How Does the Position in Business Group Hierarchies Affect Workers' Wages?*

Hartmut Egger§
University of Bayreuth
CESifo, GEP and IfW

Elke Jahn
IAB Nuremberg, University of Bayreuth and IZA

Stefan Kornitzky
University of Bayreuth

May 13, 2021

Abstract

We merge firm-level data on ownership linkages with administrative data on German workers to analyze how the position in a business group hierarchy affects workers' wages. To acknowledge that ownership linkages are not one-directional, we propose an index of hierarchical distance to the ultimate owner that accounts for the complex network structure of business groups. After controlling for unobserved heterogeneity and selection into the business group hierarchy, we find a positive effect of larger hierarchical distance to the ultimate owner of a business group on workers' wages. To explain this finding, we develop a monitoring-based theory of business groups. Our model predicts higher wages to prevent shirking by workers if a larger hierarchical distance to the ultimate owner is associated with lower monitoring efficiency.

JEL-Classifications: C23, J31, L23

Keywords: Business groups, ownership networks, workers wages, difference-in-difference, hierarchical distance

*We would like to thank participants of the Graduate Seminar at the University of Bayreuth and the Bavarian Graduate Program in Economics Workshop for valuable discussion. The authors acknowledge financial support by the German Research Foundation (DFG) grants EG308/3-1 (Egger and Kornitzky) and JA1727/2-1 (Jahn).

§Corresponding author: University of Bayreuth, Department of Economics, Universitaetsstr. 30, 95447 Bayreuth, Germany; *hartmut.egger@uni-bayreuth.de*.

1 Introduction

Although it is widely perceived that business groups account for a major part of economic activity (UNCTAD, 2016), as a hybrid form between firms and markets they are not well-defined objects of economic theory and have therefore so far received little attention in academic research (see Baker et al., 2002; Khanna and Yafeh, 2007; Altomonte et al., 2018). In particular, it is not understood yet, how the specific organizational form of a business group influences its economic performance and workers' wages. This is surprising given the vast empirical evidence showing that firm organization is a key determinant of productivity and wages (see Caroli and Van Reenen, 2001; Rajan and Wulf, 2006; Bloom et al., 2010, 2018). A prominent strand of the literature points to the vertical position in firm hierarchy as a crucial determinant of workers' wages (see Caliendo et al., 2015; Bastos et al., 2018; Friedrich, forthcoming). Due to the lack of data combining detailed information on business groups, firms, and workers, empirical research on the role of business group hierarchy for workers' wages is missing so far, despite the material role of these groups for economic activity. To fill this gap, we construct and use a new dataset that allows us to study in a systematic way how the position of the employer in a business group hierarchy affects workers' wages.[1]

For our analysis, we define business groups as ownership networks, in which the ultimate owner exercises hierarchical control over the decisions made in all affiliated firms. We extract the relevant ownership information from the Bureau van Dijk global firm database Orbis. This database provides insights on worldwide ownership linkages and thus gives detailed information on the hierarchical position of firms in their business groups. To determine how the hierarchical position impacts workers' wages, we merge Orbis with administrative data on German employees from the Institute for Employment Research (IAB) in Nuremberg. This gives us a novel dataset with detailed information on the business group German establishments are part of as well as their workforce. However, ownership linkages are not one-directional. Thus, a simple count of hierarchical layers between an establishment and its ultimate owner would give at best an imprecise measure of vertical distance. To capture the complex structure of ownership networks, we develop a hierarchical distance index, which is motivated by recent work on sectoral input-

[1] Business groups play a prominent role in a sizable, mostly empirical literature on foreign ownership wage premia (see Girma et al., 2001; Girma and Görg, 2007; Balsvik and Haller, 2010; Hijzen et al., 2013; Egger et al., 2020; Egger and Jahn, 2020). This literature emphasizes the geographical location of the ultimate owner as an important determinant of wages in foreign subsidiaries.

output relationships (cf. Antràs and Chor, 2013) and measures hierarchical distance more consistently than a pure count of ownership layers.

To explain how the position in business groups affects workers' wages, we set up a theoretical model, in which production requires consecutive performance of a continuum of stages along the value chain of the business group. The value chain is split into two segments of endogenous length which are operated by an upstream and a downstream firm (cf. Costinot et al., 2013). Crucial for our analysis, we assume that the production process is prone to a loss of control problem due to limited monitoring capacity of the downstream firm (see Calvo and Wellisz, 1979; Chen, 2017). Focussing on the problem of a single business group, we show that the optimal labor allocation and the wage profile depend on relative monitoring efficiency in the upstream and downstream firm. We assume that the value chain follows the hierarchical structure of the business group, which is common practice, for instance, in the context of vertical multinational enterprises (see Helpman, 1984; Antràs and Yeaple, 2014). That is, we associate the upstream producer with a firm in larger hierarchical distance to the owner of the business group. In this case, lower monitoring efficiency in the upstream firm than in the downstream firm leads to a positive impact of hierarchical distance on wages. In contrast, if monitoring efficiency were lower in the downstream firm, the impact of hierarchical distance on wages would be negative.[2]

In the empirical analysis, we control for observable worker and establishment characteristics to isolate the effects of hierarchical distance from other factors that are important for wage payments. Including these controls, we identify a positive effect of larger hierarchical distance to the ultimate owner of a business group on individual wages in German establishments. Estimates from a parsimonious OLS specification show that an increase in the hierarchical distance by one standard deviation amounts to a sizable increase in wages of almost two log points. Although this estimate is reduced when additionally controlling for unobserved worker, establishment, and business group heterogeneity by fixed-effects, a positive and significant

[2]Our model bears close resemblance to the monitoring-based theory of firm hierarchies that has been put forward by Calvo and Wellisz (1978, 1979), Qian (1994), and Chen (2017). According to this theory, hierarchical layers can alleviate the loss of control problem inside the firm by increasing monitoring capacity and thereby reducing the incentive pay necessary to align workforce behavior with the objective of the owner. Our model is also related to the knowledge-based theory of firm hierarchies, in which hierarchical layers facilitate the information flow between workers and their superiors and thereby reduce the number of unsolved problems in the production process (see Garicano, 2000; Garicano and Rossi-Hansberg, 2006; Caliendo and Rossi-Hansberg, 2012). Chen and Suen (2019) discuss differences and similarities between monitoring-based and knowledge-based theories of firm hierarchies.

effect of hierarchical distance on wages still exists.

To make sure that the hierarchical distance variable does not erroneously pick up other features of business groups, we control for the total number of subsidiaries, as suggested by rich evidence for a firm size-wage premium (see, for instance, Brown and Medoff, 1989; Idson and Oi, 1999; Winter-Ebmer and Zweimüller, 1999; Colonelli et al., 2018). In addition, we combine information on the horizontal and the vertical dimension of business groups to an entropy index, measuring business group complexity (see Altomonte and Rungi, 2015). Adding these covariates, the impact of hierarchical distance on wages remains positive. We also address the potential problem of selection bias by combining propensity-score matching with a difference-in-difference estimator. This two-stage procedure gives a picture that is broadly in line with our baseline results: A larger hierarchical distance to the ultimate owner of the business group increases workers' wages.

Against the background of our theoretical model, the empirical results indicate that larger hierarchical distance is associated with lower monitoring efficiency, making higher wages necessary to provide an incentive for workers to follow the profit-maximizing objectives of the ultimate owner of the business group. Missing information on monitoring effort does not allow us to directly test the theoretical hypothesis from our model. However, we provide supportive evidence by splitting our sample into sub-groups of workers with differing levels of skills and sub-groups of occupations with differing shares of routine tasks. We find that the hierarchical distance effect is most pronounced for workers with high skills and for workers performing non-routine tasks – whose effort is the most difficult to observe. This indicates that our monitoring-based theory of business groups provides a suitable explanation for the positive effect of larger hierarchical distance to the ultimate owner on workers' wages.

The remainder of this paper is organized as follows. In Section 2, we outline a theoretical model for explaining wage payments along the business group hierarchy. In Section 3, we explain how we merge global firm data from Orbis with administrative data of German workers from the IAB. There, we also report summary statistics and show descriptive evidence on the relationship between hierarchical distance to the ultimate owner of the business group and workers' wages. In Section 4, we present the empirical analysis and report our estimation results. Section 5 concludes.

2 A monitoring-based theory of business group hierarchies

We consider a single business group in a competitive market that can sell its output at a given price equal to one. The business group operates a continuum of consecutively performed production stages with measure one (see Costinot et al., 2013) and faces the trade-off between monitoring workers or paying higher wages to reduce shirking (cf. Calvo and Wellisz, 1979; Chen, 2017). To facilitate our analysis, we assume a simple structure with two firms, which are associated with an upstream (intermediate goods) producer, $j = u$, and a downstream (final goods) producer, $j = d$, respectively. The value chain of the business group is split between these two firms into two disjoint segments with endogenous length. Capturing the value chain by the unit interval, we denote by $S \in (0, 1)$ the segment performed by the upstream producer and by $1 - S$ the segment performed by the downstream producer.[3] The ultimate owner of the business group makes all relevant decisions on production, hiring, and monitoring for both firms.

Production technology

Following Costinot et al. (2013), we consider a Leontief technology that combines one unit of labor input with one unit of intermediate good from the previous stage to produce (intermediate) output. Thereby, $\lambda \in (0, 1)$ captures a Poisson rate at which mistakes occur and destroy output in the production of the two firms. For an infinitesimal ds, we can express the technology of producing stage $s + ds$ as

$$q(s + ds) = (1 - \lambda ds)q(s). \tag{1}$$

In the limit of $ds \to 0$, Eq. (1) establishes the differential equation $q'(s) = -\lambda q(s)$, whose solution is given by $q(s) = q(0) \exp(-\lambda s)$ and determines business group output at stage s as a function of the initial input $q(0)$, which we associate with a cost-free intangible asset of the business group.

We denote the accumulated production cost for one unit of output at stage s in firm j by $c_j(s)$. Accordingly, for an infinitesimal ds the costs of producing $q(s + ds)$ in firm j can be expressed as $c_j(s)q(s) + w_j q(s)ds$, with w_j as the wage rate paid by producer j. Substituting $q(s + ds) = (1 - \lambda ds)q(s)$ from Eq. (1) gives

[3] Since in the limit the ultimate owner of the business group can assign a segment of length zero to either firm, the number of hierarchical layers is endogenous in our model.

$c_j(s+ds) = [c_j(s) + w_j ds]/(1-\lambda ds)$, which in the limit can be expressed as the differential equation $c_j'(s) = \lambda c_j(s) + w_j$. Solving this differential equation for either firm and making use of the boundary conditions $c_u(0) = 0$ and $c_u(S) = c_d(S)$, we can compute the labor costs of producing one unit of final output of the business group at $s = 1$ according to

$$c_d(1) = -\frac{w_d}{\lambda} + \left\{ \frac{w_u}{\lambda} [\exp(\lambda S) - 1] + \frac{w_d}{\lambda} \right\} \exp[\lambda(1-S)] \equiv c.$$

Due to the Leontief technology, we can determine labor demand of the upstream and the downstream firm according to $\ell_u = \int_0^S q(s)ds$ and $\ell_d = \int_S^1 q(s)ds$, respectively. Solving these two integrals gives

$$S = -\frac{1}{\lambda} \ln \left[\frac{q(0) - \lambda \ell_u}{q(0)} \right] \quad \text{and} \quad \lambda(\ell_u + \ell_d) = q(0) \{1 - \exp(-\lambda)\}, \tag{2}$$

where $q(0) - \lambda \ell_u$ is intermediate output of the upstream producer and $q(0) - \lambda(\ell_u + \ell_d) = q(0)\exp(-\lambda)$ is final output of the downstream producer. In view of Eq. (2), the unit cost of production simplifies to

$$c = \frac{w_u \ell_u + w_d \ell_d}{q(0)\exp(-\lambda)}. \tag{3}$$

Hiring, monitoring, and incentive pay

Firms hire workers at a convex cost of $\zeta \ell_j^2$.[4] Workers have a disutility of effort equal to one and thus an incentive to shirk, which would lower their labor productivity to zero. The probability of a shirker to be detected by the ultimate owner of the business group (resulting in immediate job loss and zero income) is firm specific and given by $p_j = \alpha_j m_j / \ell_j$, where m_j is monitoring input while $\alpha_j > 0$ captures monitoring efficiency. The participation constraint of workers can be written as $w_j \geq 1/p_j = \ell_j/(\alpha_j m_j)$ and it holds with equality if the ultimate owner chooses the profit-maximizing wage. Similar to other models featuring a loss of control problem, we assume that the monitoring capacity of the ultimate owner of the business group is limited and normalized to one: $m_u + m_d = 1$ (see Calvo and Wellisz, 1979; Chen, 2017).

[4]Whereas it is important for our analysis that hiring costs can differ between the upstream and downstream firm, a quadratic form is not necessary and imposed for the sake of analytical tractability.

The optimization problem

We can study the business group's optimization problem in two steps. In step one, we solve for cost-minimizing labor and monitoring inputs, ℓ_j, m_j, holding output $q(0)\exp(-\lambda)$ constant. In step two, we then determine the profit-maximizing level of output, $q(0)\exp(-\lambda)$, given the business group's cost function.

Making use of the binding participation constraint $w_j = \ell_j/(\alpha_j m_j)$ and the production technology in Eq. (2), total (production plus hiring) costs can be expressed as

$$C(m_u, \ell_u, q(0)) \equiv \zeta \left\{ \left(\frac{a_u}{m_u} + 1\right) \ell_u^2 + \left(\frac{a_d}{1-m_u} + 1\right) \left[\frac{q(0)}{\lambda}\{1-\exp(-\lambda)\} - \ell_u\right]^2 \right\},$$

where $a_j \equiv (\alpha_j \zeta)^{-1}$ is an auxiliary variable, inversely related to monitoring efficiency. Minimizing $C(m_u, \ell_u, q(0))$ for a given level of $q(0)$ establishes

$$\ell_u = \frac{a_d m_u + m_u(1-m_u)}{a_u(1-m_u) + a_d m_u + 2m_u(1-m_u)} \frac{q(0)}{\lambda}\{1-\exp(-\lambda)\} \tag{4}$$

and

$$m_u = \frac{\sqrt{a_u}(1+a_d) - \sqrt{a_d}a_u}{\sqrt{a_u} + \sqrt{a_d}}. \tag{5}$$

Thereby, an interior solution with $m_u \in (0,1)$ requires $\sqrt{a_u a_d} < 1 + \min\{a_u, a_d\}$ and thus the difference between a_u and a_d to be not too large. Using the solution to the cost-minimization problem, we can express total profits of the business group as

$$\Pi = q(0)\exp(-\lambda) - \zeta \left(\frac{q(0)}{\lambda}\right)^2 \{1-\exp(-\lambda)\}^2 \frac{(a_d + 1 - m_u^*)(a_u + m_u^*)}{a_u(1-m_u^*) + a_d m_u^* + 2m_u^*(1-m_u^*)},$$

where an asterisk is used to indicate the solution to the cost-minimization problem. Maximizing profits over $q(0)$ then gives an interior solution with $p_j < 1$ if ζ is sufficiently small.

Hierarchical wage profile

Differences in exogenous monitoring efficiency lead to differences in endogenous monitoring, according to Eq. (5). Moreover, it follows from Eqs. (2), (4), and (5) that higher monitoring efficiency is associated with higher labor input. We have $\ell_u > \ell_d$ if $\alpha_u > \alpha_d$, $\ell_u < \ell_d$ if $\alpha_u < \alpha_d$, and $\ell_u = \ell_d$ in the symmetric case of $\alpha_u = \alpha_d$. The effect of monitoring efficiency on labor allocation follows from its effect on wages, which can be determined when noting that the optimal allocation of monitoring input is characterized by the condition

$$\frac{\ell_u}{\alpha_u m_u} = \sqrt{\frac{\alpha_d}{\alpha_u}} \frac{\ell_d}{\alpha_d(1-m_u)}. \tag{6}$$

Making use of the binding participation constraint, Eq. (6) can be reformulated to $w_u = w_d\sqrt{\alpha_d/\alpha_u}$.[5]

Next, we impose the assumption that the position of a firm in the value chain is decisive for its position in business group hierarchy, which is common practice, for instance, in the literature on vertical multinational enterprises.[6] In this case, wages decrease along the value chain, i.e. $w_u > w_d$, if a larger hierarchical distance to the ultimate owner of the business group is associated with lower monitoring efficiency, i.e. $\alpha_d > \alpha_u$. Research on organization networks gives good reason to believe that larger (hierarchical) distance is associated with higher costs of supervision (see Gumpert, 2018) – with the cost-saving motive providing a plausible explanation for the observed flattening of firm hierarchies over recent years (see Rajan and Wulf, 2006).

3 Data source and descriptives

In the following two subsections we introduce and describe our dataset. In the first one, we explain how we combine information on business groups, firms, and workers from two different sources. There, we also introduce the main variables and provide summary statistics of them. In the second subsection, we show descriptive evidence for the link between the hierarchical distance to the ultimate owner of a business group and workers' wages.

[5] Although we have assumed that firms have identical labor productivity, the fundamental condition in Eq. (6) governing the hierarchical wage profile in the business group would remain unchanged if productivity differences existed. Productivity differences would affect, however, the labor allocation within the business group, with production increased, ceteris paribus, in the firm showing the lower Poisson rate of mistake, λ.

[6] Starting with the seminal work by Helpman (1984) the theory of multinational firms associates a vertical investment with imports from foreign affiliates to the country hosting the headquarters of the multinational enterprise for final assembly of consumer goods (see Grossman and Helpman, 2003; Antràs and Helpman, 2004; Antràs and Yeaple, 2014).

3.1 Construction of the dataset

For our empirical analysis, we rely on two datasets. The first one covers the years 2013-2017 of Bureau van Dijk's commercial firm database Orbis. Orbis reports balance sheet information for several 100 million companies and their ownership linkages worldwide.[7] Orbis covers all firms that are subject to reporting obligations. For Germany, these are all corporate enterprises and cooperatives as well as large private companies with total assets or revenues above thresholds defined by law.[8] We select for each observation year German firms from Orbis that fulfill some minimum quality criteria and determine their ultimate owner, who can be German or not.[9]

To build the relevant business group, we follow Altomonte and Rungi (2013) and associate business groups with ownership networks of legally autonomous firms. We then extract the whole business group of the ultimate owner and keep firms with valid information on a unique ultimate owner. Thereby, we restrict attention to *major shareholders*, which are the owners with the highest fraction of shares above a 25 percent threshold.[10]

Controlling ownership linkages are hierarchical. They must be unique and can be used to divide the business group hierarchy into different layers of ascending order, assigning the ultimate owner a layer number of zero. Within the thus determined business groups, we also observe linkages of minor shareholders. Due to their existence, ownership needs not to be one-directional. For instance, a subsidiary can hold shares of its major shareholder and thus be minor shareholder of its owner.

Figure 1 shows the ownership structure of a typical business group in Orbis. In this example, firm A is the ultimate owner of the business group, which directly

[7] Orbis data have been used previously to study business groups, e.g. by Belenzon et al. (2013), Cravino and Levchenko (2017), and Altomonte et al. (2018). Altomonte et al. (2018) validate the Orbis data by comparing for the year 2010 the numbers of parents and subsidiaries of business groups by country with the respective numbers from UNCTAD. For these two key business group determinants they report a correlation well above 0.90.

[8] A firm has extensive reporting obligations if it exceeds two of the following three criteria: (1) net turnover above 12 million EUR, (2) total assets of 6 Mio EUR, (3) annual average of more than 50 employees.

[9] Firms must be active and their legal form as well as their independence indicator have to be known. Moreover, operating revenues and the number of employees have to be available for at least one year between 2010 and 2017.

[10] Using this threshold, we follow the German regulation according to which the Federal Cartel Office (Bundeskartellamt) has to review acquisitions if an acquiring firm takes over 25 percent or more of the shares of the acquired firm. In this case, the law assumes that there is a concentration of ownership giving the acquiring firm a material influence on the business of the acquired firm (see Montag and Wilson, 2011). Barbosa and Louri (2002) show evidence that the choice of the threshold does not play a role for the structure of multinational ownership networks.

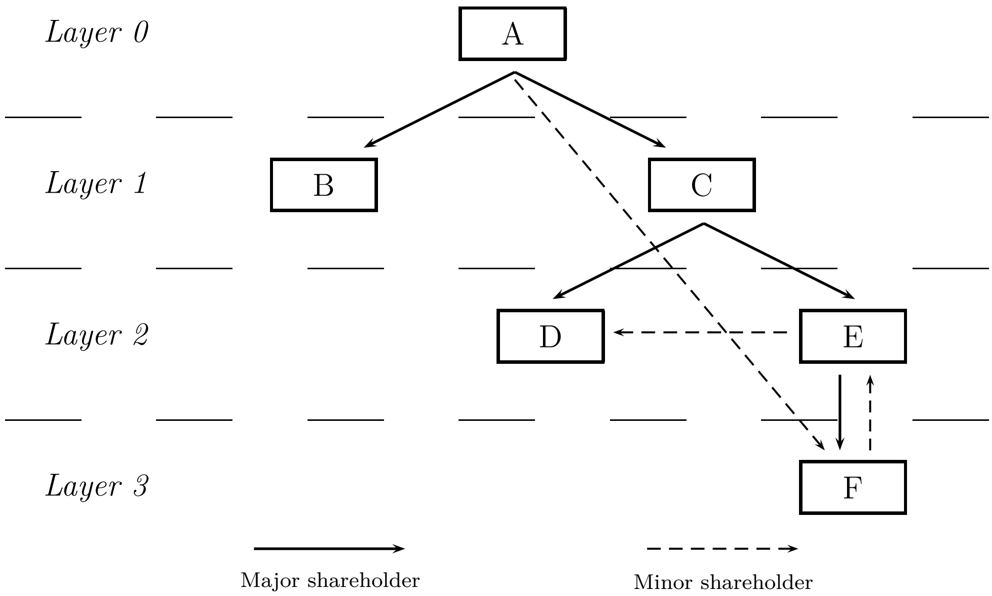

Figure 1: The business group as hierarchical ownership network

owns firms B and C and indirectly owns firms D, E, and F through its subsidiary C. Controlling ownership linkages of major shareholders are captured by solid arrows, whereas ownership by minor shareholders is captured by dashed arrows in Figure 1. Firms E and F are an example, in which ownership linkages are not one-directional. Using our network definition, we identify for each year about 40,000 different business groups, which cover at least one firm in Germany and represent in total almost one million firms worldwide.

As a second dataset, we use the Integrated Employment Biographies (IEB) from the Institute for Employment Research (IAB) of the German Federal Employment Agency. This dataset contains administrative records on all employees who are subject to social security contributions and covers about 80 percent of the German workforce. The IEB provides detailed information about age, gender, nationality, occupation, education, and the daily wage of workers employed in German establishments (see Klosterhuber et al., 2016). The IEB does not contain exact information on hours worked. Moreover, since worker information comes from social security records, wages are top-coded at the social security contribution ceiling. To deal with these issues, we consider only full-time workers aged 16–65 for our analysis and impute wages above the social security contribution ceiling, using the two-step Tobit procedure suggested by Card et al. (2013).[11]

[11]This procedure has been implemented for the IEB data by Dauth and Eppelsheimer (2020).

To merge information on administrative data of German workers from IEB with firm-level information on business groups from Orbis, we employ the procedure of the IAB, establishing linkages between observations in the two datasets relying on company names, addresses, and legal forms. Details on this procedure are provided by Antoni et al. (2018). To make sure that we correctly allocate establishments to firms over the whole sample period, we link them separately for each year between 2013 and 2017. The resulting record linkage keys allow us to merge on average 50,000 firms belonging to one of the business groups in Orbis with 86,000 establishments in the IEB per year. Finally, to ensure that each full-time worker is uniquely linked to a business group, we only keep employment spells that are valid on the 31st of December of a given year.

Firm-level variables from Orbis

Table 1 summarizes descriptive statistics on the main firm-level variables in our final dataset, which by construction do not vary over establishments or workers merged with the same firm. For the whole sample period, we count 250,494 firm-year observations.

The first two variables reported in Table 1 capture characteristics of the entire business group and are therefore identical for all firms belonging to the same group. The first variable is the total number of subsidiaries of a business group and therefore refers to *group size*. For the smallest business group, we count only one subsidiary, whereas, for the largest one, we count more than 13,000 subsidiaries. The average group size is 93 and thus fairly large.

The second variable combines information on the horizontal and vertical dimensions of business groups, that is the number of subsidiaries and the number of ownership layers, to an entropy index, which we refer to as *group complexity* (GC). It is constructed following Altomonte and Rungi (2015): $GC \equiv \sum_{l=1}^{L} l \frac{n_l}{N-1} \ln\left(\frac{N-1}{n_l}\right)$, where $N-1$ is the total number of subsidiaries, L is the total number of ownership layers, and n_l is the number of subsidiaries at ownership layer $l \in \{1, ..., L\}$. Group complexity picks up how the number of subsidiaries is spread over different ownership layers of a business group. It increases with the number of layers and places a higher weight on hierarchically more distant subsidiaries. Group complexity takes a minimum value of zero for business groups with only one layer and is unbounded from above. Its maximum level in our dataset is 28.4.

The main variable of interest in our analysis is the hierarchical distance of a firm

Table 1: Business group characteristics and hierarchical distance

	Mean	*Std. Dev.*	*Min*	*Max*
Group size	93.272	389.587	1	13,434
Group complexity	1.376	2.243	0	28.448
Hierarchical distance	1.245	1.178	0	18.770

Notes: Business group characteristics are constructed for the years 2013-2017, using firm-level information on ownership linkages from Bureau van Dijk's Orbis database. Group size is given by the total count of subsidiaries of a business group. Group complexity (GC) is defined following Altomonte and Rungi (2015). Data moments are reported for 250,494 firm-year observations.

to its ultimate owner. To construct a sensible measure of hierarchical distance, we have to acknowledge that more than nine percent of the firm-year observations in our dataset show ownership linkages that are not one-directional. For instance, in Figure 1 ownership linkages are circular for firms E and F. Firm E is a major shareholder of firm F, which in turn is a minor shareholder of firm E. However, circular ownership linkages can be even more complicated than that, spanning over multiple layers of hierarchy and including many different firms. We account for this complex pattern by developing an index of hierarchical distance that comprehensively captures circular network structures. To construct our index, we build on a method that has recently been applied for determining the vertical position of industries in global value chains (cf. Antràs et al., 2012; Antràs and Chor, 2013).

As a point of departure, we use the available ownership information, denote by ρ_{jk} the share of firm j that is owned by firm k, and express the chain of ownership in a business group as follows:

$$\sum_{k=1}^{N} \rho_{jk} + \sum_{k=1}^{N} \sum_{h=1}^{N} \rho_{jh}\rho_{hk} + \sum_{k=1}^{N} \sum_{h=1}^{N} \sum_{l=1}^{N} \rho_{jl}\rho_{lh}\rho_{hk} + \dots , \tag{7}$$

where N is the total number of firms in a business group, including the ultimate owner and all its subsidiaries. Setting ρ_{jj} equal to zero, the first element of the series in (7) measures how the *direct* ownership of firm j is spread in the business group and therefore refers to the first level of outside control. The other elements refer to *indirect* ownership, taking into account that firms holding shares of subsidiary j can in turn be owned by other firms in the business group. Since we observe $\sum_{k=1}^{N} \rho_{jk} > 1$, the series in (7) can be divergent. To address this problem and to disregard all ownership linkages outside the business group, we replace ρ_{jk} by $\hat{\rho}_{jk} \equiv \rho_{jk} / \sum_{k=1}^{N} \rho_{jk}$, and define hierarchical distance of firm j to its ultimate owner according to

$$H_j = \sum_{k=1}^{N} \hat{\rho}_{jk} + \sum_{k=1}^{N}\sum_{h=1}^{N} \hat{\rho}_{jh}\hat{\rho}_{hk} + \sum_{k=1}^{N}\sum_{h=1}^{N}\sum_{l=1}^{N} \hat{\rho}_{jl}\hat{\rho}_{lh}\hat{\rho}_{hk} + \ldots .$$

Using matrix notation, we can then summarise the hierarchical distance of all firms in a business group to their common ultimate owner by a single vector:

$$\mathbf{H} = \mathbf{R} \cdot \mathbf{1} + \mathbf{R}^2 \cdot \mathbf{1} + \mathbf{R}^3 \cdot \mathbf{1}... = [\mathbf{I} - \mathbf{R}]^{-1}\mathbf{1} - \mathbf{1}, \tag{8}$$

where $\mathbf{1}$ is an $N \times 1$ column vector of ones and $\mathbf{R}$ is an $N \times N$ matrix with $\hat{\rho}_{jk}$ as its (j,k)-th element.[12] The hierarchical distance of j from its ultimate owner is then given by the j-th row of the $N \times 1$ column vector $\mathbf{H}$. Eq. (8) determines the hierarchical distance of a subsidiary as a value-weighted count of the number of ownership layers between j and the ultimate owner in the business group (see Johnson, 2018). Higher values of H_j refer to a longer hierarchical distance and the index is normalized to give the ultimate owner a hierarchical distance value of zero. In our dataset, the hierarchical distance has a maximum of 18.8. If all ownership linkages were one-directional, index H_j would coincide with a simple layer count.

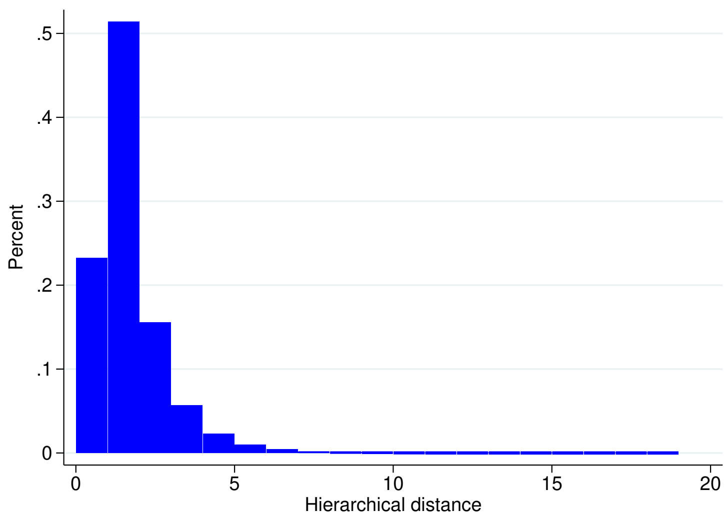

Figure 2: Frequency of hierarchical distances in firm-year observations

In contrast to the other two business group controls discussed above, hierarchical distance varies over firms within the same business group. Table 1 shows that both the mean and the standard deviation of our hierarchical distance variable are fairly

[12]Note that $[\mathbf{I} - \mathbf{R}]^{-1}$ is commonly known as the Leontief inverse matrix.

small, indicating a concentration of the firm-year observations at the bottom of its domain. Figure 2 illustrates this pattern, showing the frequencies of hierarchical distances in our dataset. Since H_j is a continuous variable, we report its frequencies for symmetric intervals of unit length. The first bar of the histogram captures the 23.2 percent ultimate owners among our firm-year observations. The remaining bars refer to the 76.8 percent subsidiaries, for which we find a strongly right-skewed density. Thus, a significant fraction of firm-year observations shows a low hierarchical distance to their ultimate owner.

Establishment and worker variables from IEB

Table 2 reports key summary statistics for the establishments and workers linked from the Integrated Employment Biographies (IEB) to the Orbis data. We count 21,609,088 worker-year observations that can be aggregated to 430,699 establishment-year observations over the sample period 2013–2017. The variation in log establishment size is fairly high and a major part of establishments come from three broad sector categories, namely manufacturing, retail & repair, and finance & insurance. Moreover, we observe considerable variation in (imputed) log daily wages, sizable age differences of workers, and underrepresentation of females. Classifying workers with no vocational training and no high-school degree as low-skilled, workers with a high-school degree and/or vocational training as medium-skilled, and workers with a degree from a university or a university of applied sciences as high-skilled, we find strong differences in the coverage of skill groups, with medium-skilled workers accounting for more than 70 percent of the worker-year observations.

3.2 Hierarchical distance and wages

Before turning to the econometric analysis, we use the linked IEB-Orbis dataset to provide descriptive evidence on how hierarchical distance to the ultimate owner affects workers' wages. To cancel out the impact of other covariates that have shown to be important wage determinants by previous empirical research, we first run a Mincer (1958)-type regression, in which we explain the log (daily) wage by worker observables on age (as a proxy for experience), age squared, and dummies for three skill groups, German nationality, female gender, and 16 federal states. We additionally control for time dummies and the six broad sector categories listed in Table 2.

To illustrate the correlation between the residual wage of workers and their hier-

Table 2: Establishment and worker characteristics

	Mean	Std. Dev.
(a) Establishment characteristics		
Log employment	2.997	1.497
Agriculture	0.008	0.087
Manufacturing	0.161	0.368
Mining, utilities & construction	0.080	0.271
Retail & repair	0.444	0.497
Finance & insurance	0.211	0.408
Private & public services	0.096	0.295
(b) Worker characteristics		
Log wage	4.815	0.496
Age	42.6	11.2
Female	0.265	0.442
Low-skilled	0.048	0.046
Medium-skilled	0.716	0.451
High-skilled	0.236	0.425

Notes: Establishment and worker descriptives are constructed for the years 2013-2017, using the Integrated Employment Biographies (IEB) from the Institute for Employment Research in Nuremberg. Establishment characteristics are computed for 430,699 establishment-year observations. Worker characteristics are computed for 21,609,088 worker-year observations. Low-skilled workers have no vocational training and no high-school degree. Workers with a high-school degree and/or vocational training are medium-skilled, whereas workers holding a degree from a university or a university of applied sciences are high-skilled.

archical distance to the ultimate owner, we assign establishments to the hierarchical distance we have computed for the firm they are merged with. We then cluster establishments into deciles of hierarchical distances and compute averages of hierarchical distances and residual wages for these deciles. Finally, we plot each pair of averages as an individual data point in Figure 3.[13]

Figure 3 shows a positive relationship between the hierarchical distance to the ultimate owner of a business group and workers' wages. According to our theoretical model outlined in Section 2, monitoring efficiency decreases with a larger hierarchical distance to the ultimate owner. Consequently, more distant establishments must pay higher wages to prevent shirking by workers. However, the evidence reported in Figure 3 is far from being conclusive and does not allow for causal inference on

[13]Since a high frequency of firms (and thus establishments) show a hierarchical distance of zero or one according to Figure 2, the number of distinct data points is less than 10.

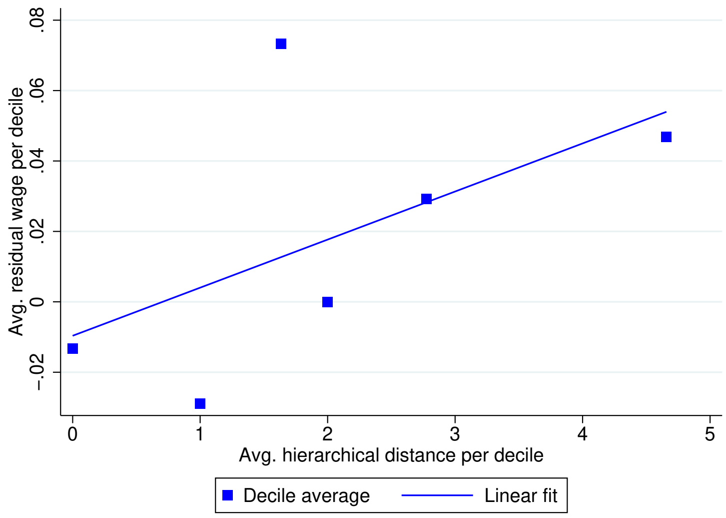

Figure 3: Hierarchical distance and wages in business groups

how changes in hierarchical distance affect workers' wages. In the empirical analysis of Section 4, we analyze the relationship between hierarchical distance and wage payments in business groups in further detail.

4 Estimation and empirical results

To study the role of business groups for individual wages in a systematic way, we first run OLS and fixed-effects regressions, in which we control for observable and unobservable worker and establishment characteristics. To avoid selection bias, we use in a second step propensity-score matching and select a control group that is (ex-ante) comparable to our treatment group. We then determine the effect of changes in hierarchical distance to the ultimate owner on wages by a difference-in-difference approach.

4.1 Baseline estimations

In the subsequent analysis, we estimate a model of the following form:

$$w_{ijkt} = \alpha + \mathbf{X}_{it} \cdot \beta + \mathbf{C}_{jt} \cdot \gamma + \mathbf{N}_{jkt} \cdot \nu + \mu_t + \epsilon_{ijkt}, \tag{9}$$

where w_{ijkt} is the log daily wage of worker i in establishment j, business group k, and year t and α is a constant. $\mathbf{X}_{it}$ is a (row) vector of the (time-varying) worker covariates age, age squared, and dummies for three skill groups, German nationality, and female gender, with β as the corresponding (column) vector of coefficients. $\mathbf{C}_{jt}$ is a vector of the (time-varying) establishment covariates log employment, log employment squared, and dummies for 16 German federal states and six broad sector categories, with γ as the corresponding vector of coefficients. Moreover, $\mathbf{N}_{jkt}$ is a vector of business group determinants with ν as the vector of coefficients, respectively. Depending on the specification $\mathbf{N}_{jkt}$ includes group size, group complexity, and our main variable of interest, the hierarchical distance index. While group size and group complexity capture characteristics of the entire business group, the hierarchical distance index varies across establishments, business groups, and time. Finally, μ_t is a vector of time dummies and ϵ_{ijkt} is the error term.

Of course, the baseline specification in Eq. (9) is prone to omitted variable bias if our set of controls does not cover all important worker, establishment, and business group determinants of wages. We capture unobserved, time-invariant determinants by adding worker-establishment-(business-)group fixed-effects. This gives a modified regression model of the following form:

$$w_{ijkt} = \alpha + \mathbf{X}_{it} \cdot \beta + \mathbf{C}_{jt} \cdot \gamma + \mathbf{N}_{jkt} \cdot \nu + \mu_t + \phi_{ijk} + \epsilon_{ijkt}, \tag{9'}$$

where ϕ_{ijk} denotes worker-establishment-group fixed-effects. By including these fixed-effects, we time-demean each worker-establishment-group observation and identify the effects of changes in the business group covariates through their variation over time. A change in the hierarchical distance variable can then only exert an effect on wages if a worker-establishment observation changes its hierarchical position in a given business group (by adding or dropping hierarchical layers). However, the effects arising from time-invariant worker, establishment, and business group determinants as well as the effects of workers switching the establishment or of establishments switching the business group are eliminated. This allows us to isolate the effect of changes in hierarchical distance from other factors influencing workers' wages, such as firm-size or foreign-ownership wage premia. As a result, the regression model in Eq. (9′) gives consistent estimates of ν, but it may underestimate the overall importance of business group variables for wages in our dataset.

Table 3 shows our estimation results. In all regressions, we control for the full set of worker and establishment covariates reported in Table 2 and additionally include time and federal state dummies. In Models (1), (3), and (5) we estimate Eq. (9) using OLS, whereas the remaining models refer to fixed-effects regressions based on Eq. (9′). Model (1) captures the most parsimonious specification and only includes the hierarchical distance of establishments to their ultimate owner as a business group control. The estimated effect is positive and significant at the one percent level. Increasing the hierarchical distance by one standard deviation ($\widehat{=}$ 1.20) increases wages by 1.71 log points. Abstracting from circular ownership linkages, one can interpret the size of this effect more intuitively as follows. Moving down one layer in a business group hierarchy would increase worker's wages by almost two percent. This effect is of similar magnitude as the foreign ownership premium typically found in the literature (see Girma and Görg, 2007; Egger et al., 2020). Model (2) shows that the size of this effect decreases when controlling for worker-establishment-group fixed-effects.

Table 3: Business groups, ownership hierarchy, and wages

Dependent variable: log daily wage	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Hierarchical distance	0.0145***	0.0013*	0.0130***	0.0028***	0.0025	0.0040***
	(0.0014)	(0.0007)	(0.0015)	(0.0008)	(0.0026)	(0.0014)
Group size			0.0103***	0.0013***		
			(0.0007)	(0.0003)		
Hierarchical distance ×Group size			-0.0012***	-0.0002***		
			(0.0002)	(0.0001)		
Group complexity					0.0269***	0.0015*
					(0.0013)	(0.0008)
Hierarchical distance ×Group complexity					-0.0030***	-0.0005***
					(0.0003)	(0.0002)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Worker Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Establishment Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Worker-establ.-group FE	No	Yes	No	Yes	No	Yes
R-sq. (within)	0.4440	0.0740	0.4504	0.0740	0.4500	0.0740
Observations	21,609,088	21,609,088	21,609,088	21,609,088	21,609,088	21,609,088

Notes: Worker covariates include age, age squared, dummies for three skill groups, German nationality, and gender. Establishment covariates include log employment, log employment squared, dummies for 16 German federal states, and six broad sector categories. In all models, we estimate a constant as well as a full set of time dummies. Hierarchical distance and the group index of complexity are constructed as outlined in Section 3. Standard errors in parentheses are clustered at the establishment-level. ***, ** and * denote significance at the 1, 5 and 10 percent levels, respectively.

In Models (3) to (6) we add further business group covariates. In Models (3) and (4) these are group size as well as its interaction with hierarchical distance. Adding these controls has rather small effects on our hierarchical distance estimate.

Moreover, the positive direct effect of group size is well in line with evidence on size-wage premia at the firm level (cf. Colonelli et al., 2018). The negative sign of the interaction term indicates that hierarchical distance is less important for wages in larger business groups. In Models (5) and (6) we investigate the role of group complexity and its interaction with hierarchical distance. In the OLS regression, we find that the impact of hierarchical distance, while staying positive, becomes considerably smaller than in the parsimonious specification of Model (1) and loses its statistical significance. In contrast, the direct effect of higher group complexity on wages is positive, sizable, and significant. This result changes considerably in the fixed-effects regression. Controlling for time-invariant unobserved heterogeneity of workers, establishments, and business groups, we find a positive and significant impact of larger hierarchical distance to the ultimate owner on workers' wages, while the impact of group complexity falls considerably. The negative and significant interaction term indicates that hierarchical distance plays a less important role for wages in more complex business groups.

Summing up, the results from Table 3 show that the omitted variable bias in estimating the link between hierarchical distance to the ultimate owner of a business group and workers' wages with simple OLS can be severe so that controlling for unobserved heterogeneity appears important. Moreover, the results from fixed-effects regressions are broadly in line with the descriptive evidence and support the conclusion that larger hierarchical distance reduces monitoring efficiency so that higher wages are needed to prevent shirking by workers in establishments at low hierarchical layers of the business group. However, although fixed-effects regressions are an effective remedy for time-invariant omitted variable bias, our estimates may still be prone to a selection bias that exists, for instance, if the position of subsidiaries and their workers in business group hierarchy is not random. To rule out a selection bias and to make sure that the positive effect of hierarchical distance reported in Table 3 is causal, we exploit in the next subsection a two-stage regression procedure, combining propensity-score matching with a difference-in-difference approach.

4.2 Selection into business groups

In line with our analysis in Section 4.1, we specify the treatment as an increase in the hierarchical distance (HD) between the ultimate owner and a worker-establishment pair. To isolate the hierarchical distance effect from other wage determinants associated with employer effects, we focus on workers who stay within the same establishment and business group around the treatment event. Therefore, we define the treatment as an increase in hierarchical distance between a worker-establishment pair and its ultimate owner within a given business group. Accordingly, we classify worker-establishment pairs as untreated if they show no change of hierarchical distance to their ultimate owner within a given business group.

Following the matching literature, we collapse the observation period 2013 to 2017 into two-year windows around the treatment period. We then eliminate all observations that are not classified as treated or untreated, that is workers switching their employer or establishments switching their business group between two time periods. Moreover, to avoid an influence on our estimate from worker heterogeneity, we specify the treatment at the worker level and capture this treatment by a binary indicator

$$D_{ijk} = \begin{cases} 1 & HD_{ijk,t=0} < HD_{ijk,t=1} \\ 0 & HD_{ijk,t=0} = HD_{ijk,t=1}, \end{cases}, \tag{10}$$

which takes a value of one, if the hierarchical distance to the ultimate owner of worker i from establishment j and business group k increases between periods $t = 0$ and $t = 1$. In contrast, the indicator takes a value of zero if the hierarchical distance to the ultimate owner of worker i from establishment j and business group k does not change.

To select for each treated observation a suitable control from the pool of untreated worker-establishment pairs, we rely on nearest-neighbor propensity-score matching (Rosenbaum and Rubin, 1983). For this purpose, we determine the probability in $t = 0$ that an observation is subject to treatment between periods 0 and 1 and estimate the following probit model:

$$P(D_{ijk} = 1) = \Phi(\nu \cdot \mathbf{N}_{jk,0} + \gamma \cdot \mathbf{C}_{j,0} + \beta \cdot \mathbf{X}_{i,0}), \tag{11}$$

where $\mathbf{N}_{jk,0}$, $\mathbf{C}_{j,0}$, $\mathbf{X}_{i,0}$ are vectors of business group, establishment, and worker covariates in period $t = 0$, with ν, γ and β being the corresponding vectors of coefficients. Business group covariates are hierarchical distance and group complexity. Establishment covariates include the log of employment to control for establishment size, sector dummies indicating the establishments' industry affiliation, and federal-state dummies to control for establishment location. Finally, worker covariates are dummies for females and three skill levels, workers' age, and their log daily wages.

To exclude time effects, we estimate the probit model in Eq. (11) within treatment cohorts, i.e. we match observations from the same year.[14] Since 206 observations are off support, we eliminate them from the treatment group after the probit estimation.

Using the estimates from our probit model, we then assign to each worker-establishment-group triple from the treatment group the worker-establishment-group triple from the pool of untreated observations with the smallest absolute difference in the propensity-score. This forms our control group, which contains fewer unique observations than the treatment group because we match with replacement (see Caliendo and Kopeinig, 2008). Moreover, since we match individuals, workers from a single establishment of the treatment group can be assigned to workers from different establishments belonging to different ownership networks in the control group. Overall, our matching procedure gives us a sample of 597,432 unique worker-establishment observations in the treatment group and 448,780 unique worker-establishment observations in the control group.

To evaluate the success of our matching procedure, we compare averages of all covariates used in the probit estimation before and after matching and report the results in the Appendix. There, we show two diagnostics that are commonly used to assess the matching quality. The first one is the standardized percentage bias introduced by Rosenbaum and Rubin (1985). Matching reduces the mean bias considerably from 12.6 percent to 1.9 percent. We also report the normalized difference between individual covariates from the treatment and control group, as put forward by Imbens and Wooldridge (2009) and Imbens and Rubin (2015). Imbens and Rubin (2015) suggest an upper limit of one quarter of the normalized difference to consider a variable as balanced. This critical threshold is not surpassed by any of our covariates after matching. The two diagnostics, therefore, indicate that we were successful in matching observations from the treatment group to similar untreated observations.

With the matched sample at hand, we can quantify the causal effect of larger hierarchical distance to the ultimate owner on wages using a difference-in-difference approach. In doing so, we contrast wages before and after treatment and compare the change in wages between workers from the treatment and the control group by

[14] As a robustness check, we account for the change in log employment before treatment as an additional control in the probit model. We include this variable to take employment dynamics before the treatment into account and report the results along with those from three further robustness checks in the Appendix.

estimating the following equation:

$$w_{ijkt} = \alpha_i + \mu + \eta \cdot D_{ijk} \cdot \mu + \epsilon_{ijkt}, \tag{12}$$

where w_{ijkt} is the log daily wage of worker i in establishment j, business group k, and year t, α_i is a worker fixed-effect to control for any remaining, time-invariant unobserved heterogeneity of workers, and μ is a time dummy that takes a value equal to one in the post-treatment period $t = 1$. D_{ijk} is the treatment indicator equal to one for each stayer i, whose establishment j has been subject to treatment between $t = 0$ and $t = 1$, and zero otherwise. Coefficient η captures the wage effect for workers, whose establishment increases its hierarchical distance to the ultimate owner within a given business group. Finally, ϵ_{ijkt} is the error term.

Table 4: Wage effect of larger distance in business group hierarchy

Dependent variable: Log daily wage	*All workers*	*Low-skilled*	*High-skilled*
Higher *HD* in $t = 1$	0.0134***	0.0138***	0.0201***
	(0.0013)	(0.0025)	(0.0017)
Observations	2,389,728	120,260	589,120

Notes: The treatment is defined as an increase in the hierarchical distance within a given business group. The estimation includes a time dummy and worker fixed-effects. Standard errors in parentheses are clustered at the establishment-level. ***, ** and * denote significance at the 1, 5 and 10 percent levels, respectively.

The first column of Table 4 summarizes the results for the pooled sample of all workers. In line with the results of Section 4.1, we find that a larger hierarchical distance to the ultimate owner of the business group increases workers' wages by 1.3 percent. According to our theoretical model, this result indicates that monitoring efficiency decreases with higher hierarchical distance. Consequently, ultimate owners have to increase incentive payments for workers employed by establishments at comparably low layers in the business group hierarchy.

Lack of information on monitoring effort prevents a direct test of the specific mechanism for explaining the positive link between larger hierarchical distance and workers' wages put forward by our theoretical model. However, we can use the observation of Jones (1984, p. 689) that "[i]n general, the more unstructured or ambiguous the task and the more specialized the skills of the job incumbent, the greater will be the difficulty of measuring performance" as an argument that the positive effect of hierarchical distance on wages should be strongest for workers with high skills and workers performing non-routine tasks since their performance is most difficult to observe. Accordingly, we interpret evidence for this argument as indirect support for the monitoring-based mechanism in our model.

To show evidence for the first part of this argument, we split our sample and estimate the effect of hierarchical distance to the ultimate owner of the business group on workers' wages separately for the sub-groups of high- and low-skilled workers. For these workers, Jones (1984) provides a clear prediction regarding the costs to monitor their workplace performance, while the prediction is less clear for median-skilled workers for whom specialization might vary considerably across vocational degrees. Columns 2 and 3 of Table 4 report the results. There, we see that the effect of larger hierarchical distance on workers' wages is 50 percent higher for high-skilled than for low-skilled workers, with the difference being highly statistically significant. This result lends support to the monitoring-based theory of business group hierarchies outlined in Section 2. Showing evidence for (or against) the second part of the argument requires information on the task content of occupations, which we do not directly observe in our dataset. Hence, we have to rely on task data from a different source.

For Germany, the task content of occupations can be constructed from employment surveys conducted by the Federal Institute for Vocational Education and Training (BIBB) and the Federal Institute for Occupational Safety and Health (BAuA) every six to seven years since 1979. These surveys cover about 20,000 – 30,000 individuals in each wave and they provide detailed information on the tasks performed by the respondents in their workplaces. Since it is shown by Becker and Muendler (2015) that the task content of occupations varies considerably over time, we only use the survey information from 2012 and thus the year prior to the first observation period in the dataset. Based on this survey, we distinguish 10 broad task categories, such as *manufacture and produce goods*; *repair and maintain*; or *purchase, procure, and sell.* Following Spitz-Oener (2006), Gathmann and Schönberg (2010), and Becker et al. (2013), we then classify tasks as either routine or non-routine. In the Appendix, we provide a list of all tasks and their classification as either routine (three tasks) or non-routine (seven tasks).[15]

[15] The category of routine tasks comprises the three activities *manufacture and produce goods*; *measure, inspect, and control quality*; and *oversee and control machinery and technical processes.* These tasks are easily codifiable and their usage has thus been negatively affected by the diffusion of computers and recent trends of automation (see Spitz-Oener, 2006; Becker and Muendler, 2015; de Vries et al., 2020).

Based on our classification, we compute for each respondent in the survey the fraction of routine and non-routine tasks conducted. We then determine for the 136 occupations observed in our dataset a routineness and non-routineness index, by averaging the previously computed task fractions over all individuals reporting to be employed in that occupation. We finally label the 33 occupations as routine for which the share of routine tasks is above while the share of non-routine tasks is below the median of all occupations. Conversely, we label the 33 occupations as non-routine for which the share of routine tasks is below while the share of non-routine tasks is above the median of all occupations. The remaining 70 occupations cannot be classified as being predominantly routine or non-routine, which is the reason we exclude workers in these occupations from our estimations. We finally link this classification to our dataset, relying on occupation codes.

Table 5: Distance effect by predominant task

Dependent variable: Log daily wage	*All workers*	*Predominant tasks* *routine*	*non-routine*
Higher HD in $t = 1$	0.0134***	0.0089***	0.0170***
	(0.0013)	(0.0016)	(0.0019)
Observations	2,389,728	777,200	211,548

Notes: The treatment is defined as an increase in the hierarchical distance within a given business group. The estimation includes a time dummy and worker fixed-effects. Standard errors in parentheses are clustered at the establishment-level. ***, ** and * denote significance at the 1, 5 and 10 percent levels, respectively.

Table 5 shows the treatment effects for all workers as well as the two sub-groups of workers employed in routine and non-routine occupations. Contrasting Columns 2 and 3, we find a sizable difference in the distance effect between occupations with a predominant share of routine tasks and a predominant share of non-routine tasks, respectively. The hierarchical distance effect for workers employed in non-routine occupations is 1.7 percent and almost twice as high as for workers performing routine tasks who nevertheless receive a distance premium of 0.9 percent. These results further support the hypothesis of our monitoring-based theory of business groups that a larger hierarchical distance to the ultimate owner increases workers' wages.

4.3 Robustness checks

To complete our empirical analysis, we investigate in a final step whether our results are robust to changes in the treatment definition. In Table 6, Model (1), we set a threshold for the hierarchical distance of 0.25 and drop observations with changes in the hierarchical distance smaller than this threshold from the treatment group. With this refinement, we eliminate small changes in ownership shares that could be the results of mismeasurement. Table 6 shows that introducing a lower threshold for the hierarchical distance variable lowers the sample size considerably and reduces the treatment effect. However, it does not change our results qualitatively.[16] In Model (2), we drop all observations showing an increase in hierarchical distance larger than two. This makes the treatment group more homogeneous and ensures that our results are not driven by a small number of outliers. Introducing the upper bound does not affect the estimation result.

Table 6: Hierarchical distance and wages: alternative specifications

Dependent variable:	*All workers*			
Log daily wage	Model (1)	Model (2)	Model (3)	Model (4)
Higher *HD* in $t = 1$	0.0115***	0.0133***	0.0029**	0.0063***
	(0.0015)	(0.0012)	(0.0015)	(0.0007)
Observations	1,868,148	2,342,520	3,005,852	4,875,328

Notes: In Model (1), we confine the treatment to increases in the hierarchical distance by at least 0.25. In Model (2), we confine the treatment to increases in the hierarchical distance by at most two. In Model (3), we consider worker-establishment pairs that change their business group around the treatment period. In Model (4), we broaden the definition of treatment and control group, including worker-establishment pairs that stay in their business group as well as worker-establishment pairs that change their business group. All estimations include a time dummy and worker fixed-effects. Standard errors in parentheses are clustered at the establishment-level. ***, ** and * denote significance at the 1, 5 and 10 percent levels, respectively.

In two further exercises, we no longer restrict the analysis to worker-establishment observations that stay in the same business group around the treatment event. In Model (3), the treatment is defined as a change in the hierarchical distance of an establishment to its ultimate owner *when* changing the business group. This lowers the treatment effect by more than 50 percent. In Model (4), we use all worker-establishment observations. In this case, the treatment is defined by a change in the hierarchical distance to the ultimate owner, irrespective of whether the establishment changes its business group or not. Similar to Model (3), changing the definition of treated and untreated observations lowers the treatment effect considerably.

However, one should be cautious when contrasting the estimates from Model (3) and (4) in Table 6 with those from Table 4. First, by changing the definition of

[16] Increasing the threshold to 0.5 or 0.75 would further reduce sample size but not substantively change our results.

the treatment group, we have also changed the sample of untreated observations in the control group, hampering the comparison of parameter estimates. Second, we cannot rule out that the treatment effects in Models (3) and (4) of Table 6 capture at least partially the impact of changing the business group, which has been put forward to exert sizable wage effects in the context of multinational enterprises. Of course, changing the business group is not confined to foreign takeover in our analysis, so that the observed drop of the treatment effect does not contradict the existence of a foreign ownership wage premium.[17]

5 Conclusion

Although the largest companies covered by the Fortune 500 list are all organized as business groups (see Altomonte et al., 2018) and some of these groups generate yearly revenues higher than the GDP of entire economies, business groups and their effects on workers' wages have received surprisingly little attention in economic research so far. The main reason for the lack of research is that existing datasets do not provide the detailed information needed for such an analysis. Our paper contributes to the literature by merging firm-level data on ownership linkages in business groups with administrative worker and establishment data for Germany. This gives a unique dataset that allows analyzing in a systematic way how the position in business group hierarchy affects workers' wages. Since the ownership linkages are not one-directional, we propose a measure of hierarchical distance that acknowledges the complex structure of business groups in our data. In the pooled sample of all workers, we find clear evidence for a positive impact of larger hierarchical distance on wages.

To shield our estimates from selection bias, we consider in a further step a two-stage estimation approach, in which we first select a treatment and control group based on propensity-score matching and then estimate the effect of an increase in the hierarchical distance to the ultimate owner on workers' wages using a difference-in-difference estimator. The results from this more elaborate empirical approach are similar to those from OLS and fixed-effects regressions. Larger hierarchical distance exerts a positive effect on workers' wages, with the effect being remarkably robust to changes in the composition of the treatment group.

[17] In two extensions to Models (3) and (4), we have added a dummy for foreign takeover and its interaction term with the treatment indicator. In these extensions, which are available upon request, we find evidence for both a positive wage effect of a larger hierarchical distance to the ultimate owner and a positive wage effect of a foreign takeover.

Overall, our results speak for a sizable impact of larger hierarchical distance to the ultimate owner of a business group on workers' wages. In a parsimonious specification, we show that increasing hierarchical distance by one standard deviation or approximately one layer increases wages by almost two log points. This effect remains fairly stable when controlling for selection bias by matching similar worker-establishment pairs.

To explain the positive link between the hierarchical distance to the ultimate owner of the business group and workers' wages, we propose a monitoring-based theory of business group hierarchies. Lack of information on monitoring effort prohibits a direct test of the theoretical hypotheses derived from our model. However, the finding that the hierarchical distance effect is most pronounced for workers with high skills and workers performing non-routine tasks – whose performance is difficult to observe – indicates that monitoring inefficiency indeed provides a reasonable explanation for the positive effect of a larger hierarchical distance on workers' wages.

A Appendix

A.1 Balancing test for the matching procedure

Table A.1: Balancing test for the matching procedure with replacement

Variable	*Sample*	*Mean*		*Stand.*	*Bias*	*Normal.*
		Treated	*Control*	*bias %*	*reduction*	*diff.*
(a) Group characteristics						
Hierarchical distance	Unmatched	2.313	1.028	113.6		
Hierarchical distance	Matched	2.313	2.306	0.6	99.5	0.002
Group complexity	Unmatched	4.173	2.063	78.3		
Group complexity	Matched	4.173	4.043	4.8	93.8	0.029
(b) Establishment characteristics						
Log employment	Unmatched	6.010	6.003	0.3		
Log employment	Matched	6.010	5.904	5.8	-1583.4	0.042
Agriculture	Unmatched	0.000	0.003	-7.2		
Agriculture	Matched	0.000	0.000	0.1	98.2	0.003
Manufacturing	Unmatched	0.538	0.453	17.0		
Manufacturing	Matched	0.538	0.531	1.4	91.8	0.010
Mining, util. & constr.	Unmatched	0.059	0.082	-9.0		
Mining, util. & constr.	Matched	0.059	0.055	1.3	85.0	0.011
Retail & repair	Unmatched	0.211	0.236	-6.0		
Retail & repair	Matched	0.211	0.217	-1.3	77.8	-0.010
Finance & insurance	Unmatched	0.164	0.121	12.4		
Finance & insurance	Matched	0.164	0.183	-5.4	56.1	-0.036
Priv. & publ. services	Unmatched	0.028	0.105	-31.2		
Priv. & publ. services	Matched	0.028	0.014	5.7	81.7	0.070
Schleswig-Holstein	Unmatched	0.023	0.021	1.3		
Schleswig-Holstein	Matched	0.023	0.023	0.2	82.7	0.002
Hamburg	Unmatched	0.053	0.029	11.9		
Hamburg	Matched	0.053	0.046	3.3	72.0	0.022
Lower Saxony	Unmatched	0.080	0.087	-2.4		
Lower Saxony	Matched	0.080	0.073	2.5	-6.3	0.019
Bremen	Unmatched	0.015	0.012	2.0		
Bremen	Matched	0.015	0.012	2.2	-8.9	0.015
North Rhine-Westphalia	Unmatched	0.200	0.201	-0.2		
North Rhine-Westphalia	Matched	0.200	0.205	-1.2	-481.5	-0.009
Hesse	Unmatched	0.123	0.081	13.8		
Hesse	Matched	0.123	0.131	-2.6	81.4	-0.016
Rhineland-Palatinate	Unmatched	0.035	0.044	-4.7		
Rhineland-Palatinate	Matched	0.035	0.033	0.6	86.4	0.005
Baden-Württemberg	Unmatched	0.138	0.168	-8.4		
Baden-Württemberg	Matched	0.138	0.138	0.1	98.3	0.001
Bavaria	Unmatched	0.170	0.175	-1.4		
Bavaria	Matched	0.170	0.177	-1.9	-35.1	-0.013
Saarland	Unmatched	0.006	0.014	-7.8		
Saarland	Matched	0.006	0.010	-3.5	54.7	-0.028
Berlin	Unmatched	0.031	0.035	-2.4		
Berlin	Matched	0.031	0.032	-0.5	78.7	-0.004
Brandenburg	Unmatched	0.028	0.023	3.6		

Table A.1 – *continued from previous page*

Variable	*Sample*	*Mean*		*Stand.*	*Bias*	*Normal.*
		Treated	*Control*	*bias %*	*reduction*	*diff.*
Brandenburg	Matched	0.028	0.029	-0.5	87.3	-0.003
Mecklenburg West-Pomerania	Unmatched	0.012	0.013	-0.9		
Mecklenburg West-Pomerania	Matched	0.012	0.010	1.5	-66.4	0.011
Saxony	Unmatched	0.044	0.049	-2.6		
Saxony	Matched	0.044	0.041	1.4	48.0	0.010
Saxony-Anhalt	Unmatched	0.026	0.022	2.5		
Saxony-Anhalt	Matched	0.026	0.024	1.2	49.5	0.009
Thuringia	Unmatched	0.016	0.024	-5.7		
Thuringia	Matched	0.016	0.016	0.4	93.5	0.003
(c) Worker characteristics						
Female	Unmatched	0.229	0.253	-5.4		
Female	Matched	0.229	0.226	0.8	85.3	0.006
Age	Unmatched	42.9	42.9	0.3		
Age	Matched	42.9	42.8	0.5	-81.5	0.004
Low skilled	Unmatched	0.051	0.045	2.6		
Low skilled	Matched	0.051	0.051	0.0	98.9	0.000
Medium skilled	Unmatched	0.703	0.740	-8.3		
Medium skilled	Matched	0.703	0.692	2.6	68.9	0.018
High skilled	Unmatched	0.246	0.215	7.5		
High skilled	Matched	0.246	0.258	-2.8	63.3	-0.019
Log wage	Unmatched	4.915	4.819	20.5		
Log wage	Matched	4.915	4.929	-2.9	85.6	-0.021
Sample				*Mean bias*	*Median bias*	
Unmatched				12.6	5.7	
Matched				1.9	1.4	

Notes: All variables are measured in $t = 0$ and averaged at the worker-level in the treated and control group respectively.

A.2 Further robustness checks

To make sure that the positive effect of larger hierarchical distance on wages reported in Section 4.2 is robust to different specifications of the propensity-score matching, we report in Table A.2 the results for the pooled sample of all workers, relying on four alternatives to our main matching procedure. In Model (1), we match without replacement and find that this has a comparably small impact on the treatment effect. In Model (2), we add the difference in log establishment employment between period $t = -1$ and $t = 0$ as a further covariate in the probit model. This allows us to control for differences in the employment dynamics prior to the treatment. Adding this covariate somewhat reduces sample size and slightly lowers the treatment effect, while leaving unchanged the main insight from our baseline specification in Table 4. In Model (3), we replace the continuous log employment variable in the probit model by dummies for five establishment size categories. Thereby, we distinguish establishments with less than ten, between ten and 49, between 50 and 249, between 250 and 999, and with more than 1000 employees. Additionally, we control for employment dynamics prior to treatment by introducing two dummies equal to one if the establishment has either increased or decreased its workforce by at least three percent between $t = -1$ and $t = 0$ (with the omitted category referring to establishments with an absolute change in workforce size by less than three percent). This modification increases the estimated treatment effect.

Table A.2: The effect of an increase in hierarchical distance on wages

Dependent variable:	*All workers*			
Log daily wage	Model (1)	Model (2)	Model (3)	Model (4)
Higher *HD* in $t = 1$	0.141***	0.0127***	0.0146***	0.0064*
	(0.0012)	(0.0013)	(0.0014)	(0.0034)
Observations	2,389,728	2,372,852	2,389,800	2,615,608

Notes: In Model (1), we match without replacement. In Model (2) we take into account employment dynamics prior to treatment by additionally matching on the difference in log (establishment) employment between $t = -1$ and $t = 0$. In Model (3), we match on five establishment size categories and two dummy variables indicating an absolute change in log (establishment) employment between $t = -1$ and $t = 0$ of at least three percent. In Model (4), we define the treatment at the establishment level and match accordingly. All estimations include a time dummy and worker fixed-effects. Standard errors in parentheses are clustered at the establishment-level. ***, ** and * denote significance at the 1, 5 and 10 percent levels, respectively.

In the final robustness check of Model (4), we define the treatment at the establishment and not the worker level, thereby imposing the constraint that all workers from a given establishment in the treatment group are matched with workers from a single establishment of the control group. As expected, imposing the additional constraint lowers matching quality. Moreover, it reduces the estimated treatment effect by about 50 percent, while leaving the main insight from our empirical analysis that larger hierarchical distance increases wages intact.

A.3 Tasks and their classification as routine or non-routine

Table A.3: Routine and non-routine tasks

Tasks	Routine	Non-routine
Manufacture and produce goods	yes	
Measure, inspect, and control quality	yes	
Oversee and control machinery and technical processes	yes	
Repair and maintain; or entertain, accommodate, and prepare food		yes
Purchase, procure, and sell		yes
Organize, plan, and prepare (others' work)		yes
Train, teach, instruct, and educate		yes
Consult and inform		yes
Gather information, develop, research, and construct		yes
Apply legal knowledge		yes

Notes: Bibb-BAuA Employment Survey 2012. Classification of tasks as routine and non-routine according to Spitz-Oener (2006) and Becker et al. (2013).

References

Altomonte, C., G. Ottaviano, and A. Rungi (2018): "Business Groups as Knowledge-Based Hierarchies of Firms," *Working Paper*.

Altomonte, C. and A. Rungi (2013): "Business groups as hierarchies of firms: determinants of vertical integration and performance," *ECB Working Paper 1554*.

——— (2015): "Firm Performance and the Optimal Organization of Business Groups," *Working Paper*.

Antoni, M., K. Koller, M.-C. Laible, and F. Zimmermann (2018): "Orbis-ADIAB: From record linkage key to research dataset," *FDZ Methodenreport 04/2018*.

Antràs, P. and D. Chor (2013): "Organizing the Global Value Chain," *Econometrica*, 81, 2127–2204.

Antràs, P., D. Chor, T. Fally, and R. Hillberry (2012): "Measuring the Upstreamness of Production and Trade Flows," *American Economic Review: Papers and Proceedings*, 102, 412–416.

Antràs, P. and E. Helpman (2004): "Global Sourcing," *Journal of Political Economy*, 112, 552–580.

Antràs, P. and S. R. Yeaple (2014): "Multinational Firms and the Structure of International Trade," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 4, chap. 2, 55–130.

Baker, G., R. Gibbons, and K. J. Murphy (2002): "Relational Contracts and the Theory of the Firm," *The Quarterly Journal of Economics*, 117, 39–84.

Balsvik, R. and S. A. Haller (2010): "Picking "Lemons" or Picking "Cherries"? Domestic and Foreign Acquisitions in Norwegian Manufacturing," *The Scandinavian Journal of Economics*, 112, 361–387.

Barbosa, N. and H. Louri (2002): "On the Determinants of Multinationals' Ownership Preferences: Evidence from Greece and Portugal," *International Journal of Industrial Organization*, 20, 493–515.

Bastos, P., N. P. Monteiro, and O. R. Straume (2018): "Foreign Acquisition and Internal Organization," *Journal of International Economics*, 114, 143–163.

Becker, S. O., K. Ekholm, and M.-A. Muendler (2013): "Offshoring and the Onshore Composition of Tasks and Skills," *Journal of International Economics*, 90, 91–106.

Becker, S. O. and M.-A. Muendler (2015): "Trade and Tasks: An Exploration over Three Decades in Germany," *Economic Policy*, 30, 589–661.

Belenzon, S., T. Berkovitz, and L. A. Rios (2013): "Capital Markets and Firm Organization: How Financial Development Shapes European Corporate Groups," *Management Science*, 59, 1326–1343.

Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter (2018): "The Disappearing Large-Firm Wage Premium," *AEA Papers and Proceedings*, 108, 317–322.

Bloom, N., R. Sadun, and J. Van Reenen (2010): "Recent Advances in the Empirics of Organizational Economics," *Annual Review of Economics*, 2, 105–137.

Brown, C. and J. Medoff (1989): "The Employer Size-Wage Effect," *Journal of Political Economy*, 97, 1027–1059.

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015): "The Anatomy of French Production Hierarchies," *Journal of Political Economy*, 123, 809–852.

Caliendo, L. and E. Rossi-Hansberg (2012): "The Impact of Trade on Organization and Productivity," *The Quarterly Journal of Economics*, 127, 1393–1467.

Caliendo, M. and S. Kopeinig (2008): "Some Practical Guidance For The Implementation Of Propensity Score Matching," *Journal of Economic Surveys*, 22, 31–72.

Calvo, G. A. and S. Wellisz (1978): "Supervision, Loss of Control, and the Optimum Size of the Firm," *Journal of Political Economy*, 86, 943–952.

——— (1979): "Hierarchy, Ability, and Income Distribution," *Journal of Political Economy*, 87, 991–1010.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

Caroli, E. and J. Van Reenen (2001): "Skill-biased Organizational Change? Evidence from a Panel of British and French Establishments," *The Quarterly Journal of Economics*, 116, 1449–1492.

Chen, C. (2017): "Management Quality and Firm Hierarchy in Industry Equilibrium," *American Economic Journal: Microeconomics*, 9, 203–244.

Chen, C. and W. Suen (2019): "The Comparative Statics of Optimal Hierarchies," *American Economic Journal: Microeconomics*, 11, 1–25.

Colonelli, E., J. Tåg, M. Webb, and S. Wolter (2018): "A Cross-Country Comparison of Dynamics in the Large Firm Wage Premium," *American Economic Review: Papers and Proceedings*, 108, 323–327.

Costinot, A., J. Vogel, and S. Wang (2013): "An Elementary Theory of Global Supply Chains," *The Review of Economic Studies*, 80, 109–144.

Cravino, J. and A. A. Levchenko (2017): "Multinational Firms and International Business Cycle Transmission," *The Quarterly Journal of Economics*, 132, 921–962.

Dauth, W. and J. Eppelsheimer (2020): "Preparing the Sample of Integrated Labour Market Biographies (SIAB) for Scientific Analysis: A Guide," *Journal for Labor Market Research*, 54.

de Vries, G. J., E. Gentile, S. Miroudot, and K. M. Wacker (2020): "The Rise of Robots and the Fall of Routine Jobs," *Labour Economics*, 66, 101885.

Egger, H. and E. Jahn (2020): "Ownership and the multinational wage premium," *Applied Economics Letters*, 27, 422–425.

Egger, H., E. Jahn, and S. Kornitzky (2020): "Reassessing the foreign ownership wage premium in Germany," *The World Economy*, 43, 302–325.

Friedrich, B. U. (forthcoming): "Trade Shocks, Firm Hierarchies, and Wage Inequality," *Review of Economics and Statistics*.

Garicano, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874–904.

Garicano, L. and E. Rossi-Hansberg (2006): "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 121, 1383–1435.

Gathmann, C. and U. Schönberg (2010): "How General Is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 28, 1–49.

Girma, S. and H. Görg (2007): "Evaluating The Foreign Ownership Wage Premium Using a Difference-in-Differences Matching Approach," *Journal of International Economics*, 72, 97–112.

Girma, S., D. Greenaway, and K. Wakelin (2001): "Who Benefits from Foreign Direct Investment in the UK?" *Scottish Journal of Political Economy*, 48, 119–133.

Grossman, G. M. and E. Helpman (2003): "Outsourcing versus FDI in Industry Equilibrium," *Journal of the European Economic Association*, 1, 317–327.

Gumpert, A. (2018): "The organization of knowledge in multinational firms," *Journal of the European Economic Association*, 16, 1929–1976.

Helpman, E. (1984): "A Simple Theory of International Trade with Multinational Corporations," *Journal of Political Economy*, 92, 451–471.

Hijzen, A., P. S. Martins, T. Schank, and R. Upward (2013): “Foreign owned firms around the world: A comparative analysis of wages and employment at the micro-level,” *European Economic Review*, 60, 170–188.

Idson, T. L. and W. Y. Oi (1999): “Firm Size and Wages,” in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter and D. Card, Elsevier, vol. 3, Part B, chap. 33, 2165–2214.

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference for Statistics, Social and Biomedical Sciences*, Cambridge University Press.

Imbens, G. W. and J. M. Wooldridge (2009): “Recent developments in the econometrics of program evaluation,” *Journal of Economic Literature*, 47, 5–86.

Johnson, R. C. (2018): “Measuring Global Value Chains,” *Annual Review of Economics*, 10, 207–236.

Jones, G. R. (1984): “Task Visibility, Free Riding, and Shirking: Explaining the Effect of Structure and Technology on Employee Behavior,” *The Academy of Management Review*, 9, 684–695.

Khanna, T. and Y. Yafeh (2007): “Business Groups in Emerging Markets: Paragons or Parasites?” *Journal of Economic Literature*, 45, 331–372.

Klosterhuber, W., P. Lehnert, and S. Seth (2016): “Linked-Employer-Employee-Daten des IAB: LIAB-Querschnittsmodell 2 1993-2014 (LIAB QM2 9314),” *FDZ Datenreport 05/2016.*

Mincer, J. (1958): “Investment in Human Capital and Personal Income Distribution,” *Journal of Political Economy*, 66, 281–302.

Montag, F. and T. Wilson (2011): “Differences in German and EU Merger Control Law – Should a Common EU Rulebook for Mergers Follow the German or European Model?” 18th St.Gallen International Competition Law Forum ICF.

Qian, Y. (1994): “Incentives and Loss of Control in an Optimal Hierarchy,” *Review of Economic Studies*, 61, 527–544.

Rajan, R. G. and J. Wulf (2006): “The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies,” *The Review of Economics and Statistics*, 88, 759–773.

Rosenbaum, P. R. and D. B. Rubin (1983): “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41–55.

——— (1985): “Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score,” *The American Statistician*, 39, 33–38.

Spitz-Oener, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure," *Journal of Labor Economics*, 24, 235–270.

UNCTAD (2016): *World Investment Report 2016: Investor Nationality: Policy Challenges*, UNCTAD, Geneva.

Winter-Ebmer, R. and J. Zweimüller (1999): "Firm-Size Wage Differentials in Switzerland: Evidence from Job-Changers," *The American Economic Review*, 89, 89–93.