Build or Buy?
Human Capital and Corporate Diversification

Paul Beaumont Camille Hebert Victor Lyonnet*

21st May, 2019

Abstract

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? Using a novel data set constructed by merging French employer payrolls with commercial M&A data sets, we show that firms are more likely to buy when their existing workforce does not include skills needed in the sector of entry. This relationship is more pronounced when labor market frictions make it difficult to hire key workers. Firms that enter by building realize lower entry sales when their existing workforce is not adapted to the sector of entry, especially in the presence of labor market frictions. Our results suggest that firms buy to acquire their targets' human capital when adapting their existing workforce is too costly.

Keywords: Mergers and Acquisitions, Corporate diversification, Product market entry, Occupational human capital, Tight labor market

*Paul Beaumont is at Université Paris Dauphine and CREST (paul.beaumont@dauphine.eu). Camille Hebert is at Université Paris Dauphine and Tilburg University (c.hebert@tilburguniversity.edu). Victor Lyonnet is at the Ohio State University (lyonnet.1@osu.edu). We thank Manuel Adelino, Philippe Aghion, Fabio Braggion, Edouard Challe, Gilles Chemla, Espen Eckbo, Simon Gervais, Edith Ginglinger, John Graham, Denis Gromb, Johan Hombert, Julien Martin, Adrien Matray, Isabelle Méjean, Mario Milone, Tamara Nefedova, Paige Ouimet, Manju Puri, Amiyatosh Purnanandam, Adriano Rampini, Luc Renneboog, David Robinson, René Stulz, Geoffrey Tate, Vish Viswanathan and Sheng-Jun Xu for valuable comments and helpful discussions. We thank also seminar participants at CREST, Duke University, the 5th ECGC workshop on Governance and Control, HEC Paris, INSEE, the 28th Mitsui Labor and Finance Symposium, NFA, Ohio State University, Université Paris-Dauphine, UT Austin Corporate Finance Forum, Tilburg University. This paper was awarded best student paper award at the 5th ECGC workshop on Governance and Control. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'avenir" program (reference: ANR-10-EQPX-17). We are responsible for all remaining errors and omissions.

1. Introduction

In 2017, more than twelve thousand cross-industry merger and acquisition (M&A) deals occurred worldwide for a total value of more than 900 billions of dollars.[1] Cross-industry M&As allow firms to diversify, to obtain specific assets and access new markets.[2] However, the magnitude of these figures can be misleading, concealing the fact that most diversifications result from firms entering new sectors directly by building on their existing capabilities. Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? This question has major implications for understanding the role played by M&As in the reallocation of resources in the economy.

In this paper, we compare firms that enter a new sector by building on existing resources ("build") to firms that acquire an already existing company operating in the sector of entry ("buy"). We focus on the role of human capital as a key resource to successfully operate in the new sector. Although our analysis applies to other assets, the availability of employee-level information and the existence of specific labor market frictions motivate our focus on human capital. We study the costs and benefits associated with each alternative. When a firm buys to enter a new sector, it has to incur both the costs of acquiring and restructuring the target, but gets access to its productive resources.[3] When a firm builds on its existing resources to enter a new sector, it faces the search costs of hiring new workers to complement its existing workforce.[4] Therefore, the more key workers in a firm's existing workforce, the lower the adjustment costs to the new sector and the more profitable it is for the firm to build.

We find that firms that have a workforce more adapted to the sector of entry are more likely to "build" rather than to "buy". The relation between internal human capital and the mode of entry in a new sector is stronger when specific skills are difficult to obtain on the external job market, i.e., when key worker occupations are in short supply. Firms that build have lower entry sales when their workforce is not adapted to the sector of entry, and especially when key workers are in short supply. These findings are consistent with the idea that firms must search for key assets to expand into a new sector. When search costs increase, buying an existing firm becomes relatively more attractive than building from scratch and hiring new workers.[5]

[1] Cross-sector M&As represent about 30% of the universe of M&A deals both in frequency and in volume. See KPMG M&A Predictor 2018 Annual Report.

[2] The motivations behind horizontal M&As are different and relate mostly to market power consolidation (see e.g., Eckbo, 1983; Farrell and Shapiro, 1990).

[3] Maksimovic, Phillips and Prabhala (2011) shows that acquiring firms enter a costly restructuring process and sell or close 46% of the plants they buy.

[4] Hiring costs have been estimated to be between one and two quarters of wage payments, and increasing in the specificity of skills (Abowd and Kramarz, 2003; Blatter, Muehlemann and Schenker, 2012).

[5] Cisco's acquisition strategy illustrates this trade-off. In 1997, a Cisco analyst described the strategy of the

One key challenge to test this hypothesis consists in defining human capital and measuring it at the firm level. Human capital is neither directly observable nor easily defined. To overcome this difficulty, we propose a measure based on a model of diversification with endogenous choice of teams. In the model, the firm hires workers from different occupations that differ in their sector-specific skills. If the firm decides to build, it can draw workers from its existing pool of workers (internal labor market) or from the external job market. Instead if the firm decides to buy, it can pick workers in the workforce of the acquired firm but has to pay the fixed costs of acquiring another firm. The model draws two predictions. First, in equilibrium, the firm chooses to build when its existing *internal human capital* is adapted to operate in the sector of entry. Instead, the firm buys when its workforce does not include the key worker occupations needed to operate in the sector of entry, and this despite restructuring costs. Second, the relationship between human capital and the decision to build or buy is stronger when it is difficult to hire key workers in the external job market.

The model yields a firm-level measure of human capital that stems from the relation between the fraction of the wage bill that goes to a given type of workers and the contribution of these workers to the firm's output. Empirically, we construct our measure of human capital in two steps. First, we use the matched employer-employee dataset from France. In the spirit of Abowd, Kramarz and Margolis (1999), we regress the (log) fraction of the wage bills of all firms in a sector that goes to a given occupation, on occupation $\times$ sector fixed effects. We interpret the occupation $\times$ sector fixed effects estimates as a score reflecting the sector-specific human capital of worker occupations. The higher the fixed effect of a given occupation in a sector, the larger the *average* share of the firms' wage bill that goes to the corresponding worker type in the sector. Second, we compute human capital at the firm level. Firm's internal human capital is the average of the (exponentiated) occupation $\times$ sector fixed effects present in the firm's workforce and the sector of entry. Our measure of human capital captures the extent to which the existing workforce is adapted to the sector of entry.[6]

Our measure of human capital has several advantages. First, because the French matched employer-employee dataset contains a detailed occupation code at the individual level, our measure of human capital encompasses various determinants of individuals' human capital, including skills, education and experience (Becker, 1962; Gibbons and Waldman, 2004; Autor and Dorn,

firm as follows: "in today's economy, building work teams from scratch can be yesterday's luxury. So, when you can't build fast enough, you buy". By 2017, Cisco had undertaken more than 200 M&As to "provide a capability, acceleration potential or earlier sectoral entry compared to partnering or developing in-house" (Wysocki, 1997; Romanski, 2017).

[6]Thereafter, we refer to this variable as "internal human capital" without mentioning that it is specific to the sector of entry.

2009). Second, the fixed effects strategy allows to rank occupations within sector following their contribution to firms' outputs. Third, our measure of human capital does not reflect firm-level personnel policies or influence of unions that may alter the returns to observable and unobservable dimensions of human capital.

Another empirical challenge of our paper is to identify how firms diversify. We use the firms' detailed breakdown of sales across sectors to identify entries in new sectors.[7] A firm enters a new sector if (i) at least one of its subsidiary starts selling in the new sector (ii) none of the other subsidiaries controlled by the same ultimate owner operates in the sector of entry. We identify buy entries by linking M&A deals retrieved from SDC Platinum and Bureau van Dijk Zephyr to the French administrative data.[8] The entry is a "buy" if the subsidiary which starts selling in the new sector has been acquired by the firm. By contrast, a firm "builds" if the entry is made though an existing subsidiary. Our final data set consists of 75,000 build or buy decisions in France from 2003 to 2014.

We provide the first cross-industry statistics on firms' decisions to build or buy. We find that 98% of entries in a new sector consisted of building firms, or 90% when weighting by entry sales. On average, firms invest a million euros at the time of entry, and about half of the firms in our sample continue to operate in the sector of entry after a year. Median (mean) entry sales are equal to 270 thousands euros (2.8 millions euros).

We test whether firms' existing workforce composition explain to their decision to build or buy. We compare firms that operate within the same sector of origin, diversify within the same sector of entry, the same year. This specification neutralizes potential unobservable time-varying synergies between the sector of entry and that of origin. We find that firms are more likely to buy when their human capital is not adapted to the sector of entry. A one standard deviation lower human capital is associated with an increase in the likelihood of buying by 1.1 percentage points. This relationship is sizeable, equal to 50% of the unconditional probability to buy. This finding holds when controlling for a wide variety of firm characteristics such as size, profitability, capital intensity, cash holdings and other assets.

We perform a battery of robustness tests on our definition of human capital and on its explanatory power on the firms' decision to build or buy. First, we test alternative measures of human capital. We show that our findings are robust to using alternative weights to account for the exact composition of the workforce, to focusing on key occupations (highest fixed effects), to

[7]The firms' detailed breakdown of sales across sectors is available in the ESA survey, also maintained by the French Bureau of Statistics (Insee). In our main analysis, we define sectors as a 5-digit code of the French SIC.

[8]We develop a web-crawler that takes acquiring and target firms' names and addresses, to link SDC Platinum and Zephyr deals to the French administrative data.

excluding CEOs, whose wages may not only reflect productivity but also agency conflicts, and to varying the unit of observation from firms to plants. Second, we exploit heterogeneity in the workforce composition of subsidiaries within firm and we show that the entry into the new sector is more often made though the subsidiary that has the more adapted internal human capital, certainly to minimize internal labor market reallocation costs (Cestone et al., 2016). Third, we check that building firms with a low internal human capital hire relatively more new workers on the external job market than firms with a high preexisting internal human capital. Fourth, we investigate the effects of financial constraints on the decision to build or buy. We show that our results hold irrespective of the firm's size, cash holding levels and access to public markets. Fifth, we mitigate concerns related to complementarities between human and physical capital. A firm with a poorly adapted human capital is also likely to have inadequate physical capital (equipment, machinery, etc.). To isolate the role of human capital, we compare firms that invest similar amounts at the time of entry. We still find that firms are more likely to buy when their internal human capital is not adapted to the sector of entry.

In an additional set of robustness tests, we try to mitigate concerns relative to the joint decision to diversify and how to diversify. First, we investigate how the complementarities between the sector of entry and the sector of origin (e.g., production synergies, common customers) interacts with human capital relatedness. We build on Bloom, Schankerman and Van Reenen (2013) to construct a distance between firms' initial activities and the sector of entry.[9] We find that firms are more likely to buy rather than to build when it enters a distant sector. However, using interacted coefficients, we show that the magnitude of the human capital coefficient increases with the distance to the sector of entry. This finding suggests that buying is a more attractive option when the cost of building is high. Second, we focus on two plausible scenarii in which firms are likely to jointly select the sector of entry as well as the type of entry.[10] In a first scenario, a firm that operates in a declining sector may be willing to shift operations and to reallocate workers to a better performing sector (Tate and Yang, 2016*b*; Baghai et al., 2018). Thus, the firm would pick a sector in which it can easily redeploy its workforce. This scenario

[9]In the spirit of Bloom, Schankerman and Van Reenen (2013), the distance is based on firms' breakdown of sales. It measures the correlation between the product market portfolio of a firm prior to entry and product market portfolios of firms already operating in the sector. We are not able to replicate the Hoberg and Phillips (2010)'s product market distance, as this measure uses textual analysis on 10k filings of U.S. firms that are not available for the universe of French firms.

[10]A classic solution to address potential selection issues is to find variables that affect the selection stage in a sector but that are plausibly orthogonal to the main dependent variable. For instance, Tate and Yang (2016*b*) uses the Tobin's Q as an instrument in a two-stage Heckman selection model. Applied to our context, we would need to find an instrument that affects the choice to diversify but not the type of entry. In addition, this instrument must be sector-specific, as we would need to instrument not only the choice of diversification but the choice to diversify in a given sector. Finding such an instrument appears to be a difficult task, hence we choose to focus on specific scenarii in which selection issues could plausibly arise.

would lead to a positive association between the likelihood to build and the human capital of the firm. In a second scenario, a serial acquirer may always build irrespective of human capital considerations (Golubov, Yawson and Zhang, 2015). In additional set of tests, we exclude serial acquirers and firms that shift a substantial part of their operations in another sector. Our main results remain unchanged, which suggests a limited role for selection issues.

Our second key finding is that firms are more likely to buy when it is costly to hire key workers on the external job market, i.e., when key worker occupations are in short supply. We test this prediction using occupation-level data on local labor market (LLM) tightness obtained from the French unemployment agency (Pôle Emploi). The data reports worker occupations in short supply in 350 commuting zones. We document a wide geographic heterogeneity in worker availability across local labor markets (Moretti, 2010). We define a measure of LLM tightness as the human-capital weighted average of occupations in short supply in the zip code where the expanding firm enters. Our measure of LLM tightness takes higher values if key occupations for the sector of entry are in short supply in the LLM. We test whether the LLM tightness measure explains the build or buy decision and we interact it with the measure of human capital. We find that firms are more likely to buy when labor markets are tight, and that the relationship between human capital and the choice to build or buy is driven by the highest tercile of labor market tightness.

Finally, we explore the implications in terms of performance of firms that build on their own resources. In presence of tight labor markets, hiring is costly. Therefore, firms with less adapted human capital grow relatively more slowly in the sector of entry than firms with a highly adapted human capital. In particular, entry sales are smaller for firms with a less adapted internal human capital. Moreover, the results show that this relationship is driven by entries in tight LLMs. Taken together, these findings suggest that labor market frictions are crucial to understand the role of human capital in the decision to build or buy.

Our paper contributes to the literature on the determinants of the decision to build or buy across sectors. Very few papers have studied together these alternatives to diversify. McCardle and Viswanathan (1994) shows that firms enter a new sector through acquisition when barriers of entry are high.[11] Phillips and Zhdanov (2013) shows that in equilibrium small firms invest in R&D, whereas large firms buy those small firms that have successfully innovated. Finally, Bernard, Redding and Schott (2010) document that US multi-product firms often vary their

[11]In the strategy literature, Yip (1982)'s findings support McCardle and Viswanathan (1994)'s theoretical findings and show that firms are more likely to enter internally (build) a sector with low barriers to entry. Lee and Lieberman (2010) finds that firms are more likely to buy unrelated product in the telecommunication industry. In the international economics literature, Nocke and Yeaple (2007) studies the choice of entering a new country by foreign direct investment or by acquiring an already existing company.

product mix but infrequently through the acquisition of a new business line (7% of the cases). We contribute to this small strand of the literature by documenting that, at the scale of the French economy, more than 90% of the corporate diversifications are made by firms that build from scratch using their preexisting resources rather than buying an incumbent in the sector of entry.

Our paper also relates to the theory of the firm. In the finance literature, the dominant view has been the "property rights" theory (Grossman and Hart, 1986; Hart and Moore, 1990; Rhodes-Kropf and Robinson, 2008). Value is created by binding together complementary assets under the control of a single firm.[12] However, another view, popular in early economics and strategy literature (Capron and Mitchell, 2012, see), is gaining attention in the trade and international economics literature. The "resource-based" view of the firm states that the decision to grow depends on pre-existing resources and on transferable capabilities (Penrose, 1955). Recent research has documented that firms are much more likely to produce in certain pairs of industries (see e.g., Bernard et al., 2007, 2018). Boehm, Dhingra and Morrow (2019) goes a step forward and shows that firms tend to co-produce in industries that require similar intermediate inputs. Our paper focuses on another type of input: labor and we show that firms are more likely to diversify by building on existing resources when it exists complementarities between the firm's internal human capital and key skills to produce in the new sector. Thus, our results imply that firms buy when they do not have the adequate inputs in house and even more, when these inputs are scarce on the external market.

Finally, our paper adds to the rapidly growing literature on labor and corporate finance.[13] More specifically, our paper is related to a strand of papers that link organization of firms and labor.[14] Tate and Yang (2016*a*) predicts diversified M&As by cross-industry labor flows, since human capital is transferable between these sectors. Ouimet and Zarutskie (2016) finds that firms that are likely to invest in labor tend to retain a larger fraction of skilled employees after the merger. Lee, Mauer and Xu (2018) finds that firms are more likely to merge and

[12]Rhodes-Kropf and Robinson (2008) show that the search for complementarities implies the existence of an assortative matching between acquirers and target firms, i.e.,"like buys like". An important body of empirical M&A research has found evidence supporting this prediction: Mergers are more likely to occur and more value-creating when merging firms have similar human capital (Tate and Yang, 2016*b*; Lee, Mauer and Xu, 2018), sell similar products (Hoberg and Phillips, 2010), use similar technology (Bena and Li, 2014), or share similar corporate cultures (Li et al., 2018).

[13]A growing literature has recently explored the implications of labor adjustment costs for the cross-section of expected returns (Eiling, 2013; Donangelo, 2014; Belo, Lin and Bazdresch, 2014; Kuehn, Simutin and Wang, 2017)), corporate investment (e.g., Merz and Yashiv, 2007; Xu, 2018; Bai, Fairhurst and Serfling, 2018) and capital structure decisions such as leverage (e.g., Matsa, 2010; Agrawal and Matsa, 2013; Simintzi, Vig and Volpin, 2014; Baghai et al., 2018; Serfling, 2016) and cash holdings (Ghaly, Anh Dang and Stathopoulos, 2017).

[14]Other papers study the consequences of organizational changes, such as M&As, on employment and wages (e.g., Lagaras, 2017; Ma, Ouimet and Simintzi, 2018).

have better post-merger outcomes when the target firm has a related human capital. We make three contributions to this literature. First, we propose a firm-level measure of human capital that captures complementarities between sectors. Second, we show that internal human capital resources predict how firms diversify. Third, our results emphasize the role of labor market frictions in the link between human capital and M&A decisions. Taken together, we show that employment composition shapes the boundaries of the firm.

2. Theoretical framework: "Build" or "Buy"?

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? We propose a model that predicts firms' decision to build or buy based on the adaptability of their workforce for the new sector.

2.1. Basic framework

Costs and profits. To enter a new sector, firms must develop new productive capacities, that is, combine additional inputs to produce in the new sector. In the model, we assume labor L is the only factor of production so the production function is $Y = \mathcal{L}$.[15] Firms can select workers from three different pools of workers. First, firms can reallocate workers from their *internal* labor market and have them produce in the new sector. Second, firms can hire new workers on the *external* labor market, that is, poach workers already employed or hire unemployed workers. Third, firms can *acquire* an existing company to have workers from the acquired firm produce in the new sector. The production function combines these three sources of labor with a constant elasticity of substitution:

$$\mathcal{L} = \left(\sum_i L_i^{\gamma} \right)^{\frac{1}{\gamma}}, \tag{1}$$

where $0 < \gamma < 1$ is the elasticity of substitution across worker pools, and $i \in \{I, E, A\}$ denotes the different pools: internal, external, and acquired. Firms choose their mix of workers from the three pools to minimize their marginal cost of production c, defined as

$$c = \left(\sum_i c_i^{-\frac{\gamma}{1-\gamma}} \right)^{-\frac{1-\gamma}{\gamma}}, \tag{2}$$

where c_i denotes workers' marginal cost of production in each pool $i \in \{I, E, A\}$. We provide microeconomic foundations for each labor type's marginal cost of production in section 2.3.

[15] Note that in the empirical analysis we consider the interactions between human and physical capital.

We assume firms compete monopolistically over product varieties and consumer preferences exhibit constant elasticity of substitution (CES) across products, so that firms' profit is proportional to their marginal cost of production c:

$$\Pi = Kc^{-(\sigma-1)}, \tag{3}$$

where σ is the elasticity of substitution between goods within a sector, $\sigma > 1$ and K is a constant measuring firms' profitability.[16]

Build. If firms choose to enter the new sector by using their existing resources ("build"), they can choose to combine workers from their internal labor market (L_I) as well as the external labor market (L_E). We assume building does not have any impact on firms' existing business lines. This assumption is realistic if internal workers are reallocated away from tasks that are either completed or sufficiently routinized, so that the new production does not disrupt the existing ones. When firms build, their profit can be written as

$$\Pi^{Build} = K\left(c_I^{-\frac{\gamma}{1-\gamma}} + c_E^{-\frac{\gamma}{1-\gamma}}\right)^{\frac{(1-\gamma)(\sigma-1)}{\gamma}}. \tag{4}$$

Equation (1) implies the three pools of workers are combined according to the CES parameter γ. Therefore, on top of consumers' elasticity of substitution between products σ, firms' monopolistic profit in (4) depends on the marginal costs of production of the different worker pools, aggregated according to a function of γ.

Buy. If firms choose to enter the new sector by acquiring an existing company ("buy"), they can access the target's pool of workers (L_A) on top of their internal labor market (L_I) and the external labor market (L_E). Accessing target workers is profitable, because their human capital is likely to be more adapted to the sector of entry, as discussed in section 2.3. However, we assume integrating a target's pool of workers to acquiring firms' internal labor market involves post-merger restructuring costs. We model these costs as a fixed cost F.[17] When firms buy,

[16] In standard monopolistic competition models, this constant depends on σ as well as the share of consumers' income devoted to a given product. In turn, this share depends on the relative price of a given product compared to the aggregate price index. All that matters for our analysis is that this constant does not depend on the marginal cost, as is the case in standard monopolistic competition models.

[17] In Appendix A.2, we show that our model's predictions remain the same when post-merger restructuring costs are modeled as the costs of restructuring the workforce of the acquired firm.

their profit can be written as

$$\Pi^{Buy} = K\left(c_I^{-\frac{\gamma}{1-\gamma}} + c_E^{-\frac{\gamma}{1-\gamma}} + c_A^{-\frac{\gamma}{1-\gamma}}\right)^{\frac{(1-\gamma)(\sigma-1)}{\gamma}} - F. \tag{5}$$

The acquiring firms' access to target's worker pool increases the variable part of (5). When firms buy, we can show theoretically that they reallocate relatively less workers from their internal labor market, and they hire relatively less workers on the external labor market. Moreover, firms systematically lay off a positive number of their target's workers. These findings provide a coherent interpretation of F as the post-merger restructuring costs, and relates to Lee, Mauer and Xu's (2018) findings that removing duplicate workers is a key source of economies of scope in mergers.

2.2. Testable predictions: Build or Buy?

When entering a new sector, firms compare the profit from building on their existing resources ("build") with the realized profit from acquiring an existing company ("buy"). We show in Appendix A.1 that if the different pools of workers are sufficiently substitutable ($\gamma > (\sigma-1)/\sigma$) and F is low enough, the model implies there exists a unique threshold $c_I^* > 0$ such that if $c_I > c_I^*$, we have $\Pi^{Buy} > \Pi^{Build}$. Otherwise, if $c_I < c_I^*$, we have $\Pi^{Buy} > \Pi^{Build}$. We obtain prediction 1.

Proposition 1. *Firms optimally choose to build when their marginal cost of production is low enough, that is, if their existing workforce is more adapted to operate into the sector of entry.*

As shown in Appendix A.1, our model also implies $\frac{\partial c_I^*}{\partial c_E} < 0$, from which prediction 2 follows.

Proposition 2. *Firms are more likely to buy when key workers for the sector of entry are in short supply in the external labor market.*

Prediction 2 implies that prediction 1 is stronger when key workers for the sector of entry are difficult to hire in the external labor market.

2.3. Micro-foundations of the labor cost

To take predictions 1 and 2 to the data, we need a micro-foundation for the labor costs c_i. In this section, we propose one, based on Cheng and Morrow (2018), in which labor costs depend on the byproduct of the availability and efficiency of workers in each pool.

Worker occupations. We assume there are different types of workers. Each worker type is employed in a given occupation. Each worker occupation, denoted $o \in \mathcal{O} = \{1, ..., O\}$, is available in quantity $(a_{i1}, ..., a_{iO})$. Importantly, workers availability differs across worker pools $i \in \{I, E, A\}$: The number of workers in each occupation is different in acquiring firms' internal labor market, the external labor market, and targets' internal labor market. Worker characteristics differ across occupations: workers' wage $(w_1, ..., w_O)$ and efficiency $(m_1, ..., m_O)$ are different across occupations, but not across workers within an occupation.[18]

Recruiting process. To recruit workers, firms conduct interviews with several workers from each pool $i \in \{I, E, A\}$. The match between a firm and a worker is assumed to be of random quality $h \geq 1$, and to follow a Pareto distribution with cumulative density function $\Psi(h) = 1 - h^{-k}$ with $k > 1$. This distribution is assumed equal across worker pools, that is, firms' expected match quality is the same regardless of whether workers come from inside the firm (i.e., the internal labor market) or from outside (i.e., the external labor market or the acquired firm). This assumption holds as long as workers' new tasks in the sector of entry are sufficiently different from the tasks they were previously assigned to. Indeed, in that case, there is no reason for firms to have different expected match quality with workers from different pools.[19] Firms observe the match quality during the interview.

We assume conducting interviews is labor-intensive and costs f per interview to the firm. This assumption implies the degree of information asymmetry between firms and workers does not vary across the three worker pools.There are two reasons why we assume this cost is the same when firms match with workers from different pools. First, this assumption simplifies the theoretical analysis and does not change its predictions. Second, in the data, we cannot observe interview and hiring costs across firms, nor differentiate those costs across worker pools.

Firms select interviewed workers by determining a match quality threshold $\underline{h}_{io}$ under which they do not recruit the worker they interview. As a result, after all interviews, firms hire a total number $N_i a_{io} \underline{h}_{io}^{-k}$ of workers in each occupation $o \in \mathcal{O}$. The labor cost of conducting interviews is then $f N_i c_i$.

[18] We, therefore, drop the subscript i relative to wages and efficiency.

[19] In the empirical analysis, we control for the product market distance between firms' existing business lines and the new sector of production. We find our model's predictions to continue being borne by the data. Moreover, we find stronger support for our model's predictions when the new sector is distant from firms' existing sectors of activity, that is, when it is most likely that k is the same across worker pools.

Costs and production function. Firms' total labor cost C_i in each labor pool i adds up the cost of conducting interviews fN_ic_i and wages of hired employees. It is given by:

$$C_i = N_i \left(\sum_o a_{io} w_o \underline{h}_{io}^{-k} + f c_i \right). \tag{6}$$

For a given number of interviews N_i and quality threshold $\underline{h}_{io}$, firms know the expected level of human capital of their workforce: Each worker occupation $o \in \mathcal{O}$ supplies an expected level of human capital equal to H_{io}.[20] Within a worker pool, we assume labor supplied by employed workers in a given occupation o aggregates into an input into firms' production function (1) as follows:

$$L_i = \left(H_{i1}^{\theta} + ... + H_{iO}^{\theta} \right)^{1/\theta}, \tag{7}$$

where $\theta < 1$ is the elasticity of substitution between the human capital supplied by workers in different occupations. We assume this elasticity is constant across worker pools $i \in \{I, E, A\}$, that is, the substitutability across worker occupations is the same in the internal labor market, the external labor market, and the acquired firm's labor market. This assumption can be interpreted as the fact that the substitutability across worker occupations is determined by the nature of tasks involved in the production process.

Our model imposes that both the interview cost *and* the match quality distribution are constant across worker pools. As a consequence, workers from the internal, external, and the acquired firm's labor markets, are perfectly substitutable. It follows that our model does not allow for firm-specific organizational or human capital. All the variation in labor costs across worker pools originates from the different vectors of occupation availability $(a_{i1}, ..., a_{iO})$.

Cost minimization. Firms minimize the total labor cost C_i in each worker pool, defined in (6), by choosing how many interviews N_i to conduct and the occupation-specific match quality threshold $\underline{h}_{io}$. Firms take as given the expected human capital supplied by their workers, H_{io}, and the production function (7).

Firms choose the optimal threshold $\underline{h}_{io}$ and the number of interviews N_i by trading off the quality of the hired workers and the search costs associated with recruiting process. One the one hand, hiring a large number N_i of workers enables firms to select the best matches by choosing high values for $\underline{h}_{io}$. On the other hand, firms save on search costs f by choosing a fewer number of workers N_i and low values for $\underline{h}_{io}$.

[20] The match quality follows a Pareto distribution, which implies a simple formula for each worker occupation's expected human capital: $H_{io} \equiv N_i a_{io} m_o \int_{\underline{h}_{io}}^{\infty} h d\Psi(h) = \frac{N a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}$.

We show in Appendix A.3 that this constrained minimization problem results in the following marginal labor cost:

$$c_i = \left(\sum_{o \in O} \left(\frac{a_{io} m_o^k w_o^{1-k}}{f(k-1)} \right)^{\frac{\theta}{\beta}} \right)^{\frac{\beta}{\theta(1-k)}}, \tag{8}$$

where $\beta \equiv \theta + \theta(1-k)$.[21] Equation (8) implies that the marginal labor cost in worker pool $i \in \{I, E, A\}$ increases with the interview cost f. Indeed, higher interview costs reduce the optimal number of interviews. Matches' average quality goes down, and, in turn, the human capital supplied by each worker occupation is lower.

This expression highlights that the labor cost is determined by the byproduct of workers' efficiency and availability. It is large when occupations that are efficient in the new sector (high m_o) are in scarce supply in the worker pool (low a_{io}).

Wage-bill share of each occupation. Denoting A_{io} as the total number of workers in occupation o hired to produce one unit of output,we show in Appendix A.4 that the share of the wage bill that goes to a given worker occupation o can be expressed as

$$\frac{w_o A_{io}}{\sum_o w_o A_{io}} = \frac{\left(a_{io} w_o^{1-k} m_o^k\right)^{\frac{\theta}{\beta}}}{\sum_o \left(a_{io} w_o^{1-k} m_o^k\right)^{\frac{\theta}{\beta}}}. \tag{9}$$

Equation (9) implies that within a worker pool $i \in \{I, E, A\}$, the share of a firm's wage bill that goes to workers in a given occupation depends on (i) occupation-specific wages w_o, (ii) the availability a_{io} of workers in that occupation, and (iii) workers' occupation-specific efficiency m_o. The interpretation is that within a given sector, worker occupations receive relatively higher wages compared to other occupations when they are relatively more efficient at producing in this sector.

3. Empirical strategy

To take predictions 1 and 2 to the data, we need to estimate the marginal labor cost (8) for workers from the internal labor market, i.e., c_I. In Section 4, we explain that neither the availability a_{Io} of workers in each occupation in the internal labor market, nor the occupation-

[21] The assumption that $k > 1$ implies that $\beta > 0$, in which case it is optimal for firms to hire workers in every occupation (Cheng and Morrow, 2018).

specific efficiency m_o, are observable in the data. However, based on our model, we propose a methodology to estimate c_I using administrative worker-level occupation data.

3.1. Occupation-specific human capital

Equation (9) allows us to estimate the average availability a_{Io} of workers in occupation o, and the occupation-specific efficiency m_o. We use the subscripts f, o, n, and t, for firm, worker occupation, sector of the firm, and time, respectively. At the firm-level, we denote $\text{Share}_{f,o,n,t}$ as the share of the wage bill that goes to occupation o in firm f operating in sector n, i.e., the ratio given by Equation (9). Rewriting worker availability $a_{f,o,n,t}$ as a deviation from the sectoral average $\bar{a}_{o,n,t}$, we have $a_{f,o,n,t} = \bar{a}_{o,n,t} \cdot \widetilde{a}_{f,o,n,t}$. Then, taking the logarithm of Equation (9), we obtain

$$\log\left(\text{Share}_{f,o,n,t}\right) = \frac{\theta}{\beta}\log\left(\bar{a}_{o,n,t} m_{o,n,t}^{k} w_{o,n,t}^{1-k}\right) - \log\Big(\sum_{o' \in \mathcal{O}_{gt}} \left(a_{g,o',t} m_{o',n,t}^{k} w_{o',n,t}^{1-k}\right)^{\frac{\theta}{\beta}}\Big) + \frac{\theta}{\beta}\log\left(\widetilde{a}_{f,o,n,t}\right), \tag{10}$$

where $\mathcal{O}_{ft}$ is the set of worker occupations observed in firm f at time t. Equation (10) provides a decomposition of $\text{Share}_{f,o,n,t}$, which can be estimated as the following fixed-effects regression:

$$\log\left(\text{Share}_{f,o,n,t}\right) = \mu_{o,n,t} + \nu_{f,t} + \epsilon_{f,o,n,t}, \tag{11}$$

where $\mu_{o,n,t}$ is an occupation $\times$ sector $\times$ time fixed effect capturing the average wage share that goes to occupation o at the level of sector n. $\nu_{f,t}$ is a firm $\times$ time fixed effect and $\epsilon_{f,o,t}$ is an error term capturing the deviation of occupation o's share in the firm's wage bill to the sectoral average.

The estimation of $\mu_{o,n,t}$ require firms to hire more than one type of occupation each year, and every occupation to be present in more than one firm in a given sector each year.[22] We use the estimated values, denoted $\hat{\mu}_{o,n,t}$, to construct our main explanatory variable below. We interpret them as a score reflecting the (wage- and availability-adjusted) human capital of a given occupation at the sectoral level. The higher the fixed effect of a given occupation in a sector, the larger the score for this occupation in this sector.[23]

[22] To ensure that these two conditions are met, we exclude firms that hire only one type of occupation and employ less than 20 workers from the sample. In addition, in the empirical analysis, we exclude occupation-sector-year triplets with less than 10 firms to obtain more precise estimates.

[23] As an illustration, table A1 displays the five occupations with the largest fixed effects in the sectors of pharmaceutical preparations, IT consultancy activities and manufacture of motor vehicles in 2013. The selected occupations indeed seem to play an important role in their own sectors.

3.2. Firms' internal human capital

We rely on the occupation scores to analyze the role of human capital in the decision to build or buy. We use the subscripts g, o n, and t, for the diversifying firm, worker occupation, sector of entry, and time, respectively. The marginal labor cost (8) can then be rewritten as $c_{I,g,n,t} = (f(k-1))^{\frac{1}{k-1}} (\psi_{I,g,n,t\text{-}1})^{\frac{\beta}{\theta(1-k)}}$, where

$$\psi_{I,g,n,t} = \sum_{o} \left(a_{I,g,o,n,t} w_{g,o,n,t}^{1-k} m_{g,o,n,t}^{k} \right)^{\frac{\theta}{\beta}}. \tag{12}$$

We interpret $\psi_{I,g,n,t}$ as a proxy for the human capital of workers from firms' internal labor market. Indeed, $c_{I,g,n,t}$ is a decreasing function of $\psi_{I,g,n,t}$, that is, the marginal cost of labor of operating in sector n is lower when the human capital of firms' existing workers is high for that sector.

Rewriting $\psi_{I,g,n,t}$ in terms of occupation $\times$ sector $\times$ time fixed effect $\mu_{o,n,t}$ and firm-level deviation from the sectoral average availability of worker occupation, $\widetilde{a}_{g,o,t}$, we have

$$\psi_{I,g,n,t} = \sum_{o} (\widetilde{a}_{g,o,n,t})^{\frac{\theta}{\beta}} \exp(\mu_{o,n,t}). \tag{13}$$

Predictions 1 and 2 consist in predicting firms' choice to build or buy, based on their existing workforce composition. Therefore, we need to empirically estimate firm g's human capital $\psi_{I,g,n,t-1}$ one year before entering the new sector.

The key difficulty in estimating (13) is that we don't observe firms' internal reallocation of existing workers toward the sector of entry, that is, we don't observe $\widetilde{a}_{g,o,n,t}$. Therefore, as soon as occupation o is present in g's existing workforce at $t-1$, we assume $\widetilde{a}_{g,o,n,t} = 1$ so that workers' availability in the internal labor market is equal to the average availability in the new sector. Otherwise, if $a_{g,o,n,t} = 0$, we assume $\widetilde{a}_{g,o,n,t} = 0$.[24] We obtain our main independent variable, which is an empirical estimation of the value of (13) at $t-1$:

$$\text{Internal Human Capital}_{g,n,t-1} = \frac{1}{\#O_{g,t-1}} \cdot \sum_{o \in \mathcal{O}_{g,t-1}} \exp\left(\hat{\mu}_{o,n,t-1}\right), \tag{14}$$

where $O_{g,t-1}$ is the set of worker occupations present in firm g's internal labor market prior to entering the new sector n (i.e., occupations for which $a_{g,o,n,t-1} > 0$), and the values of $\hat{\mu}_{o,n,t-1}$

[24] In the empirical analysis, we test several alternative assumptions as robustness checks (see Section 6.1). In particular, we use the fraction of each worker occupation as weights in the computation of $\text{Internal Human Capital}_{g,n,t-1}$. Our results are robust to alternative assumptions in our estimation of equation (13).

come from the estimation of equation (11).

Internal Human Capital$_{g,n,t-1}$ is a measure of the human capital of firm g's existing workforce for the sector of entry n. It is defined as the average value of the (exponentiated) occupation $\times$ sector $\times$ year fixed effects of the occupations present in the workforce of the firm prior to entry, given by (11). We use the average value instead of the sum to avoid human capital being mechanically larger for firms employing more worker occupations.[25]

Internal Human Capital$_{g,n,t-1}$ takes a high value when occupations with a high (wage-adjusted) efficiency $m_{o,n,t-1}$ for the sector of entry are already present in a firm's internal labor market prior to entry. We interpret it as a measure of whether a firm's workforce is adapted for a given sector of entry. In the rest of the paper, we refer to this measure as "human capital" or "internal human capital" without explicitly mentioning that it is specific to a given sector of entry.

3.3. Empirical model

Prediction 1 states that firms are more likely to "buy", as opposed to "build", when their existing internal human capital is not adapted to the market of entry. This pattern is more pronounced when key workers for the sector of entry are difficult to hire in the external labor market (prediction 2).

We test these predictions by analyzing the link between the type of entry and the constructed measure of internal human capital. The dependent variable is $\mathbb{1}(\text{Buy})_{g,n,o,t}$, a dummy equal to one if firm f enters a new sector n through an acquisition ("buy"), and zero if it enters the new sector by building on its existing resources ("build"). o indicates firm f's main original sector of activity, i.e. the sector in which the firm realizes the largest share of its sales prior to entry. Our baseline empirical model is:

$$\mathbb{1}(\text{Buy})_{g,n,o,t} = \lambda_{n,o,t} + \delta \text{Internal Human Capital}_{g,n,t\text{-}1} + \beta X_{g,n,o,t-1} + \eta_{g,n,o,t} \qquad (15)$$

This presence of origin $\times$ entry $\times$ time fixed effects implies that the role of internal human capital is identified by comparing the diversification strategy of firms operating in the same sector of origin and entering in the same new sector. All unobservable time-varying synergies and complementarities between sectors are therefore captured by the fixed effects $\lambda_{n,o,t}$. In addition, this specification also controls for unobservable factors relative to the sector of entry

[25] We also control for firm size using the number of employed workers. In unreported regressions, we control for the number of occupations to closely follow the definition of $\psi_{I,g,n,t\text{-}1}$. Our results are robust to this change.

(e.g., fixed costs of entry, barriers of entry. See McCardle and Viswanathan (1994)) or the sector of origin (e.g., ability to collateralize assets to access external finance. See Rajan and Zingales (1998)).

The vector $X_{g,n,o,t-1}$ includes other firm's characteristics that may influence the decision to build or buy. At the firm level such determinants are firm's size, cash holding, tangibility, labor productivity and product market relatedness (i.e., the distance to the sector of entry based on sales information). We also control for the variation in labor market tightness across geographical zones and sectors.

By construction, Internal Human capital$_{g,n,t-1}$ introduces a measurement error term which generate a correlation between residuals $\eta_{g,n,o,t}$ at the level of the sector of entry. It might also be more or less precisely estimated depending on the sector of origin. We therefore double-cluster standard errors to control for correlations within the sector of entry and within the sector of origin. Finally, we standardize the Internal Human capital$_{g,n,t-1}$ variable in the rest of the analysis to interpret regression coefficients in standard deviation units.

4. Data and Summary Statistics

4.1. Data sources

4.1.1 Firm data

Our primary source of data consists of French administrative data provided by the French Bureau of Statistics (Insee). We define a firm as a set of subsidiaries controlled by the same final owner. We identify the different set of subsidiaries using ownership links (*Enquête sur les Liaisons financières entre sociétés*, LIFI).[26]

We recover the subsidiaries' main sector of activity, balance-sheet information and income statements from the tax files (*Bénéfices Industriels et Commerciaux* and *Bénéfices Non-commerciaux*). We then consolidate the different variables to obtain firm-level measures. In addition, we exclude firms from the financial, agricultural and public sectors.[27]

To identify entries into new sectors, we rely on the subsidiaries' breakdown of sales by

[26]LIFI 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. The dataset is completed by data coming from Bureau Van Dijk (Diane-Amadeus data set) to cover the universe of French firms.

[27]Firms in the financial and agricultural sectors use different accounting systems, which limits the comparability and relevance of standard variables across sectors. For instance, the notion of "value added" is difficult to measure in the public or the financial sector. Second, the decision to enter those sectors may be driven by other motives than in the rest of the economy.

sectors (*Enquête Annuelle de Production*). This data set records the detailed amount of sales realized by subsidiaries in each sector every year. The survey is exhaustive for subsidiaries with at least 20 employees and randomly include smaller subsidiaries such that at least 85% of sales of a given sector are identified in the survey. According to the French Bureau of Statistics, subsidiaries included in the survey cover 96% of sales in the manufacturing sector. For smaller subsidiaries whose the breakdown of sales is not available, we make the assumption that these subsidiaries sell only in their main sector of activity, which is retrieved from the tax files. We exclude entries occurring during the year 2008 as the construction of the survey was deeply modified that year.[28]

Sectors are defined by the French Standard Industry Classification (SIC) (*Nomenclature des activités Francaises, NAF*) which is equivalent to the US SIC. We define sectors at a 5-digit French SIC code. The dataset includes 732 different sectors.

4.1.2 Merger and acquisition deals

We merge the French administrative datasets with M&A deals retrieved from SDC Platinum and Bureau van Dijk Zephyr.[29] We collect all deals between January 2003 and December 2014 that involve a French acquirer and a French target. We exclude Leverage Buy Out and Private Equity deals from the sample. We focus on deals in which the acquirer owns less than 50% of the target shares before the acquisition date and more than 50.1% after, to identify changes of majority ownership between operating companies.

SDC Platinum and Bureau van Dijk's Zephyr do not provide French firm standardized identifying numbers (SIREN). We proceed in several steps to retrieve the unique firm identifiers. First, we use tickers (available only for publicly traded firms) and the Bureau van Dijk identifiers (available only for Zephyr deals) to recover a fraction of the firm identifiers. Second, we build a Python webcrawler on two websites, which takes as input firm's name and address.[30]: (i) www.bodacc.fr (Bulletin Officiel des Annonces Civiles et Commerciales), which is a governmental website that reports official notifications involving French companies since 2003, (ii) www.societe.com, which is a commercial website that aggregates information about French companies from various sources (mostly from Insee and Bodacc.fr). Both websites are supposed to cover the universe of French firms.

[28]Some small internal entries might be due to changes introduced in the methodology of the survey. Our results are robust to whether we keep year 2008 or not in the analysis.

[29]Note that a change of ownership in the LIFI data set can not in general be used to identify M&As. Indeed, change of ownership in LIFI sometimes occur because of previously undisclosed financial links.

[30]The webcrawler builds on the Python packages Selenium and Beautiful Soup).

After running the web-crawler, we drop companies whose the address, city or zip code is missing, as we cannot identify with certainty the corresponding company identifier among several matches. We keep only observations for which the Jaro-Winkler string distance to the original name is below a certain threshold.[31] We use a conservative value of 0.8. Finally, we validate the matches manually and, when appropriate, select the best matches ourselves. Our final sample contains 7,303 deals from 2003 to 2014. To the best of our knowledge, this is the most complete M&A data set available to date for the French economy.

4.1.3 Worker-level occupation worker data

We use the French matched employer-employee administrative dataset (*Déclarations Annuelles des Données Sociales*, DADS) to construct our measure of internal human capital. Firms are required by law to report every year detailed information of their workers when filing payroll taxes.[32] The employer must in particular report the gross and net wages, number of hours worked and an occupation code for each worker. Note that before 2003, occupations are often missing in the dataset. This is the reason why we start the empirical analysis from this date. Occupations are reported using a 4-digit code. The French nomenclature of occupations (*Nomenclatures des professions et catégories socio-professionnelles des emplois salariés des employeurs privés et publics*, PCS-ESE) include 414 different occupations including, for instance, 28 different types of engineers (e.g., logistic, IT, electric, mechanical).

4.2. Main variables

4.2.1 Type of entry: build or buy

Our main dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable equal to one if the entry of firm f in sector n at time t is made through an acquisition ("buy"), and equal to zero if the entry is made internally ("build"). We consider that a firm enters a new sector if (i) at least one of its subsidiaries starts selling in that sector and (ii) none of the other subsidiaries already operates in the sector. A firm is then said to enter a sector by acquisition if the entity that sells in the new sector is added to the set of subsidiaries through a M&A. By contrast, a firm is said to buy if the subsidiary that starts selling in the new sector was already controlled by the firm, or if it is created at the time of entry.[33]

[31] The Jaró-Winkler 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.

[32] Note that reporting of the occupation code is required for firms that employed at least 20 employees in a given year, and optional for firms below the threshold.

[33] See appendix figure A1 for an illustration of the classification of entries.

4.2.2 Distance between sector of origin and sector of entry

As an alternative to our measure of human capital, we construct a product market distance based on Bloom, Schankerman and Van Reenen (2013). The firm-level variable measures the distance between the firm's sectoral repartition of sales ("sector portfolio") prior to entry to the sector portfolios of firms already operating in the sector of entry. The idea of the metric that if the sector portfolio of the entering firm is unusual compared to firms already present in the sector of entry, then the firm is "distant" to the sector of entry - otherwise it is "close".

For a firm f entering a new sector at time t, we denote by $S_g = (S_g^1, ..., S_g^N)$ the vector of sales at time $t-1$ broken down by sectors ($n = 1, ..., N$). S_g^{-n} is the vector of sales excluding sales in sector of entry n.[34] For a given sector of entry n, we define the distance $d_{g,h}^n$ between the firm and any firm h already operating in the new sector n, as (one minus) the uncentered Pearson correlation between vectors S_g^{-n} and S_h^{-n}:

$$d_{g,h}^n = 1 - \frac{(S_g^{-n} \cdot S_h^{-n})}{\sqrt{(S_g^{-n} \cdot S_g^{-n})}\sqrt{(S_h^{-n} \cdot S_h^{-n})}}$$

Finally, we define the product market distance of firm f to sector n as the weighted average of the distance $d_{g,h}^n$ for all firms h operating in sector n at $t-1$:

$$\textit{Product Market Distance}_{g,n} = \frac{\sum_h \omega_h^n d_{g,h}^n}{\sum_h \omega_h^n}$$

where the weights ω_h^n are given by the share of sales realized by firm h in sector n at $t-1$: $\omega_h^n = \frac{\text{Sales of firm } h \text{ in sector } n}{\text{Sales of firm } h}$. The weights ensure that the distance between the diversifying firm and firm h matters more if firm h realizes a large part of its sales in sector n.

4.2.3 Local labor market tightness

We measure local labor market tightness using data from the French national unemployment agency (Pôle emploi). The unemployment agency lists job vacancies, helps unemployed people to find jobs and produces national and local unemployment statistics which we use in this paper. In particular, the unemployment agency tracks occupations in short supply in 350 different local labor markets market since 2010. Occupations are flagged as in short supply when (i) job offers exceed job applications (ii) surveyed employers anticipate they will not fill the posted offers.

[34] We exclude sales in sector n as a firm that enters sector n at time t necessarily reports zero sales in n at time $t-1$.

Figure 3 plots the number of occupations in short supply by local labor markets (LLM) in 2013.[35] Darker shades of blue indicate a higher degree of tension in the local labor market. Interestingly, we see that labor market tightness is not systematically related to geographical density as tight local labor markets can be observed both in urban and rural areas.

We define the *Local Labor Tightness*$_{n,z,t-1}$ variable as the average of the (exponentiated) occupation $\times$ sector $\times$ year fixed effects of the occupations in short supply in a local labor market z:

$$\textit{Local Labor Tightness}_{n,z,t-1} = \frac{1}{N.L_{n,t-1}} \sum_{l \in L} \mathbb{1}(l \text{ in short supply in LLM } z) \times \exp(\hat{\mu}_{o,n,t})$$

where $N.L_{n,t-1}$ is the number of occupations in sector n at time $t-1$.[36] This sector-specific measure of LLM tightness takes high values if the occupations in short supply in local labor market z at time $t-1$ are key for firms operating in sector n. Figures 4a and 4b plot the geographical distribution of the variable for the manufacturing of pharmaceutical preparations and of motor vehicles. The graphs show that firms face on average more difficulties to find key workers in the second sector than in the first. Moreover, the measure appears to significantly vary both across LLMs and sectors.

4.3. Summary statistics

Build or buy? Panel A of Table 1 and Figure 1 present the evolution of the proportion of build and buy entries between 2004 and 2013 (excluding 2008). While at the beginning of the period about 1.8% of entries are made by acquisition, this figure increases over the sample period to reach 2.49% of total entries in 2014. However, buy entries are on average bigger than build entries. When weighted by entry sales, buy entries represent about 6-8% of total entries between 2004 and 2013.

We then look at how these figures vary with the definition of sector in Panel B of Table 1. The proportion of buy entries remains stable at 1.7-1.9% whether we define a sector as a 1-digit or 5-digits code of the French SIC (our baseline). This finding suggests that firms do not tend to enter more by acquisition in sectors that are distant from their sector of origin. However, we find that build entries tend to be larger in sectors that are close to the sector of origin. Indeed,

[35] Appendix table A3 lists occupations that are in short supply in 2013.

[36] In a robustness check, we directly use the percentage of jobs in tension as a proxy for local labor market tightness:

$$\textit{Local Labor Tightness}_{n,z,t-1} = \frac{1}{N.L_{n,t-1}} \sum_{l \in L} \mathbb{1}(l \text{ in short supply in LLM } z).$$

Our results are robust to this alternative specification.

entries by acquisition represent 15% of entry sales when we use the 1-digit classification level, but only 8% when we rely on the most detailed definition of sector (5-digits code).

[Insert Table 1 here]

Internal human capital. Figure 2 plots the probability density functions of firms' human capital by type of entry. The variable Internal Human Capital$_{g,n,t-1}$ is normalized to have a mean of zero and a standard deviation of one. On average, the human capital of firms that enter a new sector by acquisition is less adapted to the sector of entry than that of firms that choose to build.

[Insert Figure 2 here]

Other variables. Table 2 Panel A reports summary statistics on the different control variables we include in our baseline specification (see section 3.3). Consistently with table 1, the average number in the panel of firms diversifying through an acquisition is about 2%. Firms report on average €2.82 million of sales in the sector of entry the first year, with a very large dispersion around the mean. The same year, we find that they invest on average €960,000. Firms enter industries that are vertically integrated to their original industry in 67% of the cases and usually in sectors that are close from their sector of origin. They employ about 63,000 workers, produce about €5,000 of value added per worker, own €4,000 of fixed assets per worker, and hold about €2,000 in cash per worker.

Panel B of Table 2 compares characteristics of firms that enter a new sector by acquisition and those which build on their preexisting internal human capital. The results show that firms that build have significantly smaller sales in the sector of entry than firms that buy, with €9.39m less in sales on average the year of entry. They are also 13% less likely to stay in the sector of entry at a one-year horizon. They invest €9.95 millions less the year of entry and employ about 1,770 less workers on average. In addition, firms that diversify by acquisition are more profitable, more capital intensive and have higher internal funds. These significant differences on observable characteristics between firms that build and firms that buy emphasizes the importance of including control variables in the empirical analysis.

[Insert Table 2 here]

5. Human capital and corporate diversification

5.1. Main results

Table 3 presents our baseline results. The dependent variable $1(Buy)_{g,n,t}$ is a dummy variable equal to one if firm f enters a new sector n at time t through the acquisition of an existing firm, and zero if it enters by building on its own resources. All our regressions include interacted sector of origin $\times$ sector of entry $\times$ year fixed effects. This specification neutralizes potential unobservable time-varying synergies between the sector of entry and that of origin. The idea is to compare firms that operate in the same sector of origin and enter the same new sector the same year, to isolate the effect of human capital on firms' decision to enter through an acquisition or by building on their own resources.

We test our model's prediction that firms are more likely to build when their human capital is more adapted to operate in the sector of entry (prediction 1). Our main independent variable *Internal Human Capital* measures to which extent the firm already employs key occupations for the sector of entry prior to diversifiying. All specifications control for firms' size (log number of workers), as well as firms' total cash holdings, tangible assets, and value added that we scale by the number of workers.

Consistent with our model's prediction and the suggestive evidence in Figure 2, we find that the internal human capital of the firm prior entry is negatively correlated with the probability to enter by acquisition. A one standard deviation increase in internal human capital is associated to a 1 percentage points drop in the likelihood to enter by acquisition (columns 1 and 2). This relationship is sizeable, equal to 50% of the unconditional probability to buy, and significant at the 1% level. The point estimate is unchanged when we add control variables in column 2, suggesting that control variables are uncorrelated with our key dependent variable *Internal Human Capital*. We conclude that firms whose human capital is adapted to the sector of entry are more likely to enter by building on their own resources.

The estimates in Table 3 also show that the likelihood of buying relative to building increases with firm size. By contrast, cash holdings, tangibility and profitability do not appear to be significantly associated with the mode of entry in a new sector.

An alternative explanation could be that firms anticipate the mode of entry by adjusting the composition of their internal human capital several years before diversifying. If this was the case, human capital would be endogenous to the mode of entry and our interpretation would be biased. To overcome this problem, we test whether lagged values of internal human capital

also predict the mode of entry. We find that firms' decision to build rather than buy is still negatively correlated with the internal capital two and three years before entry (columns 3 and 4, respectively). The point estimates on each lagged measures of human capital remain unchanged, suggesting that firms do not significantly modify their workforce composition during the years preceding the entry. According to our model, one rationale for this result could be that hiring workers in key occupations is too costly because of tight local labor markets. We discuss the role of labor market tightness in section 8.1.

[Insert Table 3 here]

5.2. The role of size and financial constraints

Firms that buy tend to be larger (Table 2). One concern could be that only large may be able to afford the fixed costs associated with an acquisition. Under this hypothesis, small firms would always end up entering by building and human capital considerations would be irrelevant. We should expect in that case no significant relationship between human capital and the type of entry for smaller firms. In columns 1 and 2 of Table 4, we interact our human capital measure with firm size. We use terciles of firm size (proxied by number of workers) to allow for non-linear effects of firm size on the decision to build or buy. We find the interaction terms to be not statistically different from zero, suggesting that human capital plays a role in the decision to build or buy both for small and large firms.

One reason why larger firms tend to grow more by acquisition could be that they are less financially constrained than smaller firms, and hence can afford the acquisition costs. In columns 3 and 4, we test whether our measure of internal human capital interacts with internal financial resources as proxied for by cash holdings. The results show that the relationship between internal human capital and the probability to build or buy is not different across firms with different levels of cash earnings.

Finally, it might be that public firms have different growth strategies than private firms because they can raise new equity to fund acquisitions more easily. We test whether the relationship between internal human capital and the choice to build or buy holds differently for firms which are privately or publicly owned. We re-estimate our baseline specification with controls separately on public and private firms in columns 5 and 6 of Table 4. We find that the negative relationship holds both for public and private firms. Note that given the small number of public firms in France, we only perform those regressions with sector of origin $\times$ sector of entry interacted fixed effects, and separate year fixed effects. To summarize, Table 4 confirms

that our central findings hold across different types of firms, suggesting they may be generalized outside our sample.

[Insert Table 4 here]

5.3. Complementarities between the sector of origin and the sector of entry

Firms do not diversify in random sectors but take into account complementarities and synergies between the sector of origin and that of entry. The baseline specification includes sector of origin × sector of entry × year fixed effects (equation 15) to remove the influence of these unobservable factors on our baseline results. In this section, we explore how sectoral complementarities interact with human capital in the decision to build or buy so as to better understand when labor costs considerations matter in diversification decisions.

In the columns 1 and 2 of table 5, we investigate how human capital interacts with the product market distance between the sector of origin and the sector of entry.[37] We use to that end our firm-level distance to the sector of entry (see section 4.2.2 for more details). The variable is based on the correlation between the product market portfolio of a firm prior to entry and product market portfolios of firms already operating in the sector. A distance close to zero, for instance, means that the entering firm has very similar activities to incumbent firms.

We include terciles of our measure of distance in the baseline specification in column 1, and interact it with the measure of internal human capital in column 2. We find that firms tend to buy more frequently in the third tercile of product market distance. The point estimate for human capital is very close to that in Table 3, which suggests that role of the composition of the workforce is not explained by sectoral similarities. Interestingly, in column 2 we find the human capital coefficient to be significantly stronger for firms in the 2^{nd} and 3^{rd} terciles of distance. This suggests that high product market distance magnifies the importance of labor costs in the decision to build or buy.

In columns 3 and 4 of Table 5, we test the role of vertical integration in the decision to build or buy. Labor costs considerations may indeed be less relevant in the presence of important vertical links. Firms might be willing to enter upstream sectors primarily to acquire suppliers so as to facilitate transfers of goods along the production chain. To account for the firm's position in the production chain, we measure vertical integration following Fan and Goyal (2006). Using the 1995 input-output (IO) table for France compiled by the Organization for Economic Cooperation

[37] Hoberg and Phillips (2010) shows that product market distance plays an important role in determining M&A patterns. Firms that have similar products tend to merge more between each other, and this create more value. the authors develop a measure using textual analysis on 10k filings available for publicly listed firms in the US.

and Development (OECD), we identify a vertical link between firms and their sector of entry whenever more than 5% of inputs used by the sector of origin comes from the sector of entry (we also use 1%, 10% and 20% as thresholds).[38] Since the variable is defined at the level of a sector of origin and sector of entry, we replace the interacted fixed effects by separate origin, entry and year fixed effects. We do not find any systematic relationship between the presence of vertical links and the decision to build or buy. In particular, firms that start operations in an upstream sector do not seem to enter more often by acquisition.

[Insert Table 5 here]

6. Robustness checks and alternative mechanisms

6.1. Alternative measures of human capital

In this section, we check the robustness of our results to alternative definitions of the measure of internal human capital. To do so, we replicate our main specification from Table 3 using different versions of *Internal Human Capital* and test whether firms that do not employ the key occupations for the sector of entry are more likely to enter by acquisition. Table 6 reports the results.

In column 1, the main independent variable is a dummy variable taking the value one if the firm does not employ any worker in the top 10 most important occupations for the sector of entry. As explained in section 3, the ranking of occupations across sectors is based on the estimated values of the occupation $\times$ sector $\times$ fixed effects. We find that firms that do not hire any worker in the top 10 occupations for the sector of entry are 40% more likely to buy.

Our baseline measure of human capital is the simple average of the (exponentiated) occupation $\times$ sector $\times$ year fixed effects of occupations that are present in the firm's workforce one year before entering the new sector (see section 3). In column 2 of Table 6, we take the weighted average of the fixed effects with weights equal to the share of workers in a given occupation in the workforce of the firm. This alternative measure of internal human capital assumes that the firm allocates workers to the new sector in proportion to the existing occupational structure (see section 3.2).

In column 3, we exclude CEOs from the set of occupations used to build the measure of

[38]Note that the input-output matrix from France we use is rather coarse (35 industries by 35 industries). The regression excludes observations where the industry of origin and the industry of entry are the same as the vertical link variable is not defined in that case.

internal human capital. One concern with CEOs' wages is that they may be determined by other factors than their contribution to the firms' output (moral hazard, information asymmetries). Finally, in column 4, we change the unit of observation used in the estimation of the occupation $\times$ sector $\times$ year fixed effects. When we estimate fixed effects at the firm level, we make the assumption that every worker is involved in the production process of the firm's main activity. In contrast, plants are less diversified entities, such that estimating the fixed effects at the plant level should yield better estimates, as plants are presumably less diversified entity than firms.[39] Our main result is not affected by any of these alternative measures of internal human capital: The negative relationship between human capital and the probability to buy remains economically and statistically significant. Moreover, the point estimates are very similar across specifications.

[Insert Table 6 here]

6.2. Does selection into diversification drive our results?

A potential limit of our analysis is that we focus on the build versus buy trade-off while abstracting from the decision to diversify itself. If the entry and the type of entry in the new sector are jointly driven by unobservable factors, our OLS estimations might lead us to draw inaccurate conclusions.

Our approach to address this issue is to identify plausible self-selection narratives that could invalidate our results and directly assess their influence in our data.[40] We focus on two scenarii in which self-selection may pose a problem for identification. First, firms with under-performing business segments may be willing to shift their activities to sectors with better growth prospects. In order to minimize firing costs and preserve organizational capital, firms are likely to choose sectors in which they can easily redeploy their existing workforce (Tate and Yang, 2016*b*). In this scenario, the likelihood of an entry by acquisition would also be negatively related to internal human capital. However, if we could perfectly control for selection, there would be no role left for human capital as the decision to build or buy would be governed by the decision to enter the new sector.

We test the influence of sectoral reallocation in columns 1 to 3. We denote firms that enter internally a new sector to shift their activities as "shifting firms". We compute for each

[39] A firm is defined as a set of subsidiaries, which are themselves composed of one or several plants.

[40] Traditional methods designed to address selection issues would involve finding an instrument that affects the choice to diversify *in a given sector* but not the type of entry. This a difficult task as most of the variables that vary both at the level of the firm and the sector of entry (*e.g.*, compatibility or exposure to the sector of entry) are unlikely to be orthogonal to the decision to build or buy.

firm the growth rate of sales between $t-1$ and t in each sector in which they were operating at $t-1$ and take the firm-level minimum of sectoral growth rates. A shifting firm is a firm (i) that enters a new sector internally and (ii) for which the minimum growth rate is negative and equal to 100% in absolute value (*i.e.*, the firm exits a sector simultaneously to entering internally a new one). We also use 50% and 25% as more conservative thresholds. Under the tested hypothesis, the negative coefficient for human capital would be driven by the sub-sample of shifting firms. Excluding them should result in non-significant coefficient for human capital. By contrast, the results show that the point estimates for our main variable are unchanged even for fairly conservative definitions of sectoral reallocation.

Consider then the other polar case. Firms with low growth prospects may be willing to acquire firms to expand their set of investment opportunities (Levine, 2017). This mechanism could lead firms to enter unrelated sectors by acquisition irrespective of human capital considerations. Since firms are unlikely to employ the right set of workers to operate in unrelated sectors, this could translate into an observed negative relationship between internal human capital and the propensity to buy. However, if we could control for selection, we would once again find no role for human capital as the objective of the entry is actually to acquire a new firm. To address this particular issue, we focus on firms that enter multiple sectors by acquisition ("serial acquirers") as they are more likely to be conducting such a strategy of external growth.[41] We still find a significant, negative coefficient for human capital. Combined together, our findings suggest a limited role for selection issues.

[Insert Table 7 here]

6.3. Complementarities between human and physical capital

It is likely that firms that do not employ the right set of occupations to operate in the sector of entry do not possess the adequate machines either. It follows that the negative relationship between human capital and the propensity to buy could potentially reflect the role of adjustment costs for capital, and not labor.

We run two sets of tests to address this issue. First, we compare the build or buy decision of firms that realized similar entry sales in the year of entry. Given that firms should enter a new sector only if they anticipate high enough entry sales to offset entry costs, we use entry sales to proxy for entry costs. In columns 1 and 2 of Table 8, we rank entry sales into ten deciles and run our baseline regressions with interacted sector of origin $\times$ sector of entry $\times$ year $\times$ sales deciles

[41] Among firms that enter externally a sector at least once, 50% are classified as serial acquirers.

fixed effects. This specification allows us to compare firms that operate in the same sector of origin, entering the same sector, the same year, and also making similar sales when entering the new sector. Although this specification creates a lot of singletons that are dropped from the sample (the number of observations drops from 75k to 45k), our point estimates remain nearly unchanged.

Second, and similarly, we proxy for physical capital adjustments using the volume of capital expenditures made by firms when entering the new sector. To do so, we compare firms investing similar amounts the year of entry. For firms that build, we measure investment using capital expenditures the year of entry. For firms that buy, we measure investment as the amount of fixed assets in the target. We run our baseline regression with interacted sector of origin $\times$ sector of entry $\times$ year $\times$ investment deciles fixed effects. The idea is to isolate the effect of human capital on the build or buy decision, irrespective of differences in capital expenditures. In columns 3 and 4 of Table 8, we find that the economic magnitude of the role of human capital is approximately lowered by half. It remains negative and statistically significant at the 5% level, with or without the inclusion of control variables. Based on these two tests, we conclude that firms seem to take into account internal human capital when making the decision to build or buy, irrespective of the associated costs or physical capital adjustments.

[Insert Table 8 here]

7. Evidence of labor adjustment costs

7.1. Within-firm human capital and diversification choice

Previous sections established that firms that do not employ the right set of workers tend to enter a new sector by acquisition. This finding, we argue, suggests that firms prefer to pay the costs associated to acquiring and restructuring a target when the costs of adjusting the existing workforce are too large. In this section, we focus on firms that diversify by building on their existing resources to highlight the existence of such adjustment costs.

We define a firm as a set of subsidiaries. Therefore, the composition of the workforce may vary within a firm, with some subsidiaries better prepared to enter the new sector than others. Within firms that build, we should expect the entry to be made through subsidiaries that already employ the right set of occupations in order to minimize reallocation costs in the internal labor market. We construct our measure of internal human capital for each subsidiary and estimate

the following model:

$$\mathbb{1}(\text{Build})_{f,n,t} = \lambda_{g,n,t} + \beta \cdot \text{Internal Human Capital}_{f,n,t\text{-}1} + \gamma \cdot \text{X}_{f,n,t\text{-}1} + \epsilon_{f,n,t}$$

The dependent variable is a dummy variable that takes the value one if the entry into the new sector is made through subsidiary f, and zero otherwise. We include firm $\times$ sector of entry $\times$ year fixed effects to compare the internal capital of the different subsidiaries of the same firm g. This specification leads mechanically to exclude stand-alone firms as we focus on the heterogeneity of human capital across subsidiaries.

In column 1 of Table 9, we find that within a firm, entry is more likely to be achieved through a subsidiary with the right set of occupations for the new sector. In column 2, we add control variables at the subsidiary level. The estimates show that large, productive and cash-rich subsidiaries are more likely to diversify in a new sector. In addition, in columns 3 and 4, we show that lag values of internal human capital are also positively correlated with the entry dummy. As for firms, internal human capital at the subsidiary level is sticky over time, suggesting that the composition the workforce is not adjusted in anticipation of a diversification. Overall, these findings are consistent with the presence of reallocation costs in the internal labor market.

[Insert Table 9 here]

7.2. Internal human capital and workforce adjustment

Our measure of human capital captures to which extent to the workforce of the firm is adapted to the sector of entry. Therefore, among firms that enter a new sector by developing their own resources, we should find that firms with lower human capital hire relatively more workers so as to adjust their workforce. We test this hypothesis by looking at employment growth within subsidiaries. More precisely, we focus on the subsidiaries that within diversifying firms, start selling in the new sector. This choice allows to observe more precisely the adjustment in labor associated with entry as it allows to abstract from employment variations in the other subsidiaries.

In Table 10, we show the existence of a negative relationship between internal human capital at $t-1$ and the growth rate of employment between $t-1$ and $t+1$ (with t the year of the entry into the new sector). This finding confirms that the higher the internal human capital, the less the subsidiary needs to hire additional workers to operate in the new sector. In column 2 and 3, we look at the timing of new hirings. We find that employment growth is not related to the

level of internal human capital prior to entry (column 2). Instead, subsidiaries tend to adjust their workforce after having entered the new sector (column 3).

[Insert Table 10 here]

8. The role of labor market frictions

8.1. The effects of local labor market tightness for key occupations

Given the costs associated with an acquisition, why don't firms hire new workers instead of buying an existing firm? The model predicts that firms are more likely to buy rather than to adjust their workforce when workers in key occupations are in short supply in the external labor market (prediction 2). Specifically, the negative relationship between internal human capital and the propensity to buy should be stronger when it is costly for the firm to hire workers on the external labor market. In this section, we empirically assess the role of labor market tightness in the decision to build or buy.

We rely to that end on data from the French unemployment agency to construct a sector-specific measure of local labor market (LLM) tightness.[42] According to this metric, a local labor market is viewed as tight for a given sector if the occupations in short supply in the LLM are important for operating in the sector (see section 4.2.3).

The results are reported in Table 11. Columns 1 and 2 show that the point estimates on the second and third terciles of LLM tightness are positive and significant, meaning that firms are significantly more likely to buy when labor markets for key occupations are tighter. Moreover, the interaction terms shows that the link between internal capital and the type of entry is stronger in presence of greater market frictions. This holds both with and without control variables (columns 1 and 2). This result is confirmed when we split the sample by tercile of labor market tightness and re-estimate the regression separately on each sub-sample: the human capital coefficient increases in absolute value and is significant at 1% only for the last tercile of labor market tightness. Overall, our results strongly confirm the prediction that human capital should play a greater role when it is more difficult to hire workers in key occupations.

[Insert Table 11 here]

[42] France is divided into 347 LLMs in 2013. Since 2010, the unemployment agency lists occupations in short supply in each LLM. An occupation is in short supply in a LLM if (i) the ratio of job offers exceeds job applications and (ii) surveyed employers anticipate that job offers for this occupation will be hard to fill in.

8.2. Labor market tightness and the value of building

Our model is based on the comparison of profits made when entering by acquisition and by building on existing resources. The model predicts that the value of building should be positively related to internal human capital. Moreover, this relationship should be stronger when key occupations are in short supply in the external labor market. We have so far tested this assumption indirectly by looking at the choice to build or buy. In this last section, we directly look at the relationship between the value of building and the human capital of the firm.

The results are reported in table 12. Profits by sector are not observable in our data set. Instead, we proxy the value of building by the logarithm of entry sales. We find that entry sales are higher when the workforce of the firm is more adapted to the sector of entry. This finding suggests that the value of building is higher when the firm has the adequate internal resources. In columns 3 to 5, we run the analysis on subsample of LLM tightness terciles and show the positive link is entirely driven by tight labor markets (third tercile, column 5). This finding suggests that the value of building depends on firms' internal human capital when local labor market frictions are important. In other words, firms' internal resources matter only when it is costly to obtain them outside from the boundaries of the firm.

[Insert Table 12 here]

9. Conclusion

Why do some firms enter a new sector by acquiring an existing company ("buy"), while others do so using their existing resources ("build")? When a firm buys to enter a new sector, it has to incur both the costs of acquiring and restructuring the target, but gets access to its productive resources. When a firm builds on its existing resources to enter a new sector, it must pay the adjustment costs needed to acquire the adequate set of capabilities.

We focus on the role of labor and construct a firm-level measure of human capital based on the occupational structure of the workforce. Our main explanatory variable measures to which extent the human capital of the firm is adapted to the sector of entry. We find robust evidence that firms choose to build (buy) when their human capital is (not) adapted to the sector of entry. In addition, using occupation-level data on local labor market tightness, we show that firms are more likely to buy when it is more costly to hire new workers, *i.e.* when labor markets for key occupations are in short supply. This paper contributes to the literature

on corporate diversification by showing that both the set of internal resources and the cost of accessing external resources play a role in explaining how firms diversify, which contributes to a better understanding of the determinants of firms' boundaries.

References

Abowd, John M, and Francis Kramarz. 2003. "The costs of hiring and separations." *Labour Economics*, 10(5): 499–530.

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

Agrawal, Ashwini K, and David A Matsa. 2013. "Labor unemployment risk and corporate financing decisions." *Journal of Financial Economics*, 108(2): 449–470.

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

Baghai, Ramin, Rui Silva, Viktor Thell, and Vikrant Vig. 2018. "Talent in distressed firms: Investigating the labor costs of financial distress." *Working paper.*

Bai, John Jianqiu, Douglas J Fairhurst, and Matthew Serfling. 2018. "Employment protection, investment, and firm growth." Working paper.

Becker, Gary S. 1962. "Investment in human capital: A theoretical analysis." *Journal of political economy*, 70(5, Part 2): 9–49.

Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch. 2014. "Labor hiring, investment, and stock return predictability in the cross section." *Journal of Political Economy*, 122(1): 129–177.

Bena, Jan, and Kai Li. 2014. "Corporate innovations and mergers and acquisitions." *Journal of Finance*, 69(5): 1923–1960.

Bernard, Andrew B, Emily J Blanchard, Ilke Van Beveren, and Hylke Vandenbussche. 2018. "Carry-along trade." *The Review of Economic Studies*, 86(2): 526–563.

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott. 2007. "Firms in international trade." *Journal of Economic perspectives*, 21(3): 105–130.

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

Blatter, Marc, Samuel Muehlemann, and Samuel Schenker. 2012. "The costs of hiring skilled workers." *European Economic Review*, 56(1): 20–35.

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

Boehm, Johannes, Swati Dhingra, and John Morrow. 2019. "The comparative advantage of firms." *Working paper.*

Capron, Laurence, and Will Mitchell. 2012. *Build, borrow, or buy: Solving the growth dilemma.* Harvard Business Press.

Cestone, Giacinta, Chiara Fumagalli, Francis Kramarz, and Giovanni Pica. 2016. "Insurance between firms: the role of internal labor markets." working paper.

Cheng, Wenya, and John Morrow. 2018. "Firm productivity differences from factor markets." *Journal of Industrial Economics*, 66(1): 126–171.

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

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

Eckbo, Espen. 1983. "Horizontal mergers, collusion, and stockholder wealth." *Journal of financial Economics*, 11(1-4): 241–273.

Eiling, Esther. 2013. "Industry-specific human capital, idiosyncratic risk, and the cross-section of expected stock returns." *The Journal of Finance*, 68(1): 43–84.

Fan, Joseph P. H., and Vidhan Goyal. 2006. "On the patterns and wealth effects of vertical mergers." *Journal of Business*, 79(2): 877–902.

Farrell, Joseph, and Carl Shapiro. 1990. "Horizontal mergers: an equilibrium analysis." *American Economic Review*, 107–126.

Ghaly, Mohamed, Viet Anh Dang, and Konstantinos Stathopoulos. 2017. "Cash holdings and labor heterogeneity: the role of skilled labor." *Review of Financial Studies*, 30(10): 3636–3668.

Gibbons, Robert, and Michael Waldman. 2004. "Task-specific human capital." *American Economic Review*, 94(2): 203–207.

Golubov, Andrey, Alfred Yawson, and Huizhong Zhang. 2015. "Extraordinary acquirers." *Journal of Financial Economics*, 116(2): 314–330.

Grossman, Sanford J, and Oliver D Hart. 1986. "The costs and benefits of ownership: a theory of vertical and lateral integration." *Journal of Political economy*, 94(4): 691–719.

Hart, Oliver, and John Moore. 1990. "Property rights and the nature of the firm." *Journal of Political Economy*, 98(6): 1119–1158.

Hoberg, Gerard, and Gordon Phillips. 2010. "Product market synergies and competition in mergers and acquisitions: A text-based analysis." *Review of Financial Studies*, 23(10): 3773–3811.

Kuehn, Lars-Alexander, Mikhail Simutin, and Jessie Jiaxu Wang. 2017. "A labor capital asset pricing model." *Journal of Finance*, 72(5): 2131–2178.

Lagaras, Spyridon. 2017. "Corporate takeovers and labor restructuring."

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

Lee, Kyeong Hun, David C. Mauer, and Emma Qianying Xu. 2018. "Human capital relatedness and mergers and acquisitions." *Journal of Financial Economics*, 129(1): 111–135.

Levine, Oliver. 2017. "Acquiring growth." *Journal of Financial Economics*, 126(2): 300–319.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan. 2018. "Measuring corporate culture using machine learning." Working paper.

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

Matsa, David A. 2010. "Capital structure as a strategic variable: evidence from collective bargaining." *Journal of Finance*, 65(3): 1197–1232.

Ma, Wenting, Paige Ouimet, and Elena Simintzi. 2018. "Technological change and inequality."

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

Merz, Monika, and Eran Yashiv. 2007. "Labor and the market value of the firm." *American Economic Review*, 97(4): 1419–1431.

Moretti, Enrico. 2010. "Local labor markets." *National Bureau of Economic Research*.

Nocke, Volker, and Stephen Yeaple. 2007. "Cross-border mergers and acquisitions vs. greenfield foreign direct investment: The role of firm heterogeneity." *Journal of International Economics*, 72(2): 336–365.

Ouimet, Paige Parker, and Rebecca Zarutskie. 2016. "Acquiring labor." working paper.

Penrose, Edith. 1955. "Limits to the growth and size of firms." *American Economic Review*, 45(2): 531–543.

Phillips, Gordon M, and Alexei Zhdanov. 2013. "R&D and the Incentives from Merger and Acquisition Activity." *The Review of Financial Studies*, 26(1): 34–78.

Rajan, Raghuram, and Luigi Zingales. 1998. "Financial Dependence and Growth." *American Economic Review*, 88(3): 559–86.

Rhodes-Kropf, Matthew, and David T. Robinson. 2008. "The market for mergers and the boundaries of the firm." *Journal of Finance*, 63(3): 1169–1211.

Romanski, Hilton. 2017. "Cisco's 200th Acquisition - A Tradition of Advancement, Disruption and Growth." *Cisco Blog*.

Serfling, Matthew. 2016. "Firing costs and capital structure decisions." *Journal of Finance*, 71(5): 2239–2286.

Simintzi, Elena, Vikrant Vig, and Paolo Volpin. 2014. "Labor protection and leverage." *Review of Financial Studies*, 28(2): 561–591.

Tate, Geoffrey A, and Liu Yang. 2016*a*. "The human factor in acquisitions: cross-industry labor mobility and corporate diversification." *Working paper*.

Tate, Geoffrey, and Liu Yang. 2016*b*. "The bright side of corporate diversification: evidence from internal labor markets." *Review of Financial Studies*, 28(8): 2203–2249.

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

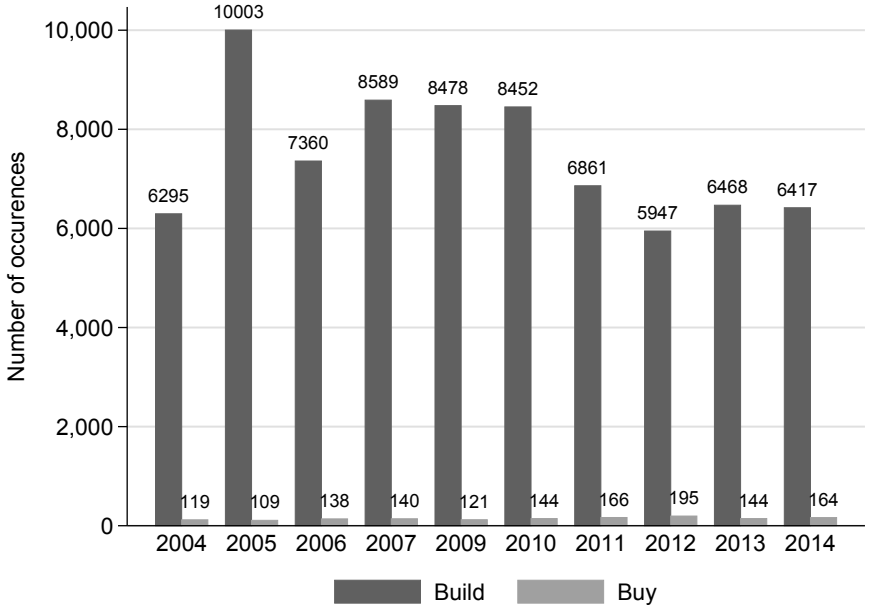
Xu, Sheng-Jun. 2018. "Skilled labor supply and corporate investment: Evidence from the H-1B visa program." *Working paper*.

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

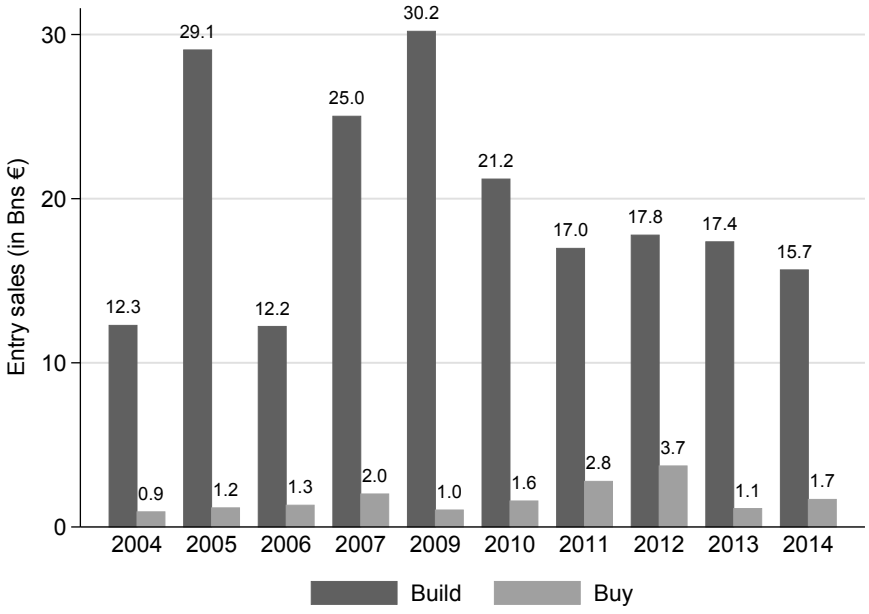
Figures and Tables

Figure 1. Number and Size of Build and Buy Entries

Source: SDC Platinum, BvD Zephyr, ESA survey. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This figure displays the number of external (Buy) and internal (Build) entries by year and the total sales by type of type. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Entries are identified with sales reported at the 5-digit level of the French SIC.



(a) Number of entries



(b) Total sales by type of entry

Figure 2. Human Capital by Type of Entry

The figure displays the probability distribution function of *Internal Human capital* by mode of entry. *Source:* SDC Platinum, BvD Zephyr, EAE survey, matched employer-employee dataset. *Sample:* Firms that enter a new sector either internally (build) or externally (buy) during the periods 2003-2007 and 2009-2014. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Build entries are identified using reported sales from the ESA survey at the 5-digits level of the French SIC. *Internal Human Capital*$_{g,n,t-1}$ measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). *Internal Human Capital*$_{g,n,t-1}$ is trimmed at the 5% level (for this graph only).

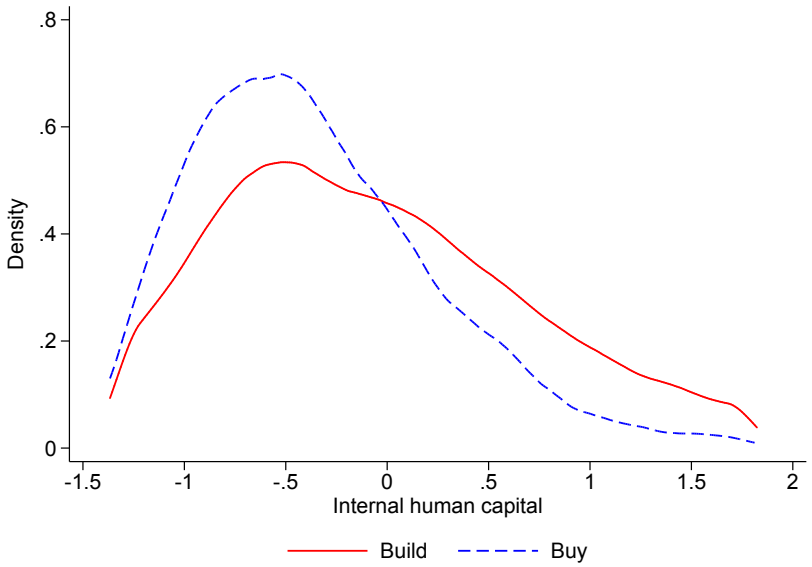
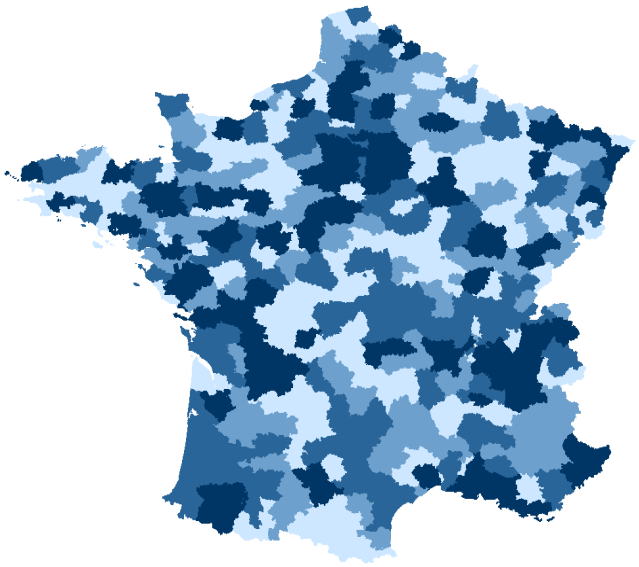


Figure 3. Occupations in Short Supply

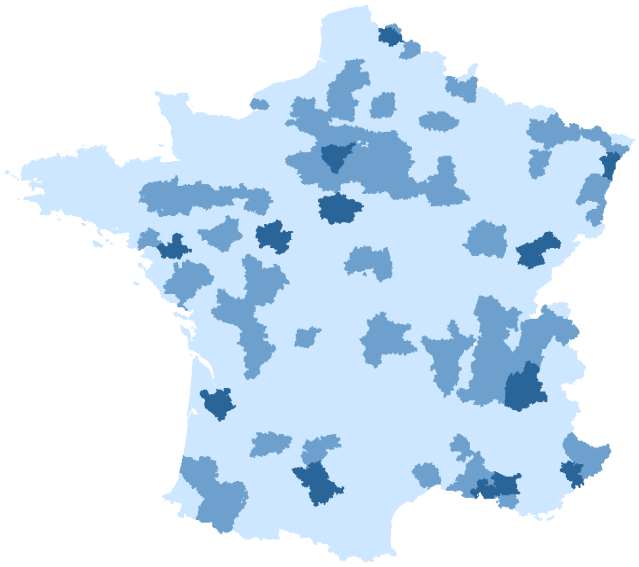
Source: French national unemployment agency. Figure 3 plots the distribution of the tightness of labor markets in 2013. Labor market tightness is measured by the number of occupations in short supply in a given local labor market. Darker shades of blue indicate a higher degree of tension in the local labor market. Occupations in short supply are identified by the French national employment agency as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers. There are 348 different local labor markets.



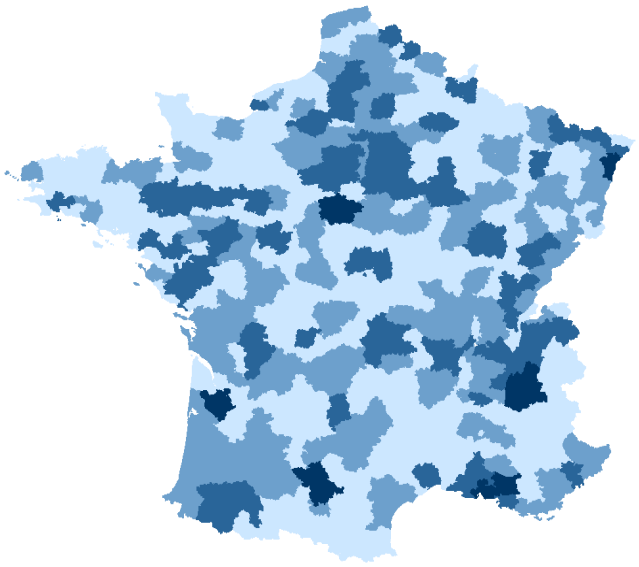
(a) Number of occupations in short supply by local labor market.

Figure 4. Local Labor Markets Tightness

Source: French national unemployment agency. The figures plot the value of $LLM\ Tightness_{n,z,t}$ of each local labor market z in $t = 2013$ for different sectors n. Figure 4a focuses on the manufacture of pharmaceutical preparations and figure 4b focuses on the manufacture of motor vehicles. Darker shades of blue indicate local labor markets with a larger number of occupations in short supply. Occupations in short supply are identified by the French national employment agency as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers. France is divided into 348 different local labor markets z.



(a) Manufacture of pharmaceutical preparations



(b) Manufacture of motor vehicles

Table 1. Evolution of the Numbers of Build and Buy Entries

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This table reports the ratio of Buy entries to Build entries. A firm is said to "buy" when it enters a new sector through a M&A (source: SDC Platinum and Bureau van Dijk Zephyr). A firm is said to "build" when it enters a new sector by one its existing subsidiaries (source: ESA survey).

Panel A. Buy and Build entries by year

	2004	2005	2006	2007	2009	2010	2011	2012	2013	2014
Build (number)	6295	10003	7360	8589	8478	8452	6861	5947	6468	6417
Buy (number)	119	109	138	140	121	144	166	195	144	164
Buy (%, frequency)	1.85	1.08	1.84	1.60	1.4	1.67	2.36	3.17	2.18	2.49
Buy (%, sales)	6.98	3.85	9.79	7.44	3.31	6.95	14.07	17.30	6.06	9.68

Panel B. Buy and Build entries by industry level

Industry level:	5 digits	4 digits	3 digits	2 digits	1 digit
Build (number)	74,870	72,250	63,176	49,570	32,431
Buy (number)	1,440	1,363	1,163	941	584
Buy (%, frequency)	1.89	1.85	1.81	1.86	1.77
Buy (%, sales)	8.05	8.55	8.31	10.57	15.72

Table 2. Summary Statistics

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. This table reports descriptive statistics for firms that are identified to enter a new sector, either internally or externally, during the periods 2003-2007 and 2009-2014. Panel A reports the distribution of the main firm characteristics. Panel B compares the means firms' characteristics that enter a new sector internally and those that enter by acquisition. A firm is said to "buy" when it enters a new sector through a M&A (source: SDC Platinum and Bureau van Dijk Zephyr). A firm is said to "build" when it enters a new sector by one its existing subsidiaries (source: ESA survey). Sectors refer to an industry at the 5-digits level of the French SIC. Description of the variables are reported in Appendix C.

Panel A. Distribution of firms characteristics

	N	Mean	St.Dev.	Percentiles				
				5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
$\mathbb{1}(\text{Buy})_{g,n,t}$	76310	0.02	0.14	0.00	0.00	0.00	0.00	0.00
$\text{Entry sales}_{g,n,t}$ (M euros)	76310	2.82	21.32	0.00	0.05	0.27	1.21	8.95
$\mathbb{1}(\text{1-year survival})_{g,n,t+1}$	76310	0.47	0.50	0.00	0.00	0.00	1.00	1.00
$\text{Investment}_{g,n,t}$ (M euros)	76310	0.96	16.24	0.00	0.03	0.10	0.35	2.44
$\text{Internal Human Capital}_{g,n,t\text{-}1}$	76310	0.14	1.04	-1.29	-0.59	-0.00	0.69	2.05
$\text{N.workers}_{g,t\text{-}1}$ (in thousands)	76310	0.63	5.16	0.02	0.04	0.07	0.19	1.29
$\text{Value added/N.workers}_{g,t\text{-}1}$	76310	0.05	0.03	0.02	0.03	0.04	0.06	0.11
$\text{Fixed assets/N.workers}_{g,t\text{-}1}$	76310	0.04	0.05	0.00	0.01	0.03	0.06	0.14
$\text{Cash holdings/N.workers}_{g,t\text{-}1}$	76310	0.02	0.02	0.00	0.00	0.01	0.02	0.06
$\text{Product market distance}_{g,n,t\text{-}1}$	76310	0.87	0.15	0.53	0.81	0.94	0.98	1.00
$\text{Vertical integration}_{g,n,t\text{-}1}$	44947	0.67	0.47	0.00	0.00	1.00	1.00	1.00

Panel B. Comparison of build and buy entries

	Build		Buy		Difference	
	Mean	St.Dev.	Mean	St.Dev.	Mean	p-value
$\text{Entry sales}_{g,n,t}$ (M euros)	2.64	20.52	12.04	45.85	$-9.39 * **$	(0.00)
$\mathbb{1}(\text{1-year survival})_{g,n,t+1}$	0.47	0.50	0.59	0.49	$-0.13 * **$	(0.00)
$\text{Investment}_{g,n,t}$ (M euros)	0.78	15.35	10.72	40.43	$-9.95 * **$	(0.00)
$\text{Internal Human Capital}_{g,n,t\text{-}1}$	0.15	1.04	-0.21	0.77	$0.36 * **$	(0.00)
$\text{N.workers}_{g,t\text{-}1}$ (in thousands)	0.59	4.82	2.37	14.33	$-1.77 * **$	(0.00)
$\text{Value added/N.workers}_{g,t\text{-}1}$	0.05	0.03	0.07	0.04	$-0.01 * **$	(0.00)
$\text{Fixed assets/N.workers}_{g,t\text{-}1}$	0.04	0.05	0.05	0.06	-0.00	(0.16)
$\text{Cash holdings/N.workers}_{g,t\text{-}1}$	0.02	0.02	0.02	0.02	$-0.01 * **$	(0.00)
$\text{Product market distance}_{g,n,t\text{-}1}$	0.87	0.16	0.92	0.10	$-0.05 * **$	(0.00)
$\text{Vertical integration}_{g,n,t\text{-}1}$	0.68	0.47	0.59	0.49	$0.08 * **$	(0.00)
N	74,870		1,440		76,310	

Table 3. Human Capital and Corporate Diversification

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and analyses the role of human capital on the type of diversification strategy. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include Sector Origin × Entry × Year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(Buy)_{g,n,t}$			
	(1)	(2)	(3)	(4)
Internal Human Capital$_{g,n,t-1}$	-0.015***	-0.010***		
	(0.004)	(0.003)		
log(N.workers)$_{g,t-1}$		0.010***	0.010***	0.011***
		(0.002)	(0.002)	(0.002)
Cash holdings/N.workers$_{g,t-1}$		0.044	0.043	0.054
		(0.045)	(0.050)	(0.058)
Fixed assets/N.workers$_{g,t-1}$		-0.021	-0.024	-0.023
		(0.028)	(0.031)	(0.034)
Value added/N.workers$_{g,t-1}$		0.049	0.046	0.038
		(0.035)	(0.038)	(0.043)
Internal Human Capital$_{g,n,t-2}$			-0.011***	
			(0.003)	
Internal Human Capital$_{g,n,t-3}$				-0.014***
				(0.004)
Sector origin × entry × year FE	Yes	Yes	Yes	Yes
R^2	0.199	0.206	0.208	0.207
N	76,354	76,296	66,145	54,230

Table 4. The Role of Size and Financial Constraints

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of firms' financial constraints on the relationship between human capital and the type of diversification strategy. Three proxy of financial constraints are considered: size, cash holding and public ownership status. Columns (1) and (2) include the second and third terciles of size (number of workers) interacted with *Internal Human Capital*$_{g,n,t-1}$. Columns (3) and (4) include the second and third terciles of cash holdings over workers (*Cash holdings/N. workers*) interacted with *Internal Human Capital*$_{g,n,t-1}$. In columns (5) and (6), we split firms between publicly and privately owned firms. Public firms are those that include at least one publicly listed firm within the firm. The dependent variable $\mathbb{1}(Buy)_{g,n,t}$ is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. Entries are identified at the 5-digit level of the French SIC. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(Buy)_{g,n,t}$					
	Size		Cash/N. workers		Type of firm	
					Public	Private
	(1)	(2)	(3)	(4)	(5)	(6)
Internal Human Capital$_{g,n,t-1}$	-0.011***	-0.011***	-0.014***	-0.010***	-0.123**	-0.007***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.051)	(0.002)
2nd tercile of N.workers$_{g,t-1}$	0.005**	0.005**				
	(0.002)	(0.002)				
3rd tercile of N.workers$_{g,t-1}$	0.025***	0.025***				
	(0.005)	(0.006)				
2nd N.workers$_{g,t-1}$× Internal HC$_{g,n,t-1}$	-0.001	-0.001				
	(0.001)	(0.001)				
3rd t. N.workers$_{g,t-1}$× Internal HC$_{g,n,t-1}$	-0.004	-0.004				
	(0.003)	(0.003)				
2nd tercile of Cash$_{g,t-1}$			0.002	0.001		
			(0.002)	(0.001)		
3rd tercile of Cash$_{g,t-1}$			0.004**	0.003		
			(0.002)	(0.002)		
2nd t. Cash$_{g,t-1}$× Internal HC$_{g,n,t-1}$			-0.000	0.000		
			(0.001)	(0.001)		
3rd t. Cash$_{g,t-1}$× Internal HC$_{g,n,t-1}$			-0.003*	-0.002		
			(0.002)	(0.002)		
Controls	No	Yes	No	No	Yes	Yes
Sector origin × entry × year FE	Yes	Yes	Yes	Yes	No	No
Sector Origin × entry FE	No	No	No	No	Yes	Yes
Year FE	No	No	No	No	Yes	Yes
R^2	0.204	0.205	0.199	0.206	0.107	0.123
N	76,354	76,296	76,354	76,354	1,198	74,204

Table 5. The Role of Sectoral Complementarities

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of complementarities between the sector of origin and sector of entry on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). In addition, columns (1) and (2) include a product market distance adapted from Bloom, Schankerman and Van Reenen (2013). It measures the distance between the sectors in which firm g operates at $t-1$ and the sector of entry n. The distance ranges from 0 to 1 (1 being the maximum). Columns (3) and (4) include the Fan and Goyal (2006)'s measure of vertical relatedness. It measures the intensity of vertical links between the main sector of activity of firm g at time $t-1$ and the sector in which g enters at time t. "Vertical" is a dummy variable that takes the value 1 if the vertical relatedness exceed 5%. Sector of origin × sector of entry × year fixed effects are included in columns (1) and (2). Models include sector of origin, sector of entry, year fixed effects separately. Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
	Product market distance		Vertical Integration	
	(1)	(2)	(3)	(4)
Internal Human Capital$_{g,n,t-1}$	-0.010***	-0.006***	-0.019***	-0.018***
	(0.003)	(0.002)	(0.005)	(0.005)
2nd t. Product Market Distance$_{g,n,t-1}$	0.003	0.007**		
	(0.003)	(0.003)		
3rd t. Product Market Distance$_{g,n,t-1}$	0.009**	0.012***		
	(0.004)	(0.004)		
2nd t. Distance$_{g,n,t-1}$ × Internal HC$_{g,n,t-1}$		-0.007***		
		(0.002)		
3rd t. Distance$_{g,n,t-1}$ × Internal HC$_{g,n,t-1}$		-0.008***		
		(0.002)		
Vertical integration$_{g,n,t-1}$			0.001	0.000
			(0.006)	(0.006)
Vertical integration$_{g,n,t-1}$ × Internal HC$_{g,n,t-1}$				-0.002
				(0.005)
Controls	Yes	Yes	Yes	Yes
Sector origin × entry × year FE	Yes	Yes	No	No
Sector of origin FE	No	No	Yes	Yes
Sector of entry FE	No	No	Yes	Yes
Year FE	No	No	Yes	Yes
R^2	0.208	0.208	0.102	0.102
N	76,226	76,226	68,531	68,531

Table 6. Alternative Measures of Human Capital

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and assess how our results are robust to the definition of human capital. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. Entries are identified at the 5-digit level of the French SIC. We test several alternative definition of our main independent variable *Internal Human Capital*$_{g,n,t-1}$. In the baseline definition, the measure is computed as the sum of occupation × sector × year fixed effects present in the workforce of firm g at time $t-1$. The measure is scaled by the number of occupations in firm g. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). In column (1), the dependent variable is a dummy equal to one if for a sector of entry n and a time of entry t, firm g has no occupation in the top 10 of occupation-sector-year fixed effects. In column (2), the sum of occupation × sector × year fixed effects is weighted by the number of employees by occupation in the firm. In column (3), CEO occupations are excluded from the sum of occupations. In column (4), occupation × sector × year fixed effects are estimated at the plant-level instead of the firm level in the baseline model. Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{No Top 10})_{g,n,t-1}$	0.008***			
	(0.003)			
Internal Human Capital$_{g,n,t-1}$ (weighted)		-0.009***		
		(0.002)		
Internal Human Capital$_{g,n,t-1}$ (no CEO)			-0.010***	
			(0.003)	
Internal Human Capital$_{g,n,t-1}$ (plant-level)				-0.009***
				(0.003)
Sector origin × entry × year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R^2	0.205	0.206	0.205	0.206
N	76,296	75,614	76,270	76,094

Table 7. Selection into Diversification

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests two potential self-selection stories that could drive the main results. The dependent variable $\mathbb{1}(\text{Buy})_{g,n,t}$ is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). In columns (1) to (3), firms that decrease their activity in a preexisting sector while entering sector n (*Shifting firms*) are excluded. *Shifting firms* are firms for which the minimum growth rate of sales is negative and greater than 100%, 50% and 25% in absolute value. We compute the growth rate of sales between $t-1$ and t in each sector in which firms were operating at $t-1$ and take the firm-level minimum of sectoral growth rates. In column (4), *Serial acquirers* are excluded. Serial acquirers are firms that enter more than one sector by acquisition during the time period. Control variables include the number of workers in logarithm as well total cash holdings, tangible assets and value added, all three scaled by the number of workers in the firm. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
Excluding:	Shifting firms			Serial acquirers
	100%	50%	25%	
	(1)	(2)	(3)	(4)
Internal Human Capital$_{g,n,t-1}$	-0.010***	-0.012***	-0.012***	-0.003***
	(0.003)	(0.003)	(0.004)	(0.001)
log(N.workers)$_{g,t-1}$	0.010***	0.010***	0.011***	0.003***
	(0.002)	(0.002)	(0.003)	(0.001)
Cash holdings/N.workers$_{g,t-1}$	0.056	0.072	0.073	0.001
	(0.046)	(0.058)	(0.057)	(0.014)
Fixed assets/N.workers$_{g,t-1}$	-0.010	0.007	0.041	-0.011
	(0.029)	(0.032)	(0.036)	(0.010)
Value added/N.workers$_{g,t-1}$	0.043	0.039	0.024	0.016
	(0.033)	(0.043)	(0.050)	(0.016)
Controls	Yes	Yes	Yes	Yes
Sector origin × entry × year FE	Yes	Yes	Yes	Yes
R^2	0.205	0.222	0.243	0.157
N	74,510	54,072	38,847	72,681

Table 8. Complementarities between Human and Physical Capital

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of labor and capital complementarities on the relationship between human capital and the type of diversification strategy. The relationship between human capital and the type of diversification strategy is compared between firms that realized similar entry sales or capital expenditures during the year of entry. In columns (1) and (2), we rank entry sales into ten deciles and run our baseline regressions with interacted sector of origin × sector of entry × year × sales decile fixed effects. Sales are defined as the total amount of sales realized in the sector of entry n by firm g at time t. In columns (3) and (4), we consider replace sales with investment deciles. In the case of build entries, investment is measured as the total amount of capital expenditures realized by firm's firms that entered in sector n at time t. In the case of Buy entries, investment is the amount of acquired fixed physical assets. The dependent variable is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). Control variables include the number of workers in logarithm as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$			
	Sales		Investment	
	(1)	(2)	(3)	(4)
Internal Human Capital$_{g,n,t-1}$	-0.009***	-0.007**	-0.004*	-0.003*
	(0.003)	(0.003)	(0.002)	(0.002)
Controls	No	Yes	No	Yes
Sector origin × entry × year FE × Sales bucket	Yes	Yes	No	No
Sector origin × entry × year FE × Investment bucket	No	No	Yes	Yes
R^2	0.232	0.235	0.267	0.267
N	45,959	45,959	31,735	31,735

Table 9. Reallocation Costs in the Internal Labor Market

Source: EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Subsidiaries of firms that enter a new sector internally during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests whether subsidiaries that enter the new sector have high human capital compared to the other subsidiaries of the same firm. The dependent variable is a dummy variable that takes the value one if the entry in sector n at time t is made through subsidiary f, zero if the entry is not made through subsidiary f within firm g. Entries are identified with sales reported at the 5-digit level of the French SIC. The main independent variable is a subsidiary-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the subsidiary is adapted to the sector of entry (see section 3.2). Column (1) and (2) are the baseline specifications. In columns (3) and (4), the measure of Internal Human Capital is calculated at $t-2$ and $t-3$ respectively. Columns (2) to (4) include the following set of control variables: the number of workers, the subsidiary's cash holdings, amount of tangible assets and value added, the last three variables being scaled by the number of workers in the subsidiary. All models are estimated with firm $\times$ sector of entry $\times$ year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Build})_{f,n,t}$			
	(1)	(2)	(3)	(4)
Internal Human Capital$_{f,n,t-1}$	0.021***	0.024***		
	(0.004)	(0.004)		
log(N. workers)$_{f,t-1}$		0.018***	0.017***	0.017***
		(0.003)	(0.003)	(0.003)
Cash holdings/N.workers$_{f,t-1}$		0.312**	0.325**	0.314*
		(0.153)	(0.161)	(0.160)
Fixed assets/N.workers$_{f,t-1}$		0.078***	0.080***	0.086***
		(0.026)	(0.026)	(0.029)
Value added/N.workers$_{f,t-1}$		0.080**	0.070**	0.062*
		(0.032)	(0.034)	(0.035)
Internal Human Capital$_{f,n,t-2}$			0.024***	
			(0.004)	
Internal Human Capital$_{f,n,t-3}$				0.023***
				(0.004)
Firm $\times$ sector of entry $\times$ year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	No
R^2	0.229	0.234	0.236	0.233
N	362,089	362,089	316,760	267,060

Table 10. Internal Human Capital and Employment Growth

Source: matched employer-employee dataset, ownership links dataset, EAE survey. *Sample:* Subsidiaries through which internal entries are realized during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the link between internal human capital and the variation of the number of workers. The main variable is the growth rate of the number of workers (computed as in Davis and Haltiwanger (1992)). The variation is computed between [t-1;t+1], [t-1;t], and [t;t+1] with t the year of entry in the new sector. The main independent variable is a subsidiary-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the subsidiary is adapted to the sector of entry (see section 3.2). Control variable include the following set of control variables: the number of workers, the subsidiary's cash holdings, amount of tangible assets and value added, the last three variables being scaled by the number of workers in the subsidiary. All models include sector of origin × entry × year fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	Δ number of workers		
	[t-1 ; t+1]	[t-1 ; t]	[t ; t+1]
	(1)	(2)	(3)
Internal Human Capital$_{f,n,t-1}$	-0.013*	0.002	-0.016**
	(0.007)	(0.001)	(0.008)
Controls	Yes	Yes	Yes
Sector origin × entry × year FE	Yes	Yes	Yes
R^2	0.639	0.102	0.659
N	19,205	26,003	19,205

Table 11. Diversification, Human Capital and Local Labor Market Tightness

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of tight local labor markets for key occupations on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). *LLM Tightness*$_{n,z,t}$ is the sum of occupations in short supply in the local labor market z at time $t-1$, weighted by the occupation × sector × year fixed effects, scaled by the number of occupations present in sector n at (see equation (4.2.3)). Terciles of LLM tightness are included in columns (1) and (2). Models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. Control variables include the number of workers in logarithm as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector origin × sector of entry × year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
Tercile of LLM Tightness$_{n,z,t-1}$:	All		≤ P33	≥ P33 and ≤ P66	≥ P66
	(1)	(2)	(3)	(4)	(5)
Internal Human Capital$_{g,n,t-1}$	-0.016***	-0.008**	-0.003	-0.013**	-0.016***
	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)
2nd tercile of LLM Tightness$_{n,z,t-1}$	0.007**	0.006**			
	(0.003)	(0.003)			
3rd tercile of LLM Tightness$_{n,z,t-1}$	0.014***	0.012***			
	(0.005)	(0.004)			
2nd t. LLM Tightness$_{n,z,t-1}$ × Internal HC$_{g,n,t-1}$	-0.005**	-0.004**			
	(0.002)	(0.002)			
3rd t. LLM Tightness$_{n,z,t-1}$ × Internal HC$_{g,n,t-1}$	-0.005**	-0.004*			
	(0.002)	(0.002)			
Controls	No	Yes	Yes	Yes	Yes
Sector origin × entry × year FE	Yes	Yes	Yes	Yes	Yes
Local Labor Market FE	Yes	Yes	Yes	Yes	Yes
R^2	0.418	0.424	0.525	0.501	0.377
N	28,598	28,957	7,953	7,796	8,067

Table 12. Local Labor Market Tightness and the Value of Building

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector by building during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of tight local labor markets for key occupations on the relationship between human capital and performance on the sector of entry. The dependent variable $log(Sales)_{g,n,z,t}$ is the logarithm of sales realized by firm g in sector n the year of entry t. The variable is defined only for Build entries, i.e., when $1(Buy)_{g,n,z,t} = 0$. The main independent variable is a firm-level measure of human capital $Internal\ Human\ Capital_{g,n,t-1}$. The variable measures to which extent the workforce of the firm is adapted to the sector of entry (see section 3.2). $LLM\ Tightness_{n,z,t}$ is the sum of occupations in short supply in the local labor market z at time $t-1$, weighted by the occupation × sector × year fixed effects, scaled by the number of occupations present in sector n at (see equation (4.2.3)). Terciles of LLM tightness are included in columns (1) and (2). Models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. Control variables include the number of workers in logarithm as well total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector origin × sector of entry × year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$log(Sales)_{g,n,t}$				
Tercile of LLM Tightness$_{n,z,t-1}$:	All		≤ P33	≥ P33 and ≤ P66	≥ P66
	(1)	(2)	(3)	(4)	(5)
Internal Human Capital$_{g,n,t-1}$	0.105***	0.099***	0.083	0.038	0.139**
	(0.021)	(0.028)	(0.051)	(0.056)	(0.061)
2nd tercile of LLM tightness$_{n,z,t-1}$		-0.014			
		(0.036)			
3rd tercile of LLM tightness$_{n,z,t-1}$		-0.012			
		(0.049)			
2nd t. LLM tightness$_{n,z,t-1}$ × Internal HC$_{g,n,t-1}$		0.003			
		(0.029)			
3rd t. LLM tightness$_{n,z,t-1}$ × Internal HC$_{g,n,t-1}$		0.012			
		(0.030)			
Controls	Yes	Yes	Yes	Yes	Yes
Sector origin × entry × year FE	Yes	Yes	Yes	Yes	Yes
Local Labor Market FE	Yes	Yes	Yes	Yes	Yes
R^2	0.498	0.498	0.571	0.502	0.491
N	27,747	27,747	7,634	7,500	7,772

Appendix A Theoretical Appendix

A.1 Predictions 1 and 2

When $\gamma > (\sigma-1)/\sigma$, the difference $\Delta(c_I, c_E, c_A) = \Pi^{Build}(c_I, c_E) - \Pi^{Buy}(c_I, c_E, c_A)$ is a decreasing function of c_I. Moreover, $\lim_{c_I \to 0} \Delta(.) = F > 0$, and

$$\lim_{c_I \to \infty} \Delta(c_I) = F + K\left(c_E^{-\frac{\gamma}{\gamma-1}}\right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} - K\left(c_E^{-\frac{\gamma}{\gamma-1}} + c_A^{-\frac{\gamma}{\gamma-1}}\right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}}, \tag{16}$$

which is negative if F is small enough relative to the marginal labor costs of workers in the target firm. This ensures the unique existence of $c_I^* > 0$ such that for any c_E and c_A, $c_I > c_I^*$ implies $\Delta(.) = 0$. This proves prediction 1. Moreover, $\Delta(.)$ is a decreasing function of c_E, thus $\Delta(.) = 0$ for a lower threshold c_I^* when c_E is large, implying prediction 2.

A.2 Endogenous restructuring costs

We propose a micro-foundation for F, based on the cost of laying off some workers in the acquired firm. We denote ϕ as the marginal cost of layoff, L_A^{before} as the labor input of the acquired firm *before* the acquisition, and L_A^{after} as the labor input *after* the merger has occurred. We assume the layoff cost is proportional to the distance between L_A^{before} and L_A^{after}:

$$F(L_A^{before}, L_A^{after}) = \phi(L_A^{before} - L_A^{after}). \tag{17}$$

Given the standard monopolistic competition framework, output Y is equal to $K(\sigma-1)c^{-\sigma}$. Since $Y = \mathcal{L}$, we can write the labor input of the acquired firm before the acquisition as

$$L_A^{before} = c_A^{-\frac{1}{\sigma}} \cdot (K(\sigma-1))^{\frac{1}{\sigma}}. \tag{18}$$

The acquiring firm minimizes the total labor cost across the three worker pools $i \in \{I, E, A\}$, equal to $\sum_i c_i L_i$, subject to the production function (1). The first-order conditions of this problem can be rewritten as

$$L_i = \frac{c_i^{-\frac{1}{1-\gamma}}}{\sum_i c_i^{-\frac{\gamma}{1-\gamma}}} \cdot \sum_i c_i L_i, \tag{19}$$

so that after the acquisition, the labor input coming from the acquired firm is

$$L_A^{after} = c_A^{-\frac{1}{\sigma}} \cdot (K(\sigma-1))^{\frac{1}{\sigma}} \cdot \left(1 + \left(\frac{c_E}{c_A}\right)^{-\frac{\gamma}{1-\gamma}} + \left(\frac{c_I}{c_A}\right)^{-\frac{\gamma}{1-\gamma}}\right)^{-\frac{1}{\gamma}+\frac{1}{\sigma}\left(\frac{1-\gamma}{\gamma}\right)}. \tag{20}$$

Comparing (18) and (20), we have $L_A^{after} \leq L_A^{before}$. It is clear from (20) that $F(L_A^{before}, L_A^{after})$ is decreasing in c_I and c_E. Recall that $\Delta(c_I, c_E, c_A)$ is also a decreasing function of c_I and c_E. We now have $\lim_{c_I \to 0} \Delta(c_I) = \phi L_A^{before} > 0$, and

$$\begin{aligned}\lim_{c_I \to \infty} \Delta(c_I) =& K\left(c_E^{-\frac{\gamma}{\gamma-1}}\right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} - K\left(c_E^{-\frac{\gamma}{\gamma-1}} + c_A^{-\frac{\gamma}{\gamma-1}}\right)^{\frac{(\gamma-1)(\sigma-1)}{\gamma}} \\ &+ \phi c_A^{-\frac{1}{\sigma}} \cdot (K(\sigma-1))^{\frac{1}{\sigma}} \left(1 - \left(1 + \left(\frac{c_E}{c_A}\right)^{-\frac{\gamma}{1-\gamma}}\right)^{-\frac{1}{\gamma}+\frac{1}{\sigma}\left(\frac{1-\gamma}{\gamma}\right)}\right), \end{aligned} \tag{21}$$

$\Delta(.)$ keeps the same properties as before. It is strictly decreasing, positive when $c_I = 0$, and negative when $c_I \to \infty$ if (21) is negative, i.e., if ϕ is not too large. Therefore, the testable predictions stated in the main text remain valid.

A.3 Optimal marginal costs (Equation (8))

Given a minimum threshold $\underline{h}_{io}$ and the random match quality $h \geq 1$ following a Pareto distribution with cf $\Psi(h)$, each worker occupation $o \in \mathcal{O}$ supplies the following amount of human capital:

$$H_{io} \equiv N_i a_{io} m_o \int_{\underline{h}_{io}}^{\infty} h d\Psi(h) = \frac{N a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}, \tag{22}$$

with $i \in \{I, E, A\}$.

We write the Lagrangian of the minimization of (6) subject to (7) and (22):

$$\begin{aligned}\mathcal{L} =& N_i \left[\sum_{o \in \mathcal{O}} a_{io} w_o \underline{h}_{io}^{-k} + f c_i\right] \\ &+ \sum_{o \in \mathcal{O}} \mu_o \left[H_{io} - \frac{N_i a_{io} m_o k \underline{h}_{io}^{1-k}}{k-1}\right] \\ &+ \lambda_{io} \left[L_i - \left(\sum_{o \in \mathcal{O}} H_{io}^{\theta}\right)^{1/\theta}\right].\end{aligned}$$

We obtain the first-order conditions

$$\mu_o = \frac{w_o}{m_o \underline{h}_{io}} \qquad \text{(wrt } \underline{h}_{io}\text{)} \tag{23}$$

$$C_i = \sum_{o \in \mathcal{O}} \mu_o H_{io} \qquad \text{(wrt } N_i\text{)} \tag{24}$$

$$\frac{w_o}{m_o \underline{h}_{io}} = \lambda H_{io}^{\theta-1} \left(\sum_{o \in \mathcal{O}} H_{io}^{\theta}\right)^{\frac{1}{\theta}-1} \qquad \text{(wrt } H_{io}\text{).} \tag{25}$$

Plugging (23) into (25), then rearranging and summing over o, we obtain

$$C_i = \left[\sum_{o \in \mathcal{O}} \left(\frac{a_{io} m_o^k w_o^{1-k}}{L_i^k (N_i k)^{-1} (k-1)} \right)^{\frac{\theta}{\beta}} \right]^{\frac{\beta}{\theta(1-k)}}. \tag{26}$$

We guess that $C_i = c_i L_i$. Using the definition of total costs C_i given by (6) we have that

$$\begin{aligned} c_i L_i &= N_i \left(\sum_{o \in \mathcal{O}} a_{io} w_o \underline{h}_{io}^{-k} + f c_i \right) \\ &= \sum_{o \in \mathcal{O}} \left(\frac{k-1}{k} \right) \frac{m_o}{w_o \underline{h}_{io}} H_{io} + N_i f c_i \\ &= \left(\frac{k-1}{k} \right) c_i L_i + N_i f c_i, \end{aligned} \tag{27}$$

where we use of Equation (22) on the second row and Equation (25) on the third row. It follows naturally from (27) that $N_i = \frac{L_i}{fk}$. Now, plugging $C_i = c_i L_i$ into (26) and using $N_i = \frac{L_i}{fk}$, we find that the optimal unit labor cost function for each labor market $i \in \{I, E, A\}$ is:

$$c_i = \left[\sum_{o \in \mathcal{O}} \left(\frac{a_{io} m_o^k w_o^{1-k}}{f(k-1)} \right)^{\frac{\theta}{\beta}} \right]^{\frac{\beta}{\theta(1-k)}}. \tag{28}$$

A.4 The relative wage share for each worker occupation (Equation (9))

Putting together the first-order conditions (24) and (25) and taking the sum over $o \in \mathcal{O}$, we obtain

$$\frac{\frac{w_o}{m_o \underline{h}_{io}} H_{io}}{C_i} = \frac{H_{io}^{\theta}}{\sum_{o \in \mathcal{O}} H_{io}^{\theta}}, \tag{29}$$

so that the share of each worker occupation in total costs is equal to the share of human capital supplied by that worker occupation.

We normalize input from each labor market at $L_i = 1$, which implies $\sum_{o \in \mathcal{O}} H_{io}^{\theta} = 1$. Plugging (22) into (29) and using $N_i = \frac{1}{fk}$, we obtain

$$\underline{h}_{io} = \left(\frac{a_{io}}{f(k-1)} \right)^{\frac{1-\theta}{\beta}} w_o^{\frac{1}{\beta}} m_0^{-\frac{\theta}{\beta}} C_i^{-\frac{1}{\beta}}, \tag{30}$$

where $\beta \equiv \theta + \theta(1-k)$.

From Equation (27), we have that the costs of conducting interviews, expressed in labor units, are equal to $N_i f c_i = \frac{c_i}{k}$. This implies that for any number of workers $\tilde{A_{io}}$ to work in an

occupation, firms must hire a larger number of workers $A_{io} = \frac{k}{k-1}\tilde{A_{io}}$ because some workers will be conducting interviews. By assumption, the number of workers in occupation $o \in \mathcal{O}$ hired writes $\tilde{A_{io}} = N_i a_{io}(1 - \Psi(\underline{h}_{io})) = N_i a_{io} \underline{h}_{io}^{-k}$. Finally, using (30) and $N_i = \frac{1}{fk}$, we find that the total number of workers in occupation o hired to produce one unit of output is

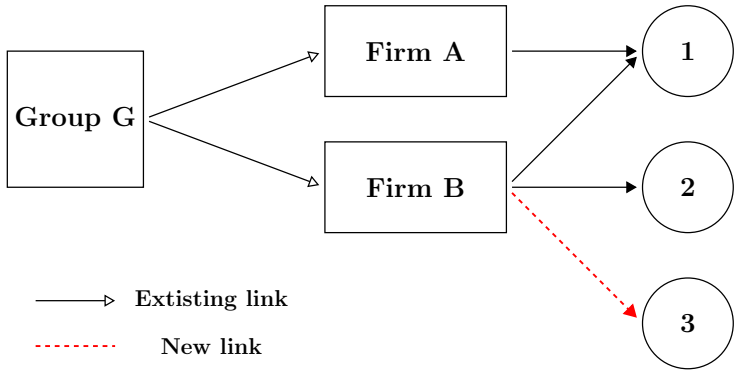
$$A_{io} = \left(\frac{a_{io}}{f(k-1)}\right)^{\frac{\theta}{\beta}} w_o^{\frac{-k}{\beta}} m_0^{\frac{k\theta}{\beta}} C_i^{\frac{k}{\beta}}. \tag{31}$$

It follows that relative share of the wage bill that goes to a given occupation o can be expressed as in Equation (9).

Appendix B Additional Tables and Figures for Internet Appendix

Figure A1. Identification of Build and Buy Entries

(a) "Build" entry in sector 3 through entity B



(b) "Buy" entry in sector 2 and 3 through the acquisition of B

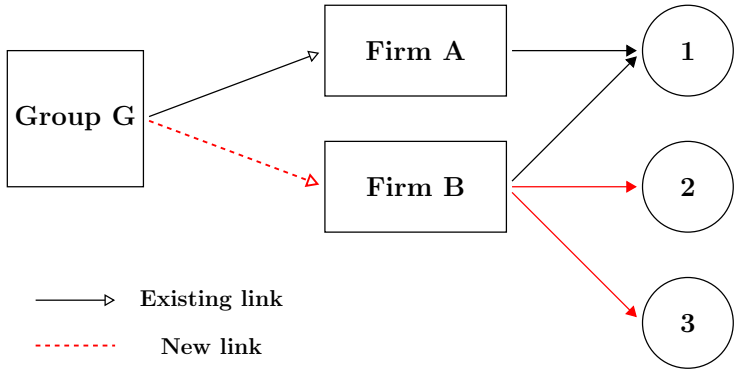


Table A1. Top 5 Occupations within Sector

Source: matched employer-employee dataset for the period 2003-2014. *Sample:* Universe of full-time French workers who earn at least 1,000 euros a year. This table reports the top five occupation × sector × year fixed effects estimated following equation (10). Occupation × sector × year fixed effects capture the average share of the wage bill that goes to an occupation within a sector a given year. The higher the fixed effects the more important the occupation for the production process of a given sector.

Panel A. Pharmaceutical preparations (2013)

Occupation	Estimated fixed effects
Technicians in production and control quality	0.99
Chemists, operators and skilled workers	0.96
R&D engineers and executives	0.61
Sales managers in SMEs	0.31
Sales representatives and technicians	0.29

Panel B. IT consultancy activities (2013)

Occupation	Estimated fixed effects
Computer science R&D engineers and executives	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
IT support engineers and executives	0.82

Panel C. Manufacture of motor vehicles (2013)

Occupation	Estimated fixed effects
Mechanical qualified assemblers in series	0.99
Technicians specialized in mechanics and metal work manufacturing and quality control	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 A2. Top 10 sectors of entry in 2013

The table reports the number of entries for sectors with the largest number of entries, by entry type in 2013. Build entries (Panel A) are identified as sales reported into a new sector (source: ESA survey). Buy entries (Panel B) are identified when firms realize diversifying acquisitions (source: SDC Platinum and Bureau van Dijk Zephyr). Sectors refer to an industry at the 4-digit level of the French SIC.

Panel A. Build entries

Sector of origin	Sector of entry	N. Pairs origin × 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

Panel B. Buy entries

Sector of origin	Sector of entry	N. Pairs origin × 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 activities	5

Table A3. Top 10 Occupations in Short Supply

Source: French unemployment agency. *Sample:* Occupations' supply and demand of 348 local labor markets in 2013. This table reports the top 10 occupations that are the reported as being in short supply the most often among local labor markets. The last column gives the percentage of local labor markets in which the occupation is reported as being in short supply. Occupations in short supply are identified by the French national employment agency as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers.

Rank	Occupation	% of local labor markets
1	Kitchen staff	77.0
2	Machining equipment operators	58.3
3	Butchers	51.7
4	Metal workers	50.0
5	Nurses	48.9
6	Technical and commercial relation managers	48.3
7	Bakers	48.0
8	Car mechanicians	48.0
9	Catering staff	48.0
10	Machining equipment maintenance workers	46.2

Table A4. Robustness Check: Alternative Measure of Local Labor Market Tightness

Source: SDC Platinum, BvD Zephyr, EAE survey, tax files, ownership links dataset, matched employer-employee dataset. *Sample:* Firms that enter a new sector during the periods 2003-2007 and 2009-2014. The table reports OLS estimates and tests the effect of tight local labor markets on the relationship between human capital and the type of diversification strategy. The dependent variable is a dummy variable that takes the value one if the entry of firm g in sector n at time t is made through an acquisition, zero if the entry is made internally. The main independent variable is a firm-level measure of human capital *Internal Human Capital*$_{g,n,t-1}$ that is the sum of occupation × sector × year fixed effects present both in the workforce of firm g at time $t-1$ and in the sector of entry n. *Internal Human Capital*$_{g,n,t-1}$ can be interpreted as the fit quality of a firm workforce to a sector prior to diversification. *LLM Tightness*$_{n,z,t}$ is the sum of occupations in short supply in the local labor market z at time $t-1$, scaled by the number of occupations present in sector n (see equation (36)). Terciles of LLM tightness are included in columns (1) and (2). Models in columns (3) to (5) are estimated on subsamples of the dataset split by terciles of LLM tightness. France is divided into 348 local labor markets. The set of control variables include the number of workers, total cash holdings of the firm, the total amount of tangible assets held by the firm and the total value added generated by the firm, all scaled by the number of workers in the firm. All models include sector origin × sector of entry × year fixed effects, as well as labor market fixed effects. Standard errors are double clustered at sector of origin and sector of entry levels and reported in parentheses. *, **, and *** mean significantly different from zero at 10, 5 and 1% levels, respectively.

Dependent variable:	$\mathbb{1}(\text{Buy})_{g,n,t}$				
Level of LLM Tightness$_{n,z,t-1}$:	All		≤ P33	≥ P33 and ≤ P66	≥ P66
	(1)	(2)	(3)	(4)	(5)
Internal Human Capital$_{g,n,t-1}$	-0.015***	-0.009**	-0.000	-0.011**	-0.019***
	(0.004)	(0.004)	(0.002)	(0.005)	(0.007)
2nd tercile of LLM Tightness$_{n,z,t-1}$	0.007*	0.006*			
	(0.004)	(0.004)			
3rd tercile of LLM Tightness$_{n,z,t-1}$	0.012**	0.012**			
	(0.005)	(0.005)			
2nd t. LLM Tightness$_{n,z,t-1}$ × Internal HC$_{g,n,t-1}$	-0.003**	-0.003			
	(0.002)	(0.002)			
3rd t. LLM Tightness$_{n,z,t-1}$ × Internal HC$_{g,n,t-1}$	-0.007**	-0.005*			
	(0.003)	(0.003)			
Controls	No	Yes	Yes	Yes	Yes
Sector origin × entry × year FE	Yes	Yes	Yes	Yes	Yes
Local Labor Market FE	Yes	Yes	Yes	Yes	Yes
R^2	0.418	0.424	0.498	0.543	0.400
N	28,957	28,957	7,897	7,778	7,882

Appendix C Description of Variables

Variables	Description
Dependent variables	
$\mathbb{1}(\text{Buy})_{g,n,t}$	Dummy variable that takes the value one if a firm g enters a sector n at year t through an acquisition, and zero if the entry is made internally by building on preexisting resources. Acquisitions are identified with SDC Platinum and Bureau van Dijk Zephyr databases. Build entries are identified using reported sales from the ESA survey at the 4-digit of the French SIC. *Source:* SDC Platinum, BvD Zephyr, Tax files, Ownership links dataset, ESA survey.
$\mathbb{1}(\text{Build})_{f,n,t}$	Dummy variable that takes the entry into sector n is made through a firm f within the firm g, and zero if the entry is *not* made through the firm f. This variable is constructed only for the subsample of Build entries for which we can identify precisely the entity responsible for the entry. *Source:* ownership links dataset, ESA survey.
Build entries	Number of occurrences in which new sales are reported in a new sector. *Source:* ESA survey.
Buy entries	Number of occurrences in which firms realize diversifying acquisitions. *Source:* SDC Platinum and Bureau van Dijk Zephyr.
Δ workers [t-i;t+j]	Growth rate of the number of workers computed following Davis and Haltiwanger (1992): $$\Delta\text{workers}_{t-i,t+j} = \frac{\#\ Workers_{t-i,t+j} - \#\ Workers_{t-i,t+j-1}}{0.5 * (\#\ Workers_{t-i,t+j} + \#\ Workers_{t-i,t+j-1})}$$. The growth rate ranges between -2 and 2. The variation is computed between [t-1;t+1], [t-1;t] and [t;t+1] with t the year of entry in the new sector. *Source:* Matched employer-employee dataset.
$\log(\text{Sales})_{g,n,t}$	logarithm of sales realized by firm g in sector n the year of entry t. *Source:* Tax files.
Independent variables	
Internal Human $\text{Capital}_{g,n,t\text{-}1}$	Sum of (exponentiated) occupation $\times$ sector $\times$ year fixed effects present both in the workforce of firm g at time $t-1$ and in the sector of entry n. The measure is scaled by the number of occupations in firm g. Fixed effects are estimated on the full French matched employer-employee dataset and are retrieved from a regression that takes the share of the wage bill that goes to a given occupation within a given sector as dependent variable. It can be interpreted as the fit quality of a firm workforce to a sector prior to diversification. *Source:* Matched employer-employee dataset.
Internal Human $\text{Capital}_{g,n,t\text{-}i}$	Lagged measure of *Internal Human Capital* at time $t-2$ or $t-3$. The variable is constructed using the same methodology as Internal Human $\text{Capital}_{g,n,t\text{-}1}$ *Source:* Matched employer-employee dataset.
$\mathbb{1}(\text{No Top 10})_{g,n,t\text{-}1}$	Dummy variable that takes the value 1 if none of the 10 most important occupation for the sector n of entry are present in the workforce of firm g at time $t-1$, and zero if there are. *Source:* Matched employer-employee dataset.
Internal Human Capital (weighted)	Variable constructed using the same methodology as Internal Human $\text{Capital}_{g,n,t\text{-}1}$, except the sum of occupation $\times$ sector $\times$ year fixed effects is weighted by the number of employees by occupation in the firm. *Source:* Matched employer-employee dataset.
Internal Human Capital (without CEO)	Variable constructed using the same methodology as Internal Human $\text{Capital}_{g,n,t\text{-}1}$, except the CEO occupations are excluded from the sum of occupations. *Source:* Matched employer-employee dataset.
Internal Human Capital (plant-level)	Variable constructed using the same methodology as Internal Human $\text{Capital}_{g,n,t\text{-}1}$, except the fixed effects are estimated at the plant-level instead of the firm level. *Source:* Matched employer-employee dataset.
Control variables	
$\log(\text{N.workers})_{g,t-1}$	Logarithm of the number of workers in firm g at time $t-1$. *Source:* Matched employer-employee dataset, Ownership links dataset.

Continued next page

Description of Variables (continued)

Variables	Description
Value added/N.workers$_{g,t-1}$	Total value added generated by firm g at time $t-1$, scaled by the number of workers in the firm. *Source:* Tax files, Ownership links dataset, Matched employer-employee dataset.
Fixed assets/N.workers$_{g,t-1}$	Total value of fixed assets held by firm g at time $t-1$, scaled by the number of workers in the firm. *Source:* Tax files, Ownership links dataset, Matched employer-employee dataset.
Cash holdings/N.workers$_{g,t-1}$	Total cash holdings of firm g at time $t-1$, scaled by the number of workers in the firm. *Source:* Tax files, Ownership links dataset, Matched employer-employee dataset.
Product market distance$_{g,n,t-1}$	Distance between the sectors in which firm g operates at $t-1$ and the sector of entry n. The distance ranges from 0 to 1. A value of 0 (resp. 1) indicates that the sales of firm g are perfectly correlated (resp not correlated) at time $t-1$ to the sales of the firms operating in the sector of entry n. The product market distance is adapted from Bloom, Schankerman and Van Reenen (2013). See section 4.2.2 for more details. *Source:* Ownership links dataset, Tax files, ESA survey.
Vertical integration$_{g,n,t-1}$	Dummy variable that is equal to one if the sector of origin of the firm (as measured in the I-O OECD classification) sources more than 5% of its inputs from the sector of entry (as measured in the I-O OECD classification). The variable is not defined when the sector of origin and the sector of entry belongs to the same item of the I-O OECD classification. The methodology follows Fan and Goyal (2006)'s measure of vertical relatedness. *Source:* Ownership links dataset, Tax files, OECD 1999 French Input-Output table.
LLM tightness$_{n,z,t-1}$	Sum over all occupations in short supply in the local labor market z at time t of the exponentiated occupation $\times$ sector $\times$ year fixed effects for sector n, scaled by the number of occupations observed in sector n at time t (see equation 4.2.3). Occupations in short supply in a LLM are identified by the French national employment agency Pole Emploi as occupations for which (i) the ratio of job offers over job applications is high (ii) surveyed employers forecast that it will be difficult to fill posted offers. France is divided into 348 different local labor markets. *Source:* Matched employer-employee dataset, Pole Emploi.
<u>Other variables</u>	
Type of firm	Dummy variable that is equal to one if at least one entity within the firm is a publicly listed company, and zero otherwise. *Source:* Bureau van Dijk Amadeus
Shifting firms	Shifting firms are firms that decrease their activity in a preexisting sector while entering sector n. They are identified with the growth rate of sales between $t-1$ and t in each sector in which firms were operating at $t-1$ that is negative and greater than 100%, 50% and 25% in absolute value. We take the firm-level minimum of sectoral growth rates. *Source:* ESA survey, Ownership links dataset.
Serial acquirers	Serial acquirers are firms that enter more than one sector by acquisition during the time period. *Source:* SDC Platinum, BvD Zephyr, Ownership links dataset, ESA survey
Entry sales$_{g,n,t}$	Total sales reported by firm g in sector n at time t. *Source:* Tax files
$\mathbb{1}$(1-year survival)$_{g,n,t+1}$	Dummy equal to one if firm g reports sales in sector of entry n one year after the entry n at time t. *Source:* Tax files
Sales bucket	Deciles of realized sales by the firm g in sector n at time t. The sales buckets are used as fixed effects to compare firms within the same entry sale bucket. *Source:* Tax files
Investment bucket	Decile of total capital expenditures relaized by firm g to enter sector n. Investment is measured in the case of build entry as the total amount of investment realized by the firm(s) of the firm that entered in sector n at time t. In the case of acquisitions, investment is set equal to the amount of fixed assets of the firm that has been acquired. The investment buckets are used as fixed effects to compare firms within the same investment bucket. *Source:* Tax files