Edoardo Di Porto[1], Marco Pagano[2], Vincenzo Pezone[3], Raffaele Saggio[4], and Fabiano Schivardi[5]

[1] *University of Naples Federico II, CSEF and INPS*
[2] *University of Naples Federico II, CSEF, ECGI and EIEF*
[3] *Tilburg University*
[4] *University of British Columbia and NBER*
[5] *Luiss University, EIEF and CEPR*

November 2023

Abstract

We investigate the compensation policies of family and non-family firms using a novel employer-employee matched dataset, comprising the universe of Italian companies, combined with information on their shareholders. Family firms pay significantly lower wages and offer slower careers: half of their wage gap is accounted for by workers' time invariant characteristics, while the rest cannot be explained by differences in productivity or in non-monetary compensating differentials. The evidence is consistent with a model where family owners protect their control over the firm by creating a "glass ceiling" that limits workers' career progression, thus hindering their human capital accumulation.

Keywords: family firms, corporate control, wage, career, workers, human capital, productivity, management.

JEL Codes: D22, D23, D24, G32, G34, J24, J31, J32, J62, M12, M51, M52, M54.

1 Introduction

Family firms employ a sizable fraction of the total workforce: between 40% and 50% of all European private employment is accounted for by family firms.[1] In this paper we investigate whether the employment policies of these firms differ from that those of non-family firms and, if so, what accounts for these differences.

We address these questions using a rich merged employer-employee dataset for the universe of Italian companies, based on social security records and detailed information about firms' ownership structure and financial accounts. The data includes almost 900,000 unique firms and over 19 million workers, over a 14-year time period. Such data presents two key advantages over those analyzed in previous work. First, it covers highly heterogeneous companies, ranging from private firms with few employees to large public corporations, whereas most previous research focuses on very large firms. Second, we have information on the careers of individual workers, rather than firm-level employment and labor cost data.

A prerequisite for our analysis is to identify family firms in our sample. In most of our analysis, a family firm is defined as one where a single family holds more than 50% of the shares. The family firms in our sample are smaller and substantially less productive than non-family ones. However, the two types of firms have very similar return on assets. The fact that family firms are not less profitable despite their lower productivity can be attributed at least partly to their ability to pay substantially lower wages. We will refer to this wage gap, which is going to be the main focus of our analysis, as the "family firm discount".

The first step in our work is to measure the gap in compensation between family and non-family firms. Controlling for gender and experience, the employees of family firms earn on average 16% less than those of non-family firms, in line with previous evidence by

[1] See the data reported by https://europeanfamilybusinesses.eu/about-european-family-businesses/, where family firms are defined as those where the majority of decision rights are in the hands of a family, and at least one family member is involved in the firm's governance.

Sraer and Thesmar (2007) and Bassanini et al. (2013) based on French data, and by Ellul et al. (2017) based on cross-country data for listed companies.

Second, upon decomposing this wage gap into the portion associated with workers' characteristics and that related to firm characteristics, based on the "AKM" method by Abowd et al. (1999), it turns out that roughly half of the total wage differential (8 percentage points) is accounted for by individual characteristics of family firms' employees and the remaining half by firm characteristics. Next, we further partition the portion of the wage differential accounted for by firm characteristics in three components, using the Oaxaca-Blinder method (Oaxaca, 1973; Blinder, 1973): (i) a component related to the firm's ownership (family or non-family); (ii) a rent-sharing component, capturing how firms split value added between wages and profits; (iii) a productivity component. The component reflecting firm ownership accounts for 10 percentage points of the wage gap: hence the family firm discount is quite sizeable. In comparison, the productivity shortfall of family firms relative to non-family ones is small: it only accounts for 2 additional points. Instead, the rent-sharing component partly offsets the first two components: workers in family firms appear to have greater bargaining power than those in non-family firms in the wage-setting process, which accounts for a 4 percentage point reduction in the wage differential.

We then explore the potential sources of the family firm discount, considering two possible explanations. The first is that the discount may reflect compensating differentials: family firms may provide superior non-monetary amenities, such as a less confrontational work environment (Mueller and Philippon, 2011; Kang and Kim, 2020) or may be more likely to uphold "implicit contracts" (Shleifer and Summers, 1988), resulting in greater job stability. Rather than focusing on a specific amenity, we adopt the revealed-preference approach developed by Sorkin (2018), that is, we measure the systematic utility provided to workers by examining their voluntary transitions. We do not find support for the compensating differential story: family firms appear to provide *lower* utility to workers,

even after controlling for the fact that they pay lower wages.

We then turn to an alternative potential explanation, which we label the "glass ceiling" hypothesis. Our conjecture is informed by a stark pattern in the data, namely the fact that the family firm discount is larger for higher percentiles of the within-firm earnings distribution. In other words, top earners in family firms earn substantially less than top earners in non-family firms, while the wage gap is much smaller for median wages or those at the bottom of the within-firm earnings distribution. This suggests that family and non-family firms differ greatly in the career prospects that they can offer to their employees.

Consistent with this hypothesis, family firms have a substantially lower fraction of managers and middle managers, and are less likely to promote their own workers. Not only are their promotions less frequent, but their "returns to promotion" are also lower: the same type of promotion (from white collar to middle manager or from middle manager to manager) leads to lower gains in earnings for workers employed in family firms, controlling for a host of firm and worker characteristics.

Our preferred explanation for this pattern is that family firms feature high "benefits of control": as families do not want to relinquish control to outsiders, they are reluctant to share technical know-how and operational knowledge with other workers. We test an implication of this theory by analyzing firm performance in the wake of CEO deaths. Our hypothesis is that the loss of the top manager should be extremely disruptive in family firms, where decision making is extremely concentrated. Conversely, in non-family firms, owing to a higher degree of delegation, more workers are ready to "step in" and take strategic decisions in a wake of the deceased CEO. Consistent with our prediction, productivity drops more in family than in non-family firms following the death of the CEO. Moreover, CEO deaths in family firms have significantly worse consequences for workers' wages, earnings, and employment outcomes. We find no effect in family firms whose CEOs do not belong to the controlling family. This dovetails with the idea that

in these firms family owners already surrendered control, so that CEOs are more easily replaced upon their demise.

In the last part of the paper, we rationalize our findings through the lens of a simple model. In our framework, workers can learn new skills on the job by exerting effort. Such effort increases the probability of acquiring skills but is unobservable, so that workers earn information rents. Firm owners promote a fraction of the workers who were successful in acquiring managerial skills. Promoted workers (the "managers") receive higher compensation and produce additional output. We assume that, even though promotions of skilled workers are profit-maximizing, they result in a loss of private benefits of control, which owners trade off against the additional profits they can generate. As family owners place a greater value on the benefits of control, they are less likely to promote their employees, who therefore have weaker incentives to invest in human capital. As a result, family firm employees earn lower average wages and are less likely to receive managerial bonuses: their expected lifetime utility is lower that that of non-family firm employees. However, the more attractive career prospects offered by non-family firms lead them to receive more job applications relative to the available vacancies than family firms do. Hence applicants to jobs in non-family firms face greater risk of unemployment than family firm applicants. This compensates for the greater utility of careers in the former than in the latter, and ensures that in equilibrium applying to either delivers the same expected utility. The model's predictions match many of our empirical results, namely, family firms being characterized by fewer managers and lower productivity, wages, returns to promotion, and utility for their employees.

The paper is organized as follows. Section 2 presents a brief review of the relevant literature. Section 3 describes the data. Section 4 presents our key empirical results. Section 5 describes our theoretical model. Section 6 makes some concluding remarks.

2 Literature

A key difference between family and non-family firms is the fact that family owners typically place a great value on retaining control over the firm. As highlighted by Burkart et al. (2003), there are three possible benefits from family control over a company: first, nonpecuniary benefits such as the pleasure from the firm bearing the family name; second, reputational benefits, such as product quality associated with the family or political connections; third, the ability to prevent expropriation by non-family professional managers. The latter may be more skilled at running the company than family members, but may extract resources from it to their own benefit, unless shareholders are adequately protected. As a result, families may delegate control to professional managers only in jurisdictions with sufficiently high shareholder protection, where they may even be willing to sell their controlling stake. These predictions are broadly in line with empirical evidence, as firms in countries with low shareholders protection tend to have more dispersed ownership (LaPorta et al., 1998; Faccio and Lang, 2002).

In principle, private benefits accruing to family owners need not be value-decreasing: the desire to preserve the reputation associated with the family name may motivate a family manager to exert more effort than a non-family manager would; however, dynastic transmission of control may place incapable heirs at the helm of family firms. Anderson and Reeb (2003) provide the first empirical investigation of the relationship between firm performance and family ownership, and find that US family firms belonging to the S&P 500 are smaller but perform better than non-family firms. Sraer and Thesmar (2007) also find family firms to feature higher profitability than non-family ones in French data. These results contrast however with those by Bennedsen et al. (2007) and Pérez-González (2006), who show that family firms tend to exhibit a worse performance in terms of operating profitability and market-to-book ratios after a succession when they appoint a family CEO relative to when they choose an external manager, consistent with wasteful nepotism in family firms. These results dovetail with family ownership being typically

associated with *worse* managerial practices: Bloom and Van Reenen (2007), Bloom and Van Reenen (2010), Bandiera et al. (2017), and Bloom et al. (2012) find that family firms tend to be poorly managed.

While poor management should lead to low productivity and low wages, some studies suggest that family firms may feature better labor relations and more risk sharing with workers, and thus provide valuable nonmonetary amenities to their employees. In implicit contract models along the lines of Baily (1974), Azariadis (1975), and Holmstrom (1983), the profit maximizing strategy for firms consists in absorbing demand shocks and grant risk-averse workers a fixed salary. Under more restrictive conditions, firms may provide workers with insurance against unemployment even in the presence of negative demand shocks. Commitments to honor these implicit contracts may be especially credible when firms are controlled by families, as they have their reputation and prestige at stake (Shleifer and Summers, 1988). Consistent with this hypothesis, family firms appear to pay lower wages and provide more "insurance" to workers vis-à-vis demand shocks (Sraer and Thesmar, 2007; Bassanini et al. (2013); Ellul et al., 2017). Moreover, Mueller and Philippon (2011) show that family ownership is more prevalent in countries in which labor relations are hostile, suggesting that this type of ownership structure may be especially beneficial in these contexts. In addition, Kang and Kim (2020) show that family firms invest more in employee relations than non-family firms. This evidence suggest that family firm employees receive amenities that compensate for their comparatively low wages.

3 Data

3.1 Data Sources

This paper relies on three main data sources. First, we obtain on-site access to social security records of all the workers employed in the Italian private sector through the *Visitinps* project. This matched employer-employee dataset has been made available to researchers

selected through a refereed call for projects issued by the Italian National Institute for Social Security (INPS). The database has detailed information regarding earnings, type of contract (open ended or temporary), part-time status, etc. It also presents a basic occupational classification, distinguishing between white collar workers, middle management (*quadri*), and managers (*dirigenti*). Finally, for individuals who sign a new contract after 2010, we also know their educational level.

Information on firms' ownership and management comes from our second data source, which we refer to as *Cerved-Infocamere*.[2] It is a digitized business registry, comprising three datasets. The first one, *Anagrafica*, includes the universe Italian limited liability firms for the period 2003–2019, with basic information, such as date of creation, date of cessation (if any), and address. It includes 2.47 million individual (unincorporated) businesses. The second file, *Soci* has information on firms' owners, which could be either individuals or other firms. For each owner, the dataset includes the name and tax ID, share owned, and address. As the tax ID in Italy is a function of name, city of birth, date of birth, and gender, we can recover this information as well. Coverage in the ownership file is not complete but increases over time, ranging from 41% in 2003 to 83% in 2019. The third file, *Esponenti*, lists the members of companies' board of directors and their top executives, and contains individuals' tax IDs, as well as their addresses and roles, as they appear in the companies' financial statements. Coverage is nearly complete.

The third data source is *Cerved*, a firm-level dataset with accounting information for all the Italian nonfinancial limited liability companies. Importantly, we can easily link all three datasets as both individuals and companies are recorded with their respective tax ID in the three data sources.

[2]Infocamere is the IT company owned by the Italian Chamber of Commerce. We purchased the data through Cerved, a private company that imported and made available to us the dataset in a user-friendly format.

3.2 Identifying Family Firms

In most of our analysis, we follow an ownership-based definition of family firm status. For this reason, in our first step we seek to identify the ultimate owners of a firm. This distinction becomes relevant whenever a firm is fully or partially owned by another firm, in which case we track the owning firm's shareholders. We repeat this procedure going up along the control chain. Each ultimate shareholder is then assigned a share computed using the so-called weakest link principle. Intuitively, if person A owns 10% of company B, which in turns owns 5% of company C, A is assumed to control 5% of company C.[3]

Neither *Cerved-Infocamere* nor *INPS* record family linkages; as a result, we need to make some assumptions to identify individuals who are likely to be relatives. First, we assume that, within the shareholders of a firm, individuals sharing the same last name belong to the same family. Second, we assume that, again within the shareholders of a firm, two individuals of different gender and last name, but sharing the same address, are partners.

Throughout the paper, we define a firm as being family owned if we can identify an individual, or a group of individuals that we classify as related, who own more than 50% of the shares of the company. Given that the vast majority of our sample is comprised by small, non-publicly listed firms, we chose a cutoff higher than those employed in most of the literature, which tends to focus on large, publicly listed firms (Volpin, 2002; Sraer and Thesmar, 2007; Ellul et al., 2010; Ellul et al., 2017).

As our classification is based on several assumptions, there are two main risks of misclassification. First, two individuals sharing the same address or last name may not actually be related. This type of error will tend to overstate the proportion of family firms. Second, when the controlling stake of a company is held by a firm for which we do not have ownership information, we make the assumption that the latter is widely held, and classify

[3]This is the standard approach followed in the governance literature (see for example La Porta et al., 1999, Faccio and Lang, 2002, Claessens et al., 2000, among others). See also Edwards et al. (2004) for a discussion.

the former as non-family firm. However, the controlling firm may be actually controlled by a single individual or firm, for which we happen not to have ownership information in our dataset, for example because it is incorporated in a foreign country. This type of error will tend to underestimate the proportion of family firms.

Relative to the previous literature, which tends to focus on small and selected samples of large firms, a key advantage of our analysis is that our dataset is an order of magnitude larger and covers a significant fraction of the private sector of a developed economy, with substantial benefits in terms of statistical power and external validity of the results. These advantages come with the cost of a higher risk of misclassification, which are unavoidable in our case, given that it is unfeasible to manually collect information on each individual firm in our sample. Fortunately, we can get a sense of the severity of the misclassification. In 2018, *Cerved* developed an in-house classification of firm ownership for a sample of large and medium-sized firms. They strive to accurately identify family relationships by consulting the real estate registry and linking individuals if they share real estate properties. The overlap of the *Cerved* sample with ours covers 68,367 firms and we find that the two classifications coincide for 93% of the observations. We also conjecture that this percentage is likely a lower bound to the true accuracy of our classification algorithm, given that this sample over-represents relatively large firms, which tend to have more complex and diversified ownership structures, and where our procedure might be more likely to deliver incorrect answers.

3.3 Descriptive statistics

Table 1 presents descriptive statistics for the data used in this paper. Our sample includes about 900,000 unique firms and 4.7 million firm-years, two thirds of which are family firms. For the worker-level analysis, our starting sample will include over 19 million unique workers and about 91 million worker-years. As family firms are smaller on average (14 employees as opposed to 31.8 for non-family firms), the distribution of workers between

the two groups of firms is roughly even.

Even though our sample is much larger than those used in previous studies, the data patterns displayed in Table 1 are broadly in line with the existing literature. Family firms are much less productive than non-family firms, as measured by operating value added per worker (45.9 versus 63.3 thousand dollars per worker), consistent with Bandiera et al. (2017). However, family firms' profitability, as measured by return-on-assets, is higher, though not by much (4.42 versus 4.37 percentage points). Although the gap is fairly small, this finding is qualitatively in line with Anderson and Reeb (2003) and Sraer and Thesmar (2007). Leverage is also fairly similar between the two groups of firms.[4]

How can family firms roughly match non-family firms' profitability despite their lower productivity? Part of the answer lies in their ability to pay lower wages. Weekly wages are 64 euros lower in family firms, corresponding to 14 log-points, a "family firm discount" that we analyze in detail in Section 4.2. The gap in the entry wage is of comparable magnitude, suggesting that the family firm discount emerges immediately in a worker's career. But the gap tends to widen over a worker's career, as family firms also display a slightly lower wage growth rate.

Interestingly, family firms are also characterized by a lower degree of within-firm wage inequality, which is apparent by looking either at the log-difference between the highest and the median wage, or at the log-difference between wages at the 90th and10th percentiles. The gaps are 21 and 10 log-points, respectively. Part of this gap is due to organizational differences between the two types of firms: non-family firms have two and a half more workers in managerial positions (1% versus 0.4%) and four times more workers in middle management positions (4% versus 1%).[5] Age and gender do not seem to play a role in

[4]As shareholders of a family firm tend to have a high portion of their wealth invested in the firm, they are generally thought as leaning towards more conservative financial choices, and so to less leverage. However, a reluctance to raise equity to prevent a dilution of their stake can lead to a *higher* leverage (and a lower size). In our data, these two channels appear to approximately offset each other.

[5]Unfortunately, Italian social security data lack a detailed occupational classification. As we discuss in Section 4.4 we know, however, whether a worker is classified as *dirigente* (manager) or *quadro* (a position that, in the Italian firms' hierarchy, lies between those of manager and rank-and-file worker, and to whom we will refer to as "middle manager").

explaining the gap, as the two types of firms do not appear to substantially differ in these dimensions. Family firms have fewer workers with full time or open ended contracts but, again, the difference is fairly small.

4 Empirical Evidence

4.1 The Origin of the Family Firm Wage Discount

In this section we outline a simple decomposition of the differences in the wages paid by family and non-family firms, so as to account for their sources. Suppose that the logarithm of the wage of worker i in period t is set according to:

$$w_{it} = \alpha_i + \psi_{j(i,t)} + X'_{it}\beta + r_{it}, \tag{1}$$

where α_i captures the portable component of productivity of worker i; $j(i,t)$ identifies the firm hiring worker i in period t and ψ_j captures the firm-level component of the wage, i.e., the premium or discount received by the employees of firm j. X_{it} represents a set of observable characteristics of worker i. The error term r_{it} captures a drift in worker productivity, random match effects and measurement error. This error term is assumed to be uncorrelated with the mobility decision of workers (see Card et al., 2013 for a discussion).

To simplify the discussion, in what follows we set $\beta = 0$ (that is, we ignore the determinant of wages included in the vector X_{it}). Let $f(j) \in \{F, NF\}$ denote an indicator for whether firm j is a family (F) or a non-family firm (NF). The observed gap in wages

between workers in family and non-family firms can be decomposed as follows:

$$\begin{aligned}\Delta w &\equiv E[w_{it}|f(j(i,t)) = F] - E[w_{it}|f(j(i,t)) = NF] \\ &= \underbrace{E[\alpha_i|f(j(i,t)) = F] - E[\alpha_i|f(j(i,t)) = NF]}_{\text{sorting component}} \\ &+ \underbrace{E[\psi_{j(i,t)}|f(j(i,t)) = F] - E[\psi_{j(i,t)}|f(j(i,t)) = NF]}_{\text{firm wage component}}.\end{aligned} \tag{2}$$

The first term of this decomposition is the "sorting component". It measures how much of the gap in wages paid by family and non-family firms is associated with differences in the unobserved abilities of workers hired by the two types of firms. This term will be negative if workers with higher productivity systematically sort into non-family firms, and positive otherwise. The second term is the firm-wage component, which captures differences in wage policies between family and non-family firms that equally affect all workers in a firm.

Table 2 reports this first decomposition. The first row shows that family firms pay substantially lower wages relative to their non-family counterparts. Specifically, we find a 0.14 log-points difference. This result is in line with evidence reported by Sraer and Thesmar (2007) and Bassanini et al. (2013) for France and with cross-country evidence by Ellul et al. (2017). We further include cubic polynomials in age and years of experience in the labor market, as well as year fixed effects (the vector X_{it}). Adding these controls slightly increases the wage gap to 16 log-points (see the "adjusted wage" in the second row of Table 2).

We go one step further than previous studies, as we decompose this gap in a worker and a firm component. To do that, we follow Abowd et al. (1999) (henceforth "AKM") and estimate equation (1), regressing the logarithm of worker's yearly wage on firm and worker fixed effects. Estimating this model is informative regarding the relative contributions of the sorting component and the firm wage component in explaining the wage gap between family and non-family firms.

As the third and fourth row of Table 2 show, we find that the two components are both equal to -0.08. Hence, non-family firms appear to hire "high quality" workers. However, this channel explains only half of the observed gap. Hence, the same worker, when moving from a family to a non-family firm, receives on average an 8% boost in wage.

Following Card et al. (2013), we can further decompose the firm-wage component using the following simple rent-sharing model:

$$\psi_j = \pi_{f(j)} NS_j + \theta_{f(j)} + \nu_j. \tag{3}$$

The coefficients $\{\pi_F, \pi_{NF}\}$ are rent-sharing coefficients, which measure how much an extra euro of net surplus (NS_j) available at firm j passes onto wages for workers in family and non-family firms, respectively. Following Card et al. (2013), the surplus is proxied by the logarithm of value added per worker. $\theta_{f(j)}$ represents the average effect on wages of being employed by a family or a non-family firm, once we account for productivity and rent-sharing differences across these type of firms. Finally, we assume that $E[\nu_j | f(j)] = 0$.

Specifically, we apply a traditional Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973) to the firm component as follows:

$$\begin{aligned} &E[\psi_{j(i,t)} | f(j(i,t)) = F] - E[\psi_{j(i,t)} | f(j(i,t)) = NF] = \\ &= \pi_F E[NS_j | f(j) = F] - \pi_F E[NS_j | f(j) = NF] + \Delta\theta \\ &= \underbrace{(\pi_F - \pi_{NF}) E[NS_j | f(j) = F]}_{\text{bargaining component}} \\ &+ \underbrace{\pi_{NF}\{E[NS_j | f(j) = F] - E[NS_j | f(j) = NF]\}}_{\text{productivity component}} + \underbrace{\Delta\theta}_{\text{unexplained component}} \end{aligned} \tag{4}$$

The bargaining component captures the idea that rent-sharing might be different between family and non-family firms. Table 2 shows that family firms engage in *more* rent sharing than non-family firms. We find that the bargaining components are equal to 0.51 and 0.47 log points in family and non-family firms, respectively. Thus, the bargaining channel

benefits more those who work in family firms. In other words, accounting for heterogeneity in rent sharing ends up *increasing* the unexplained portion of the wage gap by 4 log-points.

An alternative approach to test whether firms differ in the degree to which they engage in rent sharing consists in estimating a simple regression of the logarithm of the wage on the logarithm of value added per worker (as well as a constant), separately for family and non-family firms. We display these "rent-sharing" coefficients in the second-to-last row of Table 2. As expected, the rent-sharing coefficient is higher in family than in non-family firms, suggesting that a one percent increase in value added per worker leads to a 0.14 and 0.13 percent wage increase for workers employed in family and non-family firms, respectively. The difference between the two coefficients is very precisely estimated.

The second element of the wage gap is a productivity component. It captures to what extent wages are higher in non-family firms because these have higher productivity than family firms. We find that it only explains a small part of the observed wage gap, leading to a 2-log-point wage discount for family firms. Indeed, as shown in the last row of Table 2, family firms have substantially lower productivity, resulting in a 17-log point gap (3.56 versus 3.73 for family and non-family firms, respectively).

Finally, $\Delta\theta$ is the average treatment effect of being employed by a family firm relative to a non-family firm, netting out potential differences in net surplus (that is, productivity) and rent sharing between family and non-family firms. Adding to the 8-log-point difference in firm fixed effects the 4 log-points from the rent sharing channel and subtracting the 2 log-points from the productivity channel, leads to an unexplained component of -0.10 log points. Hence, a large part of the family wage discount cannot be explained by traditional sources of heterogeneity in wage setting, such as productivity or bargaining power. The main objective of our analysis is to uncover the sources of this family wage discount.

4.2 Wages, Productivity and Other Firm Characteristics

Figure 1 presents a graphical representation of how the firm-specific component of wages varies as a function of labor productivity. Productivity, as measured by the logarithm of value added per worker, represents the surplus available to be shared between the workers of a firm and the investors holding claims to its cash flow. As discussed in the previous section, family firms feature lower productivity than non-family firms, so that it is important to examine the firm-specific component of wages holding productivity constant.

We proceed as follows. For each firm, we compute the logarithm of value added per worker and average it over the whole sample period. We do this separately for family and non-family firms, and compute percentiles of value added per worker over the entire productivity distribution.[6]

Figure 1 shows that family firms pay consistently lower wages for virtually any level of the productivity distribution, but the gap tends to widen at high percentiles of the productivity distribution. Across percentiles, the difference in the firm component of wages is, on average, equal to -0.059 log points, but becomes -0.087 in the top five percentiles of the productivity distribution.

Table 3 presents a regression analysis of the factors that may account for the family firm discount. Column 1 presents a simple regression of the AKM firm fixed effect on a family firm dummy. The coefficient is 0.08, as already reported in Table 2. It is not only very precisely estimated but also large in magnitude, being 40% of the relevant standard deviation. In column 2 we control for the logarithm of value added per worker; again as reported in the previous section, the coefficient drops to 0.06, i.e., by 2 log-points.

In the specification of column 3 we control for industry fixed effects to take into account the possibility that family firms are concentrated in low-wage sectors. In column 4 we further include province fixed effects, as family firms may be disproportionately present in

[6]Note that "firm" should be interpreted as a combination between a given firm and its ownership status, that is, if a firm changes its status from, say, family to non-family firm, we compute two separate averages, one for the years in which it is a family firm and one for the years in which it is a non-family firm.

non-urban areas with low cost of living. Finally, the specification in column 5 also includes local labor market fixed effects, that is, a vector of industry-province dummies. Adding these controls has little effect on our estimated gap, which is now 4 log-points.

In Figure A1 of the Appendix, we further address the possibility that the family firm discount is accounted for by workers' geographical preferences. For example, working in a non-family firm may entail a higher compensation but require a longer commuting time, as non-family firms are fewer than family firms. We also regress the firm fixed effect on the logarithm of productivity and a combination of city of birth, city of residence, and city of workplace location fixed effects. Even the most saturated specification, which includes separate dummies for every possible combination of city of birth, residence, and work, cannot explain more than one log-point of the wage difference between family and non-family firms.

Thus, the presence of a family firm discount appears to be a robust feature of our data. In the next sections, we discuss potential explanations for this finding.

4.3 Testing for Compensating Differentials

The evidence so far shows that, even after controlling for sorting and productivity, family firms still pay substantially less than non-family firms. In this section we test for a "labor-supply" explanation. Specifically, workers may expect to receive more "non-wage amenities" from family firms. For example, family firms might provide greater employment stability (Sraer and Thesmar, 2007; Ellul et al., 2017). In addition, they may invest more in employee relations, reducing labor-related conflicts as a result (Mueller and Philippon, 2011; Kang and Kim, 2020). In equilibrium, these possible advantages of working in a family firm may compensate for lower wages.

As by definition these amenities are non-monetary, observing them and estimating their importance poses a challenge. To address it, we adopt the "top-down" approach developed by Sorkin (2018) to measure the utility that a typical worker draws from being

employed at a given firm. Specifically, we assume that the utility that individual i obtains for being employed at firm j is

$$U_{ij} = v_j + e_{ij}, \tag{5}$$

where e_{ij} is Type-1 distributed. Sorkin (2018) shows that the utility typically offered by firm j to its employees, i.e., v_j, can be identified from the following recursive equation:

$$\exp(v_j) = \sum_{\ell \in \mathcal{B}_j} \omega_{j,\ell} \exp(v_\ell) \qquad \text{for} j = 1, \ldots, J, \tag{6}$$

where $\omega_{j,\ell}$ is number of workers who voluntarily move from employer ℓ to employer j, scaled by the number of workers who voluntarily leave employer j. Intuitively, "good" employers, that is, employers providing higher utility to their employees attract many workers from other firms and experience few quits by their employees. If family firms were able to compensate their lower monetary compensation with non-monetary amenities, the two types of firms should give comparable levels of utility.

Figure 2 shows the utility levels typically offered by firms to their employees vis-à-vis the percentile of the logarithm of value added per worker. For a given productivity level, non-family firms' utility systematically lies above that of family firms. Hence, this evidence provides little support to the hypothesis that there are additional benefits from working in family firms over and above the wage.

In Table 4 we present the estimates of regressions of the utility that firms typically offer to their employees on a family firm dummy, controlling for other variables gradually added to the model. Column 1 shows that, the utility typically offered by family firms is significantly lower than that offered by non-family firms: the shortfall is estimated to equal 0.70, which is about half of the sample standard deviation of the dependent variable (1.45). In column 2 we test whether family firms provide lower utility simply because they pay less. After controlling for the average firm-level wage, we find that family firms still provide lower systematic utility; the coefficient on the family firm dummy is slightly smaller but

remains negative and significant, and equal to −0.58. A somewhat more important role seems to be played by the firm's industry. Once we control for industry fixed effects, the coefficient's magnitude is reduced to −0.30. Further controlling for province fixed effects or for local labor market fixed effects brings the coefficient down to −0.23. Hence, there is some evidence that family firms may cluster in industries providing lower systematic utility to workers; however, they appear to be less preferred employers even after accounting for sector and average compensation. As a result, we conclude that labor supply factors are unlikely to explain the family firm wage discount we observe in our data; if anything, family firms appear to provide *lower* non-monetary amenities.

In Figure 3, we analyze some *observable* amenities that we can measure in the INPS data. Specifically, we examine whether workers in family and non-family firms differ with respect to the use of part-time and temporary contracts (Panels A and B, respectively). In Panel C we ask whether they feature differences in their reduced-contribution weeks (a proxy for the propensity to take days off following injuries or illnesses). In these three panels we plot "AKM" firm fixed effects against the percentiles of firm productivity. Finally, in Panel D we plot the average churn rate (a measure of employment stability), again against the percentiles of firm productivity. In all four cases, we do not find stark differences between the two sets of firms. Hence, even without relying on the revealed-preference approach proposed by Sorkin (2018) and focusing on observable amenities, one cannot detect a clear advantage from working in family firms, which might offset the lower compensation they offer.

4.4 Family Firms and the "Glass Ceiling Effect"

Having established that, even after controlling for sorting, family firms consistently pay lower salaries, it is worth asking whether *all* workers are equally subject to a family firm discount, or there is heterogeneity in the family wage discount across them. In particular, does the discount affects comparatively more lower-pay or high-pay employees? This

question can be restated as one concerning within-firm wage inequality, namely: do family firms have a more compressed wage distribution than non-family ones?

We present three pieces of evidence that, taken together, strongly suggest that workers in family firms are subject to a "glass ceiling", leading to a compression of the right tail of the wage distribution. We start by plotting, in Figure 4, the within-firm log-wage distribution by log-value added per worker, for different percentiles of the wage. We show the 5^{th}, 10^{th}, 25^{th}, 75^{th}, 90^{th}, and 95^{th} percentiles of the within-firm distribution of the logarithm of wages in Panels A, B, C, D, E, and F, respectively. The log-wage percentiles for family and non-family firms are fairly close for the 5^{th} and 10^{th} percentile, and a clearly visible gap starts emerging for the 25^{th} percentile. The gap becomes more apparent, and economically large, once we focus on the 75^{th}, 90^{th}, and 95^{th} percentiles, especially for high values of log-value added per worker, hence for the most productive firms. Although this evidence should be taken as merely descriptive, as it does not account for sorting, it shows that the gap we observe in average wages can be explained primarily by the differences in compensation of workers located at the high end of the within-firm wage distribution.

To further investigate this point, we look at whether and how family and non-family firms differ in their promotion practices. In the *Visitinps* data workers are classified as white collar, *quadri* (an intermediate figure between white collars and managers), and *dirigenti* (managers). We consider all the workers employed with the same firm for two consecutive years.

For each firm and year, Panels A and B of Figure 5 display the fraction of promoted workers, either from white collar to middle management, or from middle management to manager, by percentile of value added per worker. Panel A shows that, for about three fourths of the labor productivity distribution, promotions to middle management are relatively rare and fairly similar between family and non-family firms. For more productive firms, however, they become more frequent. Moreover, a stark gap between promotion rates appears for more productive firms: zooming on the top decile of the productivity

distribution, the propensity to promote in non-family firms is two-to-three times higher than in family firms. For example, in the highest productivity percentile, the fraction of workers promoted every year is close to 0.6% in non-family firms and 0.2% in family firms.

Results are qualitatively similar when we look at Panel B, where we plot the average promotion rate from middle management to management. The average promotion rates are, not surprisingly, lower, but the difference in the promotion rates is even higher: in the top percentile of the productivity distribution, the average promotion rates for family and non-family firms are 0.02% and 0.1%, respectively. In the sample of firms subject to CEO death event, we then distinguish between family and non-family firms, defined as treatment and control firms, respectively. For every treated firm, we select a control firm in the same industry (defined using the ATECO 2-digit classification), further matching on the logarithm of size and the logarithm of value added per worker (using a 0.1 caliper), measured at the end of the year prior to the event.

While non-family firms promote more intensively to managerial positions, this might be due to a higher turnover of workers in such positions, due to firings or quits. In this case, it would be questionable to interpret the higher promotion rate as an indication of better career opportunities. Panels C and D of Figure 5 plot the fraction of middle and top managers over the total workforce. They mirror those of promotions: at higher levels of productivity, non-family firms have a larger share of workers in managerial positions. Hence, family firms appear to provide significantly lower career opportunities to their workers.

In principle, family firms might use promotions more selectively but reward promoted employees more generously, and the greater boost in compensation might compensate for the lower chance of promotion from the standpoint of employees. So our third piece of evidence compares the returns to promotion of workers employed in family and non-family firms via a matching approach. Intuitively, we select workers who have similar characteristics, including the type of promotion, but who differ in the type of employer.

Every worker promoted in a family firm is paired with a "twin" worker in a non-family firm, matching on the logarithm of firm size, the logarithm of value added per worker, and wage growth, age, (exact) promotion type, industry, contract type, full-time status, and gender. Upon imposing a two-year tenure restriction and following workers over a $(-3, +3)$-year window, we are left with 29,632 promotions from white collar to middle management and 1,416 promotions from middle management to manager.

We then estimate the following model:

$$Y_{it} = \sum_{k=-3}^{3} \alpha_k D_{it}^k + \sum_{k=-3}^{3} \beta_k D_{it}^k \times FF_i + \gamma_i + \delta_t + \varepsilon_{it}. \tag{7}$$

Here $D^k \equiv \mathbb{1}(t = t_i^* + k)$, where t_i^* is the event year for worker i and Y is an outcome of interest, that is, either the logarithm of earnings, a dummy equal to one if the worker is employed, or raw earnings. FF is a family firm dummy, γ and δ are vectors of worker and year fixed effects, respectively. We are interested in the estimated βs, which track the evolution of the dependent variable prior and after the promotion event. β_{-1} is normalized to 0; hence, our baseline year is the year prior to the event.

The results are plotted in Figure 6. Panel A shows that promotions to middle managers result in a 1.1% lower increase in earnings in family firms three year after the event: as to the level of earnings, workers promoted in family firms earn 600 euros less than their matched counterparts three years after the event. The results for employment are less precise and the estimated coefficients are relatively small, although they point out to a lower retention for workers promoted in family firms.

Panel B, which focuses on promotions from middle management to manager, reveals similar results, although with lower precision, due to the smaller sample. Workers promoted in a family firm earn 2% less three year after the promotion than those promoted in non-family firms, although this effect is only marginally statistically significant. Although in this case there is no effect on the employment rate, the difference in the earnings trajec-

tory is economically large: three years after event, a 4,000 euros-difference in the earnings emerges, relative to the baseline year.

4.5 Evidence from CEO Deaths

A potential explanation for the evidence presented in the previous sections is that members of the controlling family firms wish to retain most, if not all, managerial positions in the firms that they control, and thus retain most of the firm's technical and organizational know-how within the family circle. In other words, they are reluctant to delegate key decisions to non-family members.

This narrative can explain why family firms tend to promote fewer workers to managerial positions, as well as reward them less upon promotion. It is also consistent with the family wage discount being especially pronounced for top earners: it likely reflects family firms' unwillingness to compete keenly to attract or retain non-family employees to decision-making roles at the top of the firm's hierarchy.

An implication of this hypothesis is that losing the top decision maker should have different effects depending on whether the firm hit by this loss is a family firm or not. In a family firm, most of the technical and organizational knowledge should be concentrated in a few members of the controlling family; as a result, losing the most important member of the management team – someone who is difficult or impossible to replace – is likely to be very disruptive. Conversely, in non-family firms, the ability to take key managerial decisions is shared by several incumbents or, failing that, similar managers can be hired externally. As a result, in these firms even losing the top manager should be less detrimental to performance, as there is a larger pool of workers who should be able to "step in" and effectively replace the deceased CEO.

To test the implication of this hypothesis, we make use of the *Esponenti* file in our data, which lists the top executives and the members of the board of directors of the companies in our sample. For each firm, we identify the top decision maker, namely the chief executive

officer (*amministratore delegato*) or, if absent, the chairman of the board (*presidente*). For brevity, in what follows both will be referred to as CEOs. INPS records the death year of all the individuals who, at any point in time, paid social security contributions. We can match these records with data on CEO deaths, so as to build a panel of firms affected by a CEO death event.

This matching procedure leaves us with a sample of 1,268 events, half of which involve the death of a family CEO. We then estimate the following regression to investigate whether the loss of the firm's CEO affects the productivity of family firms differently from that of non-family firms:

$$\log(Productivity)_{it} = \sum_{k=-3}^{3} \alpha_k D_{it}^k + \sum_{k=-3}^{3} \beta_k D_{it}^k \times FF_i + \gamma_i + \delta_t + \varepsilon_{it}, \tag{8}$$

where the i and t subscripts identify firm and years, respectively, and γ and δ are firm and year fixed effects, respectively. As before, $D^k \equiv \mathbb{1}(t = t_i^* + k)$, where t_i^* is the CEO death year for firm i. As usual, productivity is measured as operating value added per worker. The coefficients of interest are the β_ks, with β_{-1} being normalized to zero.

The β coefficients obtained upon estimating equation 8, together with 95% confidence intervals, are displayed in Figure 7, Panel A.i. After the CEO deaths, productivity falls more in family firms than in non-family firms. While the effect is initially limited, two and three years after the event family firms experience productivity losses 7.6 and 5.7 percent higher than their non-family counterparts, respectively, although the coefficient β_3 is not very precisely estimated.

Our detailed board and shareholding data allow us to further refine the analysis. So far, family firms have been identified solely based solely on firm ownership, ignoring management characteristics. However, some family firms may not be "family managed", having hired an external manager to run their operations. In these firms the controlling family can be expected more inclined to delegate key decisions than in family-managed ones. For example, an incumbent managers can normally be replaced with other non-family man-

agers, without significant losses in the firm's technical know-how. In contrast, such losses can be expected in "family managed" firms, where the owner-manager's is likely harder to replace with an equally able family member. Hence, one can expect the differential effect of CEO deaths on productivity to be driven by family-managed firms.

Panel A.ii displays event-study evidence that only refers to family-managed firms, together with the corresponding matched control firms. The evidence is qualitatively similar to that of Panel A.i, but the point estimates of β_2 and β_3 are larger in magnitude. Conversely, Panel A.iii reveals no detectable differential effect of CEO death in non-family-managed firms. For completeness, Figure A2 in the Appendix displays the average productivity in the seven years surrounding the CEO death event. Both groups of firms experience substantial drops in productivity in the year of the CEO's death. However, these effects are much more persistent for family firms: even three years after the event, their productivity does not recover the level featured in the year prior to the CEO death. Conversely, non-family firms recover relatively quickly and, three years after the event, have a productivity level that is *higher* than that displayed prior to the CEO death.[7]

A potential explanation for this finding is that the drop in productivity is not driven by a loss in know-how but rather by the "emotional" impact that the loss of the owner-manager of a company can have on its workers. This explanation would be in line with the fact that results are driven by family-managed firms, where the departing CEO is a member of the family controlling the company.

To address this concern, Panel B presents a "placebo test" based on the same design as equation (8) but where the relevant event is the death of a director (*amministratore*), rather than that of the CEO. The intuition for this test is simple. The death of a family member should be an emotionally disruptive event independent of her role; however, we

[7]This improvement in productivity could stem from the new management's ability to eliminate or reduce agency problems created by the previous management: for instance, the incoming CEO may divest low-quality assets acquired by the previous CEO (Pan et al., 2016) or breaking entrenchment created by him/her (Taylor, 2010).

would expect a substantial loss in technical and operational know-how only for the loss of a family member with an executive role, namely the CEO. This is precisely what we find in Panel B: there is no detectable evidence of a differential productivity response for director death events, whether we focus on all firms, on family-managed firms, or on non-family-managed firms.

In Figure 8, we adopt the same design but now examine outcomes at the worker level. Again, we adopt a matching approach, pairing every worker in a family firm with a worker in a non-family firm, matching on the same variables used in Section 4.4. Importantly, this matching is performed within pairs of firms matched in the firm-level analysis. The final sample includes 8,306 workers, 5,802 of which are employed either in family-managed firms or in the corresponding matched firms. For brevity, and given the evidence on firm outcomes, we focus on the latter subsample.

Panel A shows that employees of family-managed firms experience larger wage losses after the CEO death, relative to employees of non-family firms. There is little effect in the short run but, three years after the event, their wages are 3.4 percent lower. Their employment probability is also lower (Panel B), but with the opposite timing: a large short-run effect, namely, a 2 percentage point lower employment probability for family-firm employees, but no significant difference three years after the CEO death.

Earnings of family firm workers are driven down by the lower employment probability in the short run, and by the lower wages in the long run. As a result, the negative effect of a family firm CEO death on the level of labor earnings is immediate and fairly stable, with coefficients implying losses between 650 and 950 euros (Panel C).

Finally, one can expect that, after the death of the CEO, firms may promote other employees to higher positions in the corporate ladder, as they may be asked to perform more managerial tasks to make up for the CEO's loss. However, these career advancement opportunities may differ in family and non-family firms. This is indeed the case, as shown in Panel D, where the dependent variable is a dummy equal to 1 if a worker is a manager

or middle manager after the death of the CEO. Family firms remain disproportionately more averse to promote their employees to managerial positions: 3 year after the CEO death, the probability of being either a manager or a middle manager is 0.65 percentage points lower for workers of family firms. Hence, the larger drop in productivity suffered by family firms following CEO deaths may be due, at least partly, to their reluctance to delegate strategically important decisions to new managers.

5 An Illustrative Model

The evidence presented so far can be broadly summarized as follows. First, careers in family firms are slower than in non-family ones, the frequency of promotions being significantly smaller in the former. Second, family firms pay lower average compensation to their employees, at each stage of their careers, even after accounting for sorting of workers into firms with different characteristics. Third, workers appear to prefer working in non-family firms, implying that family firms do not offer greater amenities to make up for the lower attractiveness of their careers. In what follows, we provide a simple theoretical framework to understand these empirical regularities.

In the model, firm owners determine their promotion policies by balancing the efficiency gains from promoting skilled workers against the personal advantages of retaining control over the firm. Specifically, in family-owned firms the owners' preference for retaining decision-making power is assumed to be greater than in non-family firms. Consequently, they are less likely to promote their employees, who therefore have weaker incentives to invest in human capital, leading them to accumulate less human capital and earn commensurately lower average wages and managerial bonuses. This leads to more job applications for positions in non-family firms relative to the available vacancies, increasing the probability of applicants being rejected and remaining unemployed. This higher probability of unemployment balances the greater expected utility of being employed in a non-family

firm: in equilibrium, applying to family and non-family firms delivers the same expected utility.

The economy is populated by identical risk-neutral workers and heterogeneous firms. There are N_{NF} non-family firms and N_F family firms, which differ in the weight they place on the benefits of corporate control. Each firm hires a mass 1 of workers, so that total employment is $N = N_{NF} + N_F$. The labor force L exceeds total employment N. The utility of the $L - N$ workers who remain unemployed is normalized to zero. Firm j's skilled employees produce y_j if they acquire skills on the job; they do so with probability e_w, which coincides with their effort, whose cost is $c_w e_w^2/2$. Effort is unobservable while, once acquired, skills can be observed. Each firm draws its productivity parameter y_j from a probability distribution $F(y)$, assumed to be the same for family and non-family firms. Unskilled workers produce a low output, which for simplicity we normalize to 0.

Each firm j chooses to promote a fraction ϕ_j of skilled workers to managerial positions. Promoted workers produce additional output δy_j only if they become skilled managers; this occurs with probability e_m, which coincides with their additional effort at a cost $c_m e_m^2/2$.[8] Firm j chooses its promotion rate ϕ_j by trading off its expected profits with the loss of private benefits $\beta_j \phi_j^2/2$ associated with promotion rate ϕ_j, where the parameter β_j captures the firm's "taste for control over promotions" and is assumed to be larger for family than for non-family firms ($\beta_F > \beta_{NF}$) and for simplicity constant within each of the two groups. The quadratic loss function captures the idea that incremental control losses generate increasing disutility for controlling shareholders.

The timeline of the model, illustrated in Figure 9, comprises four stages:

- **Matching and hiring**. At $t = 0$ each worker files a single job application for a randomly chosen family or non-family firm, taking into account the expected lifetime income respectively offered by them. Firms hire workers by randomly picking them from their applicant pool, with unsuccessful applicants remaining unemployed. Upon

[8] To keep our notation simple, we do not index by j the effort chosen by the employees and the managers of firm j.

hiring workers, each firm j offers them a (possibly state-contingent) wage contract.

- **On-the-job learning.** At $t = 1$ the employees of firm j choose effort e_w, anticipating the promotion probability ϕ_j of their employer j and the bonus contract that it will offer at stage $t = 2$.
- **Workers' promotion.** At $t = 2$, each firm j promotes a fraction ϕ_j of skilled workers to managerial positions and offers them a (possibly state-contingent) bonus contract. Managers can generate additional output δy_j only if they acquire managerial skills *on top of* worker skills. They do so by exerting unobservable effort e_m.
- **Production and compensation.** At $t = 3$, skilled workers and skilled managers respectively produce output y_j and $(1 + \delta)y_j$, and firms pay the agreed wages and bonuses.

At $t = 0$ firms are assumed to be unable to commit to a specific promotion probability nor to offering a certain bonus contract to promoted workers at $t = 2$. Hence, the model is solved by backward induction: at each stage, workers and firms act optimally, taking into account their time-consistent policies. In what follows, we present the choice problems of workers and firms at each stage and illustrate the main results that obtain in equilibrium. Detailed derivations are presented in Appendix A.2:

Result 1. Family firms pay lower wages than non-family firms and the difference in wages increases with firm productivity. The optimal wage paid by firm j is

$$w_j^* = \frac{1}{2}\left[y_j + \frac{\phi_j^*}{2c_m}\left(\frac{\delta y_j}{2}\right)^2\right] = \frac{1}{2}\left[y_j + \frac{1}{2\beta_j c_m^2}\left(\frac{\delta y_j}{2}\right)^4\right]. \tag{9}$$

The first expression indicates that the wage is not only increasing in the firm productivity parameter (y_j) but also in the probability of promotion (ϕ^*) and in workers' incremental future productivity as managers (δy_j), as all of these raise workers' optimal effort. Upon substituting for the optimal promotion probability, one sees that family firms, which fea-

ture a higher β_j, pay lower wages, because they are less likely to exert effort due to their lower probability of promotion (as we show next). This is consistent with the results shown in Table 2 and in Figure 1.

Result 2. Family firms promote fewer workers than non-family firms, and the gap in promotion rates increases with firm productivity. The optimal promotion probability chosen by firm j is

$$\phi_j^* = \min\left[\frac{1}{\beta_j c_m}\left(\frac{\delta y_j}{2}\right)^2, 1\right], \tag{10}$$

so that family firms, which feature higher taste for private benefits β_j, promote fewer employees to managerial positions. Since in expression (10) the parameter β_j interacts with firm j's productivity parameter y_j, the gap in promotion rates is greater for more productive firms. This is in line with the evidence of Figure 5, which describes the promotion policies of family and non-family firms for different classes of value added per worker.[9]

Result 3. Promotions in family firms lead to a lower increase in compensation. To prove this result, we make the additional assumption that y_j follows a lognormal distribution. First, notice that the expected bonus b for a worker who is promoted, conditional on y_j, is equal for family and non-family firms, and given by:

$$\mathbb{E}\left[b|y_j\right] = x^* e_m^* = \frac{\delta y_j}{2}\frac{\delta y_j}{2c_m} = \frac{1}{c_m}\left(\frac{\delta y_j}{2}\right)^2 \tag{11}$$

[9]The figure plots the fraction of promotions against the value added the unobserved y_j but only against the observed value added per worker, and we do so in Figure 4. But, since (i) from Figure 1 we know that, for given value added per worker, family firms promote fewer workers and (ii) promotions are increasing in y_j, for a family firm to have the same abscissa (same value added per worker) as a non-family firm, it must have a higher value of y_j (that is, a higher value y_j must compensate the negative effect on promotions due to the higher β_j). Hence, if we were to plot promotion rates against y_j, the data points corresponding to family firms would be shifted to the right relative to that shown in Figure 5, making the difference between promotion rates in family and non-family firms larger than it looks in the diagram.

However, the expected *unconditional* bonus is going to be different for family and non-family firms. We formally prove this result in Appendix A.2. However, the intuition is simple. Recall from our Result 1 that non-family firms are more likely to promote workers than family firms, and that the gap in the promotion rate increases with the productivity y_j. Hence, workers promoted in non-family firms are going to be disproportionately concentrated in high-productivity firms. As the expected bonus b is increasing in the firm's productivity, integrating b over the productivity distribution will result in a higher expected value for non-family firms.

Result 4. Family firms employees have lower expected utility than non-family firm ones, but applicants to jobs in family firms have greater probability of being hired. The equilibrium expected utility of workers from working in firm j, computed as of $t = 1$, is also increasing in the probability of skilled workers' promotions:

$$\mathbb{E}(U_{1j}^*) = \frac{1}{8c_w}\left[y_j + \frac{3}{2}\frac{\phi_j^*}{c_m}\left(\frac{\delta y_j}{2}\right)^2\right]^2 = \frac{1}{8c_w}\left[y_j + \frac{3}{2\beta_j c_m^2}\left(\frac{\delta y_j}{2}\right)^4\right]^2. \tag{12}$$

Recalling that y_j is identically distributed across family and non-family firms, the cross-sectional average of $\mathbb{E}(U_{1j}^*)$, taken over j for each of the two sets of firms, is larger for non-family firms than for family ones: by offering more rewarding careers, on average the former foster greater human capital accumulation and offer higher lifetime expected utility than the latter. Formally, $\mathbb{E}(U_{NF}^*) > \mathbb{E}(U_F^*)$, where $\mathbb{E}(U_{NF}^*) = \mathbb{E}_{j\in NF}[\mathbb{E}(U_{1j}^*)]$ and $\mathbb{E}(U_F^*) = \mathbb{E}_{j\in F}[\mathbb{E}(U_{1j}^*)]$, NF and F respectively being the set of non-family and family firms. Note that, since expression (12) is increasing in y_j, the inequality $\mathbb{E}(U_{NF}^*) > \mathbb{E}(U_F^*)$ would hold *a fortiori* if non-family firms were more uniformly productive than family ones, i.e. the cross-sectional distribution of y_j of the former first-order stochastically dominated that of the latter, rather than being identical as assumed here. This prediction is consistent with our in Figure 2, based on the methodology proposed by Sorkin (2018), and on Figure 3, based on observable amenities.

The model helps understand that this fact can be reconciled with labor market equilibrium: even though family firm employees enjoy on average lower utility than their non-family counterparts, at the job application stage the expected utility of the two groups of workers can be the same. This is because those to apply to jobs in non-family firms run a greater risk of unemployment, as such jobs are more eagerly sought for and thus have a greater number of applicants than jobs in family firms.

6 Conclusion

This paper asks how and why the compensation and promotion policies of family and non-family firms differ, and what do these differences imply for their performance. We confirm, on a much larger sample of Italian firms, the finding of the existing literature that there is a meaningful gap in average wages, with family firms paying substantially lower wages. However, our comprehensive employer-employee matched data enable us to explore the possible reasons for this gap. We find that about half of the gap, corresponding to 8 log-points, is due to workers' time invariant characteristics, but the remaining difference cannot be explained by sorting. Accounting for differences in rent sharing, productivity, industry, and location does not fully eliminate the gap.

Moreover, the gap is more pronounced in high-productivity firms and is mainly driven by differences in compensation for relatively high-pay workers, i.e., those in the highest percentiles of the within-firm earnings distribution. In addition, family firms feature fewer top and middle management positions, fewer promotions and lower returns to promotions, controlling for observable worker and firm characteristics. These findings suggest the presence of a "glass ceiling" in family firms: the controlling family may be reluctant to share organizational and technical know-how with other employees. To test this hypothesis, we conduct event studies around CEO deaths. It turns out that productivity, workers' earnings and employment are more negatively affected by CEO deaths in family firms, and

especially in family-managed ones, than in non-family firms, matching for a host of observable characteristics. This finding corroborates our glass ceiling hypothesis: if in family firms the organizational and technical know-how is not shared with other employees, part of it is irremediably lost absent the top decision maker.

We also investigate an alternative "compensating differential" hypothesis, namely the possibility that the lower wage paid by family firms may be compensated by superior non-monetary amenities, such as a less confrontational work environment (Mueller and Philippon, 2011) or greater job security (Ellul et al., 2017). We measure the systematic utility associated with family and non-family firms following Sorkin (2018)'s revealed-preference approach, but find that family firms appear to be inferior employers. Moreover, we do not detect meaningful differences in observable amenities, such as the type of contract offered to workers.

We complement this analysis with a simple theoretical model. In our setting, the only difference between family and non-family firms is that that the shareholders of the former suffer a "disutility" from promoting workers, consistent with the presence of higher benefits from control in family firms. Despite its simplicity, our model rationalizes most of our empirical findings, namely that family firms display lower productivity and wages, have fewer managers, deliver lower wage gains upon promotions, and lead to lower expected utility for their workers. The model also suggests that, while offering lower wages and slower careers, family firms may still be able to compete for job applicants with non-family firms, by offering a greater likelihood of getting a job upon applying for it.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, 1999, "High Wage Workers and High Wage Firms", *Econometrica* 67, 251–333.

Anderson, Ronald C., and David M. Reeb, 2003, "Founding-Family Ownership and Firm Performance: Evidence from the S&P 500", *Journal of Finance* 58, 1301–1328.

Azariadis, Costas, 1975, "Implicit Contracts and Underemployment Equilibria", *Journal of Political Economy* 83, 1183–1202.

Baily, Martin Neil, 1974, "Wages and Employment under Uncertain Demand", *Review of Economic Studies* 41, 37–50.

Bandiera, Oriana, Renata Lemos, Andrea Prat, and Raffaella Sadun, 2017, "Managing the Family Firm: Evidence from CEOs at Work", *The Review of Financial Studies* 31, 1605–1653.

Bassanini, Andrea, Thomas Breda, Eve Caroli, and Antoine Rebérioux, 2013, "Working in Family Firms: Paid Less but More Secure? Evidence from French Matched Employer-Employee Data", *ILR Review* 66, 433–466.

Bennedsen, Morten, Kasper Meisner Nielsen, Francisco Perez-Gonzalez, and Daniel Wolfenzon, 2007, "Inside the Family Firm: The Role of Families in Succession Decisions and Performance", *The Quarterly Journal of Economics* 122, 647–691.

Blinder, Alan S., 1973, "Wage Discrimination: Reduced Form and Structural Estimates", *The Journal of Human Resources* 8, 436–455.

Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen, 2012, "Management practices across firms and countries", *Academy of Management Perspectives* 26, 12–33.

Bloom, Nicholas, and John Van Reenen, 2007, "Measuring and Explaining Management Practices Across Firms and Countries", *The Quarterly Journal of Economics* 122, 1351–1408.

Bloom, Nicholas, and John Van Reenen, 2010, "Why do management practices differ across firms and countries?", *Journal of Economic Perspectives* 24, 203–24.

Burkart, Mike, Fausto Panunzi, and Andrei Shleifer, 2003, "Family Firms", *The Journal of Finance* 58, 2167–2201.

Card, David, Jörg Heining, and Patrick Kline, 2013, "Workplace Heterogeneity and the Rise of West German Wage Inequality", *The Quarterly Journal of Economics* 128, 967–1015.

Claessens, Stijn, Simeon Djankov, and Larry H. P. Lang, 2000, "The separation of ownership and control in East Asian Corporations", *Journal of Financial Economics* 58, 81–112.

Edwards, Jeremy, Alfons J. Weichenrieder, and Alfons Weichenrieder, 2004, "How Weak is the Weakest-Link Principle? On the Measurement of Firm Owners' Control Rights", CESifo Working Paper Series 1255, CESifo.

Ellul, Andrew, Marco Pagano, and Fausto Panunzi, 2010, "Inheritance Law and Investment in Family Firms", *American Economic Review* 100, 2414–50.

Ellul, Andrew, Marco Pagano, and Fabiano Schivardi, 2017, "Employment and Wage Insurance within Firms: Worldwide Evidence", *The Review of Financial Studies* 31, 1298–1340.

Faccio, Mara, and Larry H.P Lang, 2002, "The ultimate ownership of Western European corporations", *Journal of Financial Economics* 65, 365–395.

Holmstrom, Bengt, 1983, "Equilibrium Long-Term Labor Contracts", *The Quarterly Journal of Economics* 98, 23–54.

Kang, Jun-Koo, and Jungmin Kim, 2020, "Do Family Firms Invest More than Nonfamily Firms in Employee-Friendly Policies?", *Management Science* 66, 1300–1324.

La Porta, Rafael, Florencio Lopez-De-Silanes, and Andrei Shleifer, 1999, "Corporate ownership around the world", *The Journal of Finance* 54, 471–517.

LaPorta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny, 1998, "Law and Finance", *Journal of Political Economy* 106, 1113–1155.

Mueller, Holger M., and Thomas Philippon, 2011, "Family Firms and Labor Relations", *American Economic Journal: Macroeconomics* 3, 218–45.

Oaxaca, Ronald, 1973, "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review* 14, 693–709.

Pan, Yihui, Tracy Yue Wang, and Michael S. Weisbach, 2016, "CEO Investment Cycles", *The Review of Financial Studies* 29, 2955–2999.

Pérez-González, Francisco, 2006, "Inherited Control and Firm Performance", *American Economic Review* 96, 1559–1588.

Shleifer, Andrei, and Lawrence H. Summers, 1988, "Breach of Trust in Hostile Takeovers", in Alan J. Auerbach, ed., *Corporate Takeovers: Causes and Consequences*, 33–56 (University of Chicago Press, Chicago).

Sorkin, Isaac, 2018, "Ranking Firms Using Revealed Preference", *The Quarterly Journal of Economics* 133, 1331–1393.

Sraer, David, and David Thesmar, 2007, "Performance and Behavior of Family Firms: Evidence from the French Stock Market", *Journal of the European Economic Association* 5, 709–751.

Taylor, Lucian A, 2010, "Why Are CEOs Rarely Fired? Evidence from Structural Estimation", *The Journal of Finance* 65, 2051–2087.

Volpin, Paolo F., 2002, "Governance with poor investor protection: evidence from top executive turnover in Italy", *Journal of Financial Economics* 64, 61–90.

Figure 1
Firm-Level Component of Wages and Labor Productivity

Figure 1 presents the average "AKM" firm fixed effect visà-vis the percentile of the logarithm of value added per worker. We regress the logarithm of worker's wage on year fixed effects, cubic polynomials of age and years of experience, worker fixed effects, and firm fixed effects. We then compute the average firm fixed effect for each percentile of the logarithm of value added per worker, separately for family and non-family firms. Red and blue squares refer to family and non-family firms, respectively.

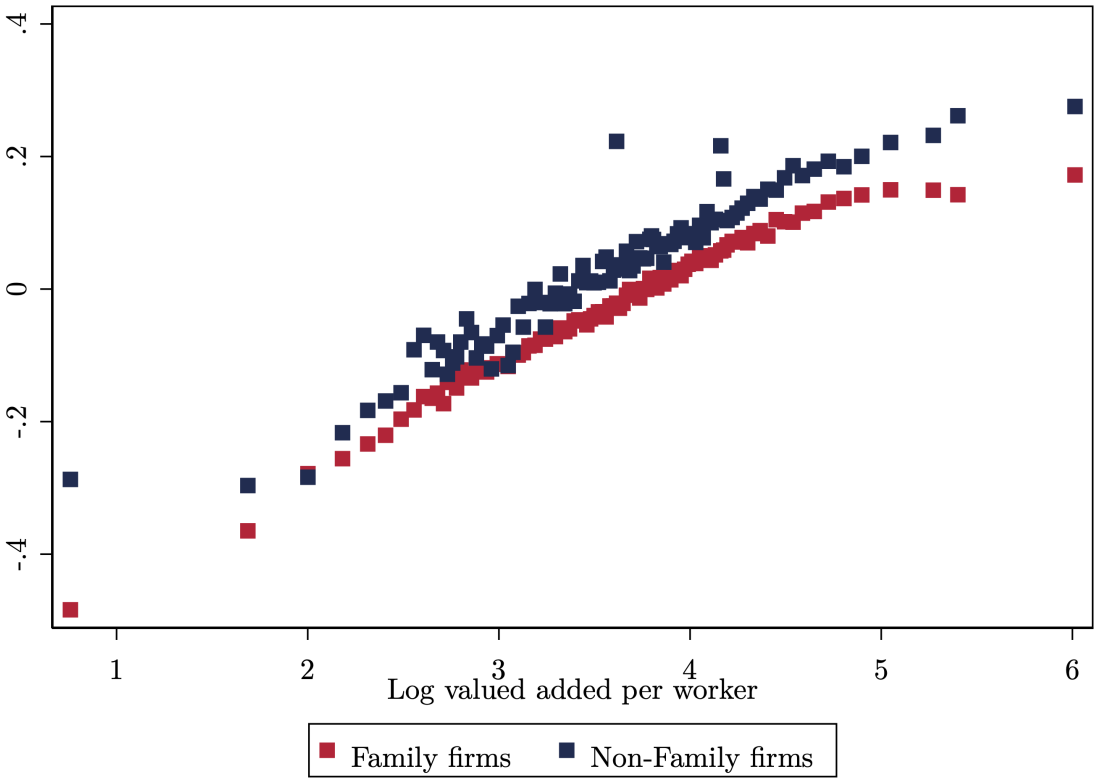


Figure 2
Worker's Utility and Labor Productivity

Figure 2 shows the average worker's typical utility from being employed in the firm visà-vis the percentile of the logarithm of value added per worker, computed using Sorkin (2018)'s approach.

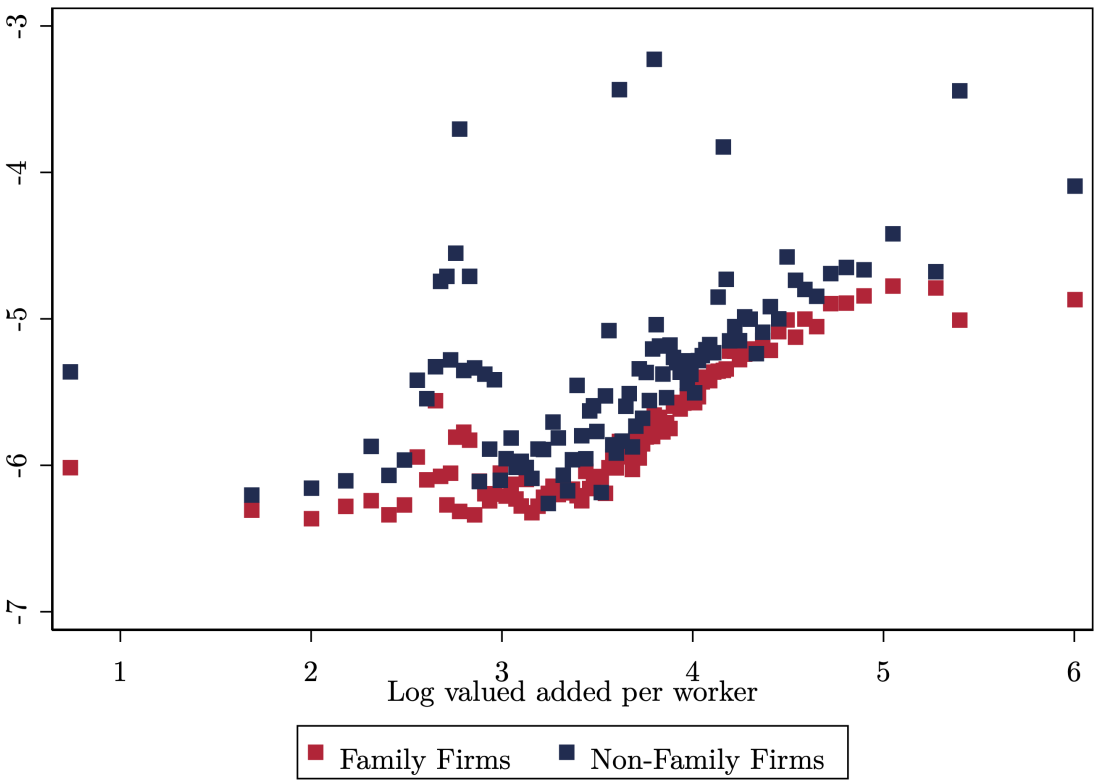


Figure 3
Observable Amenities

Figure 3 presents averages of several variables vis-à-vis the percentile of the logarithm of value added per worker. Panels A, B, and C display "AKM" fixed effects of a dummy equal to one if the worker has a part-time contract, of a dummy equal to one if the worker has a temporary contract, and of the number of reduced-contribution weeks, respectively. Panel D displays the average churn rate.

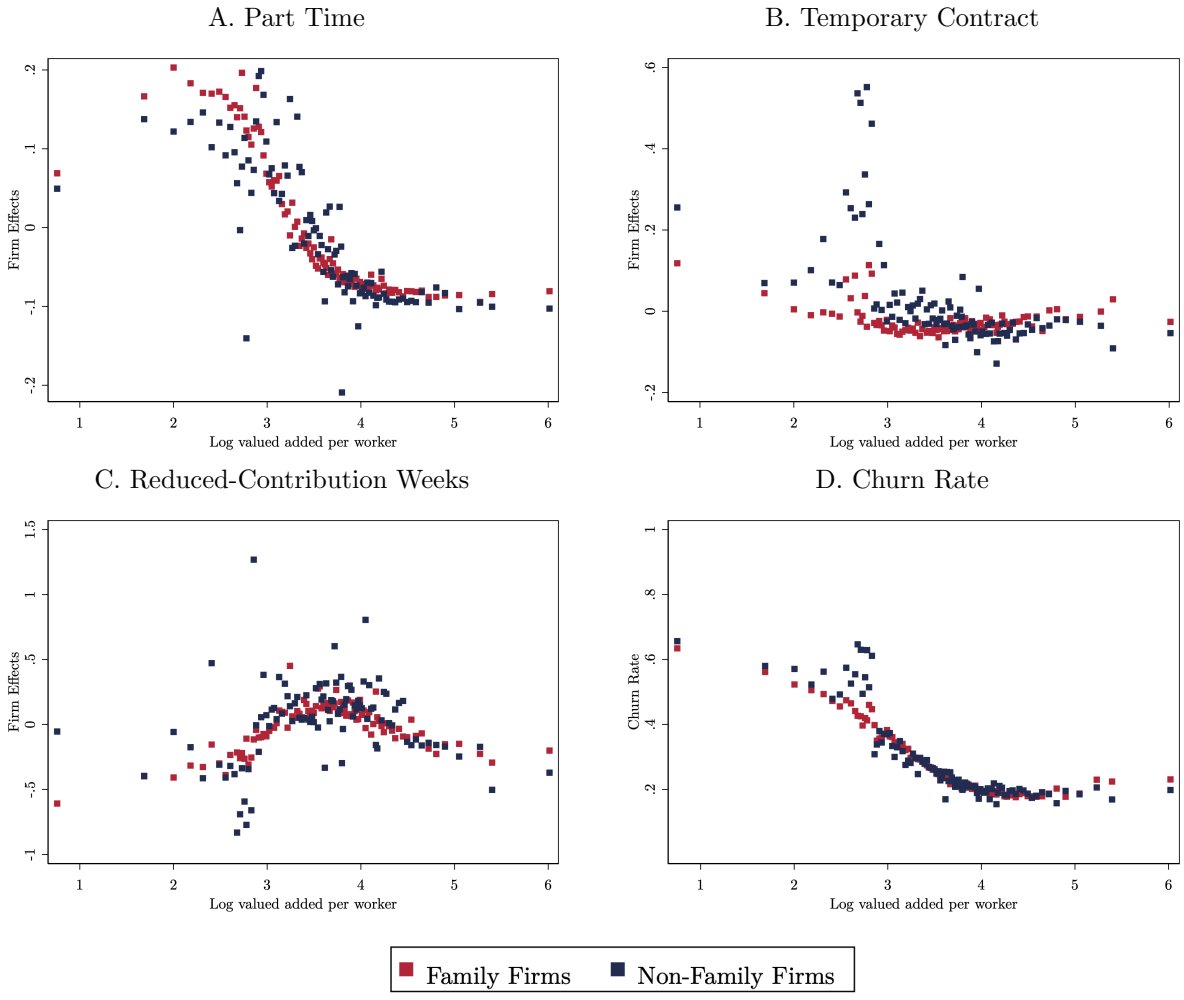


Figure 4
Within-Firm Wage Distribution

Figure 4 presents averages of the logarithm of weekly wages vis-à-vis the percentile of the logarithm of value added per worker. The variables on the $y-$axis are the 5^{th}, 10^{th}, 25^{th}, 75^{th}, 90^{th}, and 95^{th} percentiles of the within-firm distribution of the logarithm of wages. Red and blue squares refer to family and non-family firms, respectively.

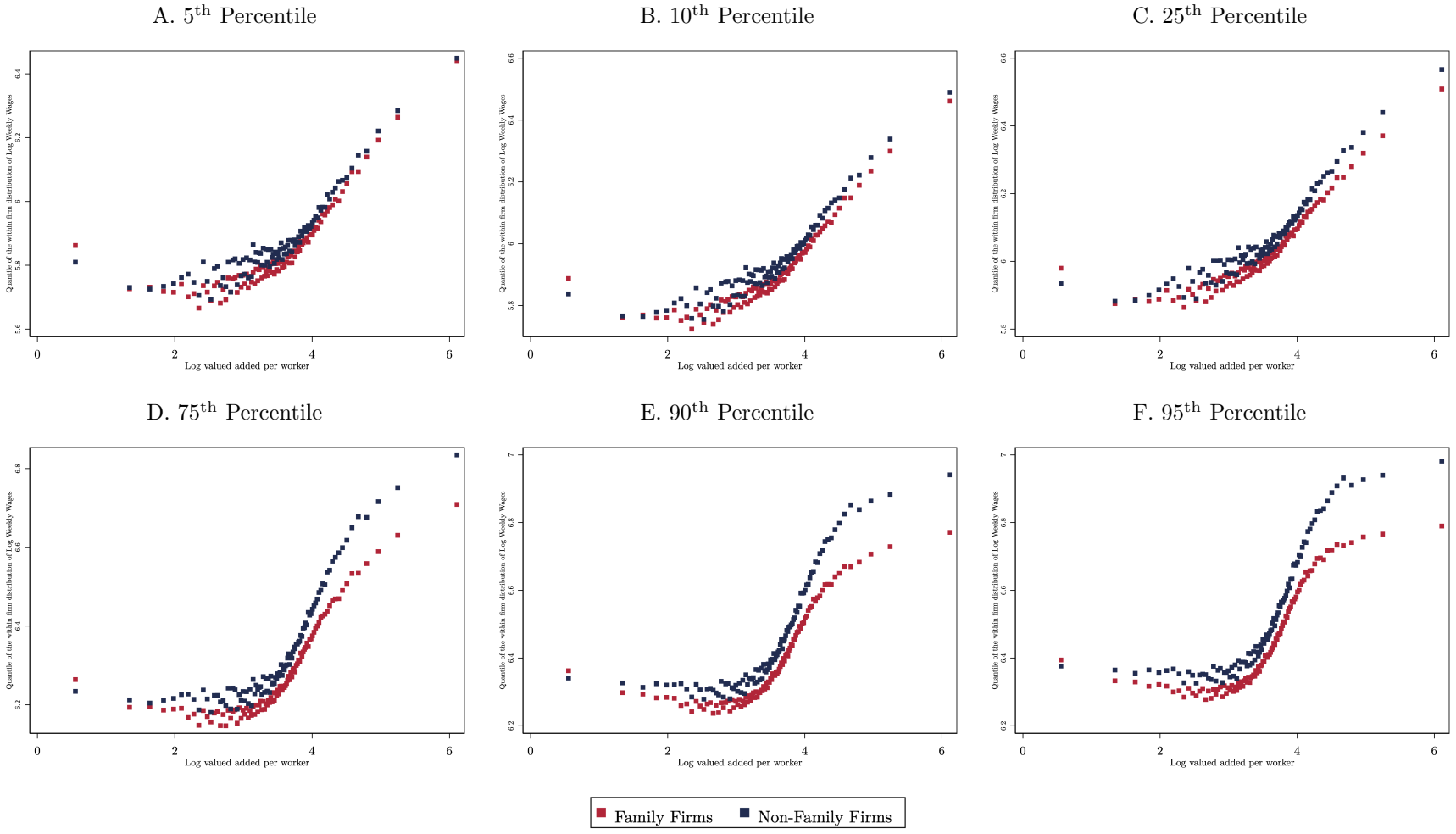
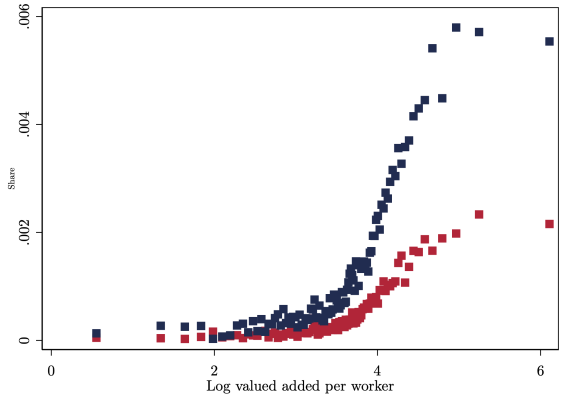


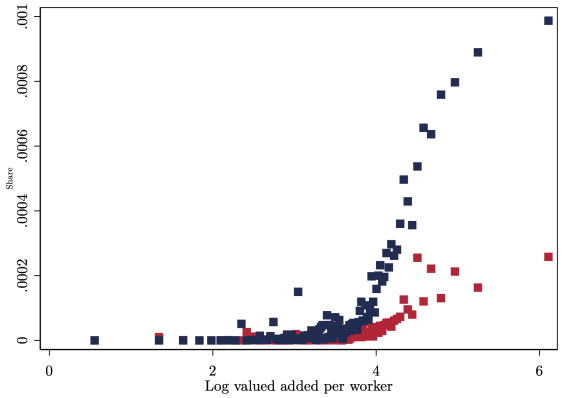
Figure 5
Promotion Rates and Fraction of Managerial Positions

Figure 5 presents averages of the promotion rates and of the fractions of managerial positions in family and non-family firms, by percentile of the logarithm of value added per worker. Panels A and B display the fractions of workers who are promoted (i.e., transitions) from a white-collar job to a middle-management position and from a middle-management position to a managerial position, respectively. Panels C and D display the fraction of middle managers and managers, respectively.

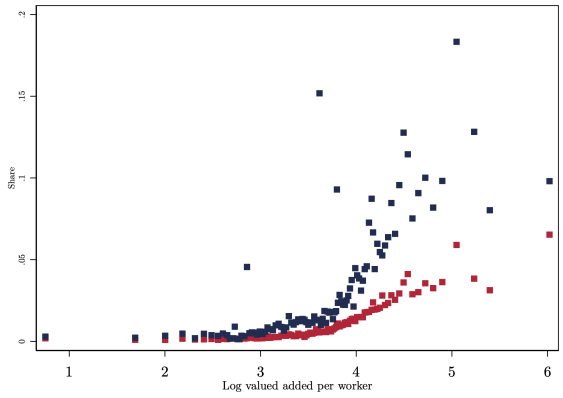
A. Transitions from W.C. to Middle Manager



B. Transitions from Middle Manager To Manager



C. Fraction of Middle Managers



D. Fraction of Managers

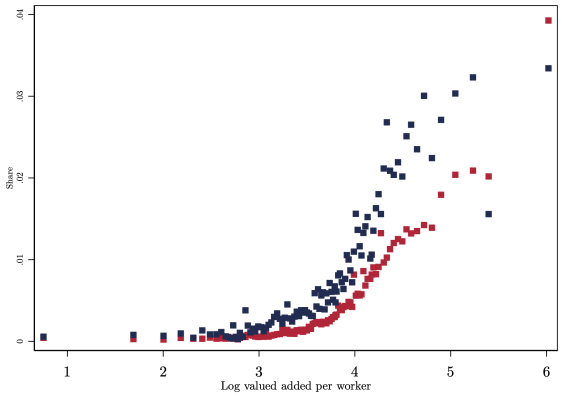


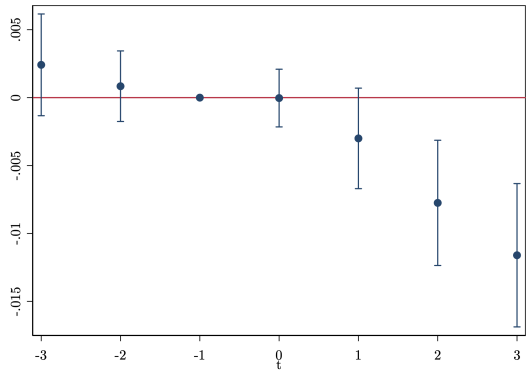


Figure 6
Returns to Promotions: Event-Study Evidence

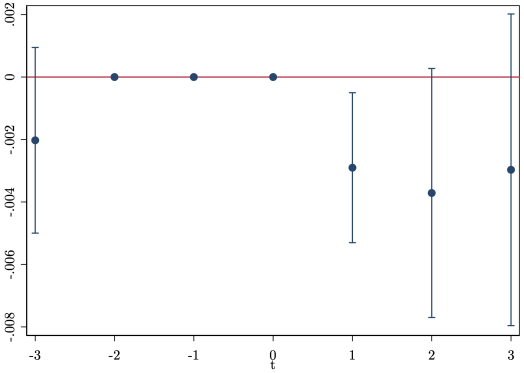
Figure 6 presents event-study evidence of the differential effects of a promotion depending on whether the worker was promoted by a family vs. non-family firm. We consider two types of promotion. Panel A includes workers promoted from white collar to middle management positions. Panel B includes workers promoted from middle management to managerial positions. For every worker promoted in a family firm, we find a worker who obtained the same type of promotion in a non-family firm, matching on the following characteristics: logarithm of firm size, the logarithm of value added per worker, and wage growth, age, and (exactly), industry, contract type, full time status, and gender. Workers are kept in the sample for a $(-3, +3)$-year window, where 0 is the year of the promotion event. The dependent variable is then regressed on year fixed effects, worker fixed effects, event-year dummies, and event-year dummies interacted with a family firm dummy. The figures plot the coefficients on these interaction terms with the associated 95% confidence intervals.

A. From White Collar to Middle Management

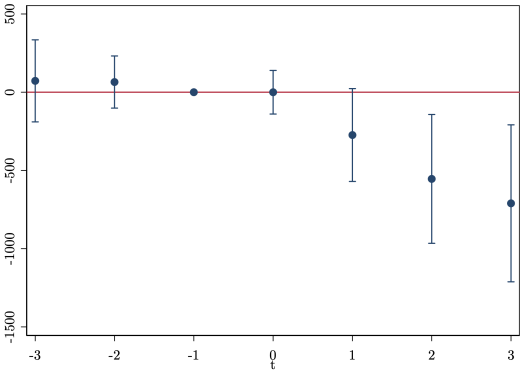
i. log(Earnings)



ii. Employment

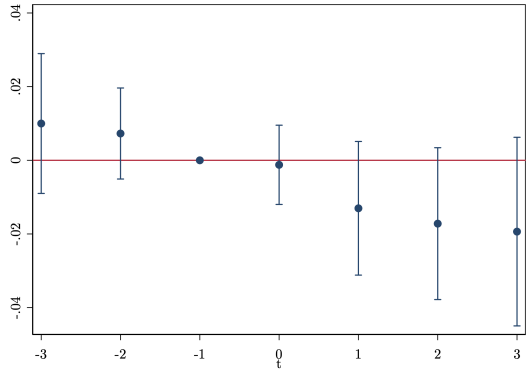


iii. Earnings

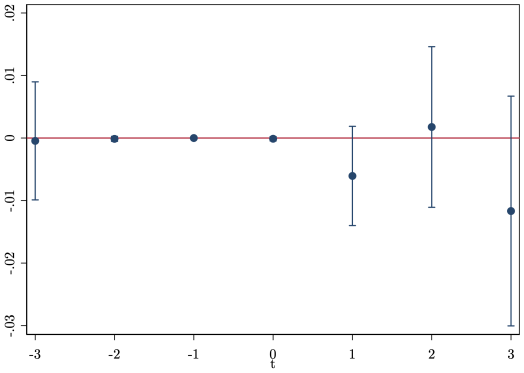


B. From Middle Management to Manager

i. log(Earnings)



ii. Employment



iii. Earnings

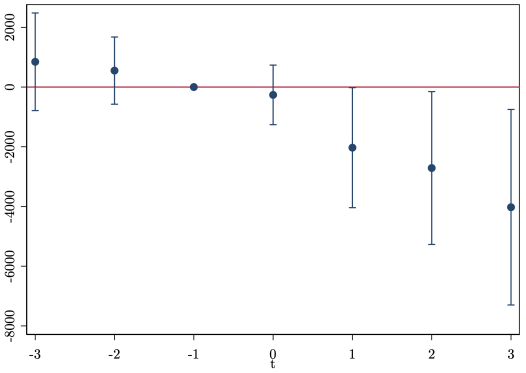
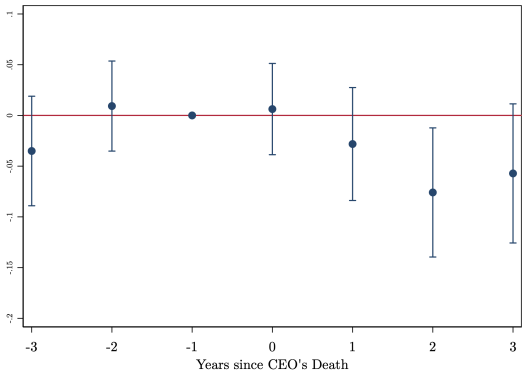


Figure 7
Productivity around CEO and Director Deaths

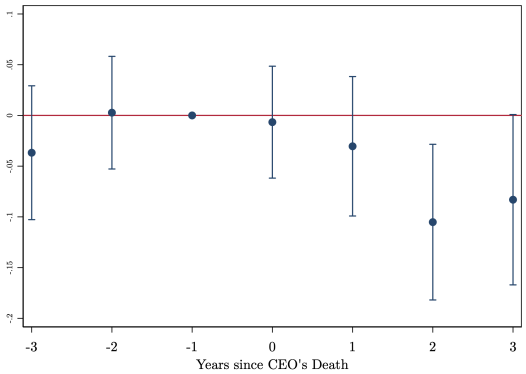
Figure 7 presents event-study evidence of the differential effects of a CEO (Panel A) or director (Panel B) death depending on whether the firm is a family or a non-family firm. For every family firm in which the CEO dies, we find a non-family firm matching on industry, the logarithm of firm size, and the logarithm of value added per worker. Firms are kept in the sample for a (-3,+3)-year window, where 0 is the CEO / director event. The logarithm of value added per worker is then regressed on year fixed effects, firm fixed effects, event-year dummies, and event-year dummies interacted with a family firm dummy. The figures plot the coefficients on these interaction terms with the associated 95% confidence intervals. Subpanels A.i and B.i include all firms. Subpanels A.ii and B.ii include only treated firms in which the deceased CEO / director was a member of the controlling family, together with the respective matched control firms. Subpanels A.iii and B.iii include only treated firms in which the deceased CEO / director was not a member of the controlling family, together with the respective matched control firms.

A. CEO Deaths

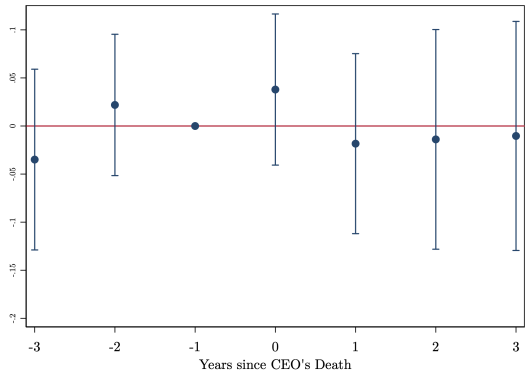
i. All Firms



ii. Family-Managed Firms

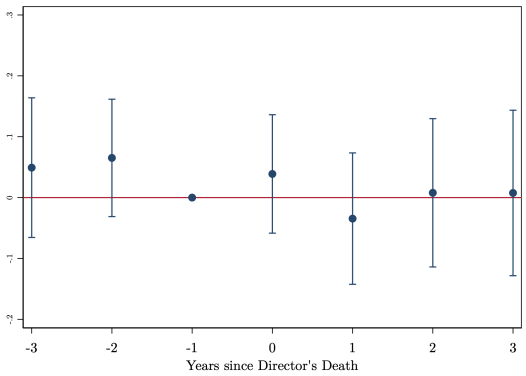


iii. Non-Family-Managed Firms

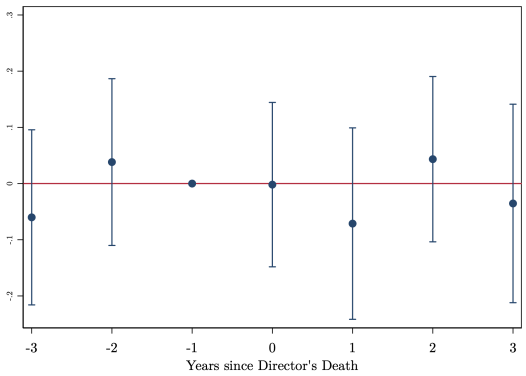


B. Director Deaths

i. All Firms



ii. Family-Managed Firms



iii. Non-Family-Managed Firms

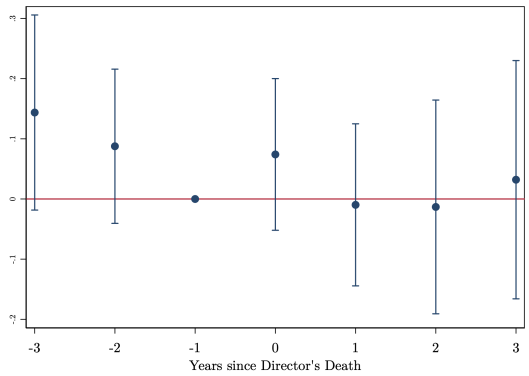


Figure 8
Worker-Level Outcomes around CEO Deaths

Figure 8 presents event-study evidence of the differential effects of a CEO death depending on whether the firm is a family or a non-family firm. For every family firm in which the CEO dies, we find a non-family firm matching on industry, the logarithm of firm size, and the logarithm of value added per worker. Within these paired firms, we match each worker in a family firm with a worker in a non-family firm based on wage growth, age, and (exactly), contract type, full time status, and gender. Workers are kept in the sample for a (3, +3)-year window, where 0 is the year of the promotion event. The dependent variable, indicated at the top of each panel, is then regressed on year fixed effects, worker fixed effects, event-year dummies, and event-year dummies interacted with a family firm dummy. The figures plot the coefficients on these interaction terms with the associated 95% confidence intervals.

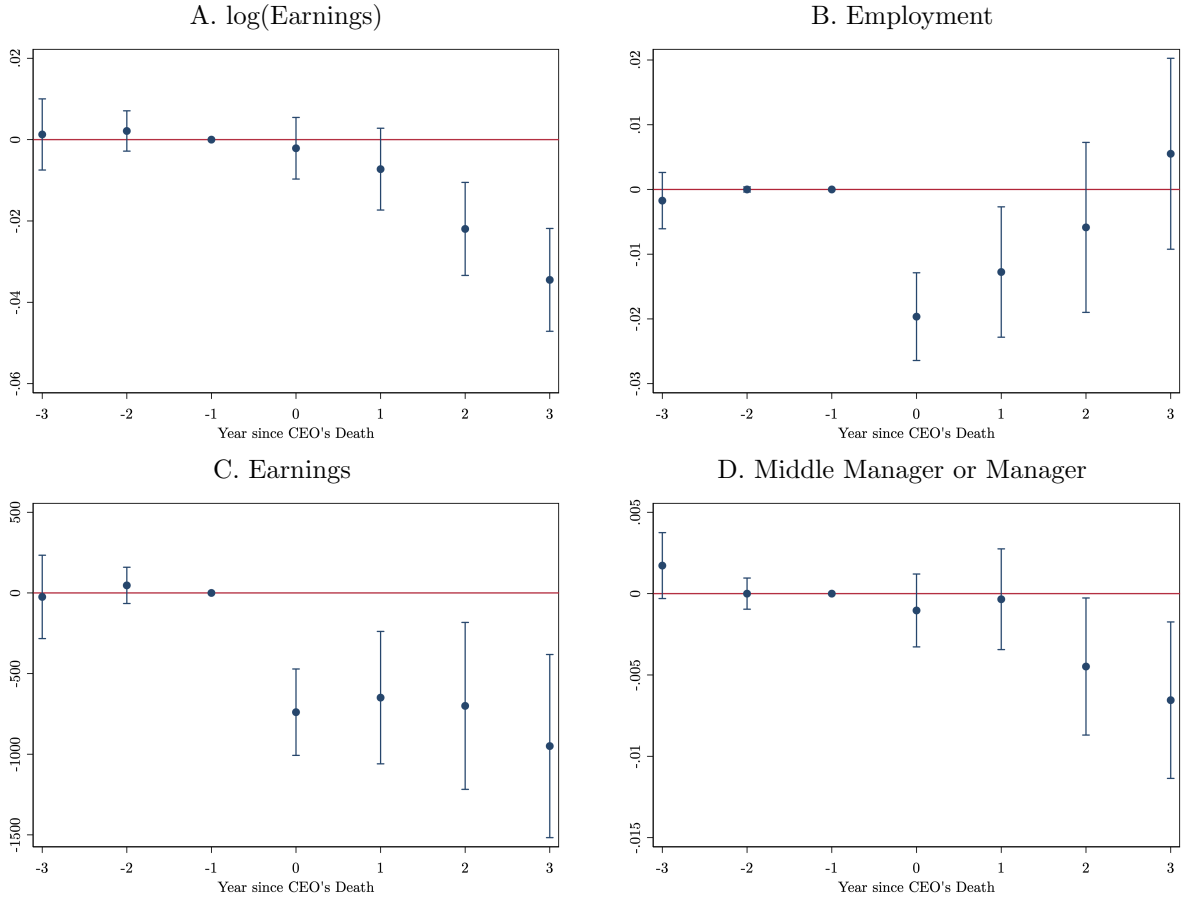


Figure 9
Timeline of the Model

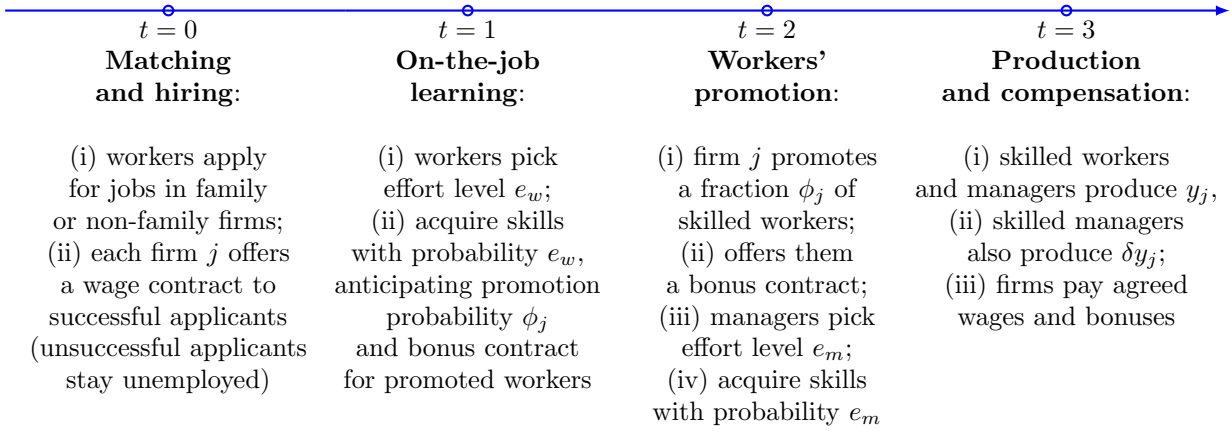


Table 1
Descriptive Statistics

Table 1 presents means of some of the firm-level variables observed in the sample, broken down between family and non-family firms. Firm size is the number of workers employed in the firm. Value added per worker is operating value added, in thousand euros, scaled by the number of workers. ROA is earnings before interests and taxes scaled by total assets, all multiplied by one hundred. Leverage is one minus common equity scaled by total assets. Wage is the weekly wage. Entry Log(Wage) is the logarithm of the entry wage (that is, the wage earned in the first year the worker has been employed with the firm). Log(Wage) (Max - Median) is the logarithm of the highest wage minus the logarithm of the median wage. Log(Wage) (90 - 10 Perc.) is the logarithm of the wage measured at the 90^{th} percentile minus the logarithm of the wage measured at the 10^{th} percentile. Share managers and share middle managers are the fractions of managers and middle maangers, respectively. Age is the average age of the firm's workforce. Experience is the number of years the worker has been employed in the current firm. Share males, part time, and temporary, are the fractions of male workers, part-time workers, and workers with temporary contracts, respectively. All the means, except for firm size, are weighted by the number of workers. The last four rows report the number of uniquer firms and workers, and the number of observations, at the firm-year level and the worker-year level.

	Family Firms	**Non-Family Firms**
Firm Size	14.056	31.782
Value Added per Worker	45.919	63.303
ROA	4.421	4.372
Leverage	0.788	0.762
Wage	436.833	510.696
Log(Wage)	6.034	6.177
Entry Log(Wage)	5.914	6.06
Log(Wage) Growth	0.042	0.046
Log(Wage) (Max - Median)	0.821	1.061
Log(Wage) (90 - 10 Perc.)	0.719	0.817
Share Managers	0.004	0.01
Share Middle Managers	0.01	0.04
Age	39.555	39.839
Experience	18.345	18.585
Share Males	0.651	0.634
Share Part Time	0.184	0.168
Share Temporary	0.21	0.221
# Firms	607,200	292,213
# Workers	9,972,643	9,421,821
# Firm-Years	3,319,919	1,394,195
# Worker-Year Obs.	46,664,781	44,310,305

Table 2
The Anatomy of the Family Firm Wage Discount

Table 2 shows averages of various variables across family and non-family firms. Person effect refers to the worker fixed effects from an "AKM" regression (Abowd et al., 1999) estimated on the sample of Italian workers employed by a firm in Cerved from 2005–2015. Firm effect refers to the firm fixed effects from the AKM regression. The regression further includes cubic polynomials of age and years of experience and year fixed effects as controls; after controlling for these variables we obtain the adjusted wage. Rent-sharing coefficient refers to the coefficient from a regression of the firm effects on a constant (denoted as theta in the table) and value-added per worker, estimated separately for family and non-family firms. Using these two regressions, we then decompose average differences between firm effects in family and non-family into a productivity channel (differences in productivity in firm effects between family and non-family firms weighted by the non-family rent-sharing coefficient) and a bargaining component (differences in rent-sharing weighted by average log value added per worker by family firms).

	Family Firm	**Non-Family Firm**	**Difference**
log(Wage)	6.03	6.18	-0.14
Adjusted Wage	5.90	6.07	-0.16
Person Fixed Effect	-0.03	0.05	-0.08
Firm Effect	-0.03	0.05	-0.08
Decomposed in...			
Bargaining Channel	0.51	0.47	0.04
Productivity Channel	0.47	0.49	-0.02
θ	-0.54	-0.44	-0.10
Rent Sharing Coefficient	0.14	0.13	0.01
log(Value Added / Worker)	3.56	3.73	-0.17

Table 3

Family Firm Wage Discount after Controlling for Productivity, Location and Sector

This table shows the results from a regression where the outcome variable is the firm effect estimated from a two-way (AKM) fixed effects decomposition on log wages. The regression is run on the micro, person-year, data and always controls for time fixed effects. Log value Added per worker is the average value added per worker observed for a given firm. Sector is defined as 2-digit Ateco code. Standard errors are clustered at the firm level.

	(1)	(2)	(3)	(4)	(5)
Family Firm	-0.082***	-0.059***	-0.047***	-0.042***	-0.038***
	(0.006)	(0.004)	(0.003)	(0.002)	(0.001)
log(Productivity)		0.137***	0.134***	0.113***	0.108***
		(0.002)	(0.002)	(0.002)	(0.002)
Std Dep. Variable	0.20	0.20	0.20	0.20	0.20
# of Observations	89,507,579	89,507,579	89,507,579	89,507,579	89,507,579
# of Firms	753,154	753,154	753,154	753,154	753,154
Industry FE			X	X	
Province FE				X	
LLM FE					X

Table 4
Systematic Utility and the Firm-level Wages

Table 4 shows coefficients from regressions of the systematic utility from working in the firm on several variables. The systematic utility is calculated following Sorkin (2018)'s approach. Family firm is a dummy equal to one if the firm is a family firm. AKM FE is the firm fixed effect obtained from running an the "AKM" regression. Column 3 includes industry fixed effects, where industry are defined using the ATECO two-digit classification. Column 4 further includes province fixed effects. Column 5 includes province × industry fixed effects.

	(1)	(2)	(3)	(4)	(5)
Family Firm	-0.703***	-0.581***	-0.299***	-0.282***	-0.231***
	(0.076)	(0.070)	(0.020)	(0.018)	(0.012)
AKM FE		1.511***	1.318***	1.169***	0.935***
		(0.208)	(0.101)	(0.085)	(0.049)
Std. of Dep. Var.	1.45	1.45	1.45	1.45	1.45
# of Firms	313,375	310,584	310,584	310,584	310,584
Industry FE			X	X	
Province FE				X	
LLM FE					X

A Appendix

A.1 Additional Results

Figure A1
The Family Firm Discount
and Workers' Places of Birth, Residence, and Location

Figure A1 shows regression coefficients obtained after regressing the firm "AKM" fixed effect on a family firm dummy, the logarithm of value added per worker, year fixed effects, and different sets of fixed effects. Every horizontal bar corresponds to the coefficient estimated on the family firm dummy in a different specification, with the associated 95% confidence interval in red. Going from the top bar to the bottom bar, the specifications include: no additional fixed effects, city of birth fixed effects, city of residence fixed effects, city of work location fixed effects, city of birth × city of residence fixed effects, city of residence × city of work location fixed effects, city of birth × city of work location fixed effects, and city of birth × city of residence × city of location fixed effects.

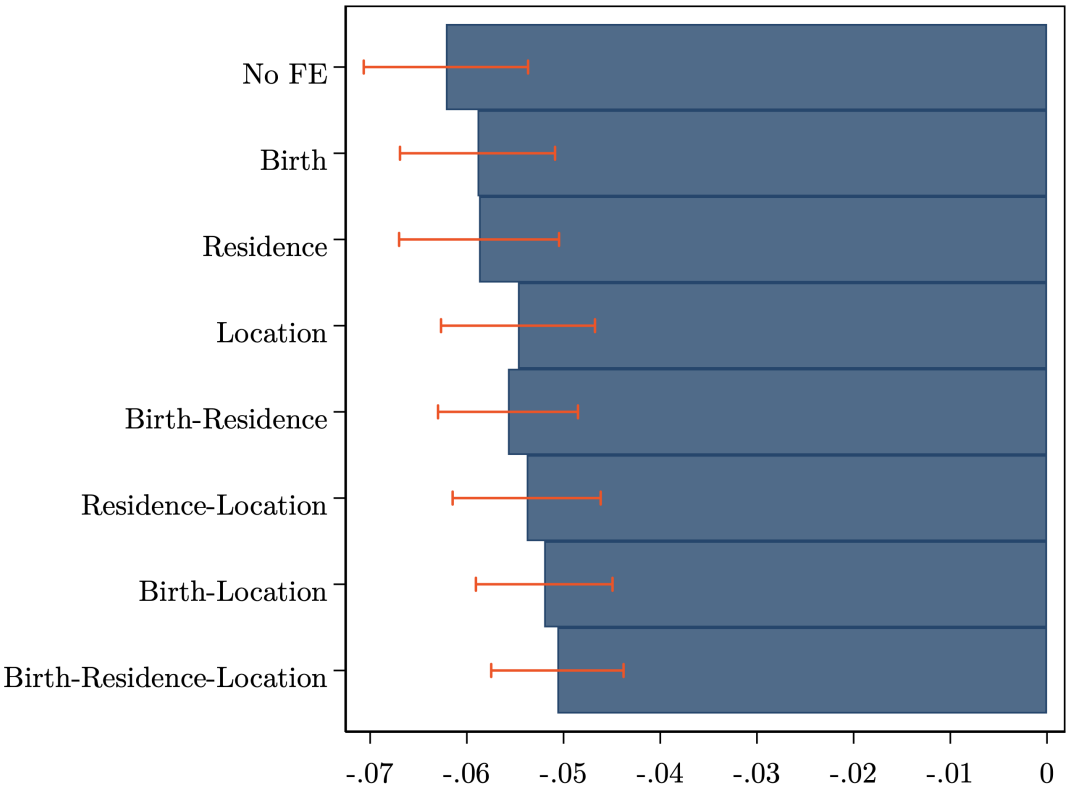
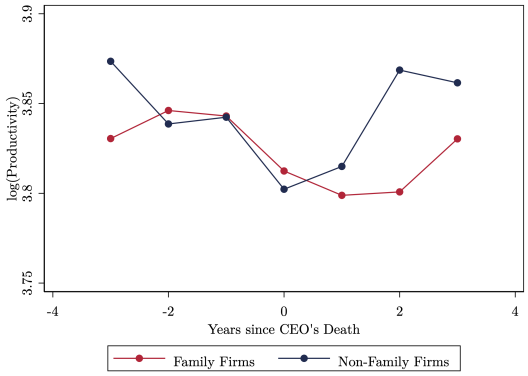


Figure A2
Productivity around CEO and Director Deaths: Sample Averages

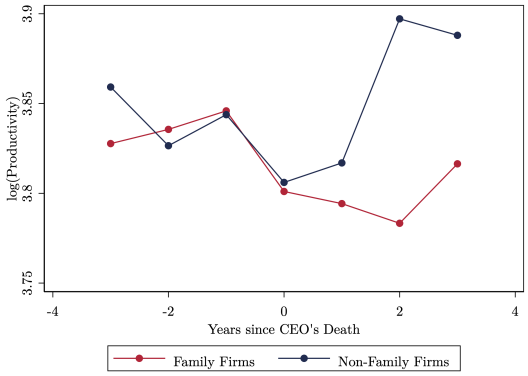
Figure A2 plots the average productivity in the seven years surrounding CEO (Panel A) or director (Panel B) deaths, depending on whether the firm is a family or a non-family firm. For every family firm in which the CEO dies, we find a non-family firm matching on industry, the logarithm of firm size, and the logarithm of value added per worker. Firms are kept in the sample for a (-3,+3)-year window, where 0 is the CEO / director event. Productivity is defined as the logarithm of value added per worker. Subpanels A.i and B.i include all firms. Subpanels A.ii and B.ii include only treated firms in which the deceased CEO / director was a member of the controlling family, together with the respective matched control firms. Subpanels A.iii and B.iii include only treated firms in which the deceased CEO / director was not a member of the controlling family, together with the respective matched control firms. Average productivity for family and non-family firms is plotted in red and blue, respectively.

A. CEO Deaths

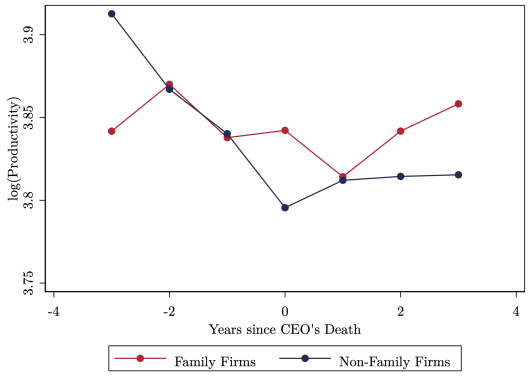
i. All Firms



ii. Family-Managed Firms

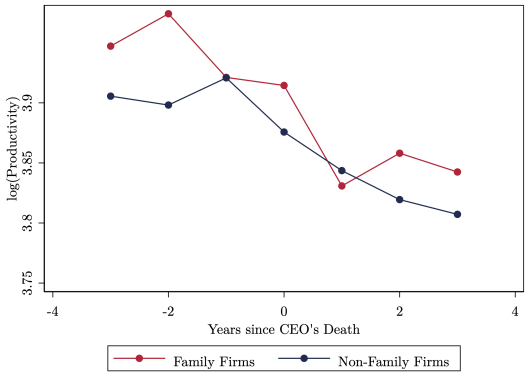


iii. Non-Family-Managed Firms

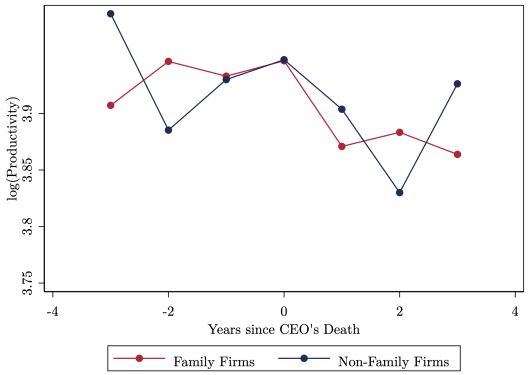


B. Director Deaths

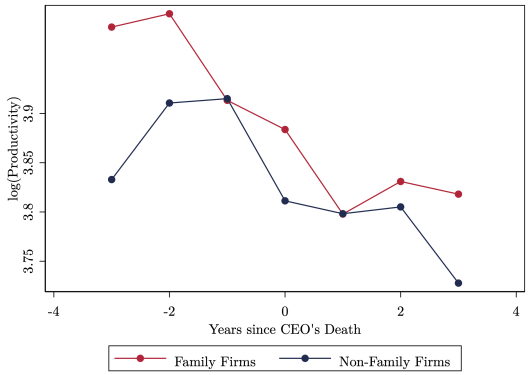
i. All Firms



ii. Family-Managed Firms



iii. Non-Family-Managed Firms



A.2 Backup of the model

To provide a conceptual framework to understand the empirical regularities documented so far, this section presents a model of the labor market and of firms' compensation policy that combines four features: (i) workers choose unobservable effort by balancing their compensation with their effort cost; (ii) firms incentivize employees via pay-for-performance contracts; (iii) firms choose the likelihood of promotions by balancing the efficiency gains of opening managerial positions for skilled workers with the implied reduction of control by company insiders; (iv) workers apply for jobs in family or non-family firms by comparing the expected utility of the corresponding expected career paths.

The economy is populated by risk-neutral workers and firms. There are N_{NF} non-family firms and N_F family firms, which differ in the weight they place on the benefits of corporate control. Each firm hires a mass 1 of workers, so that total employment is $N = N_{NF} + N_F$. The labor force L exceeds total employment N; as a result, $L - N$ workers remaining unemployed with utility normalized to zero. Each firm j produces y_j from skilled workers, and 0 from unskilled ones, where skills are acquired on the job with probability e_w if workers exert costly effort e_w. Effort is unobservable while skills are observable. Each firm draws y_j from a probability distribution F(y), assumed to be the same for family and non-family firms. Each firm i chooses to promote a fraction ϕ_j of skilled workers to managerial positions. Skilled workers promoted to these positions produce additional output δy_j if they also become skilled managers, which occurs with probability e_m if they exert costly effort e_m.

The timeline of the model, which is illustrated in Figure 9, is as follows:

- **Matching and hiring stage**. At stage $t = 0$ workers file a job application for a randomly chosen family or non-family firm, rationally taking into account the expected lifetime income respectively offered by the two types of firms, including the possible promotion to a managerial position. Firms hire workers by randomly picking them from the applicant pool, with unsuccessful applicants staying unemployed.

Upon hiring workers, firms offer them a (possibly state-contingent) wage contract and announce a promotion probability ϕ, which balances the efficiency gains from career concerns with the dislike for lower corporate control.

- **Workers' production stage.** At stage $t = 1$ workers engage in on-the-job learning by choosing costly effort e_w, they acquire the skills required for production with probability e_w.
- **Workers' promotion stage.** Each firm i chooses a fraction ϕ_j of skilled workers to be promoted to managerial positions and offers them a (possibly state-contingent) bonus contract. Managers can generate additional output δy_j only if they acquire costly managerial skills *on top of* worker skills. By exerting costly effort e_m, they acquire these additional skills with probability e_m.
- **Production and compensation stage.** Skilled workers, including managers, produce output y_j, and skilled managers produce an additional output δy_j. At this final stage, firms pay the agreed wages and bonuses to their employees.

Firms are assumed to be unable to commit to future actions: at $t = 0$ they cannot commit either to a specific promotion probability nor to offering a certain bonus contract to promoted workers $t = 2$. Hence, the model is solved by backward induction.

Stage $t = 2$. The optimal compensation scheme that any firm i can offer to its newly appointed managers is one that pays a bonus x_j if the managers succeed to produce the additional output δy_j at $t = 3$ and 0 otherwise (as limited liability prevents a negative bonus). Given this state-contingent contract, workers promoted to managerial positions in firm i choose effort e_m to hone their managerial skills, maximizing their stage-2 utility:

$$\max_{e_m} \ \mathbb{E}(U_{2i}) = e_m x_j - \frac{c_m}{2} e_m^2, \tag{A.1}$$

where the parameter c_m captures their cost of acquiring managerial skills. Hence, their optimal effort is $e_m^* = x_j/c_m$. Firm i will pick the bonus x_j so as to maximize the

incremental profits that it can extract from this effort e_m^*:

$$\max_{x_j} \ \mathbb{E}(\pi_{2i}) = \phi_j e_m^*(\delta y_j - x_j) = \frac{\phi_j x}{c_m}(\delta y_j - x_j), \tag{A.2}$$

where the expression takes into account that at stage 2 firm i promotes a fraction ϕ_j of its successful employees. Hence, the optimal bonus is $x_j^* = \delta y_j/2$. The resulting equilibrium managerial effort in firm i is increasing in the firm's productivity y_j:

$$e_m^*(x_j^*) = \frac{\delta y_j}{2c_m}, \tag{A.3}$$

where it is assumed that $\delta y_j \leq 2c_m$, so that the probability $e_m^*(x_j^*) \leq 1$. So the incremental profits that managers are expected to generate for firm i in equilibrium at stage 2 are

$$\mathbb{E}(\pi_{2i}^*) = \phi_j e_m^*(\delta y_j - x_j^*) = \frac{\phi_j}{c_m}\left(\frac{\delta y_j}{2}\right)^2. \tag{A.4}$$

This expression suggests that firm i should promote the largest feasible fraction ϕ_j of qualified workers.[10] But in their promotion policy the firm's owners may not only value the resulting expected profit $\mathbb{E}(\pi_{2i}^*)$ but also their private benefits of control: offering a high chance of promotion ϕ_j to qualified workers may also entail a cost, insofar as it constrains the owners' freedom to promote other favorite candidates, such as family members. To capture this possibility, we assume that firm i chooses its promotion probability by solving

$$\max_{\phi_j} \ \mathbb{E}(\pi_{2i}^*) - \frac{\beta_j}{2}\phi_j^2, \tag{A.5}$$

where the parameter β_j captures firm i's "taste for control over promotions" and the quadratic loss function captures the idea that incremental control losses generate increasing disutility for controlling shareholders. Hence, firm i's optimal promotion probability of

[10] As $\delta y_j \leq 2c_m$, a sufficient condition for $\phi_j^* \leq 1$ is that $\beta_j \geq \delta y_j/2$.

skilled workers is decreasing in the firm owners' taste for the benefits of control:

$$\phi_j^* = \frac{1}{\beta_j c_m}\left(\frac{\delta y_j}{2}\right)^2, \tag{A.6}$$

which, being a probability, is assumed not to exceed 1. In fact, the firm's technology and organizational structure may impose an upper bound smaller than 1 on ϕ_j^*: even highly meritocratic (low-β_j) firms may not wish to promote all the skilled workers who qualify for a promotion. It is natural to assume family firms to have a higher taste for corporate control, hence a lower optimal promotion probability: $\phi_F^* < \phi_{NF}^*$. But expression (A.6) indicates that, even within each of the two subsets of firms, the probability of promotion varies across them: it is higher in firms featuring greater productivity y_j, as this raises the efficiency gain from promoting qualified workers. Thus, even within the set of family firms, higher productivity ones will feature more frequent promotions. However, since ϕ^* is convex in productivity, the promotion probability should differ more across high-productivity family and non-family firms than across low-productivity ones.

Stage $t = 1$. At this stage as well, each firm i finds it optimal to offer workers a two-valued state-contingent compensation: w_j if the worker succeeds in acquiring productive skills on the job, and 0 otherwise. The employees of firm i choose effort e_w so as to maximize their stage-1 expected utility:

$$\max_{e_w} \ \mathbb{E}(U_{1i}) = e_w(w_j + \phi_j^* e_m^* x_j^*) - \frac{c_w}{2}e_w^2 - \phi_j^* e_w \frac{c_m}{2}e_m^{*2}, \tag{A.7}$$

where the parameter c_w captures the effort cost of acquiring worker skills on the job. The first term of expression (A.7) indicates that, when choosing their effort e_w, workers rationally anticipate that, if successful, they not only gain the wage w_j but are also given a chance ϕ_j^* of being promoted to a managerial position and earn the corresponding bonus x_j^*, while bearing the cost of acquiring the necessary managerial skills (the last term). Substituting from previous expressions for x_j^* and e_m^* and computing the first-order con-

dition with respect to e_w, one finds that the workers' optimal effort is increasing both in the wage and in the probability of promotion:

$$e_w^*(w_j) = \frac{1}{c_w}\left[w_j + \frac{\phi_j^*}{2c_m}\left(\frac{\delta y_j}{2}\right)^2\right] = \frac{1}{c_w}\left[w_j + \frac{1}{2\beta_j c_m^2}\left(\frac{\delta y_j}{2}\right)^4\right]. \tag{A.8}$$

The second step of equation A.8, obtained by replacing ϕ_j^* from (A.6), shows that the firm owners' taste for the benefits of control leads employees to produce less effort due to weaker career concerns, for a given entry-level wage w_j, and that this effect is stronger in more productive firms, being convex in y_j.

At stage 1, each firm i sets the wage so as to maximize its overall profits over both stages 1 and 2, anticipating its optimal future choice of the bonus x_j^* and the managers' future optimal choice of effort e_m^*:

$$\begin{aligned}\max_{w_j}\ \mathbb{E}(\pi_{1i}) &= e_w^*[(y_j - w_j) + \phi_j^* e_m^*(\delta y_j - x_j^*)] \\ &= \frac{1}{c_w}\left[w_j + \frac{\phi_j^*}{2c_m}\left(\frac{\delta y_j}{2}\right)^2\right]\left[y_j - w_j + \frac{\phi_j^*}{c_m}\left(\frac{\delta y_j}{2}\right)^2\right],\end{aligned} \tag{A.9}$$

yielding the optimal wage:

$$w_j^* = \frac{1}{2}\left[y_j + \frac{\phi_j^*}{2c_m}\left(\frac{\delta y_j}{2}\right)^2\right] = \frac{1}{2}\left[y_j + \frac{1}{2\beta_j c_m^2}\left(\frac{\delta y_j}{2}\right)^4\right]. \tag{A.10}$$

This expression shows that the optimal wage is not only increasing in the firm-level productivity parameter y_j but also in the probability of promotion ϕ^* and in workers' incremental future productivity as managers δy_j, as both of these raise workers' optimal effort. Upon substituting for the optimal promotion probability, in the second step of expression (A.10), one sees that high-β_j firms, where owners place great value on corporate control, pay lower wages, as less likely promotions reduce the power of their compensation package. Hence, family firms promote their workers less frequently *and* pay them lower wages.

Replacing the firm's optimal wage (A.10) in (A.8) yields workers' equilibrium effort:

$$e_w^*(w_j^*) = \frac{1}{2c_w}\left[y_j + \frac{3\phi_j^*}{2c_m}\left(\frac{\delta y_j}{2}\right)^2\right] = \frac{1}{2c_w}\left[y_j + \frac{3}{2\beta_j c_m^2}\left(\frac{\delta y_j}{2}\right)^4\right], \tag{A.11}$$

which, like the wage, is increasing in the workers' and managers' productivity and promotion probability, and thus decreasing in the firm owners' taste for the private benefits of control. Hence, the model predicts that family firms' employees exert less effort than those of non-family ones. As one might expect, this implies that family firms generate lower equilibrium expected profits than non-family ones:

$$\mathbb{E}(\pi_{1i}^*) = \frac{1}{4c_w}\left[y_j + \frac{3\phi_j^*}{2c_m}\left(\frac{\delta y_j}{2}\right)^2\right]^2 = \frac{1}{4c_w}\left[y_j + \frac{3}{2\beta_j c_m^2}\left(\frac{\delta y_j}{2}\right)^4\right]^2. \tag{A.12}$$

The equilibrium expected utility of workers from working in firm i is also increasing in the probability of skilled workers' promotions:

$$\mathbb{E}(U_{1i}^*) = \frac{1}{8c_w}\left[y_j + \frac{3}{2}\frac{\phi^*}{c_m}\left(\frac{\delta y_j}{2}\right)^2\right]^2 = \frac{1}{8c_w}\left[y_j + \frac{3}{2\beta_j c_m^2}\left(\frac{\delta y_j}{2}\right)^4\right]^2. \tag{A.13}$$

Recalling that y_j is identically distributed across family and non-family firms, the cross-sectional average of $\mathbb{E}(U_{1i}^*)$, taken over i for each of the two population of firms, is larger for non-family firms than for family ones: by more rewarding careers, the former are expected to foster human capital accumulation and offer higher lifetime expected utility than the latter. Formally, $\mathbb{E}(U_{NF}^*) > \mathbb{E}(U_F^*)$, where $\mathbb{E}(U_{NF}^*) = \mathbb{E}_{i\in NF}[\mathbb{E}(U_{1i}^*)]$ and $\mathbb{E}(U_F^*) = \mathbb{E}_{i\in F}[\mathbb{E}(U_{1i}^*)]$, NF and F respectively being the set of non-family and non-family firms. Note that, since expression (A.13) is increasing in y_j, the inequality $\mathbb{E}(U_{NF}^*) > \mathbb{E}(U_F^*)$ would hold *a fortiori* if non-family firms were more uniformly productive than family ones, i.e. the cross-sectional distribution of y_j of the former first-order stochastically dominated that of the latter, rather than being identical as assumed here.

Stage $t = 0$. In the initial stage of the game, the labor market allocates workers

between firms. Each worker chooses whether to file a job application with a family or a non-family firm, conditioning on rational expectations of future compensation and promotion prospects in the two types of firms, as computed above.

In equilibrium, workers will allocate their applications so as to equalize their expected utility from applying to family or non-family firms. Denoting the mass of applicants to family and non-family firms' jobs by A_F and A_{NF} respectively, and recalling that both types of firms hire workers by randomly drawing from their respective applicant pools, applicants obtain a job in a family firm with probability $\lambda_F \equiv N_F/A_F$ and in a non-family firm with probability $\lambda_{NF} \equiv N_{NF}/A_{NF}$. As employment in either type of firm yields a positive expected utility by equation (A.13), while unemployment entails zero utility, anyone in the labor force L applies for a job ($L = A_F + A_{NF}$), but not all applicants are hired, i.e. $\lambda_F < 1$ and/or $\lambda_F < 1$.

Applicants will be indifferent between applying in either set of firms if

$$\lambda_F \cdot \mathbb{E}(U_F^*) = \lambda_{NF} \cdot \mathbb{E}(U_{NF}^*). \tag{A.14}$$

Since in equilibrium careers in non-family firms have been shown to offer higher expected utility than those in family firms, the indifference condition (A.14) implies that $\lambda_{NF} < \lambda_F$: intuitively, the probability of being hired upon applying to a non-family firm must be low enough compared to that of being hired upon applying to a family firm. Otherwise said, in equilibrium applying for a job in non-family firms entails greater unemployment risk than applying for a job in family firms.

Finally, we prove that the expected bonus due to a promotion is higher in non-family firms than in non-family firms. The expected bonus b for a worker who is promoted, conditional on y is given by:

$$\mathbb{E}\left[b|y\right] = x^* e_m^* = \frac{\delta y}{2}\frac{\delta y}{2c_m} = \frac{1}{c_m}\left(\frac{\delta y}{2}\right)^2 \tag{A.15}$$

where for simplicity we have dropped the suffix j.

Using the law of iterated expectations, we have that the unconditional expected bonus B is equal to:

$$B = \mathbb{E}_y \left[\mathbb{E} \left[b|y \right] \right] = \frac{\int \frac{1}{c_m} \left(\frac{\delta y}{2} \right)^4 \left[y + \frac{3}{2\beta c_m^2} \left(\frac{\delta y}{2} \right)^4 \right] dF(y)}{\int \left(\frac{\delta y}{2} \right)^2 \left[y + \frac{3}{2\beta c_m^2} \left(\frac{\delta y}{2} \right)^4 \right] dF(y)} \tag{A.16}$$

Showing that the expected bonus is lower in family firms than in non-family firms is equivalent to showing that $\frac{\partial B}{\partial \beta} < 0$. Define $\int y^n dF(y) \equiv \mathbb{E}\, y^n$. Then, the numerator $\mathbb{N}$ of expression (A.16) is:

$$\mathbb{N} = \frac{1}{c_m} \left(\frac{\delta}{2} \right)^4 \mathbb{E}\, y^5 + \left(\frac{\delta}{2} \right)^8 \frac{3}{2\beta c_m^3} \mathbb{E}\, y^8$$

and therefore:

$$\frac{\partial \mathbb{N}}{\partial \beta} = - \left(\frac{\delta}{2} \right)^8 \frac{3}{2\beta^2 c_m^3} \mathbb{E}\, y^8.$$

The denominator $\mathbb{D}$ is instead:

$$\mathbb{D} = \left(\frac{\delta}{2} \right)^2 \mathbb{E}\, y^3 + \left(\frac{\delta}{2} \right)^6 \frac{3}{2\beta c_m^2} \mathbb{E}\, y^6$$

Therefore:

$$\frac{\partial \mathbb{D}}{\partial \beta} = - \left(\frac{\delta}{2} \right)^6 \frac{3}{2\beta^2 c_m^2} \mathbb{E}\, y^6.$$

Then, $\frac{\partial B}{\partial \beta}$ has the same sign as $\frac{\partial \mathbb{N}}{\partial \beta}\mathbb{D} - \frac{\partial \mathbb{D}}{\partial \beta}\mathbb{N}$, that is:

$$\begin{aligned} \frac{\partial \mathbb{N}}{\partial \beta}\mathbb{D} - \frac{\partial \mathbb{D}}{\partial \beta}\mathbb{N} = & - \left(\frac{\delta}{2} \right)^{10} \frac{3}{2\beta^2 c_m^3} \mathbb{E}\, y^3\, \mathbb{E}\, y^8 - \left(\frac{\delta}{2} \right)^{14} \frac{9}{4\beta^3 c_m^5} \mathbb{E}\, y^6\, \mathbb{E}\, y^8 \\ & + \left(\frac{\delta}{2} \right)^{10} \frac{3}{2\beta^2 c_m^3} \mathbb{E}\, y^5\, \mathbb{E}\, y^6 + \left(\frac{\delta}{2} \right)^{14} \frac{9}{4\beta^3 c_m^5} \mathbb{E}\, y^6\, \mathbb{E}\, y^8 \end{aligned} \tag{A.17}$$

Grouping the common terms, we have:

$$sign \left(\frac{\partial \mathbb{N}}{\partial \beta}\mathbb{D} - \frac{\partial \mathbb{D}}{\partial \beta}\mathbb{N} \right) = sign \left(\mathbb{E}\, y^5\, \mathbb{E}\, y^6 - \mathbb{E}\, y^3\, \mathbb{E}\, y^8 \right) \tag{A.18}$$

In the lognormal case we have that

$$\mathbb{E}\, y^n = e^{n\mu + n^2\sigma^2/2}.$$

Hence:

$$\mathbb{E}\, y^5\, \mathbb{E}\, y^6 - \mathbb{E}\, y^3\, \mathbb{E}\, y^8 = e^{11\mu + 61\sigma^2/2} - e^{11\mu + 73\sigma^2/2} < 0,$$

which in turn implies $\frac{\partial B}{\partial \beta} < 0$.