Unpacking the negative impact of initial informality on innovation:

The mediating roles of investments in R&D and employee training

Dawit Z. Assefa
davezerihun@gmail.com
Post-doctoral Researcher
University of Messina
Department of Economics
Messina, Italy

Ching T. Liao
ching.liao@neoma-bs.fr
Assistant Professor
NEOMA Business School
1 Rue du Maréchal Juin – BP 215, 76825
Mont-Saint-Aignan Cedex, France

Bisrat A. Misganaw
bisrat.misganaw@neoma-bs.fr
Assistant Professor
NEOMA Business School
1 Rue du Maréchal Juin – BP 215, 76825
Mont-Saint-Aignan Cedex, France
Tel. +33 (0)2 32 82 1707

Declarations of interest: none

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Names of authors are listed alphabetically.

Abstract

This study examines how initial informality negatively affects firm innovation by identifying the underlying mechanisms. We argue that due to persisting resource allocation patterns developed during their informal years, firms with initial informality (i.e. firms started operation without registration) tend to allocate resources to transactions certain to deliver gains and essential to survival. Hence, these firms are less likely to engage in innovation-related investments, specifically research and development (R&D) and employee training. Using data from the World Bank Enterprise Survey, we analyze firms from 30 countries in Sub-Saharan Africa. The empirical evidence supports our hypotheses that investments in R&D and employee training mediate the negative relationship between initial informality and innovation. Theoretical as well as policy and managerial implications are discussed.

Keywords: initial informality, innovation, employee training, R&D investment, mediation, Sub-Saharan Africa

1. Introduction

The general topic of this study is initial informality and innovation at the firm level. It specifically investigates why firms with initial informality — formal firms that started out without registering (Assenova and Sorenson, 2017; Bu and Cuerve-Cazurra, 2020; Mendi and Mudida, 2018) — are less innovative as compared to firms that started out formally from the outset. Informal firms, that is, firms that are unregistered and unregulated but not illegal otherwise (McGahan, 2012; La Porta and Shleifer, 2008), are omnipresent and constitute a large part of the economy in developing countries (Charmes, 2016; Schneider, 2002). In Sub-Saharan African (SSA) countries, such type of firms may account more than 75% of micro, small, and medium enterprises (Nyaware, 2019; Turkson et al., 2020). This high proportion of informal firms could

partially explain the low economic progress in the region because, compared to their formal counterparts, informal firms are less productive (Amin et al., 2019; Benjamin and Mbaye, 2014; Dabla-Norris et al., 2008), mainly due to their low innovativeness (Fu et al., 2018; Ullah et al., 2019). Consequently, scholars call for policy intervention to formalize informal firms through registration (Nordman et al., 2016; Benjamin and Mbaye, 2014), and several countries in SSA have indeed developed such policy, aiming to enhance productivity and attain growth (Amin and Islam, 2015; Ingram et al., 2007). Yet we still know little about how these firms with initial informality behave after formalization.

As rightly argued by Krafft et al. (2020), the registration of informal firms cannot be considered the same as the founding of new firms, as aspects of firms' informal past will extend through to their new life as a formal firm. In fact, prior studies have shown that even after formalizing, firms with initial informality struggle with the innovation process compared to firms that started formally (Costamagna et al., 2019). For instance, studying Kenyan firms, Mendi and Mudida (2018) found that firms with initial informality are less innovative compared to firms that started out formally. Bu and Cuervo-Cazurra (2020) also found similar evidence in 71 countries. Although these studies advanced our understanding of the impact of initial informality on innovation performance, prior research did not investigate why firms with initial informality are less innovative and the mechanisms underlying such a relationship. In this study, we attempt to narrow this gap by identifying and testing two mechanisms that may explain the relationship between initial informality and innovation at the firm level.

Insights from imprinting theory (Marquis and Tilcsik, 2013; Stinchcombe, 1965) suggest that the history of organizations determines some aspects of their present structure and behavior. Hence, we contend that firms with initial informality are likely to avoid investing in highly

uncertain and non-essential transactions because the way they coped with the environment during their informal years could have left imprints. In particular, we argue that imprinted features would constrain such firms' decisions to invest in R&D and employee training, two factors identified as important in determining firm innovation (Abubakar et al., 2019; Gallié and Legros, 2012; van Uden et al., 2017). Thus, we hypothesize that low investments in R&D and employee training mediate the negative relationship between initial informality and innovation.

We test and confirm our hypotheses using a cross-sectional sample of 13,689 firms in 30 SSA countries from the World Bank Enterprise Survey, supplemented with country level indicators from datasets such as World Development Indicator from the World Bank, and Economic Freedom Index from the Heritage Foundation. Our results illustrate the need for firms with initial informality to adjust their resource allocation patterns in order to lessen the negative impact of initial informality. Moving from operating informally to formally, together with increasing investments in R&D and employee training, is likely to enhance their propensity for innovation.

This study contributes to the literature on informality and innovation. First, it demonstrates how initial informality negatively affects firms' propensity to innovate by identifying and testing the mechanisms underlying such a negative relationship. To the best of our knowledge, this is the first study to do so. Second, by showing that firms with initial informality could carry some imprints from their informal years, our study responds to calls for research to analyze informal and formal sectors as a continuum (De Castro et al., 2014; Galdino et al., 2018), rather than two unrelated segments in the economy. The persistence of resource allocation patterns bridges firms' informal and formal stages.

The rest of the paper is structured as follows. The next section reviews the relevant literature and develops hypotheses. We then describe our method and variables, after which we present the results. We conclude by highlighting the findings, discussing the implications and pointing out the limitations, which open up avenues for future studies.

2. Literature and Hypothesis Development

2.1. The resource allocation patterns of firms with initial informality

The founding stage of firms is a sensitive period during which organizations are likely to develop certain characteristics that will have a long-lasting effect (Johnson, 2007; Marquis and Tilcsik, 2013; Stinchcombe, 1965). These characteristics may persist even after the environment has changed (Johnson, 2007). Marquis and Tilcsik (2013) point out that organizational imprinting could originate from factors such as regulative and economic factors; indeed, the imprinting features could manifest in different organizational aspects. In line with this perspective, we suggest that firms that started informally (without registration) may develop certain organizational behaviors during their informal years, and these behaviors may persist even after formalization. In particular, we argue that firms with initial informality are likely to have developed persistent resource allocation patterns guided by short-term orientation.

Compared to their formal counterparts, informal firms often employ atypical resources and management practices (Darbi and Knott, 2016), which may lead them to develop a short-term orientation when it comes to resource allocation. Long-term investments require slack resources and stable sources of resources (Lee and Wu, 2016). Informal firms mostly rely on financing from sources such as relatives, personal connections, and other informal sources (Turkson et al., 2020; Benjamin and Mbaye, 2014), which are often of small amounts, short duration, and occasional. In addition, these firms are not able to access public services and resources as much as their formal

counterparts (Demenet et al., 2016; Benjamin and Mbaye, 2014), nor able to enter into formal contracts to leverage on resources of other firms (Harris, 2014a). Thus, informal firms are less likely to have additional and stable resources to make long-term investments. Moreover, these firms often adopt a flexible employment policy where they can adjust their work force easily since they are not governed by labor regulations (Ingram et al., 2007), which may further reduce their competence in long-term planning. All these may lead informal firms to develop a short-term orientation which guides their resources allocation. Indeed, prior studies confirm that a short-term orientation affects firms' investment decision (Birhanu et al., 2015).

The resource allocation pattern that firms develop while operating informally, we argue, would follow them even after they have formalized. Consequently, firms with initial informality could be less innovative compared to firms having started out formally (Mendi and Mudida, 2018; Bu and Cuervo-Cazurra, 2020). Partly because of their short-term orientation, they may tend to allocate resources to investments that can be realized quickly, while avoiding investments not essential to their survival. On the one hand, engaging in R&D (Kuo et al., 2018) and providing employee training (Borah et al., 2019) are among those investments identified as long-term and non-essential for survival. On the other hand, both types of investments are critical for firms to improve their propensity to innovation (Abubakar et al., 2019; Gallié and Legros, 2012; van Uden et al., 2017). Thus, we argue that firms' initial informality is likely to constrain their R&D and employee training investments, due to the persistence of resource allocation patterns developed during their early stage. This in turn leads to a lower propensity to innovate.

2.2. Initial informality, R&D investment, and innovation

Innovation — both generating groundbreaking technologies and implementing existing ones — drives economic progress (Aghion and Howitt, 2006). Compared to advanced countries,

the nature of innovation is different in developing countries. Since firms in developing countries tend to have insufficient knowledge, a low-skilled labor force, and limited access to credit, groundbreaking innovation is less likely to occur (Zanello, Fu, Mohnen, and Ventresca, 2016). Innovation often occurs as firms adopt existing technology and knowledge (Zanello et al., 2016). Thus, external information is an importance source of innovation. The ability to exploit relevant external information determines a firm's propensity to innovate (Cohen and Levinthal, 1990).

To fully unleash the potential of external knowledge and technology, firms need to have the absorptive capacity to identify, assimilate, and utilize relevant information (Cohen and Levinthal, 1990). In their seminal paper, Cohen and Levinthal (1989) suggest that investing in R&D can develop this capacity. For instance, studying the effect of R&D spillover from foreign to domestic firms in the Indonesian manufacturing sector, Todo and Miyamoto (2006) find that the spillover effect is only significant when local R&D activities are involved. Although foreign direct investment (FDI) brings valuable knowledge and technology, Fu (2008) finds that in the Chinese context, the extent to which FDI promotes innovation depends on local R&D intensity. Goedhuys (2007) also shows that the combination of sourcing external information and conducting internal R&D enhances local Tanzanian firms' innovation propensity. Accordingly, R&D investment is essential for firms to exploit external knowledge and technology.

Nevertheless, we argue that firms with initial informality are less likely to invest in R&D due to the imprinting effect (Johnson, 2007) from their informal years. Informal firms tend to focus on survival and make investments that can generate returns in the short-term (Harris, 2014b). As returns from R&D investments are often long-term and highly uncertain (Kuo et al., 2018), informal firms are less likely to invest on R&D. Indeed, Schipper (2014) indicates that informal firms are likely to develop a risk preference constraining their innovation-related investments.

Even if they manage to introduce innovative products or services, informal firms might be disadvantaged regarding access to distribution channels, as well as other marketing media. Lack of marketing channels and media limits the extent to which firms can profit from innovation (Teece, 1986). This may further increase the uncertainty of generating returns from R&D investments.

Moreover, informal firms operate in a sector where competitors steal innovative products or services without apology (Harris, 2014a). This is partly because these informal firms will not be protected by the Intellectual Property Rights (IPR) regime. In fact, by definition, informal firms "do not have full governance rights as they cannot access formal systems of contractual enforcement of property ownership" (McGahan, 2012, p. 17). Also, to enhance knowledge flow and technology transfer, developing countries have traditionally refrained from having a strong IPR regime (Forero-Pineda, 2006). Although SSA countries have made steady progress in building a strong IPR regime in recent decades, its enforcement remains challenging (Ferguson and Schneider, 2015). Thus, IPR vulnerability due to irregular enforcement means that informal firms still engage in activities ranging from passive imitation to active replication of their competitors (Harris, 2014a).

Because of all these reasons, we contend, informal firms may not be able to develop and build knowledge and competence related to R&D or to navigate the IPR regime. This lack of knowledge and competence may follow them through the imprinting process even after formalization. Given the importance of investing in R&D in building up absorptive capacity and the persistence of resource allocation patterns developed during a firm's early stage, we postulate that reduced R&D investment explains the negative relationship between initial informality and innovation.

H1. R&D investment mediates the negative relationship between initial informality and innovation.

2.3. Initial informality, employee training investment, and innovation

In addition to R&D investment, firms' prior knowledge also decides their absorptive capacity because prior knowledge allows firms to recognize, evaluate, and utilize external knowledge and information (Cohen and Levinthal, 1990). While a firm's absorptive capacity is due to a complex combination of factors, certainly employees' skills and knowledge constitutes a critical component (Cohen and Levinthal, 1990). Through a systematic literature review, Zanello et al. (2016) conclude that lack of education and managerial skills are two important factors that prevent firms in developing countries from adopting a new technology. Empirical evidence from SMEs in Tanzania shows that adequate knowledge facilitates adoption of existing knowledge and technology (Mahemba and De Bruijn, 2003). The presence of skilled workers influences the effect of FDI on regional innovation (Fu, 2008). Beyond utilizing existing knowledge, a higher quality labor force also allows firms to generate their own new products and processes (Collins, 2015).

However, due to the inadequacy of formal education systems in developing countries (Borah et al., 2019), relying on such systems as sources of skilled and knowledgeable labor presents a challenge for firms in these countries (George et al., 2016; van Uden et al., 2017). To mitigate this constraint, firms invest in internal training programs (van Uden et al., 2017). In fact, Oyelaran-Oyeyinka (2004) shows that providing training to employees is a major means by which employers in small African firms improve their employees' skill and knowledge. In addition, employee training improves firm access to external information (Saka-Helmhout et al., 2020) and enhances the possibility of external collaborations (Markovic et al., 2020), which in turn leads to an increased propensity to innovate (Goedhuys, 2007). Thus, providing employee training not only

improves employees' skills (Mano et al., 2012), but is also positively associated with the firm's innovation (Abereijo et al., 2007; van Uden et al., 2017).

Despite this association, informal firms may not perceive the importance of investing in employee training. Since most jobs in the informal sectors are low skilled (Yuki, 2007), employers perceive little need to train their employees. In addition, informal firms do not comply with labor regulation (Schneider and Enste, 2000), which may exacerbate the problem further. Without formal employment contracts, employers in informal sectors enjoy the flexibility to terminate employment; they can easily lay off workers with short notice and little, if any, compensation (Harris, 2014a). For instance, analyzing the worker flow in Brazil and Mexico, Bosch and Maloney (2008) found that compared to workers in formal sectors, informal workers are more likely to leave their jobs during a recession, likely because employers in the informal sector can terminate employment easily.

Limited commitment from employees may also further reduce employer incentive to provide training. For example, in their study of firms in Madagascar, Nordman et al., (2016) found that informal sector workers are the most mobile in the labor market. Interestingly, in a study of informal firms in Kenya, Harris (2014b) showed that, by providing training to their employees, informal firms may also increase their future competitors as the workers themselves may start their own similar business next door. In fact, Harris (2014b) shows that many of the entrepreneurs in the informal sectors in Kenya began as workers of other informal firms. Neither is it uncommon for informal firms to use labor-pooling (Harris, 2014c). This means that one informal firm's training investment could benefit its competitor(s), as these trained employees could use their newly acquired skill in their other job. Hence, informal firms may have less incentive to invest in employee training.

Aside from their employers, employees in the informal sector themselves may not perceive a need for training and are less likely to ask for it. Compared to workers in formal sectors, informal workers tend to be less qualified (La Porta and Shleifer, 2014). This might be because either less qualified workers self-select into the informal sector, or because low-skilled jobs are most prevalent in the informal sector. In the case of self-selection, low skilled workers are probably less interested in skill improvement. When it comes to the prevalence of low-skilled jobs, training might be deemed unnecessary. Furthermore, since employers can terminate employment easily, employees may have little incentive to acquire firm-specific skills through training.

All in all, for reasons deriving from both the employers themselves and their employees, informal firms are less likely to invest in employee training. In keeping with the prediction of imprinting theory, we contend that this behavior persists so that firms with initial informality are less likely to invest in employee training even after they formalize. Thus, we suggest that lack of investment in employee training explains why initial informality reduces firms' innovation.

H2. Investment in employee training mediates the negative relationship between initial informality and innovation.

3. Method

3.1. Data and variables

To test our hypotheses, we utilized data from the World Bank Enterprise Survey (WBES), which covers over 141 countries since its first survey in 2006. The data comprise a representative sample of firms from each country's non-agricultural formal, private economy, excluding public utility firms, government services, healthcare, or financial services. The World Bank employed stratified random sampling by taking into account the firm size (small, medium, and large), business sector

(manufacturing, retail, and other services), and geographic region.[1] In addition, we obtained our country-level variables from the Heritage Foundation[2] and the World Bank World Development Indicators (WDI). After merging the datasets, our final working sample includes over 13,689 firms from 30 countries in the Sub-Saharan African region. Since we are interested in understanding how initial informality affect firms' innovation, all firms in our samples were registered at the time of data collection. Among these firms, 2,516 firms (18.4% of our sample) started operation without registration, and 11,173 firms were registered from the outset. Table 1 depicts the sample distribution by country and year, and table 2 shows the sectors in which the sample firms operate.

"Insert Table 1 about here"

"Insert Table 2 about here"

Dependent variable. In the WBES, firms were asked whether they had introduced new or significantly improved products or services. Furthermore, for firms having introduced new or significantly improved products, they were asked a follow-up question about whether the new or significantly improved product or service was also new for the firm's main market. Consistent with prior literature (Bu and Cuervo-Cazurra, 2020; Mendi and Mudida, 2018), we define *Innovation* as a bivariate variable, coded as 1 if a firm has introduced new or significantly improved products or services that are also new to the firm's market, and as 0 otherwise.

Independent variable. Our independent variable, *Initial Informality*, indicates whether the firm started operating without formal registration. It is constructed from the firm's response in the WBES to the question of whether the establishment was formally registered when it began

[1] Detailed sampling methodology of WBES can be found at http://www.enterprisesurveys.org
[2] The data can be accessed here: https://www.heritage.org/index/explore

operations. Thus, *Initial Informality*, is measured as a dummy variable with a value of 1 if the firm started operating without registration, and with a value of 0 otherwise. This measure has been used in several recent publications in the informality literature (Assenova and Sorenson, 2017; Bu and Cuerve-Cazurra, 2020; Costamagna et al., 2019; Mendi and Mudida, 2018), and hence is well established.

Mediating variables. We have two mediating variables: *R&D Investment* and *Employee Training*. In the WBES, firms were asked whether they spend on formal research and development activities, either in-house or contracted with other companies, excluding market research surveys. Hence, *R&D Investment* is a dummy variable with a value of 1 if the firm invests in formal research and development activities, and a value of 0 otherwise. Firms were also asked if they have formal training for their permanent full-time employees. Thus, *Employee Training* is coded as 1 if the firm has invested in formal training for its permanent full-time employees, and as 0 otherwise. Both measures are consistent with prior research using the same data source (e.g., Abubakar et al., 2019; Bu and Cuervo-Cazurra, 2020).

Control variables. We include control variables in our regression. Firm size (Abubakar et al., 2019; Protogerou et al., 2017; Tavassoli, 2015) and firm age (Kato and Zhou, 2018; Yi et al., 2017) have been found to impact the likelihood of a firm's innovation. Firms with access to financing are more likely to innovate than firms under financial constraint (Abubakar et al., 2019). We also controlled for the top manager's experience in the industry, as it affects a firm's propensity for innovation (Ayyagari et al., 2011; Maksimov et al., 2017). Engagement in exportation or internationalization has been identified as important for the innovation propensity of firms (Abubakar et al., 2019; Protogerou et al., 2017; Tavassoli, 2015). Access to recent technology, and hence innovation propensity, varies between foreign- and domestic-owned firms (Tavassoli, 2015;

Yi et al., 2017). Moreover, innovation varies across ownership structures — *Open Share, Closed Share, Partnership, Limited Partnership, Sole Proprietorship*, and *Others* (Chen et al., 2014). A firm's trust in the quality of the legal system (*Legal System Quality*) is expected to influence innovation, as the ability to profit from innovation requires the protection of intellectual property rights. Furthermore, we include country-level indicators such as a log of GDP per capita and the Heritage Economic Freedom Index to capture technology and institutional disparities across countries. Since difference in sector is also an important factor affecting a firm's likelihood of innovation (Abubakar et al., 2019; Wan et al., 2005; Yi et al., 2017), we controlled for *sector* fixed effect. We used country-fixed effects to control for unobserved country level or omitted factors that could possibly influence innovation in relation to informality. Last but not the least, we controlled for year fixed effect to capture a possible exogenous shock impacting all firms for that specific year. This is critical since our data are cross-sectional across different years. We explain our definitions and measurements of control variables in Table 3.

"Insert Table 3 about here"

3.2. Empirical model

We estimate four models using logistic regression. Equation (1) estimates innovation on independent and control variables without mediators. Equations (2) and (3) estimate the two meditating variables on independent and control variables, respectively. Equation (4) estimates innovation on independent, mediating variables and control variables.

$$Innovation_i = \beta_0 + \beta_1 Initial\ Informality_i + \gamma X_i + \varepsilon_i \quad (1)$$

$$R\&D\ Investment_i = \theta_0 + \theta_1 Initial\ Informality_i + \delta X_i\ + \pi_i \quad (2)$$

$$Employee\ Training_i = \alpha_0 + \alpha_1 Initial\ Informality_i + \mu X_i + \epsilon_i \quad (3)$$

$$Innovation_i = \sigma_0 + \sigma_1 Initial\ Informality_i + \sigma_2 R\&D\ Investment_i + \sigma_3 Employee\ Training_i + \varphi X_i + u_i \quad (4)$$

Where, X_i represents our control variables, ε_i, π_i, ϵ_i and u_i represent error terms for each respective regression.

In order to test our hypotheses, we employed non-linear mediation analysis (Imai et al., 2010a, 2010b; Imai et al., 2011). Since both our dependent and mediating variables are binary, the marginal effects are not constant. The magnitude of the marginal effects varies, depending on the value of the covariates. In such a situation, the conventional calculation of the mediation effects – by taking the product of coefficients as is the case in linear regression – is inappropriate (Buis, 2010; Imai et al., 2010a). Therefore, we undertook this analysis by using the user-written package, *ldecomp*, on STATA (Buis, 2010), to conduct the mediation analysis. In the non-linear regression, the causal mediation effect (or indirect effect) is calculated as:

$$\delta_i(t) = Y_i\big(t, M_i(1)\big) - Y_i\big(t, M_i(0)\big) \quad (5)$$

for each unit of analysis i and treatment status 0 and 1, where δ_i, Y_i, and M_i refer to causal mediation effect, outcome variable (dependent variable), and mediating variable, respectively. The causal mediation effect is the change in the outcome variable as a result of change in the mediator. Thus, it reflects the change from the value realized under no treatment or control condition, $M_i(0)$, to what would be observed under the treatment condition, $M_i(1)$, while the status of treatment would hold constant at t. Furthermore, we also calculate the direct effect of the treatment (often referred as direct effect) as:

$$\varsigma_i(t) = Y_i\big(1, M_i(t)\big) - Y_i\big(0, M_i(t)\big) \quad (6)$$

for each unit of analysis i and treatment status 0 and 1.

4. Findings

4.1. Descriptive statistics and regression results

Table 4 shows the mean, standard deviation, minimum, and maximum of the variables. Among our sample, 4,442 firms (about 32.45%) have introduced new or significantly improved products or services that are also new to the firm's market. 2,516 firms (18.38% of the total sample) started operation without formal registration. About 30.16% and 19.93% of the total sample firms invested in employee training and R&D activities, respectively. Table 5 shows that variables used in our regression analysis strongly correlate with our measure of innovation. The sign of these correlations is consistent with our hypotheses and literature.

"Insert table 4 about here"

"Insert table 5 about here"

Table 6 reports the logistic regression results corresponding to equations (1) to (4). Model 1 shows that *Initial Informality* has a significant negative effect on *Innovation* ($\beta = -0.134, p < 0.05$). This provides empirical support to the negative relationship between initial informality and innovation, consistent with previous literature (Mendi and Mudida, 2018). Models 2 and 3 show that *Initial Informality* has a negative effect on *R&D Investment* ($\beta = -0.214, p < 0.01$) and *Employee Training* ($\beta = -0.185, p < 0.01$). When the two mediator variables are included in Model 4, the significance of our independent variable *Initial Informality* vanishes, whereas both *R&D Investment* and *Employee Training* are positive and significant ($\beta = 1.014, p < 0.01$; $\beta = 0.579, p < 0.01$). Although these coefficients suggest that *R&D Investment* and *Employee Training* indeed play mediating roles, the magnitude of their effects cannot be calculated directly from the coefficients. Buis (2010) and Imai et al. (2010a) demonstrate that the approach of taking the product of regression coefficients as its mediation effect no longer works in non-linear models.

Hence, we carried out non-linear mediation analysis using equations (5) and (6) as suggested by Imai et al. (2010a).

We rely on the user-written package on STATA, *ldecomp*, to conduct the mediation analysis. We carried out simple mediation analysis by taking one mediator at a time. Table 7 depicts the result of the mediation analysis for each mediator. The total effect is the sum of the indirect effect (the causal mediation effect in equation (5)) and the direct effect ($\varsigma_i(t)$ in equation (6)). When *R&D Investment* is the mediator, the total effect is negative and significant ($\beta = -0.169, p < 0.01$). The indirect effect accounts for 42.6% of the total effect ($\beta = -0.072, p < 0.01$). This indicates that *R&D Investment* alone has a significant mediation effect. Hence, Hypothesis 1 is supported. Similarly, when we consider only *Employee Training* as a mediator, the total effect is negative and significant ($\beta = -0.177, p < 0.01$). Out of this, 37.85% is explained by the indirect effect ($\beta = -0.067, p < 0.01$), which confirms the significant mediation effect of *Employee Training*. Hence, Hypothesis 2 is supported.

"Insert table 7 about here"

We also estimate the parallel mediation effect by including both mediators: *R&D Investment* and *Employee Training* at once. Table 8 reports the result of parallel mediation analysis (Panel A reports observed coefficients while Panel B reports observed odd ratios). The coefficient of the total effect is negative and significant ($\beta = -0.197, p < 0.01$). The size of the indirect effect ($\beta = -0.114, p < 0.01$) is about 57.87% of the total effect, indicating a significant mediation effect by *R&D Investment* and *Employee Training* together. This result is consistent with the simple mediation effect analysis.

"Insert table 8 about here"

4.2. Robustness analysis

We conducted robustness checks to ensure that our results hold for alternative specifications and estimation technique. First, we conducted a sensitivity analysis to check whether the Sequential Ignorability (SI) assumption, which is needed for identification in mediation sensitivity analysis models (Imai et al., 2010a, 2010b; Tingley et al., 2014), is violated or not. The analysis shows to what degree this key identifying assumption must be violated for the conclusions obtained in the mediation analysis to be reversed. It estimates a sensitivity parameter ρ—the correlation between the error terms of the outcome and the mediation models. The mediation model estimates the mediator variable on independent and control variables. Similarly, the outcome model estimates the dependent variable on independent, mediation, and control variables. The SI assumption is said to be violated when ρ is large, which implies the existence of a strong confounding effect between mediating and dependent variables (Tingley et al., 2014; Vaznyte et al., 2020). We utilized a *medsens* package from STATA (Hicks and Tingley, 2011) to carry out the sensitivity analysis. Figure 1 plots the estimated Average Causal Mediation Effect (ACME) of the *R&D Investment* mediator against different values of sensitivity parameter ρ. As it can be seen, the ACME equals zero when ρ is 0.3, meaning the negative mediation effect of *R&D Investment* will be reversed only if ρ is greater than 0.3. This indicates a strong degree of robustness. Similarly, Figure 2 plots the estimated ACME of the *Employee Training* mediator against different values of sensitivity parameter ρ. The ACME equals zero when ρ is 0.2. The negative mediation effect of *Employee Training* will be reversed only if ρ is greater than 0.2, which again indicates a strong degree of robustness.

"Insert figure 1 about here"

"Insert figure 2 about here"

Second, since our independent variable (*Initial Informality*) was determined at some point in the past, there is no simultaneity and reverse causality concern between our dependent and independent variables. However, there may be a potential concern of endogeneity in our results because there may be a feedback effect from our mediators, *R&D Investment* and *Employee Training*, on our dependent variable, *Innovation*. It is possible that less innovative firms are less likely to invest in R&D and employee training. To address this concern, we adopted two-stage least square instrumental variable regression (2SLS IV), based on the suggestion of Angrist and Pischke (2008) and Wooldridge (2010). Angrist and Pischke (2008) demonstrate that 2SLS (two-stage least square) can still capture the average treatment effect with endogenous regressors, regardless of whether the dependent variable is binary, non-negative, or continuous. Wooldridge (2010: see p. 903–953) shows the possibility of using the fitted probabilities of a potential endogenous binary variable as an instrumental variable.

Thus, we used the fitted probabilities of our potential endogenous binary variable, *R&D Investment*, as an instrumental variable and named it as *R&D Investment 2*. We estimated *R&D Investment 2* using *probit* model with the same set of control variables included in the main analysis and one additional variable—*Corporate Tax Rate* from World Bank World Development Indicators. The variable of *Corporate Tax Rate* is measured as the total corporate tax as a percentage of commercial profits. An ideal instrument for our case would have been an R&D tax credit; however, we could not find such data for our sample. Thus, we used the corporate tax rate as a proxy. We then applied 2SLS IV regressions using *R&D Investment 2* as the instrumental variable of *R&D Investment*. Since our goal was to detect the potential reverse causality between

our mediators and dependent variable, we did not include the variable of Initial Informality in the regression. The validity of our instrument is affirmed, as the F-statistics of exclusion restriction of the instrument from the first stage and the Eigen value meet the minimum requirement.

Following a similar procedure, we used the fitted probability of *Employee Training*, another potential endogenous binary variable, as an instrumental variable, and named it *Employee Training 2*. To estimate *Employee Training 2*, we use *probit* model, including the same set of control variables as in the main analysis, with two additional variables. We added the firm's investment in fixed assets and the average human capital index of the country from Penn World Tables database (Feenstra et al., 2015). Then, we applied the 2SLS IV regressions. The F-statistics of exclusion restriction of the instrument from the first stage and the Eigen value meet the minimum requirement affirming the validity of our instrument. Table 9 presents the second stage results of 2SLS IV regression, which are consistent with our main results.

"Insert Table 9 about here"

Finally, although our regression includes a range of control variables identified from the literature, unobservable variables may confound with our independent variable, *Initial Informality*. To evaluate this concern of unobservable heterogeneity, we followed recent studies (Altonji et al., 2005; Assenova and Sorenson, 2017) and undertook ratio tests as $\rho = |\hat{\beta}^F|/|(\hat{\beta}^R - \hat{\beta}^F)|$, where $\hat{\beta}^F$ denotes the estimated coefficient of *Initial Informality* of the unrestricted (full) model and $\hat{\beta}^R$ denotes the estimated coefficient of *Initial Informality* in a restricted model (with one less regressor). The restricted model is nested in the full model. We undertook this ratio test on model 1— the model that estimates *Innovation* on *Initial Informality* and control variables. The ratio shows "how much larger would selection on unobserved variables need to be, relative to selection

on observed factors, to account for the results" (Assenova and Sorenson, 2017, p. 811). For an estimate to be robust, Altonji et al. (2005) suggest that the ratio test (i.e., the value of ρ) should be greater than 1. The value of ρ in our analysis starts from 3.22. This confirms that our results are not driven by unobserved heterogeneity.

5. Discussion and Conclusions

This study investigates how initial informality negatively affects firm innovation by identifying and testing the mechanisms underlying such a relationship. Drawing insights from imprinting theory (Stinchcombe, 1965; Marquis and Tilcsik, 2013), we argue that informal firms are likely to develop a certain resource allocation pattern that would persist even after formalizing. Considering the importance of absorptive capacity for firms to innovate, we hypothesize and confirm that firms with initial informality are less likely to invest in R&D and employee training, which in turn reduces their propensity to innovate. Because of the persistence of resource allocation patterns developed during firms' informal years, formalization through registration may not guarantee changes in firm behavior. Only when firms adjust their resource allocation patterns to reflect the change in their formal status can they lessen the negative effects of initial informality, and then reap fully the benefits of formalization.

Our contribution is twofold. First, we contribute to the literature on informality and innovation by identifying and testing two mechanisms through which initial informality affects firms' innovation. The two mechanisms — investment in R&D and employee training — have been identified as crucial factors determining firm level innovation (van Uden et al., 2017; Njiraini et al., 2018). We found that these two factors not only affect firm level innovation, but they also mediate the negative impact of initial informality on firm innovation. Our estimations specifically

show that R&D investments and employee training explain about 42.6% and 37.85% of the impacts of initial informality on innovation, respectively; and their combination can count up to 57.87 % of the effect. This demonstrates that firms with initial informality need to increase investments in R&D and employee training if they want to overcome the negative effect of initial informality on innovation. Second, we contribute to the literature on organizational imprinting by applying the theory to explain the negative relationship between initial informality and innovation. Informal firms tend to be short-term oriented in their resource allocation. They would avoid investments that are not essential to their survival. We argue that the resource allocation pattern that these firms developed while operating informally would be imprinted and persist even after formalization. Our analysis confirms that firms with initial informality are less likely to allocate their resources to long term investments such as R&D and employee training, and hence they are less innovative.

Outlining reduced investments in R&D and employee training as the mechanisms not only advances our understanding regarding the relationship between initial informality and innovation, but also has policy implications. Aiming to encourage informal firms to transform to formal, several SSA countries have introduced laws and regulations ranging from implementing penalty and incentive programs to simplifying the registration process (Amin and Islam, 2015; Ingram et al., 2007). Although these policy actions are encouraging, and may help to transform some of the firms operating informally into formal firms, our results suggest that these firms may not be as innovative as those that started formally from the outset. Thus, only easing the registration process or related costs may not be enough to foster innovation. Implementing additional policy instruments to help initially informal firms change their resource allocation patterns and invest in R&D and employee training would increase their innovativeness and promote economic progress.

Our findings are also managerially relevant in pointing out how entrepreneurs' early decisions in starting a business without registration may affect future development of their ventures. Although starting a business without registration may make it easy for entrepreneurs to set up firms with limited resources (Webb et al., 2014), initial informality may leave firms with some characteristics that affect their behavior after formalization. In particular, compared to firms without initial informality, firms with initial informality are less likely to invest in R&D and employee training, which in turn leads to reduced innovation performance. Thus, we suggest potential entrepreneurs take into consideration the impact of initial informality while evaluating the benefits of starting a business unregistered. The optimal resource allocation patterns at the initial informal stage might not be suitable once the firms become formal. Since such patterns persist, it would be costly for firms with initial informality to adjust. Overlooking this cost may impede entrepreneurs in making informed decisions.

Our study has its limitations, which opens up avenues for future studies. First, we focus on product and service innovation, i.e., we measure innovation using the introduction of new products and services. However, the determinants of other types of innovation might not necessarily be the same as product and service innovation (Reichstein and Salter, 2006). Thus, future studies may investigate whether initial informality impacts other types of innovation, such as process and organizational innovation, and identify some of the underlying mechanisms. Second, we acknowledge that the two mediators we have identified in this study might not be the only ones. Future research may explore other potential mechanisms linking initial informality with innovation. Furthermore, future research may also investigate whether and how various forms of employee training may have a different effect in mediating the relation between initial informality and innovation. In this study we only looked at the impact of investment on employee training in

general. Third, although the WBES dataset enabled us to cover a large number of observations from several SSA countries, which enhances the generalizability of our study, our study may suffer from the cross-sectional nature of the dataset. By using panel data collected pre and post formalization, future research may compare innovation, R&D, and employee training investments before and after firms formalize.

All in all, our study unpacks the negative relationship between initial informality and innovation by identifying and testing its underlying mechanisms. Covering a large number of firms across 30 countries in Sub-Saharan Africa, we supplement existing literature by focusing on an understudied region where informal firms are predominant. Besides contributing to the informality literature, the findings of our study are relevant for policymakers in refining their instrumental design, and for entrepreneurs in evaluating the consequences of starting a business unregistered.

References

Abereijo, I. O., Llori, M. O., Taiwo, K. A., & Adegbite, S. A. (2007). Assessment of the capabilities for innovation by small and medium industry in Nigeria. *African Journal of Business Management, 1*(8), 209-217.

Abubakar, Y. A., Hand, C., Smallbone, D., & Saridakis, G. (2019). What specific modes of internationalization influence SME innovation in Sub-Saharan least developed countries (LDCs)? *Technovation*, *79*, 56-70.

Aghion, P., & Howitt, P. (2006). Appropriate growth policy: A unifying framework. *Journal of the European Economic Association*, *4*(2-3), 269-314.

Altonji, J.G., Elder, T.E. & Taber, C.R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1),151-184.

Amin, M., & Islam, A. (2015). Are large informal firms more productive than the small informal firms? Evidence from firm-level surveys in Africa. *World Development*, *74*, 374–385.

Amin, M., Ohnsorge, F.L., & Okou, C. (2019). Casting a shadow: productivity of formal firms and informality. Policy Research Working Paper 8945, The World Bank, Washington.

Angrist, J.D. & Pischke, J.S., (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Assenova, V. A., & Sorenson, O. (2017). Legitimacy and the benefits of firm formalization. *Organization Science*, *28*(5), 804-818.

Ayyagari, M., Demirgüç-Kunt, A., & Maksimovic, V. (2011). Firm innovation in emerging markets: The role of finance, governance, and competition. *Journal of Financial and Quantitative Analysis*, *46*(6), 1545-1580.

Benjamin, N.C., Mbaye, A.A. (2014). Informality, growth, and development in Africa. In: UNU-WIDER Working Paper 2014–52. UNU-WIDER, Helsinki (Finland).

Birhanu, A. G., Gambardella, A., & Valentini, G. (2015). Bribery and investment: Firm-level evidence from Africa and Latin America. *Strategic Management Journal 37*(9), 1865-1877.

Borah, D., Malik, K., & Massini, S. (2019). Are engineering graduates ready for R&D jobs in emerging countries? Teaching-focused industry-academia collaboration strategies. *Research Policy, 48*(9). 103837.

Bosch, M., & Maloney, W. F. (2008). *Cyclical movements in unemployment and informality in developing countries.* Policy Research Working Paper Series 4648: World Bank.

Bu, J., & Cuervo-Cazurra, A. (2020). Informality costs: informal entrepreneurship and innovation in emerging economies. *Strategic Entrepreneurship Journal*, *14*(3), 329-368. https://doi.org/10.1111/1744-1633.12020.

Buis, M. L. (2010). Direct and indirect effects in a logit model. *The Stata Journal*, *10*(1), 11-29.

Charmes, J. (2016). The Informal Economy: Definitions, Size, Contribution and Main Characteristics, in Kraemer-Mbula, E and Wunsch-Vincent, S. (Eds.), *The Informal Economy in Developing Nations*, *Hidden engine for innovation*? pp. 13-44. Cambridge University Press, Cambridge.

Chen, V. Z., Li, J., Shapiro, D. M., & Zhang, X. (2014). Ownership structure and innovation: An emerging market perspective. *Asia Pacific Journal of Management*, *31*(1), 1-24.

Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: The two faces of R&D. *The Economic Journal*, *99*(September), 569–596.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity : A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Collins, T. (2015). Imitation: A catalyst for innovation and endogenous growth. *Economic*

Modelling, *51*, 299–307. https://doi.org/10.1016/j.econmod.2015.08.023.

Costamagna, R., Carlier, S., & Mendi, P. (2019). Initial informality as an obstacle to intellectual capital acquisitions: Empirical evidence from Latin America. *Journal of Intellectual Capital*, *20*(4), 472–487. https://doi.org/10.1108/JIC-12-2018-0218.

Dabla-Norris, E., Gradstein, M. & Inchauste, G. (2008), What Causes Firms to Hide Output? The Determinants of Informality, *Journal of Development Economics*, *85*(1-2), 1-27.

Darbi, W. P. K., & Knott, P. (2016). Strategising practices in an informal economy setting: A case of strategic networking. *European Management Journal*, *34*(4), 400-413.

De Castro, J. O., Khavul, S., & Bruton, G. (2014). Shades of grey: How do informal firms navigate between macro and meso institutional environments? *Strategic Entrepreneurship Journal*, *8*(1), 75–94. https://doi.org/10.1002/sej.

Demenet, A., Razafindrakoto, M., & Roubaud, F. (2016). Do informal businesses gain from registration and how? Panel data evidence from Vietnam. *World Development*, *84*, 326–341.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, *105*(10), 3150-82.

Ferguson, V., & Schneider, M. (2015). Enforcement of intellectual property rights in Africa. *Journal of Intellectual Property Law & Practice*, *10*(4), 269-279.

Forero-Pineda, C. (2006). The impact of stronger intellectual property rights on science and technology in developing countries. *Research Policy*, *35*(6), 808-824.

Fu, X. (2008). Foreign direct investment, absorptive capacity and regional innovation capabilities: Evidence from China. *Oxford Development Studies*, *36*(1), 89–110. https://doi.org/10.1080/13600810701848193.

Fu, X., Mohnen, P., & Zanello, G. (2018). Innovation and productivity in formal and informal firms in Ghana. *Technological Forecasting and Social Change, 131*, 315-325.

Galdino, K. M., Kiggundu, M. N., Jones, C. D., & Ro, S. (2018). The informal economy in pan-Africa: Review of the literature, themes, questions, and directions for management research. *Africa Journal of Management*, *4*(3), 225–258.

Gallié, E. P., & Legros, D. (2012). Firms' human capital, R&D and innovation: a study on French firms. *Empirical Economics*, *43*(2), 581-596.

George, G., Corbishley, C., Khayesi, J. N., Haas, M. R., & Tihanyi, L. (2016). Bringing Africa in: promising directions for management research. *Academy of Management Journal, 59*(2),

377-393.

Goedhuys, M. (2007). Learning, product innovation, and firm heterogeneity in developing countries: Evidence from Tanzania. *Industrial and Corporate Change*, *16*(2), 269–292. https://doi.org/10.1093/icc/dtm003.

Harris, J. (2014a). The messy reality of agglomeration economies in urban informality: Evidence from Nairobi's handicraft industry. *World Development*, *61*, 102-113.

Harris, J. (2014b). Meeting the challenges of the handicraft industry in Africa: Evidence from Nairobi. *Development in Practice*, *24*(1), 105-117.

Harris, J. (2014c). The confounding influence of urban informality on innovation and production specialisation in production clusters: evidence from Nairobi. *African Journal of Science, Technology, Innovation and Development*, *6*(6), 529-539.

Hicks, R. & Tingley, D., (2011). Causal mediation analysis. *The Stata Journal*, *11*(4), 605-619.

Imai, K., Keele, L., & Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309–334. https://doi.org/10.1037/a0020761.

Imai, K., Keele, L., & Yamamoto, T. (2010b). Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science, 25*(1), 51–71. https://doi.org/10.1214/10-STS321.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box: learning about causal mechanisms from experimental and observational studies. *American Political Science Review, 105*(4), 765–789. https://doi.org/10.1017/S0003055411000414.

Ingram, M., Ramachandran, V., & Desai, V. (2007). Why do firms choose to be informal? Evidence from enterprise surveys in Africa. Manuscript. *The World Bank.*

Johnson, V. (2007). What is organizational imprinting? Cultural entrepreneurship in the founding of the Paris opera. *American Journal of Sociology*, *113*(1), 97–127.

Kato, M., & Zhou, H. (2018). Numerical labor flexibility and innovation outcomes of start-up firms: A panel data analysis. *Technovation*, *69*, 15-27.

Krafft, C., Assaad, R., Rahman, K. W., & Cumanzala, M. (2020). How Do Small Formal and Informal Firms in the Arab Republic of Egypt Compare? Policy Research Working Paper 9423. Equitable Growth, Finance and Institutions Practice Group, World Bank Group.

Kuo, H. C., Wang, L. H., & Yeh, L. J. (2018). The role of education of directors in influencing firm R&D investment. *Asia Pacific Management Review, 23*(2), 108-120.

La Porta, R., & Shleifer, A. (2008). The unofficial economy and economic development. Brookings Papers on Economic Activity, 2, 275–352.

La Porta, R., & Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives*, *28*(3), 109-26.

Lee, C. L., & Wu, H. C. (2016). How do slack resources affect the relationship between R&D expenditures and firm performance?. *R&D Management, 46*(S3), 958-978.

Mahemba, C. M., & De Bruijn, E. J. (2003). Innovation activities by small and medium-sized manufacturing enterprises in Tanzania. *Creativity and Innovation Management*, *12*(3), 162–173. https://doi.org/10.1111/1467-8691.00279.

Maksimov, V., Wang, S. L., & Luo, Y. (2017). Institutional imprinting, entrepreneurial agency, and private firm innovation in transition economies. *Journal of World Business*, *52*(6), 854-865.

Mano, Y., Iddrisu, A., Yoshino, Y., & Sonobe, T. (2012). How Can Micro and Small Enterprises in Sub-Saharan Africa Become More Productive? The Impacts of Experimental Basic Managerial Training. *World Development*, *40*(3), 458–468.

Markovic, S., Bagherzadeh, M., Dubiel, A., Cheng, J., & Vanhaverbeke, W. (2020). Do not miss the boat to outside-in open innovation: Enable your employees. *Industrial Marketing Management*, *91*, 152-161.

Marquis, C., & Tilcsik, A. (2013). Imprinting: Toward a Multilevel Theory. *Academy of Management Annals*, *7*(1), 195–245. https://doi.org/10.5465/19416520.2013.766076.

McGahan, A. M. (2012). Challenges of the informal economy for the field of management. *Academy of Management Perspectives*, *26*(3), 12-21.

Mendi, P., & Mudida, R. (2018). The effect on innovation of beginning informal: Empirical evidence from Kenya. *Technological Forecasting and Social Change*, *131*, 326–335.

Njiraini, P., Gachanja, P., & Omolo, J. (2018). Factors influencing micro and small enterprise's decision to innovate in Kenya. *Journal of Global Entrepreneurship Research*, *8*(1), 1-9.

Nordman, C. J., Rakotomanana, F., & Roubaud, F. (2016). Informal versus formal: A panel data analysis of earnings gaps in Madagascar. *World Development*, *86*, 1-17.

Nyaware, B. (2019). Technology Acquisition and Innovations in Kenya's Informal Sector. KIPPRA Discussion Paper No. 225.

Oyelaran-Oyeyinka, B. (2004). Learning and local knowledge institutions in African industry.

UNU-INTECH, Maastricht. DP 2004-2.

Protogerou, A., Caloghirou, Y., & Vonortas, N. S. (2017). Determinants of young firms' innovative performance: Empirical evidence from Europe. *Research Policy*, *46*(7), 1312-1326.

Reichstein, T., & Salter, A. (2006). Investigating the sources of process innovation among UK manufacturing firms. *Industrial and Corporate Change*, *15*(4), 653–682. https://doi.org/10.1093/icc/dtl014.

Saka-Helmhout, A., Chappin, M., & Vermeulen, P. (2020). Multiple paths to firm innovation in Sub-Saharan Africa: How informal institutions matter. *Organization Studies*, *41*(11), 1551-1575.

Schipper, T., 2014. Informality, Innovation, and Aggregate Productivity Growth. MPRA Paper 69647. University Library of Munich, Germany.

Schneider, F. (2002). Size and measurement of the informal economy in 110 countries around the world, Paper presented at a Workshop of Australian National Tax Centre, ANU, Canberra.

Schneider, F., and Enste, D. H. (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, *38*(1), 77-114.

Stinchcombe, A. L. (1965). Social structure and organizations. In J. G. March (Ed.), *Handbook of Organizations* (pp. 142–193). Chicago: Rand McNally.

Tavassoli, S. (2015). Innovation determinants over industry life cycle. *Technological Forecasting and Social Change*, *91*, 18-32.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, *15*(6), 285–305.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1-38.

Todo, Y., & Miyamoto, K. (2006). Knowledge spillovers from foreign direct investment and the role of local R&D activities: Evidence from Indonesia. *Economic Development and Cultural Change*, *55*(1), 173–200. https://doi.org/10.1086/505729.

Turkson, F. E., Amissah, E., & Gyeke-Dako, A. (2020). The role of formal and informal finance in the informal sector in Ghana. *Journal of Small Business & Entrepreneurship*, 1-24. DOI: 10.1080/08276331.2020.1724002.

Ullah, S., Williams, C. C., & Arif, B. W. (2019). The impacts of informality on enterprise

innovation, survival and performance: Some evidence from Pakistan. *Journal of Developmental Entrepreneurship, 24*(3), 1-19.

van Uden, A., Knoben, J., & Vermeulen, P. (2017). Human capital and innovation in Sub-Saharan countries: a firm-level study. *Innovation*, *19*(2), 103-124.

Vaznyte, E., Andries, P., & Demeulemeester, S. (2020). "Don't leave me this way!" Drivers of parental hostility and employee spin-offs' performance. *Small Business Economics*, 1-29.

Wan, D., Ong, C. H., & Lee, F. (2005). Determinants of firm innovation in Singapore. *Technovation*, *25*(3), 261-268.

Webb, J. W., Ireland, R. D., & Ketchen, D. J. (2014). Toward a greater understanding of entrepreneurship and strategy in the informal economy. *Strategic Entrepreneurship Journal*, *8*, 1–15. https://doi.org/10.1002/sej.

Wooldridge, J.M. (2010). Econometric analysis of cross section and panel data MIT press. *Cambridge, MA*.

Yi, J., Hong, J., chung Hsu, W., & Wang, C. (2017). The role of state ownership and institutions in the innovation performance of emerging market enterprises: Evidence from China. *Technovation*, *62*, 4-13.

Yuki, K. (2007). Urbanization, informal sector, and development. *Journal of Development Economics*, *84*(1), 76-103.

Zanello, G., Fu, X., Mohnen, P., & Ventresca, M. (2016). The creation and diffusion of innovation in developing countries: A systematic literature review. *Journal of Economic Surveys*, *30*(5), 884–912. https://doi.org/10.1111/joes.12126.

Table 1: Working sample by country and year

	Year of survey							
Country name	2011	2013	2014	2015	2016	2017	2018	Total
Benin	0	0	0	0	141	0	0	141
Burundi	0	0	156	0	0	0	0	156
Cameroon	0	0	0	0	309	0	0	309
Central African Republic	146	0	0	0	0	0	0	146
Chad	0	0	0	0	0	0	144	144
Côte d'Ivoire	0	0	0	0	310	0	0	310
Democratic Republic of Congo	0	482	0	0	0	0	0	482
Eswatini (Swaziland)	0	0	0	0	113	0	0	113
Ethiopia	515	0	0	807	0	0	0	1,322
Gambia	0	0	0	0	0	0	136	136
Ghana	0	681	0	0	0	0	0	681
Guinea	0	0	0	0	130	0	0	130
Kenya	0	690	0	0	0	0	957	1,647
Lesotho	0	0	0	0	125	0	0	125
Liberia	0	0	0	0	0	147	0	147
Malawi	0	0	403	0	0	0	0	403
Mali	0	0	0	0	150	0	0	150
Mauritania	0	0	137	0	0	0	0	137
Mozambique	0	0	0	0	0	0	588	588
Namibia	0	0	461	0	0	0	0	461
Niger	0	0	0	0	0	128	0	128
Nigeria	0	0	1,860	0	0	0	0	1,860
Rwanda	224	0	0	0	0	0	0	224
Senegal	0	0	515	0	0	0	0	515
Sierra Leone	0	0	0	0	0	150	0	150
Tanzania	0	554	0	0	0	0	0	554
Togo	0	0	0	0	147	0	0	147
Uganda	0	609	0	0	0	0	0	609
Zambia	0	663	0	0	0	0	0	663
Zimbabwe	561	0	0	0	550	0	0	1,111
Total	1,446	3,679	3,532	807	1,975	425	1,825	13,689

Table 2: Sector and sub-sector in which sample firms operate.

Sector	Sub-sector	
Manufacturing	Food	1,614
	Tobacco	11
	Textiles	275
	Garments	686
	Leather	130
	Wood	270
	Paper	76
	Publishing, printing, and recorded media	555
	Refined petroleum product	29
	Chemicals	334
	Plastics & rubber	262
	Non-metallic mineral products	411
	Basic metals	141
	Fabricated metal products	574
	Machinery and equipment	158
	Electronics	105
	Precision instruments	13
	Transport machines	71
	Furniture	716
	Recycling	19
Retail	Retail	2,478
Other service	Wholesale	1,088
	Construction	631
	Services of motor vehicles	805
	Hotel and Restaurant	1,328
	Transportation, Storage, and Communications	677
	IT and other professional services	232

Table 3: Definition and measurement of control variables included in the model

Explanatory variables	Definition	Measurement
Log of Employees	Measuring firm size using the logarithm of the number of employees.	Logarithm of the amount
Log of Firm Age	Number of years since the firm was established.	Logarithm of years
Financial Constraint	Firms that did not apply for a loan for the following reasons: complex application procedure, unfavorable interest rates, very high collateral requirement, insufficient size of loan and maturity, perception it would not be approved, and other reasons.	1 = yes; 0 = otherwise
Top Manager's Experience	Top manager's years of experience working in the firm's sector.	Years
Exporter	Whether at least 10 percent of the firm's annual sales is derived from direct export.	1 = yes; 0 = otherwise
Foreign Owned	Firm is foreign owned.	1 = yes; 0 = otherwise
Ownership Structure	Whether firm ownership structure is: open-shareholding, closed-shareholding, sole proprietorship, partnership, limited partnership, or other.	1 = yes; 0 = otherwise for each ownership structure
Legal System Quality	Perception of the judiciary, based on answer to the following question: Does the firm strongly agree or tend to agree that the court system is fair, impartial, and uncorrupt?	1 = yes; 0 = otherwise
Log of GDP per Capita	Logarithm of Current GDP per capita (constant 2010 in USD).	Logarithm of the amount
Heritage Economic Freedom Index	The Index covers four main aspects of the economic and entrepreneurial environment: rule of law, government size, regulatory efficiency, and market openness.	0 to 100
Sector	Sector classifications of the firm based on ISIC classification.	1 = yes; 0 = otherwise for each sector

Table 4: Summary Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Innovation	13,689	0.324	0.468	0	1
R&D Investment	13,689	0.199	0.399	0	1
Employee Training	13,689	0.302	0.459	0	1
Initial Informality	13,689	0.184	0.387	0	1
Log of Employees	13,689	2.964	1.185	0	10.309
Log of Firm Age	13,689	2.541	0.820	0	4.828
Financial Constraint	13,689	0.422	0.494	0	1
Top Manager's Experience	13,689	15.587	10.115	0	60
Exporter	13,689	0.115	0.319	0	1
Foreign Owned	13,689	0.105	0.306	0	1
Legal System Quality	13,689	0.419	0.493	0	1
Log of GDP per Capita	13,689	6.955	0.700	5.503	8.725
Heritage Economic Freedom Index	13,689	52.585	8.490	22.100	62.700

Table 5: Correlation Matrix

Variables	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Innovation	1									
2. R&D Investment	0.27***	1								
3. Employee Training	0.18***	0.26***	1							
4. Initial Informality	-0.02**	-0.06***	-0.07***	1						
5. Log of Employees	0.11***	0.19***	0.24***	-0.21***	1					
6. Log of Firm Age	0.05***	0.08***	0.09***	-0.01	0.29***	1				
7. Financial Constraint	-0.06***	-0.10***	-0.12***	0.08***	-0.19***	-0.04***	1			
8. Top Manager's Experience	0.04***	0.02**	0.07***	-0.03***	0.18***	0.48***	-0.02***	1		
9. Log of GDP per capita	0.05***	0.06***	0.04***	0.11***	-0.09***	0.07***	-0.01	-0.02*	1	
10. Heritage Economic Freedom Index	0.02*	-0.05***	0.05***	0.10***	-0.10***	-0.15***	-0.09***	-0.04***	0.25***	1

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Results of Logistic Regression

	(1)	(2)	(3)	(4)
	Innovation	R&D Investment	Employee Training	Innovation
R&D Investment				1.014***
				(0.049)
Employee Training				0.579***
				(0.045)
Initial Informality	-0.134**	-0.214***	-0.185***	-0.089
	(0.054)	(0.068)	(0.058)	(0.055)
Log of Employees	0.172***	0.289***	0.362***	0.088***
	(0.019)	(0.022)	(0.020)	(0.020)
Log of Firm Age	0.013	0.061*	0.027	-0.001
	(0.029)	(0.036)	(0.030)	(0.030)
Financial Constraint	-0.135***	-0.328***	-0.302***	-0.062
	(0.041)	(0.051)	(0.043)	(0.042)
Top Manager's Experience	0.004*	-0.004	0.007***	0.004*
	(0.002)	(0.003)	(0.002)	(0.002)
Exporter	0.261***	0.431***	0.209***	0.171***
	(0.061)	(0.070)	(0.063)	(0.063)
Foreign Owned	-0.096	-0.270***	0.048	-0.067
	(0.065)	(0.078)	(0.066)	(0.069)
Ownership Structure				
Open Share	-0.057	0.617***	0.454***	-0.236
	(0.162)	(0.208)	(0.175)	(0.164)
Closed Share	-0.107	0.425**	0.334**	-0.237
	(0.148)	(0.193)	(0.162)	(0.149)
Sole Proprietorship	-0.298**	-0.104	-0.205	-0.291**
	(0.139)	(0.187)	(0.156)	(0.139)
Partnership	-0.211	0.177	0.078	-0.271*
	(0.152)	(0.199)	(0.165)	(0.153)
Limited Partnership	0.149	0.506***	0.177	0.044
	(0.143)	(0.190)	(0.159)	(0.143)
Legal System Quality	0.042	0.079	-0.035	0.037
	(0.040)	(0.049)	(0.042)	(0.041)
Log of GDP per Capita	-0.763***	-0.272	-0.382**	-0.735***
	(0.145)	(0.166)	(0.160)	(0.150)
Heritage Economic Freedom Index	0.005	-0.006	-0.031**	0.009
	(0.014)	(0.017)	(0.014)	(0.014)
Constant	4.058***	0.356	0.790	3.544***
	(1.242)	(1.440)	(1.345)	(1.277)
Sector	YES	YES	YES	YES
Country Fixed Effect	YES	YES	YES	YES
Year Fixed Effect	YES	YES	YES	YES
Log-Likelihood	-8,121.09	-5,975.63	-7,518.01	-7,758.13
Number of observations	13,689	13,689	13,689	13,689
Chi2	914.183	1,470.646	1,472.286	1,494.407
Pseudo R2	0.059	0.126	0.103	0.101

Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Results of Simple Mediation Analysis

	Observed Coefficient	Bootstrap std. err.	Normal-based [95% Conf. Interval]	
Simple Mediation (*R&D Investment*) †				
Total Effect	-0.169***	0.054	-0.276	-0.063
Average Indirect Effect	-0.072***	0.011	-0.093	-0.051
Average Direct Effect	-0.098*	0.053	-0.201	0.006
Simple Mediation (*Employee Training*) ††				
Total Effect	-0.177***	0.055	-0.284	-0.070
Average Indirect Effect	-0.067***	0.009	-0.084	-0.050
Average Direct Effect	-0.109**	0.054	-0.215	-0.004

†Bootstrap replications: 4991

††Bootstrap replications: 4984

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8: Results of Parallel Mediation Analysis (R&D Investment & Employee Training)

	Panel A†			
	Observed Coefficient	Bootstrap std. err.	Normal-based [95% Conf. Interval]	
Total Effect	-0.197***	0.054	-0.303	-0.092
Average Indirect Effect	-0.114***	0.013	-0.139	-0.089
Average Direct Effect	-0.083	0.052	-0.185	0.019
	Panel B††			
	Observed Odd Ratio	Bootstrap std. err.	Normal-based [95% Conf. Interval]	
Total Effect	0.821***	0.044	0.739	0.912
Average Indirect Effect	0.892***	0.011	0.870	0.915
Average Direct Effect	0.920	0.048	0.831	1.019

†Bootstrap replications: 4985

†† Bootstrap replications: 4984

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

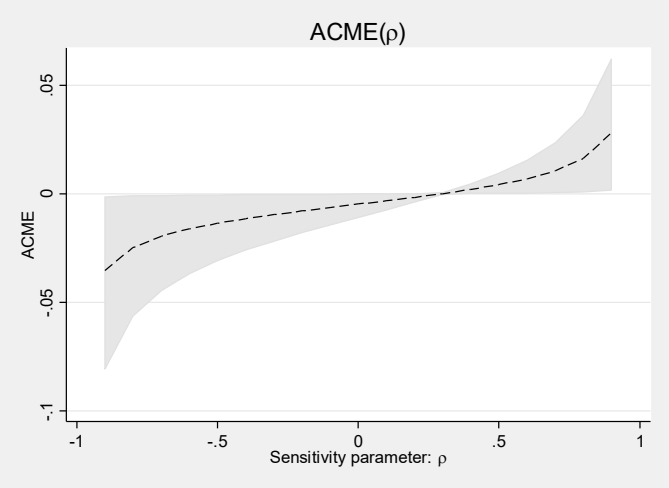

Figure 1: Average causal mediation effect (ACME) of R&D Investment as a function of degree of violation of SI assumption

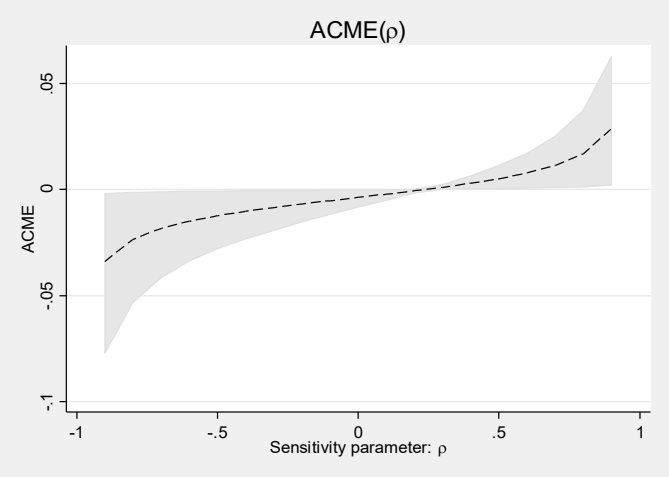

Figure 2: Average causal mediation effect (ACME) of Employee Training as a function of degree of violation of SI assumption

Table 9: The Second Stage Results of Two-Stage Least Square Instrumental Variable Regression (2SLS IV)

	(5)	(6)
	Innovation	Innovation
R&D Investment	0.434***	
	(0.165)	
Formal Training		0.835***
		(0.072)
Log of Employee	0.017**	-0.022***
	(0.009)	(0.007)
Log of Firm Age	-0.002	-0.002
	(0.006)	(0.008)
Financial Constraint	-0.009	0.011
	(0.011)	(0.011)
Top Manager's Experience	0.001**	-0.000
	(0.000)	(0.001)
Exporter	0.024	0.015
	(0.018)	(0.018)
Foreign Owned	-0.002	-0.016
	(0.015)	(0.019)
Ownership Structure		
Open Share	-0.055	-0.124***
	(0.037)	(0.044)
Closed Share	-0.052	-0.089**
	(0.033)	(0.039)
Sole Proprietorship	-0.059**	-0.039
	(0.029)	(0.036)
Partnership	-0.056*	-0.066*
	(0.032)	(0.039)
Limited Partnership	-0.002	0.011
	(0.034)	(0.037)
Legal System Quality	0.004	0.019*
	(0.008)	(0.010)
Log of GDP per Capita	-0.160***	-0.125***
	(0.032)	(0.038)
Heritage Economic Freedom Index	0.000	0.005*
	(0.002)	(0.003)
Constant	1.277***	1.069***
	(0.269)	(0.312)
Sector	YES	YES
Country Fixed Effect	YES	YES
Year Fixed Effect	YES	YES
Number of observations	13,689	11,734
Chi^2	1,283.485	872.913
F-test	50.5637	291.982
Eigen Values	51.8317	324.412
Shea's R^2	0.0038	0.027

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.