The Journal of Technology Transfer (2026) 51:3047–3078
https://doi.org/10.1007/s10961-025-10268-w

RESEARCH

The *'underdog'* informal Vietnamese enterprises: R&D, innovation and productivity performances

Marco Di Cintio[1] · **Claudio Petti**[1]

Accepted: 31 July 2025 / Published online: 26 August 2025
© The Author(s) 2025, corrected publication 2025

Abstract
This paper analyzes the different performances of formal vs. informal firms in terms of R&D, Innovation, and Productivity. Using data from the 2011, 2013, and 2015 results of the Micro, Small, and Medium Enterprise Survey in Vietnam, we also evaluate the effects of the transition from an informal to a formal status on firms' innovativeness and productivity. By accounting for the under-reporting of R&D investments, we show that (i) informal firms invest less in R&D, and (ii) those who switch to a formal status have a higher probability of introducing innovations. However, our study provides evidence of hump-shaped R&D-innovation and innovation-productivity relationships. Also, it identifies the existence of an innovation probability threshold, beyond which informal firms outperform formal ones in productivity. Although the transition to formal status exerts a positive direct effect on productivity, the formalization's benefits diminish as innovation probability intensifies. Hence, these findings underscore an interesting and hidden trade-off in the linkage between firms' flexibility and productivity. Even though formal firms have a higher probability of introducing innovations, our results also reveal that—against all odds—(iii) the'underdog' informal Vietnamese firms may have a higher level of productivity, the more likely they are to introduce radical or incremental innovations. Results also hold when we control for a reverse causality effect of firms' productivity that may result in a driven factor of their formalizations. Thus, our outcomes challenge one-size-fits-all approaches to formalization, and policymakers should recognize that formalization's benefits also depend on the innovative profile of firms.

Keywords CDM · Informal firms · Innovation · Productivity · R&D · Transition

JEL Classification D22 · E24 · E26 · J46 · L26

✉ Marco Di Cintio
marco.dicintio@unisalento.it

Claudio Petti
claudio.petti@unisalento.it

[1] University of Salento, Lecce, Italy

1 Introduction

In sports language, the word *'underdog'* refers to a team or player in a game, who—in the end—wins the competition against all odds. Our paper shows that, when engaged in innovative activities, informal high-innovative Vietnamese enterprises achieve higher productivity levels than would have been reasonably expected.

Informal (legal) economic activities involve workers employed without an official contract and unregistered firms. Informal firms are typically small, use limited external finance, employ low-skilled workers and are usually less productive compared to formal enterprises. Most informal enterprises are engaged in unproductive activities that are expected to disappear as highly productive formal enterprises grow and drive economic development (La Porta & Shleifer, 2008, 2014). Therefore, one would not expect such firms to be so productive. Also, the disadvantages in terms of resources' access and availability inherent to a smaller size, often exacerbated by a race to innovation, make competition a daily bet for many small and medium-sized enterprises (SMEs). Nonetheless, evidence of higher-than-expected productivity levels is more than anecdotal, especially in smaller firms. Actually, Cefis et al. (2025) found Italian micro-firms to over-perform larger SMEs counterparts in terms of policy-induced productivity effects. These findings mirror Baumann and Kritikos (2016) conclusions about R&D intensity and downstream productivity in German firms engaging in innovative activities, which increase as size decreases. Both studies highlight a more than marginal role of micro, smaller *'underdogs'*, equivalent to that of larger SMEs and, in some cases, even better. Consistent with the above findings, this paper extends this framework to the informal-innovative Vietnamese enterprises and shows that, surprisingly, these firms are more productive than their formal counterparts.

Nowadays, the informal economy is a global phenomenon that plays a significant role in many developing economies worldwide. In such economies, the informal sector contributes, on average, to about 35% of the gross domestic product (Loayza, 2016) and the informal sector employs a large portion of the population, particularly in urban areas (Nguimkeu, 2014). The International Monetary Fund (2017) reports that approximately 70% of jobs are informal. Yet, the informal economy is a relevant phenomenon in developed economies, too. In most OECD countries, it makes between 10 and 20% of GDP, and even more in Southern and Central Europe (Schneider & Enste, 2000). Worldwide, two-thirds of all enterprises start businesses without registration (Autio & Fu, 2015), and more than half of all existing enterprises operate without being formally registered (Acs et al., 2013). However, the nature of informal activity often differs between developing and developed economies. While in these latter it primarily involves tax evasion and the resort to undeclared labor, in developing and emerging economies, informal activity serves as a vital source of employment for a substantial portion of the workforce, compensating for the limited capacity of the formal sector to generate sufficient job positions. Indeed, informal businesses often have lower entry barriers than formal ones, which makes them an important source of employment for many individuals, who are almost forced to work in the informal sector (La Porta & Shleifer, 2008). However, since informal firms tend to evade taxes and hurt formal firms, for example, by not complying with patent law or other regulations (Baily et al., 2006; Schneider et al., 2011), they are seen as harmful to the economy. Therefore, governments and policymakers have recently placed a greater emphasis on formalization policies.

The benefits of formalization would seem uncontroversial. At the macro level, a larger number of formal enterprises has a positive impact on GDP level and growth rate. Countries with larger formal economies tend to have higher levels of GDP per capita (La Porta & Shleifer, 2008). In addition, employees working in formal firms generally receive higher wages and experience better working conditions than informal colleagues, both of which promote higher labor productivity. Moreover, increasing the number of formal enterprises is likely to lead to a virtuous cycle of mutual reinforcement between formalization and economic growth. For instance, the increase in tax revenues collected from formalized businesses may fund greater public investment in infrastructure and education, which in turn can lead to improved workforce skills and competencies. On the other hand, better working conditions and high-skilled workers increase firms' productivity by incentivizing their formalization and supporting economic growth. Furthermore, according to Klapper et al. (2006), improved regulation encourages firms to operate in the formal sector, and formalization may enhance the public perception of government quality (De Mel et al., 2013). In turn, it could also lead to a positive impact on the willingness to adhere to regulations, making formalization self-sustaining. Good governance, in fact, plays a prominent role in determining the kind of entrepreneurship undertaken (Baumol, 1990). Nonetheless, firms' formalization has positive effects at the micro-level as well. Informal firms are often situated in underdeveloped locations with limited economic and physical infrastructure, relying on low-skilled workers (Oyelaran-Oyeyinka & Lal, 2006), labor-intensive production, and outdated machinery (Yuki, 2007). Hence, earlier research has demonstrated that informal firms usually do not introduce innovative products or services (Assefa et al., 2022). Thus, compared to those in informal sectors, firms operating within formal sectors generally have better performance in terms of innovation, growth, and survival rates (La Porta & Shleifer, 2014).

Following the above, we look at the case of Vietnam for two reasons. First, because in this country firms have been formalizing faster than elsewhere. The ongoing acceleration of formalization among those firms is partly due to relatively recent policies and laws (2013 Government Decree) that apply pecuniary penalties (approximately from five to ten million of Vietnamese dong) on violations of business registration obligations for household businesses. Second, because not only the relationship between innovation and productivity is at the core of the Country's catch-up strategy (Korean Development Institute, 2020; World Bank Group & MPI, 2016), but also for the peculiar urgence put in these plans to develop the necessary knowledge stock and absorptive capacity to make the most of international knowledge and technology transfer. Indeed, it is to inefficiencies in this key process that their labor productivity, R&D and innovation gaps against foreign counterparts' productivity gains are attributed to. More specifically, by using data from the 2011, 2013, and 2015 results of the Micro, Small, and Medium Enterprise Survey in Vietnam, we build a three-stage empirical model, where *(i)* we account for the under-reporting of R&D in informal establishments; *(ii)* estimate the innovation likelihood of firms due to R&D and to the *switch to a formal status*; and *(iii)* evaluate the simultaneous effects of both innovations and the *transition from an informal to a formal status* on firms' productivity, shedding light on labor productivity differences between formal and informal innovative firms.

In this context, our paper extends earlier studies along three main dimensions. First, our study integrates both R&D investments and innovation outcomes into an empirical model of formal and informal firms' productivity in the context of a developing economy. Indeed,

the richness of our data allows us to apply the *Crépon-Duguet-Mairesse* (CDM) structural model (Crépon et al., 1998), in which firm productivity is a function of *radical* or *incremental* innovations, which in turn are explained by the R&D intensity of firms. Although the CDM model has been already employed to study firms' productivity in many emerging countries (see Zhang & Islam, 2022; Gaglio et al., 2022; Rukundo, 2017), to the best of our knowledge the differences between informal and formal firms and the effects of formalization on their productivity has not been investigated in previous literature. Second, our contribution adds the informality status to the evidence collected about *'underdogs'* unexpected performance linked to firms' size (see Baumann & Kritikos, 2016; Cefis et al., 2025). Thus, we extend the extant micro-firms' empirical evidence to informal firms in a developing context. From this standpoint, the lack of data on informal firms and the difficulties in measuring innovations in developing countries have limited the empirical research on the connection between innovations and firm productivity, especially with respect to informal firms. Hence, as a third contribution, our analysis also complements the collection of existing findings on developed countries, consistently with the early adoption of the CDM model (see, for instance, Hall et al., 2009; Antonelli & Colombelli, 2015; Medda, 2020; Antonelli & Feder, 2021).[1]

In particular, our study extends the CDM framework by integrating informal sector dynamics and revealing hump-shaped R&D-innovation and innovation-productivity relationships. Also, it identifies the existence of an innovation probability *threshold*, beyond which informal firms outperform formal ones in productivity. In some way, this challenges CDM's traditional focus on formal firms, demonstrating how the reduction in firms' flexibility could moderate the innovation-productivity linkages. To the best of our knowledge, these are novel mechanisms absent in prior CDM analyses. Specifically, we show that: *(i)* informal firms invest less in R&D, and *(ii)* those who *switch to a formal status* have a higher probability of introducing innovations. Although the transition to formal status also exerts a positive direct effect on productivity, the formalization's benefits diminish as firms' innovation probability intensifies. However, while we find that formal firms are more likely to introduce innovations, our results also reveal that – against all odds – *(iii)* the *'underdog'* informal Vietnamese firms may have a higher level of productivity, the more likely they are to introduce radical or incremental innovations. The discovery of this interesting and hidden trade-off challenges one-size-fits-all approaches to formalization, and policymakers should recognize that formalization's benefits also depend on the innovativeness levels of firms. Overall, the results of our paper suggest economic policies designed to encourage the transition of firms from an informal to a formal status. In practice, benefits may be expected from interventions that stimulate an innovative environment where informal enterprises grow as successful innovators. Moreover, since high innovative informal firms are found to be more productive than formal ones, greater economic advantages can be expected in terms of within-firm human capital growth and knowledge development by supporting informal firms to operate in the formal sector, but leaving them maintaining their high flexibility level.

The remainder of the paper is organized as follows. Section 2 reviews the literature and discusses previous empirical research on informal firms' formalization and their performance in terms of R&D, Innovation and Productivity. Section 3 specifies our empirical

[1] For a recent review of works that have adopted or extended the CDM model, see Mairesse et al. (2025).

model. Data are described in Sect. 4, while Sect. 5 debates the empirical findings. Discussion and conclusions are reported in Sect. 6.

2 Literature background

Since formalization is expected to bring about positive outcomes for the economy, such as greater R&D investments, more innovations, increased economic growth, higher productivity of firms, and improved welfare (Tijdens et al., 2015), addressing the informal economy has received so much emphasis as being included in Sustainable Development Goals (SDGs). Consequently, policymakers place a high priority on formalization strategies, deploying interventions ranging from financial incentives (e.g., tax exemptions, credit guarantees) to capacity-building programs (e.g., worker training, managerial upskilling). These efforts align with the broader consensus that informality poses a structural barrier to inclusive development, particularly in economies with weak institutional frameworks. To systematically tackle this challenge, governments typically adopt three complementary approaches. First, they may reduce formalization costs by simplifying registration procedures, reducing bureaucratic delays, and offering targeted subsidies to offset compliance expenses. Second, they may enhance formalization benefits by expanding access to formal-sector advantages such as public procurement contracts, innovation grants, and legal protections for intellectual property. Third, they may strengthen enforcement mechanisms by increasing penalties for non-compliance or conducting regular fiscal inspections. However, despite numerous attempts by policymakers to bring the informal firms into the formal sector, there is still a lack of clear results on the impact of these efforts. The results about the effectiveness of such policies are conflicting, varied, and often modest. Apart from the notable exception of De Mel et al. (2013) for Sri Lanka, the empirical evidence on the effects of financial and monetary benefits on firms' formalization is inconsistent. Rocha et al. (2018) found that cutting taxes and reducing the costs of formalization procedures were an incentive for Brazilian informal enterprises to formalize. Differently, Fandl and Bustamante (2016) reported no significant change in business registrations in Colombia, while Goldszmidt et al. (2022) found that the microcredit offered by the Brazilian public bank did not increase formalization. Moreover, some studies show conflicting results even for the same kind of reform in different countries, or for the same country in different cutoff dates (Fajnzylber et al., 2011 vs. Monteiro & Assunção, 2012).

However, the linkages among R&D investments, innovation and increased firms' productivity, via the increase of a firm's knowledge stock (Griliches, 1979), make the push towards formalization of local firms a key step in developing countries' catching-up efforts. In catching-up contexts, the quest for innovation-led productivity increase is particularly urgent due to the lack of technologies and market access inherent to the latecomer nature of domestic firms (Hobday, 1995; Mathews, 2002). To fill these gaps, the endowment of an essential stock of knowledge is a necessary condition for benefit from technology transfer (Kedia & Bhagat, 1988) and knowledge spillovers brought by import–export dynamics and foreign direct investments (Guo et al., 2019). Thus, formalization is a necessary step to create another virtuous cycle between export-driven and innovation-driven growth. Therefore, R&D and innovation, no matter whether 'secondary' (Wu et al., 2009), 'market-oriented' (Liu, 2008), or technology-driven, play a key role in domestic latecomer firms' learning and

catch-up. This is because it allows for the benefit of the transfer, absorption and exploitation of technologies and knowledge from foreign enterprises to develop stronger manufacturing capabilities (Choi et al., 2019) and, hopefully, lead to the increase of productivity needed to catch-up.

For all these reasons, innovation has become increasingly important in developing countries, leading to new fields of study. A recent branch of the literature specifically concentrates on the introduction of new products and services by agents not included in the formal economy. In this respect, the resourcefulness, ingenuity, and adaptability of individuals living in poverty conditions are recognized to be the main sources of local innovation. Thus, when evaluating the impact of innovation on a firm's performance in many developing nations, it is crucial to acknowledge the presence of a dual-economy system consisting of formal and informal businesses (La Porta & Shleifer, 2014).

The literature on the role of innovation on firms' performance in low-income countries has also grown in recent years. For example, Mahemba and Bruijn (2003) surveyed SMEs in Tanzania and found a positive association between firms' innovativeness and their productivity. Similarly, Gebreeyesus (2009) found that Ethiopian innovative SMEs grew faster in terms of employment. In Sri Lanka, De Mel et al. (2009) found that innovations increase the profits of SMEs. Mathew and Paily (2022) found the same in India, also in terms of sales growth. Nevertheless, a growing body of research has also shown a weaker relationship between innovations and firms' performances in emerging economies. For example, Goedhuys et al., (2008, 2014) discovered that firms' productivity in Tanzania was not improved by firms' R&D and innovations, but rather by the business environment. These findings reveal that R&D, innovation, and productivity are more weakly correlated in emerging countries than developed ones. According to the widely accepted belief, while the effect of innovation on firm growth in developed economies has been widely recognized, its impact in emerging countries is only partially documented.

In addition, the lack of data and difficulties in measuring innovation in low-income countries have limited the empirical research on the connection between innovation and firm growth, especially with respect to informal firms (Wunsch-Vincent & Kraemer-Mbula, 2016). In emerging countries, measuring the impact of firms' innovations on their productivity is challenging, also because of the marginal role that R&D investments play for the majority of micro and small informal firms. However, as La Porta and Shleifer (2014) noticed, since economic growth can help in reducing informality, investigating the innovative practices of informal firms and their impact on growth could be a critical factor in this process of formalization. In their survey on informality and development, the authors presented five key characteristics of informal firms in emerging countries. The informal firms (a) account for a significant number of enterprises at the aggregate level and (b) tend to avoid some tax payments and government regulations. On the other hand, the informal sector often includes (c) small and (d) inefficient firms, also because (e) informal entrepreneurs and employees have a limited education. Thus, it leads to low productivity levels and weaker economic growth.

The different likelihood in investing in R&D and adopting and diffusing innovations between formal and informal firms is therefore shaped by their characteristics and capabilities, as demonstrated by Wunsch-Vincent and Kraemer-Mbula (2016). Mainly because of their resource constraints, informal firms often struggle to fully exploit innovation opportunities (Coccia, 2017). Formal establishments, with their greater human and financial

resources, are more likely to engage in collaborative innovation activities with other firms and foreign organizations and institutions (Oyelaran-Oyeyinka et al., 1996). On the other hand, informal firms typically lack financial resources and may rely more on entrepreneurs' drive and the constraints of their operating environment (Robson et al., 2009). Although entrepreneurs' ingenuity has been viewed as the driving force behind innovation for informal firms (Prahalad, 2012), the literature, with the notable exception of De Mel et al. (2009), has mostly focused on observable indicators, such as the firm's size, and workers' and entrepreneurs' education, as factors that determine innovations.

Small firm size is one of the defining features of informal establishments. This aspect, characterizing the extent of a company's operations, has been deemed a hindrance to innovation in several studies (De Mel et al., 2009; Gebreeyesus, 2009; Robson et al., 2009). Moreover, the absence of adequate resources in the education system in many low-income countries prompts non-formal training to become the primary means of acquiring knowledge, along with on-the-job learning (Oyelaran-Oyeyinka & Lal, 2006). Nevertheless, as indicated by Yuki (2007), the majority of workers in the informal sector are low-skilled, causing employers to view employee training as unnecessary and unhelpful. Moreover, informal firms, free from formal employment contracts, have the freedom to dismiss employees with little to no compensation (Harris, 2014a). Thus, the lack of stability in the workforce, along with the possibility of trained employees benefiting competitors, could reduce the incentive for informal firms to invest in employee formal training (Harris, 2014b). However, since informal firms are not bound by labor regulations, their employment strategies are usually flexible, allowing them to easily adjust their workforce in case of necessity (Ingram et al., 2007). In turn, flexible employment adjustments might mitigate the negative effects of product market uncertainty on firms' innovative efforts both in developing and developed countries (Di Cintio & Grassi, 2017).

In addition, informal firms are vulnerable to having their innovative products or services stolen by competitors due to the lack of protection offered by intellectual property rights (Harris, 2014b). Moreover, even though informal firms could be able to introduce new and innovative products or services, they might face challenges in terms of accessing distribution channels. Indeed, the limited access to marketing channels and media restricts the capacity of these enterprises to gain the benefits of their innovations (Teece, 1986). Anyway, informal firms often concentrate on survival and make investments that bring quick returns (Harris, 2014c). As a result, they are less likely to allocate financial resources toward R&D. According to Bu and Cuervo-Cazurra (2020), informal firms that began as such may prioritize short-term goals over innovation, resulting in lower levels of innovation compared to formally established firms. Similarly, Mendi and Mudida (2018) found that past informality status hurts technological innovations in Kenya. Therefore, long-term investments—such as those in R&D activities—may be less common among informal firms. Indeed, the negative correlation between informality and innovation is due to low investments in R&D (Assefa et al., 2022).

However, research has also shown that small firms in the informal sector are able to innovate and acquire new knowledge. Indeed, Von Hippel (2005) explains that informal firms are encouraged to innovate even when the economic returns of the product they are creating are low because, anyway, they benefit from the process of developing or modifying an existing product. As a matter of fact, R&D is not the only input leading to innovation (Conte & Vivarelli, 2005; Hall et al., 2009; Mathew & Paily, 2022). Especially SMEs tend

to resort to external intermediate outputs, such as new machinery, equipment and technologies more frequently than to internal sources (Conte & Vivarelli, 2005; Santarelli & Sterlacchini, 1990) and often conduct non-permanent R&D using resources from departments different than R&D (Ortega-Argiles et al., 2009). Therefore, it should not be surprising that Fu et al. (2018) offer empirical insights into the impact of innovation on the firms' productivity in Ghana. Their study revealed that formal firms tend to experience a stronger effect on productivity from technological innovations, but both formal and informal firms benefit equally from non-technological innovations. This latter result can thus be attributed to informal R&D and, more specifically, to what Petti et al. (2017) called embedded innovation capacity (EIC). Thus, EIC represents the set of investments in physical/overall capital that induce an endogenous innovation process of technology updating through the replacement of old machinery and equipment (Ortega-Argiles et al 2009) and through the purchasing of external technology incorporated in licenses, consultancies and know-how. Especially in low-tech sectors, EIC often leads to incremental innovations (Santarelli & Sterlacchini, 1990). Such innovation without R&D mode has been argued to be particularly suited to SMEs (Cowan & van de Paal, 2000). As a matter of fact, informal R&D activities have been argued as an explanation for the unexpected productivity of micro-firms (Baumann & Kritikos, 2016) and positively correlated to small Indian firms' performances (Mathew & Paily, 2022). In particular, these latter authors recognized that informal innovation processes occur from *'learning-by-DUI'* mode (e.g. *D*oing, *U*sing, and *I*nteracting) as key mechanisms for latecomer SMEs.

Summing up, although the contribution of innovation to firms' productivity in developed countries is well documented, its impact is still mostly unclear in developing economies, mainly across informal firms. Empirical studies have been limited in emerging economies due to issues related to data availability and measuring innovation inputs and outputs for informal firms. Moreover, the literature has focused either on the effects of policy interventions aimed at formalizing informal firms or on the effects that different levels of firms' productivity may have on the *switching to the formal status* of existing informal firms. Therefore, to the best of our knowledge, there are no studies, instead, that analyze the impact of the *transition from an informal to a formal status* on firms' innovativeness and productivity in emerging countries.

This paper aims to close this gap by exploring data from the Enterprise Survey in Vietnam. In particular, as will be explained in detail in the following sections, we will try to answer three main research questions that inspired our work: *1) "In the context of emerging countries with both formal and informal firms, how does R&D and innovation affect firms' productivity?"*; *2) "are there differences between formal and informal firms?"; 3) "could the transition from informal to formal status enhance firms' innovativeness and productivity?"*.

3 Empirical model

Our empirical strategy applies a structural model (Crépon et al., 1998) to answer our research questions. Specifically, the empirical model is constructed in three stages and includes three equations to account for the endogeneity of firms' innovation efforts and outcomes. In the first stage, we address the concern of under-reporting of R&D expenditures and estimate an

R&D input function at time t. Then, in the second stage, we consider a *knowledge generation function* at time $t+2$ that reflects the firms' probability of introducing innovations in the next two years. Finally, in the last stage, we estimate a *productivity function* at time $t+4$ to evaluate how innovation affects firms' labor productivity in the following two years. Hence, our empirical methodology takes into account a two-year lagged time between each step to alleviate potential problems of endogeneity and reverse causality. In this way, we can also be more confident that R&D investments have enough time to fully deploy their effects on innovation outcomes and that innovations, in turn, have a useful time span to improve labor productivity.

Since a firm's productivity represents the ability of a firm to generate greater outputs with fewer inputs, we suppose, firms have a (CRS) *Cobb–Douglas* production function $Y_i = A_i * K_i^{\alpha_k} * L_i^{\alpha_L}$, where firm-level output, Y_i, is a function of inputs of capital (K_i), labor (L_i) and knowledge (A_i). Thus, the term A_i is the portion of output that cannot be directly explained from physical inputs. It represents a proxy of the total know-how and knowledge incorporated into the firm's production process. Hence, by using this model, we can easily estimate the effects of innovations (i.e. knowledge) on the firm's productivity, by taking into account that higher R&D investments may help firms to introduce innovations. To answer our research questions, these relationships are evaluated by considering the effects of informality on the different firms' performances in the specific way described in what follows.

As proposed by earlier studies, self-reported R&D investments for small and medium enterprises may significantly underestimate their innovation efforts (Kleinknecht, 1987; Santarelli & Sterlacchini, 1990). For instance, Roper (1999) suggested that factors such as the presence of informal innovation activities, the peculiar nature of R&D undertaken, or the presence of unstructured R&D divisions can influence the reported R&D effort, and this could be especially the case for informal firms in developing countries. Consequently, self-reported R&D expenditure often inadequately reflects the actual innovative investments of such firms. So, estimating the R&D intensity equation becomes crucial, and it is a necessary stage to take a more accurate proxy for the R&D investments made by firms in our sample.[2] Therefore, given the high likelihood of under-reporting of R&D by informal firms, we assume that a latent variable $\widetilde{R\&D}$ is linked to the firm's informality status, INF_i, and a set of independent control variables, x_{1i}.[3] Thus, we use a standard *type II Tobit model*:

$$\begin{aligned} \widetilde{R\&D}_i &= \gamma_1 * INF_i + \beta_1 * x_{1i} + \alpha + v_i \\ R\&D_i &= \begin{cases} \widetilde{R\&D}_i, & \widetilde{R\&D}_i > c \\ 0, & \widetilde{R\&D}_i \leq c \end{cases} \end{aligned} \tag{1}$$

where γ_1 and β_1 are coefficients to be estimated, α is a constant term, v_i is the error term, and c represents the threshold above which R&D becomes observable.[4] The choice of the

[2] A comparable methodology can be found in Crépon et al. (1998) and Hall et al. (2009).

[3] An exhaustive explanation of the independent variables is specified in Sect. 4. It is worth noting that, among regressors, we also include the correction term for selectivity.

[4] The censoring point is—a priori – unknown, and an arbitrary choice in setting this threshold is not innocuous. Following Carson and Sun (2007), a consistent procedure is to set $c = min\left\{\widetilde{R\&D}\right\}$.Thus, in our

Tobit model is a prevalent practice in innovation studies, especially when some variables are exclusively observed for firms involved in innovation. Therefore, we interpret the firm's R&D efforts ($R\&D_i$) as a corner solution response variable, taking value of 0 with positive probability, while remaining a continuous variable with strictly positive values. This implies that some firms may not find it optimal to engage in R&D activities.

The CDM approach has been adapted to account for a different process underlying the likelihood to undertake R&D and the intensity of R&D expenditure (Medda, 2020), and to avoid also any risk of selection bias in predicting firms' R&D intensity. Nevertheless, by showing the differences between the empirical strategy adopted by Crépon et al. (1998) and Mairesse et al. (2005), some studies have also shown that the inclusion of a correction term in the first stage is not strictly necessary when innovations are measured by a binary indicator (see, for instance, Giotopoulos et al., 2023; Medda, 2020). However, since those firms engaged in R&D activities may not be randomly selected, we still apply a Heckman (1979) two-step procedure to address potential selection bias in R&D intensity estimation.[5] Following Antonelli and Colombelli (2015), we first estimate a *Probit* model for the probability of engaging in R&D.[6] We then estimate the above R&D intensity equation, including the *Inverse Mills Ratio* (IMR) to correct for selectivity. From this *R&D input function*, we derive the predicted values as a firm-level proxy of actual R&D intensity ($\widehat{R\&D}$).[7]

In the subsequent stage, we explore the relationship between innovation output and R&D intensity (i.e. innovation input). Specifically, we will refer to the introduction of both *radical* (i.e. firms introduced new products) and *incremental innovations* (i.e. firms made any improvements to existing products). In particular, in considering the *knowledge generation function* that reflects the firms' probability of introducing innovations, we run a *Probit model* where we regress the unobservable latent variable ($\widetilde{Inn}$) on the predicted R&D intensity ($\widehat{R\&D}$), the firm's informality status, INF_i, the *"transition to formality status"* ($Trans_FORM_i$) in the time span between time t and time $t+2$ and a set of independent control covariates (x_{2i}):

$$\widetilde{Inn}_i = \theta * \widehat{R\&D}_i + \gamma_1 * INF_i + \gamma_2 * Trans_FORM_i + \beta_2 * x_{2i} + \delta + \varepsilon_i$$
$$Inn_i = \begin{cases} 1, \widetilde{Inn}_i > 0 \\ 0, otherwise \end{cases} \tag{2}$$

where θ, γ_1, γ_2 and β_2 are coefficients to be estimated, δ is a constant term, and ε is the error component. Thus, for each firm, Inn_i is the observed dummy variable equal to 1 if a firm introduced innovations.[8] We use these estimates to obtain predicted values of the firms'

case, this threshold is equal to zero.

[5] The model is identified without exclusion restrictions due to the non-linearity of the Probit residuals at the preliminary step (Vella, 1998).

[6] Different from them, the structure of our data prevents us from applying the estimation method proposed by Wooldridge (1995) with panel data.

[7] Note that, implicitly, predicted values of R&D reflect both official and informal R&D investments made by firms. Moreover, as pointed out by Mairesse (2005), predicted R&D intensity already serves as a strong instrument for innovation.

[8] While we acknowledge that external knowledge acts as a complementary input to internal R&D in the *knowledge generation function*, our dataset does not allow us to incorporate measures such as the cumulative stock of patents (Antonelli & Colombelli, 2015) or to include the sector-level R&D knowledge stocks available to firms in the *R&D input function* (Antonelli & Feder, 2021).

probability to introduce innovations ($\widehat{Inn_i}$), which are then included in the third stage of our structural model as the key determinants of firms' labor productivity.[9]

The last stage of our empirical model is dedicated to studying the effects of both firms' innovation ($\widehat{Inn_i}$) and *"transition to formality status"* ($Trans_FORM_i$) on their labor productivity (y_i). Therefore, in order to capture the differences in productivity between formal and informal innovative firms, we include an interaction term between these two main regressors. It is worth noticing that while the firm's productivity is evaluated at time $t+4$, the probability of innovation refers to time $t+2$, and the dummy related to the transition into a *formal status* indicates whether the same firm switched from informal to formal status in the time span between time t and time $t+4$. Thus, we estimate the firms' *productivity function* through the following OLS equation:

$$y_i = \vartheta * \widehat{Inn_i} + \gamma_2 * Trans_FORM_i + \gamma_3 * \left(Trans_FORM_i * \widehat{Inn_i}\right) + \beta_3 * x_{3i} + \varphi + \mu_i \quad (3)$$

where ϑ, γ_2, γ_3, and β_3 are coefficients to be estimated, $varphi$ is a constant term and μ is the error term.

To capture potential non-linearities in the innovation process, we include the squared term of R&D intensity in the *knowledge generation function* and the square of innovation probability in the *productivity function*. This specification allows us to test for diminishing returns of R&D on the likelihood of innovation, and of innovations on firm productivity. In contexts with limited absorptive capacity or inefficiencies in innovation systems—typical of many developing and emerging economies—returns to R&D may not be constant (Hall & Maffioli, 2008). Moreover, evidence from Latin America and other emerging economies suggests that returns to innovation and productivity gains may be strongly decreasing because of institutional constraints, financial frictions, or low research complementarities (Crespi et al., 2016).

It is worth noting that we also apply an exclusion restriction as an identification strategy. Specifically, there should be at least one significant explanatory variable in the *knowledge generation function* that is not used in the *productivity function*. This variable should have a direct impact on innovation and affect productivity solely indirectly through innovation.[10] Similarly to Calza et al. (2018), also referring to Vietnamese enterprises, we propose using *"customer certification requests"* as an instrumental variable for innovation. This approach addresses potential endogeneity between innovation and productivity. Specifically, while *"customer certification requests"* are clearly exogenous, they should be strongly correlated with a firm's likelihood of innovating. Customers demanding certifications often push technical or procedural requirements that incentivize firms to innovate—whether through new processes, products, or organizational practices—to meet these standards. On the other hand, *"customer certification requests"* are supposed influencing productivity only indirectly through innovation. Their direct effect on productivity, such as

[9] The use of predicted values also serves as a method to tackle potential endogeneity concerns associated with the knowledge inputs.

[10] In their original model, Crépon et al. (1998) used just two stages instead of three stages as in our model. However, in our opinion and as already mentioned, estimating the R&D intensity equation becomes crucial to take a more accurate proxy for the R&D investments made by informal firms in developing countries.

cost reductions or efficiency gains unrelated to innovation, is negligible.[11] Any productivity impact linked to certifications would stem from the innovation efforts they induce. This satisfies the exclusion-restriction properties, as there is no plausible channel for *"customer certification requests"* to directly affect productivity outside the innovation mechanism. Thus, as an exclusion-restriction, we refer to a dummy variable that reflects whether or not firms have been required by customers to obtain certifications, being confident that it would enhance firms' innovations and affect productivity only through innovations.

At the same time, since an increase in firms' productivity could stimulate the transition of firms from an informal to a formal status, we also need to consider that the variables y_i and $Trans_FORM_i$ could be both endogenous. Indeed, since more productive firms may have better opportunities to operate inside the formal and regulatory framework, a higher level of productivity could have a reverse causality effect that may result in a driven factor of their *"transition to formality status"*. Therefore, as one of the main *robustness checks*, we perform our analysis with a different estimation methodology. In particular, we estimate a simultaneous-equations system by using a *full-maximum-likelihood* estimator (instead of multi-separated steps), where both firms' productivity and the *"transition to formality status"* are supposed to be endogenous.[12] Hence, in the *productivity function* at time $t+4$, $Trans_FORM_i$ is instrumented with the ratio between the amount of *"total fees-and-taxes"* paid by firms and firms' employment at time t.

We are aware that the identification of causal effects would require some further and more in-depth investigation, and that we will only be able to provide suggestive evidence of the correlation among variables included in our analysis. Nevertheless, given the lack of previous work on these relationships, we believe our empirical results provide preliminary insight that could be very useful in shaping future research on this topic.

4 Data and variables description

Our research is based on data from the 2011, 2013, and 2015 waves of the Micro, Small, and Medium Enterprise (MSME) survey in Vietnam, conducted by the United Nations University World Institute for Development Economics Research (UNU-WIDER). The survey comes from the collaboration between the University of Copenhagen and two Vietnamese partners: the Central Institute of Economic Management (CIEM, Vietnam) and the Institute of Labor Studies and Social Affairs (ILSSA, Vietnam). Each wave contained detailed data on more than 2500 formal and informal MSMEs in the Vietnamese private manufacturing sector.

The surveys specifically consist of three modules: (1) main enterprise questionnaire; (2) employee questionnaire; and (3) economic accounts questionnaire. In this study, we especially focused on (1) the enterprise questionnaire which provided detailed information on enterprises performances, firms history, labor force, business environment, and background characteristics of managers/owners, and (2) the economic accounts questionnaire which provided information on firms' revenues, investments, assets, liabilities and taxes.

[11] We tested this formally. While pure correlation is below 0.08, the instrumental variable does not affect firms' productivity. In an auxiliary estimation, the *p-value* signals the absence of significance at conventional levels.

[12] Simultaneous estimations, accounting for a full-covariance structure, are in general also more efficient.

The original sample includes 2512 firms in the wave of 2011, 2542 firms in the wave of 2013, and 2650 firms in the last wave of 2015. Each wave reports information related to the previous two years. The content of the questionnaire remained almost the same, notwithstanding some slight adjustments over time. Specifically, the sample contained a wide range of enterprises (private enterprises, cooperatives, partnerships, limited liability companies, and joint stock companies) operating in Vietnam's 10 provinces. The surveys were administered in the northern, central, and southern regions of the country: *Hanoi, Ho Chi Minh City, Ha Tay, Hai Phong, Long An, Phu Tho, Quang Nam, Nghe An, Khanh Hoa,* and *Lam Dong*. The stratified sample guarantees representation by ownership type for the 10 provinces by incorporating non-state locally-owned manufacturing firms and excluding state-owned joint stock companies and foreign-owned joint ventures. Nevertheless, while other datasets for Vietnam (such as the GSO or the World Bank Survey) exclusively focus on formal firms (i.e., those that are registered and have a tax identification number), this survey is unique because it includes both formal and informal domestic private firms.

To implement our empirical model, we rely on three *key dependent variables*, i.e., the *firms' labor productivity*, the *radical* or *incremental innovations* introduced by firms, and their *R&D intensity*. In particular, we compute R&D intensity (measured in the fiscal year 2010) as the ratio between R&D expenditure and sales, while innovation (measured in the fiscal year 2012) is defined as a dummy equal to one if a firm introduced new products or made any improvements to existing products. Finally, as a measure of labor productivity (measured in the fiscal year 2014), we refer to the *sales-per-employee* at the firm level, which is widely used in innovation literature to capture firms' productivity.

Similar to other studies, we incorporate controls for specific firm characteristics. The firm's *manager/owner education* and the implementation of *training activities for employees* may have a specific impact on R&D intensity. Moreover, since larger firms are expected to better exploit economies of scope and scale, we also control for the firm size, measured by the *level of employment*. Additional controls used in all the estimation stages refer to: a dummy to distinguish between (family business) *households and non-household firms*; a dummy variable indicating if firms incurred *bribes* or other informal payments; *firm liquidity* measured as the ratio between the amount of cash-and-deposit and turnover; *firm age*; *sectors* and *provinces* dummies. Moreover, in the 2015 survey, we were also able to find information on the firm's *manager risk propensity*. Hence, we consider this variable (eleven graduations from 0 to 10; risk aversion = 0—risk loving = 10) as an additional regressor in estimating the productivity function. Moreover, since only a small minority of firms were involved in internationalization activities, we were unable to find a significant number of firms that declared their share of export over turnover. Nonetheless, we use a dummy variable to control whether a firm has an *internationally recognized quality certificate*. Obviously, in the estimation of the firm's productivity, we finally include the capital input measured through the value of the *firm's total assets-per-employee*.

We merge information from the three mentioned waves and remove observations with inconsistent or missing information on the relevant variables. Due to the lack of these pieces of information, the sample in our estimation was limited to 1050 firms, for which we can track the dynamics from a firm's R&D investment in 2010, to innovations introduced in 2012, to productivity in 2014. Table 1 summarizes the descriptive statistics.

As in other developing countries, on average, R&D intensity is particularly low and with high variability, mainly because many enterprises do not report official R&D investments,

Table 1 Summary statistics

Variable	Mean	Std. dev	Min	Max
R&D intensity 2010	0.0023	0.0443	0	1.635
Innovation 2012	0.1724	0.3779	0	1
Radical innovation 2012	0.0067	0.0814	0	1
Incremental innovation 2012	0.1695	0.3754	0	1
Sales-per-employee 2014 (million VDN)	251.128	307.5848	3.955	4304
Informal 2010	0.468	0.4992	0	1
Transition to formal 2012	0.0895	0.2856	0	1
Transition to formal 2010–2014	0.377	0.4849	0	1
Manager education	0.109	0.3112	0	1
On-the-job training	0.037	0.1892	0	1
Employment	9.164	20.4337	1	367
Household	0.919	0.2729	0	1
Bribes	0.675	0.4685	0	1
Liquidity	0.136	0.2827	0	3.695
Age	19.103	9.6022	2	61
Risk propensity	3.725	2.448	0	10
International certification	0.025	0.1555	0	1
Asset-per-employee (million VDN)	366.629	542.9659	3.914	6050
Observations	1050			

and just about 1.74% of firms declare a positive amount. However, in our sample, about 17% of firms introduced new products or improved the existing ones. However, a large proportion of these innovations are *incremental* innovations (16.95%) rather than *radical* innovations (0.67%). Labor productivity (sales-per-employee) is also not particularly high, and a very high standard error can be observed (from a minimum of about 4 million to a maximum of 4.303 million VDN). Moreover, in 2010, nearly 46% of enterprises were informal. However, while 8.95% of informal enterprises transited into the formal economy in 2012, the transition rate is shown to be much higher after 4 years (37.71%). This implies that the government policies devoted to this goal have been quite successful. Managers with a college or university education are just under 11%, and only 3.71% of enterprises provide job training to their employees. On average, there are about 9 employees per enterprise, with a maximum of 367 and a minimum of only one entrepreneur with no employees. Consistently, around 92% of enterprises in the sample belong to family business households. Even though not reported in the table to save space, a significant proportion of households dominate the following three industries: Food Processing (34.29%), Wood Processing (11.11%), and Fabricated Metal Products (17.91%). The 67.52% of firms incurred bribes or other informal payments and the level of liquidity in terms of cash and deposits is particularly modest. While, as expected, the low engagement in international activity is clearly shown by a low level of international quality certification (2.48%), surprisingly Vietnamese entrepreneurs show, on average, that they do not particularly care about risk (the mean of risk aversion is 3.7 on a scale from 0 to 10).

5 Empirical results

As mentioned previously, in examining the relationship among R&D, innovation, and productivity, we developed an adapted version of the model introduced by Crépon et al. (1998). To encompass innovations that may not solely stem from formal R&D activities, especially relevant for informal MSMEs, our estimations are not limited to officially recognized R&D-performing firms. Thus, our first step consists in estimating an *R&D input function* at time t (fiscal year 2010), that is the process that leads formal and informal firms to decide whether to undertake R&D investments or not and how much to spend on R&D. Differently, the second step involves a *knowledge generation function* at time $t + 2$ (fiscal year 2012), in which the predicted R&D intensity (that considers both official and informal investments in R&D) is one of the key inputs and *radical* or *incremental* innovations are the outputs. Finally, the third step concerns a simple *productivity function* at time $t + 4$ (fiscal year 2014) in which knowledge (i.e., the predicted probability to introduce *radical* or *incremental* innovations) is among the main inputs. Moreover, to evaluate the effect of the *transition in the formal status* on firms' performances, in the last two steps, we consider an additional regressor that indicates whether a firm switched *from an informal to a formal status.*

Table 2 reports the results of the selection equation in panel *(a)* and the estimation of the *R&D input function* in panel *(b)*. The first column reports the determinants of R&D, while the first and second column of each panel reports estimated coefficients and standard errors, respectively. To account for potential sample selection into R&D activities, we estimate a preliminary probit model and compute the Inverse Mills Ratio (IMR), which is then included in the R&D intensity equation to correct for selectivity. However, the inspection of the coefficients reveals that many regressors are not statistically significant. In particular, informal firms do not significantly differ from formal ones in their probability of engaging in R&D. Conversely, firm size emerges as a positive and significant predictor of R&D participation, while older firms and those exposed to bribes show a lower likelihood of engaging

Table 2 R&D input function

	Panel (a)		Panel (b)	
	Coefficient	Std. err	Coefficient	Std. err
Informal	−0.17	(0.109)	−0.506***	(0.087)
Manager education	−0.097	(0.157)	0.328***	(0.102)
On-the-job training	0.294	(0.236)	0.432***	(0.079)
Household	−0.207	(0.225)	−0.735***	(0.101)
Bribes	−0.179*	(0.097)	−0.327***	(0.096)
Liquidity	0.149	(0.252)	0.583***	(0.175)
Age	−0.009**	(0.004)	−0.036***	(0.006)
Employment	0.446***	(0.079)	0.101***	(0.037)
Constant	−0.276	(0.368)	−4.556***	(0.108)
IMR			0.639***	(0.117)
Pseudo R2	0.135		0.337	
sigma	0.045		0.765	(0.045)
lambda	−0.025	(0.033)		
rho (ρ)	−0.552			
Industry dummies	Yes		Yes	
Province dummies	Yes		Yes	
Observations	1050		1050	

In Panel (a), the dependent variable is a dummy equal to one if a firm invests in R&D, while in Panel (b), the dependent variable is the (log) R&D intensity. The likelihood of conducting R&D activities is estimated through a Probit model. The R&D Input Function is estimated through a Tobit model. All continuous variables are in logs. Robust standard errors in parentheses. ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1050 observations observed in the fiscal year 2010

in such activities. Most importantly, the assumption of correlation between the error terms in the selection and intensity equations is not supported. In the Heckman framework, this is reflected by the absence of significance of the coefficient ρ, which is defined as the ratio of the *lambda* coefficient to the standard error of the residuals (*sigma*). It suggests limited evidence of selection bias in R&D engagement. Therefore, the self-selection hypothesis appears to be rejected, and the inclusion of a correction term in the intensity equation is not strictly necessary. With respect to the R&D intensity (panel *b)*, and as expected, the results clearly show that informal firms allocate fewer resources to R&D compared to formal counterparts. Differently, the higher education of managers and on-the-job training positively affect the firm's R&D intensity. Since greater financial resources can push R&D investments, firms' liquidity also increases firms' R&D. Moreover, younger firms are more likely to produce a more intense effort in innovative activities. Inversely, household firms and those who incurred bribes show a lower R&D intensity. Finally, the size of the firm—measured as the level of employment—has a significant and positive impact on R&D. Nonetheless, the positive and significant coefficient of the IMR in the *R&D input function* indicates that, without the correction term, the estimated coefficients would have been larger in absolute value. In such a way, our prediction of the R&D intensity represents a more accurate and consistent proxy for the actual R&D investments made by formal and informal firms in Vietnam.

We now turn to the evaluation of the R&D-innovations relationship. The estimated coefficients of the *knowledge generation function* are reported in Table 3. First of all, our instrument—*customer certification requests*—is a significant predictor of the firms' probability to introduce *Radical* or *Incremental* innovation. As expected, the positive and significant effect of the R&D intensity on the firms' likelihood to innovate is evident, even though the negative sign of the squared term indicates a decreasing positive effect. This result is not a

Table 3 Knowledge generation function

	Coefficient	Std. err
Instrument	0.639*	(0.382)
R&D intensity 2010	42.841**	(19.540)
R&D intensity 2010 (squared)	−992.971***	(367.505)
Informal	0.056	(0.133)
Transition to formal 2010–2012	0.384**	(0.169)
Household	−0.032	(0.203)
Bribes	−0.114	(0.103)
Liquidity	0.312	(0.255)
Age	0.004	(0.005)
Employment	0.303***	(0.084)
Constant	−1.831***	(0.384)
Pseudo R2	0.128	
Industry dummies	Yes	
Province dummies	Yes	
Observations	1050	

The dependent variable is a dummy equal to one if a firm introduced Radical or Incremental innovations. The Knowledge Generation Function is estimated through a Probit model. As an instrument, we use the customer certification requests. R&D intensity is given by the predicted values of the first stage. All continuous variables are in logs. Robust standard errors in parentheses. ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1050 observations observed in the fiscal year 2012

surprise since most mainstream studies have already emphasized the importance of R&D investments on innovation. We also find that the dummies relative to household firms and bribe payments are not among the significant determinants of innovations, as well as their liquidity and age. Differently, larger firms show a higher probability of innovating. More important for our study are the two regressors linked to the firms' *informal* status and the firm's *transition to a formal status*. Indeed, while informal firms do not significantly differ from formal ones in their probability of introducing new products or improving the existing ones, those informal firms that switched to the formal economy show better performances in terms of innovation. At the same time, this result clearly shows that those firms that remain informal are less likely to introduce innovation in the two years following their R&D investments.

As far as the relationship between the R&D intensity and the probability of introducing innovation is concerned, we formally test this hump-shaped pattern by keeping the results of Table 3 and following the procedure proposed by Lind and Mehlum (2010). According to Haans et al. (2016), we confirm such a relationship because: (i) the coefficient of the squared term is significant and of the expected sign; (ii) the slope is sufficiently steep at both extremes of the data range; and (iii) the turning point is located within the data range. To account for the last two points, Lind and Mehlum's (2010) test first indicates that the R&D level (2.22%) at which the U-shaped relationship would peak falls in the data range (from 0 to about 11.07%). Then, it calculates the slope at the lower (+42.84) and upper limit (−164.99). Finally, it compares the null hypothesis of a monotone against a hump-shaped relationship. The result (*t-statistic* is 2.19) rejects the null hypothesis at any *p-value* higher than 0.014. So, with a significance level at 5%, this is another confirmation of our prior understanding that we have decreasing marginal returns of the R&D intensity.

Thus, our findings on the informality-R&D and R&D-innovation relationships are, hence, fully in line with the conclusions of Assefa et al. (2022), who found that Sub-Saharan firms are less likely to engage in innovation investments and that informality negatively affecting their ability to innovate due to lower R&D. Moreover, they also showed that formalization is likely to enhance firms' innovations.

Table 4 reports the estimated coefficients of the third stage. Consistent with most of the literature on the role of innovation, our empirical model shows a positive and significant impact of innovations on firms' productivity. In particular, the average marginal effect of innovations on productivity turns out to be + 1.006, which is the ratio of the percentage change in our dependent variable to the percentage change in the innovation probability (i.e. an elasticity measure).[13] Thus, it means that, for instance, a 10% increase in the innovation probability induces a 10.06% increase in the firm's labor productivity. Indeed, we support the widespread idea that more innovative firms experience higher labor productivity. However, we find evidence of a hump-shaped relationship also between the probability of introducing innovation and firms' productivity. Indeed, the positive coefficients of the first-order term and the negative one of the squared term show that, after a turning point (46.67%), the effect of innovations on productivity is still positive but in a decreasing way. Based on the estimated coefficients in Table 4, Fig. 1 provides a graphical representation of the overall effect of innovation on firms' productivity. We plot the productivity pattern for increas-

[13] To measure this elasticity, we calculated marginal effects through the partial derivative of the dependent variable with respect to the probability of innovation, keeping the other covariates at their sample means and using the average of the predicted probability of innovation.

Table 4 Productivity function

	Coefficient	Std. err
Innovation probability (radical and incremental)	1.864***	(0.544)
Innovation probability (squared)	−1.998***	(0.744)
Transition to formal 2010–2014	0.132*	(0.079)
Transition to formal 2010–2014 * Innovation	−0.680**	(0.326)
Employment	−0.012	(0.045)
Asset-per-employee	0.328***	(0.023)
International certification	0.967***	(0.220)
Risk propensity	0.317***	(0.088)
Bribes	0.230***	(0.052)
Liquidity	−1.677***	(0.181)
Age	−0.007***	(0.002)
Constant	7.805***	(0.315)
R2	0.357	
Industry dummies	Yes	
Province dummies	Yes	
Observations	1050	

The dependent variable is the (log) sales-per-employee. The *Productivity Function* is estimated through an *OLS model*. Innovation Probability is given by the predicted values of the second stage. All continuous variables are in logs. Robust standard errors in parentheses. ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1050 observations observed in the fiscal year 2014

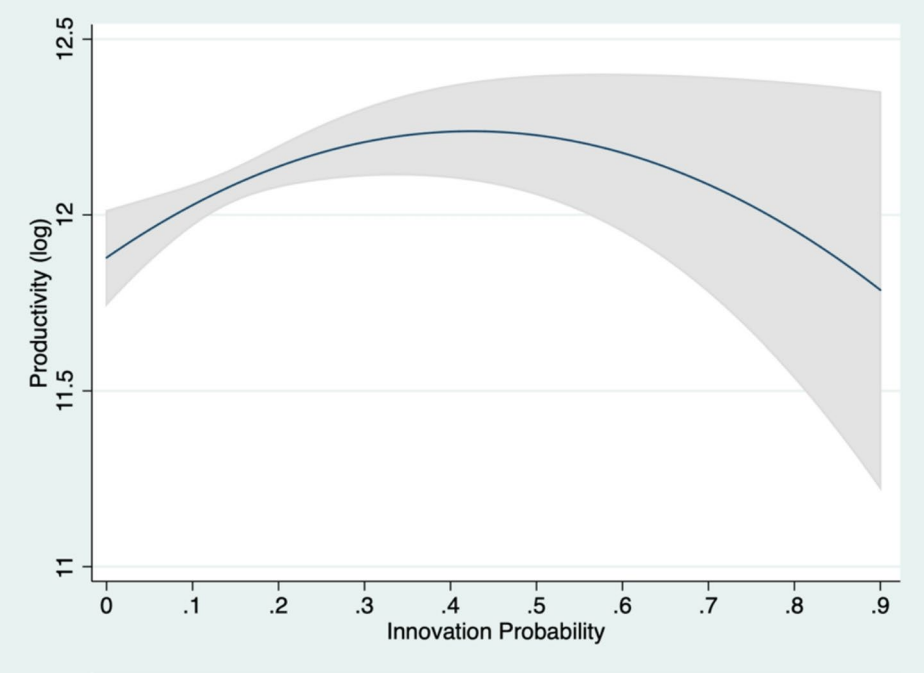


Fig. 1 Productivity trajectory and innovation

ing levels of the former. In particular, the probability of innovation ranges from 0 to its maximum (90.59%) while holding all other covariates at their sample means. The resulting figure displays the trajectory of the estimated productivity as a function of the probability of innovation, with a shaded confidence interval (95%). Nonetheless, to be absolutely sure of our finding, as before, we verify this hump-shaped relationship through Lind and Mehlum's (2010) test. The test yields three key findings. First, it confirms that the peak of the estimated inverted U-shaped relationship occurs when—*ceteris paribus*—the innovation probability equals 46.67%, which lies within the observed data range. Second, it computes the slopes at the lower extreme (+1.86) and the upper extreme (−1.75). Finally, by testing a monotonic relationship against a hump-shaped one, the analysis produces a *t-statistic* of 1.9, allowing us to reject the null hypothesis of monotonicity at a *p-value* < 0.029. Thus, at the 5% significance level, this provides further evidence of diminishing marginal returns of innovation on productivity.

Although of a small amount, the "*Transition to Formal*" status also exerts a positive direct effect on productivity, maybe reflecting gains from improved access to formal markets, credit, or legal protections.[14] However, this effect is moderated by a penalty of interaction with the probability of innovation. Indeed, the negative coefficient of the interaction term implies that formalization's benefits diminish as firms' innovation probability intensifies. This is potentially due to bureaucratic rigidities, compliance costs, or misalignment between formal status, the agility required for high-innovative environments, as well as the marginality of the effect measured, consistently with explanations argued by found by Cefis et al. (2025) for the same phenomenon found as size increases from micro-to small and medium size. Thus, both findings, jointly considered, underscore an interesting and hidden trade-off.

To explore this aspect more in-depth, we plot again the estimated productivity as a function of the probability of innovation, distinguishing between firms that remain informal and those that switched to the formal status. In both cases, productivity follows the already observed hump-shaped pattern. Nonetheless, Fig. 2 reveals a critical innovation probability *threshold*, T, (about 20%) where the two productivity trajectories of formal and informal firms intersect. Due to the direct advantages of formality, formal firms outperform below T. Beyond T, however, the interaction penalty dominates, and informal firms' productivity is higher than that of their formal counterparts. Thus, while formalization enhances productivity at low innovation levels, it imposes a hidden cost at high innovation intensities, where informal firms' flexibility becomes advantageous. These findings challenge one-size-fits-all approaches to formalization, and policymakers should recognize that formalization's benefits depend on the innovative profile of firms. For highly innovative firms, relaxing regulatory burdens or offering specific incentives for innovation may mitigate the negative trade-offs. Meanwhile, managers of informal firms should balance the maintenance of flexibility and compliance with official regulations to avoid productivity losses at high innovation levels. Hence, our results reveal that—against all odds—the *'underdog'* informal firms may have a higher level of productivity, the more likely they are to introduce radical or incremental innovations. Therefore, our results are highly compatible with those of Shekar

[14] By computing the marginal effect of the *transition to a formal status* on productivity, we find a value of +0.014. Thus, an informal firm increases its productivity by 0.014 percentage points by switching from an informal to a formal status.

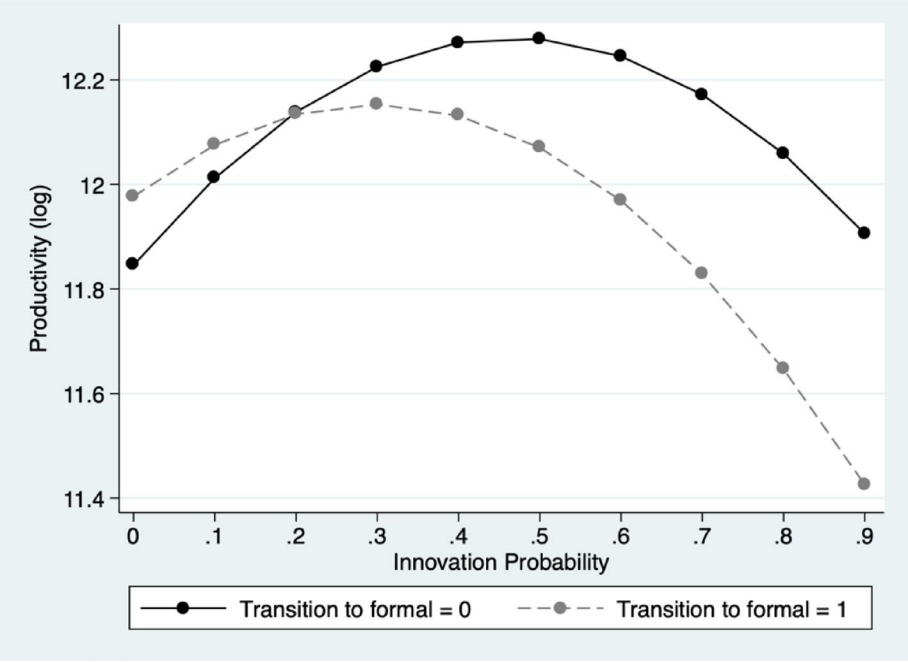


Fig. 2 Productivity Trajectories and Innovation by 'transition status'

(2025), who shows that the longer an Indian firm remains unregistered (i.e. informal), the greater is the probability of experiencing a productivity growth.

As a first robustness check, we replicate our estimation by calculating bootstrapped standard errors with 100 replications.

Results are reported in the appendix (Tables 8 and 9). The statistical significance of our estimates persists, reinforcing the reliability of our findings.[15] Since conventional robust standard errors may fail to capture true estimator variability in small samples or weakly identified models, the bootstrap method is particularly suited to our analysis because it mitigates possible underestimation of variability. Also, it amplifies potential instability caused by collinearity, outliers, or weak significance. Nevertheless, it produces confidence intervals that reflect true heterogeneity and, lastly, bootstrapped standard errors account for hypothetical misspecification of the models (e.g. omitted variables correlated with regressors). Thus, the stability of our results suggests our specification adequately captures the innovation-productivity dynamics of Vietnamese firms. In particular, the hump-shaped R&D-innovation and innovation-productivity relationships are still valid. At the same time, we remark on the existence of a *threshold T*, beyond which informal firms with high innovation intensities outperform formal ones in productivity. This is aligned with the evidence that formality imposes rigidity costs (e.g., compliance, reduced flexibility) detrimental to high-innovative firms. This mirrors findings in Ghana, where formal firms perform well in technological innovation, but informal firms exploit non-technological adaptations effectively (Fu et al.,

[15] While noting that in some estimates the instrumental variable is not significant at conventional levels (the p-value is right on the 10% significance threshold), we would like to recall that, as pointed out by Mairesse (2005), predicted R&D intensity already serves as a strong instrument for innovation.

2018). Our results also extend explanations for similar effects observed in studies on firm size (e.g., Baumann & Kritikos, 2016; Cefis et al., 2025) and the role of informal innovation (Mathew & Paily, 2022).

To explore these results more in-depth, we investigate how our estimations change if we distinguish between *radical* and *incremental* innovation. Specifically, we re-estimate both the second and third stages of the empirical model by disentangling how R&D investments affect the probability of introducing new products separately from those of improving the existing ones. Table 5 shows the results of the second stage, where the coefficients of the main regressors continue to hold the same signs as before, although their magnitudes are slightly different from those shown in Table 3. However, they still remain appropriately significant.

As for the third stage, Table 6 reports the estimation results on the *productivity function*, by distinguishing between the probability of introducing *radical* or *incremental* innovations. The signs of the coefficients are the same as those previously shown in Table 4. It is evident, however, that the magnitude of the effects of *radical* innovations is greater than that of *incremental* innovations. Specifically, while the average marginal effect of radical innovations on productivity amounts to 6.3, that of the incremental innovations sums to 0.95. On the other hand, the marginal effect of the *transition to a formal status* is pretty similar with respect to panel-b (+0.014) and panel-a (+0.010).

Hence, the hump-shaped relationship between innovation probability and productivity is confirmed and, also in this case, the negative interaction between innovation probability and formalization suggests that mandatory formalization is likely to dampen productivity gains for firms nearing or exceeding the above-mentioned innovation *threshold* T. This highlights critical challenges for informal firms. While informal enterprises benefit from organiza-

Table 5 Knowledge generation function—radical and incremental innovations

	(a) Radical innovations		(b) Incremental innovations	
	Coefficient	Std. err	Coefficient	Std. err
Instrument	0.756*	(0.424)	0.651*	(0.381)
R&D intensity 2010	96.637**	(39.373)	44.198**	(19.811)
R&D intensity 2010 (squared)	−1479.425**	(732.952)	−1018.047***	(379.814)
Informal	0.264	(0.395)	0.058	(0.134)
Transition to formal 2012	0.744**	(0.312)	0.370**	(0.169)
Household	−0.265	(0.408)	0.026	(0.206)
Bribes	0.053	(0.260)	−0.122	(0.103)
Liquidity	0.443	(0.407)	0.278	(0.261)
Age	0.014*	(0.008)	0.004	(0.005)
Employment	0.182	(0.233)	0.327***	(0.084)
Constant	−3.000***	(0.836)	−1.912***	(0.388)
Pseudo R2	0.172		0.124	
Industry dummies	Yes		Yes	
Province dummies	Yes		Yes	
Observations	1050		1050	

The dependent variable is a dummy equal to one if a firm introduced just *Radical* (Panel *a*) or just *Incremental* (Panel b) innovations. The *Knowledge Generation Function* is estimated through a *Probit model*. As an instrument, we use '*certification required by customers'*. R&D intensity is given by the predicted values of the first stage. All continuous variables are in logs. Robust standard errors in parentheses. ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1050 observations observed in the fiscal year 2012

Table 6 Productivity function—radical and incremental innovations

	(a) Radical innovations		(b) Incremental innovations	
	Coefficient	Std. err	Coefficient	Std. err
Innovation probability	8.229***	(2.231)	1.808***	(0.545)
Innovation probability (squared)	−40.143***	(13.002)	−1.994***	(0.750)
Transition to formal 2010–2014	0.048	(0.053)	0.138*	(0.079)
Transition to formal 2010–2014 * Innovation	−5.630***	(1.710)	−0.729**	(0.331)
Employment	0.022	(0.039)	−0.009	(0.045)
Asset-per-employee	0.330***	(0.023)	0.328***	(0.023)
International certification	1.001***	(0.194)	0.972***	(0.219)
Risk propensity	0.313***	(0.088)	0.318***	(0.088)
Bribes	0.238***	(0.052)	0.230***	(0.052)
Liquidity	−1.675***	(0.179)	−1.675***	(0.181)
Age	−0.007***	(0.002)	−0.007***	(0.002)
Constant	7.909***	(0.310)	7.808***	(0.315)
R2	0.357		0.357	
Industry dummies	Yes		Yes	
Province dummies	Yes		Yes	
Observations	1050		1050	

The dependent variable is the (log) sales-per-employee. The *Productivity Function* is estimated through an *OLS model*. Innovation Probability is given by the predicted values of the second stage: Panel (*a*) considers *Radical* innovations, Panel (b) *Incremental* innovations. All continuous variables are in logs. Robust standard errors in parentheses. ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1050 observations observed in the fiscal year 2014

tional flexibility, their productivity trajectory is constrained by structural barriers such as limited access to technology, financing, and institutional support. These findings are consistent with Montresor and Vezzani (2022) concerning the hampering role of the financial constraints inherent to their smaller size, likely exacerbated by their informal status. This latter may also prevent those firms from getting institutional support destined for formal counterparts. For example, due to the ineligibility for government funding and related programs. Therefore, as suggested by Colombelli et al. (2013), innovation policies oriented toward the promotion of clusters or technological platforms may mitigate technological uncertainty and other constraints and, in the case of informal firms, for instance, these policies would enable informal firms to tackle the uncertainty related to investments in innovations without forcing early formalization. That may explain the distinctive role of co-working spaces, a peculiar characteristic of Vietnam's entrepreneurial eco-system, against Western and neighboring counterparts (i.e., China, Japan, and Korea) in the Country's technology start-up race (Petti et al., 2021). Thus, even if small in magnitude, our results suggest that, while some formalization incentives may transform many informal firms into formal ones, such policies alone are neither sufficient nor necessary to increase these firms' productivity.

We have also re-estimated the empirical model to ensure its persuasiveness, especially for readers who may be skeptical about the validity of the earlier findings. As already mentioned in Sect. 3, we perform a further *robustness check* of our analysis using a different estimation methodology. In particular, we refer to a *full-maximum-likelihood* estimator of a simultaneous-equations system, where both firms' productivity and the *"transition to formality status"* are supposed to be endogenous. Specifically, to estimate the *productivity*

function at time $t+4$, we instrumented the *"transition to formality status"* with the *"total fees-and-taxes-per-employee"* paid by firms at time t.[16]

Table 7 clearly shows the stability of our results, even considering that a higher level of productivity could have a reverse causality effect, which may result in a driven factor of firms' *"transition to formality status"*. Thus, similarly to previous results, while innovations positively impact firms' productivity, formalization's benefits diminish as firms' innovation probability increases.

Hence, since our study found successful high-innovative informal firms more productive than formal ones, economic policies should encourage their transition to the formal sector, but they also should leave them a high level of flexibility. In this way, such policies could enhance firms' innovation, increase their productivity, and drive economic progress in Vietnam. In this context, our paper suggests policies fostering an economic environment where informal firms grow as successful innovators.

Table 7 Productivity function—simultaneous-equations system (full-maximum-likelihood estimator)

	Panel (a)		Panel (b)		Panel (c)	
	Coefficient	Std. err	Coefficient	Std. err	Coefficient	Std. err
Second stage: productivity function						
Innovation probability	1.918***	(0.627)	10.215***	(2.997)	1.826***	(0.631)
Innovation probability (squared)	−2.041**	(0.851)	−42.869**	(19.269)	−1.992**	(0.873)
Transition to formal 2010–2014	0.486**	(0.226)	0.415**	(0.211)	0.480**	(0.235)
Transition to formal 2010–2014 * Innovation	−0.696**	(0.349)	−7.344**	(2.965)	−0.750**	(0.359)
Employment	−0.024	(0.039)	−0.023	(0.036)	−0.020	(0.040)
Asset-per-employee	0.321***	(0.021)	0.319***	(0.021)	0.322***	(0.021)
International certification	1.042***	(0.314)	1.030**	(0.482)	1.044***	(0.314)
Risk propensity	0.318***	(0.086)	0.328***	(0.087)	0.319***	(0.086)
Bribes	0.234***	(0.051)	0.246***	(0.051)	0.234***	(0.051)
Liquidity	−1.662***	(0.137)	−1.665***	(0.138)	−1.660***	(0.137)
Age	−0.008***	(0.002)	−0.008***	(0.002)	−0.008***	(0.002)
Constant	7.811***	(0.291)	8.009***	(0.292)	7.815***	(0.291)
First stage: transition to formal 2010–2014						
Total fees and taxes per employee	0.149***	(0.023)	0.149***	(0.023)	0.149***	(0.023)
Sigma	−0.392***	(0.035)	−0.391***	(0.035)	−0.394***	(0.035)
Rho	−0.337*	(0.204)	−0.338*	(0.200)	−0.323	(0.212)
Industry dummies	Yes		Yes		Yes	
Province dummies	Yes		Yes		Yes	
Observations	1050		1050		1050	

The dependent variable in the *first stage* is a dummy equal to one if a firm switches from the informal to the formal status (2010–2014). The transition into a formal status is instrumented by the *"total fees-and-taxes-per-employee"* paid by firms at time t. Further details on the other regressors included in the first stage are available upon request. The dependent variable in the *second stage* is the (log) sales-per-employee. Innovation Probability is given by the predicted values of the *Knowledge Generation Function*: Panel (*a*) considers both *Radical* and *Incremental* innovation, Panel (b) just *Radical* innovations, Panel (c) just *Incremental* innovations. All continuous variables are in logs. Standard errors in parentheses. ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1050 observations observed in the fiscal year 2014

[16] As expected, the instrument used at the first stage (i.e. *Total fees and taxes*) hurts the probability of switching from a formal to an informal status.

Summing up, we found three significant results. First, informal firms—on average—invest less in R&D than formal firms. However, the higher education of managers and employee on-the-job training can push R&D investments, as well as greater financial resources. Second, those informal firms that do not switch to the formal sector are less likely to introduce innovation in the two years following their R&D investments. Third, we identify the existence of an innovation probability *threshold*, beyond which informal firms outperform formal ones in productivity.

Given our robust empirical evidence, we suggest policies that address the specific challenges of informal enterprises. For instance, these policies may encourage innovative behavior of informal firms and preserve their high degree of flexibility and adaptability. Since informal firms have close relationships with their local customer, this proximity provides valuable insights into customer preferences, enabling them to introduce product improvements that are better suited to the needs of the local target market. Then, they are usually even more responsive to changes in the market than formal firms. On the other hand, their lower costs could allow them to introduce innovations with less effort. Nevertheless, since informal firms face lower entry costs than formal firms, the fiercer competition of informal firms is often argued as potentially shrinking the incentives for formal firms to invest in innovative activities. However, if competition spurs innovation races, the net effect may be even positive. This is what Goel and Nelson (2025) very recently found in their extensive study of more than 100,000 firms in 132 countries, including Vietnam. While these authors referred to the need for formal firms to stay ahead of informal firms' competition, our findings offer another compelling reason.

Nonetheless, recognizing the positive contributions of such effects does not mean that governments should lean on legitimizing a kind of *laissez-faire*, cut-throat competition based on the *'survival of the fittest'*. This would indeed be a distorted signal that may hamper the virtuous cycle of improved regulation, formalization, perception of government quality, social cohesion, and ultimately sustained innovation and growth acquired in extant literature. It rather means that policymakers should look to reforms that balance the need for the integration of informal enterprises into the formal sector with the preservation of the lean, flexible, and adaptable characteristics typically associated with informal enterprises. At the same time, such policies should take into account any adverse effects on the firms' productivity in the formal sector. By doing so, such policies could have the chance to boost innovation, improve productivity, and drive the whole economic growth in emerging countries.

6 Discussion and conclusions

It is crucial to gain a deeper understanding of the impact of formalization and innovation on firms' performance in emerging countries. Informal firms in these countries, which employ a significant portion of the population, face many difficulties due to their low productivity and efficiency, as well as socio-economic and political obstacles that limit their access to resources and knowledge. In this challenging environment, innovation could be a key factor for their survival and improved productivity. Indeed, informal firms can play a significant role in terms of innovations, as our results reveal.

Overall, our results mainly show that: *(i)* informal firms invest less in R&D, and *(ii)* those who *switch to a formal status* have a higher probability of introducing innovations. Although the transition to formal status also exerts a positive direct effect on productivity, the formalization's benefits diminish as firms' innovation probability intensifies. However, while we found that formal firms are more likely to introduce innovations, our results also revealed that– against all odds—*(iii)* the *'underdog'* informal Vietnamese firms may have a higher level of productivity, the more likely they are to introduce radical or incremental innovations. The discovery of this interesting and hidden trade-off challenges one-size-fits-all approaches to formalization, and policymakers should recognize that formalization's benefits depend on the innovativeness levels of firms.

From the theoretical point of view, we hope to have provided three key contributions to the existing literature. First, our study extends the CDM framework by integrating informal sector dynamics and revealing hump-shaped R&D-innovation and innovation-productivity relationships. Also, it identifies the existence of an innovation probability *threshold*, beyond which informal firms outperform formal ones in productivity. In some way, this complements CDM's traditional focus on formal firms, demonstrating how formalization—by reducing firms' flexibility—could moderate the innovation-productivity linkages. To the best of our knowledge, these are novel mechanisms absent in prior CDM analyses. Second, our contribution adds the informality status to the recent evidence collected about *'underdogs'* unexpected performance linked to firms' size. Third, by expanding the CDM framework to informal firms in a developing context, our results also provide novel empirical evidence that integrates existing studies focused on advanced economies.

Often, firms' flexibility comes at the expense of long-term growth, as informal firms face many barriers, such as limited access to credit, restricted access to formal markets, or lack of legal protections. However, while some policy actions, such as easing the registration process or related costs, reducing or sustaining administrative and compliance provisions, may transform informal firms into formal ones, alone they are not sufficient to foster productivity in such firms. Hence, while formalization is a fundamental step in overcoming inefficiencies in developing economies, our findings suggest that complementary policies are required. Additional policies should be based on the recognition that in an informal economy, innovation is informal too. Therefore, it is as much as important to promote those systems of non-R&D innovation, such as the acquisition of advanced machinery, external technology and knowledge through licenses, consulting, outsourcing, market knowledge, and employee training and welfare expenditures (Ortega-Argiles et al., 2009; Mathew & Paily, 2022).

Thus, we suggest some direct and indirect ad hoc policies to reconcile formality and flexibility. Informal firms' *operational flexibility* may stem from labor market informality (ability to adjust workforces without rigid contracts), fiscal adaptability (avoiding tax compliance costs that divert resources from innovation), or regulatory avoidance (for instance, by quick implementation of business models not bound by licensing delays). As a consequence, informal firms are more responsive to changes in the market than formal firms, and they may quickly introduce new products or improve the existing ones. Therefore, policymakers should be aware that innovation is not just a result of economic growth but rather a way of development. Especially in the informal economy, innovations are often overlooked and not adequately supported. Therefore, flexibility can be a tool that policymakers can shape. By targeting specific channels of flexibility (labor, tax, regulation) with ad hoc policy instruments, governments could avoid the adverse effects of formality on

productivity. The first possibility is the introduction of tax systems in which enterprises pay reduced tax rates for initial formalization, which gradually increase as they grow. This model can increase formalization without reducing business flexibility. A second option is to create an *"innovation-driven formalization"* path by offering the formal status only if a firm participates in collaborative R&D centers. It could allow for the preservation of flexibility and granting access to potential public R&D funds. Another way would be to give a tax credit or exemption of formalization costs according to profits reinvested in innovative activities.

Even more importantly, the emergent empirical evidence about the more-than marginal role played by micro informal firms in innovation, productivity, and catching-up efforts opens to the need for innovation policies adapted to the peculiarity of those firms, like the assets' hyper and super depreciation reported in Cefis et al. (2025). If informal innovation patterns actually become highly productive, they could produce a strong impact and make government support worthwhile. With specific reference to the key results of this work about the productivity of *'underdogs'*, it seems particularly relevant and reasonable to assume that an injection of public resources in these firms could allow them to undertake an effective technological upgrading process and eventually become more significant than formal innovators. According to our findings, such *'underdogs'* should become the target of the ad hoc, and *'time-limited'*, designed policy measures. The *'time-limited'* attribute of these policies refers to the need to avoid that the short-term positive effects on informal firms' innovation produce, in the long-run, a business stealing effect on formal firms' innovation (Goel & Nelson, 2025) or an unfair competition against them (Baily et al., 2006). In this last regard, it would be also important to leverage on other resources and initiatives aimed at both informal and formal firms, like promoting specific capacity building and vocational training programs, supporting the linkages with local foreign invested enterprises or the access to foreign markets, as well as to traditional R&D-based targeted incentives like Italy's Innovative Start-up/SMEs or, closer to Vietnam's context, the Chinese small giants programs. The collaborative programs with specific countries and banks set up by the Vietnamese government are in the right direction, as well as the Vietnamese incubator model and the promotion of co-working spaces reported in Petti et al. (2021). Pro-collaboration measures of this latter kind may preserve and even leverage the flexibility of informal firms while reducing their costs, enabling partnering, distributed-open innovation, and flexible production models. These actions provide not only further funding, but also develop innovation programs, foster technical training, and provide business mentoring. With the same target, programs such as the Guangdong Technology Expert Secondment Program (TESP) foresee the secondment of university personnel in SMEs.

Turning to the limitations of this study, we admit that our results are, at the moment, limited to the context of Vietnam only. Second, we believe that cross-sectional data limit causal linkages about innovation- and productivity- formalization trajectories. Third, although it does not affect the credibility of the correlation between the key variables, we must recognize that the availability of an observed measure of innovation intensity would also make the magnitude of the estimated coefficients and the innovation probability *threshold* more reliable.

Finally, we would suggest three main lines of future research. On the one hand, an analysis applied to other emerging countries and/or with more recent data could assess the spatial–temporal extent of the R&D-innovation-productivity dynamics highlighted in our study. On the other hand, studies on longitudinal data could analyze how informal firms balance the benefits of innovation with the constraints of formalization over time. Finally, it would be very timely to investigate the role of artificial intelligence and advanced digital tools in helping informal firms to avoid traditional formalization barriers while maintaining their high flexibility.

Appendix

See Tables 8 and 9.

Table 8 Knowledge generation function—bootstrapped standard errors

	(a) Innovation		(b) Radical innovations		(c) Incremental innovations	
	Coefficient	Std. err	Coefficient	Std. err	Coefficient	Std. err
Instrument	0.639	(0.458)	0.756*	(0.447)	0.651	(0.456)
R&D intensity 2010	42.841*	(23.439)	96.637**	(47.993)	44.198*	(23.450)
R&D intensity 2010 (squared)	−992.971*	(565.815)	−1479.425*	(873.163)	−1018.047*	(583.266)
Informal	0.056	(0.172)	0.264	(0.556)	0.058	(0.173)
Transition to formal 2012	0.384**	(0.174)	0.744**	(0.368)	0.370**	(0.170)
Household	−0.032	(0.238)	−0.265	(0.497)	0.026	(0.242)
Bribes	−0.114	(0.096)	0.053	(0.241)	−0.122	(0.096)
Liquidity	0.312	(0.319)	0.443	(1.906)	0.278	(0.340)
Age	0.004	(0.005)	0.014	(0.011)	0.004	(0.005)
Employment	0.303***	(0.088)	−0.182	(0.241)	0.327***	(0.090)
Constant	−1.831***	(0.431)	−3.000***	(0.880)	−1.912***	(0.436)
Pseudo R2	0.128		0.172		0.124	
Industry dummies	Yes		Yes		Yes	
Province dummies	Yes		Yes		Yes	
Observations	1050		1050		1050	

The dependent variable is a dummy equal to one if a firm introduced *Radical* or *Incremental Innovations* (Panel *a*), just *Radical* (Panel *b*), or just *Incremental* (Panel *c*) innovations. The *Knowledge Generation Function* is estimated through a *Probit model*. As an instrument, we use '*certification required by customers'*. R&D intensity is given by the predicted values of the first stage. All continuous variables are in logs. Bootstrapped Standard Errors in parentheses (100 complete replications). ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1,050 observations observed in the fiscal year 2012

Table 9 Productivity function—bootstrapped standard errors

	(a) Innovation		(a) Radical innovations		(b) Incremental innovations	
	Coefficient	Std. err	Coefficient	Std. err	Coefficient	Std. err
Innovation probability	1.864***	(0.560)	8.229***	(2.508)	1.808***	(0.574)
Innovation probability (squared)	−1.998**	(0.913)	−40.143**	(19.266)	−1.994**	(0.943)
Transition to formal 2010–2014	0.132*	(0.075)	0.048	(0.051)	0.138*	(0.075)
Transition to formal 2010–2014 * Innovation	−0.680**	(0.312)	−5.630***	(2.110)	−0.729**	(0.315)
Employment	−0.012	(0.049)	0.022	(0.042)	−0.009	(0.050)
Asset-per-employee	0.328***	(0.020)	0.330***	(0.021)	0.328***	(0.020)
International certification	0.967***	(0.257)	1.001***	(0.219)	0.972***	(0.254)
Risk propensity	0.317***	(0.088)	0.313***	(0.091)	0.318***	(0.088)
Bribes	0.230***	(0.057)	0.238***	(0.056)	0.230***	(0.057)
Liquidity	−1.677***	(0.167)	−1.675***	(0.165)	−1.675***	(0.167)
Age	−0.007***	(0.002)	−0.007***	(0.002)	−0.007***	(0.002)
Constant	7.805***	(0.302)	7.909***	(0.299)	7.808***	(0.301)
R2 (Adjusted)	0.341		0.342		0.342	
Industry dummies	Yes		Yes		Yes	
Province dummies	Yes		Yes		Yes	
Observations	1050		1050		1050	

The dependent variable is the (log) sales-per-employee. The *Productivity Function* is estimated through an *OLS model*. Innovation Probability is given by the predicted values of the second stage: Panel (*a*) considers *Radical* or *Incremental* innovation, Panel (b) just *Radical* innovations, Panel (c) just *Incremental* innovations. All continuous variables are in logs. Bootstrapped Standard Errors in parentheses (100 complete replications). ***, **, *denote, respectively, significance levels at 1%, 5% and 10%. Estimates are carried out on a sample of 1050 observations observed in the fiscal year 2014

Author contributions M.D. wrote the main manuscript text and performed the econometric and data sections. Both Authors contributed to literature background and discussions. All authors reviewed the manuscript.d the manuscript. All authors reviewed the manuscript.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 872618. Open access funding provided by the University of Salento within the CRUI-CARE Agreement.

Disclaimer The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed lies entirely with the authors.

Data availability The data that support the findings of this study are openly available in "UNU-Wider—Viet Nam Small and Medium Enterprises data" at: https://www.wider.unu.edu/database/viet-nam-sme-database.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material.

If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acs, Z., Desai, S., Stenholm, P., & Wuebker, R. (2013). Institutions and the rate of formal and informal entrepreneurship across countries. *Frontiers of Entrepreneurship Research, 35*(15), 1–24.

Antonelli, C., & Colombelli, A. (2015). External and internal knowledge in the knowledge generation function. *Industry and Innovation, 22*(4), 273–298.

Antonelli, C., & Feder, C. (2021). Knowledge appropriability and directed technological change: The Schumpeterian creative response in global markets. *Journal of Technology Transfer, 46*, 686–700.

Assefa, D. Z., Liao, C. T., & Misganaw, B. A. (2022). Unpacking the negative impact of initial informality on innovation: The mediating roles of investments in R&D and employee training. *Technovation, 114*, 102455.

Autio, E., & Fu, K. (2015). Economic and political institutions and entry into formal and informal entrepreneurship. *Asia Pacific Journal of Management, 32*(1), 67–94.

Baily, M., Farrell, D., & Remes, J. (2006). The hidden key to growth. *The International Economy, 20*(1), 48.

Baumann, J., & Kritikos, A. S. (2016). The link between R&D, innovation and productivity: Are micro firms different? [Article]. *Research Policy, 45*(6), 1263–1274.

Baumol, W. J. (1990). Entrepreneurship: Productive, unproductive, and destructive. *Journal of Political Economy, 98*, 893–921.

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

Calza, E., Goedhuys, M., & Trifković, N. (2018). Drivers of productivity in Vietnamese SMEs: The role of management standards and innovation. *Economics of Innovation and New Technology, 28*(1), 23–44.

Carson, R. T., & Sun, Y. (2007). The tobit model with a non-zero threshold. *The Econometrics Journal, 10*, 488–502.

Cefis, E., Scrofani, S., & Tubiana, M. (2025). Innovation policies and firms' productivity: The Italian Industry 4.0 Plan for digital transformation. *The Journal of Technology Transfer*. https://doi.org/10.1007/s10961-024-10179-2

Choi, S. B., Feng, Y., Liu, J., & Zhu, Q. (2019). Motivating corporate social responsibility practices under customer pressure among small-and medium-sized suppliers in China: The role of dynamic capabilities. *Corporate Social Responsibility and Environmental Management, 26*(1), 213–226.

Coccia, M. (2017). Sources of technological innovation: Radical and incremental innovation problem-driven to support competitive advantage of firms. *Technology Analysis and Strategic ManagemeNt, 29*(9), 1048–1061.

Colombelli, A., Krafft, J., & Quatraro, F. (2013). Properties of knowledge base and firm survival: Evidence from a sample of French manufacturing firms. *Technological Forecasting and Social Change, 80*(8), 1469–1483.

Conte, A., Vivarelli, M. (2005). One or many knowledge production functions? Mapping innovative activity using microdata, IZA Discussion Papers, No.1878.

Cowan, R., & van de Paal, G. (2000). *Innovation policy in the knowledge-based economy*. European Commission DG-Enterprise.

Crépon, B., Duguet, E., & Mairesse, J. (1998). Research, innovation and productivity: An econometric analysis at the firm level. *Economics of Innovation and New Technology, 7*(2), 115–158.

Crespi, G., Tacsir, E., Vargas, F. (2016). Innovation Dynamics and Productivity: Evidence for Latin America. In: Inter-American Development Bank, Grazzi, M., Pietrobelli, C. (eds) Firm Innovation and Productivity in Latin America and the Caribbean. *Palgrave Macmillan*, New York.

De Mel, S., McKenzie, D., & Woodruff, C. (2013). The demand for, and consequences of, formalization among informal firms in Sri Lanka. *American Economic Journal: Applied Economics, 5*(2), 122–150.

Di Cintio, M., & Grassi, E. (2017). Uncertainty, flexible labour relations and R&D. *Metroeconomica, 68*, 91–120.

Fajnzylber, P., Maloney, W. F., & Montes-Rojas, G. V. (2011). Does formality improve micro- firm performance? Evidence from the Brazilian SIMPLES program. *Journal of Development Economics, 94*(2), 262–276.

Fandl, K. J., & Bustamante Izquierdo, J. (2016). Incentivizing gray market entrepreneurs in emerging markets. *Journal of International Law and Business, 37*, 16–34.
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.
Gaglio, C., Kraemer-Mbula, E., & Lorenz, E. (2022). The effects of digital transformation on innovation and productivity: Firm-level evidence of South African manufacturing micro and small enterprises. *Technological Forecasting and Social Change, 182*, 121785.
Gebreeyesus, M. (2009). Innovation and microenterprises growth in Ethiopia. In: Szirmai, W.A., Naudé, A., Goedhuys, M. Entrepreneurship, Innovation, and Economic Development. Oxford University Press, Oxford (UK).
Giotopoulos, I., Kritikos, A. S., & Tsakanikas, A. (2023). A lasting crisis affects R&D decisions of smaller firms: The Greek experience. *The Journal of Technology Transfer, 48*, 1161–1175.
Goedhuys, M., Janz, N., & Mohnen, P. (2008). What drives productivity in Tanzanian manufacturing firms: Technology or business environment? *European Journal of DevElopment Research, 20*, 199–218.
Goedhuys, M., Janz, N., & Mohnen, P. (2014). Knowledge-based productivity in "low-tech" industries: Evidence from firms in developing countries. *Industrial and Corporate Changes, 23*, 1–23.
Goel, R. K., & Nelson, M. A. (2025). Is adversity the mother of innovation? International evidence at the firm level. *Journal of Technology Transfer*. https://doi.org/10.1007/s10961-024-10159-6
Goldszmidt, R., Behr, P., Zucco, C., Lenz, A. K., Farias, L. E. G., Valdivia, M., & Grade, P. (2022). Microcredit impacts: Evidence from a large-scale observational study in Brazil. *European Journal of Development Research, 34*, 1940–1966.
Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics., 10*(1), 92–116.
Guo, L., Zhang, M. Y., Dodgson, M., & Gann, D. (2019). Huawei's catch-up in the global telecommunication industry: Innovation capability and transition to leadership. *Technology Analysis and Strategic ManageMeNt, 31*(12), 1395–1411.
Haans, R. F., Pieters, C., & He, Z. L. (2016). Thinking about U: Theorizing and testing U-and inverted U-shaped relationships in strategy research. *Strategic Management Journal, 37*(7), 1177–1195.
Hall, B. H., Maffioli, A. (2008). Evaluating the Impact of Technology Development Funds in Emerging Economies: Evidence from Latin America. *NBER Working Paper* No. 13835.
Hall, B. H., Lotti, F., & Mairesse, J. (2009). Innovation and productivity in SMEs: Empirical evidence for Italy. *Small Business Economics, 33*(1), 13–33.
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). The confounding influence of urban informality on innovation and production specialization in production clusters: Evidence from Nairobi. *African Journal of Science, Technology, Innovation and Development, 6*(6), 529–539.
Harris, J. (2014c). Meeting the challenges of the handicraft industry in Africa: Evidence from Nairobi. *Development in Practice, 24*(1), 105–117.
Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*. https://doi.org/10.2307/1912352
Hobday, M. (1995). East Asian latecomer firms: Learning the technology of electronics. *World Development, 23*(7), 1171–1193.
Ingram, M., Ramachandran, V., Desai, V. (2007). Why Do Firms Choose to Be Informal? Evidence from Enterprise Surveys in Africa. Manuscript. *The World Bank*.
International Monetary Fund (2017). Perspectives économiques régionales: Afrique subsaharienne, faire redémarrer la croissance. *Technical report, International Monetary Fund* (IMF).
Kedia, B. L., & Bhagat, R. S. (1988). Cultural constraints on transfer of technology across nations: Implications for research in international and comparative management. *The Academy of Management Review, 13*(4), 559–571.
Klapper, L., Laeven, L., & Rajan, R. (2006). Entry regulation as a barrier to entrepreneurship. *Journal of Financial Economics, 82*(3), 591–629.
Kleinknecht, A. (1987). Measuring r&d in small firms: How much are we missing? *The Journal of Industrial Economics, 36*(2), 253–256.
Korean Development Institute. (2020). Vietnam National Productivity Plan – Toward Innovation-driven Development. *Technical Report, Korean Development Institute* (KDI).
La Porta, R., & Shleifer, A. (2014). Informality and development. *The Journal of Economic Perspectives, 28*(3), 109–126.
Lind, J. T., & Mehlum, H. (2010). With or without u? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics, 72*(1), 109–118.

Liu, X. (2008). China's Development Model: An Alternative Strategy for Technological Catch-Up. SLPTMD Working Paper Series No.20.
Loayza, N. V. (2016). Informality in the process of development and growth. *World Economy, 39*(12), 1856–1916.
Mahemba, C. M., & Bruijn, E. J. (2003). Innovation activities by small and medium-sized manufacturing enterprises in Tanzania. *Creativity and Innovation Management, 12*, 162–173.
Mairesse, J., Mohnen, P., Kremp, E., & Kremp, E. (2005). The importance of R&D and innovation for productivity: A reexamination in light of the French innovation survey. *Annales D'économie Et De Statistique, 79*(80), 487–527.
Mairesse, J., Mohnen, P., & Notten, A. (2025). Innovation and productivity: The recent empirical literature and the state of the art. *Eurasian Business Review, 15*, 1–27.
Mathew, N., & Paily, G. (2022). STI-DUI innovation modes and firm performance in the Indian capital goods industry: Do small firms differ from large ones? *The Journal of Technology Transfer, 47*(2), 435–458.
Mathews, J. A. (2002). Competitive advantages of the latecomer firm: A resource-based account of industrial catch-up strategies. *Asia Pacific Journal of Management, 19*(4), 467–488.
Medda, G. (2020). External R&d, product and process innovation in European manufacturing companies. *Journal of Technology Transfer, 45*, 339–369.
De Mel, S., McKenzie, D., Woodruff, C. (2009). Innovative firms or innovative owners? Determinants of innovation in micro, small, and medium enterprises. *World Bank Policy Research* - Working Paper 4934. The World Bank, Washington, DC (US).
Mendi, P., & Mudida, R. (2018). The effect on innovation of beginning informal: Empirical evidence from Kenya. *Technological Forecasting and Social Change, 131*, 326–335.
Monteiro, J. C., & Assunção, J. J. (2012). Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises. *Journal of Development Economics, 99*(1), 105–115.
Montresor, S., & Vezzani, A. (2022). Financial constraint.s to investing in intangibles: Do innovative and non-innovative firms differ? *Journal of Technology Transfer, 47*, 1–32.
Nguimkeu, P. (2014). A structural econometric analysis of the informal sector heterogeneity. *Journal of Development Economics, 107*, 175–191.
Ortega-Argilés, R., Vivarelli, M., & Voigt, P. (2009). R&D in SMEs: A paradox? *Small Business Economics, 33*, 3–11.
Oyelaran-Oyeyinka, B., Laditan, G. O. A., & Esubiyi, A. O. (1996). Industrial innovation in sub-Saharan Africa: The manufacturing sector in Nigeria. *Research Policy, 25*, 1081–1096.
Oyelaran-Oyeyinka, B., & Lal, K. (2006). Learning new technologies by small and medium enterprises in developing countries. *Technovation, 26*, 220–231.
Petti, C., Nguyen Dang Tuan, M., Nham Phong, T., Pham Thi, M., Ta Huong, T., & Perumal, V. V. (2021). Technological catch-up and innovative entrepreneurship in Vietnamese firms. *Administrative Sciences, 11*(3), 100.
Petti, C., Rubini, L., & Podetti, S. (2017). Government support and R&d investment effectiveness in Chinese SMEs: A complex relationship. *Asian Economic Papers, 16*(1), 201–226.
La Porta, R., Shleifer, A. (2008). The unofficial economy and economic development. *Brookings Papers on Economic Activity*. 275–352.
Prahalad, C. K. (2012). Bottom of the pyramid as a source of breakthrough innovations. *Journal of Product Innovation Management, 29*, 6–12.
Robson, P. J. A., Haugh, H. M., & Obeng, B. A. (2009). Entrepreneurship and innovation in Ghana: Enterprising Africa. *Small Business Economics, 32*, 331–350.
Rocha, R., Ulyssea, G., & Rachter, L. (2018). Do lower taxes reduce informality? Evidence from Brazil. *Journal of Development Economics, 134*, 28–49.
Roper, S. (1999). Under-reporting of R&d in small firms: The impact on international R&d comparisons. *Small Business Economics, 12*(2), 131–135.
Rukundo, J. B. (2017). Firm performance and innovation in the developing countries: Evidence from firm-level survey. *Corporate Ownership & Control, 15*(1), 235–245.
Santarelli, E., & Sterlacchini, A. (1990). Innovation, formal vs informal R&D, and firm size: Some evidence from Italian manufacturing firms. *Small Business Economics, 2*(3), 223–228.
Schneider, F., Buehn, A., Montenegro, C.E. (2011). Shadow Economies all Over the World: New Estimates for 162 Countries from 1999 to 2007. Handbook on the Shadow Economy 9–77.
Schneider, F., & Enste, D. (2000). Shadow economies: Size, causes and consequences. *Journal of Economic Literature, 38*(1), 77–114.
Shekar, K. C. (2025). Informality, innovation, and firm performance: evidence from World Bank Enterprise Survey in India, 2013–2014. *Journal of Innovation and Entrepreneurship, 14*, 17.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy, 15*(6), 285–305.
Tijdens, K., Besamusca, J., & Van Klaveren, M. (2015). Workers and labour market outcomes of informal jobs in formal establishments. A job-based informality index for nine sub-Saharan African countries. *European Journal of Development Research, 27*(5), 868–886.
Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources, 33*(1), 127–169.
Von Hippel, E. (2005). *Democratizing innovation*. MIT Press.
Wooldridge, M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics, 68*(1), 115–132.
World Bank Group and the Ministry of Planning and Investment in Vietnam. (2016). Vietnam 2035. Toward Prosperity, Creativity, Equity and Democracy. *Technical Report, World Bank Group.*
Wu, X., Ma, R., & Xu, G. (2009). Accelerating secondary innovation through organizational learning: A case study and theoretical analysis. *Industry and Innovation, 16*(4/5), 389–409.
Wunsch-Vincent, S., & Kraemer-Mbula, E. (2016). *The Informal economy in developing nations: Hidden engine of innovation?* Cambridge University Press.
Yuki, K. (2007). Urbanization, informal sector, and development. *Journal of Development EconOmics, 84*(1), 76–103.
Zhang, J., & Islam, M. S. (2022). The role of market power in driving innovation and productivity: A firm-level study of emerging ASEAN. *International Journal of Emerging Markets, 17*(8), 1865–1888.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.