics, productivity growth, and job creation in developing countries: The role of micro-and small enterprises, *The World Bank Research Observer*, **30**, 3–38.

Lucas, R. E. (1978) On the size distribution of business firms, *The Bell Journal of Economics*, **9**, 508–523.

Malesky, E. and Taussig, M. (2009) Out of the gray: The impact of provincial institutions on business formalization in Vietnam, *Journal of East Asian Studies*, **9**, 249–290.

Maloney, W. F. (2004) Informality revisited, *World Development*, **32**, 1159–1178.

McCaig, B. and Pavcnik, N. (2013) Moving out of agriculture: Structural change in Vietnam, Working Paper 19616, National Bureau of Economic Research.

McCaig, B. and Pavcnik, N. (2015) Informal employment in a growing and globalizing low-income country, *American Economic Review*, **105**, 545–50.

McCaig, B. and Pavcnik, N. (2018) Export markets and labor allocation in a low-income country, *American Economic Review*, **108**, 1899–1941.

McCaig, B., Pavcnik, N. and Wong, W. F. (2021) FDI inflows and domestic firms: Adjustments to new export opportunities.

McCasland, J. and Hardy, M. (2020) Are small firms labor constrained? Experimental evidence from Ghana.

McKenzie, D. (2017) Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition, *American Economic Review*, **107**, 2278–2307.

McKenzie, D. and Paffhausen, A. L. (2019) Small firm death in developing countries, *The Review of Economics and Statistics*, **101**, 645–657.

McKenzie, D. and Woodruff, C. (2014) What are we learning from business training and entrepreneurship evaluations around the developing world?, *The World Bank Research Observer*, **29**, 48–82.

Mead, D. C. and Liedholm, C. (1998) The dynamics of micro and small enterprises in developing countries, *World Development*, **26**, 61–74.

Mondragón-Vélez, C. and Peña, X. (2010) Business ownership and self-employment in developing economies: The Colombian case, in *International differences in entrepreneurship*, University of Chicago Press, pp. 89–127.

Nataraj, S. (2011) The impact of trade liberalization on productivity: Evidence from India's formal and informal manufacturing sectors, *Journal of International Economics*, **85**, 292–301.

Pages, C. (2012) What are the most important research questions on the economics of growth and labor markets in low--income countries?, First GLM—LIC workshop: Setting the research agenda.

Paulson, A. L. and Townsend, R. (2004) Entrepreneurship and financial constraints, *Journal of Corporate Finance*, **10**, 229–262.

Paulson, A. L., Townsend, R. and Karaivanov, A. (2006) Distinguishing limited liability from moral hazard in a model of entrepreneurship, *Journal of Political Economy*, **114**, 100–144.

Roberts, M. J. and Tybout, J. R. (1996) *Industrial evolution in developing countries : Micro patterns of turnover, productivity, and market structure (English)*, The World Bank, Washington, D.C.

Samphantharak, K. and Townsend, R. M. (2012) Measuring the return on household enterprise: What matters most for whom?, *Journal of Development Economics*, **98**, 707–720.

Startz, M. (2016) The value of face-to-face: Search and contracting problems in Nigerian trade, *Available at SSRN 3096685*.

Taussig, M. and Hang, P. T. T. (2004) Private enterprise formality and the role of local government: Making markets work better for the poor, *Asian Development Bank Discussion Paper*.

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

Ulyssea, G. (2018) Firms, informality, and development: Theory and evidence from Brazil, *American Economic Review*, **108**, 2015–47.

Ulyssea, G. (2020) Informality: Causes and consequences for development, *Annual Review of Economics*, **12**.

Vijverberg, M. W. and Haughton, J. (2002) Household enterprises in vietnam: Survival, growth, and living standards, Policy Research Working Paper 2773, The World Bank.

Woodruff, C. (2012) The other half: What we do(n't) know about self-employment in LICs?, First GLM—LIC workshop: Setting the research agenda.

Online Appendix (Not for Publication)

A Supplementary Analysis

In this appendix, we provide supplementary analysis that either extends or confirms the results presented in the main text.

A.1 Trends over time and variation over space

In this subsection, we extend the results from section 4. We do so by discussing additional evidence on trends over time and variation across space in informal business performance.

As the share of households operating an informal business has fallen, the informal businesses look marginally more successful, as measured by the share that hire workers, operates as the owner's primary job, and operates as a year-round business. Table 2 displays the share of informal businesses that hire workers, operates as the owner's primary job, and operates as a year-round business. Hiring of workers is measured with reports of positive labor expenses, while year-around operation is measured as the share of businesses that report operating for all 12 months during the previous year.[46] It shows increases over time in these three dimensions of business performance.[47] The share of informal businesses operated as the owner's primary job has grown from 0.76 in 2006 to 0.84 in 2018, while the share of year-round informal businesses increased from 0.59 to 0.73. Although very few informal businesses hire paid workers (only 9% in 2006), the share has been increasing over time to 12% in 2018. The existing literature has emphasized hiring of paid workers (outside of the family) as a particularly important milestone of business performance and expansion (de Mel *et al.*, 2010, 2013). The low share at the end of the sample highlights how rare it is for an informal business to hire paid workers (outside of the family members).

The most common industries in which informal businesses operate have not changed much over time. Figure A1 reports the share of informal businesses operating in the five

[46] The business module of the 2006 and 2008 surveys asked directly about the number of paid workers and positive labor expenses. Across these two surveys, 96% of informal businesses give consistent answers across these two questions.

[47] We do not include standard errors in the table given the already large size of the table, but the change between the start and end surveys are statistically significant at the 5% level.

most common industries. Retail remains the most common industry by far, although its share fell from 39 to 35%. In total, the share of informal businesses in the five most common industries fell from 75 to 67% between 2006 and 2018. In section 5, we return to these industries to compare their entry and exit rates.

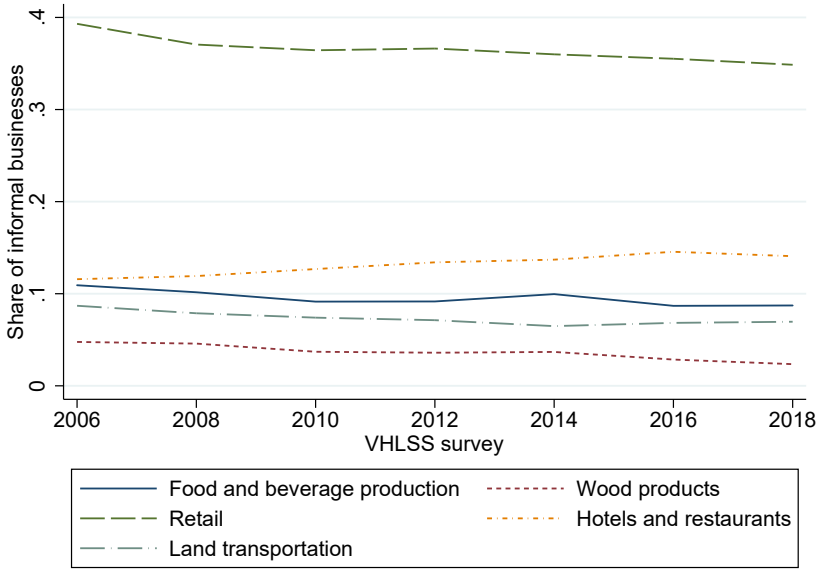
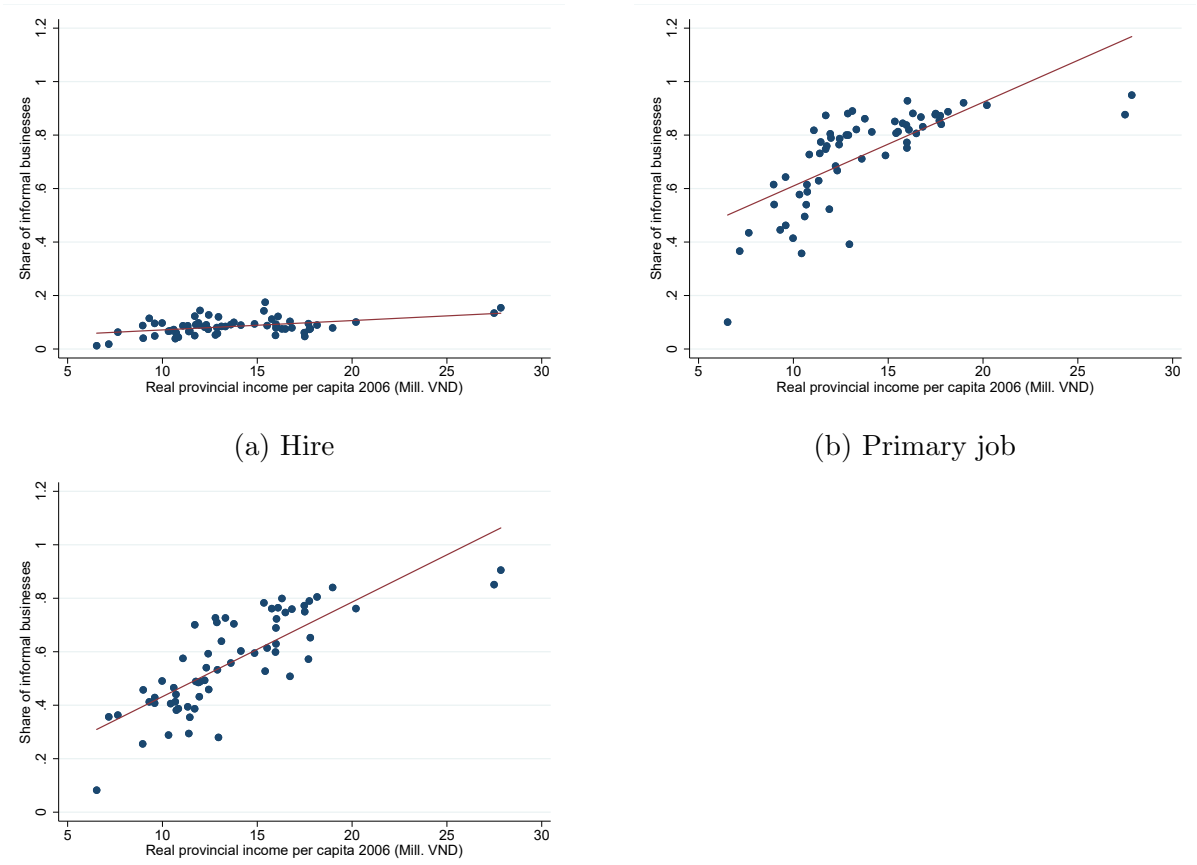


Figure A1: Most common industries for informal businesses

Provinces within Vietnam range in their level of income and differ in their underlying economic structure. This provincial variation enables us to compare informal business characteristics across space, ranging from GDP per capita PPP 1796 USD in 2017 dollars to 12,053 USD over a period of 15 years.[48]

Figure A2 displays that the share of informal businesses that hire outside labor, that operate as the owner's primary job, and that operate year-round is higher in provinces with higher income. Hence, not only the incidence of operating an informal business, but also the characteristics of these businesses vary with the level and structure of development across

[48]These estimates are derived from estimates of GDP by province in 2010 from statistical yearbooks. The weighted sum of provincial estimates of GDP per capita exceeds the national estimate. As such, we adjust the provincial estimates of GDP per capita downward such that they sum to the national estimate of GDP per capita as reported by the General Statistics Office and reproduced in the World Development Indicators (WDI) database. We then convert the estimates to PPP adjusted 2017 constant USD using an adjustment factor derived from the WDI database. The range is from Vietnam's poorest province, Ha Giang, to Vietnam's second richest province, Binh Duong. We exclude Ba Ria-Vung Tau due to its significantly higher level of GDP per capita from oil deposits.



(a) Hire

(b) Primary job

(c) Operate 12 months

Figure A2: Share of informal businesses that hire, are the primary job, and operated for the last 12 months, by provincial income

Note: The sample is all informal businesses in the 2006 survey.

provinces.

A.2 Robustness and extension of revenue overlap among entrants and exiters

In this subsection, we describe robustness checks and extensions to the evidence provided in subsection 5.2 that the performance of entrants and exiters overlaps.

We have examined (and ruled out) several explanations for the large overlap in performance between entering and exiting informal businesses. The creation of informal non-farm business as a coping mechanism to smooth agricultural shocks as in Adhvaryu *et al.* (2019) cannot explain most of the overlap. If non-farm informal businesses mainly help household

smooth agricultural shocks, the new businesses might not necessarily be more productive. The overlap would then be predominately driven by informal businesses whose owners enter from agriculture and/or exit to agriculture. However, the overlap exists for businesses of owners that enter/exit from agriculture and non-agriculture. The overlap is present in rural and urban samples (although is somewhat stronger in rural areas) and in all industries.

Second, the overlap does not simply reflect adding and dropping of business activities that are complementary to existing household activities or mismeasurement of businesses within a household. It persists when we repeat the analysis using a household (rather than a business) as the unit of observation. The overlap persists when a household adds or drop a new business (when it did not have one) versus adding an additional businesses. Finally, the overlap does not simply reflect churning in underlying household composition, that is, it is not driven by people that own a business joining or leaving the household (see Figure A3).

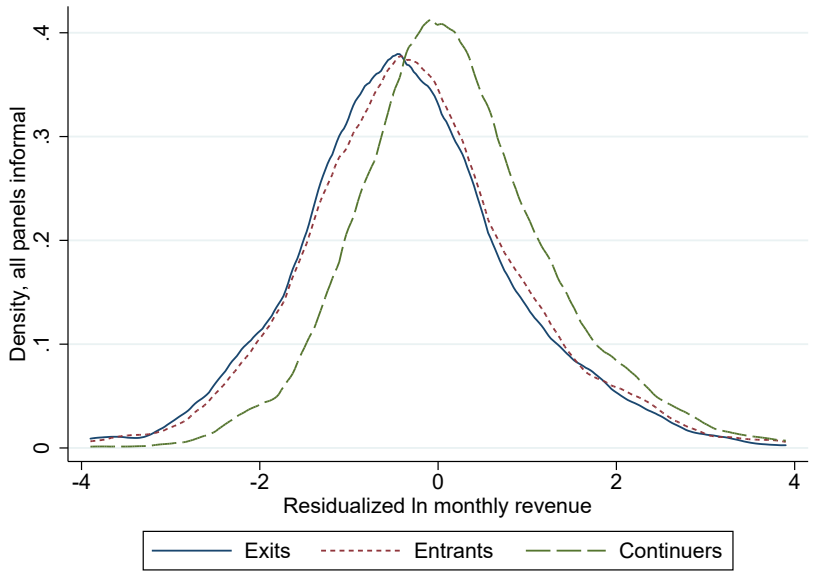


Figure A3: Revenue distribution for entering, exiting, and continuing informal businesses, excluding businesses run by owners joining and leaving the household

Note: The sample is all informal businesses in the middle survey of the 3-survey panels, excluding businesses run by owners who either leave or join the household. We removed survey fixed effects from the revenue distribution by regressing ln monthly revenue on survey fixed effects. We plot the residuals.

In Figure A4 we plot the revenue distributions for continuers, entrants, and exiters with confidence intervals. The confidence intervals statistically confirm the conclusions presented

in subsection 5.2 that the distribution for continuers is shifted to the right of those for both exiters and entrants and that the distributions for exiters and entrants are statistically indistinguishable.

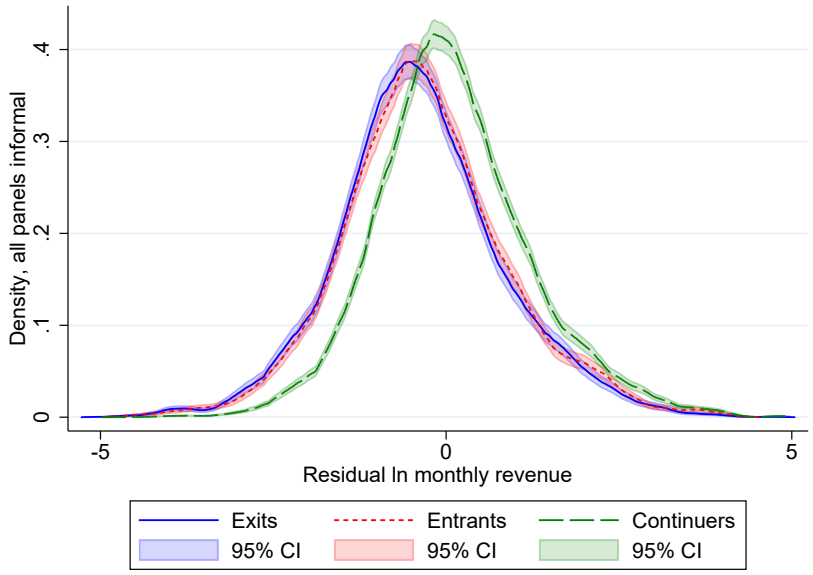


Figure A4: Revenue distribution for entering, exiting, and continuing informal businesses with 95% confidence intervals

Note: The sample is all informal businesses in the middle survey of the 3-survey panels, excluding businesses run by owners who either leave or join the household. We removed survey fixed effects from the revenue distribution by regressing ln monthly revenue on survey fixed effects. We plot the residuals.

In Figure A5 we display the distribution of monthly hours worked by the owner. Recall that a large majority of informal businesses have only the owner as a worker and thus the hours spent working by the owner is the only labor input for the majority of these businesses. Here too we see strong evidence of an overlap in hours worked by the owner for entering and exiting businesses and the owners of exiting and entering businesses work fewer hours in a month, on average, than owners of continuing businesses.

In addition to similar hours worked by owners in entering and exiting businesses, the human capital of the owners, as measured by highest grade completed, also substantially overlap. In Figure A6 we report the share of owners by highest grade completed. We pool across 2-survey panels. Entering owners are those that operated a business at the end of the panel, but not at the beginning, and exiting owners are those that operated a business at the beginning of the panel, but not at the end. The share of exiting owners with a specific

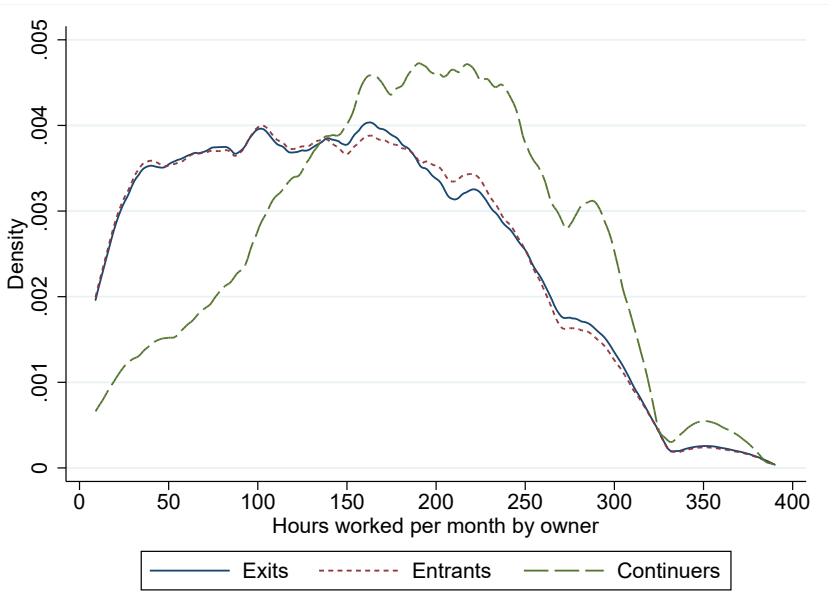


Figure A5: Monthly hours worked by owner distribution for entering, exiting, and continuing informal businesses

Note: The sample is informal businesses in the middle survey of all 3-survey panels. A continuer is a business that operated in all three surveys. An entrant is a business that operated in the middle survey, but not in first survey. An exiter is a business that operated in middle survey, but not in the last survey. We removed survey fixed effects from the hours distribution by regressing hours on a constant and survey fixed effects for 2012, 2014, and 2016. We then subtracted the estimated survey fixed effects for 2012, 2014, and 2016.

highest completed grade is extremely similar to the share of entering owners with the same highest completed grade. The results are very similar for each 2-survey panel and thus we do not report them. Thus, the evidence for two important factors of production, hours worked by the owner and the human capital of the owner, show the same high degree of overlap as revenue for entrants and exiters.

We conclude this subsection by examining the robustness of Figure 6b which shows the high degree of overlap between entrants and exiters in terms of the number of household members working in the business. The 2004-06-08 panel allows us to address whether the same pattern is observed using total number of workers in the business. Figure A7 demonstrates that using the total number of workers leads to the same conclusions: the vast majority of entrants and exiters have only one worker (the owner) and more broadly that the distribution of the number of workers is very similar for entrants and exiters. This confirms the results in Figure 6b.

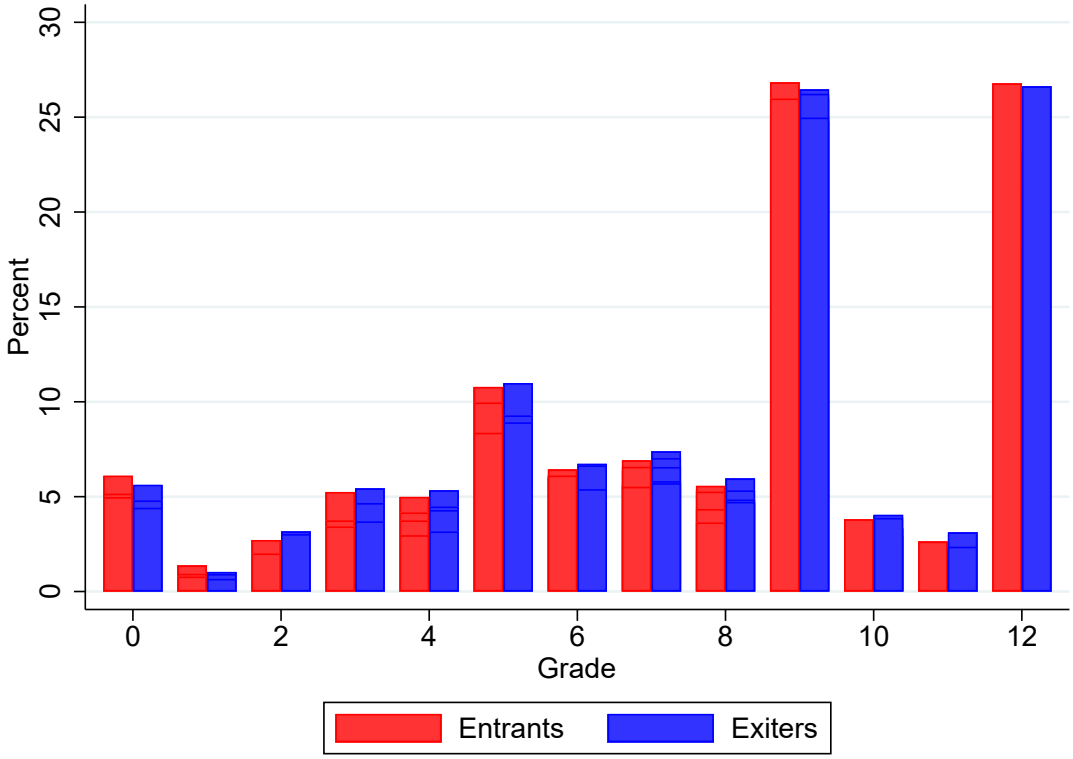


Figure A6: Distribution of highest grade completed by entrants and exiters

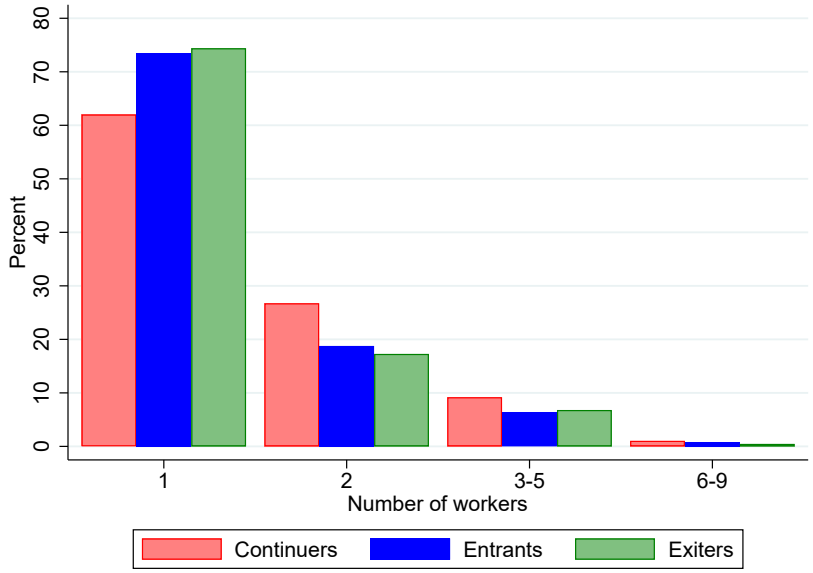


Figure A7: Number of workers in continuers, exiters and entrants

Note: The sample is all informal businesses in 2006 operated by households in the 2004-06-08 panel. A continuer is a business that oeprated in all three surveys. An entrants is a business that operated in 2006, but not in 2004. An exiter is a business that operated in 2006, but not 2008.

A.3 Residual based measure of productivity

In subsection 5.2 we showed that the revenue distribution of entrants and exiters nearly perfectly overlapped. Here, we pursue a residual-based way of estimating total factor productivity (TFP).

Measuring TFP is challenging in most firm datasets, but the challenges are even greater in this context where we do not have reported capital for the businesses. Nonetheless, we can incrementally move closer to TFP by using information on the number of reported workers in the business and information on expenses. We have information on the number of workers in the business (both household and non-household members) in the 2006 and 2008 surveys and thus limit the analysis to these surveys. We estimate a residual based measure of TFP based on the following specification:

$$y_{ik} = \alpha_k + \beta l_i + \sum_j \theta^j D_i^j + \epsilon_i \tag{2}$$

where y_{ik} is the ln of real revenue of business i in industry k, α_k is an industry fixed effect, l_i is ln number of workers, and D_i^j is a series of indicator variables for positive expenses for expense item j (materials, non-durable tools, repairs, depreciation, rent, labor, interest, taxes, energy and water, and other).[49] When we do this in the 2006 and 2008 repeated cross sections, we find that the correlation between ln revenue and the residual from the above equation is 0.74 in 2006 and 0.75 in 2008. The high correlation suggests that using revenue may be a reasonable proxy for TFP in this context.

Figure A8 shows the distribution of this residual-based estimate of TFP for entrants, exiters, and continuers in the 2004-06-08 panel. It is consistent with the revenue distributions shown in Figure 6 as the distributions for entrants and exiters are essentially identical and shifted to the left of that for continuing informal firms.

[49] We remove observations in industries with fewer than 10 businesses. In the 2006 cross section this leads to the removal of 42 out of 20,001 informal businesses.

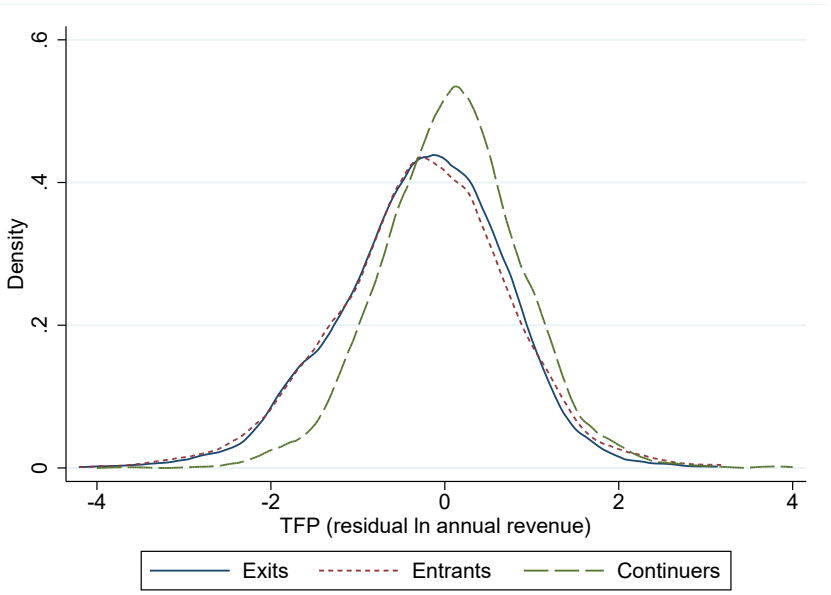


Figure A8: TFP distribution for entering, exiting, and continuing informal businesses

Note: The sample is informal businesses in the 2006 survey that are part of the 2004-06-08 panel. A continuer is a business that operated in all three surveys. An entrant is a business that operated in 2006, but not in 2004. An exiter is a business that operated in 2006, but not in 2008.

A.4 Why do informal firms exit?

The analysis in subsections 5.1 and 5.2 highlights that informal business entry and exit are in part systematically related to the level of economic development and productivity of informal business. More generally, a decision to exit could be driven by firm-specific shocks that influence its profitability, such as low firm productivity, increases in costs of inputs, or increased product market competition. These are common reasons for exit of formal firms. Because informal businesses usually have only one worker, the owner, the decision to close down the business can also depend on alternative employment opportunities of the owner (such as wage employment), or shocks to the owner health or household (such as age and health of the owner or illness in the family).

In order to further examine the reasons for informal business exit, we rely on a survey of former business owners that asked the owners about specific reasons for exit. Table A1 displays the results. It reports the reasons for all businesses, as well as for businesses located in provinces above and below the median provincial per capital income.[50] First,

[50]These results are based on all exiting businesses owned by the household surveyed in a special module

Table A1: Reason for exit

Reason	All	Above median provincial per capita income	Below median provincial per capita income
Lack of capital/credit	.096	.116	.065
Drop in product price	.043	.037	.057
Increase in production cost	.067	.066	.075
Loss of main source of inputs	.047	.029	.068
Too old and not strong enough	.101	.093	.124
Drop in labourers in the household	.096	.067	.135
Other activities with more income	.277	.303	.221
Other	.272	.289	.255
Number of businesses	386	221	151

more profitable economic opportunities are the most common reported reason for exit. In particular "Other activities with more income" is the most common reason stated for exit, accounting for 27% of the exits. This answer is more prevalent in richer provinces. To the extent that "Other activities with more income" also influence entry, this evidence is consistent with the negative relationship between exit and entry rates and provincial income per capita in Figure 5.[51] Production-level shocks are also important jointly as a reason for exit, including because of "Lack of capital/credit" (9.2%), "Increase in production cost" (6.5%), and "Loss of main source of inputs" (4.5%). 4.2% of owners mention lower price for the reason to exit. Finally, owner and household-specific factors also play a role. 10% of businesses exit because the owner is "Too old and not strong enough" and 9.9% due to "the drop in number of laborers within the household." Both of these reasons are more prevalent in poorer provinces, where we would more likely expect imperfections in the labor market, in part due to outmigration of younger workers to more prosperous provinces. Finally, 27% of exiting business report the residual category "i.e. other," as a reason to exit, highlighting the heterogeneity in reasons for exiting across owners.

of the 2004 VHLSS. The survey includes both informal and formal businesses, but the vast majority of businesses owned by the household at the time are informal. Even two years after this survey, only 2.2% of businesses owned by households were formal.

[51] Unfortunately, the surveys do not provide information on reasons for why informal businesses enter. The decision to enter could be due to the owner having a good business idea or foreseeing a market opportunity, lack of other employment opportunities, or coping with negative household shocks.

A.5 Growth of surviving entrants

In this subsection, we expand on the results shown in subsection 5.3 on the performance of entering informal firms.

Survival among entrants is less likely than the survival of equally performing continuing firms. Figure A9 plots the probability of survival of entrants and continuing informal firms in t+2 in relationship to their monthly revenue. Although the probability of survival increases with revenue for both set of firms, informal entrants are less likely to survive than continuing informal firms with the same revenue. This illustrates the difficulties of survival for new informal businesses.

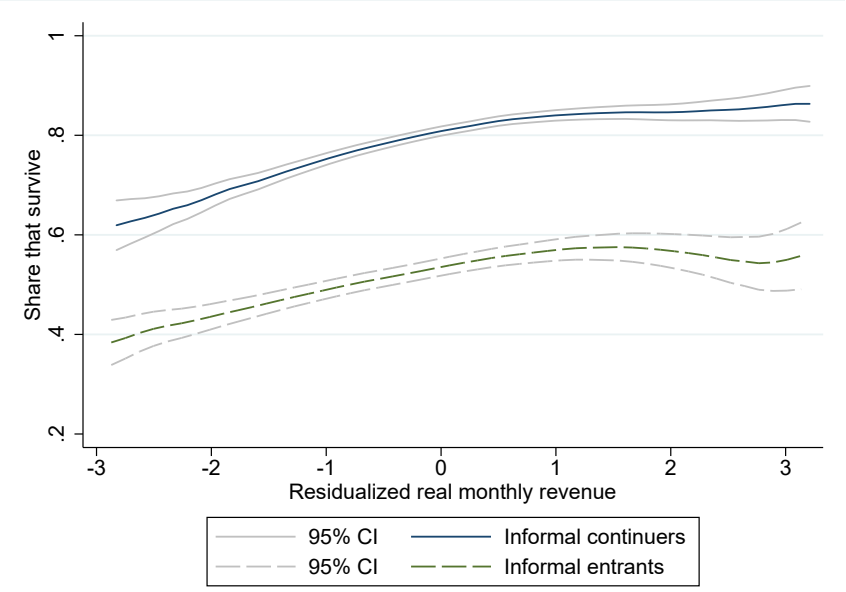
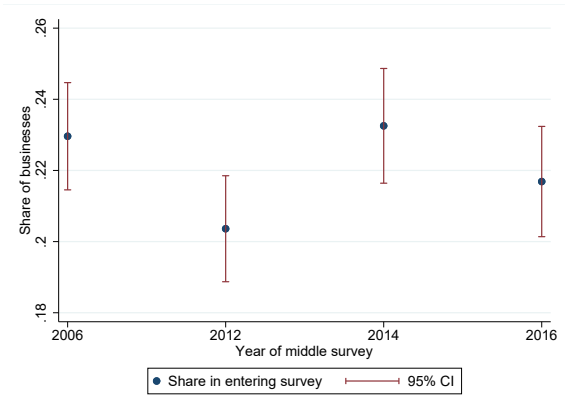


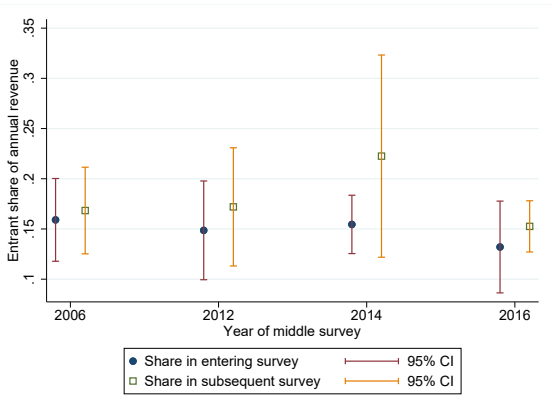
Figure A9: Share of informal businesses that survive by initial revenue

Note: The sample is based on all informal businesses that operated in the first survey in the 2-survey panels. We removed survey fixed effects from the revenue distribution by regressing ln monthly revenue on survey fixed effects. We plot the residuals.

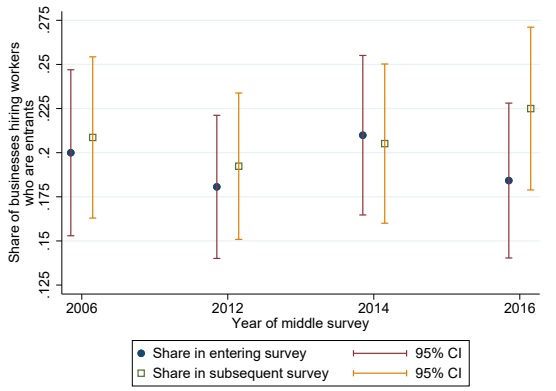
Finally, although initially more successful entrants survive, they do not contribute much to aggregate outcomes in the informal sector, including the share of firms, revenue, or employment. Figure A10 (a) plots the entrants that survive to t+2 as a share of all informal firms. Surviving entrants constitute 15-17.5% of all informal firms. Panel (b) plots the share of total informal sector revenue in the year of entry, survey t, and in the subsequent survey at t+2 for entrants that will survive up to t+2. The surviving informal entrants account on



(a) Share of businesses



(b) Share of revenue



(c) Hiring

Figure A10: Share of informal sector outcomes due to surviving entrants

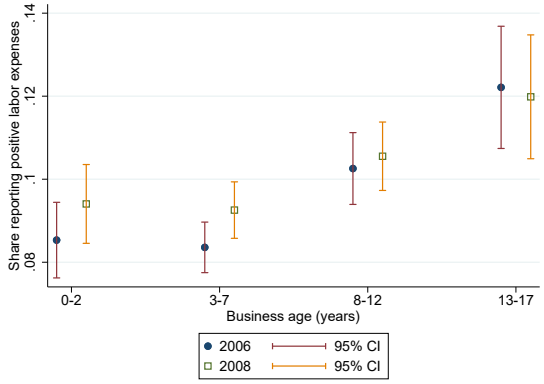
average for about 12-13% of total revenue among informal businesses, less than their share of informal businesses. Interestingly, while the surviving informal entrants do not expand substantially between 2006 and 2008 or between 2014 and 2016, there is some evidence of relative expansion in the other years, consistent with positive growth subsequent to entry. This perhaps highlights the positive selection into entrants over time (and their subsequent higher growth).

Figure A10 (c) plots the share of informal businesses that hire workers that is accounted for by surviving entrants. About 15% of informal businesses that hire workers are surviving entrants, right around their respective share of informal businesses. Hence, surviving entrants are smaller on average in terms of revenue, but about equally likely to be engaged in hiring workers outside the household as other informal businesses. There is no clear pattern that the share of businesses hiring workers moves disproportionately toward entrants over time,

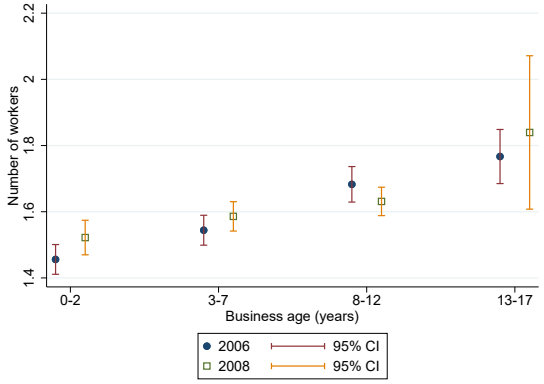
with the share not changing in a statistically significant way over time.

A.6 Informal business performance and age

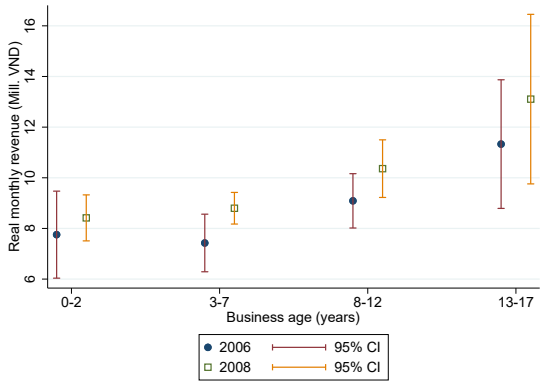
In this subsection, we build on the results in subsection 5.3 by focusing on the relationship between the age of informal businesses and several performance measures (see Hsieh and Klenow (2014) for analysis with formal and informal firms in India and Mexico). In particular, we plot the relationship between informal business age and five outcomes: the share of businesses that report positive labor expenses, the average number of workers, average monthly revenue, the share of businesses that operated for the past 12 months, and the share that are the owner's primary job. We also report the age distribution of informal firms. We use the 2006 and 2008 repeated cross sections as the age of the business is only available in these surveys. These plots are presented in Figure A11. The informal firms are grouped into 0-2, 3-7, 8-12, and 13-17 year old firms. Of course, unlike the panel-data analysis above, this analysis cannot distinguish whether the observed trends are related to changes within informal businesses or simply reflect the selection of better firms over time. With that in mind, all five figures show a gradual increase in the performance measure with the age of the business. Importantly, there is no statistically significant difference between the performance of 0-2 year old informal businesses and 3-7 year old businesses for positive labor expenses, number of workers, and monthly revenue. However, businesses that survive 8-12 and 13-17 years post entry are more likely to have positive labor expenses, have a higher average number or workers, and higher monthly revenue than businesses that 0-2 years old. The estimates suggest a 33% increase in the incidence of positive labor expenses, a 20% increase in the number of workers, and almost a doubling of monthly revenue over a period of 11-15 years. However, note that only 10% of businesses survive 13-17 years post entry. Operating for the past 12 months and being the owner's primary both show the same positive relationship with age as the other measures, but with a more pronounced increased between 0-2 and 3-7 years old.



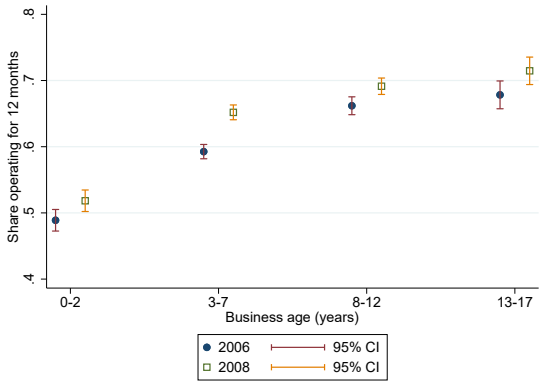
(a) Positive labor expenses



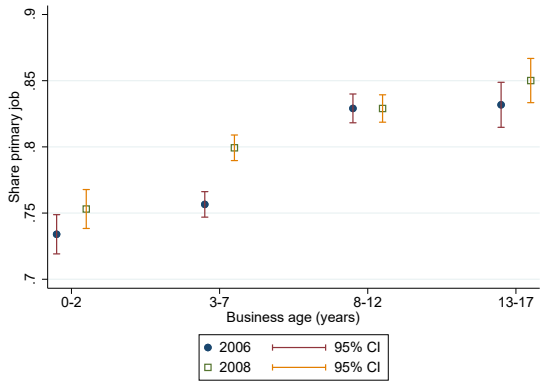
(b) Number of workers



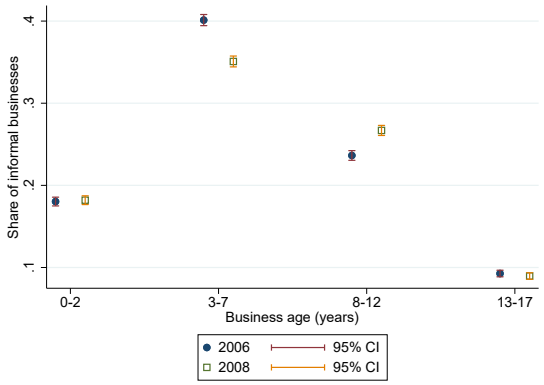
(c) Monthly revenue



(d) Operate all 12 months



(e) Primary job



(f) Share of informal businesses by age

Figure A11: Informal business outcomes by age

A.7 Do informal businesses get better over time?

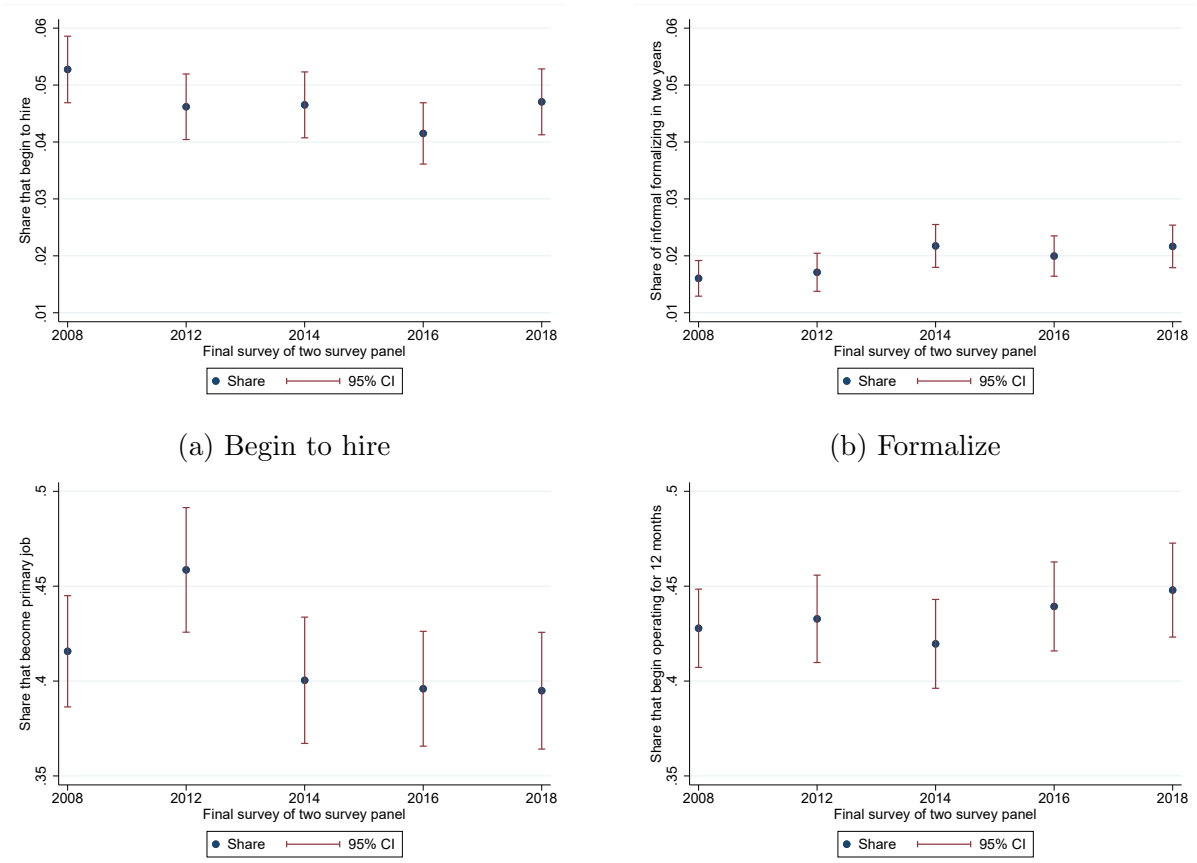
In this subsection, we expand the analysis from subsection 5.3, which examined growth among surviving entrants, to focus on growth among continuing informal businesses.

How likely are continuing informal firms to grow and transition to formal businesses over time and across provinces that vary in initial prosperity? The existing literature has emphasized hiring of paid workers (outside of the family) and formalization as indicators of business expansion (de Mel *et al.*, 2010, 2013). Recall from Table 2 that the share of informal businesses that hire workers beyond family labor in Vietnam, that is, have positive labor expenses, is very low, but increases from 9.0% in 2006 to 11.6% in 2018. In addition, the share of formal businesses is also very low (Figure A14). Operating as the manager's main job and all year round are more common (and also increases over time). However, that analysis does not distinguish how transitioning from no workers to hiring non-family workers (or formalization) changes over time and space with economics development.

Further analysis indicates that transitioning to employing paid workers is rare among continuing informal businesses. Only 4.5% of surviving informal firms that do not initially hire paid employees report doing so within 2 years (Figure A12a).

Becoming a formal business is even more rare. Figure A12b plots the share of informal businesses that become formal. Interestingly, this probability increases slightly over time from 1.5% in 2008 to about 2% in 2018, but it still remains very low.

In contrast, it is more common for an informal business to operate as the owner's main job and operate all year around, as well as to become the owner's primary job or to start operating all year (Figures A12c and A12d). Around 40% of surviving informal businesses that were not the manager's primary job become so within two years. Similarly, between 42 and 44% of surviving informal businesses that were not operating all year begin to do so within two years. The higher likelihood of transitioning to being the manager's primary job or to operating all year is consistent with the aggregate trends shown in Table 2. However, over the span of 15 years (during which the share of households having an informal business is declining), these relationships are relatively stable over time, suggesting that even in times of economic growth, transitioning to hiring and formalization is hard.



(a) Begin to hire

(b) Formalize

(c) Become the owner's primary job

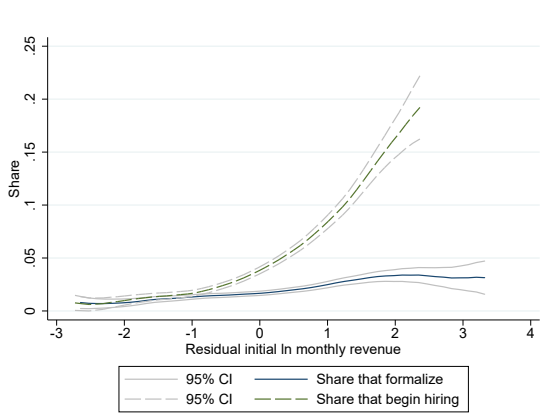
(d) Begin to operate all round

Figure A12: Share of informal firms that start to hire, formalize, become the primary job, and begin to operate all year

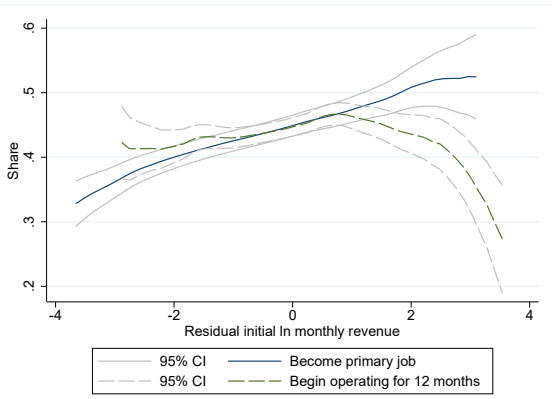
The probability of starting to hire workers, formalize, become the primary job of the owner, and operate all year is closely related to informal business performance. In particular, it increases with past revenue (see Figure A13a).[52] This figure further highlights that even the best performing informal firms face a low likelihood of formalizing, while the likelihood of starting to hire workers outside the family is more responsive to firm performance (although continues to be low). Transitioning into a business being the primary activity of the owner (and operating it all year around) also increases with initial business revenue, but this transition is more likely to occur at all levels of initial revenue.[53]

In sum, our analysis suggests that continuing informal firms that are more successful

[52]Our longitudinal analysis proceeds as follows. We pool the data from the all the two-survey panels and in each panel focus on continuing informal firms. Among the firms that did not hire workers in the initial survey, we examine the probability that the firm hires workers in the subsequent survey as a function of its revenue in the initial survey. We do the same for formalization, becoming the primary job of the owner, and



(a) Begin hiring and formalize



(b) Become primary job and begin operating all year

Figure A13: Share of informal firms that start to hire, formalize, become primary job, and begin to operate all year relative to initial revenue

Note: The sample is all businesses in the 2-survey panels that are informal in the start survey. Depending on the figure, the sample is further restricted to those that did not hire, were not the owner's primary job, or did not operate for all 12 months in the start survey. We removed survey fixed effects from the revenue distribution by regressing ln monthly revenue on survey fixed effects. We plot the residuals.

are more likely to start hiring workers and register as an enterprise, although in general these changes are uncommon. However, informal businesses that start hiring workers, and transition to the formal sector are rare. This is consistent with the finding on the role of selection in hiring and formalization by Maloney (2004), La Porta and Shleifer (2008), and de Mel *et al.* (2013),

A.8 Household income and informal business entry and exit

Similar to section 5.4, we examine the impact of entry and exit at the household level. For although individual income fluctuates considerably with opening or closing an informal business, this may be smoothed within a household.

We focus on households exiting and entering the informal business sector relative to households continuing to operate in the informal business sector. We use households that continue to operate an informal businesses as a reference group. We define an exiter as a household that operated one or more informal business at the start of 2006-08 panel

operating a business all year.

[53]Most informal businesses with high initial revenue operate year-around, hence the estimates of operating full year at high levels of revenue are very noisy and based on a few observations.

and no informal businesses at the end of the panel. An entrant is defined as a household that operated no informal businesses at the start of the 2006-08 panel and one or more informal businesses at the end of the panel.[54] We exclude all households that operated any formal businesses in 2006 or 2008. The exiters and entrants are compared to households that operated at least one informal business in both 2006 and 2008. Table A2 reports the results. In columns 1 and 2 we focus on the change in household income per capita as a dependent variable, with column 2 using values windsorized at the 1st and 99th percentiles. Households that start an informal business observe an increase in household income, while informal business exit is associated with a reduction in household income. The reduction in income in connection with exit is consistent with results from McKenzie and Paffhausen (2019). Since informal business earnings are measured with potentially large errors (de Mel *et al.*, 2009), we also report a specification based on an indicator for a positive change in household income in column 3. We find that households households starting an informal business are more likely to experience a positive change in income, while households exiting the informal business sector are less likely to experience a positive increase in income relative to continuing households. These results confirm the importance of informal businesses for income in households, with entry significantly adding to income and exit contributing to declines in income.

[54]Our household income estimates come from Benjamin *et al.* (2017). We restrict our household income analysis to the 2006-08 panel because Benjamin *et al.* (2017) only had access to smaller samples of the 2012 and 2014 household surveys.

Table A2: Changes in household income due to informal business entry and exit

	(1) Change	(2) Change	(3) Positive change
Exiter	-12.34	-10.95	-0.11
	(3.10)	(1.32)	(0.01)
Entrant	6.27	6.80	0.09
	(1.75)	(1.31)	(0.01)
Change in household size	6.32	6.47	0.05
	(1.20)	(0.63)	(0.00)
Change in houeshold workers	7.65	6.54	0.07
	(1.37)	(0.66)	(0.01)
Observations	8861	8861	8861

The dependent variable in columns 1 and 2 is the change in household income and in column 3 is an indicator for a positive change. The sample is all households that operated an informal business in either or both of the start or end survey of the 2006-08 panel. We exclude households that operated a formal business. Observations in column 2 are winsorized at the 1st and 99th percentile. All income are reported in million VND (Jan 2012 prices).

A.9 Additional Tables

Table A3: Observations in business panels

	Number of businesses		
Panel	First survey	Last survey	Matched across all surveys
2004-06	10248	9923	6743
2006-08	9416	9468	6344
2010-12	8582	8311	5882
2012-14	8053	8404	5905
2014-16	8282	8512	6117
2016-18	8208	8356	5966
2004-08	4664	4487	2370
2010-14	4021	4013	2286
2012-16	3473	3676	2098
2014-18	3685	3897	2187

Table A4: Business and manager characteristics, repeated cross sections

	2004	2006	2008	2010	2012	2014	2016	2018
Business characteristics								
Primary industries	.009	.008	.008	.006	.005	.004	.002	.002
Secondary industries	.275	.264	.257	.241	.228	.233	.217	.219
Tertiary industries	.716	.727	.735	.753	.767	.763	.781	.779
Urban	.334	.351	.355	.393	.405	.398	.406	.395
Formal		.022	.021	.021	.026	.026	.029	.032
Operated for past 12 months	.575	.596	.634	.643	.673	.679	.708	.736
Real revenue (million VND)	90.1	107.7	117.7	136.7	152.7	165.8	186.1	204.9
Real expenses (million VND)	59.6	71	75.6	84.1	92.7	99.3	105.2	118.5
Labor expense share	.028	.033	.035	.041	.042	.044	.049	.05
Positive labor expenses	.091	.098	.105	.117	.124	.12	.127	.128
Workers	1.67	1.67	1.7					
Paid workers	.32	.36	.38					
Age	7.69	8.18	8.48					
Manager characteristics								
Primary job	.733	.769	.793	.791	.817	.793	.808	.813
Female	.587	.586	.583	.575	.575	.577	.573	.565
Age	40.5	41.7	42.8	41.9	42.9	43.8	44.8	45.8
Grade	7.57	7.61	7.67	7.81	7.93	8	8.08	8.2
Less than primary	.183	.183	.176	.166	.155	.153	.151	.143
Completed primary	.299	.293	.292	.29	.284	.275	.265	.255
Completed lower secondary	.346	.34	.338	.331	.337	.336	.329	.335
Completed upper secondary	.171	.184	.194	.211	.221	.233	.25	.263
Ethnic minority	.077	.085	.089	.092	.084	.089	.093	.098
Number of businesses	21458	20458	20465	19544	18029	18314	18346	18049

Table A5: Before and after operating a business

Panel		Informal mangers age 15 to 64				All age 15 to 64	
		Before entering (share)	After exiting (share)	Before entering (mean grade)	After exiting (mean grade)	First survey (share)	Second survey (share)
2006-08	Wage worker in agriculture	.028	.023	5.8	5.5	.036	.034
	Wage worker in informal sector	.129	.124	8.1	7.7	.069	.078
	Wage worker in formal sector	.077	.073	9.8	9.6	.111	.124
	Self-employed in agriculture	.417	.363	6.5	7.1	.375	.364
	Self-employed in informal sector	.165	.15	7.9	7.8	.127	.129
	Self-employed in formal sector	.004	.005	7.8	10.1	.003	.003
	Not in the workforce	.094	.119	7.8	7.7	.225	.147
	Not in the household	.086	.143			.054	.12
	All	1	1	7.4	7.6	1	1
2016-18	Wage worker in agriculture	.011	.006	6.3	7.8	.036	.035
	Wage worker in informal sector	.156	.147	8.3	8.3	.096	.102
	Wage worker in formal sector	.124	.121	10.1	9.8	.166	.177
	Self-employed in agriculture	.365	.279	6.9	7.1	.322	.297
	Self-employed in informal sector	.144	.167	9.0	8.6	.141	.146
	Self-employed in formal sector	.005	.007	9.5	10.6	.004	.003
	Not in the workforce	.11	.109	8.5	7.8	.178	.129
	Not in the household	.085	.164			.059	.111
	All	1	1	8.1	8.1	1	1

Table A6: Employment growth and exit by age and number of workers

	Number of workers				
Age (years)	1	2	3-5	6-9	Total
Panel A: Mean employment growth rate conditional on survival					
0-2	.292	-.148	-.075	.422	.178
3-5	.192	-.093	-.098	.447	.11
6-8	.182	-.048	-.155	.001	.104
9-12	.199	-.046	-.079	-.026	.105
13+	.22	-.026	-.185	.057	.089
Total	.212	-.064	-.127	.163	.113
Panel B: Share of businesses that exit by 2008					
0-2	.483	.427	.381	.21	.464
3-5	.38	.283	.298	.132	.353
6-8	.346	.253	.283	.19	.321
9-12	.266	.237	.219	.303	.255
13+	.264	.166	.167	.121	.222
Total	.356	.264	.254	.184	.325
Panel C: Number of informal businesses in 2006					
0-2	1217	314	105	13	1649
3-5	1686	476	134	16	2312
6-8	1311	386	131	12	1840
9-12	1057	420	154	21	1652
13+	993	473	201	30	1697
Total	6264	2069	725	92	9150
Panel D: Number of businesses that exit by 2008					
0-2	587	138	42	3	770
3-5	642	128	38	3	811
6-8	468	97	40	2	607
9-12	291	97	34	5	427
13+	263	76	34	5	378
Total	2251	536	188	18	2993

The sample is informal businesses in 2006 that are operated by households in the 2006-08 panel.

Table A7: Household worker transitions among surviving entrants

Household workers at entry	Household workers 2 years later: 1	2	3	4+	Number of businesses
1	.884	.106	.008	.002	1621
2	.368	.589	.032	.011	375
3	.25	.346	.346	.058	52
4+	0	.176	.353	.471	17

The sample includes surviving informal entrants across all 3-survey panels. A surviving informal entrant is a business that appeared in the middle and end surveys of the 3-survey panel, but not the start survey and was informal in the middle survey. In each row, we report the share of businesses that had the indicated number of household workers at entry by the number of household workers 2 years later. The shares sum to 1 across a row.

Table A8: Worker transitions among surviving entrants

Workers at entry	Workers 2 years later: 1	2	3	4+	Number of businesses
1	.827	.133	.027	.013	480
2	.309	.604	.06	.027	149
3	.226	.419	.29	.065	31
4+	.048	.095	.333	.524	21

The sample includes surviving informal from the 2004-06-08 panel. A surviving informal entrant is a business that operated in 2006 and 2008, but not in 2004, and was informal in 2006. In each row, we report the share of businesses that had the indicated number of workers at entry by the number of workers 2 years later. The shares sum to 1 across a row.

Table A9: Estimates of exit rates in the literature: Developing countries, small firms

Source	Country	Firm type	Period	Annual exit rate
Davies and Kerr (2018)	Ghana	Small (0-9 workers) manufacturing firms	10 years	3.5
Nagler and Naude (2017)	Uganda	Household non-farm	2005/06-2009/10	7.6
			2009/10-2010/11	30.8
			2010/11-2011/12	26.2
	Nigeria		2010/11 + 6 months	26.3
			2010/11 + 1.5 years	3.0
			2012 + 6 months	3.8
Krafft (2016)	Egypt	Household non-farm (formal and informal)	1998-2006	6.4
			2006-2012	7.6
Vijverberg and Haughton (2002)	Vietnam	Household non-farm	1992/93-1997/98	12.2
Vijverberg et al. (2006)	Vietnam	Household non-farm	2002-2004	18.4
Cling et al. (2010)	Vietnam	Informal businesses	2007-2009	13.5
Field et al. (2013)	India	Microenterprises	3 years	12.9

Table A10: Estimates of exit rates in the literature - developing countries, large and formal firms

Source	Country	Firm type	Period	Annual exit rate
Soderbom, Teal, Harding (2006)	Ghana	Manufacturing	5 years	3.8
	Kenya	Manufacturing	5 years	7.8
	Tanzania	Manufacturing	5 years	8.8
Davies and Kerr (2018)	Ghana	Small and large manufacturing firms	10 years	3.3
Klapper and Richmond (2011)	Cote d'Ivoire	All formal firms	1977-1997	10
Shiferaw (2009)	Ethiopia	Manufacturing (10+ workers)	1996-2002	16
Lay (2003)	Taiwan	Manufacturing plants	1987-1998	12.8
Tybout (1996)	Chile	Manufacturing plants (10+ workers)	1980-85	10.8
Haddad, de Melo, and Horton (1996)	Morocco	Manufacturing (10+ workers)	1984-1989	6
Roberts (1996)	Colombia	Manufacturing plants (all until 1982, 10+ workers after)	1977-1985	11.1
Bartelsman, Haltiwanger, and Scarpetta (2009)	Argentina	Manufacturing	1988-1997	8
		All sectors	1988-1997	9
	Latvia	Manufacturing	1996-2002	5
		All sectors	1996-2002	5
	Estonia	Manufacturing	1995-2001	6
		All sectors	1995-2001	6
	Colombia	Manufacturing (20+ workers)	1982-1998	4
	Romania	Manufacturing	1992-2001	6
		All sectors	1992-2001	7.5
	Chile	Manufacturing (20+ workers)	1996-2001	4
	Mexico	Manufacturing	1985-2001	10
		All sectors	1985-2001	11
	Brazil	Manufacturing	1996-2001	11
	Slovenia	Manufacturing	1992-2001	5
		All sectors	1992-2001	4
	Hungary	Manufacturing	1992-2001	6
		All sectors	1992-2001	6
Authors own calculation	Vietnam	Manufacturing	2000-2017	17
		All sectors	2000-2017	14

Table A11: Estimates of exit rates in the literature - developed countries

Source	Country	Firm type	Period	Annual exit rate
Bartelsman, Haltiwanger, and Scarpetta (2009)	Germany (West)	Manufacturing	1977-1999	6
		All sectors	1977-1999	6
	Finland	Manufacturing	1988-1998	7
		All sectors	1988-1998	6
	Italy	Manufacturing	1986-1994	7
		All sectors	1986-1994	7.5
	Netherlands	Manufacturing	1987-1997	5
		All sectors	1987-1997	6
	Portugal	Manufacturing	1983-1998	6
		All sectors	1983-1998	6
	USA	Manufacturing	1988-1997	8
		All sectors	1988-1997	10
	Denmark	Manufacturing	1981-1994	10
		All sectors	1981-1994	11
	France	Manufacturing	1989-1997	9
		All sectors	1989-1997	7.5
	UK	Manufacturing	1980-1988	10
	Canada	Manufacturing	1984-1998	7.5

Table A12: Estimates of entry rates in the literature: Developing countries, small firms

Source	Country	Firm type	Period	Annual entry rate
Mead and Liedholm (1998)	Botswana	All sectors, up to 50 workers	1 year	32.0
	Kenya	All sectors, up to 50 workers	1 year	19.7
	Malawi	All sectors, up to 50 workers	1 year	24.6
	Swaziland	All sectors, up to 50 workers	1 year	22.0
	Zimbabwe	All sectors, up to 50 workers	1 year	23.5
	Dominican Republic	All sectors, up to 50 workers	1 year	27.1
Nagler and Naude (2017)	Uganda	Household non-farm	2005/06-2009/10	9.0
			2009/10-2010/11	30.0
			2010/11-2011/12	23.3
	Nigeria	Household non-farm	2010/11 + 6 months	28.9
			2010/11 + 1.5 years	13.1
			2012 + 6 months	21.4
Krafft (2016)	Egypt	Household non-farm (formal and informal)	1998-2006	7.7
			2006-2012	7.6
Vijverberg et al. (2006)	Vietnam	Household non-farm	2002-2004	18.4

Table A13: Estimates of entry rates in the literature - developing countries, large and formal firms

Source	Country	Firm type	Period	Annual entry rate
Klapper and Richmond (2011)	Cote d'Ivoire	All formal firms	1977-1997	12.4
Shiferaw (2009)	Ethiopia	Manufacturing (10+ workers)	1996-2002	20
Lay (2003)	Taiwan	Manufacturing plants	1987-1998	13.5
Tybout (1996)	Chile	Manufacturing plants (10+ workers)	1980-85	6
Haddad, de Melo, and Horton (1996)	Morocco	Manufacturing (10+ workers)	1984-1989	13.0
Roberts (1996)	Colombia	Manufacturing plants (all until 1982, 10+ workers after)	1977-1985	12
Bartelsman, Haltiwanger, and Scarpetta (2009)	Argentina	Manufacturing	1988-1997	7
		All sectors	1988-1997	10
	Latvia	Manufacturing	1996-2002	23
		All sectors	1996-2002	22
	Estonia	Manufacturing	1995-2001	10
		All sectors	1995-2001	11
	Colombia	Manufacturing (20+ workers)	1982-1998	4
	Romania	Manufacturing	1992-2001	17
		All sectors	1992-2001	20
	Chile	Manufacturing (20+ workers)	1996-2001	5.5
	Mexico	Manufacturing	1985-2001	12
		All sectors	1985-2001	15
	Brazil	Manufacturing	1996-2001	15
	Slovenia	Manufacturing	1992-2001	16
		All sectors	1992-2001	21
	Hungary	Manufacturing	1992-2001	15
	All sectors	1992-2001	20	
Authors own calculation	Vietnam	Manufacturing	2000-2017	26
		All sectors	2000-2017	26

Table A14: Estimates of entry rates in the literature - developed countries

Source	Country	Firm type	Period	Annual entry rate
Bartelsman, Haltiwanger, and Scarpetta (2009)	Germany (West)	Manufacturing	1977-1999	5
		All sectors	1977-1999	6
	Finland	Manufacturing	1988-1998	8
		All sectors	1988-1998	9
	Italy	Manufacturing	1986-1994	7
		All sectors	1986-1994	8
	Netherlands	Manufacturing	1987-1997	8
		All sectors	1987-1997	10
	Portugal	Manufacturing	1983-1998	12
		All sectors	1983-1998	15
	USA	Manufacturing	1988-1997	9
		All sectors	1988-1997	11
	Denmark	Manufacturing	1981-1994	7
		All sectors	1981-1994	9
	France	Manufacturing	1989-1997	10
		All sectors	1989-1997	720
	UK	Manufacturing	1980-1988	11
	Canada	Manufacturing	1984-1998	8

A.10 Additional Figures

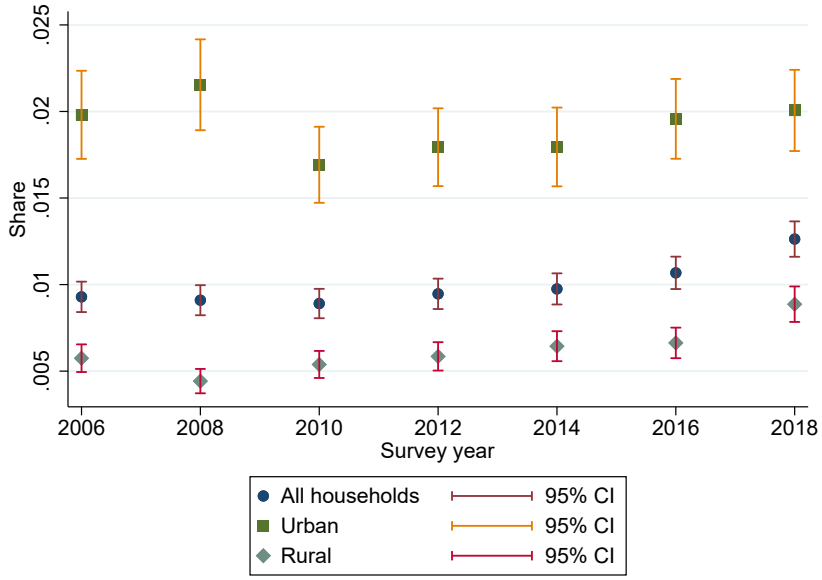


Figure A14: Share of households operating a formal business by survey

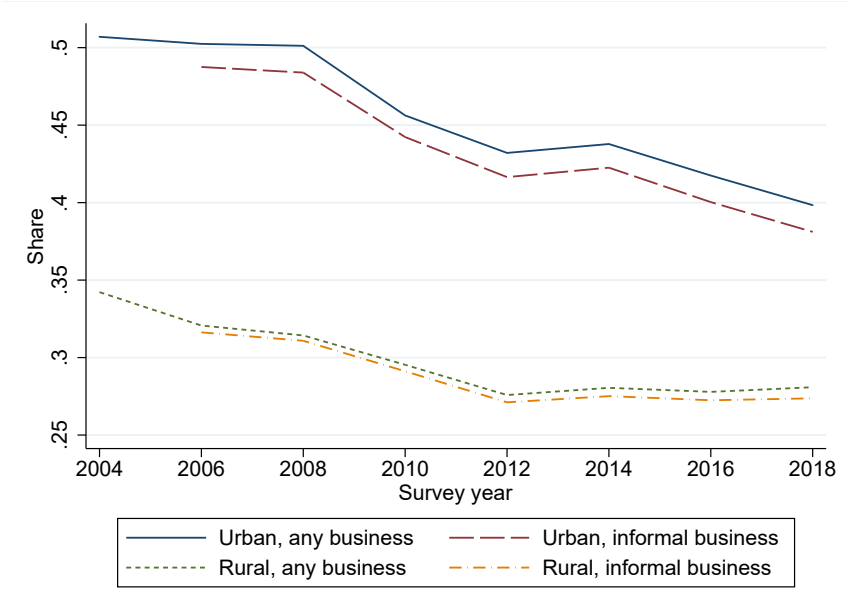


Figure A15: Share of households operating a business by survey, urban vs rural

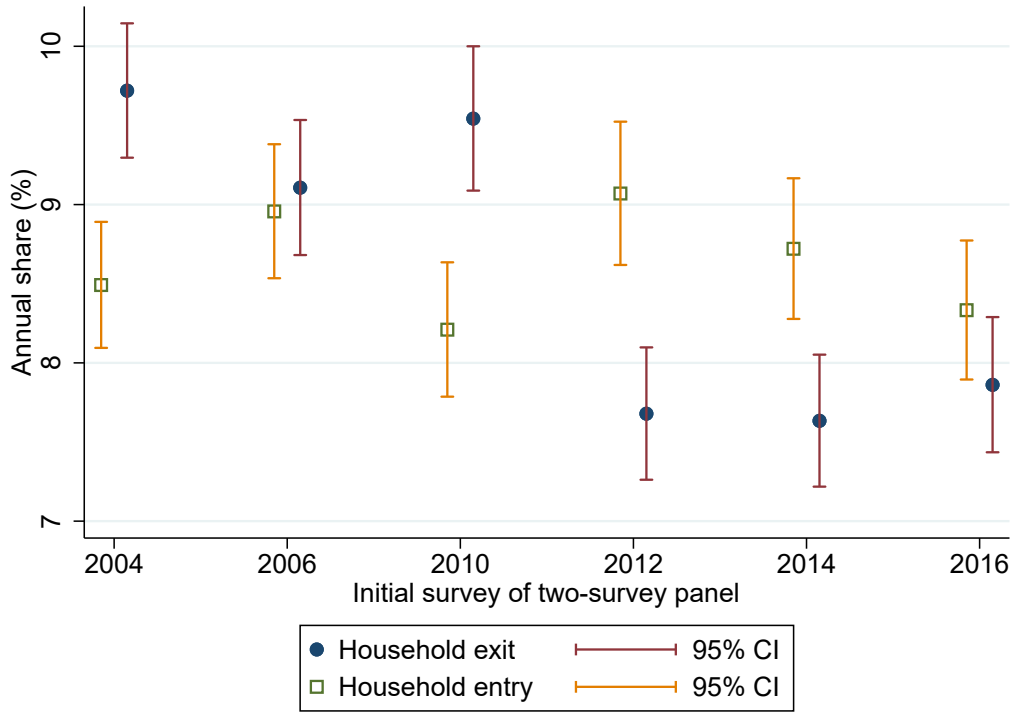


Figure A16: Annual household-level exit and entry rates

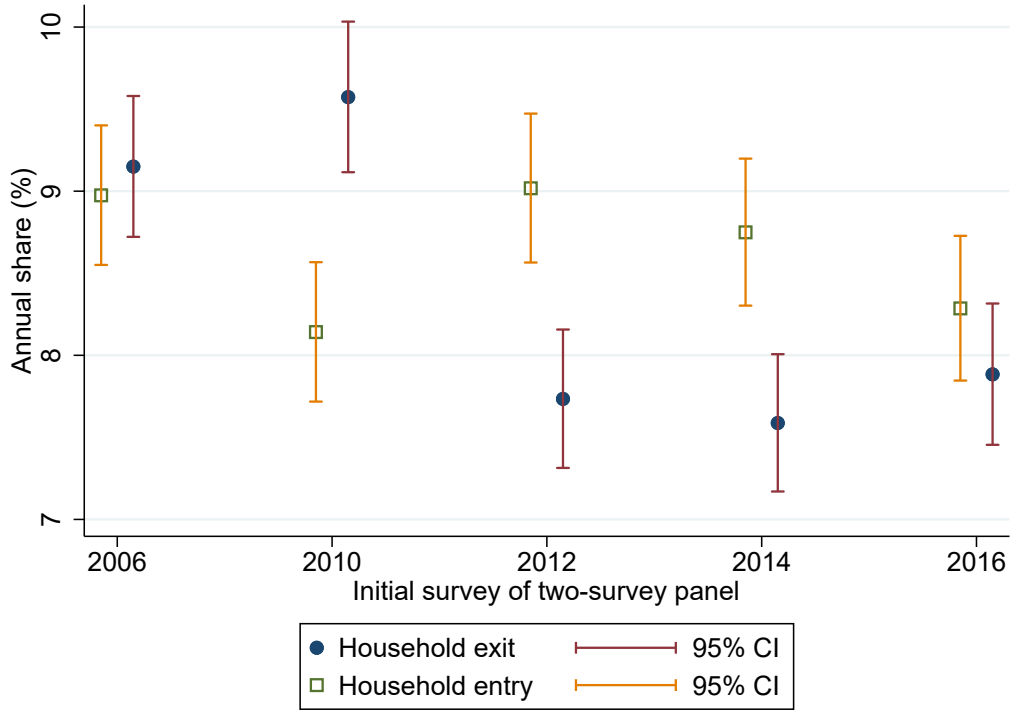


Figure A17: Annual household-level exit and entry rates, informal businesses

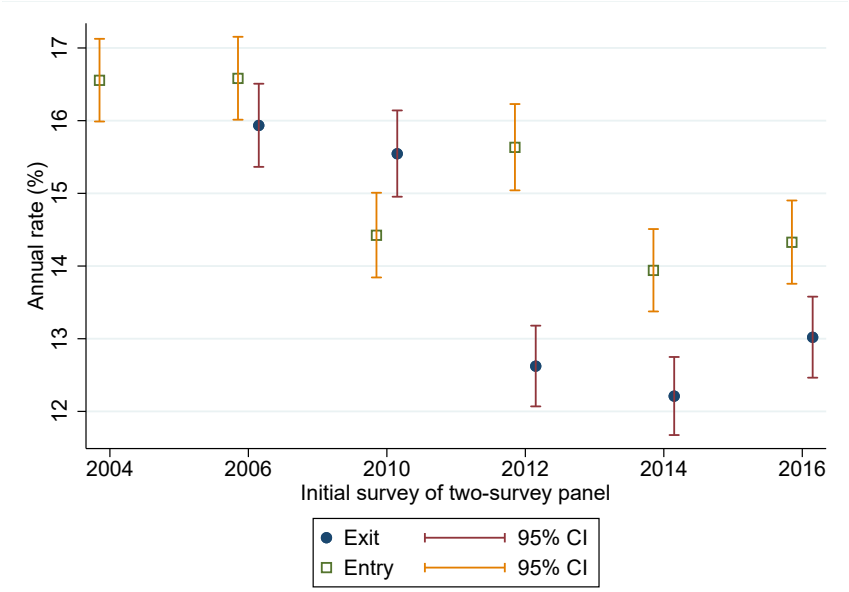


Figure A18: Annual exit and entry rates of informal businesses, excluding owners joining or leaving the household

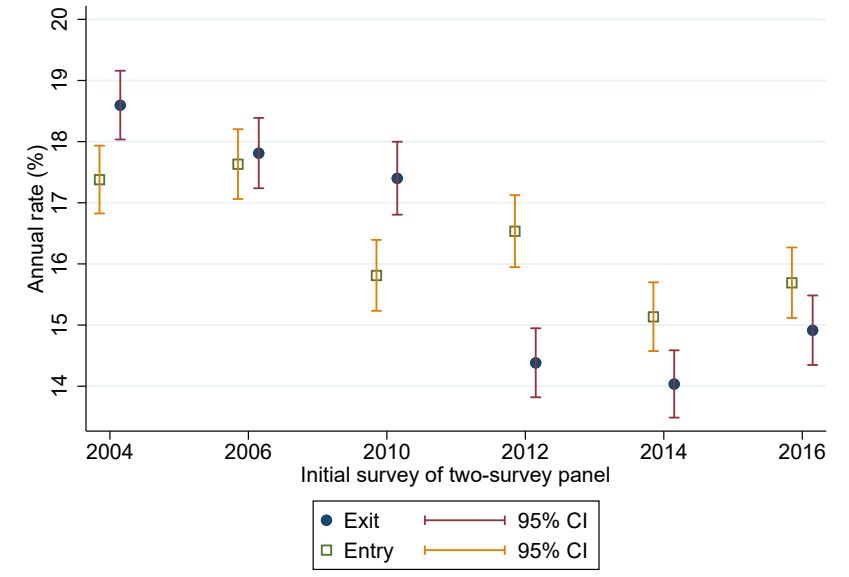


Figure A19: Annual exit and entry rates of all household businesses

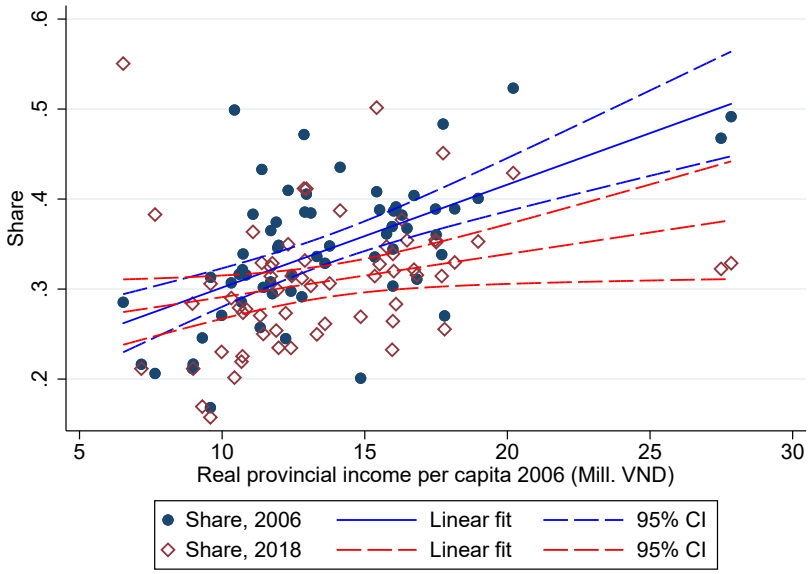
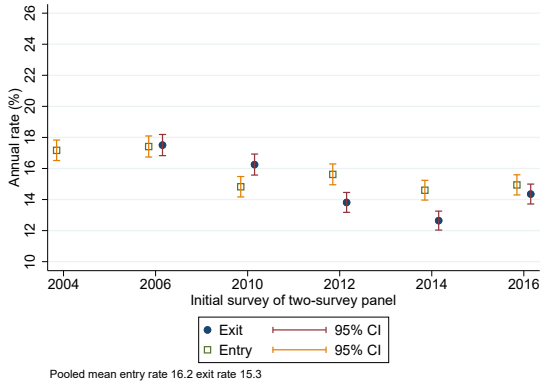
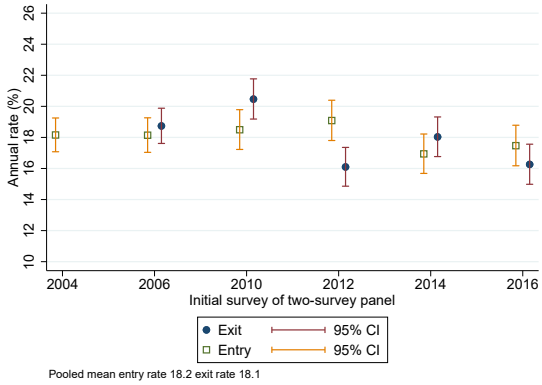


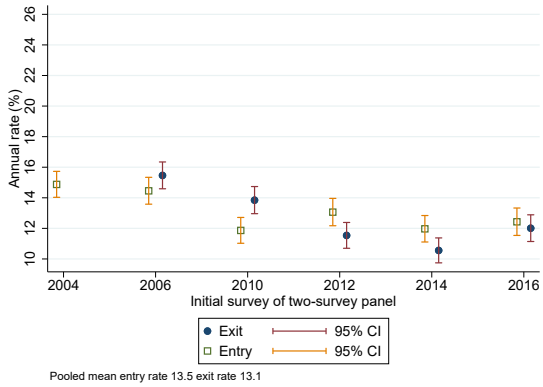
Figure A20: Share of households within a province operating an informal business



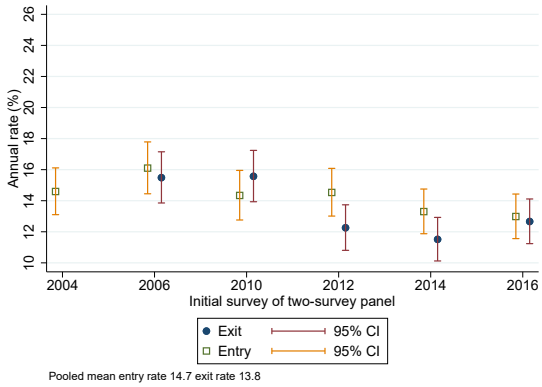
(a) Services



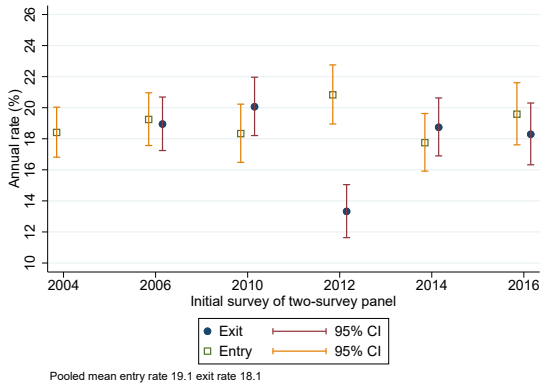
(b) Manufacturing



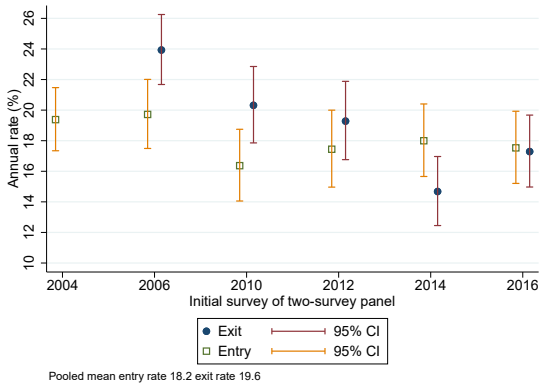
(c) Retail



(d) Hotels and restaurants



(e) Food and beverage



(f) Land transportation

Figure A21: Annual exit and entry rates by industries, informal businesses

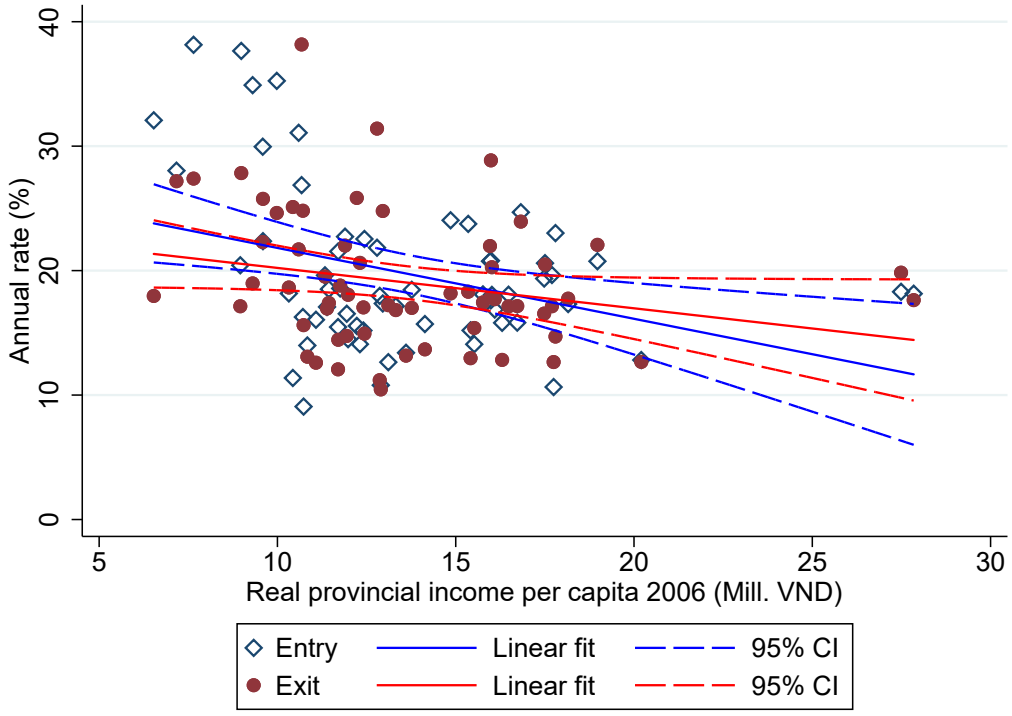


Figure A22: Annual informal entry and exit rates by province with confidence intervals, 2006-08 panel

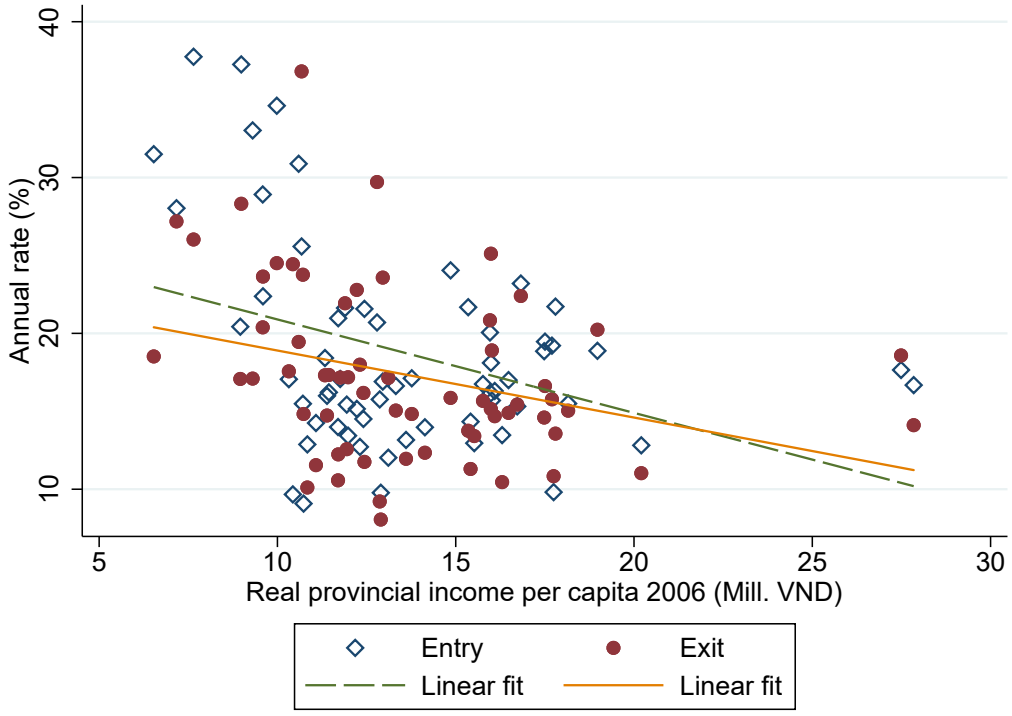


Figure A23: Annual informal entry and exit rates by province, 2006-08 panel, excluding owners joining or leaving the household

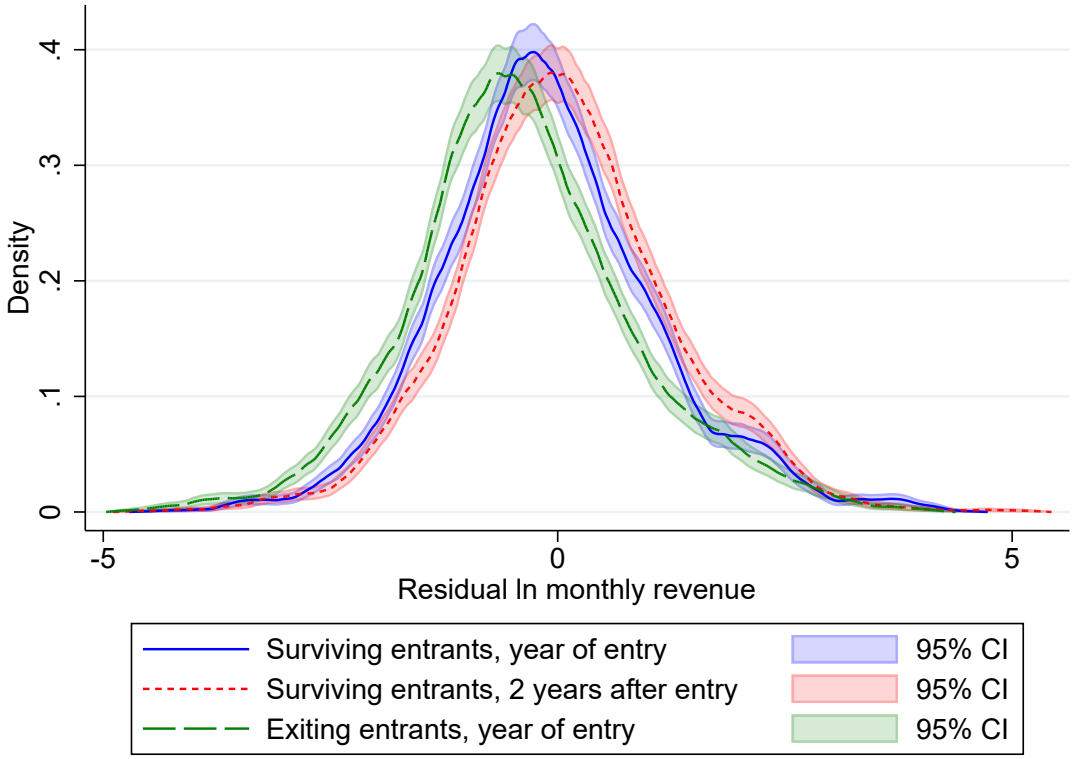


Figure A24: Revenue distributions of surviving and exiting informal entrants with confidence intervals

Note: We removed panel fixed effects from the revenue distribution by regressing ln monthly revenue on a series of indicators for which 3-survey panel the business belongs to. We plot the residuals.

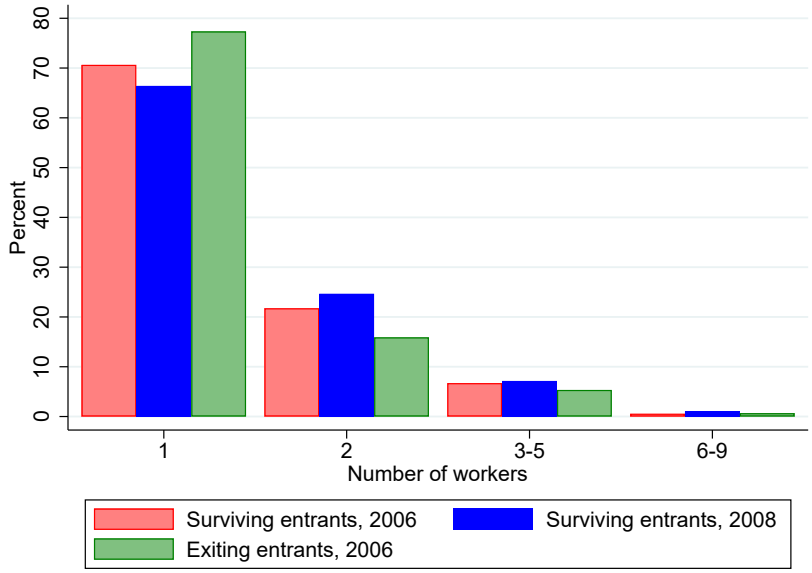


Figure A25: Number of workers in exiting and surviving entrants, 2004-06-08 panel

Note: The sample is all informal businesses that began operations in 2006 within households in the 2004-06-08 panel. A surviving entrant is a business that operated in 2006 and 2008, but not 2004. An exiting entrants is a business that operated in 2006, but not in 2004 or 2008.

Appendix References

Bartelsman, E. J., Haltiwanger, J., and Scarpetta, S. (2009) Measuring and analyzing cross-country differences in firm dynamics, in *Producer Dynamics: New Evidence from Micro Data* (Eds.) T. Dunne, J.B. Jensen, and M.J. Roberts, University of Chicago Press, pp. 15-76.

Cling, J.P., Nguyen, H.C., Razafindrakoto, M. and Roubaud, F. (2010) Urbanization and Access to Labour Market in Vietnam: Weight and Characteristics of the Informal Sector.

Davies, E. and Kerr, A. (2018) Firm Survival and Change in Ghana, 2003–2013, *Journal of African Economies*, **27**, 149–171.

Field, E., Pande, R., Papp, J., and Rigol, N. (2013) Does the classic microfinance model discourage entrepreneurship among the poor? Experimental evidence from India, *American Economic Review*, **103**, 2196–2226.

Haddad, M., de Melo, J., and Horton, B. (1996) Morocco, 1984–1989: Trade Liberalization, Exports, and Industrial Performance, in *Industrial Evolution in developing Countries* (Eds.) M. Roberts and J. Tybout, Oxford University Press.

Klapper, L., and Richmond, C. (2011) Patterns of business creation, survival and growth: Evidence from Africa, *Labour Economics*, **18**, S32–S44.

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

Lay, T. J. (2003) The determinants of and interaction between entry and exit in Taiwan's manufacturing, *Small Business Economics*, **20**, 319–334.

Mead, D. C., and Liedholm, C. (1998) The dynamics of micro and small enterprises in developing countries, *World development*, **26**, 61–74.

Nagler, P. and Naudé, W. (2017) Non-farm entrepreneurship in rural sub-Saharan Africa:

New empirical evidence, *Food Policy*, **67**, 175–191.

Roberts, M. (1996) Colombia 1977–1985: Producer Turnover, Margins, and Trade Exposure, in *Industrial Evolution in developing Countries* (Eds.) M. Roberts and J. Tybout, Oxford University Press, pp. 227-259.

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

Söderbom, M., Teal, F., and Harding, A. (2006) The determinants of survival among African manufacturing firms, *Economic Development and Cultural Change*, **54**, 533–555.

Tybout, J. (1996) Chile, 1979–86: Trade Liberalization and Its Aftermath, in *Industrial Evolution in developing Countries* (Eds.) M. Roberts and J. Tybout, Oxford University Press, pp. 200-225.

Vijverberg, W.P.M. and Haughton, J. (2002) Household enterprises in Vietnam: Survival, Growth, and Living Standards, World Bank Policy Research Working Paper 2773..

Vijverberg, W.P.M., Hoang. T.T.H., Nguyen, C.T., Nguyen N.Q., Nguyen, T.Q., Phung, D.T., and Vu T.K.M. (2006) Non-Farm Household Enterprises in Vietnam A Research Project using Data from VHLSS 2004, VHLSS 2002 and AHBS 2003.

B Data Appendix

In this appendix, we describe six data preparation steps:

1. Verifying the reported business manager
2. Cross checking business and employment information within households
3. Predicting the manager for businesses in surveys in which the manager was not reported
4. Matching the manager's job to the business
5. Verifying the household and individual panels
6. Creating business panels

B.1 Verifying the reported business manager

Both the 2006 and 2008 VHLSSs identified the most knowledgeable household member for each business, hereafter referred to as the owner. We checked whether the owner reported information in the employment module that is consistent with owning the business. In particular, the owner should report being self-employed in the same industry as the business. Both the 2006 and 2008 VHLSSs follow the same structure in the employment module. There is a question identifying whether the individual was self-employed in a non-farm business in any of their jobs during the past 12 months. Additionally, the employment module collected detailed information on the most and second most time consuming job during the past 12 months, including the industry of employment and whether the individual is self-employed or works for others. They also report whether the individual worked a third job during the past 12 months, but no detailed information is provided concerning this job. Across the 2006 and 2008 VHLSSs, we find that the vast majority of reported managers provided consistent information in the labor module (B1). In the 2006 and 2008 surveys, we found 2.8 and 2.8% of businesses for which the reported owner did not provide consistent information in the employment module. The inconsistency could be due to reporting or coding errors in either the business module or the employment module. Table B1 provides an example of

Table B1: Reported Business-owner inconsistencies

Problem	2006	2008
No manager reported	3	1
Manager did not report working	0	105
Manager did not report working in a business	388	281
Manager's job(s) did not match the business	189	191
No inconsistency	19,878	19,887
Total	20,458	20,465

Table B2: Example of incorrect manager

Household member	Primary industry	Primary ownership	Secondary industry	Secondary ownership
1	Clothing	Employee	None reported	None reported
2	Retail	Self-employed	None reported	None reported

where the available evidence suggests that the wrong household member was identified at the owner in the business module. The hypothetical business in this example operates in retail and household member 1 was reported as the manager. However, based on the employment information, it appears as though household member number 2 is more likely to be the manager. Neither household member reports working more than one job in this example, household member 1 reports working in the wrong industry and not in self-employment, whereas household member 2 reports working in the same industry as the business and in self-employment.

Thus, we have changed the manager for only 2.2 and 2.4 percent of businesses in 2006 and 2008 respectively. We find it very reassuring that the vast majority of reported managers provided consistent data in the labor module.

B.2 Cross checking business and employment information within households

More generally, we can compare the employment module data with the business module data to verify the industry of the business and to check for possible missing businesses (i.e., one or

more household member reports being self-employed in a business, but there is no matching business).

B.3 Predicting the manager for businesses in surveys in which the manager was not reported

With the exception of the 2006 and 2008 surveys, the VHLSS business module does not include information on the most knowledgeable person/business manager. Hence, we predicted the manager of each business in the other surveys. This is useful for two reasons. First, part of our analysis requires looking at the relationship between various business outcomes and characteristics of the manager. Second, knowing the manager of the business helps to facilitate the matching of businesses over time. We developed an algorithm for predicting the manager and tested it using the 2006 VHLSS. We correctly predicted the manager for 92.9% of businesses. For each survey, we combine data from the employment and business modules that can be matched. In particular, from the employment module we identify individuals that reported being self-employed in a household business for either their primary or secondary job during the past year. All of these jobs should have a matching business and vice versa.

In Table B3, we provide a summary of the matches by the step within the manager prediction algorithm at which the match was made for the 2004 VHLSS along with the corresponding success rate using the 2006 VHLSS. The table is organized sequentially such that the first step of the algorithm was to identify the manager for businesses in which only one household member reported being self-employed in the industry of the business and then only businesses without a predicted manager would proceed to the next row. The first step of the algorithm matches an individual as the manager for the business for 70.5% of all businesses in the 2004 VHLSS. The corresponding rate of success using the 2006 VHLSS is 98.9%. Thus, for a large share of businesses we have a very high degree of confidence in our predicted manager. Next, we identified a manager for any remaining businesses when there was only one household member that reported being a manager of a business in the same industry in the 2006 VHLSS, taking advantage of the panel information, and so on down

Table B3: 2004 manager prediction results with 2006 success rate

Match type	Number	Success rate
No job matched business	131	0
Businesses matched to a primary or secondary job		
Only job that matched business	15122	.989
Only manager in 06	1595	.755
Only job that matched by year, months and days	232	.912
Only job that matched by months and days	1250	.794
Only job that matched by months	191	.755
Highest number of years in the job	742	.789
Highest number of days worked in the past year in the job	180	.659
Only one of the head or spouse matched	186	.831
Highest number of hours per day in the job	313	.681
Highest ranked individual	968	.703
Only primary job	3	1
Businesses not matched to a primary or secondary job		
Only job that matched business	421	.952
Only manager in 06	33	.667
Highest ranked individual	91	.443

The success rate is based on applying the same algorithm to the 2006 VHLSS.

the rows of the table. In sum, the algorithm correctly identified the manager for 92.9% of businesses in the 2006 VHLSS. Thus, our manager prediction algorithm is doing a very good job of identifying the manager of the business.

The employment modules in the 2010 through 2018 VHLSSs did not ask the same combination of questions related to months, days, and hours worked. As such, minor changes were required for the algorithm. Table B4 reports the number of matches by step for these surveys. As for the 2004 survey, the vast majority of managers are predicted based on there only being one worker in the household with a job that matches the business.

B.4 Matching the manager's job to the business

Each survey collected detailed information on the most time consuming (i.e., primary) and second most time consuming (i.e., secondary) jobs during the past 12 months. The surveys also asked whether the individual worked more than two jobs, but did not collect detailed information on the third job.

Table B4: 2010 to 2018 manager prediction results

Match type	2010	2012	2014	2016	2018
No job matched business	222	207	214	298	312
Businesses matched to a primary or secondary job					
Only job that matched business	13623	12609	12911	12901	12683
Highest number of days worked in the past year	3027	2860	2798	2739	2632
Only one of the head or spouse matched	399	386	368	346	294
Highest number of hours per day in the job	441	415	404	363	373
Highest ranked individual	1267	1197	1113	1106	1138
Only primary job	1	1	2	2	
Businesses not matched to a primary or secondary job					
Only job that matched business	416	260	381	464	464
Highest ranked individual	148	94	123	127	153

We begin by creating a list of all possible jobs, primary, secondary, or other, among managers that could be consistent with managing a business. Next, we loop through the businesses reported by the household, starting with business 1 and ending, if applicable, with business 4. For each business, we identify the best matched job, update the jobs available to be matched to businesses, and then move onto the next business in the household. By looping through businesses one at a time, this ensures that an individual that manages two businesses in the same industry, for example, has one job assigned to each business.

B.5 Verifying the household and individual panels

We previously verified the 2004-06 and 2006-08 panels for McCaig and Pavcnik (2015). For this project we further verified the panels for 2010-12, 2012-14, 2014-16, and 2016-18. We identified inconsistencies in both the household panels (e.g., the same household being matched to two households, a household being matched to a household that does not exist, etc.) and the individual panels (e.g., the same individual being matched to two individuals, an individual being matched to an individual that does not exist, an individual match with inconsistent gender or year of birth information). We then used a combination of algorithms and visual matching to fix the inconsistency where possible or to remove the match. We subsequently searched for missing matches, both at the household level and at the individual level within households. The verified panels, both household and individual, are available

for download on McCaig's website.

B.6 Creating business panels

The surveys contain both household and individual-level panels. The household surveys were not designed to directly follow businesses and thus we use characteristics of the business that should not change for most businesses in order to match them over time. We use the longitudinal dimension of our data at the household and individual level.

Note that we cannot match over time all businesses run by a panel household. For example, any household that reports running a different number of businesses across the two years has experienced net entry or exit of businesses and thus we cannot match at least one business within the household. Thus, for any given household the maximum number of matched businesses is the minimum of the number of businesses run in either year. Table B5 summarizes the number of businesses run by panel households in each of the two-survey panels. Many households experienced net entry or exit. For example in the 2004-06 panel, a total of 1,603 operated no businesses in 2004 but at least one business in 2006. None of these 2006 businesses can be matched to a 2004 business within the household. More generally, the total number of possible matches is equal to the sum over all cells of the number reported times the minimum number of businesses operated at the start or end of the cell.[55]

We start by matching businesses using information on the industry of operation and the manager of the business. We subsequently relax the matching criteria and consider matching the remaining unmatched businesses first by industry (allowing the manager of the business to change over time) and then by manager (allowing the industry of the business to change over time). Table B6 shows the distribution of match types by panel.

[55]For example, in the 2004-06 panel, the maximum possible number of matches is 1*(4222+581+54+7+637+46+7) + 2*(647+75+9+81+10) + 3*(42+3+8) + 4*2 = 7,365.

Table B5: Number of households by number of businesses

		Number of businesses in end year				
Panel	No. in start year	0	1	2	3	4
2004-06	0	0	1466	126	10	1
	1	1643	4222	581	54	7
	2	163	637	647	75	9
	3	13	46	81	42	3
	4	4	7	10	8	2
2006-08	0	0	1428	133	9	0
	1	1466	3911	588	57	5
	2	119	541	639	58	16
	3	8	41	80	42	6
	4	2	5	10	8	3
2010-12	0	0	1232	85	5	0
	1	1404	3832	451	38	4
	2	116	490	528	61	6
	3	5	35	55	26	4
	4	1	6	7	2	3
2012-14	0	0	1303	101	2	0
	1	1124	3862	498	36	4
	2	73	412	494	65	8
	3	2	31	55	24	11
	4	0	2	11	1	0
2014-16	0	0	1253	94	7	0
	1	1108	3835	478	26	1
	2	79	419	556	71	12
	3	5	34	63	44	6
	4	0	3	8	10	5
2016-18	0	0	1163	95	12	1
	1	1109	3724	489	40	2
	2	89	448	528	72	6
	3	4	36	68	48	6
	4	0	2	6	5	5

Table B6: Number of businesses by match type

Match type	2004-06	2006-08	2010-12	2012-14	2014-16	2016-18
Manager and industry	5186	4506	4071	4064	4284	4273
Industry only	564	732	905	935	942	950
Manager only	993	1106	906	906	891	743