Build or Buy?
Human capital and corporate diversification

Paul Beaumont - Camille Hebert - Victor Lyonnet*

June 30, 2018

Abstract

Why do some firms enter a new market by acquiring an existing company operating in this market (external entry), while more than 90% of firms do so by building on their existing resources (internal entry)? This paper shows that human capital acquisition is a key motive behind firms' decision to grow externally. We develop a simple production model in which firms use heterogeneous worker teams to execute tasks. We derive a firm-level measure of human capital, which we estimate using a novel dataset constructed by merging French employer payrolls with commercial M&A data sets. Human capital acquisition explains firms' chosen mode of entry: firms whose human capital is adequate for the new market favor internal entry, while those whose human capital is not adequate for the new market favor external entry.

*Paul Beaumont is from Université Paris Dauphine, and CREST. Camille Hebert is from Université Paris Dauphine and Tilburg University. Victor Lyonnet is from Ohio State University. We thank Manuel Adelino, Philippe Aghion, Fabio Braggion, Edouard Challe, Gilles Chemla, Espen Eckbo, Simon Gervais, Edith Ginglinger, John Graham, Denis Gromb, Julien Martin, Adrien Matray, Isabelle Mejean, Mario Milone, Tamara Nefedova, Paige Ouimet, Manju Puri, Adriano Rampini, Luc Renneboog, David Robinson, Geoff Tate (discussant), and Vish Viswanathan for valuable comments and helpful discussions. We thank also seminar participants at the 2018 ECGC workshop, the 2018 Mitsui Finance Symposium, CREST, Duke Fuqua, INSEE, Université Paris Dauphine, and Tilburg.

1 Introduction

Merger and acquisitions are typically associated with the large premiums paid in acquisitions of publicly traded targets (e.g., Schwert, 1996), and negative bidder returns on the announcement of acquisitions (Roll, 1986). Yet, under the neoclassical view of M&A activity, efficiency-seeking corporate takeovers can create value (Maksimovic et al., 2011). One important line of argument is that M&As can facilitate the efficient allocation of human capital when assembling new worker teams is too costly or takes too much time. Gaughan (2013) analyses the M&A strategy of Cisco Systems, Inc., explicitly aimed at acquiring human capital, sometimes paying up to 2 million per worker. As described by a Cisco analyst, *"in today's economy, building work teams from scratch can be yesterday's luxury. So, when you can't build fast enough, you buy."* (Wysocki, 1997).

In this paper, we study how human capital shapes firms' decisions when entering a new market. A firm can enter the new market by building on its existing resources (internal entry), or by acquiring an existing company already operating in this market (external entry). The trade-off is as follows. With internal entry, the firm's human capital might not be adapted to the market of entry. With external entry, the firm has to pay the acquisition premium to takeover an existing firms' human capital. We test the hypothesis that firms choose to enter a new market internally (externally) when their pre-entry human capital is (not) adapted to the market of entry.

Why would human capital considerations lead firms to enter new markets through an acquisition? This prediction can be rationalized by three main mechanisms. First, worker skills are sector-dependent, therefore the transferability of "occupation-specific human capital" is uncertain (Autor and Dorn, 2009), making internal entry risky. Second, given worker heterogeneity and costly search for firms, hiring the right workers for the new market's sector can be excessively costly (e.g., Albrecht and Axell (1984)). Third, firm-specific organizational capital can only be partially transferable across firms (e.g., Atkeson and Kehoe (2005)), so that acquiring an existing firm's human capital might turn out to be a better alternative than poaching workers while separating them from their current team or employer.

One first difficulty consists in defining human capital at the firm level: human capital is neither directly observable nor easily defined. To overcome this problem, we propose a simple production model in which firms use heterogeneous labor teams to execute tasks in a given sector. The model yields a firm-level measure of human capital, that is readily estimable using firm-level data on workforce at the occupation level. We construct it in two steps.

First, in the model, we assume that worker occupations depend on their skills. We find that in any given sector, firms split the wage bill across all occupations in accordance with an occupations' efficiency in the production of goods in this sector. Therefore in the first step, we use data from the French Statistical Institute (INSEE),[1] and regress the fraction of the wage bill of all firms in a sector devoted to each occupation, on occupation $\times$ sector fixed effects (following the methodology of Abowd et al. (1999)).[2] We obtain fixed-effects estimates, which through the lens of our model we interpret as a score: the higher the fixed effect of a given occupation in a sector, the more efficient this occupation in this sector. Thereby, our first step obtains a first ranking of all worker occupations, according to their importance for every sector of the French economy.

Second, we use our ranking of occupations in each sector to construct a firm-level measure of human capital. For each firm, we use the fixed effects estimated in the first step to compute their average value across all occupations available at the firm prior to entry. We obtain a first measure of human capital at the firm-level, that is specific to the market of entry. Henceforth for clarity, we refer to this measure as human capital without explicitly mentioning that it is specific to a given firm relative to a product market of entry.

Our measure of human capital presents several advantages. First, it is comprehensive of all worker occupations across all sectors of the French economy without any selection bias. Second, the fixed-effect estimation encompasses various determinants of sector-specific human capital, including worker skills, education or experience which are all reflected in wages at equilibrium. Third, it controls for firm fixed effects in estimating the occupation-sector effects, and is immune to firm-specific compensation policies.

Using our constructed measure of human capital at the firm level, we test whether this measure explains firms' decisions when entering a new market. We find evidence consistent with the view that human capital drives the decision of product market entry. Firms favor external entry when their *pre-existing* human capital is not adapted to the market of entry. This finding holds when controlling for size, profitability, capital intensity, access to internal funds and dependence on external finance. We find that the link between human capital and entry type is more pronounced for larger groups, consistent with the view that acquisition costs are more affordable for larger groups.[3] These results are robust to the introduction of time fixed effects, sector of entry, and sector of origin fixed effects. Our results are therefore not driven by sector-specific characteristics, such as entry costs or asset collateralizability. We interpret this finding as evidence that human capital drives firms' choice between internal and external entry. This finding is consistent with the view that M&As enable firms to acquire and adapt their human capital to their market of entry. In contrast, firms entering a market internally can use their preexisting human capital,

[1] The data contains the distribution of wages across occupations and sectors on all employer-employee relationships in France in the period 2008-2014. See below for a detailed description of the dataset.

[2] This methodology enables us to purge these estimates from firm fixed-effects (see Section 4.1).

[3] This can stem from access to internal capital markets, or easier access to external finance.

which is relatively more adapted the new market of entry.

We conduct additional tests to investigate the role of human capital in determining firms' mode of entry. To study whether the effect of human capital varies according to the market of entry, we control for the endogenous choice of the sector of diversification arguably driven by sectoral synergies between the market of entry and the acquiror's initial activities.[4] We use a modified version of sectoral distances by Bloom et al. (2013).[5] In the data, firms in a sector that is distant (close) to the market of entry are more likely to enter this market externally (internally). The more distant firms are from the sector of entry, the more significant and quantitatively large the role of human capital in determining entry type. Finally, we find that firms entering internally adjust the composition of their workforce in the years preceding diversification, adapting their human capital to the sector of entry. In contrast, firms entering externally do not.

Our results are robust to a range of robustness tests. First, we test varieties of our human capital measure: (i) we focus on "top ten occupations" for the market of entry, (ii) we regress the mode of entry on a dummy equal to one if the firm does not employ any of the ten occupations with the highest score for the sector of entry, (iii) we withdraw CEO-like occupations, (iv) we focus solely on occupation's fraction of the total (sectoral) wage bill instead of the fixed-effect estimation, and (v) we estimate fixed effects at the business group level and at the establishment level (vi). Our main result holds true across those different measures of human capital. In addition, for the sub-sample of internal entries, we consider human capital of each firms within a group and predict that the entry into the market of diversification is made though the entity whose human capital fits the most the required work force of the sector of diversification. We find empirical validation for this prediction and this adds credit to ability of our measure to capture human capital similarities between a firm and a market.

Second, we address the concern that the choice of the sector of diversification is not random, reflecting selection for potential unobservable synergies with the sector of origin. In the spirit of Paravisini et al. (2014), we control for sector of entry $\times$ sector of origin fixed effects $\times$year, dropping singletons of sector pairs and allowing the comparison of the alternative modes of entry driven by the same diversification considerations at the same point in time. Our results continue to hold and we find that even across firms of a given sector that enter in the an identical other sector, those firms that lack human capital enter the new market through an acquisition.

Finally, we try to account for the endogeneity of firms' human capital to the mode of

[4]Examples of potential sectoral synergies include product market complementarities, customer bases, production synergies, etc.

[5]This measure captures product market similarities by estimating the most likely combinaison of product markets.

entry. Since firms modify their workforce prior to an internal entry, we regress firms' mode of entry on *lagged* measures of human capital. The implicit assumption is that firms do not adjust the composition of their workforce years in advance but only marginally the year before the entry. Again, we find that past human capital predicts firms' mode of entry. Despite the workforce adjustments that we observe, the evidence suggests that firms' human capital is persistent and explains firms' mode of diversification across sectors at least three years in advance.

There are two main strands of the literature this paper contributes to: the literature on corporate diversification, and that on human capital.

Prior research has focused on market-level characteristics to predict the mode of entry. Similarly to Yip (1982), who shows that firms are more likely to enter internally a market with low barriers to entry, we find that firms are more likely to enter highly concentrated markets externally. Lee and Lieberman (2010) develop a time-varying measure of market relatedness and show that firms are more likely to enter unrelated sectors by acquisition. In the same spirit, we use a version of Bloom et al. (2013)'s sectoral distance measure and find that acquisitions are more frequent when firms are distant from their market of entry. In contrast to the existing literature, we show that human capital at the firm level can explain how firms enter a new market. Our conclusions are in line with McCardle and Viswanathan (1994), who show that firms with a relatively higher entry cost in a new market – in our context, firms whose human capital is not adapted to the market of entry – do not find it optimal to enter internally.

Second, we document that external entries represent around 7% of market entries in France. This is consistent with Bernard et al. (2010) who use US data. In light of this figure, external entries have arguably been the focus of a disproportionate number of papers in the finance literature[6]. The link between corporate diversification and labor has only been studied from the M&A perspective. Tate and Yang (2015) predict diversified M&A operations with cross-sectoral labor flows, and interpret human capital transferability as synergies across sectors. Ouimet and Zarutskie (2016) find a positive correlation between references to "'skilled workers"' in target firms' 10-K fillings and post-merger employment outcomes. Our conclusions are in line with these existing papers, suggesting that human capital acquisition is a motive behind M&As. We contribute to the existing literature in two ways: we study the link between firm's entry type and human capital, and we propose a firm-level measure of human capital based on the comprehensive workforce composition.

Prior research has relied on indirect proxies of human capital, such as aggregated cross industry labor flows (e.g. Tate and Yang (2015), Donangelo (2014)), the use of human-capital related words in 10-K fillings (Ouimet and Zarutskie (2016)), the amount

[6]See Eckbo (2014) for a recent survey of the M&A literature.

of life insurance firms take on against their workers' death (Israelsen and Yonker (2015)), accumulated SG&A expense to assets (Eisfeldt and Papanikolaou (2013)), or key workers' human capital defined by education, work experience, or personality (e.g. Kaplan et al. (2012), Custódio and Metzger (2013)). Instead, our measure is specific to a given market of entry, and based on the empirical predictions of a production model with endogenous choice of worker teams. In the model, workers' wages are correlated with observable as well as unobservable dimensions of human capital. Our empirical estimation of this human capital measure is inspired by Abowd et al. (1999)'s high-dimensional fixed effects regressions, and is based on employer-employee data. It has two main advantages. First, it yields a first ranking of top (to bottom) worker occupations, for every sector of the French economy. Second, it is a first measure of human capital at the firm-level, that is specific to a given sector of entry. Our human capital measure is consistent with growing evidence that human capital is driven by the level of occupation-specificity of human capital (Autor and Dorn (2009), Kambourov and Manovskii (2009a) and Kambourov and Manovskii (2009b)).

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 describes the data set. Section 4 explains the empirical strategy. Section 5 presents the empirical results and Section 6 concludes.

2 Model

We propose a standard monopolistic competition model, in which consumers have CES preferences. We assume that a continuum of tasks has to be executed by worker teams. Each worker team is defined as one exogenous worker occupation, and one endogenous number of workers.

2.1 Model setup

2.1.1 Endowments and preferences

There are N sectors in the economy, each of which consists of one single product. The preferences of a representative consumer are given by a CES utility function over the different products $n \in [0, 1]$:

$$U = \sum_{n \in N} \left(\int_{\omega \in \Omega_n} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

where $q(\omega)$ denotes the demand for product n's variety $\omega \in \Omega_n$, and $\sigma > 1$ is the elasticity of substitution between a product's varieties.[7] The demand function for variety ω of product n writes

$$q(\omega) = \frac{E_n}{P_n}\left(\frac{p_{fn}}{P_n}\right)^{-\sigma} \tag{2}$$

where P_n is the sectoral price index and E_n is the expenditure in sector n.[8]

2.1.2 Production

We assume that each firm is a monopolist for its own variety ω among a product n's varieties Ω_n. For simplicity, we further assume that sectors are perfectly segmented and derive firms' optimal production decision independently for each product variety.[9] We study the program of a single firm f for a single product variety n, and drop the f and n subscripts for clarity.

2.1.3 Worker occupations

Firms' production technology uses labor as the only input, supplied by L occupations of workers who execute a unit mass of tasks denoted $t \in [0, 1]$. We denote each worker type by $l \in L$ where L is the set of worker occupations, where worker occupations differ as described in Assumption 2.1.

Assumption 2.1 (**Worker occupations and team productivity**). *The marginal cost of l-occupation workers to execute production tasks is denoted h_{lt}. It is* iid, *and such that:*

$$P(h_{lt} \leq h) = 1 - \exp\left(-e_l s_l^{\beta} h^{\theta}\right) \tag{3}$$

where e_l is the (market-specific) efficiency of l-occupation workers, s_l is the team size of l-occupation workers, and θ determines the level of specificity of l-occupation workers.

We assume decreasing returns to scale in workforce size:

$$\beta < \min\left(\frac{\theta}{\sigma}, 1\right). \tag{4}$$

[7] In a previous version we have solved a model with sector-specific elasticity of substitution between varieties, and imperfect substitutability between products. The model's conclusions remain unchanged.

[8] As originally shown by Dixit and Stiglitz (1977), $P_n = \left(\int_{\omega\in\Omega_n} p(\omega)^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}}$, and $E_n = P_n\left(\int_{\omega\in\Omega_n} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right)^{\frac{\sigma}{\sigma-1}}$, $R = \sum_{n\in N} E_n$, where R is the representative consumer's (exogenous) income. σ pins down the price elasticity of demand for product n's variety ω.

[9] A more general model in which firms make their production decisions when operating in multiple sectors yields qualitatively similar results, and identical testable predictions. We therefore keep the model simple by assuming segmented markets.

Finally, total wages must not exceed a fraction $\alpha \in [0,1]$ of total variable costs per firm:

$$\sum_{l \in L} w_l s_l \leq \alpha q(\omega) c(\omega) \tag{5}$$

where the vector of wages w is exogenous

In condition (3), we assume that worker occupations $l \in \{0, \ldots, L\}$ differ in their efficiency e_l at executing tasks in a given sector. Labor input specificity differs across markets, as encompassed by parameter θ.

To prevent corner solutions in which a firm hires a unique worker occupation, we impose in condition (4) that worker productivity is decreasing in team size. Parameter θ determines the level of skill specificity in a given sector. When θ is high, skills are "generic" in that most teams have low marginal costs in executing tasks, and firms are likely to hire multiple worker occupations. Parameter β determines the marginal increase in productivity when the workforce size increases. The lower β, the smaller the workforce and the larger the number of worker occupations firms hire.

Finally, condition (5) is akin to a positive profits condition, ensuring that firms cannot share with their workers more than a given fraction of their value added.

2.2 Model analysis

Firms face monopolistic competition. They first produce, and then sell their products at an optimally chosen price. They take the demand function $q(\omega)$ as given, and maximize their profit by choosing: (i) which worker occupations to hire, (ii) the size of each worker team, and (iii) their product's optimal price $p(\omega)$. Their program writes

$$\max_{p(\omega), c(\omega)} \pi(\omega) = p(\omega) q(\omega) - q(\omega) c(\omega), \tag{6}$$

subject to constraint 5, where $c(\omega)$ is the marginal cost of production resulting from the optimal choice of worker occupations and team sizes.

We solve for firms' optimal decisions backwards. First, we take the marginal cost of production as given, and solve for firms' optimal price. The optimal price set by the firm is equal to $p(\omega) = \frac{\sigma}{\sigma-1} c(\omega)$, where the mark-up over marginal cost is given by $\sigma/(\sigma-1)$.

Second, we derive firms' marginal costs in two steps. In a first step, we solve for firms' choice between worker occupations. Firms select occupations with the lowest marginal cost of production. We denote this lowest marginal cost by $h_t^* \equiv \min_l h_{lt}$, where h_{lt} is l-occupation workers' marginal cost in executing a given task t.

In a second step, we solve for optimal team sizes. Fims maximize (6) subject to (5). We obtain Proposition 1.

Proposition 1. *Firms' optimization program yields a unique solution for the optimal team size s_l^* of each worker occupation. As a result, the share $Share_l$ of worker occupation l of a firm's total wage bill is:*

$$Share_l \equiv \frac{w_l s_l^*}{\sum\limits_{k \in L} w_k s_k^*} = \frac{e_l^{\frac{1}{1-\beta}} w_l^{-\frac{\beta}{1-\beta}}}{\sum\limits_{k \in L} e_k^{\frac{1}{1-\beta}} w_k^{-\frac{\beta}{1-\beta}}} \tag{7}$$

Proof. See Appendix A.2.3. □

According to Proposition 1, relatively more efficient worker teams are devoted a higher share of firms' total wage bill as long as their relative wage is not too high. This cost/benefit analysis leads to an endogenous composition of firms' workforce, and of the share of tasks executed by each worker occupation in equilibrium.

From Proposition 1, one can rewrite optimal marginal cost as a function of worker occupation-specific efficiency and wages,[10] so that equation (7) relates a given worker occupation's fraction of a firm's wage bill to this occupation's relative efficiency. In Section 4.1 we estimate equation (7) using data on wages at the worker occupation level. We first obtain a ranking of (top to bottom) worker occupations in every sector. We then use our ranking of occupations to construct a firm-level measure of human capital.

2.3 Internal versus external growth

We now study firms' choice of entry type. When entering a new market, firms choose to enter either internally or externally by comparing expected profits. $\pi(\omega)$ denotes firm's profits *prior to* entry.

When entering internally, expected profits write:

$$\mathbb{E}\left[\pi^I\right] = \pi(\omega) + \mathbb{E}_f\left[\pi_{f,n'}\right]$$

where $\mathbb{E}\left[\pi_{f,n'}\right]$ denotes the firm's expected profits in the new market n'. These profits are uncertain because the firm's existing workers execute tasks in sector n' at an uncertain marginal cost.

When entering externally, expected profits write:

$$\mathbb{E}\left[\pi^E\right] = \pi(\omega) + \pi_g + S_{fg}$$

where π_g are target's profits, and S_{fg} are match-specific synergies between the acquiror and the target.

[10]See Appendix 6.

The comparison of expected profits from internal versus external growth yields Proposition 2.

Proposition 2. *A firm chooses to grow internally if expected profits from internal entry are higher than expected profits from external entry, which writes:*

$$\mathbb{E}\left[\left(\sum_{l\in L} e_{l,n'}^{\frac{1}{1-\beta}} w_l^{\frac{-\beta}{1-\beta}}\right)^{\frac{(1-\beta)(\sigma-1)}{\theta-\beta(\sigma-1)}}\right] \geq \left(\sum_{k\in L_g} e_{kn'}^{\frac{1}{1-\beta}} w_k^{\frac{-\beta}{1-\beta}}\right)^{\frac{(1-\beta)(\sigma-1)}{\theta-\beta(\sigma-1)}} + K_{fg}, \tag{8}$$

where $E_{l,n'}$ is efficiency of l-occupation workers in the new sector n', L_g is the set of worker occupations hired by the target firm, and K_{fg} a constant.

Proof. See Appendix A.2.4. □

The left-hand side term in equation 8 represents the expected efficiency of the firm's existing teams in the new market n', which we interpret as the firm's existing human capital. The first term on the right-hand side represents the efficiency of teams in the target g in the new market, i.e. the target's human capital. The last term K_{fg} is a constant (positive or negative) that is specific to an acquiror-target match. It represents potential synergies that might arise from the merger between two specific firms.

We obtain two empirical predictions from Proposition 2. First, firms with poor human capital match to the new sector should grow externally. In Section 4, we propose a firm-level measure of human capital based on equation (7), and test whether firms with poor human capital favor external entry.

According to Proposition 2, firms with poor human capital match choose to grow by acquiring a target already existing in the new sector. This strategy is more profitable when the target's human capital is higher. Therefore the second prediction of our model is that targets already operating in the new sector should have a relatively higher human capital compared to their peers.

Consistent with our model and the existing literature, we find empirical support for our model's predictions. We interpret these findings as evidence that external entries are a means for firms to acquire their targets' human capital.

3 The data

3.1 Firm data

Our primary source of data consists in French administrative data provided by the French Statistical Institute (INSEE). We work at the business group level, and identify the set of firms that belong to the same business group using data on ownership links (Enquête sur les Liaisons financières entre sociétés/ Lifi[11]). Even though we work at the business group level, for expositional simplicity we henceforth refer to "firms".

We focus on firms in the private sector and exclude the financial and agricultural sectors.[12] We recover balance-sheets and operational performance data from tax returns files (which include all profit and loss and balance sheet statements).

Our observations of firms' activity comes from an extensive yearly survey from the French statistic institute (INSEE). These data sets display the amount of sales realized in any market for virtually every French firm.[13]

We use the French industry classification (nomenclature of activity, or NAF) which is the equivalent of the US Census Bureau's Standard Industry Classification (SIC). We define a "market" as a 4-digits French-equivalent SIC code. Our data set contains 525 different markets. We focus on the 2008-2014 period due to major modifications in industry classification in 2008.

3.2 Merger and acquisitions

We merge the French administrative datasets listed above with data bases on mergers and acquisitions, SDC Platinum and Bureau van Dijk's Zephyr. We collect all deals between January 2008 and December 2014. We select deals in which the acquiror owns less than 50% of the target shares before the acquisition date and more than 50.1% after. We focus on deals involving both a French acquirer and target. LBO deals and operations made by private equity firms are excluded from the sample.

Note that the firm standardized ID numbers are not provided in commercial databases. We proceed in several steps to retrieve French firm standardized identifying numbers (called Siren). First, we use the stock tickers (available only for publicly traded firms) and the Bureau van Dijk identifier (available only for Zephyr deals) to download their identifying number from the Bureau Van Dyck Orbis data set.[14] Second, we build a Python robot to

[11] The survey is exhaustive on the set of firms that employ more than 500 employees, that generate more than 60 millions euros in revenues or that hold more than 1.2 million euros of shares, but is completed by data coming from Bureau Van Dijk (Diane-Amadeus data set) to cover the whole universe of French business groups.

[12] We exclude these sectors for two reasons: first, they use very different accounting systems, which limits the comparability and relevance of standard variables across sectors: the notion of "value added", for instance, is difficult to apprehend and to measure in the public or the financial sector. Second, entry in these sectors are presumably determined by other economic rationales than in the rest of the economy.

[13] The survey is exhaustive for firms whose turnover or number workers exceed a given threshold. We treat firms that are not included in the surveys as operating only in their main sector of activity, which we observe using tax returns data.

[14] Bureau van Dijk is a private data provider which gives access to Financial data (Amadeus file), legal entity details and ownership links (Orbis dataset), and financial transactions including M&A activities (Zephyr file) of millions of private and public companies over the world.

look for a company's name and address on two websites.[15]

Finally, we drop observations with several matches which do not report an address, city, zip code nor any other information. We impose a maximum Jaró-Winkler string distance between the original and retrieved names of 0.8 for both the name and address and drop the matches that do not meet this restriction.[16]

Our final sample contains 7,303 deals from 2008 to 2014. To the best of our knowledge, this is the largest M&A data set available to date for the French economy.

3.3 Occupation-level worker data

We use matched employers-employees data (Déclarations Annuelles des Données Sociales/DADS). These data is available for all workers. Firms are required by law to report information on each of their employees in order to establish payroll taxes. For each worker, we observe the gross and net wages as well as the number of days and hours worked in a given year. We restrict to workers' main occupation, with an annual gross wage higher than 1000 euros. Importantly, we observe the occupation of each worker. We interpret the 4-digit occupation code as the empirical equivalent of a team in our model.[17]

4 Empirical strategy

4.1 Human capital

Our strategy is based on structural equations derived from the model of Section 2.2, to propose a firm-level measure of human capital that we can take to the data. The construction of our variable comes in two steps.

In line with the model, our first step builds on the idea that if a large share of the wage bill is devoted to a given worker occupation in a given market, then this occupation is key

[15] The robot is built as using the Python packages Selenium and Beautiful Soup). We use our web-crawler on (i) www.bodacc.fr (Bulletin Officiel des Annonces Civiles et Commerciales), a governmental website which has collected and published official notifications involving French companies since 2008, and on (ii) www.societe.com, a commercial website that aggregates and reshapes information about French companies from various sources (mostly from INSEE and Bodacc.fr). Both websites claim to cover the universe of French firms.

[16] This distance measures the number of characters in common between strings with the idea that differences near the start of the string are more significant than differences near the end of the string (Christen, 2006).

[17] This is the most detailed level of the French classification of worker occupations (PCS-ESE), with 414 different occupations.

to produce in this market. We rewrite Equation (7) with sector, firm and time subscripts:

$$Share_{lfnt} = \frac{e_{lfnt}^{\frac{1}{1-\beta}} w_{lt}^{-\frac{\beta}{1-\beta}}}{\sum\limits_{k \in L} e_{kfnt}^{\frac{1}{1-\beta}} w_{kt}^{-\frac{\beta}{1-\beta}}}, \tag{9}$$

where $lfnt$ denotes an occupation l in firm f that operates in sector n at date t. We assume that an occupation's efficiency, e_{lfnt}, can be decomposed into sector-specific and firm-specific components as: $e_{lfnt} = \tilde{e}_{lnt} * \tilde{e}_{lfnt}$, with $\tilde{e}_{lfnt}$ iid across firms in a given (l, n, t) triplet. After simple manipulation, and in logarithm, Equation (9) delivers a fixed-effects decomposition of the share of worker occupations in firms' total wage bill:

$$\log Share_{lfnt} = \log \epsilon_{lnt} - \log \epsilon_{fnt} + \log \epsilon_{lfnt} \tag{10}$$

where:

$$\begin{aligned}
\epsilon_{lnt} &= \tilde{e}_{lnt}^{\frac{1}{1-\beta}} w_{lt}^{-\frac{\beta}{1-\beta}} \\
\epsilon_{fnt} &= \sum_{k \in L} e_{kfnt}^{\frac{1}{1-\beta}} w_{kt}^{-\frac{\beta}{1-\beta}} \\
\epsilon_{lfnt} &= \tilde{e}_{lfnt}^{\frac{1}{1-\beta}} w_{lt}^{-\frac{\beta}{1-\beta}}
\end{aligned}$$

Each year, we obtain a decomposition of worker occupations' shares in firms' total wage bill into three terms: an occupation × sector component, a firm × sector component, and a residual specific to the occupation-firm-sector triplet.[18] The first term of equation (10) can be directly estimated with an occupation × market × time fixed effect;[19] and the second term with a firm × time fixed effect. In our first step, we are interested in estimating the first term, which proxies for the importance of every worker occupation in any sector.[20]

We estimate Equation (10) using a two-way fixed effect estimator, following the method developed by Abowd et al. (1999). Identification hinges on the following exo-

[18]We let the fixed effects depend on time to capture potential changes in the relative importance of occupations in the production processes (Autor and Dorn, 2009).

[19]We do not observe the market to which each worker is assigned in a firm. As an approximation, we allocate workers to the production process of their firm's main sector (defined as the market in which the firm realizes the majority of its sales). As a robustness check, we apply the same methodology with plants as units of observation. Our main results remain. This alternative, though seemingly more precise, has the drawback to rely to a plant-level sector code that is less accurate than the firm-level one.

[20]The second and third terms encompass firm-specific personnel policies of the firm, respectively for all workers and for specific occupations, that may alter the returns to observable and unobservable dimensions of human capital.

geneity condition:[21]

$$\mathbb{E}\left(\tilde{e}_{lfnt}|(f,t)(l,t)\right)=0,$$

which in our context, implies that worker occupations are not matched to specific firms' production needs. We argue that this is a reasonable assumption, given occupations are classified by the French statistical institute, and widespread across firms and sectors.

Note that for the two fixed effects to be identifiable, firms must hire more than one type of occupation each year, and every occupation must be present in more than one firm in a given sector each year (Abowd et al., 1999). Hence, we focus on firms that hire more than one worker occupation.[22] We also restrict to a minimum of 10 firms for every occupation-market-year triplet, to obtain precise estimates.

We obtain a fixed-effect estimate for each occupation,[23] which we interpret as a score which reflects the average wage share of a given occupation within a sector. Indeed the higher the estimated fixed effect for an occupation in a given sector, the more efficient and therefore the more paid this occupation in this sector. Thereby, our first step obtains a ranking of (top to bottom) worker occupations, for every sector of the French economy.

[Insert Table 3b here]

Table 3b provides examples of top occupations, i.e. occupations with the highest scores computed from the fixed-effects in (10), for the "manufacture of pharmaceutical preparations" and "computer consultancy activities" sectors in 2013. The ranking of occupations correspond to tasks that indeed, intuitively seem key to this sector's production process.

In a second step, we use our ranking of occupations in every sector to construct a firm-level measure of human capital. For each firm, we use the fixed effects estimated in the first step to compute their average value across all occupations prior to entry (denoted $t-1$ by convention). Our definition of human capital at the firm level is:

$$\text{Human capital}_{fn,t-1}=\frac{1}{\#L_{f,t-1}}\sum_{k\in L_{f,t-1}}\exp\left(\widehat{\log}\left(\epsilon_{lnt}\right)\right)$$

where $\#L_{t\text{-}1}$ denotes the number of different occupations hired by the firm prior to entry. We obtain a first measure of human capital at the firm-level, that is specific to the market of entry n (consistent with Autor and Dorn, 2009). We test several variations of this formula in robustness checks (see Section 5.2).

[21] Note that estimating Equation (10) year-by-year in the panel dimension does not impose any restriction on the time correlation of the various components.

[22] We also restrict to firms with more than 20 reported workers, because occupations are badly reported for firms with less than 20 employees.

[23] Fixed effects coefficients are estimated and extracted using the Stata package Reghdfe (Correia et al. (2016)).

Henceforth, for clarity, we refer to this measure as human capital without explicitly mentioning that it is specific to a given market of entry.

4.2 Internal versus external entry

We impose that new sales in the market of entry, as well as total fixed investment by the firm, are above 50 000 euros. This threshold is relatively low compared to the average entry sales and investment (1,5 and 2,5 millions euros respectively), and enables us to remove "small" entries or false positives, i.e. cases in which sales across sectors are imperfectly reported.

Table 1 displays the number of internal and external entries in new markets. We find a ratio of external to internal entries of approximately 6-7%, which is very close to the 6.8 % estimate found by Bernard et al. (2010) in a US setting. The ratio increases when we use less precise definitions of markets, suggesting that external entries occur relatively more often in more distant markets. Table 2 shows the markets in which most firms entered - both internally and externally - in 2013.

[Insert Table 2 here]

4.3 Distance between sector of origin and sector of entry

Building on Bloom et al. (2013), we construct a measure of the distance between the firm (at the business group level) and its market of entry. This metric is based product market similarities. It is based on the idea that if firms in a given market tend to also operate in another market, then those two markets are "close" – otherwise they are "distant".

For a firm entering a market at time t, we denote $\mathbb{S} = (S^1, ..., S^N$ the vector of its sales at time $t-1$, broken down by market $(n = 1, ..., N)$. $\mathbb{S}^{-n}$ is the vector of sales excluding sales in market n. For a given market of entry n, we define the distance d_g^n between the firm and any firm g already operating in the new market n, as (one minus) the uncentered Pearson correlation between vectors $\mathbb{S}^{-n}$ and $\mathbb{S}_g$:

$$d_g^n = 1 - \frac{(S_g^{-n} \cdot S^{-n})}{\sqrt{(S_g^{-n} \cdot S_g^{-n})}\sqrt{(S^{-n} \cdot S^{-n})}}$$

This measure explicitly excludes the sales of both firms in market n, because the firm enters market n at time t, hence necessarily $S^n = 0$: a simple correlation of the sales repartition between the firm and any firm g already operating in market n would mechanically be zero.

Finally, we measure the distance of a firm to a market of entry n as the weighted average of d_g^n across all firms operating in market n at $t-1$:

$$\text{Distance to market of entry } n = \frac{\sum_g \omega_g^n d_g^n}{\sum_g \omega_g^n}$$

where the weights are given by the share of sales realized by each firm operating in market n at $t-1$: $\omega_g^n = \frac{\text{Sales of firm } g \text{ in market } n}{\text{Sales of group } g}$.

The weights ensure that the distance between the growing firm and a firm already operating in market n matters more if the firm in market n realizes a large part of its sales in sector n.

4.4 Empirical model

Our dependent variable is External entry$_{f,n,o,t}$, a dummy equal to one if firm f enters a new market n externally, i.e through an acquisition, and zero if it enters the new market internally. o indicates firm f's market of origin, i.e. the market in which the firm's group realizes the largest share of its sales prior to entry. Our baseline empirical model is:

$$\text{External entry}_{f,n,o,t} = \lambda_n + \lambda_o + \lambda_t + \delta \text{Human capital}_{f,n,t\text{-}1} + \beta X_{f,n,o,t-1} + \epsilon_{f,n,o,t} \quad (11)$$

The fixed effects λ_n and λ_o remove unobserved stable factors at the level of the markets of entry (e.g., magnitude of fixed costs of entry) and origin (e.g., ability to collateralize assets to access external finance). The time fixed effects c_t capture aggregate macroeconomic factors. $X_{f,n,o,t-1}$ denotes the set of control variables.

By construction, Human capital$_{f,n,t\text{-}1}$ introduces a measurement error term which generate a correlation between residuals $\eta_{f,n,o,t}$ at the level of the sector of entry. We therefore cluster standard errors to control for correlations within the sector of entry. Finally, we standardize our constructed variable Human capital$_{f,n,t\text{-}1}$ in the rest of the analysis, to interpret regression coefficients in standard deviation.

4.5 Control variables and descriptive statistics

Table 6 shows that firms growing internally and externally differ on several dimensions. External entries are associated with higher entry sales and a higher probability of staying in the new market (+ 9 percentage points at the one-year horizon). Firms entering externally have more workers, are more profitable, more capital intensive and have higher internal funds. Table 6 stresses the importance of controlling for observable factors using a multivariate analysis.

[Insert Table 6 here]

The first set of control variables is related to the profitability and the financial health of the firm. The variable log(# of workers $)_{f,t\text{-}1}$ is the (log) total number of workers in the firm, and controls for the firm size. We measure the profitability as the value added per worker (Value added/#workers$_{f,t\text{-}1}$), the intensity in capital of the production process as the total amount of fixed assets per worker (Fixed assets/#workers$_{f,t\text{-}1}$) and the size of internal funds as the amount of cash holdings per worker (Cash holdings/#workers$_{f,t\text{-}1}$).

The type of entry may also depend on firms' ability to finance the acquisition with external funding. Firms in sectors that rely heavily on external finance might find it more difficult to to finance an acquisition through additional funding. We measure a firm's dependence on external finance as follows. First, we obtain sectoral measures of dependence on external finance by computing the average share of capital expenditures that is not financed by operating cash-flows in a given sector (Rajan and Zingales, 1998). Second, for each firm, we take the sales-weighted average of our measure. We obtain variable External finance dependence$_{f,n,t\text{-}1}$, which measures a firm's dependence on external finance based on the dependence on external finance in the markets it operates.

We then control for barriers to entry in the new market. First, we include the Herfindahl index in that market (Herfindahl$_{n,t}$). Second, we control for the size of the market of entry by computing the (log) total sectoral sales (log(Sales in the new sector)$_{n,t}$). The model suggests that the larger expected profits, the less important the fixed costs of acquisitions, hence the higher the probability of external entry.

Table 5 gives an overview of the distribution of the variables of interest and control variables used in the analysis.

[Insert Table 5 here]

Figure 4 plots the density of firms' human capital$_{f,n,t\text{-}1}$ by entry type, for our entire sample. One can already observe in Figure 4 that firms entering a new market externally appear not to own the right human capital to operate in the market of entry.

[Insert Figure 4 here]

5 Estimations

5.1 Main results

Table 7 presents our baseline results. Columns 1 to 3 analyze the role of human capital for the entire sample of firms. All specifications control for firm size (as measured by the number of workers): we use tercile dummies rather than a continuous specification to allow for non-linear effects. Interactions between our measure of human capital and tercile dummies account for potential heterogeneity across firms of different size, and the correlation between human capital and entry type.

[Insert Table 7 here]

Consistently with Figure 4, Human capital$_{f,n,t-1}$ coefficients are all negative and significantly different from zero. The probability of external entry increases with the size of the firm; moreover, column 2 shows that the relationship between the probability of external entry and human capital is significantly more negative in the last tercile of firm size. The baseline coefficient for human capital coincidentally decreases in absolute value from -0.005 to -0.001, suggesting that the negative relationship between human capital and the probability of external entry is mainly driven by large firms. We observe that firms that are cash-rich, more profitable, rely less on fixed assets and are more dependent on external finance more often enter a new market externally. Still, our findings are robust to the introduction of control variables.

Columns 4 to 6 replicate the previous analysis on the three terciles of firm size. The relationship between the probability of external entry and the control variables is weakened once we control for firm size. The Human capital$_{f,n,t-1}$ coefficient, however, is negative and significative at 5% for all terciles of size. In line with columns 2 and 3, regression coefficients increase with firm size.

Overall, these results suggest that firms that do not own the right human capital, i.e. that do not hire the key worker teams to produce in the market of entry, tend to enter the new market through an acquisition. This finding is more pronounced for larger firms, which can presumably better afford to acquire existing firms in the new market.

5.2 Robustness checks: Alternative measures of human capital

In Table 8 we check the robustness of our results to changes in the definition of human capital. All specifications control for size, profitability, capital intensity, cash holdings, dependence on external finance, sector of origin, sector of entry and year fixed effects.

[Insert Table 8 here]

A first source of concern about the methodology of construction of our main variable is that it may underestimate the human capital for diversified firms. These firms are likely to hire a large range of worker occupations; and since our baseline human capital measure is computed as an average value of fixed effects, it can be driven down by occupations that are irrelevant to the sector of entry.

To tackle this issue, we build a dummy variable that is equal to one if before diversifying, the group does not hire any of the top 10 occupations for the sector of entry. We interpret these "top 10" occupations as the most important worker occupations for the sector of entry. Since the other occupations do not play a role in the computation of the variable, this measure is immune to the bias described above. In line with our previous results, we find in column 1 of Table 8 that the coefficient of this alternative measure of human capital is positive and statistically significant: firms that do not hire any of the top 10 occupations in the sector of entry tend to grow externally in the new market.

Alternatively, in column 2 of Table 8, we compute human capital considering *only* in the summation the ten most specific occupations to the sector of entry. We confirm the results of our baseline model and find a negative correlation between the probability of an external entry and human capital.

In column 3, we take a very simple measure of human capital, measured as the average share of the total wage bill that is devoted to the occupation in the market of entry. We do not use fixed effects. The results remain true, and even one order of magnitude larger economically: the coefficient on Human capital$_{f,n,t\text{-}1}$ goes from -1bp to -10bp.

In column 4, we weigh the estimated occupation × sector × year fixed effects by the number of workers employed by the firm in the occupation. This specification incorporates information on the relative size of occupations within a firm. Our results remain.

In column 5, we exclude CEOs from the measure of human capital, with the concern that CEO wages may be set differently from those of other worker occupations. The negative correlation between the probability to enter the new market through an acquisition and human capital remains.

In column 6, we exclude occupations that have been manually coded from the measure of Human capital$_{f,n,t\text{-}1}$, to assess the role of potential measurement errors. Occupations marked as automatically coded include jobs which are presumably broad and include unfocused tasks (e.g. 'secretaries', 'qualified administrative staff', 'commercial executive at SMEs'), as opposed to well-defined occupations (e.g. 'stewardesses and stewards', 'veterinarians', 'bombers') which are seldom identified as automatically coded. The results are barely modified.

Finally, in column 7 we change the unit of observation used in the estimation of fixed effects. In our baseline methodology, we assume that each worker is employed in the main sector of its employing firm. Estimating fixed effects at the plant-level may yield better estimates as plants are necessarily less diversified than firm, although the sector of a plant is presumably less accurate. The negative relationship between human capital and external entry still holds under this alternative specification.

5.3 Interactions between sector of origin and sector of entry

Firms endogenously choose to enter a new market given complementarities and synergies between the sector of entry and their original activities. For instance, firms may choose to enter distant markets externally, and close markets internally. In this subsection, we try to control for interactions between sector of origin and sector of entry.

[Insert Table 10 here]

We include our Distance to the sector of entry$_{f,n,t\text{-}1}$ variable in columns 1 and 2, to check that our human capital coefficient only captures complementarities arising due to human capital considerations. We find that the probability of external entry increases with our measure of product market distance, which confirms that firms are more likely to enter distant sectors externally. Surprisingly, we find external entries to occur more frequently in smaller sectors (in terms of total sales): we would have expected acquisitions to occur in large markets as fixed costs of acquisition should be offset by higher expected profits. External entries also occur more often in concentrated markets, but the relationship is only significant at the 10% level.

We check in column 3 that the relationship between sectoral distance and entry type is monotonous in our distance measure.

In column 4, we reestimate Equation 11 replacing origin and entry sector fixed effects with origin $\times$ entry $\times$ year fixed effects. This specification is identified by variations in the human capital across firms *within the same sectors of origin and destination* and within a year. By looking within a sector pair, we can remove any potential economic factors based on complementarities between the two sectors. This method, however, has the drawback to drop all the origin $\times$ destination singletons, i.e. approximately 4,000 observations. Even with origin $\times$ entry fixed effects, we find that human capital is negatively correlated with external entry (at the 5% level).

5.4 Within-group human capital

Our previous results show that firms with the right human capital tend to enter a new market internally. One implication of this is that *within the group of a firm growing internally*, the firm that enters the new market should be the one with the workforce that is the most adequate to the sector of entry. We test this additional hypothesis in this subsection and focus therefore on groups that grow internally and that are composed of more than one firm.

[Insert Figure 3 here]

For a group g entering sector n internally, we measure the distance of any firm f belonging to group g to the "best" human capital of firms in g as:

$$\text{Distance to best}_{f,n,t\text{-}1} = \frac{\max_{f \in g} \text{Human capital}_{f,n,t\text{-}1} - \text{Human capital}_{f,n,t\text{-}1}}{\max_{f \in g} \text{Human capital}_{f,n,t\text{-}1}}$$

This distance is equal to zero if firm f in group g is the firm owing the "best" human capital to enter sector n. It approaches one when Human capital$_{f,n,t\text{-}1}$ tends to zero, so that firm f does not own the right human capital to grow in sector n. Figure 3 shows that within a group, a large share of internal entries indeed are made by firms that own the best human capital compared to the other firms belonging to the group.

[Insert Table 9 here]

In Table 9, we define a dummy Internal entry$_{f,n,t\text{-}1}$ equal to one if firm f enters sector n internally. We regress this dummy on Distance to best$_{f,n,t\text{-}1}$. Standard errors are clustered at the group level; the sector of origin is defined as the sector where the firm realizes most of its sales, and the sector of entry as the sector to which belongs the market in which the group enters internally.

All specifications show a large and significant negative relationship between a firm's probability to enter internally and its Distance to closest$_{f,n,t\text{-}1}$ within the group it belongs to. Column 2 shows that internal entry is positively associated with capital intensity and cash holdings. We include origin $\times$ destination fixed effects in column 3; the correlation coefficient decreases in absolute value but remains significantly different from zero at the 1% level.

The results from this additional test essentially show that on average, groups diversify via their firm that owns the best human capital to operate in the new market. This finding supports not only the importance of human capital as a determinant of firms' diversification, but also the validity of our measure of human capital.

5.5 Change in human capital around entry

So far we have considered the composition of human capital one year before entry. However, entries might be anticipated years before they occur, so that firms might change their human capital accordingly. This raises the issue of endogeneity of the human capital to the type of entry.

To investigate this question, we take all firms that enter a new market – regardless of their entry type – and compute the change in their workers number as:

$$\Delta Workers_{f,t} = \frac{\# \ Workers_{f,t} - \# \ Workers_{f,t-1}}{0.5 * (\# \ Workers_{f,t} + \# \ Workers_{f,t-1})}$$

Values of this growth rate range between -2 and 2 (Davis and Haltiwanger, 1992).

We benchmark the occupation-sector-year fixed effects at their 2012 value in order to compute variations in firms' human capital around entry. Our definition of Human capital becomes:

$$\overline{\text{Human capital}}_{f,n,t} = \sum_{k \in L_{f,t}} \omega_{f,k,t} \exp\left(\widehat{\log}\left(\phi(e_{n,k,2012}, w_{k,2012})\right)\right),$$

where $\omega_{f,k,t}$ are weights computed as the relative number of workers of occupation k in firm f at t. This modification ensures that $\overline{\text{Human capital}}_{f,n,t}$ only reflects the evolution of weights $\omega_{f,k,t}$ over time – reflecting changes in human capital immune to changes in occupation-sector-year fixed effects. In contrast with our baseline human capital measure, we sum the fixed effects to ease interpretation of the results.

The growth rate $\Delta\overline{Workforce}_{f,n,t}$ is computed similarly as $\Delta Workers_{f,t}$. Due to data limitations, we restrict ourselves to a symmetric time interval $[T-2, T+2]$ where T is the entry date. We only keep firms for which we have 8 consecutive observations between 2007 and 2014 and focus therefore on entries occurring between 2010 and 2011.

Figure 5 displays average values of $\Delta Workers_{f,t}$ and $\Delta\overline{Workforce}_{f,n,t}$ for firms entering new markets internally. We see that in the two years prior an internal entry, the number of workers stays stable on average while human capital increases by more than 4% on average. New hires are made in occupations that are key for the market of entry. The increase in human capital decreases post entry suggests a slowdown in the workforce adjustment. $\Delta\overline{Workforce}_{f,n,t}$ becomes negative two years after entry but stays greater than $\Delta Workers_{g,t}$, which implies that layoffs occur mainly in occupations that are not key for the market of entry.

[Insert Figure 5 here]

Figure 6 shows the same figure for firmsentering new markets externally. These firms' human capital of increases sharply (10-15%) in the two years before the acquisition, while the total number of workers increases only marginally. This suggests that targeted firms rationalizes its workforce to focus on key occupations for their market. After the acquisition, the growth rates of the number of workers and human capital follow similar paths, with a positive growth rate of 2% on average the year following the acquisition, and a decrease of 3% the year after.

[Insert Figure 6 here]

These findings suggest the existence of changes in human capital that are initiated by firms prior to entry. Hence human capital is not entirely exogenous to entry type. We try to address this issue, by using the 2- and 3-year lags of Human capital$_{g,n}$ in our estimations. The results are in Table 11.

[Insert Table 11 here]

The number of observations drops mechanically, but ou results remain. We interpret this finding as a consequence of the persistence of human capital over time. Even though firms do adjust their human capital to adjust to the market of entry, human capital remains a key factor determining firms' entry type.

6 Conclusion

What determines a potential entrant's choice between internal versus external entry in a new market? This paper proposes a model in which firms face the following trade-off when they diversify. On the one hand, if they enter the new market internally, firms face uncertainty on how adapted their existing stock of human capital is for the sector of entry. On the other hand, if they enter the new market externally, firms have to pay an acquisition premium to takeover an existing firms' human capital. The model offers a structural estimation of human capital at the firm level that measures how adapted a diversifying firm's human capital to the sector of entry. We use data from the French Statistical Institute on the distribution of wages across all occupations and sectors to test the predictions of our model, and robustly find that firms choose to enter a new market internally (externally) if they own an existing stock of human capital that is adapted (unadapted) for the sector of entry. This paper contributes to the literature on human capital and corporate diversification, showing that firms' choice between building or buying human capital plays a role in explaining how firms diversify, which contributes to a better understanding of the determinants of firms' boundaries.

References

John M Abowd, Francis Kramarz, and David N Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999.

James W Albrecht and Bo Axell. An equilibrium model of search unemployment. *The Journal of Political Economy*, pages 824–840, 1984.

Andrew Atkeson and Patrick J Kehoe. Modeling and measuring organization capital. *Journal of Political Economy*, 113(5):1026–1053, 2005.

David Autor and David Dorn. This job is "getting old": Measuring changes in job opportunities using occupational age structure. *American Economic Review*, 99(2):45–51, 2009.

Andrew B Bernard, Stephen J Redding, and Peter K Schott. Multiple-product firms and product switching. *The American Economic Review*, 100(1):70–97, 2010.

Nicholas Bloom, Mark Schankerman, and John Van Reenen. Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393, 2013.

Peter Christen. A comparison of personal name matching: Techniques and practical issues. In *Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on*, pages 290–294. IEEE, 2006.

Sergio Correia et al. Reghdfe: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. *Statistical Software Components*, 2016.

Cláudia Custódio and Daniel Metzger. How do ceos matter? the effect of industry expertise on acquisition returns. *Review of Financial Studies*, 26(8):2008–2047, 2013.

Steven J Davis and John Haltiwanger. Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics*, 107(3):819–863, 1992.

Avinash K Dixit and Joseph E Stiglitz. Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3):297–308, June 1977. URL .

Andres Donangelo. Labor mobility: Implications for asset pricing. *The Journal of Finance*, 69(3): 1321–1346, 2014.

Jonathan Eaton and Samuel Kortum. Technology, geography, and trade. *Econometrica*, 70(5): 1741–1779, 2002.

B Espen Eckbo. Corporate takeovers and economic efficiency. *Annu. Rev. Financ. Econ.*, 6(1): 51–74, 2014.

Andrea L Eisfeldt and Dimitris Papanikolaou. Organization capital and the cross-section of expected returns. *The Journal of Finance*, 68(4):1365–1406, 2013.

Patrick A Gaughan. *Maximizing Corporate Value Through Mergers and Acquisitions: A Strategic Growth Guide*. John Wiley & Sons, 2013.

Ryan D Israelsen and Scott E Yonker. Key human capital. 2015.

Gueorgui Kambourov and Iourii Manovskii. Occupational mobility and wage inequality. *The Review of Economic Studies*, 76(2):731–759, 2009a.

Gueorgui Kambourov and Iourii Manovskii. Occupational specificity of human capital. *International Economic Review*, 50(1):63–115, 2009b.

Steven N Kaplan, Mark M Klebanov, and Morten Sorensen. Which ceo characteristics and abilities matter? *The Journal of Finance*, 67(3):973–1007, 2012.

Gwendolyn K Lee and Marvin B Lieberman. Acquisition vs. internal development as modes of market entry. *Strategic Management Journal*, 31(2):140–158, 2010.

Vojislav Maksimovic, Gordon Phillips, and Nagpurnanand R Prabhala. Post-merger restructuring and the boundaries of the firm. *Journal of Financial Economics*, 102(2):317–343, 2011.

Kevin F McCardle and S Viswanathan. The direct entry versus takeover decision and stock price performance around takeovers. *Journal of Business*, pages 1–43, 1994.

Paige Parker Ouimet and Rebecca Zarutskie. Acquiring labor. *Working paper*, 2016.

Daniel Paravisini, Veronica Rappoport, Philipp Schnabl, and Daniel Wolfenzon. Dissecting the effect of credit supply on trade: Evidence from matched credit-export data. *Review of Economic Studies*, pages 1–26, 2014.

Raghuram G. Rajan and Luigi Zingales. Financial dependence and growth. *The American Economic Review*, 88(3):559–586, 1998.

Richard Roll. The hubris hypothesis of corporate takeovers. *The Journal of Business*, 59(2): 197–216, 1986. URL .

G. William Schwert. Markup pricing in mergers and acquisitions. *Journal of Financial Economics*, 41(2):153–192, June 1996. URL .

Geoffrey A Tate and Liu Yang. The human factor in acquisitions: Cross-industry labor mobility and corporate diversification. *US Census Bureau Center for Economic Studies Paper No. CES-WP-15-31*, 2015.

Bernard Jr. Wysocki. Many mergers driven by search for fresh talent. *Wall Street Journal*, December 1997.

George S Yip. Diversification entry: Internal development versus acquisition. *Strategic Management Journal*, 3(4):331–345, 1982.

Appendix A Model

A.1 Demand with monopolistic competition

Consumers maximize their utility 1, which is CES between product varieties and additive preferences across goods, subject to a budget constraint. This problem yields the classical demand function

$$q(\omega) = \frac{E}{P}\left(\frac{p(\omega)}{P}\right)^{-\sigma}, \tag{12}$$

where P is the sectoral price index (substripts n are omitted for clarity), E denotes consumer expenditures for this sector, and σ is the elasticity of substitution *and* the price-elasticity of demand.

A.2 Firms' maximization program

We study the maximization progam of firms which produce using labor as an input. Firms first jointly chose which worker occupations they hire and worker team sizes. Then, they chose the optimal price of their product variety. We solve this problem by backward induction.

A.2.1 Optimal price

The assumption of monopolistic competition simplifies the optimal pricing formula. Firms maximize profits (6) and a classical result from monopolistic competition is that the optimal monopoly price is a constant markup over marginal cost:

$$p(\omega) = \frac{\sigma}{\sigma - 1} c(\omega). \tag{13}$$

A.2.2 Optimal worker teams - Lemma A.1

We now propose Lemma A.1, which we will then use to prove Proposition 1.

Lemma A.1. *Firms optimally choose between worker occupations so that their optimal production cost writes*

$$c(\omega) = \Theta^{\frac{1}{\theta}},$$

where $\Theta = \left(\sum_{k \in L} e_k s_k^{\beta}\right)^{-1} \left(\Gamma\left(\frac{\theta+1-\rho}{\theta}\right)\right)^{\frac{\theta}{1-\rho}}$.

Proof. Proof of Lemma A.1 (Eaton and Kortum (2002)).

First we rewrite assumption (3) by denoting $G_l(h) \equiv Pr(h_{lt} \leq h) = 1 - e^{-e_l s_l^{\beta} h^{\theta}}$ the distribution of team l's marginal costs.

The firm minimizes its costs by selecting worker occupations with the lowest marginal cost at executing tasks. Therefore the firm's marginal cost is the realization of the random variable $h^* = min\{h_l; l \in L\}$ which distribution is given by:

$$\begin{aligned} G(h) &= Pr\left(h^* \leq h\right) = 1 - \Pi_{l \in L} Pr\left(h_{lt} > h\right) \\ &= 1 - \Pi_{l \in L}\left(1 - G_l(h)\right) \\ &= 1 - e^{-h^\theta \sum_{l \in L} e_l s_l^\beta} \end{aligned}$$

The firm's marginal cost is a CES aggregator over tasks, $c(\omega) = \left(\int_0^1 h^*(t)^{1-\rho} dt\right)^{\frac{1}{1-\rho}}$, where ρ denotes the elasticity of substitution between tasks. We rewrite:

$$\begin{aligned} c(\omega)^{1-\rho} &= \int_0^\infty h^{1-\rho} dG(h) \\ &= \int_0^\infty \left(\frac{x}{\sum_{l \in L} e_l s_l^\beta}\right)^{\frac{1-\rho}{\theta}} e^{-x} dx \\ &= \left(\sum_{l \in L} e_l s_l^\beta\right)^{\frac{\rho-1}{\theta}} \Gamma\left(1 + \frac{1-\rho}{\theta}\right), \end{aligned}$$

where the second row uses a change of variable with $x = h^\theta \sum_{l \in L} e_l s_l^\beta$. Finally, we obtain the following optimal marginal cost for the firm:

$$c(\omega) = \left(\sum_{l \in L} e_l s_l^\beta\right)^{-\frac{1}{\theta}} \left(\Gamma\left(1 + \frac{1-\rho}{\theta}\right)\right)^{\frac{1}{1-\rho}}. \tag{14}$$

□

A.2.3 Proof of Proposition 1

Proof. Proof of Proposition 1. Once firms have chosen which worker occupations they hire, they chose worker team sizes to maximize profits. Using the demand function (12), the optimal price formula (13) and the optimal marginal cost of production (14), one can solve for the Lagrangian corresponding to firms' constrained-maximization program of (6) subject to the (binding) constraint (5). From the first-order condition and after some tedious calculations, one obtains:

$$\frac{w_l s_l}{\sum_{l \in L} w_l s_l} = \frac{e_l s_l^\beta}{\sum_{l \in L} e_l s_l^\beta}.$$

Plugging the binding constraint (5) into the above equation, one can solve for the optimal size of each worker team s_l^*:

$$s_l^* = \left(\alpha E \left(\frac{P\sigma}{\sigma - 1} \right)^{1-\sigma} \Gamma \left(1 + \frac{1-\rho}{\theta} \right)^{\frac{\sigma-1}{1-\rho}} \right)^{\frac{\theta}{\theta-\beta(\sigma-1)}} \left(\frac{e_l}{w_l} \right)^{\frac{1}{1-\beta}} \left(\sum_{l \in L} e_l^{\frac{1}{1-\beta}} w_l^{-\frac{\beta}{1-\beta}} \right)^{\frac{\sigma-1-\theta}{\theta-\beta(\sigma-1)}}. \tag{15}$$

Finally, multiplying both sides of (15) by w_l and dividing by the corresponding expression for $\sum_{l\in L} w_l s_l$, on obtains equation (7). QED.

□

A.2.4 Proof of Propostion 2

Proof. Proof of Propostion 2.

Firms choose between the two entry types by comparing expected profits of either strategy, and they choose to grow internally iff $\mathbb{E}\left[\pi^I\right] \geq \mathbb{E}\left[\pi^E\right]$, which rewrites:

$$\mathbb{E}_f\left[\pi_{f,n'}\right] \geq \pi_g + S_{fg}. \tag{16}$$

Plugging equations (12), (13) and (14) into the profits formula (6), and using the optimal worker team size (15), one can rewrite firms' profits for a given sector as:

$$\pi(\omega) = \left(\alpha^\beta E \sigma^{\beta-\sigma} \left(\frac{P}{\sigma - 1} \right)^{1-\sigma} \Gamma \left(1 + \frac{1-\rho}{\theta} \right)^{\frac{(1-\sigma)(1-2\beta)}{(1-\rho)}} \right)^{\frac{1}{1-\beta}} \left(\sum_{l \in L} e^{\frac{1}{1-\beta}} w^{\frac{-\beta}{1-\beta}} \right)^{\frac{(1-\beta)(\sigma-1)}{\theta-\beta(\sigma-1)}},$$

We use this expression to rewrite the condition (16) for firms to choose internal over external entry:

$$\mathbb{E}_f\left[\left(\sum_{l\in L} e_{l,n'}^{\frac{1}{1-\beta}} w_l^{\frac{-\beta}{1-\beta}} \right)^{\frac{(1-\beta)(\sigma-1)}{\theta-\beta(\sigma-1)}} \right] \geq \left(\sum_{k \in L_g} e_{k,n'}^{\frac{1}{1-\beta}} w_k^{\frac{-\beta}{1-\beta}} \right)^{\frac{(1-\beta)(\sigma-1)}{\theta-\beta(\sigma-1)}} + K_{fg},$$

where $K_{fg} = \frac{S_{fg}(\sigma-1)^{\frac{1-\sigma}{1-\beta}}}{\left(\alpha^\beta E \sigma^{\beta-\sigma} P^{1-\sigma} \Gamma\left(1+\frac{1-\rho}{\theta}\right)^{\frac{(1-\sigma)(1-2\beta)}{(1-\rho)}}\right)^{\frac{1}{1-\beta}}}$. Equation (8) obtains. QED. □

Appendix B Internal and external entries

[Insert Figure 1 here]

We analyse firms' vectors of sales and identify a business group G's entry in a new market at t when observing the appearance of a new sector in this vector at t compared to $t-1$.

We adopt the following approach to further distinguish internal from external entry in the new sector. As illustrated in Figure 1, the identification of internal *versus* external entry depends on whether a given firm B belongs to group G one period before we observe the appearance of a new sector in the vector of group G's sales.

On the one hand, Figure 1a illustrates that internal entries occur when group G is composed of several firms at $t-1$ and that one of them, firm B, is observed to enter a new sector at t while remaining in group G.

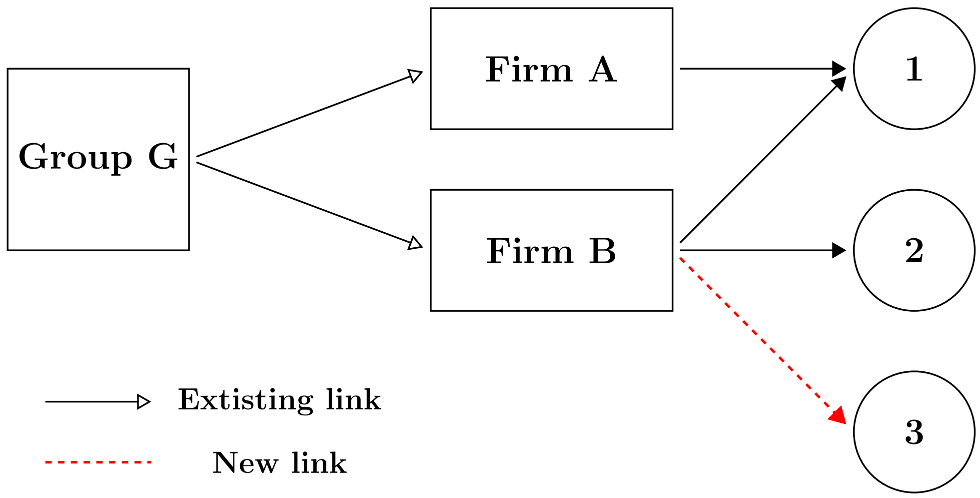
On the other hand, Figure 1b illustrates that external entries occur when a new sector appears in the vector of group G's sales through a given firm B at time t *and* when firm B newly appears in business group G.

As explained in section 3, a change of ownership in the ownership (LIFI) data does not always match with an acquisition in the M&A data (SDC Platnium or BvD Zephyr). Indeed, the appearance of ownership links in LIFI is sometimes due to previously undisclosed financial links. We consider those observations internal entries, and reintegrate this new firm's past vector of sales in the history of group g. This is a conservative choice, which might only lead to a *downward* bias of our coefficient of interest as it may mistake external entries for internal ones but not the opposite.

Appendix C Tables and figures

Figure 1: Example of internal and external entries for group G

(a) Internal entry in sector 3 through firm B



(b) External entry in sector 2 and 3 through acquisition of B

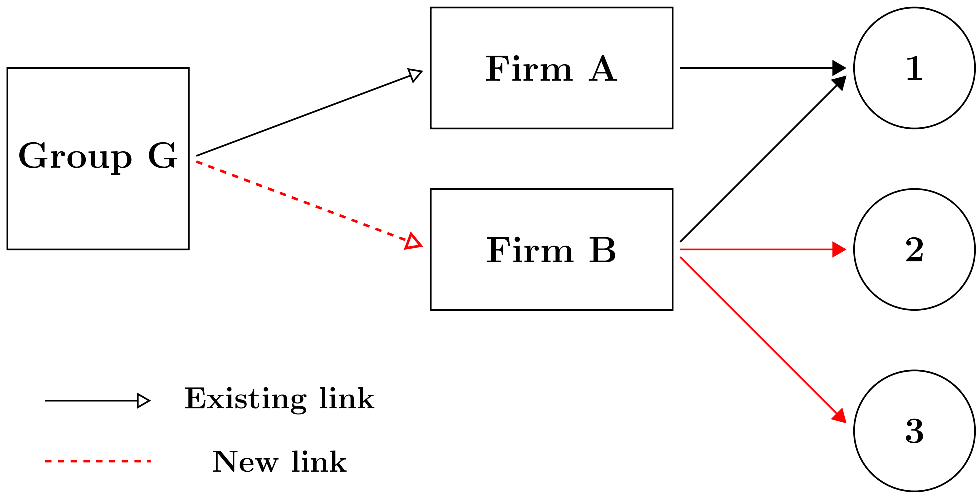


Table 1: Evolution of the external/internal entry ratio

The table reports the number of entries by entry type in the period 2008-2014. Standard industry classification uses the Nomenclature des Activite Francaise (NAF). Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification. External entries are identified with SDC Platinum and Bureau van Dijk Zephyr databases.

(a) Panel A: External/internal ratio by industry

Level sector	4-digit	3-digit	2-digit	1-digit
N. sectors	615	272	88	18
External	2762	2649	2410	2032
Internal	57069	53508	46344	37506
Ratio (%)	4.62	4.72	4.94	5.14

(b) Panel B: External/internal entry ratio by year

Year	2009	2010	2011	2012	2013	Total
External	391	448	552	858	513	2762
Internal	12519	13987	11060	9272	10231	57069
Ratio (%)	3.03	3.10	4.75	8.47	4.77	4.62
Total	12910	14435	11612	10130	10744	59831

(c) Panel B: External/internal entry ratio by year if sales>50k€

	2009	2010	2011	2012	2013	Total
External	256	337	413	617	341	1964
Internal	5341	6945	4783	4001	4153	25223
Ratio (%)	4.57	4.63	7.95	13.36	7.59	7.22
Total	5597	7282	5196	4618	4494	27187

Table 2: Top 10 sectors of entry in 2013

The table reports the number of entries for the top ten sectors by mode of entry in 2013. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification. External entries are identified with SDC Platinum and Bureau van Dijk Zephyr databases.

(a) Panel A: Internal entries

Sector of origin	Sector of entry	N. Pairs origin x entry
Retail sale in stores with food or tobacco	Manufacture of bread, pastry and cakes	880
Retail sale in stores with food or tobacco	Renting and leasing of cars and light motor vehicles	565
Retail sale in stores with food or tobacco	Agents involved in the sale of food	488
Retail sale in stores with food or tobacco	Retail sale of automotive fuel in specialized stores	395
Retail sale in stores with food or tobacco	Retail sale via mail order or Internet	338
Sale of cars and light motor vehicles	Maintenance and repair of motor vehicles	332
Retail sale in stores with food or tobacco	Renting or leased real estate	296
Freight transport by road	Other transportation support activities	295
Freight transport by road	Warehousing and storage	278
Freight transport by road	Maintenance and repair of motor vehicles	272

(b) Panel B: External entries

Sector of origin	Sector of entry	N. Pairs origin x entry
Computer consultancy activities	Computer programming activities	14
Activities of head offices	Wholesale of machinery and equipment	8
Software publishing	Computer programming activities	8
Computer consultancy activities	Software publishing	7
Computer consultancy activities	Business and other management consultancy activities	6
Activities of head offices	Wholesale of other household goods	6
Engineering activities	Programming activities	6
Wholesale of machinery and equipment	Manufacture of lifting and handling equipment	5
Computer consultancy activities	Wholesale of computers, peripheral equipment and software	5
Computer consultancy activities	Other business support service activitiesï£¡	5

Table 3: Top 5 occupations × sector × year fixed effects

The table reports the top five exponentiated occupation × sector × year fixed effects.

(a) Panel A: Manufacture of pharmaceutical preparations (2013)

Occupation	**Estimated fixed effects**
Technicians in production and control quality, processing industries	0.99
Operators and skilled workers, chemistry	0.96
R&D engineers and executives, processing industries	0.61
Commercial managers in SMEs	0.31
Sales representatives and technicians	0.29

(b) Panel B: Computer consultancy activities - Top 5 occupations (2013)

Occupation	**Estimated fixed effects**
R&D engineers and executives, computer science	2.34
IT project manager	1.63
CEOs of service companies (1-49 workers)	0.99
CEOs of commercial companies (1-49 workers)	0.92
Engineers and executives, IT support	0.82

(c) Panel C: Manufacture of motor vehicules - Top 5 occupations (2013)

Occupation	**Estimated fixed effects**
Mechanical qualified assemblers in series	0.99
Manufacturing and quality control technicians in mechanics and metal work	0.85
CEOs of companies (50-499 workers)	0.40
R&D engineers specialized in mechanics and metal work	0.40
Unskilled workers in assembly lines of metal work	0.14

Table 4: Description of variables

Variables	Description
Dependent Variable	
External entry$_{f,n,t}$	Dummy indicating that group g entered sector n at time t through an acquisition. *Source: SDC Platinum, BvD Zephyr, LIFI, EAE, BRN-RSI.*
Independent Variable	
Human capital$_{f,n,t\text{-}1}$	Sum of the exponentiated occupation-sector-year fixed effects for sector n in group g at time $t-1$, scaled by the number of occupations in group g. Higher values indicate a greater fit of the group's employement structure for the sector of entry. *Source: LIFI, DADS.*
Control Variables	
log(#workers)$_{f,t-1}$	Logarithm of the number of workers in group g at time $t-1$. *Source: LIFI, DADS.*
Value added/#workers$_{f,t\text{-}1}$	Total value added generated by group g at time $t-1$, scaled by the number of workers in the group. *Source: LIFI, DADS, BRN-RSI.*
Fixed assets/#workers$_{f,t\text{-}1}$	Total fixed assets held by group g at time $t-1$, scaled by the number of workers in the group. *Source: LIFI, DADS, BRN-RSI.*
Cash holdings/#workers$_{f,t\text{-}1}$	Total cash holdings of firm f at time $t-1$, scaled by the number of workers in the group. *Source: LIFI, DADS, BRN-RSI.*
Distance to the sector of entry$_{f,n,t\text{-}1}$	We explain the variable construction in Section 4.3. A value of 0 (resp. 1) indicates that the sales of the group g are perfectly correlated (resp not correlated) at time $t-1$ to the sales of the firms operating in sector n. *Source: LIFI, DADS, BRN-RSI.*
External finance dependence$_{f,t\text{-}1}$	Sales-weighted average of the external finance dependence ratios of the sectors in which group g operates at time $t-1$. This ratio is computed as the sectoral share of investment of fixed assets that is not financed by operating cash flows. Higher values indicate that the group operates on average in sector with a high dependence to external finance. *Source: LIFI, DADS, BRN-RSI.*
log(Market size)$_{n,t}$	Logarithm of the sum of sales in sector n at time t. *Source: BRN-RSI.*
Herfindahl$_{n,t}$	Herfindahl index in sector n at time t. *Source: BRN-RSI.*

Table 5: Descriptive statistics

Sample: Sample of business groups with an observed internal growth (observed from EAE) or external growth (observed from SDC Platinum or BvD Zephyr) in the 2008 – 2014 period. *Source:* LIFI, DADS, BRN-RSI, SDC Platinum, BvD Zephyr, EAE.

				Percentiles				
Name	#Obs.	Mean	St.Dev.	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
log(Sales in the new sector)$_{f,n,t}$	59571	5.82	2.31	1.75	4.29	5.94	7.42	9.46
Survival probability (1 year)$_{f,n,t\text{-}1}$	59571	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Human capital$_{f,n,t\text{-}1}$	59571	1.20	0.82	0.20	0.64	1.05	1.54	2.77
log(Number of workers)$_{f,t\text{-}1}$	59571	4.95	1.57	3.26	3.78	4.53	5.65	8.18
Value added/#workers$_{f,t\text{-}1}$	59571	0.06	0.03	0.02	0.04	0.05	0.07	0.13
Fixed assets/#workers$_{f,t\text{-}1}$	59571	0.06	0.06	0.00	0.02	0.04	0.07	0.17
Cash holdings/#workers$_{f,t\text{-}1}$	59571	0.02	0.02	0.00	0.00	0.01	0.03	0.07
External finance dependence$_{f,n,t\text{-}1}$	59571	-7.28	3.18	-11.70	-8.96	-7.23	-5.34	-2.76

Table 6: First difference between groups characteristics

Group-level data is taken for the 2008-2014 period. The table display means of business groups' characteristics. The sample is split by mode of entry. The first differences of means are reported in column 5.

	Internal		External		Difference	
	Mean	SD	Mean	SD	Mean	*p*-value
log(Sales in the new sector)$_{f,n,t}$	5.77	2.29	7.54	2.24	-1.77***	(-33.21)
Survival probability (1 year)$_{f,n,t-1}$	0.38	0.49	0.69	0.46	-0.31***	(-28.55)
Human capital$_{f,n,t-1}$	1.21	0.82	0.77	0.58	0.44***	(31.48)
log(Number of workers)$_{f,t-1}$	4.91	1.54	6.13	1.84	-1.22***	(-28.07)
Value added/#workers$_{f,t-1}$	0.06	0.03	0.07	0.04	-0.01***	(-11.00)
Fixed assets/#workers$_{f,t-1}$	0.06	0.06	0.06	0.07	-0.01***	(-3.34)
Cash holdings/#workers$_{f,t-1}$	0.02	0.02	0.03	0.03	-0.01***	(-8.46)
External finance dependence$_{f,n,t-1}$	-7.28	3.19	-7.36	2.97	0.08	(1.09)
Observations	57750		1821		59571	

Figure 2: Density of human capital by mode of entry

The figure displays the PDF of *Human capital* by mode of entry. *Human capital* is the sum of the exponentiated occupation × sector × year fixed effects for sector n in group g at time $t-1$, scaled by the number of occupations in group g. Higher values indicate a greater fit of the group's employment structure to the sector of entry. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification.

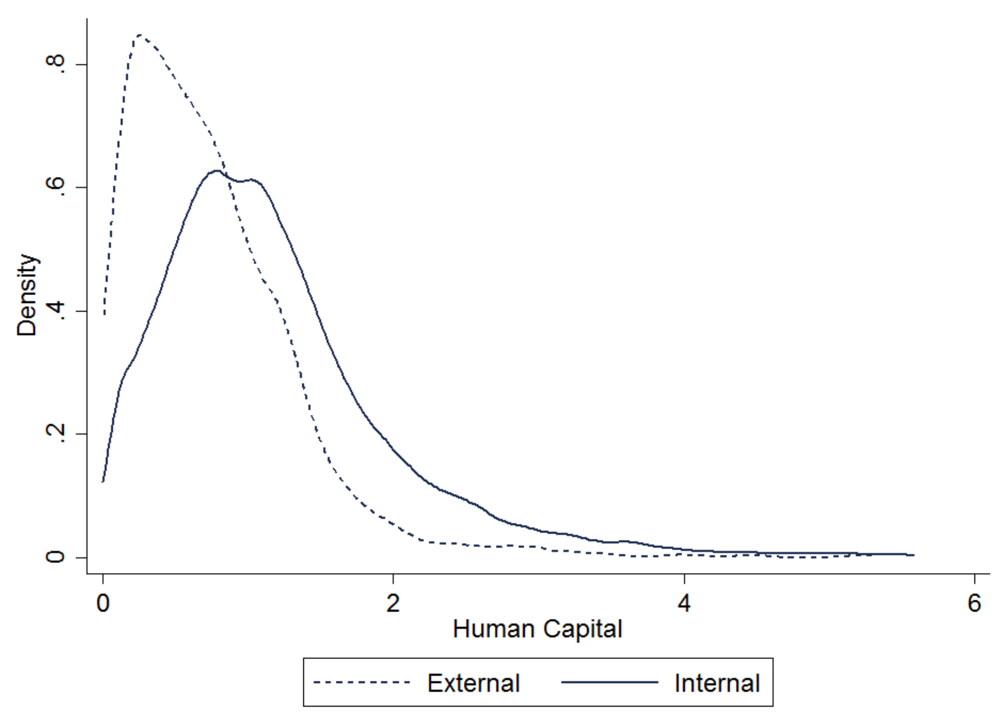


Table 7: Firms entry type and human capital

Group-level data is taken for the 2008-2014 period. External entry by group g in sector n year t is a dummy variable which takes the value one if the entry is made through an acquisition, zero if the entry is made internally. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification. *Human capital* is the sum of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector n in group g at time $t-1$, scaled by the number of occupations in group g. Higher values indicate a greater fit to the group's employment structure for the sector of entry. *Terciles of #workers* are dummy variables corresponding to terciles of firm f's number of employees at time $t-1$; *Cash holdings/#workers* is the total cash holding of firm f at time $t-1$, scaled by the number of workers in the group; *Fixed assets/#workers* is the total fixed assets held by group g at time $t-1$, scaled by the number of workers in the group; *Value added/#workers* is the total value added generated by group g at time $t-1$, scaled by the number of workers in the group; *External finance dependence* is sales-weighted average of the external finance dependence ratios of the sectors in which group g operates at time $t-1$. In columns 4-5-6 we run the baseline model on subsamples corresponding to each tercile of the employee number of firm f at time $t-1$. Reported regressions control for *sector of entry*, *sector of origin*, and *year* fixed effects. All standard errors are clustered at the level of the sector of entry. *, **, and *** denote statistical significance at 10, 5 and 1%. Standard errors are given in parentheses.

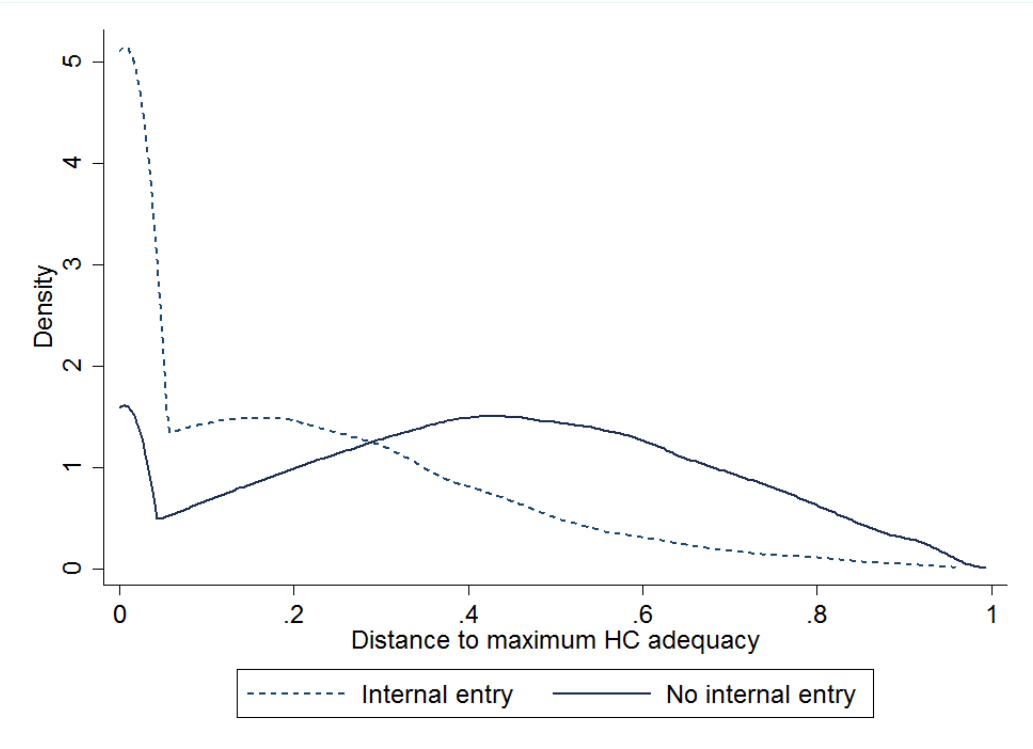
	(1)	(2)	(3)	(4)	(5)	(6)
Human capital$_{f,n,t\text{-}1}$	-0.005^{***}	-0.001	-0.001	-0.003^{***}	-0.005^{***}	-0.008^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
2nd tercile of #workers$_{f,t\text{-}1}$	0.013^{***}	0.014^{***}	0.014^{***}			
	(0.002)	(0.002)	(0.002)			
3rd tercile of #workers$_{f,t\text{-}1}$	0.045^{***}	0.041^{***}	0.042^{***}			
	(0.003)	(0.003)	(0.003)			
2nd t. #workers*Human capital$_{f,n,t\text{-}1}$		-0.002^{*}	-0.002^{*}			
		(0.001)	(0.001)			
3rd t. #workers*Human capital$_{f,n,t\text{-}1}$		-0.013^{***}	-0.013^{***}			
		(0.003)	(0.003)			
Cash holdings/#workers$_{f,t\text{-}1}$			0.158^{***}	0.118^{**}	0.120^{**}	0.173^{*}
			(0.047)	(0.059)	(0.054)	(0.095)
Fixed assets/#workers$_{f,t\text{-}1}$			-0.013	-0.001	-0.051^{*}	0.022
			(0.018)	(0.023)	(0.030)	(0.040)
Value added/#workers$_{f,t\text{-}1}$			0.023	-0.003	0.111^{**}	0.001
			(0.039)	(0.059)	(0.054)	(0.074)
External finance dependence$_{f,n,t\text{-}1}$			-0.000	0.001^{*}	-0.001	-0.001
			(0.000)	(0.000)	(0.001)	(0.001)
Sector of origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector of entry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.090	0.091	0.091	0.093	0.122	0.133
Observations	59601	59601	59601	20129	19556	19777

Table 8: Robustness check: definition of human capital

Group-level data is taken for the 2008-2014 period. *External entry* by group g in sector n year t is a dummy variable which takes the value one if the entry is made through an acquisition, zero if the entry is made internally. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification. The table presents regression of *External entry* on alternative measure of the variable *Human capital*. *No Top 10* is a dummy which takes the value one if the group g does not employ any of the 10 most important occupation for sector n at time $t-1$. In *Top 10 only*, the summation in the computation of *Human capital* is made only over the top 10 occupations of the sector. In *Share* we replace the exponentiated fixed effects by the average share of the payroll attributed to the occupation in the sector. *Weighted* is the average of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector n in group g at time $t-1$, weighted by the number of employees by occupation in the group. *Without CEOs* exclude CEO-type occupations from the sum of occupation $\times$ sector $\times$ year fixed effects. *Manual* excludes manually coded occupation codes from the sum of occupation $\times$ sector $\times$ year fixed effects. *Plant-level* is the sum of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector n in group g at time $t-1$ with occupation $\times$ sector $\times$ year fixed effects estimated at the plant level. Reported regressions control for *terciles of #workers*, *Cash holdings/#workers*, *Fixed assets/#workers*, *Value added/#workers*, *External finance dependence*, *sector of entry*, *sector of origin*, and *year* fixed effects. All standard errors are clustered at the level of the sector of entry. *, **, and *** denote statistical significance at 10, 5 and 1%. Standard errors are given in parentheses.

	Dependent variable: External entry$_{f,n,t}$						
Measure of Human capital:	No Top 10	Top 10 only	Share	Weighted	Without CEOs	Manual	Plant-level
Human capital$_{f,n,t\text{-}1}$	0.024***	−0.001*	−0.010***	−0.004***	−0.004***	−0.004***	−0.006***
	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Sector of origin FE	Yes	Yes	Yes	Ye s	Yes	Yes	Yes
Sector of entry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.100	0.035	0.102	0.090	0.092	0.090	0.091
Observations	64261	21251	62473	59508	63535	58306	57461

Figure 3: Density of the within-group distance to closest human capital



The figure displays the PDF of firms' *Distance to max Human Capital* depending on whether firms enter internally a new sector. *Distance to max Human Capital* is the difference between the highest *Human capital* of firm $f*$ in group g and the *Human capital* of a given firm f. The distance is equal to zero for the firm $f*$ in group g which has the highest *Human capital* and approaches one when *Human capital* tends to zero.

Table 9: Within group internal entries and firms' human capital

Firm-level data is taken for the 2008-2014 period. *Internal entry* by firm f of firm f in sector n year t is a dummy variable which takes the value one if the internal entry is made though the firm, zero otherwise. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification. *Distance to max Human Capital* is the difference between the highest *Human capital* of firm $f*$ in group g and the *Human capital* of a given firm f. The distance is equal to zero for the firm $f*$ in group g which has the highest *Human capital* and approaches one when *Human capital* tends to zero. Reported regressions control for *log of #workers*, *Cash holdings/#workers*, *Fixed assets/#workers*, *Value added/#workers* and *External finance dependence* computed at firm f level. In addition, columns 1-2-3 include *sector of entry*, *sector of origin* and *year* fixed effects. Column 4 includes *sector of entry* $\times$ *sector of origin* and *year* fixed effects. All standard errors are clustered at the level of the group. *, **, and *** denote statistical significance at 10, 5 and 1%. Standard errors are given in parentheses.

	Dependent variable: Internal entry in the new sector$_{f,n,t}$		
	(1)	(2)	(3)
Distance to max Human Capital$_{f,n,t\text{-}1}$	-0.274^{***}	-0.275^{***}	-0.193^{***}
	(0.014)	(0.014)	(0.011)
log(Number of employees)$_{f,t\text{-}1}$		0.014^{***}	0.018^{***}
		(0.002)	(0.002)
Cash holdings/#workers$_{f,t\text{-}1}$		1.355^{***}	0.994^{***}
		(0.283)	(0.264)
Fixed assets/#workers$_{f,t\text{-}1}$		-0.019	0.028
		(0.028)	(0.027)
Value added/#workers$_{f,t\text{-}1}$		-0.025	0.027
		(0.027)	(0.028)
Sector of origin FE	Yes	Yes	No
Sector of entry FE	Yes	Yes	No
Year FE	Yes	Yes	Yes
Origin x Entry x Year FE	No	No	Yes
Observations	197612	197612	185686
R^2	0.115	0.120	0.221

Figure 4: Density of human capital by mode of entry

The figure displays the PDF of *Human capital* by mode of entry. *Human capital* is the sum of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector n in group g at time $t-1$, scaled by the number of occupations in group g. Higher values indicate a greater fit of the group's employment structure to the sector of entry. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification.

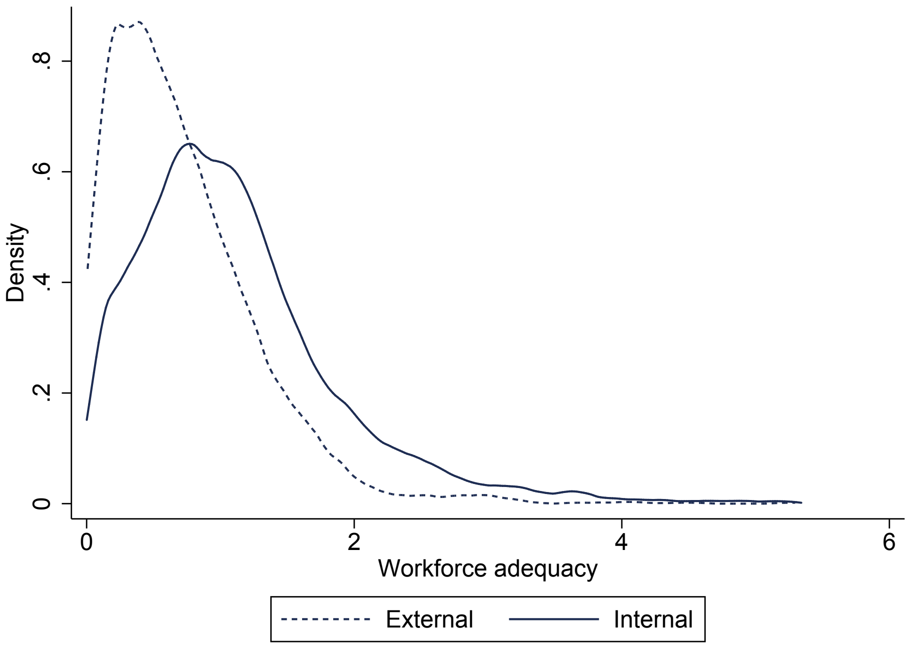


Table 10: Firms' entry type: interactions between market of origin and market of entry

Group-level data is taken for the 2008-2014 period. External entry by group g in sector n year t is a dummy variable which takes the value one if the entry is made through an acquisition, zero if the entry is made internally. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification. *Human capital* is the sum of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector n in group g at time $t-1$, scaled by the number of occupations in group g.*Log(Market size)* is the logarithm of the sum of sales in sector n at time t; *Herfindahl* is the Herfindahl index in sector of entry n at time t. Regressions control for *terciles of #workers*, *Cash holdings/#workers*, *Fixed assets/#workers*, *Value added/#workers*, *External finance dependence*. Reported models include *sector of entry*, *sector of origin*, and *year* fixed effects. All standard errors are clustered at the level of the sector of entry. *, **, and *** denote statistical significance at 10, 5 and 1%. Standard errors are given in parentheses.

	(1)	(2)	(3)	(4)
Human capital$_{f,n,t\text{-}1}$	-0.007^{***}	-0.004^{***}	-0.004^{***}	-0.003^{**}
	(0.001)	(0.001)	(0.001)	(0.001)
log(Market size)$_{n,t}$	-0.004^{***}			
	(0.001)			
Herfindahl$_{n,t}$	0.021			
	(0.019)			
Distance to the sector of entry$_{f,t\text{-}1}$	0.048^{***}	0.036^{***}		
	(0.008)	(0.009)		
2nd tercile of distance$_{f,t\text{-}1}$			0.006^{***}	
			(0.002)	
3rd tercile of distance$_{f,t\text{-}1}$			0.008^{***}	
			(0.002)	
Control variables	Yes	Yes	Yes	Yes
Sector of origin FE	No	Yes	Yes	No
Sector of entry FE	No	Yes	Yes	No
Year FE	Yes	Yes	Yes	No
Origin x Entry x Year FE	No	No	No	Yes
R^2	0.032	0.091	0.091	0.180
Observations	59704	59567	59601	41456

Table 11: Firms' entry type: lag values of human capital

Group-level data is taken for the 2008-2014 period. External entry by group g in sector n year t is a dummy variable which takes the value one if the entry is made through an acquisition, zero if the entry is made internally. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Internal entries are identified using reported sales from the ESA survey at the four-digit standard industrial classification. Variables *Human capital* are the sum of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector n in group g at, respectively time $t-1$, $t-2$ and $t-3$, scaled by the number of occupations in group g. Regressions control for *terciles of #workers*, *Cash holdings/#workers*, *Fixed assets/#workers*, *Value added/#workers*, *External finance dependence*. Reported models include *sector of entry*, *sector of origin*, and *year* fixed effects. All standard errors are clustered at the level of the sector of entry. *, **, and *** denote statistical significance at 10, 5 and 1%. Standard errors are given in parentheses.

	Dependent variable: External entry$_{f,n,t}$				
	(1)	(2)	(3)	(4)	(5)
Human capital$_{f,n,t\text{-}1}$	-0.005^{***}			-0.006^{***}	-0.003^{*}
	(0.001)			(0.002)	(0.002)
Human capital$_{g,n,t\text{-}2}$		-0.002^{**}		0.001	
		(0.001)		(0.001)	
Human capital$_{g,n,t\text{-}3}$			-0.004^{***}		-0.002
			(0.001)		(0.001)
Sector of origin FE	Yes	Yes	Yes	Yes	Yes
Sector of entry FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
R^2	0.090	0.104	0.119	0.104	0.119
Observations	59601	47190	34053	46534	33469

Figure 5: Evolution of human capital around internal entries

The figure displays the evolution of firm f's number of employees and human capital around an internal entry in sector n at year 0.

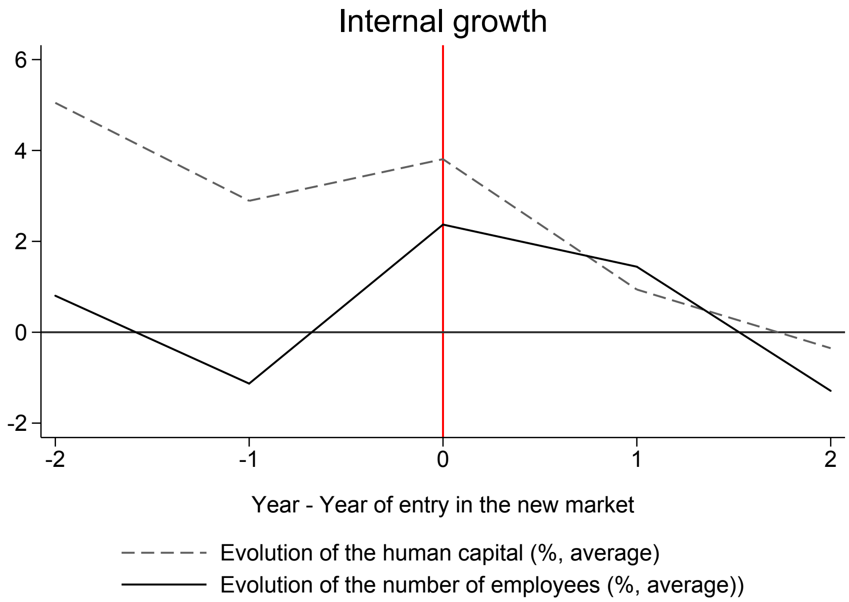


Figure 6: Evolution of human capital around external entries

The figure displays the evolution of firm f's number of employees and human capital around an external entry in sector n at year 0.