The Organization of Knowledge in Multinational Firms

Anna Gumpert (University of Munich)

Discussion Paper No. 29

April 25, 2017

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

The organization of knowledge in multinational firms

Anna Gumpert*

April 2017

Abstract

This paper provides the first in-depth study of the organization of knowledge in multinational firms. In the theory, knowledge is a costly input for firms that they can acquire at their headquarters or their production plants. Communication costs impede the access of the plants to headquarter knowledge. The model shows that multinational firms systematically acquire more knowledge at both their foreign and domestic plants than non-multinationals if their foreign plants face higher communication costs with headquarters than their domestic plants. This theoretical prediction helps understand why multinational firms pay higher wages to workers than non-multinational firms, and why their sales decrease across space. The empirical analyses show that higher communication costs indeed decrease multinational firms' foreign sales. Consistent with model-specific comparative statics, the decrease is stronger in sectors with less predictable production processes. Novel data on corporate transferees allow shedding light on one tool of multinational firms' organization of knowledge.

JEL codes: D21, D24, F21, F23.

Keywords: multinational firm, knowledge hierarchy, organization, geography of FDI, multinational wage premium, corporate transferees.

*LMU Munich, Seminar for Comparative Economics, Akademiestrasse 1, 80799 Munich, Germany, and CESifo, e-mail: anna.gumpert@econ.lmu.de. I am grateful to Pol Antràs and Esteban Rossi-Hansberg for helpful discussion and suggestions throughout the development of this paper, and to Monika Schnitzer for her constant encouragement and support. I thank Lorenzo Caliendo, Kerem Cosar, Arnaud Costinot, Nadja Dwenger, Carsten Eckel, Teresa Fort, Stefania Garetto, Inga Heiland, Elhanan Helpman, James R. Hines Jr., Wolfgang Keller, Michael Koch, Wilhelm Kohler, Chen Li, Yanping Liu, James Markusen, Marc Melitz, Eduardo Morales, Raffaella Sadun, Heiner Schumacher, Ludwig Straub, Uwe Sunde, Alexander Tarasov, Felix Tintelnot, Martin Watzinger, Stephen Yeaple, and seminar participants at Harvard University, LMU Munich, MIT, Princeton University, Yale University, as well as the participants of various conferences for their comments. I am grateful to Darren Lubotsky for sharing the code of Lubotsky and Wittenberg (2006). I thank the CESifo Network for the 2014 Distinguished CESifo Affiliate Award in the Area Global Economy and the RIEF Network for their best paper prize. This paper was partly written while I was a visitor at the Economics Department of Harvard University. I thank the Economics Department for its hospitality, and am particularly thankful to Pol Antràs for making my visit possible. I conducted the empirical analyses during visits to the research center of the Deutsche Bundesbank and gratefully acknowledge the hospitality of the Bundesbank, the constructive support of the staff, as well as the access to its Microdatabase Direct investment (MiDi). The project has benefited from financial support through the German Science Foundation (DFG) under GRK 801, and through the German National Academic Foundation (Studienstiftung des deutschen Volkes) under the ERP Fellowship Programme. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged.

1 Introduction

In today's economy, knowledge is an essential production factor. Knowledge is a costly input for firms because it is typically tacit and employees have to acquire it through costly learning. Production processes are complex and involve many different employees. The efficient organization of knowledge is therefore a key ingredient for firms' success. It determines which employees specialize in which part of the production process, and to whom they turn for help if they encounter a problem that they are not able to solve. Firms organize knowledge to match the problems that arise in production to the employees with the knowledge to solve them, taking into account both the costs of learning and the costs of communication between employees.

Communication costs are an important, but understudied determinant of the organization of knowledge in firms. Existing papers assume that the communication costs are constant throughout a firm, so searching for help is equally costly for all employees. This assumption is a good approximation for the interaction of employees in small firms, active at a single location. However, it is likely to be overly simplistic in the study of large firms with production plants in different locations, and it certainly does not apply to multinational firms, a very important subgroup of firms.[1] Multinational firms have headquarters in their home country that communicate with plants in the home and in foreign countries. The communication costs between the headquarters and the plants vary across countries. Language barriers, time zone differences, and lack of face-to-face interaction render cross-border communication within a multinational firm more difficult than communication within a domestic firm. Such communication frictions impede the diffusion of knowledge within multinational firms and hamper the access of foreign plants to headquarter knowledge. Yet, the question of how multinational firms optimally organize knowledge in the presence of heterogeneous communication costs is so far unexplored.

This paper provides an in-depth analysis of the organization of knowledge in multinational firms. I develop a theory to show that heterogeneous communication costs in multinational firms lead to systematic differences between the optimal organization of knowledge in multinational and non-multinational firms. These differences explain both the geographic distribution of sales and investments of multinational firms and the emergence of multinational firm wage premiums. Prior theories explain only either of the two stylized facts. The empirical analyses confirm model-specific predictions on the impact of communication costs on the foreign sales of multinational firms. Novel data on the flows of corporate transferees between countries show that the use of a specific tool for knowledge transfer by multinational firms is also consistent with the model.

Specifically, I construct a stylized model of multinational firms in the spirit of the knowledge hierarchies framework (e.g., Antràs et al., 2006; Caliendo and Rossi-Hansberg, 2012; Garicano, 2000). In this framework, production involves labor and knowledge. The labor input generates problems that are solved using knowledge to produce output. I assume that the total knowledge level of firms is exogenously given and heterogeneous. The higher its total knowledge level

[1] To illustrate, less than 1% of U.S. manufacturing firms are multinationals, but they account for a third of manufacturing output and 26% of manufacturing employment (Bernard and Jensen, 2007).

is, the more output a firm can produce per unit of labor. The prerequisite is that the firm's employees learn the knowledge. Firms consist of two layers: managers in the domestic corporate headquarters and workers in production plants that can be located in the domestic country, or the domestic and a foreign country. The managers and workers can communicate and leverage differences in their knowledge. Firms endogenously choose the number of managers and workers, as well as the proportion of the total knowledge that they learn. The organization of knowledge yields endogenous marginal production costs. Due to the heterogeneity of the total knowledge, the marginal production costs are heterogeneous across firms. To derive the consequences of the organization of knowledge for firm behavior, I embed the model of the organization of knowledge in a heterogeneous firm model of foreign direct investment (FDI) similar to Helpman et al. (2004). Firms choose whether to serve the foreign country through exporting or FDI. The model yields predictions on firms' sales and their self-selection into FDI.

Three results summarize the main insights on the optimal organization of knowledge in multinational firms. First, the optimal knowledge level at a plant increases with the communication costs between the plant and the headquarters. A multinational firm thus assigns more knowledge to its foreign plant than to its domestic plant to avoid the higher cross-border communication costs. Foreign plants master a higher share of the production process by themselves and approach the headquarters for help less frequently than domestic plants. The increase of plant knowledge is the stronger, the less predictable the production process is.

Second, multinational firms assign less knowledge to their headquarters than if they were non-multinational firms (i.e., purely domestic firms or exporters). This result is more than the inverse of the first statement: it stems from multinationals' balancing the costs of headquarter knowledge and its utilization in domestic and foreign production. Foreign plants use headquarter knowledge less frequently than if they were domestic plants due to their higher knowledge level. Consequently, the utilization rate of headquarter knowledge in multinational firms is lower than if they were not multinational. As providing knowledge at headquarters is costly, a multinational firm chooses to maintain a lower level of knowledge at its headquarters to balance its utilization rate and its costs.

Third, the lower level of knowledge at the headquarters of a multinational firm also affects its domestic production plants: Multinational firms assign more knowledge to their domestic plants than non-multinational firms. Multinationals' headquarters have less knowledge than the headquarters of non-multinational firms, so multinationals' domestic plants have to learn more knowledge to ensure the efficiency of production. The knowledge level of a multinational's domestic plants is typically still lower than the knowledge level of the foreign plants, so the optimal knowledge levels at the different plants of a multinational firm are heterogeneous.

How does the organization of knowledge help us understand the nature of multinational production? The optimal organization of knowledge yields endogenous marginal production costs that depend on the total knowledge level of the firm and home and foreign country characteristics. It thus helps explain distinct stylized facts concerning multinational firms. A special feature of multinational firms is that their marginal costs are interdependent across countries. This result arises because the foreign and the domestic production plant share common headquarters. In consequence, and consistent with the empirical evidence (Antràs and Yeaple, 2014; Tomiura, 2007), multinational and non-multinational firms with the same marginal costs endogenously coexist in the home and the foreign country, unlike in models that assume firms to be heterogeneous in productivity.

It is well-known and empirically documented that multinational firms pay higher wages to their production workers than equally productive domestic firms (so-called "residual multinational firm wage premiums", see, e.g., Aitken et al., 1996). The organization of knowledge helps explain the residual multinational wage premiums: Multinationals assign more knowledge to their production plants than non-multinationals with the same marginal costs, and this knowledge is remunerated. The wage premiums vary with home and foreign country characteristics because these affect the organization of knowledge. The model thus explains why multinational wage premiums depend on the nationality of the acquirer (as found by Girma and Görg, 2007). The self-selection of firms into FDI reinforces the wage premiums.

Likewise, it is a well-known stylized fact that the foreign sales and investment probability decrease with the distance of a country from a multinational's home country (e.g., Antràs and Yeaple, 2014). The organization of knowledge provides a novel explanation for this empirical regularity. The endogenous marginal costs increase with the communication costs between a foreign plant and the headquarters of a multinational firm. The increase is the stronger, the less predictable the production process is. Foreign sales and the probability of foreign entry correspondingly decrease with the communication costs, that are correlated with geographic distance. The organization of knowledge thus helps understand distinct features of multinational firms' behavior that have hitherto been analyzed separately in the literature.

Is there evidence for the model in the data? Providing direct evidence is difficult as knowledge is intangible and typically proprietary. Knowledge flows within multinational firms are very hard to observe. Neither data on the organization of multinationals nor data on their wage payments across different countries are available. To overcome this problem, I exploit the model's predictions on multinational firms' foreign marginal costs that are reflected in the foreign sales. Using comprehensive firm-level data for German multinational firms, I show that German multinationals have lower sales in countries that are characterized by higher communication costs with Germany, as measured by the overlap in office hours, linguistic proximity, communication technology and flight time. This finding is robust to controlling for firm heterogeneity and to including geographic distance, trade cost measures as well as further determinants of foreign sales, e.g., the quality of the investment climate. To show that firms' organization of knowledge drives this effect, I use the model prediction that the impact of the communication costs varies with the predictability of the production process. I construct a new measure of the predictability of the production process in a sector and study how the predictability of the production process interacts with the communication costs.[2] Consistent with the model, the negative impact of higher communication costs on sales is stronger in sectors with

[2]This type of strategy has been employed by Rajan and Zingales (1998), Keller and Yeaple (2013) and others.

a less predictable production process. This result supports that the effect of commmunication costs on foreign sales reflects the organization of knowledge in multinational firms.

To shed light on specific strategies that multinational firms use to organize knowledge across countries, I use unique data on the flows of corporate transferees between countries. Corporate transferees are employees who multinational firms transfer from their regular place of work to one of their units in another country for a limited period of time. Multinationals use corporate transferees predominantly to transfer know-how (e.g., Djanani et al., 2003). To the best of my knowledge, I am the first to exploit corporate transferees as a visible reflection of firms' organization of knowledge.[3] I find that the proportion of corporate transferees in the employment of multinationals systematically increases with the communication costs between two countries, in line with the model's predictions.

Though the paper focuses on multinational firms, the insights on the optimal organization of knowledge apply more generally. They are transferable to situations where different groups of agents collaborate with one group of experts at varying collaboration costs. Such situations may arise in many contexts, including production networks with several plants within a country or the organization of the public administration.

The paper contributes to several strands of literature. First, the paper adds to the literature on firms as knowledge hierarchies (Garicano, 2000; for a survey, see Garicano and Rossi-Hansberg, 2015). Within this literature, the paper is closest to that of Antràs et al. (2006), who study the formation of cross-country teams, a form of vertical FDI, and to the work on the organization of exporters by Caliendo and Rossi-Hansberg (2012). To the best of my knowledge, this paper is the first to study heterogeneity in the communication costs within firms and to show that this heterogeneity can be useful to understand the specific features of the behavior of firms with several plants.

Second, the paper contributes to the understanding of multinational firm wage premiums (Harrison and Rodríguez-Clare, 2010, and Malchow-Moller et al., 2013, survey the empirical literature).[4] By focusing on the particular features of the organization of knowledge, the paper proposes an explanation that is specific to multinationals and distinct from the scale-based arguments related to exporter wage premiums. The paper thus adds to prior explanations based on fair wage preferences (Egger and Kreickemeier, 2013) or positive assortative matching (Davidson et al., 2014).

Third, the paper contributes to the literature on the role of headquarter inputs for local affiliate production (Keller and Yeaple, 2013; Irarrazabal et al., 2013). Previous papers in this literature focus on the geography of FDI and extend the framework in Helpman et al. (2004) to incorporate productivity-shifting mechanisms. This paper is distinct in modeling the organization of multinational firms. It introduces a novel angle to the study of multinational firms as it endogenously determines how firms adjust the characteristics of their headquarters

[3] More recently, Astorne-Figari and Lee (2016) study knowledge hierarchies with corporate transferees data.

[4] Many papers document that affiliates of multinational firms pay higher wages than domestic firms. The wage premium tends to be higher in developing than in developed countries (e.g., Aitken et al., 1996; Hijzen et al., 2013). Worker heterogeneity does not fully explain the wage premium (Malchow-Moller et al., 2013). Likewise, multinational parent companies pay higher wages than domestic firms (Heyman et al., 2007).

to their mode of internationalization.[5] The paper thus provides a coherent rationale for both the geography of FDI and multinational firm wage premiums.

Fourth, the paper adds a theoretical perspective to a series of predominantly empirical papers showing that communication costs inhibit investments by multinational firms (Bahar, 2016; Cristea, 2015; Defever, 2012; Oldenski, 2012). Relatedly, Fort (2017) studies the impact of communication technology on the fragmentation of production processes.

Finally, the paper contributes to the literature on the spatial diffusion of knowledge (for a survey, see Keller, 2004). Investments by multinational firms are an important channel of international knowledge diffusion (e.g. Arnold and Javorcik, 2009; Harrison and Rodríguez-Clare, 2010). This paper highlights that spatial communication frictions have a substantial impact on multinational firms. Consequently, investment promotion policies should not only improve the business climate inside a country, but also reduce communication costs with source countries of FDI. Improving language training, investing in the communication infrastructure and other targeted measures to facilitate bilateral communication may prove useful in attracting FDI and thus bringing new technologies to a country.

The following section develops the model of the organization of knowledge and constitutes the core of the paper. Section 3 derives the model implications for multinationals' sales and the probability of investment, as well as their wage setting behavior. Section 4 contains the empirical evidence concerning the geography of multinational firms' sales, and explains how the predictability of the production process helps infer the organization of knowledge. Section 5 uses data on corporate transferees to shed light on a specific tool for the organization of knowledge. Section 6 discusses the relation of the organization of knowledge and a monitoring based model. The last section concludes.

2 The optimal organization of knowledge

2.1 Set up

The model economy consists of two countries, the home country $j = 0$ and the foreign country $j = 1$. The countries are populated by N_j agents each endowed with one unit of time. The analysis abstracts from capital market and contractual imperfections for clarity.

Establishing firms. Agents choose between supplying their time in the labor market and being entrepreneurs. An entrepreneur in the home country hires f units of labor in the domestic labor market to pay the entry cost and establish a firm. The entry cost is thereafter sunk. Upon paying the entry cost, each entrepreneur receives the blueprint of a differentiated product, a level of knowledge $\bar{z}_i$ and the option to establish a corporate headquarters. The knowledge level $\bar{z}_i$ corresponds to the state of a firm's technology. Mathematically, knowledge is an interval

[5] Earlier papers assume that headquarter services are public goods, i.e., foreign affiliates can use them without additional investment (e.g., Helpman et al., 2004; Keller and Yeaple, 2013; Irarrazabal et al., 2013), or study the impact of constraints to the managerial capacity or span of control (e.g., Ramondo, 2014; Yeaple, 2013). Ethier and Horn (1990) study adjustments to managerial capacity, but in a monitoring hierarchy.

ranging from zero to a firm-specific upper bound $\bar{Z}_i$. $\bar{z}_i$ denotes the length of a knowledge interval $[0, \bar{Z}_i]$ (i.e., its Lebesgue measure). Knowledge levels $\bar{z}_i$ follow a known distribution $G(\bar{z})$, which is symmetric in the two countries. The entrepreneur does not know how to employ the knowledge in production—he would have to learn it first—but can assess whether the knowledge interval is large or small. Given this assessment, the entrepreneur decides whether to establish headquarters and produce, or instead to provide his time in the labor market.

If the entrepreneur decides to set up a corporate headquarters and produce, he spends his unit of time providing leadership services in the headquarters. He decides whether to sell in the domestic country, the foreign country, or both, and whether to set up a production plant only at home or in both countries. He determines the number of employees in the headquarters and the production plant(s) and the organization of knowledge. These activities capture non-rival and non-delegable headquarter services similar to those in Markusen (1984) and the subsequent multinational firm literature. The entrepreneur receives the market wage as well as profits.

To study the differences between the optimal organization of knowledge of domestic and multinational firms (MNEs) in a transparent manner, I restrict the parameter space so the entrepreneur always finds it optimal to hire employees in the headquarters and the production plant(s) (see Appendix A). All firms thus consist of the headquarters and at least one production plant. The n_h employees hired in the headquarters are called managers and the n_j employees working in the production plant in country j are called workers.[6] To simplify the exposition, section 2 focuses on a single firm established to produce output using the knowledge level $\bar{z}$. Section 3 extends the analysis to many firms indexed by i with heterogeneous knowledge levels $\bar{z}_i$.

Producing output. Production is a problem solving process based on labor and knowledge (as in Caliendo and Rossi-Hansberg, 2012; Garicano, 2000). For each unit of labor employed in production, problems are realized with a mass 1. Transforming labor into output requires that the problems be solved. An agent solves a problem if it is realized within his knowledge interval. The problems are distributed according to an exponential probability distribution function:

$$f(z) = \lambda e^{-\lambda z}$$

where $z \in [0, \infty)$ refers to the domain of possible problems and $\lambda > 0$ denotes the predictability of the production process. A higher value of λ implies that the mass of the probability distribution is concentrated close to zero. This means that the production process is more predictable as problems in the tail of the probability distribution occur with lower probability, so more output can be produced with a given amount of labor and knowledge.

[6]The entrepreneur obtains the option to set up headquarters by paying the sunk costs of entry. Setting up headquarters with managers in the foreign country would entail an equally high fixed costs, so managers are hired only in the domestic country in optimum. It is possible to extend the model to a three-layer structure with headquarter managers, intermediate managers and workers without altering the main results. In this case, the firm could hire intermediate managers in the foreign plant. As the additional implications of such an extension are not testable with the available data, the paper builds on a two-layer structure.

The output q_j of n_j units of labor input with knowledge $\bar{z}$ can be calculated as n_j times the value of the cumulative distribution function:

$$q_j = n_j(1 - e^{-\lambda \bar{z}}).$$

Learning and communicating. The firm's knowledge $\bar{z}$ is only useful if its employees learn it. The underlying idea is that employees have to know how to employ production technologies to use them fruitfully. The knowledge can be learned by workers or managers. The entrepreneur fully uses her time to provide leadership services. Learning knowledge is costly: Employees have to hire teachers to train them. Teachers spend $c_j z_k$ units of time to train an employee to learn a knowledge interval of length z_k, $k = h, j$. In equilibrium, all agents receive the market wage w_j per unit of time they spend working. Correspondingly, employees pay teachers the remuneration $w_j c_j z_k$. The entrepreneur remunerates his employees for the time they spend in production and for their learning expenses (as in Caliendo and Rossi-Hansberg, 2012).

The workers and the managers can communicate and leverage the potentially different knowledge levels. Communication is costly. Analogous to the results for firms with a single production plant in Garicano (2000), only workers supply labor and managers use their time solely for communication because this specialization makes it possible to achieve the optimal utilization rate of costly knowledge. As is standard in the literature (e.g., Bolton and Dewatripont, 1994; Garicano, 2000), the managers bear the communication costs: they have to spend time listening. The communication costs, i.e., the amount of time that a manager spends listening, depends on whether the workers are located in the same or another country. A manager in country j spends $\theta_{kj} \geq 0$ units of time listening to workers in country k. The assumption that $\theta_{10} > \theta_{00}$ and $\theta_{11} = \theta_{00}, \theta_{01} = \theta_{10}$ captures the fact that there are frictions in cross-border communication compared to communication within a country.

Organizing knowledge. The entrepreneur designs the optimal organization of knowledge, i.e., he decides which part of the firm's knowledge is learned by the workers and which part is learned by the managers. The production process thus works as follows. During each unit of time that they spend in production, the workers immediately solve the problems realized in their knowledge interval and produce output. The workers communicate all problems that are not covered by their knowledge interval to the managers. The managers solve all problems covered by their knowledge interval. Any problems that are not covered by the knowledge intervals of either the workers or the managers remain unsolved.[7]

Both workers and managers are optimally characterized by knowledge levels that are uniform within each group and different between the two groups. Uniform knowledge levels reduce

[7]The model applies to production processes in which workers execute routine tasks and consult an expert if non-routine problems arise. For example, manufacturing firms teach workers the common features of machines, but employ experts for exceptional applications. In the context of service firms, Garicano and Hubbard (2007) show that the framework describes how law firms split tasks between associates and lawyers. Bloom et al. (2012) and Bloom et al. (2014) study the allocation of decisions, such as the purchase of equipment, through the lens of knowledge hierarchy models. Some of their evidence is consistent with this model's predictions.

communication time by diminishing the time spent searching for a competent contact. Workers know that only managers may know solutions to problems that they themselves cannot solve, and that it does not matter which manager they approach. To minimize the probability that costly communication is necessary, the knowledge level of workers covers the solutions to more frequently occurring problems, whereas managers know the solutions to problems that occur more rarely (Garicano, 2000). The knowledge interval of workers correspondingly starts at 0, where the mass of the problem density is highest, and ranges to an endogenous country specific upper bound Z_j, $j = 0, 1$. z_j denotes the length of the knowledge interval of workers $[0, Z_j]$. The managers learn to solve infrequent problems. Under the parameter restrictions imposed above (see Appendix A), it is never optimal that the employees do not learn part of the firm's knowledge interval $[0, \bar{Z}]$. More knowledge enables the firm to produce more output with a given amount of labor input and thus decreases marginal costs. The upper bound of managerial knowledge and the upper bound of the knowledge interval of the firm coincide. The knowledge interval of managers ranges from a lower bound Z_h to $\bar{Z}$. z_h denotes the length of this interval $[Z_h, \bar{Z}]$.[8] The entrepreneur chooses the knowledge level(s) z_j and z_h as well as the number of workers n_j and managers n_h. By choosing z_j and z_h, the firm determines the upper bound of the workers' knowledge interval(s) Z_j and the lower bound of managerial knowledge Z_h.

2.2 The optimization problem

The entrepreneur chooses the optimal organization of knowledge to minimize the production costs. The costs consist of the cost for personnel at the production plant(s) and at the headquarters, as well as the entrepreneurial wage. Each employee is remunerated with the market wage w_j per unit of time spent working for the firm and for the learning expenses $w_j c_j z_k$, $k = h, j$.

The cost minimization problem applies to an MNE, and comprises a domestic firm as special case. The entrepreneur optimally chooses the number of workers $\{n_j\}_{j=0}^{1}$, their country specific knowledge level $\{z_j\}_{j=0}^{1}$, the number of managers n_h, and the managerial knowledge level z_h. Figure 1 illustrates the optimization problem.

$$C(\bar{z}, q_0, w_0, q_1, w_1) = \min_{\{n_j, z_j\}_{j=0}^{1}, n_h, z_h} \sum_{j=0}^{1} n_j w_j (1 + c_j z_j) + n_h w_0 (1 + c_0 z_h) + w_0 \tag{1}$$

$$\text{s.t.} \quad n_j(1 - e^{-\lambda \bar{z}}) \geq q_j \qquad \forall j \tag{2}$$

$$z_j \geq \bar{z} - z_h \quad \forall j \tag{3}$$

$$n_h \geq \sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} \tag{4}$$

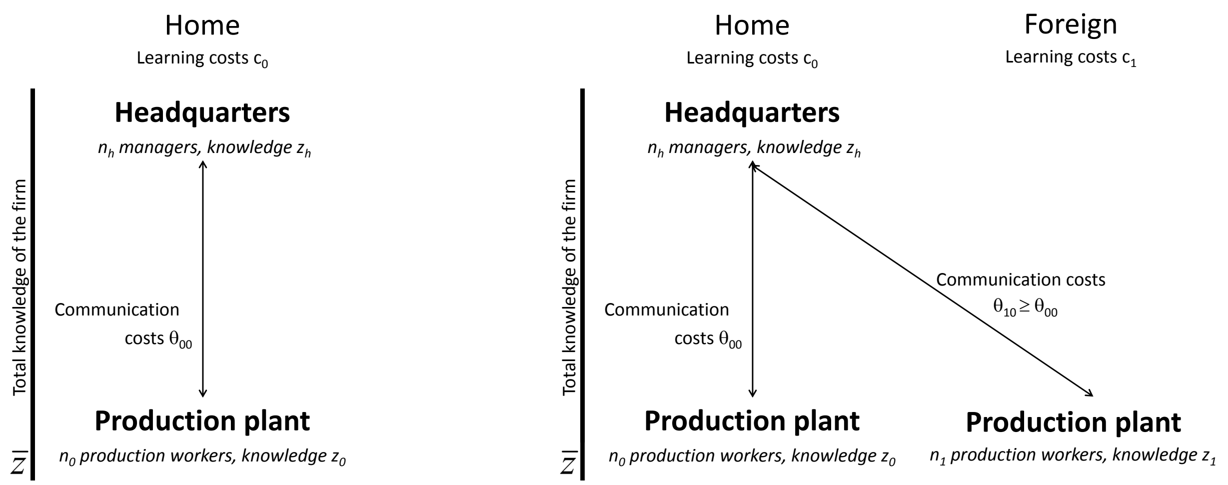
$$n_h \geq 0, \; z_h \geq 0, z_h \leq \bar{z} \tag{5}$$

$$n_j \geq 0, \; z_j \geq 0, z_j \leq \bar{z} \qquad \forall j \tag{6}$$

[8] All managers have the same knowledge z_h to capture the fact that managers have to address problems brought to them from anywhere in the corporation. This is true at least at some level of seniority even in large MNEs that have separate specialized divisions at their headquarters.

Figure 1: Optimization problem

Figure 1a: Domestic firm Figure 1b: MNE



The figures illustrate the optimization problem of the domstic firm and the MNE. Endogenous variables are in italics, exogenous variables in roman.

The production quantities $\{q_j\}_{j=0}^1$ are taken as given in the cost minimization problem, but they are endogenized in subsection 3.1. Wages $\{w_j\}_{j=0}^1$ are endogenized in subsection 3.2. The predictability of the production process λ, communication costs $\{\theta_{j0}\}_{j=0}^1$, and learning costs $\{c_j\}_{j=0}^1$ are positive exogenous parameters determined by the predictability of the production process and the geography and institutions of a country.[9]

When choosing $\{n_j\}_{j=0}^1$, n_h, $\{z_j\}_{j=0}^1$ and z_h, the entrepreneur faces four types of constraints:

Eq. (2): The firm has to produce a total output $n_j(1 - e^{-\lambda\bar{z}})$ of at least q_j units.

Eq. (3): The managers or the workers have to learn the firm's knowledge. This is ensured if the workers' knowledge level z_j and the managers' knowledge level z_h add up to at least the knowledge level of the firm $\bar{z}$.

Eq. (4): The entrepreneur has to hire a sufficient number of managers such that the managers are able to listen to all problems brought to them. The number of problems sent by each plant is calculated as the mass of problems generated through labor input n_j times the probability that the solution is not found by the workers in j, $e^{-\lambda z_j}$. This term is multiplied by the communication costs θ_{j0}.

Eq. (5, 6): All choice variables are restricted to be positive. Employees' knowledge cannot exceed the total knowledge of the firm.

Equation (3) indicates that overlaps between managerial knowledge and the knowledge of workers may occur. This is specific to MNEs. In domestic firms, overlaps cannot be optimal: The overlap of managerial and workers' knowledge increases costs, but remains unused at the

[9]Endogenizing the total knowledge level $\bar{z}$ is possible and the main results go through, but at the expense of a more complicated and less transparent analysis. A note on the model with endogenous total knowledge is available from the author upon request.

headquarters (Garicano, 2000). If the firm has two plants, overlaps between the knowledge at one plant and managerial knowledge may occur as long as the overlapping managerial knowledge is used to solve problems communicated by the workers from the other plant.[10]

The Lagrangian equation is given by

$$\mathcal{L} = \sum_{j=0}^{1} n_j w_j (1 + c_j z_j) + n_h w_0 (1 + c_0 z_h) + w_0 + \sum_{j=0}^{1} \xi_j \left[q_j - n_j (1 - e^{-\lambda \bar{z}}) \right] + \sum_{j=0}^{1} \phi_j \left[\bar{z} - z_h - z_j \right]$$
$$+ \kappa \left[\sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} - n_h \right] - \sum_{j=0}^{1} \upsilon_j n_j - \upsilon_h n_h - \sum_{j=0}^{1} \nu_j z_j - \nu_h z_h + \sum_{j=0}^{1} \bar{\nu}_j (z_j - \bar{z}) + \bar{\nu}_h (z_h - \bar{z}).$$

The Lagrangian multiplier ξ_j denotes the marginal costs of production. κ captures the marginal costs of using the headquarters. The other multipliers do not have intuitive interpretations. Appendix B.1 contains the first order conditions.

The optimal number of workers is determined by the quantity constraint (2):

$$n_j = \frac{q_j}{1 - e^{-\lambda \bar{z}}}.$$

The optimal number of managers results from the constraint on the number of managers (4):

$$n_h = \sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} = \sum_{j=0}^{1} \frac{q_j \theta_{j0} e^{-\lambda z_j}}{1 - e^{-\lambda \bar{z}}}.$$

Both n_j and n_h are positive for positive values of q_j.

The knowledge levels of the workers $\{z_j\}_{j=0}^{1}$ may differ due to asymmetries in the country characteristics. The knowledge constraint (3) is binding for at least one country:

$$z_j = \bar{z} - z_h. \tag{7}$$

If the knowledge constraint is non-binding for both countries, the overlap of managerial knowledge and workers' knowledge remains unused. This cannot be optimal.

If the knowledge constraint is non-binding in one country, the optimal knowledge level of the workers is determined by

$$e^{-\lambda z_j} = \frac{w_j c_j}{\lambda \theta_{j0} w_0 (1 + c_0 z_h)}. \tag{8}$$

Both z_j are positive by $z_j \geq \bar{z} - z_h$. $z_j < \bar{z}$ because otherwise, communication with the headquarters is not worthwhile. The characteristics of the country with the binding constraint $z_{\bar{j}} = \bar{z} - z_h$ and the non-binding constraint $z_{\hat{j}} > \bar{z} - z_h$ are related as follows:

$$\theta_{\bar{j}0} w_{\hat{j}} c_{\hat{j}} < \theta_{\hat{j}0} w_{\bar{j}} c_{\bar{j}}.$$

The knowledge constraint is, ceteris paribus, more likely to be binding in the home country due

[10] In principle, gaps between managerial knowledge and the knowledge of workers may also occur. Knowledge gaps render the analysis analytically less tractable, so they are treated in Appendix B.4.

to the lower communication costs, and in the country with higher wages and learning costs.[11]

Only firms with a sufficiently high knowledge level $\bar{z}$ choose asymmetric knowledge levels of workers. The savings due to less frequent communication with the headquarters have to outweigh the cost increase due to higher worker knowledge levels. This is more likely for higher $\bar{z}$, because managerial knowledge increases with $\bar{z}$ (see subsection 2.3). More asymmetric country characteristics also render asymmetric knowledge levels more likely (see Appendix B.1).

The managerial knowledge of a firm with two production plants is implicitly determined by

$$\sum_{j=0}^{1} [\mathbf{1}(z_j > \bar{z} - z_h) n_j \theta_{j0} e^{-\lambda z_j} w_0 c_0 + \\ \mathbf{1}(z_j = \bar{z} - z_h) n_j \left(\theta_{j0} e^{-\lambda(\bar{z} - z_h)} w_0 (c_0 + \lambda(1 + c_0 z_h)) - w_j c_j\right)] = 0. \tag{9}$$

The indicator function $\mathbf{1}(\cdot)$ determines whether the constraint $z_j = \bar{z} - z_h$ is binding.

If the firm only produces in the domestic country, z_0, n_0 and n_h are determined by the constraints (2)-(4) with $n_1 = 0$. Managerial knowledge is implicitly defined by

$$\theta_{00} e^{-\lambda(\bar{z} - z_h)} (c_0 + \lambda(1 + c_0 z_h)) - c_0 = 0. \tag{10}$$

The first order conditions (9) and (10) equate the marginal benefit and the marginal costs of z_h. The marginal benefit consists of the savings in the learning costs of the workers, $n_0 w_0 c_0$, or, for an MNE, $\sum_{j=0}^{1} \mathbf{1}(z_j = \bar{z} - z_h) n_j w_j c_j$. The marginal costs are composed of the costs of increasing managerial knowledge, $n_0 \theta_{00} e^{-\lambda(\bar{z} - z_h)} w_0 c_0$, or $\sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} w_0 c_0$, and the increase in the number of managers, $n_0 \theta_{00} e^{-\lambda(\bar{z} - z_h)} \lambda w_0 (1 + c_0 z_h)$, or $\sum_{j=0}^{1} \mathbf{1}(z_j = \bar{z} - z_h) n_j \theta_{j0} e^{-\lambda(\bar{z} - z_h)} \lambda w_0 (1 + c_0 z_h)$. The number of workers and wages drop from equation (10).

A comparison of equations (9) and (10) shows that the optimal organization of knowledge systematically differs in domestic firms and MNEs. The knowledge levels in a domestic firm depend only on variables that are exogenous to the firm. They are independent of the production quantity. In contrast, an MNE takes the production quantity into account in allocating knowledge. As is shown in subsection 3.1, an MNE organizes in such a way that results in greater cost reduction for a plant the larger its output.

The marginal costs of production consist of the product of inverse labor productivity $\frac{1}{1-e^{-\lambda \bar{z}}}$ and the personnel costs at the production plant and the headquarters per unit of labor input:

$$\xi_j = \frac{1}{1 - e^{-\lambda \bar{z}}} \left[w_j (1 + c_j z_j) + w_0 (1 + c_0 z_h) \theta_{j0} e^{-\lambda z_j} \right]. \tag{11}$$

[11]This results by $\frac{w_{\hat{j}} c_{\hat{j}}}{\lambda \theta_{\hat{j}0} w_0 (1 + c_0 z_h)} = e^{-\lambda z_{\hat{j}}} \leq e^{-\lambda(\bar{z} - z_h)}$ and $e^{-\lambda(\bar{z} - z_h)} \leq \frac{w_{\bar{j}} c_{\bar{j}}}{\lambda w_0 (1 + c_0 z_h) \theta_{\bar{j}0}}$ by $\phi_{\bar{j}} \geq 0$ if $z_{\bar{j}} = \bar{z} - z_h$.

2.3 The comparative statics results

Proposition 1. *The optimal knowledge levels vary with the characteristics of the location(s) of the production plant(s) $\{\theta_{j0}, c_j, w_j\}_{j=0}^1$, the production quantities $\{q_j\}_{j=0}^1$, the total knowledge $\bar{z}$, and the predictability of the production process λ as follows:*

Table 1: Comparative statics

Knowledge levels/ model parameters	θ_{j0}	c_j	w_j	q_j	$\bar{z}$	λ
Workers' knowledge z_0, domestic firm	+	-	0	0	+	+/-
Workers' knowledge z_j, MNE, $z_0 = z_1 = \bar{z} - z_h$	+	-*	-*	+/-	+	+/-
Workers' knowledge z_j, MNE, $z_j = \bar{z} - z_h$	+	-	-	-	+**	+/-
Workers' knowledge z_j, MNE, $z_j > \bar{z} - z_h$	+	-	-	-	+	+/-
Managerial knowledge z_h, domestic firm	-	+	0	0	+	+/-
Managerial knowledge z_h, MNE, $z_0 = z_1 = \bar{z} - z_h$	-	+*	+*	+/-	+	+/-
Managerial knowledge z_h, MNE, $z_j = \bar{z} - z_h$	-	+	+	+	+	+/-
Managerial knowledge z_h, MNE, $z_j > \bar{z} - z_h$	0	-	-	-	+	+/-

The table displays the effects of the model parameters on the optimal knowledge levels. + denotes positive effects, $-$ negative effects, $+/-$ ambiguous effects and 0 no relation. Results denoted * only apply to $j = 1$. Results denoted ** hold if $q_{\bar{j}}\theta_{\bar{j}0}e^{-\lambda(\bar{z}-z_h)}\lambda(1+c_0 z_h) > q_{\hat{j}}\theta_{\hat{j}0}e^{-\lambda z_{\hat{j}}}c_0$, where the constraint $z_j = \bar{z} - z_h$ is binding in $\bar{j}$ and slack in $\hat{j}$. Appendix B.2 contains the results for the number of workers n_j and managers n_h.

Proof. See Appendix B.2. □

The optimal organization of knowledge varies with the characteristics of the home and foreign countries. The firm may be domestic or multinational. In this case, the knowledge constraint may be binding at both plants, or binding at one and slack at the other plant. The country characteristics generally have similar effects on the organization of knowledge in the different cases. I explain the comparative statics results by model parameter for the different cases in the order in which they appear in Table 1.

Most importantly, higher communication costs θ_{j0} always increase the knowledge level of workers z_j to reduce the number of problems that need to be communicated to the headquarters. Managerial knowledge z_h decreases in the communication costs if the knowledge constraint $z_j = \bar{z} - z_h$ is binding, and is independent of the communication costs if it is slack.

Higher learning costs c_j increase the remuneration for every worker, so it is optimal to reduce the knowledge they hold to mitigate cost increases. Correspondingly, managerial knowledge increases in the learning costs, except if the knowledge constraint is not binding. This result may seem counterintuitive at first. If the knowledge level of workers decreases, the number of problems sent to headquarters increases. This entails an incentive to reduce the marginal costs of using the headquarters $w_0(1 + c_0 z_h)$, which is achieved by decreasing managerial knowledge. This is possible as the knowledge constraint is not binding.

Higher wages w_j decrease the knowledge level of workers and affect managerial knowledge in an MNE for the same reasons.[12]

[12]If the knowledge constraint is binding at both plants, the comparative statics only apply to foreign workers' knowledge. Managerial knowledge decreases in domestic wages. The domestic workers' knowledge level thus increases. The domestic learning costs have an ambiguous effect on managerial and domestic workers' knowledge.

If a larger quantity q_j is to be produced, more workers need to be hired, each of whom receives $w_j(1 + c_j z_j)$. An MNE can mitigate this cost increase by adjusting the optimal organization of knowledge within its organization.[13] The production quantity does not affect the workers' optimal knowledge level in a domestic firm. An increase in the production quantity leads to a proportional increase in the number of workers, which causes a proportional increase in the number of managers. Similarly, wages scale the total costs of production. The effect of learning costs and communication costs is different. The entrepreneur faces a trade-off also if he produces at a single location: Assigning more knowledge to the workers increases the costs at the production plant, but decreases the costs that accrue due to communication between workers and managers.[14]

The knowledge level of the workers and the knowledge level of managers both increase with the total knowledge of the firm $\bar{z}$. The predictability of the production process λ has an ambiguous effect on the knowledge level of workers and managers. A higher value of λ decreases the probability that the workers do not find the solution to a problem for a given value of z_j. This sets an incentive to reduce workers' knowledge to save costs. At the same time, a higher value of λ implies that the number of managers responds more strongly to changes in z_j. More managers need to be hired if z_j is decreased, which dampens the negative effect of λ on z_j.

The predictability of the production process has an unambiguous effect on the rate at which the knowledge of production workers increases with the communication costs θ_{j0}.

Corollary 1. *The increase of the foreign workers' knowledge level z_1 with the communication costs θ_{10} is the stronger, the less predictable the production process (i.e. the lower λ) is.*

Proof. See Appendix B.2. □

Intuitively, if the production process is less predictable, fewer problems are solveable with a given amount of knowledge. A firm thus has to assign even more knowledge to the workers to mitigate the effect of higher communication costs on the number of problems sent to the headquarters.

Taking the first order conditions for managerial knowledge (9) and (10) and the comparative statics together reveals that the optimal level of managerial knowledge in an MNE is systematically different from the optimal managerial knowledge in a domestic firm.

Proposition 2. *If $\theta_{00} w_1 c_1 < \theta_{10} w_0 c_0$, a firm with a given level of knowledge $\bar{z}$ systematically chooses a lower level of managerial knowledge when it is multinational than when it is domestic.*

Proof. See Appendix B.3. □

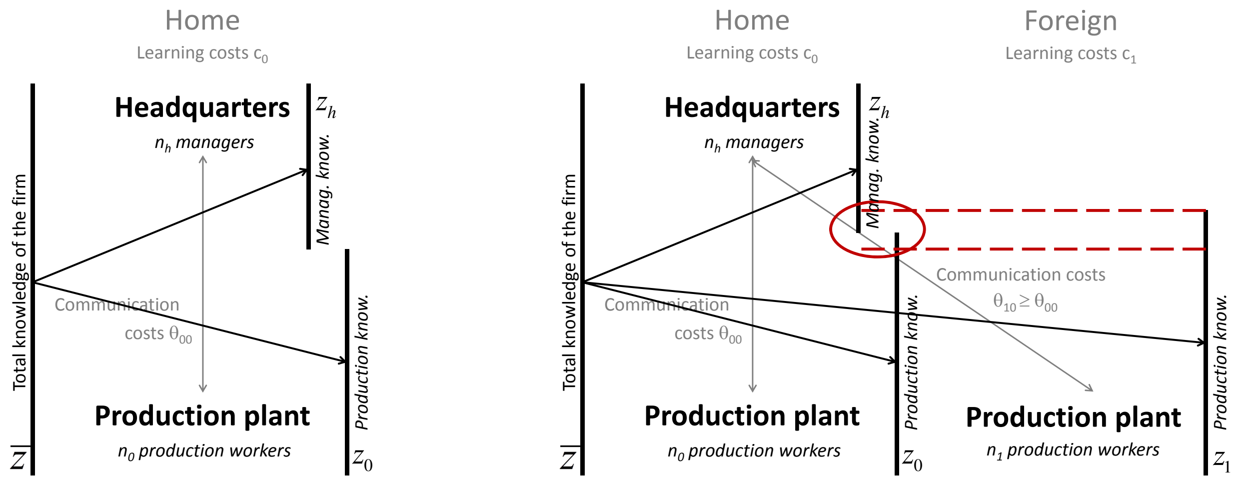
Figure 2 illustrates Propostion 2. It displays the optimal organization of knowledge of a firm with a given total amount of knowledge $\bar{z}$ as a domestic firm and an MNE. Intuitively,

[13] MNEs with asymmetric worker knowledge levels always decrease the workers' knowledge z_j when q_j increases. MNEs with symmetric knowledge levels decrease the workers' knowledge if z_j is the country with the higher ratio of $\frac{w_j c_j}{\theta_{j0}}$ and increase it otherwise. They thereby reorganize towards asymmetric workers' knowledge.

[14] The results for domestic firms correspond to the results derived in Bloom et al. (2014).

Figure 2: Optimal organization of knowledge in MNEs vs. domestic firms

Figure 2a: Domestic firm Figure 2b: MNE



The figure illustrates the optimal organization of knowledge in a domstic firm and an MNE. Endogenous variables are in italics, exogenous variables in roman. Country characteristics are in grey to put emphasis on the organization of knowledge. The red marks highlight the overlap of the knowledge of the MNE's foreign production workers and the domestic firm's managers, as well as the adjustments to the managerial and domestic production knowledge in the MNE compared to the domestic firm.

a firm chooses a lower level of managerial knowledge if it is an MNE than if it is a domestic firm to ensure an efficient utilization rate of knowledge. As Figure 2 shows, the workers in the MNE's foreign plant have higher levels of knowledge than the workers in the domestic firm's plant because of the higher cross-border communication costs. The knowledge of the foreign production workers and the optimal managerial knowledge in the domestic firm overlap. Foreign workers in an MNE thus turn to headquarters for help less frequently than the workers in a domestic firm. This decreases the utilization rate of managerial knowledge. At the same time, managerial knowledge is equally costly for a domestic firm and an MNE. An MNE consequently decreases the amount of managerial knowledge to balance its utilization rate and its costs. In consequence, the MNE assigns more knowledge to its domestic production workers to mitigate the negative effect of lower managerial knowledge on domestic production.

In summary, section 2 shows that the optimal organization of knowledge in a firm differs with the firm's multinational status. An MNE assigns systematically higher levels of knowledge to its workers and systematically lower levels of knowledge to its managers to avoid the higher communication costs with the foreign market. An MNE may choose asymmetric knowledge levels for its domestic and foreign workers. Its organization of knowledge depends on the foreign and domestic production quantities, whereas the production quantity does not influence the organization of knowledge in domestic firms.

3 The implications for MNEs' foreign sales and wages

3.1 Foreign sales and the self-selection of firms into FDI

The analysis of the choice between domestic activity, exporting, and FDI focuses on firms in the home country $j = 0$, and analogously applies to firms in the foreign country $j = 1$. There are many monopolistically competing firms in both countries (similar to Helpman et al., 2004). Each firm i produces a distinct variety and is characterized by its firm-specific knowledge $\bar{z}_i$.

Consumers have symmetric CES preferences:

$$U(x_j(\bar{z})) = \left(\int_{\Omega_j} x_j(\bar{z}_i)^{\frac{\sigma-1}{\sigma}} M_j \mu(\bar{z}) d\bar{z} \right)^{\frac{\sigma}{\sigma-1}}, \tag{12}$$

where Ω_j is the set of varieties available in country j, M_j is the mass of firms, $\mu(\bar{z})$ denotes the density of knowledge levels of the firms in country j, $\sigma > 1$ is the elasticity of substitution and $x_j(\bar{z}_i)$ is the individual consumption level in country j of the variety produced by firm i with knowledge input $\bar{z}_i$. The set of varieties Ω_i, the mass of firms M_j and the density of their knowledge levels $\mu(\bar{z})$ are determined in general equilibrium in the next subsection.

The total demand is given by the population N_j multiplied by the individual demands: $q_j(\bar{z}_i) = N_j x_j(\bar{z}_i)$. Utility maximization subject to the individual's budget constraint yields the demand function for product i:

$$p_j(\bar{z}_i) = q_j(\bar{z}_i)^{-\frac{1}{\sigma}} Q_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}}, \tag{13}$$

Q_j is the consumption basket in country j and P_j denotes the price index. I normalize the domestic price index P_0 to 1.

Each entrepreneur chooses the location(s) of the production plant(s) and the production quantities to maximize profits. The location decision affects the optimal organization of knowledge, so each choice is associated with distinct endogenous marginal production costs. Each option entails fixed costs in units of domestic labor. Firms can sell their output in the home country at fixed costs f^D ("domestic firms"). With additional fixed costs f^X, "exporters" ship output to the foreign country. To ship output from country k to country $j \neq k$, the firm incurs iceberg transport costs $\tau > 1$. MNEs serve consumers from two local plants at fixed costs $f^D + f^I$, i.e. they conduct "horizontal FDI".[15] I assume that $f^I > \tau^{\sigma-1} f^X > \frac{Q_1 P_1^{\sigma-1}}{Q_0} f^D$. It is thus never optimal to export but not to serve the domestic market.

The entrepreneur first determines the optimal production quantities and then chooses the location(s) of the production plant(s) associated with the maximum resulting profits. In what follows, optimal quantities are characterized by the mode, using the superscripts D for domestic firms, X for exporters, and I for MNEs. The quantities q_0, q_1 in section 2 comprise potential exports, i.e., $q_0 \in \{q_0^D, q_0^X + \tau q_1^X, q_0^I\}$ and $q_1 \in \{\tau q_0^V + q_1^V, q_1^I\}$.

[15] Appendix D derives the results on the geography of sales and the MNE wage premiums for vertical FDI.

Production quantities and sales. The profit maximization problem for FDI is given by

$$\max_{q_0^I, q_1^I \geq 0} \quad \pi^I(\bar{z}_i, w_0, w_1) = \sum_{j=0}^{1} p_j(q_j^I(\bar{z}_i)) q_j^I(\bar{z}_i) - C(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1). \tag{14}$$

Optimal prices are a constant mark-up over marginal costs:

$$p_j(\bar{z}_i) = \frac{\sigma}{\sigma - 1} \xi_j(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1).$$

The marginal costs ξ_j are a function of $\{q_j^I\}_{j=0}^1$ through z_h and z_j. The optimal quantities are thus implicitly defined by

$$q_j^I(\bar{z}_i) = Q_j P_j^{\sigma-1} \left(\frac{\sigma}{\sigma - 1} \xi_j(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1) \right)^{-\sigma}. \tag{15}$$

The entrepreneur analogously maximizes profits of exporting:

$$\max_{q_0^X, q_1^X \geq 0} \quad \pi^X(\bar{z}_i, w_0) = \sum_{j=0}^{1} p_j(q_j^X(\bar{z}_i)) q_j^X(\bar{z}_i) - C(\bar{z}_i, q_0^X(\bar{z}_i) + \tau q_1^X(\bar{z}_i), w_0). \tag{16}$$

Optimal prices are a constant mark-up over marginal costs, including transport costs τ where applicable. The marginal costs are constant. The optimal quantities are given by

$$q_0^X(\bar{z}_i) = Q_0 \left(\frac{\sigma}{\sigma - 1} \xi_0(\bar{z}_i, w_0) \right)^{-\sigma}; \qquad q_1^X(\bar{z}_i) = Q_1 P_1^{\sigma-1} \left(\frac{\sigma}{\sigma - 1} \tau \xi_0(\bar{z}_i, w_0) \right)^{-\sigma}. \tag{17}$$

The optimal production quantity of a domestic firm is determined by similar considerations.

Optimal quantities vary by mode. As is well-known, an exporter sells larger quantities in the domestic country than in the foreign country by $\tau > 1, \sigma > 1$, so concentrating production in one location is more profitable the lower the transport costs τ.

Quantities sold domestically by an MNE are lower than domestically sold quantities would be if the firm produced only domestically:

$$q_0^D(\bar{z}_i) = q_0^X(\bar{z}_i) \geq q_0^I(\bar{z}_i). \tag{18}$$

This result arises because the entrepreneur cannot tailor the headquarters of an MNE to the domestic plant. Correspondingly, domestic profits are lower in the case of FDI than in the case of exporting or domestic activity.

The higher fixed costs and the sales foregone in case of FDI are only worthwhile if the foreign production quantities in case of FDI exceed foreign export quantities:

$$q_1^I(\bar{z}_i) > q_1^X(\bar{z}_i). \tag{19}$$

Comparative statics. The optimal quantities of an MNE vary with country characteristics. Equation (15) indicates that the optimal quantities vary negatively with the marginal costs of production that depend on foreign country characteristics through the organization of knowledge. However, the relationship between foreign country characteristics and the optimal production quantities is complex. The complexity arises because the marginal costs depend on the domestic and foreign production quantities due to their effect on the optimal organization of knowledge. An MNE organizes knowledge in a way that favors plants with larger output: the larger the output of a plant j, the lower the marginal costs ξ_j at the expense of higher marginal costs ξ_k, $k \neq j$. That is, the foreign marginal costs ξ_1 decrease with the foreign production quantity q_1^I and increase with the domestic production quantity q_0^I. The analogous result holds for the domestic marginal costs ξ_0. This adjustment has to be taken into account in determining the effect of country characteristics on production quantities.

Proposition 3. *The foreign marginal costs $\xi_1(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1)$ of MNEs increase with the communication costs θ_{10}. In consequence, the foreign production quantities and sales are generally lower in countries with higher communication costs θ_{10}.*

Proof. See Appendix C.1.1. □

Higher communication costs increase the foreign marginal costs of production. This exerts a direct negative effect on foreign output. As the output affects the optimal organization of knowledge, higher communication costs also have an indirect positive effect on the foreign marginal costs of production. The entrepreneur adjusts the organization of knowledge due to the lower foreign production quantity, so the foreign marginal costs increase even further, depressing foreign output and foreign sales.[16]

Communication frictions between two countries arise due to foreign languages, time zone differences, or weak communication infrastructure. Some of these factors are correlated with the geographic distance between two countries. The negative effect of the communication costs between the home and the foreign country on the foreign sales thus provides a novel explanation for the stylized fact that MNEs' foreign sales decrease with the distance between the foreign country and the home country of the MNE (e.g., Antràs and Yeaple, 2014, Sec. 2).

Importantly, the model predicts that the impact of the communication costs varies across sectors.

Corollary 2. *The less predictable the production process (i.e. the lower λ) is, the stronger is the increase of the foreign marginal costs $\xi_1(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1)$ with the communication costs θ_{10} if $z_1 > \bar{z} - z_h$.*

Proof. See Appendix C.1.2. □

This comparative statics result stems from two factors. First, as corollary 1 shows, the increase of the knowledge of production workers with the communication costs is the higher,

[16]The indirect adjustment through the production quantities lead to an analytically ambiguous overall effect only for symmetric workers' knowledge levels and $w_1 c_1 \theta_{00} < w_0 c_0 \theta_{10}$. The effect is always negative in simulations.

the lower the predictability of the production process is. This affects the marginal production costs because they are increasing in the knowledge of production workers. Second, the lower predictability of the production process reduces the output per unit of labor input. The marginal production costs inversely depend on the product of labor, so this decrease also leads to a higher increase of the marginal production costs with the communication costs if λ is lower.

Although it is possible to analytically derive Corollary 2 only for $z_1 > \bar{z} - z_h$ due to the non-linearity of the model, the simulation results in Figures C.1-C.4 in the Appendix show that it holds more generally. They also show that the result transfers to MNEs' foreign sales: the decrease of the foreign sales with the communication costs is the stronger, the less predictable the production process is.

It is more difficult to determine the impact of the foreign learning costs and wages on the optimal foreign production quantities of MNEs because it is not possible to determine their effect on the foreign marginal costs of production in an unambiguous manner. Foreign wages w_1 and learning costs c_1 have a positive direct effect on the foreign marginal costs of production, but they also affect the organization of knowledge. These adjustments generally work against the direct positive effect, i.e., they decrease the marginal costs. The total effect of foreign wages and learning costs on the marginal costs is thus analytically ambiguous.

Investment decision. Given the optimal production quantities, the entrepreneur chooses the production mode (D, X, I) with the maximum total net profits.

FDI affects the organization of knowledge and thus the marginal production costs. Unlike previous models of horizontal FDI (e.g., Helpman et al., 2004), the marginal production costs are interdependent across countries. Domestic marginal costs are affected by the decision to set up a foreign plant, so total net profits—domestic and foreign net profits—have to exceed the total net profits of exporting.

The choice between exporting and purely domestic activity only depends on whether the foreign variable export profits exceed the fixed costs of exporting. The firm produces additional output without adjusting its organization, so domestic profits are not affected.

3.2 Aggregate exports and foreign sales

The general equilibrium analysis determines how the frictions in cross-border communication affect the aggregate export and foreign investment flows between countries through MNEs' organization of knowledge. For simplicity, I assume that the foreign and the domestic country are symmetric with respect to the learning costs $c_1 = c_0 = c$ and the population $N_1 = N_0 = N$. This implies that equilibrium outcomes are symmetric in both countries. Firms either sell their product only domestically, or export it to the foreign market, or conduct FDI, and have to incur the fixed costs associated with each of these options. f^D, f^X and f^I are such that domestic firms, exporters and foreign investors co-exist. Each entrepreneur draws the blueprint of a differentiated product and a firm-specific knowledge level $\bar{z}_i$ upon paying the sunk entry costs f. The knowledge levels follow a known distribution $G(\bar{z})$ that is defined for an interval

$[\bar{z}_{min}, \bar{z}_{max}]$ determined by the parameter restriction in Appendix A.

The general equilibrium conditions determine the symmetric cut-off knowledge levels for activity $\bar{z}^*$, for exporting $\bar{z}^X$, and for FDI $\bar{z}^I$, the mass of firms M, wages w, and total income Q. The domestic price index is normalized to unity, so the foreign price index is equal to one: $P_1 = P_0 = 1$. The parameters λ, c, θ_{kj} and N are exogenous. Appendix C.2 contains the proofs.

Three *zero-cut-off profit conditions* describe how firms self-select into the different options, based on the results of subsection 3.1. The least productive active firm is indifferent between producing domestically and remaining inactive: its variable profits are equal to the fixed costs of production f^D. The first *zero cut-off profit condition* determines the knowledge level $\bar{z}^*$ of the marginal entrant as a function of wages w.

$$wf^D = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\xi_0(\bar{z}^*, w)^{1-\sigma} - w \tag{20}$$

The density of the knowledge levels of the active firms is $\mu(\bar{z}) = \frac{g(\bar{z})}{1-G(\bar{z}^*)}$. The marginal exporter is indifferent between exporting and not exporting: the variable foreign export profits are equal to the fixed costs of exporting. The second *zero cut-off profit condition* determines the exporting cut-off $\bar{z}^X$.

$$wf^X = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q(\tau\xi_0(\bar{z}^X, w))^{1-\sigma} - w \tag{21}$$

The marginal MNE is indifferent between exporting and FDI. The net total export profits of exporting are equal to the net total profits earned from FDI. The multinational cut-off $\bar{z}^I$ is determined by the third *zero cut-off profit condition*:

$$\begin{aligned}
&\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\left(\xi_0(\bar{z}^I, q_0^I(\bar{z}^I), w, q_1^I(\bar{z}^I), w)^{1-\sigma} + \xi_1(\bar{z}^I, q_0^I(\bar{z}^I), w, q_1^I(\bar{z}^I), w)^{1-\sigma}\right) - wf^I = \\
&\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\,\xi_0(\bar{z}^I, w)^{1-\sigma}(1+\tau^{1-\sigma}) - wf^X
\end{aligned} \tag{22}$$

Entrepreneurs enter up to the point at which the net value of entry is zero. The *free entry condition* is given by[17]

$$\begin{aligned}
wf = &\int_{\bar{z}^*}^{\bar{z}^I} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\xi_0(\bar{z}, w)^{1-\sigma} - w(1+f^D)dG(\bar{z}) \\
&+ \int_{\bar{z}^X}^{\bar{z}^I} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q(\tau\xi_0(\bar{z}, w))^{1-\sigma} - w(1+f^X)dG(\bar{z}) \\
&+ \int_{\bar{z}^I}^{\bar{z}_{max}} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\left(\xi_0(\bar{z}, q_0(\bar{z}), w, q_1(\bar{z}), w)^{1-\sigma} + \xi_1(\bar{z}, q_0(\bar{z}), w, q_1(\bar{z}), w)^{1-\sigma}\right) \\
&-w(1+f^D+f^I)dG(\bar{z})
\end{aligned} \tag{23}$$

[17] The free entry condition assumes a unique cut-off knowledge level for FDI. As both the export and FDI profits are strictly increasing and concave in $\bar{z}$, several cut-offs may exist. The results hold, but are less tractable.

The *goods market clearing condition* determines the total income Q.

$$wN = Q \tag{24}$$

The *labor market clearing condition* determines the number of firms M. Labor is used to cover the sunk cost of entry, the fixed costs of production, exporting and FDI, and the demands for labor in production, management, and teaching. Labor demand for production, management, and teaching can be calculated by setting wages equal to 1 in the cost function $C(\bar{z}, \cdot)$.

$$\begin{aligned} N =& \frac{M}{1 - G(\bar{z}^*)} \left(f + \left(\int_{\bar{z}^*}^{\bar{z}^I} f^D + C(\bar{z}, q \in \{q_0^D, q_0^X + \tau q_1^X\}, 1) dG(\bar{z}) + \int_{\bar{z}^X}^{\bar{z}^I} f^X dG(\bar{z}) \right.\right. \\ &+ \left.\left. \int_{\bar{z}^I}^{\bar{z}_{max}} f^D + f^I + C(\bar{z}, q_0, 1, q_1, 1) dG(\bar{z}) \right) \right) \end{aligned} \tag{25}$$

By symmetry, the *trade balance condition* is fulfilled.

Inspection of the zero-cut-off profit conditions shows that MNEs have a higher knowledge level $\bar{z}_i$ than exporters, which in turn are more knowledgeable than domestic firms: $\bar{z}^I > \bar{z}^X > \bar{z}^*$. Manipulation of equations (20) and (21) permits to derive

$$\xi_0(\bar{z}^X, w) = \left(\frac{f^D}{f^X} \right)^{\frac{1}{\sigma - 1}} \frac{1}{\tau} \xi_0(\bar{z}^*, w) \quad < \xi_0(\bar{z}^*, w),$$

so exporters have lower marginal production costs than domestic firms, as in Melitz (2003). As the marginal costs $\xi_0(\bar{z}_i, w)$ strictly decrease with $\bar{z}_i$, $\bar{z}^X > \bar{z}^*$ results.

$\bar{z}^I > \bar{z}^X$ results because the fixed costs of FDI are higher than the fixed costs of exporting by a factor of more than $\tau^{\sigma-1}$, so only firms with a higher knowledge level carry out FDI profitably. Domestic profits decrease in the case of FDI as the headquarters are no longer tailored to domestic needs but balance domestic and foreign requirements. Compared to a model with independent marginal costs, the marginal costs cut-off is thus shifted downwards.

The self-selection of firms on knowledge is perfect: two firms with the same level of knowledge never make different investment decisions. Due to the reorganization of knowledge in MNEs, MNEs and exporters with the same marginal costs of production may, however, coexist. A firm organizes differently when it is an exporter than when it is a multinational. The domestic marginal costs of the firm as a MNE exceed its marginal costs as an exporter because the MNE cannot tailor its headquarters to the domestic plant. In consequence, the marginal costs of MNEs immediately above the cut-off knowledge level $\bar{z}^I$ are as high as the marginal costs of exporters immediately below it. The coexistence of MNEs and exporters with the same marginal costs is consistent with the empirical evidence (Antràs and Yeaple, 2014; Tomiura, 2007) and different from the predictions of standard models of FDI (e.g. Helpman et al., 2004).

To determine the effect of the communication frictions on aggregate export and FDI flows, it is necessary to consider the zero cut-off profit conditions together with the free entry condition. The model parameters have a direct effect on the export and FDI knowledge cut-offs, and an indirect effect through wages. Given wages, an increase in the transportation costs τ leads to an increase in the exporting cut-off knowledge level $\bar{z}^X$ and a decrease of the knowledge cut-off $\bar{z}^I$, because foreign export profits decrease. $\bar{z}^I$ increases with the communication costs θ_{10} between the home and the host country as the communication costs decrease profits from FDI. Equilibrium wages decrease both with higher transport costs τ and communication costs θ_{10}, because they decrease the net value of entry. The decrease in wages dampens the increase in the export knowledge cut-off with transport costs and the increase of the FDI cut-off with communication costs. It amplifies the negative effect of higher transport costs on the FDI cut-off, and leads to a decrease in the export cut-off with higher communication costs. In sum, the export knowledge cut-off thus increases with transport costs and decreases with the communication costs, and the minimum knowledge level required for foreign investment decreases with the transport costs and increases with the communication costs. MNEs' aggregate foreign sales thus increase relative to aggregate exports if the transport costs rise, and decrease with higher communication costs.

3.3 Multinational wage premiums

In addition to the results for the geography of MNEs' investments, the model is consistent with empirical evidence on MNE wage premiums (e.g., Harrison and Rodríguez-Clare, 2010; Heyman et al., 2007): MNEs are predicted to pay higher remuneration to workers than non-MNEs both in the home and the foreign countries.[18] The prediction results from the assumption that the knowledge of employees is remunerated (as in Caliendo and Rossi-Hansberg, 2012). The wage premiums arise via two channels: due to firm organization and due to a selection effect.

Empirical studies typically compare MNEs and non-MNEs with the same observable characteristics, such as productivity or sales, and find that MNEs pay higher wages than similar non-MNEs. In the model, MNEs and non-MNEs with the same marginal costs and sales endogenously coexist in the home and the foreign country. This result arises due to the differences in the optimal organization of knowledge between MNEs and non-MNEs and despite the fact MNEs and non-MNEs differ with respect to their total level of knowledge.

As outlined in section 2, given a knowledge level $\bar{z}$, a firm chooses an organization of knowledge with higher levels of worker knowledge if it is an MNE than if it is not. The higher communication costs involved in foreign production increase the optimal level of knowledge at their foreign plant and decrease the number of problems communicated to the headquarters. MNEs therefore decrease managerial knowledge to balance its utilization rate and costs, and increase the knowledge level of their domestic workers. Due to the self-selection of firms into FDI studied in subsection 3.2, and as the marginal costs decrease with total firm knowledge, a non-MNE with the same marginal costs and sales as an MNE has lower knowledge $\bar{z}$ than the MNE. The difference in the total amount of knowledge reinforces the difference in workers' knowledge and remuneration, because production workers' knowledge increases in the total

[18]The model abstracts from contractual imperfections, which are relevant in understanding the evolution of managerial wages (e.g., Marin et al., 2015). This section therefore focuses on predictions for workers' wages.

amount of knowledge. Thus, MNEs pay higher remuneration to workers than non-MNEs with the same marginal costs of production and the same sales. Proposition 4 summarizes the results.

Proposition 4. *MNEs pay higher remuneration to domestic workers than non-MNEs in the home country with the same marginal costs and domestic sales if $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$. They pay higher remuneration to foreign workers than non-MNEs in the foreign country with the same marginal costs and foreign sales if $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$ and $c_1 \geq c_0$. The parameter restrictions are sufficient, but not necessary conditions.*

Proof. See Appendix C.3. □

Inspection of the parameter restriction $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$ shows that the model predicts residual MNE wage premiums both for developed to developed and developed to developing country FDI. Foreign wages and learning costs must not exceed domestic wages and learning costs by more than the friction in cross-border communication: $w_1c_1 < \frac{\theta_{10}}{\theta_{00}}w_0c_0$. This includes the case $w_0c_0 = w_1c_1$ studied in the last subsection: the model predicts residual MNE wage premiums in the symmetric general equilibrium. It also comprises $w_0c_0 \approx w_1c_1$, which is likely to apply to FDI from developed countries to other developed countries. Learning costs are likely to be higher in developing than in developed countries, for example due to lower literacy rates. Market wages are typically much lower. Wage premiums occur whenever the difference in market wages outweighs the difference in learning costs. Higher communication frictions increase the likelihood that this is the case. The wage premium in the foreign country is higher the greater c_1 is. Consistent with the empirical evidence (e.g., Aitken et al., 1996; Hijzen et al., 2013), MNE wage premiums are thus predicted to be stronger for developing than for developed countries. As the communication costs and relative wages and learning costs are heterogeneous across countries, the model explains why wage premiums vary with the nationality of the acquirer, as found in Girma and Görg (2007).

The mechanism is reminiscent of but different from that of Caliendo and Rossi-Hansberg (2012), who study exporter-wage premiums using a knowledge-hierarchy model. In their framework, firms reorganize after an increase in output due to trade liberalization. In contrast, the residual MNE wage premium stems from an organizational friction—domestic headquarters for potentially multiple production plants—that is characteristic of MNEs.

In addition, the model features average MNE wage premiums due to the self-selection of firms into FDI. Only firms with a higher knowledge level $\bar{z}$ become MNEs. These firms pay on average higher wages than non-MNEs to managers and workers, both in their home country and the foreign country, due to the positive effect of $\bar{z}$ on z_h, z_0 and z_1 (see Proposition 1). This wage premium does not stem from multinationality per se, but from a firm characteristic—knowledge—that favors FDI and leads to higher wages. The channel is similar to explanations that attribute MNE wage premiums to differences in firm characteristics between MNEs and non-MNEs, such as differences in labor demand volatility or closure rates (see the survey in Malchow-Moller et al., 2013).

4 Communication costs, the predictability of the production process, and MNEs' foreign sales

The theory sections of the paper show that the organization of knowledge in MNEs is systematically different from the organization of knowledge in non-multinationals, and that these differences are useful to understand the geography of MNEs' sales and MNE wage premiums.

The goal of the following section is to provide empirical evidence for the model predictions. This is a difficult undertaking. Knowledge is intangible in nature and typically proprietary. The allocation of knowledge within MNEs is thus very hard to observe. Studies on the organization of knowledge in national firms use social security data (e.g. Caliendo et al., 2015; Friedrich, 2016) or management surveys (e.g. Bloom et al., 2014), but these data sets do not cover information on affiliates of one and the same multinational firm in different countries. To make progress, I exploit the model implications for MNEs' foreign marginal costs and foreign sales. In particular, I focus on the prediction that the effect of the communication costs between a foreign country and the home country of the MNE on the foreign marginal costs and sales varies with the predictability of the production process. This prediction is specific to the organization of knowledge and hard to explain with alternative models of MNEs.

I use comprehensive data on German MNEs that contain information on parent sectors. I develop a new sector-level measure of the predictability of the production process and study how the predictability of the production process in the parent sector and the communication costs between a foreign country and Germany jointly affect MNEs' foreign sales. The strategy thus allows offering specific evidence for the organization of knowledge in MNEs.

4.1 Empirical specification

Proposition 3 states that MNEs' foreign marginal costs increase with the communication costs between the headquarters and the foreign plant. According to Corollary 2, the increase is the stronger, the less predictable the production process is. As explained in section 3.1, the variation of the marginal costs affects MNEs' foreign sales that decrease in the marginal costs.

To provide evidence for this prediction, I exploit the rich variation between investments of the same parent in different countries. Such an approach is popular in the literature on MNEs, and used by Keller and Yeaple (2013) among others, because the administrative data on MNEs are typically only available at country level. The coarse geographic information renders it difficult to transfer the differences-in-differences identification strategies employed in the literature on national multi-plant firms to MNEs. Giroud (2013) uses the introduction of direct flight routes , for example, but with only country-level information on the location of either parent or affiliate it is not possible to determine which firms benefit from lower travel times due to new routes. The within-parent across-country identification strategy allows studying the following empirical predictions:

Prediction 1. *An MNE's foreign sales in a host country j decrease with the bilateral communication costs θ_{jk} between country j and the home country k of the MNE.*

Prediction 2. *The decrease of an MNE's foreign sales with the bilateral communication costs θ_{jk} is the stronger, the lower the predictability of the production process λ is.*

To bring Prediction 1 to the data, I log-linearize the model expression for sales $p_j(\bar{z}_i)q_j(\bar{z}_i)$: $\ln\left(p_j(\bar{z}_i)q_j(\bar{z}_i)\right) = (1-\sigma)\ln\left(\frac{\sigma}{\sigma-1}\right) + \ln Q_j + (\sigma-1)\ln P_j + (1-\sigma)\ln\xi_j(\bar{z}_i, q_0, w_0, q_1, w_1)$. The empirical analysis thus focuses on the intensive margin of MNEs' investments. I estimate a reduced-form version of the resulting equation.[19]

$$\ln\left(\text{foreign sales}_{ijt}\right) = \beta_0 + \beta_1\theta_{j0t} + \beta_2\ln Q_{jt} + \beta_3 c_{jt} + \beta_4 w_{jt} + \delta\mathbf{X}_{jt} + \alpha_{it} + \epsilon_{ijt} \tag{26}$$

The dependent variable is the natural log of the foreign sales of MNE i in country j and year t. The main covariate of interest is θ_{j0t}, the communication costs between country j and Germany, country 0, in year t. I control for the other determinants of foreign sales in the model: the market size of country j in year t, Q_{jt}, the learning costs c_{jt}, and wages w_{jt}. $\mathbf{X}_{jt}$ is a vector of additional controls, including trade costs and investment climate measures. As the estimation relies on cross-section variation, including a rich set of controls is important to ensure that coefficient estimates are not subject to omitted variables bias. α_{it} is an MNE–year fixed effect and ϵ_{ijt} is an MNE–country–year specific error term.

The MNE–year fixed effects are a central component of the empirical approach. They account for heterogeneity between different multinational firms, for example for differences in firm technology $\bar{z}_i$, because the regressions compare investments of the same parent in the same year in different countries. They also hold fixed the managerial knowledge z_h and other headquarter characteristics, as well as, more generally, any common MNE characteristic that may influence performance across destinations.

To account for correlations of the sales of the same MNE across countries and over time on the one hand and of the sales of different MNEs in the same country on the other hand the standard errors are clustered by MNE and country. I implement the two-way clustering using the `xtivreg2`-command (Schaffer, 2015). I employ a built-in option to take into account that the number of clusters may be small due to the limited number of countries.

I adjust the estimation equation by including an interaction term of the communication costs and the predictability of the production process to study Prediction 2:

$$\ln\left(\text{foreign sales}_{ijt}\right) = \gamma_0 + \gamma_1\theta_{j0t} + \gamma_2\theta_{j0t}\times\lambda_s + \gamma_3\ln Q_{jt} + \gamma_4 c_{jt} + \gamma_5 w_{jt} + \eta\mathbf{X}_{jt} + \alpha_{it} + \epsilon_{ijt} \tag{27}$$

I measure the predictability of the production process λ_s at the level of the parent sector of the MNE. Its base effect is captured by the MNE–year fixed effect.

This approach is reminiscent of Keller and Yeaple (2013), who infer the size of spatial knowledge transmission frictions from regressions of MNEs' foreign sales on an interaction of transport costs and the codifiability of knowledge in a sector. The predictability of the production process is distinct from codifiability: The latter denotes how easy it is to verbally

[19] Due to the non-linear nature of the original equation, it is not possible to provide a structural interpretation of the resulting parameter estimates.

summarize a piece of knowledge, whereas the former describes how often a piece of knowledge is likely to be used. The two papers thus study distinct dimensions of MNEs' behavior. The following empirical analyses are specific to the model proposed in the theory part of this paper.

The empirical approach controls for firm heterogeneity as source of differences in MNEs' performance across markets and thus mitigates bias due to the self-selection of firms across countries. Nonetheless, the set of locations is a choice variable of the firm and does not vary exogenously. It is difficult to guarantee that the estimation conditions all information available to the MNE, so the results may be biased due to unobservable MNE–country-specific factors. It is necessary to keep it in mind when interpreting the regression results.

4.2 Data

Foreign sales. I use detailed firm-level data on German MNEs from the Microdatabase Direct investment (MiDi) of the German central bank. The database consists of a panel of yearly information on virtually the universe of foreign affiliates of German MNEs from 1999 to 2010. German residents are legally obliged to report information on the financial characteristics of their foreign investments once these meet the reporting requirements (Schild and Walter, 2015). The database contains detailed balance sheet information, including the sales, the number of employees, and the financial structure of every affiliate. The data also include parent and affiliate sectors, mostly at the two-digit level.

I clean the data (see Appendix E.1 for details), and restrict the sample to majority-owned affiliates. The model applies to situations in which the parent is actively involved in the local production. This is unlikely if other shareholders own the majority of the affiliate. To ensure consistency in the level of analysis of the model, I aggregate the affiliate–level information at the parent–country–year level. The data set contains 164,604 parent–country–year observations.

I augment the data with measures of the communication costs, the predictability of the production process and further controls. Table 2 provides an overview of the model parameters, their empirical analogs, and the data sources. Appendix E.2 tables the summary statistics.

Communication costs. To approximate the bilateral communication costs θ_{jk}, I employ the overlap of office hours, the flight time between Frankfurt and the main economic city of the host country, measures for the similarity of languages, and the internet bandwidth as a measure of communication technologies. I refrain from generic proxies with various alternative interpretations, such as distance.

The office hours overlap captures the fact that time zone differences inhibit communication between the foreign operations and the MNEs' headquarters. Personnel at either location may have to work overtime to communicate. Using e-mail as a time-independent means of communication only mitigates the problem because questions cannot be addressed directly, which causes delay. The lower the office hours overlap, the higher the proportion of problems is that a foreign plant has to address on its own. The lower the office hours overlap, the lower the foreign sales of an MNE should be. The variable is computed as $\max\{10 - |\text{time difference in hours}|, 0\}$.

Table 2: Overview of the model parameters, their empirical analogs and data sources

Parameter	Definition	Empirical analog	Data source
Communication costs			
θ_{jkt}	Bilateral communication costs	Office hours overlap	Author using www.timeanddate.com
		Flight time (between main cities/ Frankfurt for Germany)	www.weltinfo.com, www.meine-flugzeit.de, main city: CEPII
		Linguistic proximity to German (section 4)	Author using CEPII, Ethnologue, Spolaore and Wacziarg (2009)
		Common spoken language, linguistic proximity (section 5)	Melitz and Toubal (2014)
		Internet bandwidth (Mbit/s)	ITU's ICT Indicators Database
Predictability of the production process			
$\frac{1}{\lambda_s}$	Inverse predictability	Probability of facing unexpected problems	Author estimates based on 2006 BiBB/BAuA survey
Foreign country characteristics			
Q_{jt}	Market size	GDP, GDP per capita	IMF
c_{jt}	Learning costs	Average years of schooling	Barro and Lee (2013)
w_{jt}	Wages	Unit labor costs	OECD
τ_{jt}	Trade costs	Trade costs	World Bank
		Effectively applied tariffs by sector	WITS
Additional controls			
	Geography	Distance (population weighted)	CEPII
	Investment climate	Statutory tax rate	IBFD
		Rule of law/ regulatory quality/ government effectiveness/ corruption	World Governance Indicators
	Monitoring costs	Bilateral trust	Eurobarometer 46.0

The table contains an overview of the model parameters, the variables employed in the empirical analyses and the data sources. The indices are j, k: source, host country; t: year; s: MNE parent sector.

The flight time captures how quickly managers can travel to the foreign operations and address potential issues in the production process on site. Despite technological advances, face-to-face communication is often indispensable to ensure successful production (e.g., UNCTAD, 2004; Hausmann, 2016). As higher flight times impede communication, the foreign sales should decrease with the flight time.

To capture difficulties in the direct communication between two individuals, I employ the common spoken language measure from Melitz and Toubal (2014) as well as a measure of the linguistic proximity of the language(s) of the host country and German. "Common spoken language" measures the probability $\in [0, 1]$ that two randomly chosen individuals from two countries speak the same language. The variable "linguistic proximity" captures the notion that it is easier to learn a language and to express oneself precisely the closer that language is to one's mother tongue. I construct a linguistic proximity measure as a function of the number of linguistic nodes common to German and each language spoken by at least 20% of people in the host country following Spolaore and Wacziarg (2009). I take the simple average in the case of several languages.[20] Undoubtedly, international business communication often takes place in English. Still, non-native English speakers tend to develop their own English dialect,

[20]Specifically, I use $\left(\sqrt{\frac{\#\text{ common nodes}-7}{7}}\right)$, slightly modifying the formula in Spolaore and Wacziarg (2009).

strongly influenced by their native languages and often difficult for native English speakers to understand (Gardner, 2013). Linguistic proximity is therefore appropriate to capture frictions in communication despite the use of English in business contexts.

To measure the quality of communication technologies, I use data on the internet bandwidth. Internet bandwidth is comparable across countries, which is not the case for price data that may capture the unobserved quality of service. It is available for many countries, and is arguably exogenous to bilateral FDI flows, unlike the telecommunications traffic for example.

I use the approach of Lubotsky and Wittenberg (2006) to compute the estimated effect of the unobserved communication costs from the proxy variables. I include all proxies in the regression and aggregate their coefficients using the covariance of each proxy and the dependent variable as weights. To correctly estimate the covariance, it is necessary to partial out the correlation of the dependent variable and the proxies with other control variables first. One covariance has to be normalized. I choose the office hours overlap. Formally, the estimated effect is:

$$\hat{\beta}_1 = \sum_{k=1}^{K} \frac{\text{cov}(y_{ijt}, x^{\theta}_{jt,k})}{\text{cov}(y_{ijt}, x^{\theta}_{jt,1})} b_{x^{\theta}_{jt,k}}, \tag{28}$$

where y_{ijt} and $x^{\theta}_{jt,k}$ are the residuals of regressions of log foreign sales and the kth proxy for θ_{j0t} on the other control variables, and $b_{x^{\theta}_{jt,k}}$ is the coefficient estimate of the kth proxy. Lubotsky and Wittenberg (2006) show that this estimate is less downward biased due to the measurement error of the proxy variables than estimates using only one proxy or the principal component of the proxies. Running regressions on an index of the proxy variables computed with the covariances as weights instead of the proxies leads to the same estimate of $\hat{\beta}_1$. I compute the standard error of $\hat{\beta}_1$ by re-running the regressions with the index to take the two-way clustering into account.[21]

Predictability of the production process. In the model, the predictability of the production process is given by λ, the rate parameter of the problem probability distribution function. The lower λ is, the more often "rare" problems in the tail of the distribution arise. To capture this variable, I construct a new measure based on a survey question from the "BIBB/BAuA Employment Survey 2006" administered by the German Federal Institute for Vocational Education and Training (*Bundesinistitut für Berufsbildung*, BiBB) and the Federal Institute for Occupational Safety and Health (*Bundesanstalt für Arbeitsschutz und Arbeitsmedizin*, BAuA) (Hall and Tiemann, 2006). The survey collects data on the education, career and current employment conditions of a representative sample of 20,000 working age individuals in Germany.

One of the survey questions is ideally suited to measuring the predictability of the production process λ. Survey participants report how often they have "to react to and solve unforeseeable problems" in their current job. I restrict the sample to sectors with at least 25 responses. I regress a dummy that is equal to one if participants answer "frequently", and zero if they answer "sometimes" or "never", on dummies for the two-digit sector that a participant is employed in.

[21] I also bootstrapped the standard errors of $\hat{\beta}_1$ with resampling of MNEs. These standard errors were smaller.

I use the estimated coefficients of the sector dummies as measure of the predictability of the production process in the parent sector of a German MNE. The measure is inversely related to λ: the higher λ, the less likely it is that rare problems in the tail of the distribution arise. Appendix section E.3 documents that the measure plausibly captures the predictability of the production process using summary statistics and its relation with other sector characteristics.

Other covariates. I control for the other model parameters and potential omitted variables. I use data on GDP and GDP per capita to measure variation in demand due to differences in the market size and income across countries. I measure the learning costs using the average years of schooling[22] and employ information on the unit labor costs to measure wages.

I control for bilateral geographic distance because it is known to affect MNE performance. I employ data on trade costs and the average effectively applied tariffs in the parent sector to measure barriers to trade. To account for factors that may influence sales but are not included in the model, I add measures of the investment climate: the statutory tax rate and indicators on the rule of law, government effectiveness, corruption, and regulatory quality. For the same reason, I add a measure of the bilateral trust between countries.

I take the logarithms of the covariates if their distribution in levels is skewed.

4.3 Graphical evidence

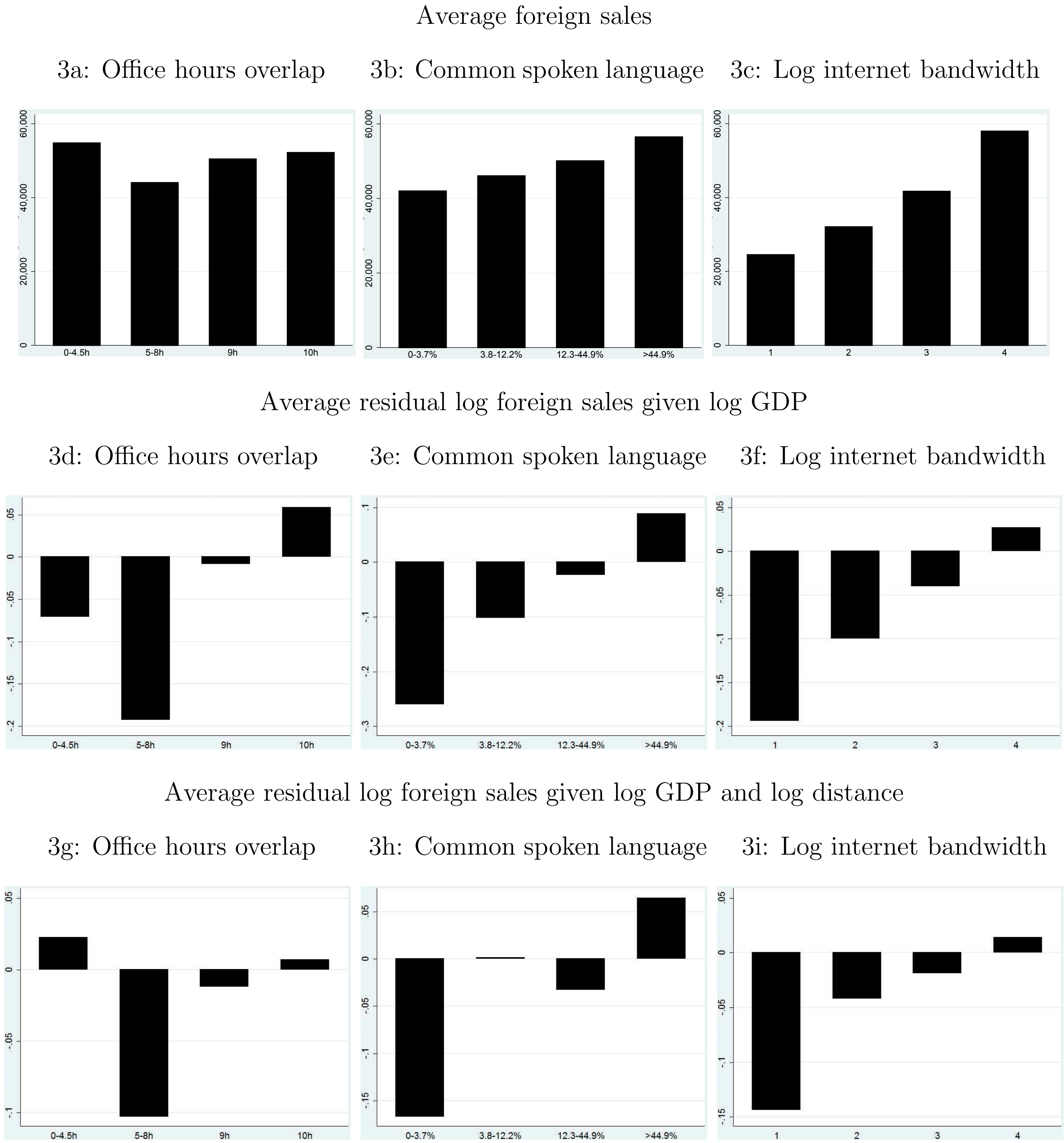
To illustrate that the bilateral communication costs are a first-order determinant of MNEs' sales, figure 3 presents graphical evidence on the relation of MNEs' foreign sales and the communication cost proxies. It includes the office hours overlap and the internet bandwidth, because these two measures have the highest predicted power as measured by a high R-squared and a low standard error (see Appendix Table E.5). It also displays the common spoken language proxy, because the proxy has a straightforward probability interpretation (unlike linguistic proximity). The figures are constructed by assigning countries to quartiles of the communication proxy distribution and computing the average observed sales per quartile.[23] The three figures in the top row show the relation of the communication cost proxies and the foreign sales in levels. The second row shows the relation with the residuals of a regression of log foreign sales on log GDP to take differences in the market size across countries into account. The sales residuals used in the three figures in the bottom row additionally condition on log distance.

The foreign sales increase with lower communication costs. Whether the figures display sales in levels or sales residuals, the sales are higher for higher values of the common spoken language and the internet bandwidth proxies. The relation also holds for the second to forth quartile of the office hours overlap. An exception is the lowest quartile of this variable. The non-linearity is largely driven by the US, as Appendix figure E.1 shows. Overall, the figures lend support to the postulated relation of foreign sales and communication costs.

[22] The data are for 2000, 2005, and 2010. I assign the value of the closest year to my 1999-2010 sample. In unreported regressions, I use the public expenditure on education and PISA scores, and obtain similar results.

[23] I use bar plots as scatter plots are difficult to interpret: the office hours overlap is categorical, the internet bandwidth takes on a limited number of values, and the common spoken language proxy has a point mass at 0.

Figure 3: Foreign sales vs. bilateral communication costs



Figures 3a-3c plot the average foreign sales of German MNEs by quartiles of the communication cost proxies office hours overlap, common spoken language and log internet bandwidth. Figures 3d-3f plot the average residuals of a regression of log foreign sales on log GDP by quartiles of the same proxies. Figures 3g-3i plot the average residuals of a regression of log foreign sales on log GDP and log distance by quartiles of the proxies.

4.4 Regression results

4.4.1 Communication costs and MNEs' foreign sales

Table 3 presents the regression results for Prediction 1. The table displays nine specifications. Columns 1 and 2 contain the model parameters: the communication and learning costs, the wages, and the market size.[24] The specifications are displayed separately because the wage data are only available for OECD countries, so the sample size decreases once wages are included. Column 3 adds distance, and Column 4 shows a specification with distance, but without communication costs. Columns 5 and 6 add trade costs and tariffs. These variables are only available for manufacturing, so the sample now only contains manufacturing firms. Columns 7 to 9 additionally include measures for the quality of the investment climate and bilateral trust between Germany and foreign countries. The Table reports the estimated effect of the communication costs computed by aggregating the coefficients of the communication cost proxies à la Lubotsky and Wittenberg (2006) along with the coefficient estimates. Appendix Table E.4 displays the weights used to aggregate the coefficients.

Table 3 displays the number of MNEs and the number of country combinations together with the number of observations. The number of country combinations is the number of distinct combinations of countries in which the MNEs in the sample are active. The number is decisive because the variation within MNEs across countries drives the regression results.[25] The number of country combinations exceeds the number of MNEs because MNEs change the set of investment destinations over time.

The regression results lend strong support for Prediction 1. Among the proxies, the office hours overlap has a positive effect on foreign sales significant at the 5% level in most specifications. The effect of flight time is mostly negative, but insignificant. Linguistic proximity increases foreign sales. The coefficients are significant at the 1% or even 0.1% level throughout. Given linguistic proximity, a higher probability that two individuals speak the same language does not have a significant effect. Its coefficient is negative, but as Lubotsky and Wittenberg (2006, p. 558) argue, a proxy may have a different sign than the true effect if it is highly correlated with another, better-measured proxy included in the regression. Higher internet bandwidth increases foreign sales, though the effect is insignificant from column 2, which probably stems from the smaller number of observations and the more homogeneous set of countries. The estimated effect of the communication costs computed from the coefficients of the proxy variables is positive, of similar size and significant at the 0.1% level throughout specifications.. Consistent with the model, the regressions show that communication costs affect MNEs' foreign sales, even if other determinants of foreign sales are controlled for.

Column 3 adds distance to the specification. Distance is correlated both with communication and trade costs. Its coefficient is positive but insignificant. As column 4 shows, distance has a significantly negative effect once the communication cost proxies are omitted, consistent with many papers that show that MNEs' foreign sales decrease with the distance of a foreign country from the home country (e.g. Irarrazabal et al., 2013). The comparison of columns 3 and 4 indicates that the negative distance coefficient often reported likely reflects the correlation of distance and omitted communication costs. As distance is correlated with the communication costs, I include the coefficient of distance in the estimated effect of the communication costs in column 3. Otherwise, the estimated effect is six times larger and thus implausibly high. This

[24]Appendix Table E.5 includes the communication cost proxies one by one.

[25]Due to missing values for the covariates, not all the countries are always included in the regression sample.

Table 3: Regression results: within-firm differences in log foreign sales across countries

Log foreign sales	1	2	3	4	5	6	7	8	9
Office hours overlap	0.047*	0.044	0.034		0.068*	0.068*	0.067*	0.081**	0.069*
	(0.020)	(0.028)	(0.023)		(0.031)	(0.031)	(0.031)	(0.027)	(0.031)
Log flight time	0.017	−0.120	−0.521+		−0.058	−0.062	−0.065	−0.014	−0.168
	(0.070)	(0.112)	(0.297)		(0.149)	(0.148)	(0.146)	(0.129)	(0.125)
Common spoken language	−0.102	−0.014	0.217		−0.154	−0.159	−0.124	−0.031	−0.379
	(0.171)	(0.174)	(0.242)		(0.207)	(0.208)	(0.216)	(0.275)	(0.219)
Linguistic proximity	0.452**	0.627***	0.538**		0.690***	0.694***	0.669***	0.603**	0.865***
	(0.135)	(0.132)	(0.156)		(0.161)	(0.161)	(0.165)	(0.190)	(0.177)
Log internet bandwidth	0.070**	0.032	0.034		0.016	0.016	0.015	−0.010	0.022
	(0.022)	(0.028)	(0.028)		(0.033)	(0.034)	(0.033)	(0.031)	(0.035)
Log GDP	0.255***	0.386***	0.364***	0.366***	0.411***	0.412***	0.404***	0.437***	0.457***
	(0.031)	(0.037)	(0.027)	(0.044)	(0.048)	(0.047)	(0.049)	(0.050)	(0.055)
Log GDP per capita	0.139***	−0.066	−0.092	0.089	−0.104	−0.101	−0.114	−0.148	−0.317*
	(0.038)	(0.066)	(0.064)	(0.073)	(0.084)	(0.084)	(0.093)	(0.087)	(0.116)
Average years of schooling	0.026	0.065***	0.061***	0.048*	0.081***	0.081***	0.081***	0.070**	0.104**
	(0.018)	(0.015)	(0.014)	(0.023)	(0.019)	(0.019)	(0.018)	(0.022)	(0.028)
Unit labor cost		−0.730+	−0.778*	−0.431	−0.743	−0.747	−0.750+	−0.765	−0.099
		(0.359)	(0.373)	(0.479)	(0.460)	(0.465)	(0.435)	(0.565)	(0.542)
Log distance			0.329	−0.200***					
			(0.218)	(0.040)					
Log trade costs					−0.038	−0.035	−0.033	−0.045	−0.020
					(0.102)	(0.101)	(0.101)	(0.088)	(0.079)
Effectively applied tariffs						0.007**	0.007**	0.006*	0.007**
						(0.002)	(0.002)	(0.002)	(0.002)
Statutory tax rate							0.004		
							(0.007)		
Regulatory quality								−0.000	
								(0.166)	
Rule of law								0.015	
								(0.146)	
Government efficiency								0.223+	
								(0.114)	
Corruption								−0.148	
								(0.101)	
Bilateral trust									0.171
									(0.122)
Estimated effect of communic. costs	0.079***	0.139***	0.144***		0.148***	0.148***	0.146***	0.143***	0.248***
	(0.012)	(0.019)	(0.017)		(0.026)	(0.026)	(0.026)	(0.025)	(0.036)
# observations	121,116	78,357	78,357	80,619	47,990	47,454	47,454	41,391	41,904
# MNEs	4,089	3,283	3,283	3,292	2,008	1,999	1,999	1,912	1,884
# country comb.	8,801	8,452	8,452	8,464	5,584	5,494	5,494	5,044	5,379
R-squared (within)	0.193	0.155	0.156	0.138	0.168	0.168	0.169	0.162	0.171

Two-way clustered standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign sales per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. *# country comb.:* number of combinations of countries with MNE activity. The *Estimated effect of communic. costs* is computed from the coefficients of the office hours overlap, log flight time, common spoken language, linguistic proximity, log internet bandwidth and, in column 3, log distance following Lubotsky and Wittenberg (2006). Appendix Table E.4 displays the weights.

is a further indication that distance proxies for omitted communication costs in column 4.[26]

[26]In unreported regressions, I include subsets of the proxy variables in the regression. I obtain very similar coefficient estimates, whether distance is among the proxies or not. As Lubotsky and Wittenberg (2006) show,

The other covariates have plausible effects. Foreign sales tend to be higher in larger countries as measured by GDP. GDP per capita is mostly insignificant. Lower learning costs, as measured by higher average years of schooling, significantly increase foreign sales. Higher unit labor costs tend to decrease sales, but the coefficients are mostly insignificant. In terms of the impact of learning costs and wages on foreign sales discussed in subsection 3.1, this finding implies that the direct negative effect of these variables is not outweighed by indirect adjustments to the organization of knowledge. Higher tariffs increase foreign sales, consistent with a horizontal motive of FDI. The trade cost estimate itself is not significant. Among the business climate measures, only higher government efficiency significantly increases MNEs' foreign sales. Higher bilateral trust has a positive effect that is marginally significant (P-value $\approx$ 16%).

4.4.2 Heterogeneity of effects with the predictability of the production process

Table 4 presents the regression results on Prediction 2. I employ two variants of the measure of predictability: Columns 1 and 2 use the estimated inverse predictability for all sectors. Columns 3 and 4 set the measure to 0 if the coefficient estimate of the sector dummy is not significant at the 20% level as robustness check. Columns 1 and 3 contain the communication cost proxies and interactions of each proxy with the predictability of the production process along with GDP, GDP per capita and years of schooling. Columns 2 and 4 add the unit labor costs. This leads to a significant reduction of the number of countries.

The table lends strong support to Prediction 2. The estimates are very similar for the two measures of the predictability of the production process. The coefficient patterns of the base effect of the communication cost proxies are very similar to the pattern in columns 1 and 2 of Table 3. Consistent with the model of the organization of knowledge, the impact of the communication costs is stronger if the production process is less predictable (i.e. the probability of facing unexpected problems is higher): the interaction terms have the same sign as the base effects, except for the flight time. The interaction terms of the predictability of the production process with the linguistic proximity and the internet bandwidth are positive and significant at the 5% and 0.1% level. The other interaction terms are not robustly significant. The control variables have similar estimated effects as in Table 3.

The method of Lubotsky and Wittenberg (2006) is only applicable to one latent unobserved variable. Table 4 features two latent variables, the communication costs and their interaction with the predictability of the production process. To nonetheless improve the interpretability of the estimated interaction term, Appendix Table E.6 reports results of complementary regressions that include the index of the communication cost proxies à la Lubotsky and Wittenberg (2006) and an interaction term of the index and the predictability of the production process (Agrawal, 2017, uses a similar approach in a different context). I construct the index using weights employed to aggregate the coefficients in Tabe 3. Unlike in Table 3, the coefficients of the index and the interaction term in Table E.6 do not contain the exact same information as the coefficients of the proxy variables and their interactions, because they are not equivalent

this is a further indication that distance proxies for the same unobserved variable as the other proxies.

Table 4: Effect of communication costs with the predictability of the production process

Log foreign sales	$\hat{\lambda}$: baseline measure 1	2	$\hat{\lambda}$: 0 if insignificant 3	4
Office hours overlap	0.037^{+}	0.037	0.034^{+}	0.038
	(0.020)	(0.029)	(0.020)	(0.029)
Office hours overlap ×	0.129	0.119	0.181	0.115
Probability of unexpected problems	(0.128)	(0.149)	(0.123)	(0.138)
Log flight time	−0.042	−0.174	−0.034	−0.157
	(0.065)	(0.119)	(0.064)	(0.119)
Log flight time ×	0.772^{+}	0.799	0.765^{*}	0.664
Probability of unexpected problems	(0.397)	(0.477)	(0.376)	(0.424)
Common spoken language	−0.092	0.026	−0.067	0.043
	(0.163)	(0.167)	(0.170)	(0.168)
Common spoken language ×	−0.348	−0.725	−0.713	−1.023
Probability of unexpected problems	(0.776)	(0.743)	(0.729)	(0.751)
Linguistic proximity	0.354^{*}	0.525^{***}	0.361^{*}	0.538^{***}
	(0.142)	(0.129)	(0.142)	(0.127)
Linguistic proximity ×	1.644^{*}	1.617^{*}	1.732^{*}	1.631^{*}
Probability of unexpected problems	(0.778)	(0.619)	(0.745)	(0.628)
Log internet bandwidth	0.035	0.003	0.036	0.004
	(0.023)	(0.029)	(0.023)	(0.027)
Log internet bandwidth ×	0.470^{***}	0.395^{**}	0.498^{***}	0.414^{**}
Probability of unexpected problems	(0.121)	(0.137)	(0.120)	(0.129)
Log GDP	0.253^{***}	0.382^{***}	0.253^{***}	0.383^{***}
	(0.031)	(0.036)	(0.031)	(0.036)
Log GDP per capita	0.143^{***}	−0.053	0.142^{***}	−0.054
	(0.037)	(0.063)	(0.038)	(0.063)
Average years of schooling	0.023	0.062^{***}	0.023	0.063^{***}
	(0.018)	(0.014)	(0.018)	(0.014)
Unit labor cost		-0.778^{*}		-0.782^{*}
		(0.345)		(0.343)
# observations	116,172	75,333	116,172	75,333
# MNEs	3,972	3,199	3,972	3,199
# sectors	38	38	38	38
# countries	102	23	102	23
R-squared (within)	0.197	0.156	0.197	0.156

Two-way clustered standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign sales per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. *# sectors:* number of parent sectors. *# countries:* number of countries. Columns 1 and 2 use the estimated probability of unexpected problems in all sectors. Columns 3 and 4 set the probability to zero if the coefficient estimate of the sector dummy is not significantly different from 0 at the 20% level.

to the weighted aggregate of the individual coefficients. Due to the lower number of covariates in the index regressions, it is possible to cluster the standard errors at the sector-country level. Both the baseline effect of the communication cost index and the interaction term are positive. Despite the conservative calculation of the standard errors, the interaction term is significant at the 10% and 5% level in columns 1 and 3.

Overall, the regression results thus confirm Prediction 2: the less predictable the production process is, the more negative is the effect of higher communication costs on foreign sales. These findings lend strong support to the theory of the organization of knowledge in MNEs developed in this paper.

4.4.3 Robustness checks

Section E.6 in the Appendix reports the results of robustness checks. Table E.7 replicates the regressions in Table 3 after dropping firms that change their investment destinations, i.e. adjust the extensive margin of their investments. The results are very similar, which suggests that the intensive margin of investments is indeed driving the results.

Appendix Table E.8 replicates the regressions in Table 3 after dropping Austria from the sample. The coefficient of the linguistic proximity turns insignificant. This is plausible, because Austria is the only country with a linguistic proximity of 1 to Germany, so dropping Austria reduces the variation in the data. The estimated effect of the communication costs is significantly positive, lower and more similar in size across specifications. Part of the increase in the estimated effect between columns 1 and 2 of Table 3 may therefore be attributable to the special role of Austria. Appendix Table E.9 re-runs the specification in Table 4 after dropping Austria. Interestingly, the interaction term of linguistic proximity and the predictability of the production process turns out positive and significant, as in the main Table.

Tables E.10 and E.11 replicate the regressions using the number of employees as dependent variable. This size measure is less sensitive to exchange rate fluctuations, for example. I restrict the sample to 2004-2010, because the number of employees was optional beforehand, so the data is partly only estimated in earlier years. There are small shifts in the significance pattern of the proxies in Table E.10 compared to Table 3, mainly because the estimation is less precise, possibly due to the lower number of observations. The estimated effect of the communication costs is highly significant and even slightly larger in magnitude, so the results confirm the main results of the paper. Similarly, the sign and significance of individual coefficients in Table E.11 differ from Table 4, but as a whole, the estimates point in the same direction.

Beyond these robustness checks, I use a time-zone shift in Venezuela in 2007 for a supplementary empirical exercise. The time zone shift decreased the office hours overlap between Venezuela and Germany. I implement a differences-in-differences approach using investments of German MNEs in the neighboring country Colombia as control group. Consistent with the predictions of the model, the foreign investments in Venezuela grow at slower rates than the control group after the time zone shift. The supplement is available upon request.

5 Corporate transferees and the organization of knowledge

Section 4 shows that higher communication costs affect the geography of MNEs' sales in ways that strongly support the model of the organization of knowledge in MNEs: the less predictable the production process, the stronger the negative effect of higher communication costs on foreign sales. Yet how do MNEs adjust the organization of knowledge within their boundaries?

To shed light on this question, I use unique and novel information on the flows of corporate transferees between pairs of countries. Corporate transferees are employees who MNEs transfer from their regular place of work to operations of the MNE in another country for a limited period of time. Transferring knowledge is the predominant motive for such within-MNE employee relocations (e.g., Bonache and Brewster, 2001). In recent surveys of Canadian and German firms, three quarters of firms state that they use corporate transferees for knowledge transfer, making knowledge transfer their most frequent purpose (Canadian Employee Relocation Council, 2013; Djanani et al., 2003, p. 34f.). Transferring knowledge is even more important for large firms: Almost 90% of German firms with 2,001 to 10,000 employees use corporate transferees to transfer knowledge compared to 79.5% in the group with 501 to 2000 employees, and 60.9% of firms with up to 500 employees. Corporate transferees are thus a visible reflection of the organization of knowledge in MNEs, and information on the flows of corporate transferees is useful to elucidate the measures that MNEs take to organize knowledge across plants.

Proposition 1 shows that the optimal knowledge level at a foreign plant increases with the communication costs between the plant and the headquarters. To get from this firm-level comparative statics result to a prediction on the flows of corporate transferees between pairs of countries, it is necessary to take the indirect effect of the communication costs on the knowledge level at a foreign plant through their impact on the profit maximizing quantities and the self-selection of firms into account. As explained Appendix F.1, higher communication costs increase the optimal level of knowledge at a foreign plant and thus the desired knowledge transfer even after accounting for changes in quantities and investments. As the number of corporate transferees necessary to transfer a certain amount of knowledge not only depends on the communication costs, but also on the size of operations, Prediction 3 focuses on the share of corporate transferees in MNEs' employment.

Prediction 3. *The share of corporate transferees from country j in country k in MNEs' employment increases with the bilateral communication costs θ_{jk} between the two countries.*

It is far from obvious to find evidence for this prediction, as higher communication costs may drive up an employee's cost of moving, and thus an MNE's costs of using corporate transferees. Further, knowledge transfer is not the only motive for using corporate transferees. MNEs also use them to support the establishment of foreign operations, address talent shortages in the local labor market and for career development, for example (Canadian Employee Relocation Council, 2013; Djanani et al., 2003). Some corporate transferees thus perform managerial services abroad. These activities do not necessarily respond to the communication costs.

5.1 Data

Obtaining evidence for Prediction 3 requires data on the bilateral flows of corporate transferees and MNEs' employment, as well as measures for the bilateral communication costs.

Corporate transferees. The data on the corporate transferees come from Finaccord, a market research company. The data contain information on the number of corporate transferees from 25 source countries in 29 host countries, as well as selected source-host country pairs with significant expatriate populations for the year 2009. The information covers transfer periods of between one and five years. The data are left-censored at 100, and do not distinguish between transferees sent from the headquarters to the foreign operations and foreign employees being trained at the headquarters. Appendix F.2 provides a list of source and host countries. Appendix F.3 provides a plausibility check of the data based on Djanani et al. (2003).

MNEs' employment. To measure bilateral FDI flows, I use data on employment by MNEs from country j in country k provided by the Organization for Economic Co-operation and Development (OECD). Although the data are comprehensive, they do not contain information on all country pairs in the corporate transferee data. Bilateral employment is available for 316 country pairs. Appendix F.2 provides details concerning the variable construction.

Communication costs. I employ the same communication cost measures as in section 4, with the exception of linguistic proximity. I use the linguistic proximity variable from Melitz and Toubal (2014) that is coarser than the one used in section 4, but available for many countries.

5.2 Empirical specification

I specify the following regression equation to provide evidence concerning Prediction 3:

$$\ln\left(\frac{\#\text{ corporate transferees}_{jk}}{\text{Employment}_{jk} + \text{Employment}_{kj}}\right) = \beta_0 + \beta_1\theta_{jk} + \beta_2 d_{\text{cens}} + \boldsymbol{\alpha}_k + \boldsymbol{\alpha}_j + \epsilon_{jk} \qquad (29)$$

The dependent variable is the share of corporate transferees in the total bilateral employment of MNEs in countries j and k. I take the log because the distribution of the share in levels is right-skewed. MNEs may send corporate transferees from the headquarters to foreign operations, or train foreign employees at the headquarters. The transferee data do not distinguish between the two modes, so I put the sum of employment by MNEs from the source country j in the host country k, Employment$_{jk}$, and employment by MNEs from country k in country j, Employment$_{kj}$, in the denominator.

The explanatory variable of interest is θ_{jk}, the communication costs between the two countries. The expected sign of β_1 depends on the measure of θ_{jk}. d_{cens} is a dummy for observations with censored information on the transferee flows. $\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_k$ denote source and host country dummies to capture other determinants of the corporate transferee flows. More generally, the host and source country fixed effects capture any factors that generally increase or decrease the number of corporate transferees sent from or to certain countries. ϵ_{jk} is an error term.

Alternatively, I could specify a Tobit model with the number of corporate transferees as the dependent variable. I prefer the above specification for two reasons. First, the specification permits the use of source and host country fixed effects. A Tobit model with fixed effects entails an incidental parameters problem: around 60 fixed effects are estimated from around 300 observations. Second, the employment of MNEs is a control variable in the Tobit specification,

Table 5: Regression results for the log share of corporate transferees

Log share of transferees	1	2	3	4	5	6	7	8
Office hours overlap	-0.122^{***}				-0.024	-0.144^{***}		-0.113^{*}
	(0.021)				(0.043)	(0.023)		(0.053)
Flight time in hours		0.080^{***}			0.037°		0.058^{**}	-0.030
		(0.013)			(0.026)		(0.018)	(0.028)
Common spoken lang.			-0.730^{**}		-0.540^{*}	0.274	-0.035	-0.069
			(0.259)		(0.246)	(0.314)	(0.324)	(0.385)
Linguistic proximity			-0.291^{***}		-0.144^{**}	-0.099°	-0.153^{*}	-0.194^{**}
			(0.061)		(0.054)	(0.063)	(0.064)	(0.063)
Log bandwidth (Mbit/s)				-0.296^{***}	-0.224^{***}	-0.543^{***}	-0.475^{***}	
				(0.067)	(0.062)	(0.069)	(0.075)	
Constant	-4.937^{***}	-6.256^{***}	-4.926^{***}	-1.460°	-2.203^{*}	1.292°	-0.456	-6.473^{***}
	(0.172)	(0.130)	(0.167)	(0.946)	(1.043)	(0.954)	(1.107)	(0.587)
# observations	316	316	316	315	315	315	315	316
# source countries	26	26	26	26	26	26	26	26
# host countries	30	30	30	29	29	29	29	30
Source country dummies	*N*	*N*	*N*	*N*	*N*	*Y*	*Y*	*Y*
Host country dummies	*N*	*N*	*N*	*N*	*N*	*N*	*N*	*Y*
R-squared	0.114	0.120	0.153	0.092	0.210	0.495	0.461	0.722

Robust standard errors in parentheses. $^{\circ}$ $p < 0.20$, $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Dependent variable:* log share of corporate transferees in MNE employment. *Covariate definitions:* see Table 2.

resulting in simultaneity bias because the size of foreign operations depends on the organization of knowledge reflected in the corporate transferees. Appendix F.5 reports the results of Tobit regressions without fixed effects that are in line with the main results.

5.3 Regression results

Table 5 presents the regression results. As stated in Prediction 3, the share of corporate transferees in the total employment of MNEs increases with higher bilateral communication costs. Columns 1 to 4 separately include the different measures for the communication costs. The regression results imply that an overlap in office hours of one hour longer is associated with a decrease in the share of corporate transferees of 12%. A one hour longer flight time leads to an 8% increase in the share of corporate transferees. A higher probability of 10 percentage points that two randomly chosen individuals speak the same language is associated with decrease in the share of corporate transferees of 7%. Linguistic proximity only takes four values: 0, 0.25, 0.5, and 0.75. Closer linguistic proximity—for example, an increase from 0.5 to 0.75 if two languages belong to the same sub-branch and not only the same branch of a language tree—increases the share of corporate transferees by 7%. An increase in the internet bandwidth of 10% leads to a 3% decrease in the share of corporate transferees.

In column 5, the covariates are jointly included. Signs are robust, but the significance levels decrease, reflecting that the different measures are correlated. The office hours overlap turns insignificant. Columns 6 and 7 add source country dummies. As overlap and flight time are most strongly correlated, the models separately include the two variables and restore their significance levels. The common spoken language variable turns insignificant, but the significant linguistic proximity variable shows that language matters. Column 8 adds host country dummies. The bandwidth drops as it is a host country characteristic. As before, only one of office hours overlap and flight time is significant. Common spoken language is insignificant, but linguistic proximity is negative and significant.

In summary, the share of corporate transferees increases with the communication costs, consistent with Prediction 3.[27] These findings shed light on one tool used by multinational firms to organize knowledge across countries.

6 The organization of knowledge and monitoring

The determinants of the performance of large multi-plant firms receive increasing attention in the literature. Two recent papers find that investment in a plant increases after a new airline route between the firm's headquarters and the plant location is introduced (Giroud, 2013), and that greater distance between the establishment and the headquarters is associated with shorter establishment survival (Kalnins and Lafontaine, 2013). Both articles attribute their findings to monitoring problems and information asymmetries between headquarters and establishments. Monitoring is likely also an important factor for the performance of MNEs. One could therefore be concerned in how far understanding the organization of knowledge improves our understanding of MNEs.

To explore this question, the regressions include bilateral trust in column 8 of Table 3. Trust is a specific proxy of monitoring costs. Higher bilateral trust decreases monitoring costs and allows firms to decentralize more easily (Bloom et al., 2012). Bilateral trust has a positive and marginally significant effect on the MNE-level distribution of sales, consistent with the predictions of a monitoring model. The estimates for the communication costs are similar to those in other columns. This suggests that monitoring and the organization of knowledge are two factors that simultaneously affect the performance of MNEs. The empirical analyses in Table 4 are a further indication of the relevance of the organization of knowledge: the effect of the communication costs varies with the predictability of the production process in ways that are characteristic of the model proposed in this paper.

Beyond the empirical evidence presented in this paper, a vast literature documents that MNEs pay higher wages than comparable domestic firms. A monitoring model does not explain why MNE wage premiums emerge. Instead, the predictions of monitoring models are hard to reconcile with the empirical evidence. If the cross-border monitoring costs exceed the within-country monitoring costs, only firms with better monitoring technology self-select into FDI. This implies that MNEs pay lower wages in the home country than domestic firms: Firms with better monitoring technology are able to implement optimal effort levels with lower wage payments. This is at odds with empirical evidence on home country wage premiums. Likewise, a monitoring model cannot explain residual MNE wage premiums in the foreign country because it predicts that foreign affiliates and domestic firms in the foreign country with the same marginal costs pay the same wages. Appendix G sketches a formal analysis of this argument.

[27] Astorne-Figari and Lee (2016) obtain similar results from micro-data on expatriates in Korean MNEs.

Overall, it is thus plausible that both monitoring and the organization of knowledge affect the performance of MNEs.

7 Conclusion

This paper provides the first systematic analysis of the organization of knowledge in MNEs. It shows that the optimal organization of knowledge in MNEs differs from the organization of knowledge in domestic firms: MNEs assign higher knowledge levels to production workers in the home and the foreign countries than if they were non-multinational. The knowledge levels of production workers increase with the communication costs between the plant and the headquarters, and this increase is the stronger, the less predictable the production process is. These features of the organization of knowledge in MNEs are useful in understanding why MNEs' sales and their probability of entry decrease with the distance of a country from the home country of the MNE. They also explain why MNEs pay higher wages than equally productive domestic firms in the home and the foreign countries, and why MNE wage premiums vary with the nationality of the parent firm. The paper provides conclusive empirical evidence in support of the predictions of the model. MNE-level analyses support the predictions regarding the relationship between the communication costs and MNEs' sales, and confirm that the relationship is stronger in sectors with a less predictable production process. Novel data on the flows of corporate transferees between countries shed light on one means of intra-MNE knowledge transfer.

More detailed (though so far unavailable) data on the organization and wage structure of MNEs across countries would allow conducting stricter tests of the model predictions. For example, one could study the relation of the size of MNE wage premiums and the difference in communication costs within and across countries and sectors, or allow for heterogeneity in the within-firm communication costs across firms and explore the effect of such heterogeneity on firm behavior.

The paper offers relevant insights for the design of policies aimed at promoting investment and the diffusion of knowledge across countries. Creating well-paid, relatively knowledge-intensive new jobs is one of the main objectives of investment promotion efforts (Javorcik, 2012). The results of this paper generally support the presumption that employment in MNEs is likely to be more knowledge intensive and better paid than employment in domestic firms. In their efforts to reap these benefits, countries may be tempted to focus on investing in targeted information campaigns and a good investment climate in terms of administration, governance, and the education of their workforce. As this paper demonstrates, targeted foreign language training and good communication infrastructures may be equally relevant in fostering FDI inflows as they facilitate multinationals' task of efficiently organizing knowledge across countries.

References

Agrawal, D. (2017). The Internet as a Tax Haven?: The Effect of the Internet on Tax Competition. *University of Kentucky*, mimeo.

Aitken, B., A. Harrison, and R. E. Lipsey (1996). Wages and foreign ownership A comparative study of Mexico, Venezuela, and the United States. *Journal of International Economics 40*(3-4), 345–371.

Antràs, P., L. Garicano, and E. Rossi-Hansberg (2006). Offshoring in a knowledge economy. *The Quarterly Journal of Economics 121*(1), 31–77.

Antràs, P. and S. R. Yeaple (2014). Multinational Firms and the Structure of International Trade. *Handbook of International Economics 4*, 55–130.

Arnold, J. and B. Javorcik (2009). Gifted kids or pushy parents? Foreign direct investment and plant productivity in Indonesia. *Journal of International Economics 79*(1), 42–53.

Astorne-Figari, C. and J. Lee (2016). An empirical investigation on the transfer of expatriates within MNCs from a knowledge perspective. *University of Memphis*, mimeo.

Bahar, D. (2016). Should I stay or should I go? Knowledge transmission, the ease of communication and location decisions of multinational firms. *Harvard University*, mimeo.

Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics 104*, 184–198.

Bernard, A. B. and J. B. Jensen (2007). Firm structure, multinationals, and manufacturing plant deaths. *The Review of Economics and Statistics 89*(2), 193–204.

Bloom, N., L. Garicano, R. Sadun, and J. Van Reenen (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science 60*(12), 2859–2885.

Bloom, N., R. Sadun, and J. Van Reenen (2012). The Organization of Firms Across Countries. *The Quarterly Journal of Economics 127*(4), 1663–1705.

Bolton, P. and M. Dewatripont (1994). The firm as a communication network. *The Quarterly Journal of Economics 109*(4), 809–839.

Bonache, J. and C. Brewster (2001). Knowledge transfer and the management of expatriation. *Thunderbird International Business Review 43*(1), 145–168.

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015). The Anatomy of French Production Hierarchies. *The Journal of Political Economy 123*(4), 809–852.

Caliendo, L. and E. Rossi-Hansberg (2012). The impact of trade on organization and productivity. *The Quarterly Journal of Economics 127*(3), 1393–1467.

Canadian Employee Relocation Council (2013). CERC Survey: Intra-Company Transfers.

Cristea, A. D. (2015). The effect of communication costs on trade in headquarter services. *Review of World Economics 151*(2), 255–289.

Davidson, C., F. Heyman, S. Matusz, F. Sjholm, and S. C. Zhu (2014). Globalization and imperfect labor market sorting. *Journal of International Economics 94*(2), 177–194.

Defever, F. (2012). The spatial organization of multinational firms. *Canadian Journal of Economics/Revue canadienne d'économique 45*(2), 672–697.

Djanani, C., M. Lösel, and C. Lösel (2003). *Entsendung von Arbeitnehmern deutscher Aktiengesellschaften ins Ausland: eine Studie zur Auslandsentsendung deutscher Mitarbeiter (Expatriates).* Katholische Universität Eichstätt-Ingolstadt, Wirtschaftswissenschaftliche Fakultät.

Egger, H. and U. Kreickemeier (2013). Why foreign ownership may be good for you. *International Economic Review 54*(2), 693–716.

Ethier, W. J. and H. Horn (1990). Managerial control of international firms and patterns of direct investment. *Journal of International Economics 28*(1–2), 25–45.

Fort, T. C. (2017). Technology and Production Fragmentation: Domestic versus Foreign Sourcing. *The Review of Economic Studies 84*(2), 650–687.

Friedrich, B. (2016). Trade shocks, firm hierarchies and wage inequality. *Yale University*, mimeo.

Gardner, J. (2013). Misused English words and expressions in EU publications. European Court of Auditors Secretariat General Translation Directorate.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *The Journal of Political Economy 108*(5), 874–904.

Garicano, L. and T. N. Hubbard (2007). Managerial leverage is limited by the extent of the market: Hierarchies, specialization, and the utilization of lawyers human capital. *Journal of Law and Economics 50*(1), 1–43.

Garicano, L. and E. Rossi-Hansberg (2015). Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics 7*(1), 1–30.

Girma, S. and H. Görg (2007). Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *Journal of International Economics 72*(1), 97–112.

Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *The Quarterly Journal of Economics 128*(2), 861–915.

Hall, A. and M. Tiemann (2006). BIBB/BAuA Employment Survey of the Working Population on Qualification and Working Conditions in Germany 2006, suf_1.0. Technical report, GESIS Cologne, Germany (data access); Federal Institute of Vocational Education and Training, Bonn doi:10.4232/1.4820.

Harrison, A. and A. Rodríguez-Clare (2010). Trade, Foreign Investment, and Industrial Policy for Developing Countries. *Handbook of Development Economics 5*, 4039–4214.

Hausmann, R. (2016, Januray). Should Buisness Travel Be Obsolete?

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004). Export versus FDI with heterogeneous firms. *The American Economic Review 94*(1), 300–316.

Heyman, F., F. Sjöholm, and P. G. Tingvall (2007). Is there really a foreign ownership wage premium? Evidence from matched employer–employee data. *Journal of International Economics 73*(2), 355–376.

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.

Irarrazabal, A., A. Moxnes, and L. D. Opromolla (2013). The margins of multinational production and the role of intrafirm trade. *The Journal of Political Economy 121*(1), 74–126.

Javorcik, B. (2012). Does FDI bring good jobs to host countries? *Background Paper for the World Development Report 2013, World Bank*.

Kalnins, A. and F. Lafontaine (2013). Too far away? The effect of distance to headquarters on business establishment performance. *American Economic Journal: Microeconomics 5*(3), 157–179.

Keller, W. (2004). International technology diffusion. *Journal of Economic Literature 42*(3), 752–782.

Keller, W. and S. R. Yeaple (2013). The Gravity of Knowledge. *American Economic Review 103*(4), 1414–44.

Lipponer, A. (2009). Microdatabase Direct Investment (MiDi) - A brief guide. Technical report, Deutsche Bundesbank.

Lubotsky, D. and M. Wittenberg (2006). Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics 88*(3), 549–562.

Malchow-Moller, N., J. R. Markusen, and B. Schjerning (2013). Foreign Firms, Domestic Wages. *The Scandinavian Journal of Economics 115*(2), 292–325.

Marin, D., J. S. Schymik, and A. Tarasov (2015). Trade in Tasks and the Organization of Firms. *University of Munich*, mimeo.

Markusen, J. R. (1984). Multinationals, multi-plant economies, and the gains from trade. *Journal of International Economics 16*(3), 205–226.

Melitz, J. and F. Toubal (2014). Native language, spoken language, translation and trade. *Journal of International Economics 93*(2), 351–363.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Oldenski, L. (2012). Export versus FDI and the communication of complex information. *Journal of International Economics 87*(2), 312–322.

Qian, Y. (1994). Incentives and loss of control in an optimal hierarchy. *The Review of Economic Studies 61*(3), 527–544.

Rajan, R. G. and L. Zingales (1998). Financial Dependence and Growth. *The American Economic Review 88*(3), 559–586.

Ramondo, N. (2014). A quantitative approach to multinational production. *Journal of International Economics 93*(1), 108–122.

Schaffer, M. E. (2015). xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. *Statistical Software Components*.

Schild, C.-J. and F. Walter (2015). Microdatabase Direct Investment 1999-2013, Data Report 2015-01 - Metadata Version 2-2. Technical report, Deutsche Bundesbank Research Data and Service Centre.

Spolaore, E. and R. Wacziarg (2009). The diffusion of development. *The Quarterly Journal of Economics 124*(2), 469–529.

Tomiura, E. (2007). Foreign outsourcing, exporting, and FDI: A productivity comparison at the firm level. *Journal of International Economics 72(1)*, 113–127.

UNCTAD (2004). World Investment Report 2004. The shift towards services.

Yeaple, S. R. (2013). Scale, scope, and the international expansion strategies of multiproduct firms. *NBER Working Paper 19166*.

Appendix

A Parameter restriction

Assumption 1. *The exogenous parameters $\bar{z}, \lambda$, and $\{c_j, \theta_{j0}\}_{j=0}^1$, as well as those exogenous parameters that are contained in $\{q_j\}_{j=0}^1$ and $\{w_j\}_{j=0}^1$, fulfil the following parameter restrictions:*

$$e^{\lambda\bar{z}} \leq \frac{\lambda\sum_{j=0}^1 q_j w_j(1 + c_j z_j + \frac{1}{\lambda}c_j) + \lambda e^{\lambda z_h} w_0(1 + c_0 z_h)\sum_{j=0}^1 \mathbf{1}(z_j = \bar{z} - z_h)q_j\theta_{j0}}{\sum_{j=0}^1 \mathbf{1}(z_j = \bar{z} - z_h)q_j w_j c_j} \tag{A.1}$$

where z_h and z_j are defined by equations (7), (8), and (9); and

$$\theta_{00}e^{-\lambda(\bar{z}-z_h)}(1 + c_0 z_h) - c_0 z_h \leq 0 \tag{A.2}$$

where z_h is defined by equation (10). Furthermore, the fixed costs of FDI f^I ensure that entrepreneurs only self-select into FDI if they draw a knowledge level $\bar{z}$ such that

$$\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(1 + c_0 z_h) - w_1 c_1 z_h \leq 0 \tag{A.3}$$

where z_h is defined by $\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0 + \lambda(1 + c_0 z_h)) - w_1 c_1 = 0$.

Assumption 1 restricts the parameter space such that the following requirements hold:

1. The set of possible values for $\bar{z}$ is such that the employees never learn knowledge that the entrepreneur would not adopt were he free to choose the overall knowledge level (upper bound).
2. Both employees at the headquarters and employees in the production plants are optimally involved in the domestic and foreign production process (lower bound).

To derive equation (A.1), assume that the entrepreneur chooses the total knowledge level z^* subject to the constraint that it cannot exceed the knowledge draw: $z^* \leq \bar{z}$. For the simplicity of exposition, I study a domestic firm, where the constraints (2) to (4) substitute z_j, n_j and n_h:

$$\begin{aligned} \min_{z_h, z^*} \quad & \frac{q_0 w_0}{1 - e^{-\lambda z^*}}\left(1 + c_0(z^* - z_h) + \theta_{00}e^{-\lambda(z^*-z_h)}(1 + c_0 z_h)\right) \\ \text{s.t.} \quad & z^* \leq \bar{z} \end{aligned}$$

The corresponding Lagrangian equation is given by

$$\mathcal{L} = \frac{q_0 w_0}{1 - e^{-\lambda z^*}}\left(1 + c_0(z^* - z_h) + \theta_{00}e^{-\lambda(z^*-z_h)}(1 + c_0 z_h)\right) + \phi(z^* - \bar{z})$$

A necessary condition for $z^* = \bar{z}$ is $\phi \geq 0$. $\phi \geq 0$ if $\bar{z} \leq \bar{z}_{\max}$, where $\bar{z}_{\max}$ is implicitly defined by

$$e^{\lambda\bar{z}_{\max}} = \frac{1}{c_0}\left(\lambda\left(1 + c_0(\bar{z}_{\max} - z_h) + \frac{1}{\lambda}c_0\right) + \lambda\theta_{00}e^{\lambda z_h}(1 + c_0 z_h)\right)$$

and z_h is the solution of (10) given $\bar{z}_{\max}$. Analogously, $\phi \geq 0$ for an MNE with two plants whenever

$$e^{\lambda\bar{z}} \leq \frac{\lambda\sum_{j=0}^1 q_j w_j(1 + c_j z_j + \frac{1}{\lambda}c_j) + \lambda e^{\lambda z_h} w_0(1 + c_0 z_h)\sum_{j=0}^1 \mathbf{1}(z_j = \bar{z} - z_h)q_j\theta_{j0}}{\sum_{j=0}^1 \mathbf{1}(z_j = \bar{z} - z_h)q_j w_j c_j}$$

Concerning the lower bound, in an organization with a headquarters and a production affiliate, each worker learns z_h units of knowledge less than in an organization in which one layer of employees learns all of the firm's knowledge. A two-layer organization is thus optimal if the resulting cost decrease $c_0 z_h$ per worker exceeds the costs of hiring managers $\theta_{00}e^{-\lambda(\bar{z}-z_h)}(1 + c_0 z_h)$ (equation A.2). The condition decreases with $\bar{z}$. It ensures $z_h > 0$, because otherwise it would be better to produce without a headquarters. An analogous condition holds for the foreign country (equation A.3).

B The optimal organization of knowledge

B.1 First order conditions and asymmetry of workers' knowledge

$$\mathcal{L} = \sum_{j=0}^{1} n_j w_j(1+c_j z_j) + n_h w_0(1+c_0 z_h) + w_0 + \sum_{j=0}^{1} \xi_j \left[q_j - n_j(1-e^{-\lambda \bar{z}})\right] + \sum_{j=0}^{1} \phi_j \left[\bar{z} - z_h - z_j\right]$$
$$+ \kappa \left[\sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} - n_h\right] - \sum_{j=0}^{1} \upsilon_j n_j - \upsilon_h n_h - \sum_{j=0}^{1} \nu_j z_j - \nu_h z_h + \sum_{j=0}^{1} \bar{\nu}_j (z_j - \bar{z}) + \bar{\nu}_h (z_h - \bar{z})$$

$$\frac{\partial \mathcal{L}}{\partial n_j} = w_j(1+c_j z_j) - \xi_j(1-e^{-\lambda \bar{z}}) + \kappa \theta_{j0} e^{-\lambda z_j} - \upsilon_j = 0$$

$$\frac{\partial \mathcal{L}}{\partial z_j} = n_j w_j c_j - \phi_j - \lambda \kappa n_j \theta_{j0} e^{-\lambda z_j} - \nu_j + \bar{\nu}_j = 0$$

$$\frac{\partial \mathcal{L}}{\partial n_h} = w_0(1+c_0 z_h) - \kappa - \upsilon_h = 0$$

$$\frac{\partial \mathcal{L}}{\partial z_h} = n_h w_0 c_0 - \sum_{j=0}^{1} \phi_j - \nu_h + \bar{\nu}_h = 0$$

$$\frac{\partial \mathcal{L}}{\partial \xi_j} = q_j - n_j(1-e^{-\lambda \bar{z}}) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \phi_j} = \bar{z} - z_h - z_j = 0$$

$$\frac{\partial \mathcal{L}}{\partial \kappa} = \sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} - n_h = 0$$

The workers' knowledge levels $\{z_j\}_{j=0}^{1}$ are asymmetric if $e^{-\lambda z_{\hat{j}}} \leq e^{-\lambda(\bar{z}-z_h)}$, with $z_{\hat{j}} > \bar{z} - z_h$. This is possible if $w_{\hat{j}} c_{\hat{j}} \leq e^{-\lambda(\bar{z}-z_h)} \lambda \theta_{\hat{j}0} w_0(1+c_0 z_h)$. The (binding) inequality implicitly defines a threshold $\bar{z}$ for an asymmetric solution. The threshold is increasing in $w_{\hat{j}}$, $c_{\hat{j}}$ and decreasing in $\theta_{\hat{j}0}$.

B.2 Comparative statics: Proposition 1, Corollary 1

Table B.1 lists the comparative statics for the number of workers and managers.

Table B.1: Comparative statics: workers and managers

# workers, managers/ Model parameters	θ_{j0}	c_j	w_j	q_j	$\bar{z}$	λ
# workers n_j	0	0	0	+	-	-
# managers n_h, domestic firm	+	+	0	+	-	-
# managers n_h, MNE, $z_0 = z_1 = \bar{z} - z_h$	+	$+^{*}$	$+^{*}$	+	-	-
# managers n_h, MNE, $z_j = \bar{z} - z_h$	$+^{**}$	+	+	+	-	-
# managers n_h, MNE, $z_j > \bar{z} - z_h$	0	+	+	+	-	-

The table displays the effects of the parameters on the number of workers and managers (+ positive, − negative, 0 none). Results denoted * only apply to $j = 1$. Results denoted ** hold if $q_{\bar{j}} \theta_{\bar{j}0} e^{-\lambda(\bar{z}-z_h)} \lambda(1+c_0 z_h) > q_{\hat{j}} \theta_{\hat{j}0} e^{-\lambda z_{\hat{j}}} c_0$, where the constraint $z_j = \bar{z} - z_h$ is binding $\bar{j}$ and slack in $\hat{j}$.

Number of production workers.

$$\frac{\partial n_j}{\partial q_j} = \frac{1}{1-e^{-\lambda \bar{z}}} > 0; \qquad \frac{\partial n_j}{\partial \bar{z}} = -\frac{q_j \lambda e^{-\lambda \bar{z}}}{(1-e^{-\lambda \bar{z}})^2} < 0; \qquad \frac{\partial n_j}{\partial \lambda} = -\frac{q_j \bar{z} e^{-\lambda \bar{z}}}{(1-e^{-\lambda \bar{z}})^2} < 0$$

Number of managers.

$$
\begin{aligned}
\frac{\partial n_h}{\partial \theta_{j0}} &= \frac{q_j e^{-\lambda z_j} - \lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial \theta_{j0}}}{1 - e^{-\lambda \bar{z}}} \\
\frac{\partial n_h}{\partial c_j} &= -\frac{\lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial c_j}}{1 - e^{-\lambda \bar{z}}} \\
\frac{\partial n_h}{\partial w_j} &= -\frac{\lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial w_j}}{1 - e^{-\lambda \bar{z}}} \\
\frac{\partial n_h}{\partial q_j} &= \frac{\theta_{j0} e^{-\lambda z_j} - \lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial q_j}}{1 - e^{-\lambda \bar{z}}} \\
\frac{\partial n_h}{\partial \bar{z}} &= -\frac{\lambda e^{-\lambda \bar{z}}}{(1 - e^{-\lambda \bar{z}})^2} \sum_j q_j \theta_{j0} e^{-\lambda z_j} - \frac{1}{1 - e^{-\lambda \bar{z}}} \sum_j \lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial \bar{z}} \\
\frac{\partial n_h}{\partial \lambda} &= \sum_{j=0}^{1} \frac{(-z_j q_j \theta_{j0} e^{-\lambda z_j} - \lambda q_j \theta_{j0} e^{-\lambda z_j} \frac{dz_j}{d\lambda})(1 - e^{-\lambda \bar{z}}) - \bar{z} e^{-\lambda \bar{z}} q_j \theta_{j0} e^{-\lambda z_j}}{1 - e^{-\lambda \bar{z}}}
\end{aligned}
$$

I determine the signs using the derivatives of the knowledge levels at the end of each subsection.

Proposition 1 distinguishes three cases: the domestic firm, the MNE with binding knowledge constraint at both plants, and the MNE with binding knowledge constraint at only one plant. The following subsections contain the proofs of Proposition 1 separately for each case.

B.2.1 Proposition 1, domestic firm

Managerial knowledge z_h. By the implicit function theorem, $\frac{dz_h}{dx_0} = -\frac{\frac{d(10)}{dx_0}}{\frac{d(10)}{dz_h}}$. The sign of $\frac{dz_h}{dx_0}$ is given by $-\frac{d(10)}{dx_0}$ because $\frac{d(10)}{dz_h} = \lambda\theta_{00} e^{-\lambda(\bar{z}-z_h)}(2c_0 + \lambda(1 + c_0 z_h)) > 0$.

$$
\begin{aligned}
\frac{d(10)}{d\theta_{00}} &= e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1 + c_0 z_h)) > 0 & \Rightarrow \frac{dz_h}{d\theta_{00}} &< 0 \\
\frac{d(10)}{dc_0} &= \theta_{00} e^{-\lambda(\bar{z}-z_h)}(1 + \lambda z_h) - 1 < 0 & \Rightarrow \frac{dz_h}{dc_0} &> 0 \\
\frac{d(10)}{dw_0} &= \frac{d(10)}{dq_0} = 0 & \Rightarrow \frac{dz_h}{dw_0} = \frac{dz_h}{dq_0} &= 0 \\
\frac{d(10)}{d\bar{z}} &= -\lambda\theta_{00} e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1 + c_0 z_h)) < 0 & \Rightarrow \frac{dz_h}{d\bar{z}} &> 0 \\
\frac{d(10)}{d\lambda} &= \theta_{00} e^{-\lambda(\bar{z}-z_h)}((1 + c_0 z_h) - (\bar{z} - z_h)(c_0 + \lambda(1 + c_0 z_h))) & \Rightarrow \frac{dz_h}{d\lambda} &\lesseqgtr 0
\end{aligned}
$$

Production knowledge. By $z_0 = \bar{z} - z_h$, $\frac{dz_0}{d\theta_{00}} > 0$, $\frac{dz_0}{dc_0} < 0$, $\frac{dz_0}{dw_0} = \frac{dz_0}{dq_0} = 0$, $\frac{dz_0}{d\lambda} \lesseqgtr 0$. $\frac{dz_0}{d\bar{z}} = 1 - \frac{dz_h}{d\bar{z}} > 0$ by $\frac{dz_h}{d\bar{z}} < 1$.

Number of managers. $\frac{\partial n_h}{\partial \theta_{00}} > 0$ by $\frac{\partial z_0}{\partial \theta_{00}} < \frac{1}{\lambda\theta_{00}}$; $\frac{\partial n_h}{\partial c_0} > 0$ by $\frac{\partial z_0}{\partial c_0} < 0$; $\frac{\partial n_h}{\partial w_0} = 0$ by $\frac{\partial z_0}{\partial w_0} = 0$; $\frac{\partial n_h}{\partial q_0} > 0$ by $\frac{\partial z_0}{\partial q_0} = 0$; $\frac{\partial n_h}{\partial \bar{z}} < 0$ by $\frac{\partial z_0}{\partial \bar{z}} > 0$; $\frac{\partial n_h}{\partial \lambda} < 0$ because ambiguous terms cancel.

B.2.2 Proposition 1, MNE, $\mathbf{z_0 = z_1 = \bar{z} - z_h}$

Managerial knowledge z_h. By the implicit function theorem, $\frac{dz_h}{dx_j} = -\frac{\frac{d(9)}{dx_j}}{\frac{d(9)}{dz_h}}$.

The sign of $\frac{dz_h}{dx_j}$ is given by $-\frac{d(9)}{dx_j}$ as $\frac{d(9)}{dz_h} = \lambda e^{-\lambda(\bar{z}-z_h)} w_0(2c_0 + \lambda(1+c_0 z_h)) \sum_{j=0}^{1} q_j \theta_{j0} > 0$.

$$
\begin{aligned}
\frac{d(9)}{d\theta_{j0}} &= q_j e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) > 0 && \Rightarrow \frac{dz_h}{d\theta_{j0}} < 0 \\
\frac{d(9)}{dc_1} &= -q_1 w_1 < 0 && \Rightarrow \frac{dz_h}{dc_1} > 0 \\
\frac{d(9)}{dc_0} &= e^{-\lambda(\bar{z}-z_h)} w_0(1+\lambda z_h) \sum_{j=0}^{1} q_j \theta_{j0} - q_0 w_0 && \Rightarrow \frac{dz_h}{dc_0} \lesseqgtr 0 \\
\frac{d(9)}{dw_1} &= -q_1 c_1 < 0 && \Rightarrow \frac{dz_h}{dw_1} > 0 \\
\frac{d(9)}{dw_0} &= e^{-\lambda(\bar{z}-z_h)} (c_0 + \lambda(1+c_0 z_h)) \sum_{j=0}^{1} q_j \theta_{j0} - q_0 c_0 > 0 && \Rightarrow \frac{dz_h}{dw_0} < 0 \\
\frac{d(9)}{dq_j} &= \theta_{j0} e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) - w_j c_j && \Rightarrow \frac{dz_h}{dq_j} \lesseqgtr 0 \\
\frac{d(9)}{d\bar{z}} &= -\lambda e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) \sum_{j=0}^{1} q_j \theta_{j0} < 0 && \Rightarrow \frac{dz_h}{d\bar{z}} > 0 \\
\frac{d(9)}{d\lambda} &= e^{-\lambda(\bar{z}-z_h)} w_0((1+c_0 z_h) - (\bar{z}-z_h)(c_0 + \lambda(1+c_0 z_h))) \sum_{j=0}^{1} q_j \theta_{j0} && \Rightarrow \frac{dz_h}{d\lambda} \lesseqgtr 0
\end{aligned}
$$

Production knowledge. By $z_j = \bar{z} - z_h$, $\frac{dz_j}{d\theta_{j0}} > 0$, $\frac{dz_1}{dc_1} < 0$, $\frac{dz_1}{dw_1} < 0$, $\frac{dz_j}{d\lambda} \lesseqgtr 0$. $\frac{dz_j}{d\bar{z}} = 1 - \frac{dz_h}{d\bar{z}} > 0$ by $\frac{dz_h}{d\bar{z}} < 1$.

Whether $\frac{d(9)}{dq_j}$ is positive or negative depends on $\frac{w_j c_j}{\theta_{j0}}$. If $\frac{w_0 c_0}{\theta_{00}} > \frac{w_1 c_1}{\theta_{10}}$, $\frac{d(9)}{dq_0} < 0$, so $\frac{dz_h}{dq_0} > 0$, $\frac{dz_0}{dq_0} < 0$, $\frac{dz_h}{dq_1} < 0$, $\frac{dz_1}{dq_1} > 0$. Analogously, if $\frac{w_0 c_0}{\theta_{00}} < \frac{w_1 c_1}{\theta_{10}}$, $\frac{d(9)}{dq_1} < 0$, so $\frac{dz_h}{dq_1} > 0$, $\frac{dz_1}{dq_1} < 0$, $\frac{dz_h}{dq_0} < 0$, $\frac{dz_0}{dq_0} > 0$.

Number of managers. $\frac{\partial n_h}{\partial \theta_{j0}} > 0$ by $\frac{\partial z_j}{\partial \theta_{j0}} < \frac{1}{\lambda \theta_{j0}}$; $\frac{\partial n_h}{\partial c_1} > 0$ by $\frac{\partial z_1}{\partial c_1} < 0$; $\frac{\partial n_h}{\partial w_1} > 0$ by $\frac{\partial z_1}{\partial w_1} < 0$; $\frac{\partial n_h}{\partial q_j} > 0$ because ambiguous terms cancel; $\frac{\partial n_h}{\partial \bar{z}} < 0$ by $\frac{\partial z_j}{\partial \bar{z}} > 0$; $\frac{\partial n_h}{\partial \lambda} < 0$ because ambiguous terms cancel.

B.2.3 Proposition 1, MNE, $\mathbf{z_{\hat{j}} > \bar{z} - z_h, z_{\bar{j}} = \bar{z} - z_h}$

The interior solution to the MNE's optimization problem is given by a system of two equations in two unknowns, $z_{\hat{j}}$ and z_h, where $z_j = \bar{z} - z_h$ is binding in $\bar{j}$ and slack in $\hat{j}$:

$$
\begin{aligned}
0 &= w_{\hat{j}} c_{\hat{j}} - \theta_{\hat{j}0} e^{-\lambda z_{\hat{j}}} \lambda w_0 (1 + c_0 z_h) \\
0 &= q_{\bar{j}} \theta_{\bar{j}0} e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) + q_{\hat{j}} \theta_{\hat{j}0} e^{-\lambda z_{\hat{j}}} w_0 c_0 - q_{\bar{j}} w_{\bar{j}} c_{\bar{j}}
\end{aligned}
$$

I differentiate the system of equations with respect to the parameters x_j and solve for $\frac{dz_h}{dx_j}$ and $\frac{dz_{\hat{j}}}{dx_j}$.

Managerial knowledge z_h. The denominator of $\frac{dz_h}{dx_j}$ is given by $d \equiv q_{\bar{j}} \theta_{\bar{j}0} e^{-\lambda(\bar{z}-z_h)} \lambda w_0 (2c_0 + \lambda(1+c_0 z_h))(1+c_0 z_h) - q_{\hat{j}} \theta_{\hat{j}0} e^{-\lambda z_{\hat{j}}} w_0 c_0^2$. A solution to the first order condition (9) with a positive value for d exists $\forall \bar{z}$ s.t.

$$
e^{\lambda \bar{z}} \geq \frac{q_{\bar{j}} \theta_{\bar{j}0} e^{\lambda z_h^*} \lambda w_0 (c_0 + \lambda(1+c_0 z_h^*))(1+c_0 z_h^*)}{q_{\bar{j}} w_{\bar{j}} c_{\bar{j}} \lambda (1+c_0 z_h^*) - q_{\hat{j}} w_{\hat{j}} c_{\hat{j}} c_0} \tag{B.1}
$$

where z_h^* is implicitly defined by $q_{\bar{j}}\theta_{\bar{j}0}e^{-\lambda(\bar{z}-z_h^*)}\lambda^2 w_0(1+c_0 z_h^*)^2(2c_0+\lambda(1+c_0 z_h^*)) = q_{\hat{j}}w_{\hat{j}}c_{\hat{j}}c_0^2$. The first order condition is a U-shaped function of z_h. Condition (B.1) ensures that the first order condition is negative at its minimum, so the roots of the first order condition exist. It is possible to ensure that only firms with values of $\bar{z}$ for which the asymmetric solution exists select into FDI by assuming that f^I is sufficiently large. Multiplied by $\lambda^2 q_{\hat{j}}\theta_{\hat{j}0}e^{-\lambda z_{\hat{j}}}w_0 > 0$, the term d is the determinant of the Hessian matrix of the optimization problem that is positive at the minimum of the optimization problem.

This implies:

$$
\begin{aligned}
\frac{\partial z_h}{\partial \theta_{\bar{j}0}} &= -d^{-1}(1+c_0 z_h)q_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)) \quad < 0 \\
\frac{\partial z_h}{\partial \theta_{\hat{j}0}} &= 0 \\
\frac{\partial z_h}{\partial q_{\bar{j}}} &= -d^{-1}(1+c_0 z_h)\left(\theta_{\bar{j}0}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)) - w_{\bar{j}}c_{\bar{j}}\right) \quad > 0 \\
\frac{\partial z_h}{\partial q_{\hat{j}}} &= -d^{-1}\frac{1}{\lambda}c_0 w_{\hat{j}}c_{\hat{j}} \quad < 0 \\
\frac{dz_h}{d\bar{z}} &= d^{-1}\lambda(1+c_0 z_h)q_{\bar{j}}\theta_{\bar{j}0}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)) \quad > 0 \\
\frac{dz_h}{d\lambda} &= d^{-1}\Big(q_{\bar{j}}\theta_{\bar{j}0}e^{-\lambda(\bar{z}-z_h)}w_0((\bar{z}-z_h)(c_0+\lambda(1+c_0 z_h))(1+c_0 z_h)-(1+c_0 z_h)^2) \\
&\quad + q_{\hat{j}}\theta_{\hat{j}0}e^{-\lambda z_{\hat{j}}}w_0 c_0\frac{1}{\lambda}(1+c_0 z_h)\Big) \quad \lesseqgtr 0
\end{aligned}
$$

Concerning the wages and the learning costs, it is necessary to distinguish two cases.

1. The knowledge constraint $z_j = \bar{z} - z_h$ is binding in $j = 0$, slack in $j = 1$.

$$
\begin{aligned}
\frac{\partial z_h}{\partial c_0} &= d^{-1}\Big(\left(q_0 w_0 - q_0\theta_{00}e^{-\lambda(\bar{z}-z_h)}w_0(1+\lambda z_h) - q_1\theta_{10}e^{-\lambda z_1}w_0\right)(1+c_0 z_h) \\
&\quad + q_1\theta_{10}e^{-\lambda z_1}w_0 c_0 z_h\Big) \quad > 0 \\
\frac{\partial z_h}{\partial c_1} &= -d^{-1}\frac{1}{\lambda}q_1 w_1 c_0 \quad < 0 \\
\frac{\partial z_h}{\partial w_0} &= d^{-1}(1+c_0 z_h)q_1\theta_{10}e^{-\lambda z_1}c_0 \quad > 0 \\
\frac{\partial z_h}{\partial w_1} &= -d^{-1}\frac{1}{\lambda}q_1 c_1 c_0 \quad < 0
\end{aligned}
$$

2. The knowledge constraint $z_j = \bar{z} - z_h$ is binding in $j = 1$, slack in $j = 0$.

$$
\begin{aligned}
\frac{\partial z_h}{\partial c_0} &= -d^{-1}\Big(\left(q_1\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(1+\lambda z_h) + q_0\theta_{00}e^{-\lambda z_0}w_0\right)(1+c_0 z_h) \\
&\quad - \frac{1}{\lambda}q_1 c_0\left(w_0 - \theta_{00}e^{-\lambda z_0}\lambda w_0 z_h\right)\Big) \quad < 0 \\
\frac{\partial z_h}{\partial c_1} &= d^{-1}q_1 w_1(1+c_0 z_h) \quad > 0 \\
\frac{\partial z_h}{\partial w_0} &= -d^{-1}(1+c_0 z_h)\left(q_1\theta_{10}e^{-\lambda(\bar{z}-z_h)}(c_0+\lambda(1+c_0 z_h)) + q_0\theta_{00}e^{-\lambda z_0}c_0\right) \quad < 0 \\
\frac{\partial z_h}{\partial w_1} &= d^{-1}q_1 c_1(1+c_0 z_h) \quad > 0
\end{aligned}
$$

Production knowledge. For the country with the binding knowledge constraint $z_{\bar{j}} = \bar{z} - z_h$, $\frac{\partial z_{\bar{j}}}{\partial x_j} = -\frac{\partial z_h}{\partial x_j}$, $x_j \in \{\lambda, \theta_{j0}, c_j, w_j, q_j\}$. Consequently, $\frac{\partial z_{\bar{j}}}{\partial \theta_{\bar{j}0}} > 0$, $\frac{\partial z_{\bar{j}}}{\partial c_{\bar{j}}} < 0$, $\frac{\partial z_{\bar{j}}}{\partial w_{\bar{j}}} < 0$, $\frac{\partial z_{\bar{j}}}{\partial q_{\bar{j}}} < 0$, $\frac{\partial z_{\bar{j}}}{\partial \lambda} \gtrless 0$.

$\frac{\partial z_{\bar{j}}}{\partial \bar{z}} = 1 - \frac{\partial z_h}{\partial \bar{z}} > 0$ if $\frac{\partial z_h}{\partial \bar{z}} < 1$, i.e., if $q_{\bar{j}}\theta_{\bar{j}0}e^{-\lambda(\bar{z}-z_h)}\lambda(1+c_0z_h) > q_{\hat{j}}\theta_{\hat{j}0}e^{-\lambda z_{\hat{j}}}c_0$.

For the country where the knowledge constraint is slack $\hat{j}$,

$$
\begin{aligned}
\frac{\partial z_{\hat{j}}}{\partial \theta_{\hat{j}0}} &= \frac{1}{\lambda\theta_{\hat{j}0}} && > 0 \\
\frac{\partial z_{\hat{j}}}{\partial q_{\hat{j}}} &= \frac{c_0}{\lambda(1+c_0z_h)}\frac{dz_h}{dq_{\hat{j}}} && < 0 \\
\frac{\partial z_{\hat{j}}}{\partial \bar{z}} &= \frac{c_0}{\lambda(1+c_0z_h)}\frac{dz_h}{d\bar{z}} && > 0 \\
\frac{\partial z_{\hat{j}}}{\partial \lambda} &= -\frac{1}{\lambda}z_{\hat{j}} + \frac{1}{\lambda^2} + \frac{c_0}{\lambda(1+c_0z_h)}\frac{dz_h}{d\lambda} && \lesseqgtr 0.
\end{aligned}
$$

For $\hat{j} = 1$,

$$
\begin{aligned}
\frac{\partial z_1}{\partial c_1} &= -\frac{1}{\lambda c_1} + \frac{c_0}{\lambda(1+c_0z_h)}\frac{\partial z_h}{\partial c_1} && < 0 \\
\frac{\partial z_1}{\partial w_1} &= -\frac{1}{\lambda w_1} + \frac{c_0}{\lambda(1+c_0z_h)}\frac{\partial z_h}{\partial w_1} && < 0
\end{aligned}
$$

Further, $\frac{\partial z_0}{\partial q_1} = -\frac{dz_h}{dq_1} > 0$ and $\frac{\partial z_1}{\partial q_0} = \frac{c_0}{\lambda(1+c_0z_h)}\frac{dz_h}{dq_0} > 0$.

For $\hat{j} = 0$,

$$
\begin{aligned}
\frac{\partial z_0}{\partial c_0} &= -\frac{w_0 - \lambda\theta_{00}e^{-\lambda z_0}w_0z_h}{\lambda^2\theta_{00}e^{-\lambda z_0}w_0(1+c_0z_h)} + \frac{c_0}{\lambda(1+c_0z_h)}\frac{\partial z_h}{\partial c_0} && < 0 \\
\frac{\partial z_0}{\partial w_0} &= \frac{c_0}{\lambda(1+c_0z_h)}\frac{\partial z_h}{\partial w_0} && < 0
\end{aligned}
$$

Further, $\frac{\partial z_1}{\partial q_0} = -\frac{dz_h}{dq_0} > 0$ and $\frac{\partial z_0}{\partial q_1} = \frac{c_0}{\lambda(1+c_0z_h)}\frac{dz_h}{dq_1} > 0$.

Number of managers. $\frac{\partial n_h}{\partial \theta_{\bar{j}0}} > 0$ by $\frac{\partial z_{\bar{j}}}{\partial \theta_{\bar{j}0}} < \frac{1}{\lambda\theta_{\bar{j}0}}$; $\frac{\partial n_h}{\partial \theta_{\hat{j}0}} = 0$ by $\frac{\partial z_{\hat{j}}}{\partial \theta_{\hat{j}0}} = \frac{1}{\lambda\theta_{\hat{j}0}}$; $\frac{\partial n_h}{\partial c_j} > 0$ by $\frac{\partial z_j}{\partial c_j} < 0$; $\frac{\partial n_h}{\partial w_j} > 0$ by $\frac{\partial z_j}{\partial w_j} < 0$; $\frac{\partial n_h}{\partial q_j} > 0$ and $\frac{\partial n_h}{\partial \lambda} < 0$ because ambiguous terms cancel; $\frac{\partial n_h}{\partial \bar{z}} < 0$ unambiguously if $\frac{\partial z_j}{\partial \bar{z}} > 0$, $-\lambda q_{\hat{j}}\theta_{\hat{j}}e^{-\lambda z_{\hat{j}}}\frac{\partial z_j}{\partial \bar{z}}$ and $\frac{\partial z_{\bar{j}}}{\partial \bar{z}}$ cancel if $\frac{\partial z_{\bar{j}}}{\partial \bar{z}} < 0$.

B.2.4 Corollary 1

To show:

$$\frac{d^2 z_0}{d\theta_{00}d\lambda} < 0$$

Domestic firm.

$$
\begin{aligned}
\frac{d^2 z_h}{d\theta_{00}d\lambda} &= \frac{\lambda\theta_{00}(1+c_0z_h)(c_0+\lambda(1+c_0z_h)) + \theta_{00}c_0(2c_0+\lambda(1+c_0z_h))}{\lambda^2\theta_{00}^2(2c_0+\lambda(1+c_0z_h))^2} && > 0 \\
\Rightarrow \quad \frac{d^2 z_0}{d\theta_{00}d\lambda} &= -\frac{d^2 z_h}{d\theta_{00}d\lambda} && < 0
\end{aligned}
$$

Multinational firm, $\mathbf{z_0 = z_1 = \bar{z} - z_h}$.

$$\frac{d^2 z_h}{d\theta_{j0} d\lambda} = \frac{(\lambda(1+c_0 z_h)(c_0 + \lambda(1+c_0 z_h)) + c_0(2c_0 + \lambda(1+c_0 z_h)))q_j \sum_{j=0}^{1} q_j \theta_{j0}}{\lambda^2 (\sum_{j=0}^{1} q_j \theta_{j0})^2 (2c_0 + \lambda(1+c_0 z_h))^2} > 0$$

$$\Rightarrow \quad \frac{d^2 z_j}{d\theta_{j0} d\lambda} = -\frac{d^2 z_h}{d\theta_{j0} d\lambda} \quad < 0$$

Multinational firm, $\mathbf{z_{\hat{j}} > \bar{z} - z_h, z_{\bar{j}} = \bar{z} - z_h}$.

$$\begin{aligned}
\frac{d^2 z_h}{d\theta_{\bar{j}0} d\lambda} = &- d^{-2} q_{\bar{j}} e^{-\lambda(\bar{z}-z_h)} w_0 (1+c_0 z_h) \Big(q_{\hat{j}} \theta_{\hat{j}0} e^{-\lambda z_{\hat{j}}} w_0 c_0^2 (c_0 + \lambda(1+c_0 z_h))(\bar{z} - z_h - z_{\hat{j}}) \\
&- q_{\bar{j}} \theta_{\bar{j}0} e^{-\lambda(\bar{z}-z_h)} w_0 (1+c_0 z_h)(c_0(2c_0 + \lambda(1+c_0 z_h)) + \lambda(1+c_0 z_h)(c_0 + \lambda(1+c_0 z_h))) \\
&- q_{\hat{j}} \theta_{\hat{j}0} e^{-\lambda z_{\hat{j}}} w_0 c_0^2 (1+c_0 z_h) \Big) > 0 \quad \text{by } z_{\hat{j}} > \bar{z} - z_h
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \quad \frac{\partial^2 z_{\bar{j}}}{\partial\theta_{\bar{j}0} \partial\lambda} &= -\frac{d^2 z_h}{d\theta_{\bar{j}0} d\lambda} \quad < 0 \quad \text{for } z_{\bar{j}} = \bar{z} - z_h \\
\frac{\partial^2 z_{\hat{j}}}{\partial\theta_{\hat{j}0} \partial\lambda} &= -\frac{1}{\lambda^2 \theta_{\hat{j}0}} \quad < 0 \quad \text{for } z_{\hat{j}} > \bar{z} - z_h
\end{aligned}$$

B.3 Managerial knowledge in MNEs and non-MNEs: Proposition 2

$\theta_{00} w_1 c_1 < \theta_{10} w_0 c_0$, so the knowledge constraint is binding in the home country: $z_1 \geq z_0 = \bar{z} - z_h$. Take a domestic firm and an MNE with the same knowledge level $\bar{z}$. A comparison of equations (9) and (10) shows that $z_h^I < z_h^D$, i.e., MNEs assign less knowledge to the headquarters than domestic producers with the same total knowledge $\bar{z}$. At $z_h = z_h^D$, equation (9) is not fulfilled, but positive. As equation (9) is increasing in z_h, $z_h^I < z_h^D$.

For the MNE wage premiums studied below, the comparison with foreign domestic firms is also relevant. The proof applies to the comparison of MNEs and domestic firms in the foreign country if $c_1 \geq c_0$ because this condition ensures that domestic firms in the foreign country assign at least as much knowledge to the headquarters as domestic firms in the domestic country.

B.4 Extension: Knowledge gaps

Set-up. The entrepreneur solves the optimization problem without the knowledge constraint (3).

$$\begin{aligned}
C(\bar{z}, q_0, w_0, q_1, w_1) = &\min_{\{n_j, z_j\}_{j=0}^{1}, n_h, z_h} \sum_{j=0}^{1} n_j w_j (1 + c_j z_j) + n_h w_0 (1 + c_0 z_h) + w_0 \\
\text{s.t.} \quad n_j(1 - e^{-\lambda \bar{z}}) &\geq q_j \qquad \forall j \text{ s.t. } z_j \geq \bar{z} - z_h \\
n_j(1 - e^{-\lambda z_j} + e^{-\lambda(\bar{z}-z_h)} - e^{-\lambda \bar{z}}) &\geq q_j \qquad \forall j \text{ s.t. } z_j < \bar{z} - z_h \\
n_h &\geq \sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} \\
n_h \geq 0,\, z_h \geq 0, z_h \leq \bar{z}; \quad & n_j \geq 0,\, z_j \geq 0, z_j \leq \bar{z} \qquad \forall j
\end{aligned}$$

Neither knowledge gaps at both locations, i.e., $z_j < \bar{z} - z_h \,\forall j$, nor overlaps at both locations, i.e., $z_j > \bar{z} - z_h \,\forall j$, are optimal: in the former case, the MNE could produce more output at the same costs by shifting managerial knowledge to close the gap; the latter case entails waste of resources.

As the choice set of the MNE is constrained—$0 < z_k < \bar{z}$, $k = j, h$—a solution featuring a knowledge gap at one and an overlap at the other location does not always exist.

Lagrangian equation and first-order conditions.

$$
\begin{aligned}
\mathcal{L} = & \sum_{j=0}^{1} n_j w_j (1 + c_j z_j) + n_h w_0 (1 + c_0 z_h) + w_0 \\
& + \sum_{j=0}^{1} \xi_j \left[q_j - n_j (1 - e^{-\lambda \bar{z}} + \mathbf{1}(z_j < \bar{z} - z_h)(e^{-\lambda(\bar{z} - z_h)} - e^{-\lambda z_j})) \right] \\
& + \kappa \left[\sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} - n_h \right] - \sum_{j=0}^{1} \upsilon_j n_j - \upsilon_h n_h - \sum_{j=0}^{1} \nu_j z_j - \nu_h z_h + \sum_{j=0}^{1} \bar{\nu}_j (z_j - \bar{z}) + \bar{\nu}_h (z_h - \bar{z})
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial n_j} &= w_j (1 + c_j z_j) - \xi_j (1 - e^{-\lambda \bar{z}} + \mathbf{1}(z_j < \bar{z} - z_h)(e^{-\lambda(\bar{z} - z_h)} - e^{-\lambda z_j})) + \kappa \theta_{j0} e^{-\lambda z_j} - \upsilon_j = 0 \\
\frac{\partial \mathcal{L}}{\partial z_j} &= n_j w_j c_j - \mathbf{1}(z_j < \bar{z} - z_h) \xi_j \lambda e^{-\lambda z_j} n_j - \lambda \kappa n_j \theta_{j0} e^{-\lambda z_j} - \nu_j + \bar{\nu}_j = 0 \\
\frac{\partial \mathcal{L}}{\partial n_h} &= w_0 (1 + c_0 z_h) - \kappa - \upsilon_h = 0 \\
\frac{\partial \mathcal{L}}{\partial z_h} &= n_h w_0 c_0 - \mathbf{1}(z_j < \bar{z} - z_h) \xi_j n_j \lambda e^{-\lambda(\bar{z} - z_h)} - \nu_h + \bar{\nu}_h = 0 \\
\frac{\partial \mathcal{L}}{\partial \xi_j} &= q_j - n_j (1 - e^{-\lambda \bar{z}} + \mathbf{1}(z_j < \bar{z} - z_h)(e^{-\lambda(\bar{z} - z_h)} - e^{-\lambda z_j})) = 0 \\
\frac{\partial \mathcal{L}}{\partial \kappa} &= \sum_{j=0}^{1} n_j \theta_{j0} e^{-\lambda z_j} - n_h = 0
\end{aligned}
$$

Insights. The knowledge level of production workers is determined by

$$
\begin{aligned}
e^{-\lambda z_{\hat{j}}} &= \frac{w_{\hat{j}} c_{\hat{j}}}{\lambda w_0 (1 + c_0 z_h) \theta_{\hat{j}0}} \quad \text{if } z_{\hat{j}} > \bar{z} - z_h \\
e^{-\lambda z_{\tilde{j}}} &= \frac{w_{\tilde{j}} c_{\tilde{j}}}{\lambda w_0 (1 + c_0 z_h) \theta_{\tilde{j}0} + \lambda \xi_{\tilde{j}}} \quad \text{if } z_{\tilde{j}} < \bar{z} - z_h
\end{aligned}
$$

where $\xi_{\tilde{j}} = \frac{w_{\tilde{j}}(1 + c_{\tilde{j}} z_{\tilde{j}}) + w_0 (1 + c_0 z_h) \theta_{\tilde{j}0} e^{-\lambda z_{\tilde{j}}}}{1 - e^{-\lambda \bar{z}} + e^{-\lambda(\bar{z} - z_h)} - e^{-\lambda z_{\tilde{j}}}}$.
The knowledge gap is more likely in the country with the higher wage, the higher learning costs and the lower communication costs by $w_{\tilde{j}} c_{\tilde{j}} \theta_{\hat{j}0} > w_{\hat{j}} c_{\hat{j}} \theta_{\tilde{j}0}$, which follows from $e^{-\lambda z_{\hat{j}}} < e^{-\lambda(\bar{z} - z_h)} < e^{-\lambda z_{\tilde{j}}}$.

Managerial knowledge is implicitly determined by

$$
w_0 c_0 \sum_{j=0}^{1} \frac{q_j \theta_{j0} e^{-\lambda z_j}}{1 - e^{-\lambda \bar{z}} + \mathbf{1}(z_j < \bar{z} - z_h)(e^{-\lambda(\bar{z} - z_h)} - e^{-\lambda z_j})} - \frac{\mathbf{1}(z_j < \bar{z} - z_h) q_j \lambda e^{-\lambda(\bar{z} - z_h)} \xi_{\tilde{j}}}{1 - e^{-\lambda \bar{z}} + e^{-\lambda(\bar{z} - z_h)} - e^{-\lambda z_j}} = 0.
$$

Managerial knowledge depends on the production quantities $\{q_j\}_{j=0}^{1}$ in both countries, which leads to an interdependence in the organization of knowledge and the marginal costs of production.

The comparative statics with respect to the communication costs are given by

$$
\frac{dz_h}{d\theta_{\hat{j}0}} = 0; \quad \frac{dz_{\hat{j}}}{d\theta_{\hat{j}0}} = \frac{1}{\lambda \theta_{\hat{j}0}} > 0 \quad \text{for } z_{\hat{j}} > \bar{z} - z_h. \tag{B.2}
$$

The communication costs $\theta_{\tilde{j}0}$ have a positive direct effect on $z_{\tilde{j}}$. Due to the non-linearities of the optimization problem, the total effect cannot be signed analytically. It is positive in simulations.

C The implications for MNEs' foreign sales and wages

C.1 Profit maximization: Proposition 3, Corollary 2

C.1.1 Proposition 3

To show: The foreign marginal costs $\xi_1(\cdot)$ of MNEs increase with the communication costs θ_{10}.

We build on the results from Proposition 1. There are two possible cases:

1. When $z_1 = \bar{z} - z_h$, so $\xi_1 = \frac{1}{1-e^{-\lambda\bar{z}}}\left(w_1(1 + c_1(\bar{z} - z_h)) + w_0(1 + c_0 z_h)\theta_{10}e^{-\lambda(\bar{z}-z_h)}\right)$:

$$\frac{\partial \xi_1}{\partial \theta_{10}} = \frac{1}{1-e^{-\lambda\bar{z}}}\left((-w_1 c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0 + \lambda(1 + c_0 z_h)))\frac{\partial z_h}{\partial \theta_{10}} + w_0(1 + c_0 z_h)e^{-\lambda(\bar{z}-z_h)}\right)$$
$$> 0$$

   This result holds both when the knowledge levels are symmetric, i.e. $z_0 = z_1 = \bar{z} - z_h$, and when they are asymmetric, i.e. $z_0 > z_1 = \bar{z} - z_h$.

2. When $z_1 > \bar{z} - z_h$, so $\xi_1(\bar{z}, q_0, w_0, q_1, w_1) = \frac{1}{1-e^{-\lambda\bar{z}}}w_1\left(1 + c_1 z_1 + \frac{1}{\lambda}c_1\right)$:

$$\frac{\partial \xi_1}{\partial \theta_{10}} = \frac{1}{1-e^{-\lambda\bar{z}}}w_1 c_1 \frac{\partial z_1}{\partial \theta_{10}} \quad > 0$$

To show: The foreign production quantities and sales are generally lower in countries with higher communication costs θ_{10}.

The profit maximization problem is an optimization problem in two variables, q_0 and q_1. q_j affects the optimal solution for q_k, $k \neq j$, through its impact on the marginal costs of production ξ_k. To determine the impact of some characteristic x_j on the optimal output, we totally differentiate the system of first order conditions:

$$\frac{\partial \pi(\cdot)}{\partial q_0(\bar{z}_i)} = \frac{\sigma-1}{\sigma}q_0(\bar{z}_i)^{-\frac{1}{\sigma}}Q^{\frac{1}{\sigma}} - \xi_0(\cdot) = 0$$
$$\frac{\partial \pi(\cdot)}{\partial q_1(\bar{z}_i)} = \frac{\sigma-1}{\sigma}q_1(\bar{z}_i)^{-\frac{1}{\sigma}}Q_1^{\frac{1}{\sigma}}P_1^{\frac{\sigma-1}{\sigma}} - \xi_1(\cdot) = 0$$

Solving for $\frac{dq_0}{dx_j}$ and $\frac{dq_1}{dx_j}$ yields:

$$\frac{dq_0}{dx_j} = \frac{\frac{d\xi_0}{dx_j}\left[\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_1^{-\frac{1}{\sigma}-1}Q_1^{\frac{1}{\sigma}}P_1^{\frac{\sigma-1}{\sigma}} - \frac{d\xi_1}{dq_1}\right)\frac{1}{\frac{d\xi_0}{dq_1}}\right] + \frac{d\xi_1}{dx_j}}{\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_1^{-\frac{1}{\sigma}-1}Q_1^{\frac{1}{\sigma}}P_1^{\frac{\sigma-1}{\sigma}} - \frac{d\xi_1}{dq_1}\right)\frac{1}{\frac{d\xi_0}{dq_1}} - \frac{d\xi_1}{dq_0}}$$

$$\frac{dq_1}{dx_j} = \frac{\frac{d\xi_1}{dx_j}\left[\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}}\right] + \frac{d\xi_0}{dx_j}}{\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_1^{-\frac{1}{\sigma}-1}Q_1^{\frac{1}{\sigma}}P_1^{\frac{\sigma-1}{\sigma}} - \frac{d\xi_1}{dq_1}\right)\frac{1}{\frac{d\xi_1}{dq_0}} - \frac{d\xi_0}{dq_1}}$$

where

$$\frac{\partial \xi_0}{\partial q_j} = \frac{1}{1-e^{-\lambda \bar{z}}}(-w_0 c_0 + \theta_{00} e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)))\frac{\partial z_h}{\partial q_j} \quad \text{for } z_0 = \bar{z} - z_h$$
$$< 0 \text{ if } j = 0, \quad > 0 \text{ if } j = 1$$
$$\frac{\partial \xi_0}{\partial q_j} = \frac{1}{1-e^{-\lambda \bar{z}}} w_0 c_0 \frac{\partial z_0}{\partial q_j} \quad \text{for } z_0 > \bar{z} - z_h$$
$$< 0 \text{ if } j = 0, \quad > 0 \text{ if } j = 1$$
$$\frac{\partial \xi_1}{\partial q_j} = \frac{1}{1-e^{-\lambda \bar{z}}}(-w_1 c_1 + \theta_{10} e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)))\frac{\partial z_h}{\partial q_j} \quad \text{for } z_1 = \bar{z} - z_h$$
$$< 0 \text{ if } j = 1, \quad > 0 \text{ if } j = 0$$
$$\frac{\partial \xi_1}{\partial q_j} = \frac{1}{1-e^{-\lambda \bar{z}}} w_1 c_1 \frac{\partial z_1}{\partial q_j} \quad \text{for } z_1 > \bar{z} - z_h$$
$$< 0 \text{ if } j = 1, \quad > 0 \text{ if } j = 0$$

The denominators of these expressions are positive transformations of the determinant of the Hessian matrix, which is positive at a maximum. The sign of $\frac{dq_1}{d\theta_{10}}$ consequently depends on the numerator. Substituting yields:

$$\operatorname{sgn}\left(\frac{dq_1}{d\theta_{10}}\right) = -\operatorname{sgn}\left(\frac{d\xi_1}{d\theta_{10}}\right) \qquad < 0$$
$$\text{for } z_1 > \bar{z} - z_h$$
$$\text{by } \frac{d\xi_0}{d\theta_{10}} = 0 \text{ and } \left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma} q_0^{-\frac{1}{\sigma}-1} Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} < 0$$
$$\operatorname{sgn}\left(\frac{dq_1}{d\theta_{10}}\right) = \operatorname{sgn}\left(\frac{d\xi_1}{d\theta_{10}}\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma} q_0^{-\frac{1}{\sigma}-1} Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} + \frac{d\xi_0}{d\theta_{10}}\right) \qquad < 0$$
$$\text{for } z_0 > z_1 = \bar{z} - z_h, \text{ and } z_1 = z_0 = \bar{z} - z_h \text{ with } w_1 c_1 \theta_{00} > w_0 c_0 \theta_{10}$$
$$\text{by } \frac{d\xi_0}{d\theta_{10}} < 0 \text{ and } \left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma} q_0^{-\frac{1}{\sigma}-1} Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} < 0$$
$$\operatorname{sgn}\left(\frac{dq_1}{d\theta_{10}}\right) = \operatorname{sgn}\left(\frac{d\xi_1}{d\theta_{10}}\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma} q_0^{-\frac{1}{\sigma}-1} Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} + \frac{d\xi_0}{d\theta_{10}}\right) \qquad \gtreqless 0$$
$$\text{for } z_1 = z_0 = \bar{z} - z_h \text{ with } w_1 c_1 \theta_{00} < w_0 c_0 \theta_{10}$$
$$\text{by } \frac{d\xi_0}{d\theta_{10}} > 0 \text{ and } \left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma} q_0^{-\frac{1}{\sigma}-1} Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} < 0$$

Results on sales follow by sales increasing in the output.

C.1.2 Corollary 2

To show: The lower λ is, the stronger is the increase of the foreign marginal costs $\xi_1(\cdot)$ with the communication costs θ_{10} if $z_1 > \bar{z} - z_h$.

We build on Corollary 1. The result follows from:

$$\frac{\partial^2 \xi_1}{\partial \theta_{10} \partial \lambda} = -\frac{\lambda e^{-\lambda \bar{z}}}{(1-e^{-\lambda \bar{z}})^2} w_1 c_1 \frac{\partial z_1}{\partial \theta_{10}} + \frac{1}{1-e^{-\lambda \bar{z}}} w_1 c_1 \frac{\partial^2 z_1}{\partial \theta_{10} \partial \lambda} \qquad < 0$$

Numerical analysis. The numerical analysis studies the impact of λ on the increase of the marginal costs with the communication costs if $z_1 = \bar{z} - z_h$ as well as the implications for sales and the production quantity.

Figure C.1: Numerical analysis: impact of communication costs on foreign marginal costs and sales for different values of the predictability of the production process (partial equilibrium)

Figure C.1a: ξ_1, $\bar{z} = 1.5, c_1 = c_0$

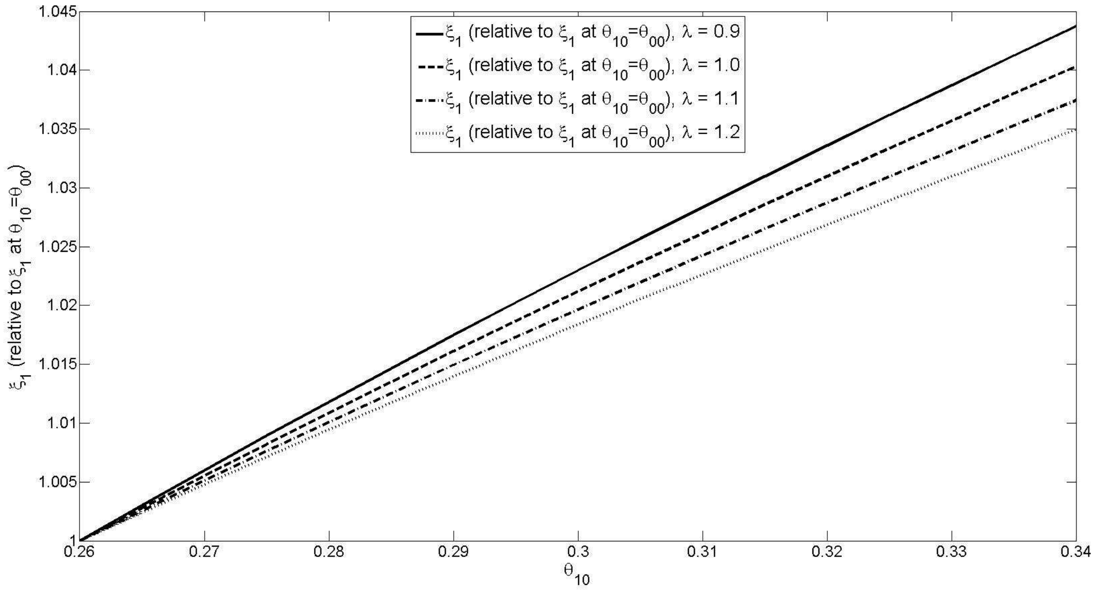
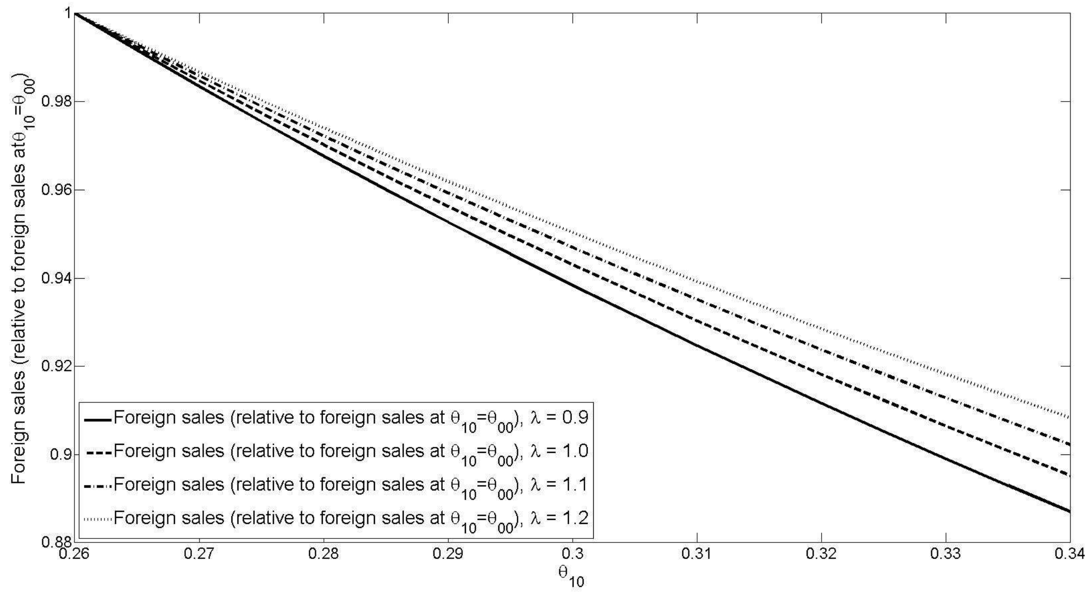


Figure C.1b: Sales, $\bar{z} = 1.5, c_1 = c_0$



The figure plots numerical analysis of the impact of the communication costs θ_{10} (x-axis) on foreign marginal costs ξ_1 (y-axis, top) and foreign sales (y-axis, bottom) for different values of the predictability of the production process λ. The foreign marginal costs and sales are plotted relative to their value if $\theta_{10} = \theta_{00}$. A higher λ describes a more predictable production process. The following parameter values are taken from Caliendo and Rossi-Hansberg (2012): $\sigma = 3.8$, $c_0 = 0.225$, $\theta_{00} = 0.26$. The other parameters are set to $w_1 = w_0 = 1$, $Q_1 = Q_0 = 100$, $P_1 = P_0 = 1$. The knowledge constraint $z_1 + z_h \geq \bar{z}$ is binding up to the following communication costs values: $\lambda = 0.9 : \theta_{10} \leq 0.375$; $\lambda = 1.0 : \theta_{10} \leq 0.365$; $\lambda = 1.1 : \theta_{10} \leq 0.355$; $\lambda = 1.2 : \theta_{10} \leq 0.345$.

Figure C.2: Numerical analysis: impact of communication costs on foreign marginal costs and sales for different values of the predictability of the production process (partial equilibrium)

Figure C.2a: ξ_1, $\bar{z} = 3, c_1 = c_0$

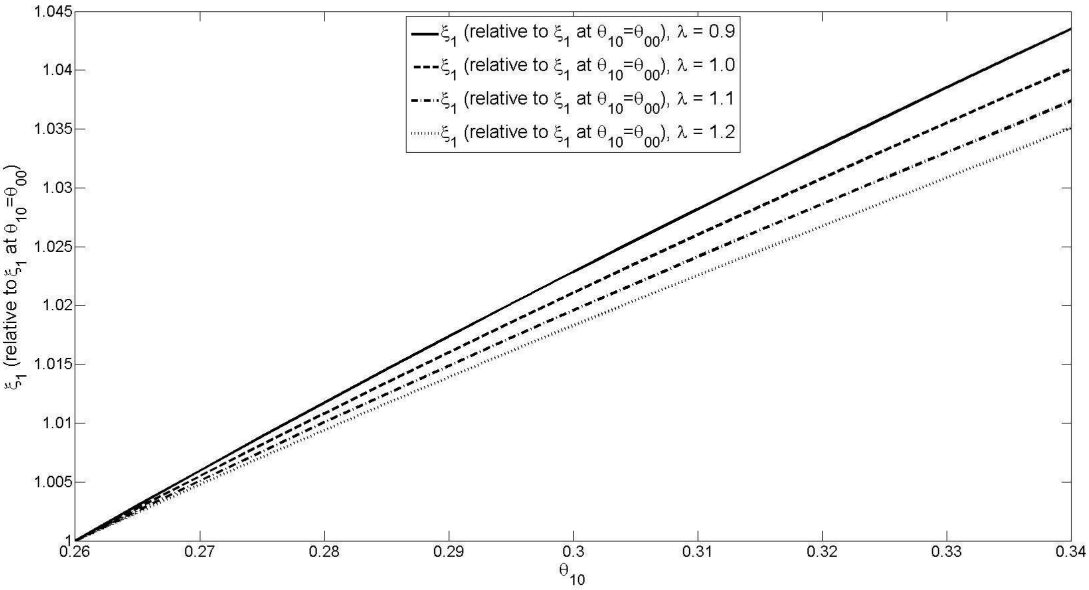
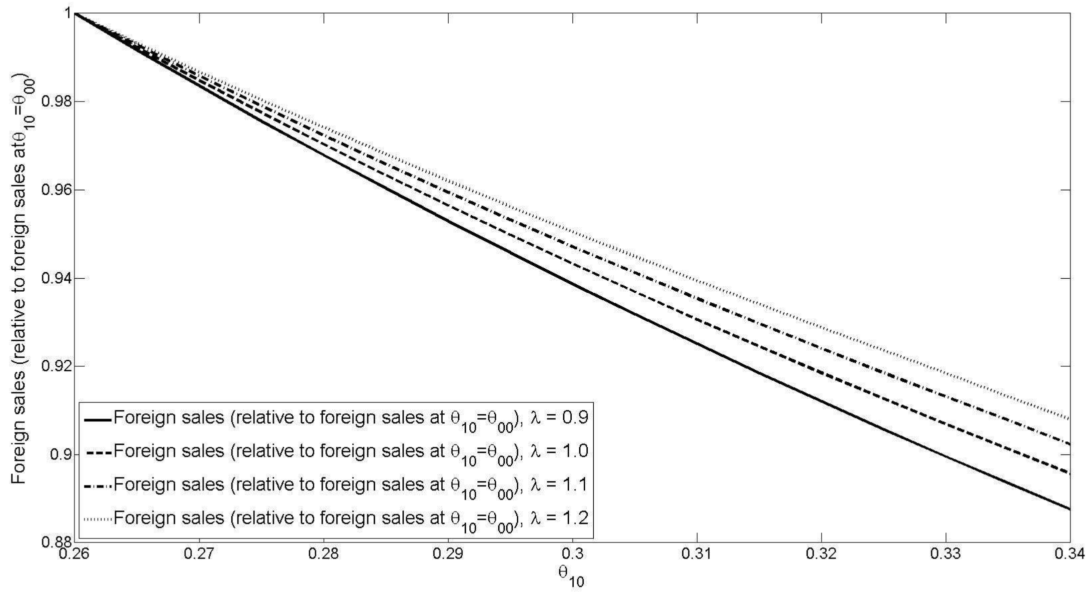


Figure C.2b: Sales, $\bar{z} = 3, c_1 = c_0$



The figure plots numerical analysis of the impact of the communication costs θ_{10} (x-axis) on foreign marginal costs ξ_1 (y-axis, top) and foreign sales (y-axis, bottom) for different values of the predictability of the production process λ. The foreign marginal costs and sales are plotted relative to their value if $\theta_{10} = \theta_{00}$. A higher λ describes a more predictable production process. The following parameter values are taken from Caliendo and Rossi-Hansberg (2012): $\sigma = 3.8$, $c_0 = 0.225$, $\theta_{00} = 0.26$. The other parameters are set to $w_1 = w_0 = 1$, $Q_1 = Q_0 = 100$, $P_1 = P_0 = 1$. The knowledge constraint $z_1 + z_h \geq \bar{z}$ is binding up to the following communication costs values: $\lambda = 0.9 : \theta_{10} \leq 0.35$; $\lambda = 1.0 : \theta_{10} \leq 0.34$; $\lambda = 1.1 : \theta_{10} \leq 0.335$; $\lambda = 1.2 : \theta_{10} \leq 0.325$.

Figure C.3: Numerical analysis: impact of communication costs on foreign marginal costs and sales for different values of the predictability of the production process (partial equilibrium)

Figure C.3a: ξ_1, $\bar{z} = 1.5, c_1 = 0.85c_0$

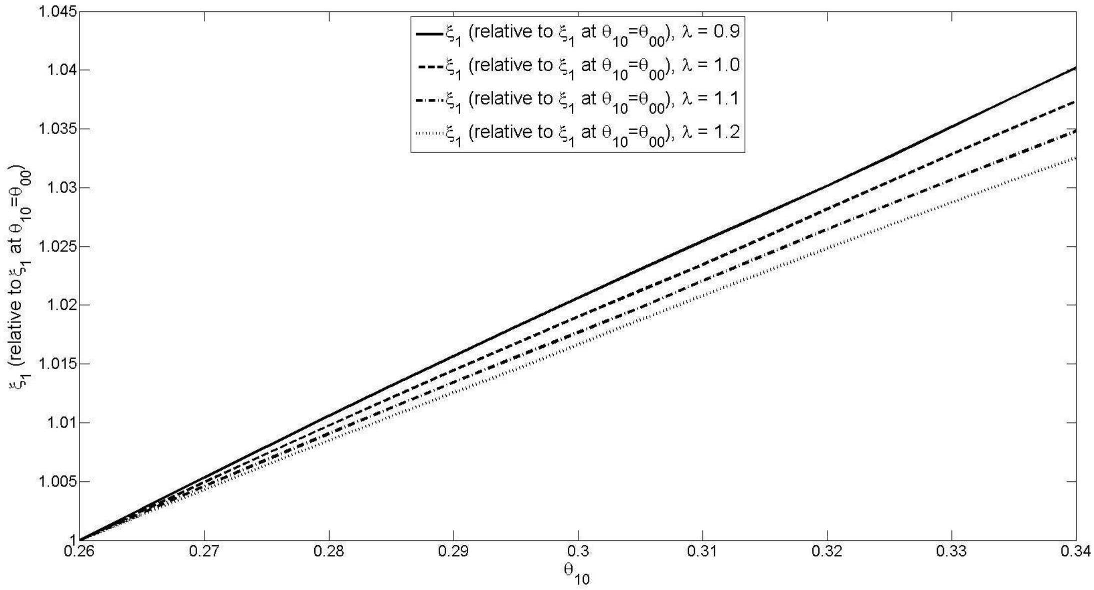
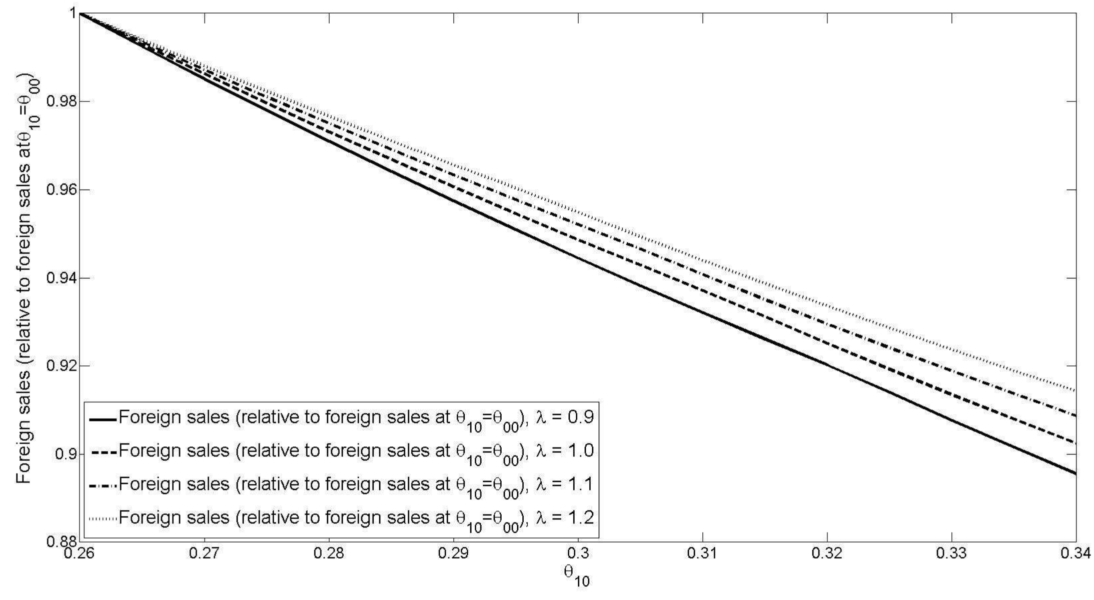


Figure C.3b: Sales, $\bar{z} = 1.5, c_1 = 0.85c_0$



The figure plots numerical analysis of the impact of the communication costs θ_{10} (x-axis) on foreign marginal costs ξ_1 (y-axis, top) and foreign sales (y-axis, bottom) for different values of the predictability of the production process λ. The foreign marginal costs and sales are plotted relative to their value if $\theta_{10} = \theta_{00}$. A higher λ describes a more predictable production process. The following parameter values are taken from Caliendo and Rossi-Hansberg (2012): $\sigma = 3.8$, $c_0 = 0.225$, $\theta_{00} = 0.26$. The other parameters are set to $w_1 = w_0 = 1$, $Q_1 = Q_0 = 100$, $P_1 = P_0 = 1$. The knowledge constraint $z_1 + z_h \geq \bar{z}$ is binding up to the following communication costs values: $\lambda = 0.9 : \theta_{10} \leq 0.32$; $\lambda = 1.0 : \theta_{10} \leq 0.31$; $\lambda = 1.1 : \theta_{10} \leq 0.30$; $\lambda = 1.2 : \theta_{10} \leq 0.295$.

Figure C.4: Numerical analysis: impact of communication costs on foreign marginal costs and sales for different values of the predictability of the production process (partial equilibrium)

Figure C.4a: ξ_1, $\bar{z} = 3, c_1 = 0.85c_0$

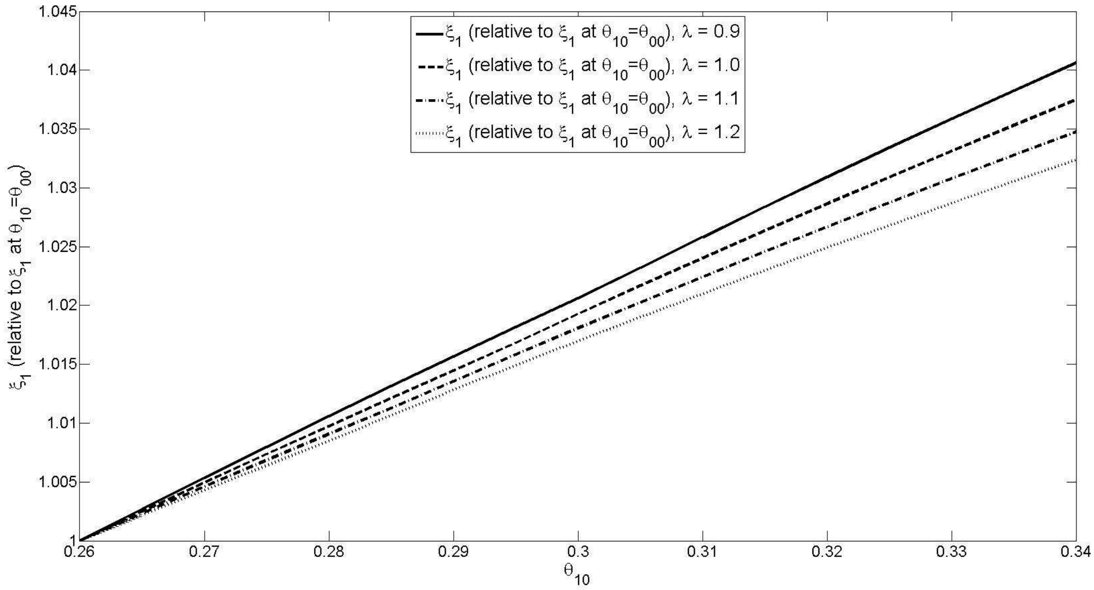
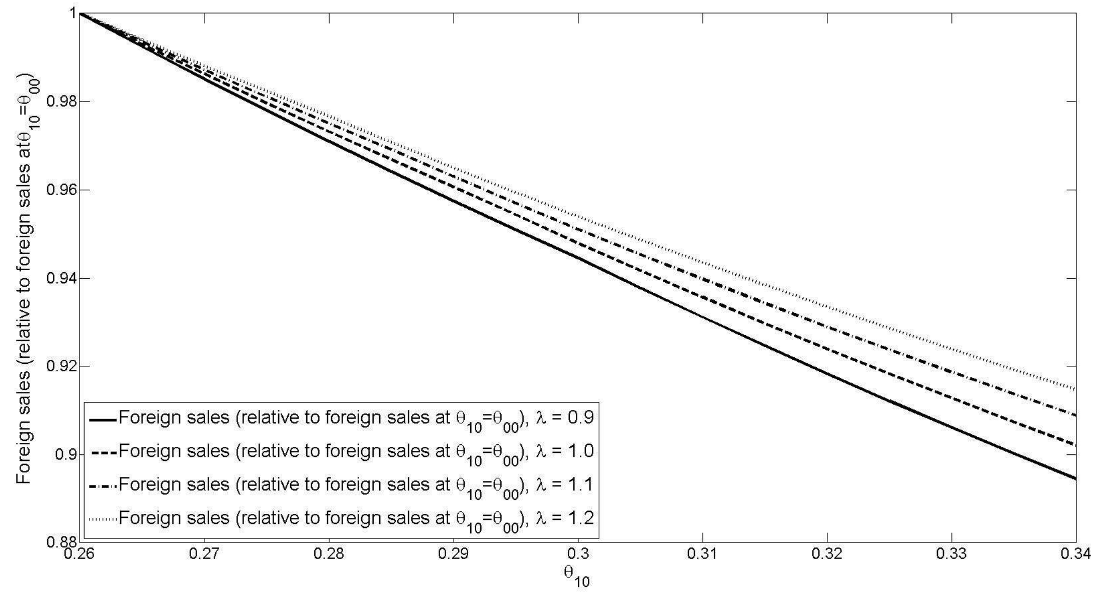


Figure C.4b: Sales, $\bar{z} = 3, c_1 = 0.85c_0$



The figure plots numerical analysis of the impact of the communication costs θ_{10} (x-axis) on foreign marginal costs ξ_1 (y-axis, top) and foreign sales (y-axis, bottom) for different values of the predictability of the production process λ. The foreign marginal costs and sales are plotted relative to their value if $\theta_{10} = \theta_{00}$. A higher λ describes a more predictable production process. The following parameter values are taken from Caliendo and Rossi-Hansberg (2012): $\sigma = 3.8$, $c_0 = 0.225$, $\theta_{00} = 0.26$. The other parameters are set to $w_1 = w_0 = 1$, $Q_1 = Q_0 = 100$, $P_1 = P_0 = 1$. The knowledge constraint $z_1 + z_h \geq \bar{z}$ is binding up to the following communication costs values: $\lambda = 0.9 : \theta_{10} \leq 0.30$; $\lambda = 1.0 : \theta_{10} \leq 0.29$; $\lambda = 1.1 : \theta_{10} \leq 0.285$; $\lambda = 1.2 : \theta_{10} \leq 0.28$.

C.1.3 Further results

To show: The impact of the learning costs c_1 and the wages w_1 on the marginal costs, production quantities and sales is generally ambiguous.

We build on the results from Proposition 1. There are two possible cases:

1. When $z_1 = \bar{z} - z_h$, so $\xi_1 = \frac{1}{1-e^{-\lambda\bar{z}}}\left(w_1(1+c_1(\bar{z}-z_h)) + w_0(1+c_0 z_h)\theta_{10}e^{-\lambda(\bar{z}-z_h)}\right)$:

$$\begin{aligned}
\frac{\partial \xi_1}{\partial c_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}\left(w_1(\bar{z}-z_h) + \frac{\partial z_h}{\partial c_1}(-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\right)\\
&> 0 \text{ if } z_1 = z_0 = \bar{z}-z_h \text{ and } w_1c_1 < \frac{\theta_{10}}{\theta_{00}}w_0c_0, \quad \gtrless 0 \text{ otherwise}\\
\frac{\partial \xi_1}{\partial w_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}\left(1+c_1(\bar{z}-z_h) + \frac{\partial z_h}{\partial w_1}(-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\right)\\
&> 0 \text{ if } z_1 = z_0 = \bar{z}-z_h \text{ and } w_1c_1 < \frac{\theta_{10}}{\theta_{00}}w_0c_0, \quad \gtrless 0 \text{ otherwise}
\end{aligned}$$

2. When $z_1 > \bar{z} - z_h$, so $\xi_1(\bar{z}, q_0, w_0, q_1, w_1) = \frac{1}{1-e^{-\lambda\bar{z}}}w_1\left(1 + c_1z_1 + \frac{1}{\lambda}c_1\right)$:

$$\begin{aligned}
\frac{\partial \xi_1}{\partial c_1} &= \frac{w_1}{1-e^{-\lambda\bar{z}}}\left(z_1 + c_1\frac{1}{\lambda}\frac{c_0}{1+c_0z_h}\frac{\partial z_h}{\partial c_1}\right) \quad \gtrless 0\\
\frac{\partial \xi_1}{\partial w_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}\left(1 + c_1z_1 + \frac{1}{\lambda}\frac{c_0}{1+c_0z_h}w_1c_1\frac{\partial z_h}{\partial w_1}\right) \quad \gtrless 0
\end{aligned}$$

We determine the impact of c_1, w_1 on the production quantities by substituting into the equations from section C.1.1:

$$\begin{aligned}
\operatorname{sgn}\left(\frac{dq_1}{dc_1}\right) &= \operatorname{sgn}\left(\frac{1}{1-e^{-\lambda\bar{z}}}w_1\left(z_1\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\right.\right.\\
&\quad \left.\left. - \; c_1\frac{c_0}{\lambda(1+c_0z_h)}\frac{dz_h}{dc_1}\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}}\right)\right) \gtrless 0 \quad \text{for } z_1 > \bar{z}-z_h\\
\operatorname{sgn}\left(\frac{dq_1}{dc_1}\right) &= \operatorname{sgn}\left(\frac{1}{1-e^{-\lambda\bar{z}}}\left(w_1(\bar{z}-z_h)\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\right.\right.\\
&\quad \left.\left. - \left(-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h))\right)\frac{dz_h}{dc_1}\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}}\right)\right)\\
&\quad \gtrless 0 \text{ for } z_1 = \bar{z}-z_h, z_0 > \bar{z}-z_h \text{ and } z_1 = z_0 = \bar{z}-z_h \text{ with } w_1c_1\theta_{00} > w_0c_0\theta_{10}\\
&\quad < 0 \text{ for } z_1 = z_0 = \bar{z}-z_h \text{ with } w_1c_1\theta_{00} < w_0c_0\theta_{10}
\end{aligned}$$

The effect of wages w_1 is analogous. Results on sales follow by sales increasing in the output.

C.2 General equilibrium

Existence. I follow Caliendo and Rossi-Hansberg (2012) to show that a unique equilibrium exists.

Zero cut-off profit condition: The zero cut-off profit condition starts at the point $(0,0)$ and is strictly increasing in the $\bar{z}^*, w$-plane by:

$$\frac{dw}{d\bar{z}^*} = -\frac{\frac{d\xi_0}{d\bar{z}}}{\frac{d\xi_0}{dw}} > 0$$

Free-entry condition: The free entry condition starts at the point $(0, \hat{w})$, where $\hat{w} > 0$. Its slope is given by:

$$\frac{dw}{d\bar{z}^*} = (*)^{-1}\left(\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} N\xi(\bar{z}^*, w)^{1-\sigma} - (1+f^D)\right) g(\bar{z})$$

The free entry condition is increasing up to the intersection with the zero cut-off profit condition and decreasing otherwise.

A unique intersection exists by the intermediate value theorem.

Comparative statics. To determine the equilibrium effects of transport costs and communication costs on the export and FDI cut-offs, I totally differentiate the equilibrium conditions (20), (21), (22) and (23). This yields, with $\xi_{j,I} \equiv \xi_j(\bar{z}^I, q_0(\bar{z}^I), w, q_1(\bar{z}^I), w)$:

Wages.

$$\frac{dw}{d\tau} = (*)^{-1}\int_{\bar{z}^X}^{\bar{z}^I} \frac{1-\sigma}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} N\tau^{-\sigma}\xi_0(\bar{z}, w)^{1-\sigma} dG(\bar{z}) < 0$$

$$\frac{dw}{d\theta_{10}} = (*)^{-1}\int_{\bar{z}^I}^{\bar{z}_{max}} \frac{1-\sigma}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} N\left(\xi_{0,I}^{-\sigma}\frac{d\xi_0}{d\theta_{10}} + \xi_{1,I}^{-\sigma}\frac{d\xi_1}{d\theta_{10}}\right) dG(\bar{z}) < 0$$

where

$$(*) = -\int_{\bar{z}^*}^{\bar{z}^I} \frac{1-\sigma}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \frac{N}{w}\xi_0(\bar{z}, w)^{1-\sigma} dG(\bar{z}) - \int_{\bar{z}^X}^{\bar{z}^I} \frac{1-\sigma}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \frac{N}{w}(\tau\xi_0(\bar{z}, w))^{1-\sigma} dG(\bar{z})$$
$$- \int_{\bar{z}^I}^{\bar{z}_{max}} \frac{1-\sigma}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \frac{N}{w}\left(\xi_{0,I}^{1-\sigma} + \xi_{1,I}^{1-\sigma}\right) dG(\bar{z}) < 0$$

Cut-off knowledge level for activity $\bar{z}^*$.

$$\frac{d\bar{z}^*}{d\tau} = -\frac{dw}{d\tau}\frac{d\xi_0(\bar{z}^*, w)/dw}{d\xi_0(\bar{z}^*, w)/d\bar{z}^*} < 0$$

$$\frac{d\bar{z}^*}{d\theta_{10}} = -\frac{dw}{d\theta_{10}}\frac{d\xi_0(\bar{z}^*, w)/dw}{d\xi_0(\bar{z}^*, w)/d\bar{z}^*} < 0$$

Cut-off knowledge level for exporting $\bar{z}^X$.

$$\frac{d\bar{z}^X}{d\tau} = -\frac{\xi_0(\bar{z}^X, w) + \tau\frac{dw}{d\tau}\frac{d\xi_0(\bar{z}^X, w)}{dw}}{\tau d\xi_0(\bar{z}^X, w)/d\bar{z}^X} > 0 \quad \text{by} \quad \frac{d\xi_0(\bar{z}^X, w)}{dw} = \frac{\xi_0(\bar{z}^X, w)}{w} \text{ and } \frac{dw}{d\tau} > -\frac{w}{\tau}$$

$$\frac{d\bar{z}^X}{d\theta_{10}} = -\frac{dw}{d\theta_{10}}\frac{d\xi_0(\bar{z}^X, w)/dw}{d\xi_0(\bar{z}^X, w)/d\bar{z}^X} < 0$$

Cut-off knowledge level for FDI $\bar{z}^I$.

$$\frac{d\bar{z}^I}{d\tau} = \frac{\tau^{-\sigma}\xi_0(\bar{z}^I, w)^{1-\sigma} - \frac{dw}{d\tau}\left(\xi_{0,I}^{-\sigma}\frac{d\xi_{0,I}}{dw} + \xi_{1,I}^{-\sigma}\frac{d\xi_{1,I}}{dw} - (1+\tau^{1-\sigma})\xi_0(\bar{z}^I, w)^{-\sigma}\frac{d\xi_0(\bar{z}^I, w)}{dw}\right)}{\xi_{0,I}^{-\sigma}\frac{d\xi_{0,I}}{d\bar{z}^I} + \xi_{1,I}^{-\sigma}\frac{d\xi_{1,I}}{d\bar{z}^I} - (1+\tau^{1-\sigma})\xi_0(\bar{z}^I, w)^{-\sigma}\frac{d\xi_0(\bar{z}^I, w)}{d\bar{z}^I}} < 0$$

by $\xi_{0,I}^{-\sigma}\frac{d\xi_{0,I}}{d\bar{z}^I}+\xi_{1,I}^{-\sigma}\frac{d\xi_{1,I}}{d\bar{z}^I}-(1+\tau^{1-\sigma})\xi_0(\bar{z}^I,w)^{-\sigma}\frac{d\xi_0(\bar{z}^I,w)}{d\bar{z}^I}<0$ and

$$\xi_{0,I}^{-\sigma}\frac{d\xi_{0,I}}{dw}+\xi_{1,I}^{-\sigma}\frac{d\xi_{1,I}}{dw}-(1+\tau^{1-\sigma})\xi_0(\bar{z}^I,w)^{-\sigma}\frac{d\xi_0(\bar{z}^I,w)}{dw}=(f^I-f^X)\frac{\sigma}{Nw}\left(\frac{\sigma-1}{\sigma}\right)^{1-\sigma}>0$$

$$\frac{d\bar{z}^I}{d\theta_{10}}=\frac{-\xi_{0,I}^{-\sigma}\frac{d\xi_0}{d\theta_{10}}-\xi_{1,I}^{-\sigma}\frac{d\xi_1}{d\theta_{10}}-\frac{dw}{d\theta_{10}}\left(\xi_{0,I}^{-\sigma}\frac{d\xi_{0,I}}{dw}+\xi_{1,I}^{-\sigma}\frac{d\xi_{1,I}}{dw}-(1+\tau^{1-\sigma})\xi_0(\bar{z}^I,w)^{-\sigma}\frac{d\xi_0(\bar{z}^I,w)}{dw}\right)}{\xi_{0,I}^{-\sigma}\frac{d\xi_{0,I}}{d\bar{z}^I}+\xi_{1,I}^{-\sigma}\frac{d\xi_{1,I}}{d\bar{z}^I}-(1+\tau^{1-\sigma})\xi_0(\bar{z}^I,w)^{-\sigma}\frac{d\xi_0(\bar{z}^I,w)}{d\bar{z}^I}}$$

It is difficult to determine the sign of $\frac{d\bar{z}^I}{d\theta_{10}}$ analytically. Suppose $\frac{d\bar{z}^I}{d\theta_{10}}<0$. Then, $\frac{d\bar{z}^I}{d\theta_{10}}<0$, $\frac{d\bar{z}^X}{d\theta_{10}}<0$ and $\frac{d\bar{z}^*}{d\theta_{10}}<0$, i.e., the expected profits at entry increase. The sunk costs of entry are constant. Therefore, the wages increase. This is a contradiction to $\frac{dw}{d\theta_{10}}<0$.

Aggregate implications.

- As the export cut-off $\bar{z}^X$ is increasing and the FDI cut-off $\bar{z}^I$ is decreasing in the transport costs, and the export sales are decreasing in the transport costs, aggregate exports decrease in the transport costs. Aggregate affiliate sales increase in the transport costs because the FDI cut-off is decreasing and wages decrease.
- As the FDI cut-off $\bar{z}^I$ is increasing in the communication costs, and the foreign sales are decreasing in the communication costs, aggregate MNE foreign sales decrease in the communication costs. Aggregate exports increase in the communication costs because the export cut-off is decreasing, the FDI cut-off is increasing and wages decrease.

C.3 Residual MNE wage premiums: Proposition 4

As shown in Proposition 2 and its proof in section B.3, a firm with given level of knowledge $\bar{z}$ chooses a lower level of managerial knowledge when it is a multinational than when it is not if $\theta_{00}w_1c_1<\theta_{10}w_0c_0$.

In consequence, if an MNE and a non-MNE with the same $\bar{z}$ coexisted, the MNE would pay higher remuneration to its production workers. As established in subsection 3.2, MNEs and non-MNEs with the same total level of knowledge do not coexist in equilibrium. Instead, MNEs have higher knowledge levels than non-MNEs. This reinforces the difference in workers' remuneration, because workers' knowledge and thus their remuneration increases in the total level of knowledge of the firm.

It is possible to find an MNE and a non-MNE with the same marginal costs of production even though MNEs have higher knowledge levels because firms reorganize when they become multinational. This drives up their domestic marginal costs of production. As the marginal costs of production decrease in the total level of knowledge, an MNE has a higher level of knowledge than a non-MNE with the same marginal costs of production, and thus, by the arguments above, pays higher remuneration to its production workers. A similar argument holds for the remuneration at MNEs' foreign plants compared to non-MNEs in the foreign country.

D Extension: Vertical FDI

This section shows that the predictions of the paper on the geography of MNEs' sales and the MNE wage premiums hold for "vertical MNEs", i.e. MNEs with domestic headquarters but only a foreign production plant.

D.1 The optimal organization of knowledge

D.1.1 First-order conditions

z_1, n_1 and n_h are given by the constraints (2)-(4) with $n_0=0$. Managerial knowledge is implicitly defined by

$$\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h))-w_1c_1=0. \tag{D.1}$$

D.1.2 Comparative statics

Proposition D.1. *The optimal knowledge levels vary with the foreign country characteristics θ_{10}, c_1, w_1, the production quantity q_1, the total knowledge $\bar{z}$, and the predictability of the production process λ as follows:*

Table D.1: Comparative statics

Knowledge levels/ model parameters	θ_{10}	c_1	w_1	q_1	$\bar{z}$	λ
Workers' knowledge z_1	+	-	-	0	+	+/-
Managerial knowledge z_h	-	+	+	0	+	+/-
# managers n_h	+	+	+	+	-	-
# workers n_1	0	0	0	+	-	-

The table displays the effects of the parameters on the endogenous variables for vertical MNEs (+ positive, − negative, 0 none).

Proof. By the implicit function theorem, $\frac{dz_h}{dx_1} = -\frac{\frac{d(D.1)}{dx_1}}{\frac{d(D.1)}{dz_h}}$.
The sign of $\frac{dz_h}{dx_1}$ is given by $-\frac{d(D.1)}{dx_1}$ because $\frac{d(D.1)}{dz_h} = \lambda\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(2c_0 + \lambda(1 + c_0 z_h)) > 0$.

$$\begin{aligned}
\frac{d(D.1)}{d\theta_{10}} &= e^{-\lambda(\bar{z}-z_h)}w_0(c_0 + \lambda(1 + c_0 z_h)) > 0 && \Rightarrow \frac{dz_h}{d\theta_{10}} < 0\\
\frac{d(D.1)}{dc_1} &= -w_1 < 0 && \Rightarrow \frac{dz_h}{dc_1} > 0\\
\frac{d(D.1)}{dw_1} &= -c_1 < 0 && \Rightarrow \frac{dz_h}{dw_1} > 0\\
\frac{d(D.1)}{dq_1} &= 0 && \Rightarrow \frac{dz_h}{dq_1} = 0\\
\frac{d(D.1)}{d\bar{z}} &= -\lambda\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0 + \lambda(1 + c_0 z_h)) < 0 && \Rightarrow \frac{dz_h}{d\bar{z}} > 0\\
\frac{d(D.1)}{d\lambda} &= \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0((1 + c_0 z_h) - (\bar{z} - z_h)(c_0 + \lambda(1 + c_0 z_h))) && \Rightarrow \frac{dz_h}{d\lambda} \lesseqgtr 0
\end{aligned}$$

$\frac{dz_1}{d\theta_{10}} > 0$, $\frac{dz_1}{dc_1} < 0$, $\frac{dz_1}{dw_1} < 0$, $\frac{dz_1}{dq_1} = 0$, $\frac{dz_1}{d\lambda} \lesseqgtr 0$ by $z_1 = \bar{z} - z_h$. $\frac{dz_1}{d\bar{z}} = 1 - \frac{dz_h}{d\bar{z}} > 0$ by $\frac{dz_h}{d\bar{z}} < 1$.

Number of managers. $\frac{\partial n_h}{\partial\theta_{10}} > 0$ by $\frac{\partial z_1}{\partial\theta_{10}} < \frac{1}{\lambda\theta_{10}}$; $\frac{\partial n_h}{\partial c_1} > 0$ by $\frac{\partial z_1}{\partial c_1} < 0$; $\frac{\partial n_h}{\partial w_1} > 0$ by $\frac{\partial z_1}{\partial w_1} < 0$; $\frac{\partial n_h}{\partial q_1} > 0$ by $\frac{\partial z_1}{\partial q_1} = 0$; $\frac{\partial n_h}{\partial\bar{z}} < 0$ by $\frac{\partial z_1}{\partial\bar{z}} > 0$; $\frac{\partial n_h}{\partial\lambda} < 0$ because ambiguous terms cancel. □

D.1.3 Managerial knowledge in MNEs and non-MNEs

Proposition D.2. *Given $\bar{z}$ and $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$, a firm systematically chooses a lower level of managerial knowledge if it is a vertical MNE than if it is a domestic firm.*

Proof. $w_1c_1 < w_0c_0$, otherwise, vertical FDI is not worthwhile, so $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$. Take a domestic firm and a vertical MNE with the same knowledge level $\bar{z}$. Domestic firms determine z_h^D via $\theta_{00}e^{-\lambda(\bar{z}-z_h^D)}w_0(c_0 + \lambda(1 + c_0 z_h^D) - w_0c_0 = 0$. Vertical MNEs determine z_h^V via $\theta_{10}e^{-\lambda(\bar{z}-z_h^V)}w_0(c_0 + \lambda(1 + c_0 z_h^V)) - w_1c_1 = 0$. The equations are increasing in z_h. As $\theta_{10} > \theta_{00}$ and $w_0c_0 > w_1c_1$, $z_h^V < z_h^D$.

For the MNE wage premiums, the comparison with foreign domestic firms is also relevant. Take a foreign domestic firm and a vertical MNE with the same knowledge level $\bar{z}$. Foreign domestic firms determine z_h^D via $\theta_{11}e^{-\lambda(\bar{z}-z_h^D)}w_1(c_1 + \lambda(1 + c_1 z_h^D) - w_1c_1 = 0$. Vertical MNEs determine z_h^V via $\theta_{10}e^{-\lambda(\bar{z}-z_h^V)}w_0(c_0 + \lambda(1 + c_0 z_h^V)) - w_1c_1 = 0$. The equations are increasing in z_h. As $\theta_{10} > \theta_{11}$ and $w_0c_0 > w_1c_1$, $z_h^V < z_h^D$. □

D.2 Profit maximization

Optimization problem:

$$\max_{q_0^V, q_1^V \geq 0} \quad \pi^V(\bar{z}_i, w_0, w_1) = \sum_{j=0}^{1} p_j(q_j^V(\bar{z}_i))q_j^V(\bar{z}_i) - C(\bar{z}_i, w_0, \tau q_0^V(\bar{z}_i) + q_1^V(\bar{z}_i), w_1) \tag{D.2}$$

Optimal quantities:

$$q_0^V(\bar{z}_i) = Q_0 \left(\frac{\sigma}{\sigma-1}\tau\xi_1(\bar{z}_i, w_0, w_1)\right)^{-\sigma} \qquad q_1^V(\bar{z}_i) = Q_1 P_1^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\xi_1(\bar{z}_i, w_0, w_1)\right)^{-\sigma} \tag{D.3}$$

Foreign quantities are higher if the firm conducts vertical FDI than if it conducts horizontal FDI:

$$q_1^V(\bar{z}_i) \geq q_1^I(\bar{z}_i). \tag{D.4}$$

Consequently, domestic quantities have to be higher if the firm conducts horizontal FDI than if it conducts vertical FDI:

$$q_0^I(\bar{z}_i) > q_0^V(\bar{z}_i). \tag{D.5}$$

Proposition D.3. *The foreign marginal costs $\xi_1(\bar{z}_i, w_0, w_1)$ of vertical MNEs increase with the communication costs θ_{10}, the foreign wages w_1, and the foreign learning costs c_1. Consequently, foreign production quantities and sales are higher in countries with lower communication costs θ_{10}, lower wages w_1 and lower learning costs c_1.*

Proof. Foreign marginal costs of production $\xi_1(\bar{z}, w_0, w_1)$, where
$\xi_1(\bar{z}, w_0, w_1) = \frac{1}{1-e^{-\lambda\bar{z}}}\left(w_1(1 + c_1(\bar{z} - z_h)) + w_0(1 + c_0 z_h)\theta_{10}e^{-\lambda(\bar{z}-z_h)}\right)$

$$\begin{aligned}
\frac{\partial\xi_1}{\partial\theta_{10}} &= \frac{1}{1-e^{-\lambda\bar{z}}}w_0(1+c_0 z_h)e^{-\lambda(\bar{z}-z_h)} \quad &> 0\\
\frac{\partial\xi_1}{\partial c_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}w_1(\bar{z}-z_h) \quad &> 0\\
\frac{\partial\xi_1}{\partial w_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}(1+c_1(\bar{z}-z_h)) \quad &> 0
\end{aligned}$$

Foreign output q_1 and sales $p_1 q_1$, where $x_1 \in \{\theta_{10}, c_1, w_1\}$:

$$\begin{aligned}
\frac{\partial\pi(\cdot)}{\partial q_1(\bar{z}_i)} &= \frac{\sigma-1}{\sigma}q_1(\bar{z}_i)^{-\frac{1}{\sigma}}Q_1^{\frac{1}{\sigma}}P_1^{\frac{\sigma-1}{\sigma}} - \xi_1(\bar{z}, w_0, w_1) = 0\\
\Rightarrow \quad \frac{\partial q_1(\bar{z}_i)}{\partial x_1} &= -\frac{-\frac{\partial\xi_1}{\partial x_1}}{-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_1(\bar{z}_i)^{-\frac{1}{\sigma}-1}Q_1^{\frac{1}{\sigma}}P_1^{\frac{\sigma-1}{\sigma}}}\\
\Rightarrow \quad \operatorname{sgn}\left(\frac{\partial q_1(\bar{z}_i)}{\partial x_1}\right) &= -\operatorname{sgn}\left(\frac{\xi_1}{x_1}\right)
\end{aligned}$$

Results on sales follow by sales increasing in the output. □

Investment decision: Firms compare the profits of domestic activity, exporting, horizontal FDI and vertical FDI. Only higher $\bar{z}$ firms select into vertical FDI if $\pi_0^V(\bar{z}^*, w_0, w_1) < (f^I + f^D)w_0$.

D.3 Residual MNE wage premiums

Proposition D.4. *Vertical MNEs pay higher remuneration to foreign workers than non-MNEs in the foreign country with the same marginal costs and local sales.*

Proof. Follows from Proposition D.2, together with the fact that production workers' knowledge is increasing in $\bar{z}$ and the marginal costs of production are decreasing in $\bar{z}$. □

E Foreign sales

E.1 Data cleaning

The Microdatabase Direct investment (MiDi) contains virtually the universe of German FDI because residents are legally obliged to report their investments to the central bank once their investments meet the reporting requirements. The reporting requirements vary across years. Until 2002, data on stakes of at least 10% in a firm with a balance sheet total of more than 5 million euro and stakes of at least 50% in a firm with a balance sheet total of more than 0.5 million euro had to be reported. Since 2002, data on stakes of at least 10% in a firm with a balance sheet total of more than 3 million euro has to be reported. The same data has to be provided on branches or permanent establishments if their operating assets exceed the reporting threshold (Lipponer, 2009).

I drop observations on 26,042 affiliates (7.9% of all observations) of investors that are government institutions, private households, agriculture or mining companies and housing enterprises. I drop agriculture and mining companies because natural resources are decisive for their investments, but ignored in the theory and the empirics. I drop housing enterprises because they often report sales of zero, even though they are not small, which would lead to measurement error in the analysis.

I restrict the data so that all observations meet a uniform threshold: I keep reports on affiliates with a balance sheet total of at least 5 million euro and a degree of participation of at least 10%, or with a balance sheet total between 3 and 5 million euro, but parent stakes of at least 50%. 36,754 observations drop from the sample (12.0% of the remaining observations).

Some affiliates are reported several times, because an investor has direct and indirect interests, or because multiple investors hold participating interests in them. I therefore aggregate direct and indirect participation shares per affiliate before restricting the sample to majority owned affiliates. 22,425 observations (8.3% of the remaining observations) drop from the sample because the affiliates are not majority owned. The resulting data set contains 246,394 affiliate–year observations.

E.2 Descriptive statistics

Table E.1: Summary statistics, section 4

	N	Mean	SD	p5	p25	p50	p75	p95
Log foreign sales	153,710	9.906	1.453	7.844	8.987	9.741	10.707	12.588
Log foreign employees	154,705	4.305	1.567	1.869	3.258	4.220	5.293	6.989
Office hours overlap	164,604	7.896	2.915	3.000	5.000	10.000	10.000	10.000
Log flight time from Frankfurt	164,192	5.154	0.959	4.174	4.317	4.654	6.292	6.600
Common spoken language	163,989	0.431	0.276	0.006	0.219	0.389	0.612	0.993
Linguistic proximity to German	162,650	-0.802	0.233	-1.000	-0.926	-0.926	-0.756	0.000
Log internet bandwidth	160,351	11.307	2.542	6.429	9.839	11.694	13.122	14.914
Log GDP	162,645	6.360	1.530	3.902	5.339	6.233	7.380	9.273
Log GDP per capita	162,638	9.832	1.038	7.658	9.276	10.168	10.568	10.870
Average years of schooling	161,865	9.873	1.944	6.475	8.939	9.787	11.165	13.086
Unit labor cost	112,901	0.643	0.075	0.521	0.598	0.654	0.690	0.748
Log distance	162,883	7.429	1.188	5.938	6.548	6.921	8.935	9.228
Log trade costs	97,781	4.023	0.518	3.299	3.677	3.995	4.354	4.812
∅ effectively applied tariffs	116,186	0.536	2.740	0.000	0.000	0.000	0.260	2.648
Statutory tax rate	162,954	28.937	7.674	15.000	25.000	30.000	35.000	39.000
Rule of law	139,511	1.024	0.809	-0.550	0.500	1.310	1.700	1.910
Regulatory quality	139,506	1.106	0.667	-0.290	0.810	1.250	1.620	1.850
Government efficiency	139,506	1.172	0.771	-0.230	0.570	1.490	1.800	2.060
Corruption	139,506	1.053	0.950	-0.590	0.270	1.320	1.960	2.240
Bilateral trust	119,979	2.549	0.420	1.744	2.307	2.729	2.856	3.091

The table provides summary statistics of the variables employed in the empirical analysis. *Variable definitions:* see Table 2. *pX,* $X \in \{5, 25, 50, 75, 95\}$*:* Xth percentile. Maximum possible number of observations: 164,604.

E.3 Predictability of the production process

Table E.2: List of sectors ordered by inverse predictability of the production process

10 sectors with lowest predictability			10 sectors with highest predictability		
	Sector	$\frac{1}{\lambda}$		Sector	$\frac{1}{\lambda}$
62	Air transport	.305	25	Man. of rubber & plastic products	.019
72	Computer & related activities	.298	21	Man. of pulp, paper & paper products	.014
80	Education	.262	52	Retail trade, except of motor vehicles; repair	.010
66	Insurance & pension funding	.197	26	Man. of other non-metallic mineral products	.000
20	Man. of wood & wood products	.196	36	Man. of furniture, manufacturing nec.	−.032
85	Health & social work	.178	90	Sewage & refusal disposal	−.049
29	Man. of machinery & equipment nec.	.152	15	Man. of food products & beverages	−.057
92	Recreational, cultural & sporting act.	.145	51	Wholesale & commission trade	−.086
73	Research & development	.132	17	Man. of textiles	−.093
64	Post & telecommunications	.116	93	Other service activities nec.	−.131

The table displays the 10 sectors with the highest and lowest predictability of the production process. The predictability is measured relative to agriculture, i.e. a value of .1 implies a 10% higher probability to encounter unforeseeable problems than in agriculture. The sectors are classified according to NACE Rev. 1.1.

Table E.3: Descriptive statistics, inverse predictability of the production process

(a) Summary statistics

	N	Mean	SD	CV	iqr
Inverse predictability of the production process $\frac{1}{\lambda}$	40	.078	.096	1.237	.088

The table provides summary statistics. *CV:* coefficient of variation; *iqr:* interquartile range.

(b) Relation with other sector characteristics

	$\frac{1}{\lambda}$	R&D int.	Skill int.	Capital int.
$\frac{1}{\lambda}$	1			
R&D intensity	0.275°	1		
Skill intensity	0.250°	0.513**	1	
Capital intensity	0.075	0.283°	0.276°	1

° $p < 0.20$, $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. The table displays the correlation coefficients of the inverse predictability of the production process and the R&D, skill and capital intensities of a sector. The P-value of the correlation with the R&D intensity is 14.9%, and the P-value of the correlation with the skill intensity is 12.4%. Data sources: R&D intensity, Center for European Economic Research; skill intensity (share of employees in science and technology), Eurostat; capital intensity, German Federal Statistical Office.

(c) Relation with sector characteristics constructed from other survey questions

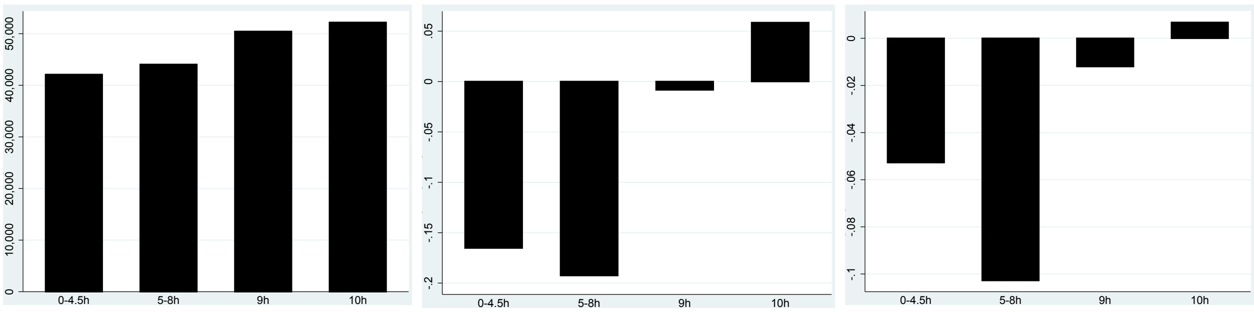
	$\frac{1}{\hat{\lambda}}$	Pr(d.)	Pr(n.t.)	Pr(c.m.)	Pr(f.w.)	Pr(m.r.)
$\frac{1}{\hat{\lambda}}$	1					
Pr (decision)	0.549***	1				
Pr (new task)	0.702***	0.470***	1			
Pr (consequential mistake)	0.106	−0.194	−0.067	1		
Pr (fixed workload)	0.084	−0.234°	0.024	0.441**	1	
Pr (minute rules)	−0.162	−0.205	−0.289+	0.542***	0.549***	1

° $p < 0.20$, + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The table displays the correlation coefficients of the inverse predictability of the production process in a sector and the probability that employees make tough decisions on their own responsibility [= Pr(decision)], confront new tasks [= Pr(new task)], risk to make mistakes that entail big financial losses [= Pr(consequential mistake)], facing fixed quantity or time requirements [= Pr(fixed workload)], or have to adhere to minute rules about the work process [= Pr(minute rules)]. The probabilities are constructed from regressions of the survey answers on sector dummy variables. Only sectors with at least 25 employees are included. Source: BIBB/BAuA Employment Survey 2006.

E.4 Supplementary material on graphical analysis

Figure E.1: Foreign sales vs. office hours overlap: Dropping the US

E.1a: Average foreign sales E.1b: Residuals | log GDP E.1c: R. | log GDP, distance



The figures plot, by quartiles of the office hours overlap, the average foreign sales of German MNEs, the residuals of a regression of log foreign sales on log GDP, and the residuals of a regression of log foreign sales on log GDP and log distance. The USA are excluded from the sample.

E.5 Supplementary material on regression analysis

Table E.4: Weights used to aggregate coefficients in Table 3

Weights for column	1	2	3	5	6	7	8	9
Office hours overlap	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Log flight time	−0.306	−0.322	−0.322	−0.234	−0.235	−0.237	−0.213	−0.303
Common spoken language	0.060	0.080	0.080	0.060	0.059	0.063	0.067	0.088
Linguistic proximity	0.055	0.079	0.079	0.107	0.105	0.104	0.103	0.183
Log internet bandwidth	0.255	0.254	0.254	0.105	0.106	0.118	0.088	0.177
Log distance			−0.387					
# observations	121,116	78,357	78,357	47,990	47,454	47,454	41,391	41,904

The table displays the covariances of the communication cost proxies and the dependent variable normalized by the covariance of the office hours overlap and the dependent variable used as weights to aggregate the coefficients of the proxy variables in Table 3 following Lubotsky and Wittenberg (2006).

Table E.5: Regression results: within-firm differences in log foreign sales across countries

Log foreign sales	1	2	3	4	5	6	7
Log GDP	0.270***	0.305***	0.295***	0.291***	0.281***	0.196***	0.135
	(0.024)	(0.026)	(0.026)	(0.022)	(0.023)	(0.031)	(0.235)
Log GDP per capita	0.274***	0.212***	0.202***	0.192***	0.229***	0.210***	0.479+
	(0.032)	(0.038)	(0.038)	(0.038)	(0.033)	(0.037)	(0.274)
Average years of schooling	0.002	0.017	0.013	−0.003	0.010	−0.000	0.019
	(0.017)	(0.018)	(0.019)	(0.017)	(0.017)	(0.017)	(0.030)
Office hours overlap		0.048***					
		(0.011)					
Log flight time			−0.144***				
			(0.032)				
Common spoken language				0.459***			
				(0.127)			
Linguistic proximity					0.387***		
					(0.094)		
Log internet bandwidth						0.095***	0.021+
						(0.027)	(0.011)
Country FE	N	N	N	N	N	N	Y
# observations	121,116	121,116	121,116	121,116	121,116	121,116	121,116
# MNEs	4,089	4,089	4,089	4,089	4,089	4,089	4,089
R-squared (within)	0.177	0.187	0.186	0.183	0.182	0.182	0.222

Two-way clustered standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign sales per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs.

Table E.6: Regression results: interaction of predictability and communication costs index

	$\hat{\lambda}$: baseline measure		$\hat{\lambda}$: 0 if insignificant	
Log foreign sales	1	2	3	4
Communication cost index	0.076***	0.139***	0.074***	0.138***
	(0.013)	(0.014)	(0.013)	(0.014)
Communication cost index ×	0.053+	0.014	0.087*	0.041
Probability of unexpected problems	(0.029)	(0.073)	(0.033)	(0.042)
# observations	116,172	75,333	116,172	75,333
# MNEs	3,972	3,199	3,972	3,199
# sectors	38	38	38	38
# countries	102	23	102	23

Two-way clustered standard errors in parentheses (sector, country). $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign sales per MNE, country, and year. *# MNEs:* number of MNEs. *# sectors:* number of parent sectors. *# countries:* number of countries. Columns 1 and 2 use the estimated probability of unexpected problems in all sectors. Columns 3 and 4 set the probability to zero if the coefficient estimate of the sector dummy is not significantly different from 0 at the 20% level. The communication cost index is computed by aggregating the variables office hours overlap, log flight time, common spoken language, linguistic proximity and log internet bandwidth using the covariance of these variables with the dependent variable as weights.

E.6 Robustness

Table E.7: Regression results: excluding adjustments at the extensive margin

Log foreign sales	1	2	3	4	5	6	7	8
Office hours overlap	0.047*	0.036	0.023	0.068+	0.071+	0.071+	0.082*	0.075+
	(0.022)	(0.033)	(0.025)	(0.035)	(0.035)	(0.035)	(0.033)	(0.037)
Log flight time	0.024	−0.123	−0.690**	−0.011	−0.012	−0.014	−0.007	−0.103
	(0.076)	(0.124)	(0.236)	(0.153)	(0.151)	(0.149)	(0.136)	(0.115)
Common spoken language	−0.165	−0.109	0.220	−0.292	−0.287	−0.275	−0.275	−0.385
	(0.177)	(0.186)	(0.221)	(0.222)	(0.218)	(0.224)	(0.296)	(0.237)
Linguistic proximity	0.422**	0.625***	0.499**	0.706***	0.709***	0.701***	0.665**	0.812**
	(0.142)	(0.144)	(0.158)	(0.178)	(0.178)	(0.172)	(0.196)	(0.207)
Log internet bandwidth	0.069**	0.042	0.045	0.044	0.041	0.041	−0.023	0.045
	(0.025)	(0.031)	(0.031)	(0.040)	(0.039)	(0.039)	(0.036)	(0.045)
Log GDP	0.275***	0.400***	0.367***	0.411***	0.415***	0.412***	0.473***	0.467***
	(0.038)	(0.043)	(0.036)	(0.053)	(0.052)	(0.056)	(0.050)	(0.060)
Log GDP per capita	0.097*	−0.088	−0.124	−0.133	−0.128	−0.134	−0.228*	−0.368*
	(0.046)	(0.087)	(0.079)	(0.101)	(0.098)	(0.116)	(0.096)	(0.135)
Average years of schooling	0.030	0.071**	0.065***	0.081**	0.080**	0.080**	0.082**	0.090**
	(0.020)	(0.019)	(0.017)	(0.022)	(0.022)	(0.022)	(0.023)	(0.031)
Unit labor cost		−0.764+	−0.793+	−0.499	−0.435	−0.434	−0.175	0.020
		(0.440)	(0.432)	(0.478)	(0.486)	(0.482)	(0.625)	(0.712)
Log distance			0.471*					
			(0.178)					
Log trade costs				−0.066	−0.063	−0.062	−0.068	−0.053
				(0.113)	(0.111)	(0.110)	(0.095)	(0.094)
Effectively applied tariffs					0.032*	0.032*	0.033*	0.053
					(0.015)	(0.015)	(0.015)	(0.036)
Statutory tax rate						0.001		
						(0.009)		
Regulatory quality							0.089	
							(0.171)	
Rule of law							−0.102	
							(0.146)	
Government efficiency							0.296*	
							(0.122)	
Corruption							−0.094	
							(0.130)	
Bilateral trust								0.206+
								(0.117)
Effect of communication costs	0.068***	0.127***	0.137***	0.132***	0.137***	0.135***	0.134***	0.212***
	(0.013)	(0.021)	(0.019)	(0.029)	(0.029)	(0.029)	(0.027)	(0.038)
# observations	50,786	34,538	34,538	21,656	21,453	21,453	18,469	19,002
# MNEs	1,789	1,371	1,371	840	837	837	808	775
R-squared (within)	0.201	0.193	0.195	0.218	0.218	0.218	0.210	0.220

The sample only includes firms that do not change their investment destinations over time. Two-way clustered standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign sales per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. The estimated effect of the communication costs is computed from the coefficients of the variables office hours overlap, log flight time, common spoken language, linguistic proximity, log internet bandwidth and, in column 3, log distance following Lubotsky and Wittenberg (2006).

Table E.8: Regression results: dropping Austria, Table 3

Log foreign sales	1	2	3	4	5	6	7	8	9
Office hours overlap	0.048^{*}	0.040	0.033		0.059°	0.059°	0.056°	0.068^{*}	0.072^{*}
	(0.021)	(0.035)	(0.026)		(0.038)	(0.037)	(0.037)	(0.032)	(0.032)
Log flight time	0.022	-0.131	-0.523^{+}		-0.080	-0.084	-0.094	-0.043	-0.168°
	(0.072)	(0.122)	(0.277)		(0.151)	(0.150)	(0.142)	(0.137)	(0.108)
Common spoken language	-0.056	0.015	0.235		-0.048	-0.052	0.012	0.111	-0.312°
	(0.175)	(0.174)	(0.243)		(0.195)	(0.196)	(0.217)	(0.301)	(0.193)
Linguistic proximity	0.292°	0.397	0.389°		0.152	0.147	0.050	-0.048	0.373°
	(0.227)	(0.356)	(0.296)		(0.326)	(0.322)	(0.345)	(0.296)	(0.271)
Log internet bandwidth	0.071^{**}	0.037°	0.037		0.032	0.033	0.033	0.005	0.041
	(0.022)	(0.027)	(0.030)		(0.033)	(0.033)	(0.033)	(0.031)	(0.035)
Log GDP	0.256^{***}	0.383^{***}	0.364^{***}	0.408^{***}	0.404^{***}	0.405^{***}	0.392^{***}	0.426^{***}	0.448^{***}
	(0.031)	(0.034)	(0.028)	(0.038)	(0.044)	(0.044)	(0.045)	(0.042)	(0.050)
Log GDP per capita	0.139^{***}	-0.053	-0.083	0.018	-0.095	-0.091	-0.110	-0.123°	-0.258^{*}
	(0.038)	(0.077)	(0.073)	(0.066)	(0.093)	(0.093)	(0.100)	(0.083)	(0.119)
Average years of schooling	0.026°	0.068^{***}	0.063^{***}	0.066^{**}	0.083^{***}	0.083^{***}	0.082^{***}	0.074^{**}	0.114^{***}
	(0.018)	(0.015)	(0.014)	(0.017)	(0.019)	(0.019)	(0.018)	(0.021)	(0.028)
Unit labor cost		-0.814^{+}	-0.852^{*}	-0.719°	-0.775°	-0.785°	-0.785^{+}	-0.550	-0.007
		(0.397)	(0.400)	(0.449)	(0.482)	(0.486)	(0.446)	(0.573)	(0.552)
Log distance			0.327°	-0.218^{***}					
			(0.214)	(0.041)					
Log trade costs					-0.019	-0.016	-0.010	-0.018	-0.007
					(0.095)	(0.094)	(0.092)	(0.076)	(0.070)
Effectively applied tariffs						0.008^{**}	0.008^{**}	0.006^{*}	0.008^{***}
						(0.002)	(0.002)	(0.002)	(0.002)
Statutory tax rate							0.006		
							(0.007)		
Regulatory quality								0.104	
								(0.138)	
Rule of law								-0.122	
								(0.117)	
Government efficiency								0.283^{*}	
								(0.108)	
Corruption								-0.163°	
								(0.104)	
Bilateral trust									0.093
									(0.104)
Estimated effect of communic. costs	0.062^{***}	0.096^{***}	0.100^{***}		0.082^{***}	0.084^{***}	0.084^{***}	0.080^{***}	0.132^{***}
	(0.011)	(0.017)	(0.015)		(0.019)	(0.019)	(0.019)	(0.018)	(0.022)
# observations	114,023	71,751	71,751	73,993	44,325	43,846	43,846	38,243	38,303
# MNEs	3,911	3,110	3,110	3,120	1,916	1,910	1,910	1,829	1,791
R-squared (within)	0.198	0.165	0.166	0.157	0.182	0.182	0.183	0.176	0.186

The sample excludes foreign investments in Austria. Two-way clustered standard errors in parentheses. $^{\circ}$ $p < 0.20$, $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign sales per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. The estimated effect of the communication costs is computed from the coefficients of the variables office hours overlap, log flight time, common spoken language, linguistic proximity, log internet bandwidth and, in column 3, log distance following Lubotsky and Wittenberg (2006).

Table E.9: Regression results: dropping Austria, Table 4

	$\hat{\lambda}$: baseline measure		$\hat{\lambda}$: 0 if insignificant	
Log foreign sales	1	2	3	4
Office hours overlap	0.036	0.022	0.033	0.026
	(0.022)	(0.034)	(0.021)	(0.034)
Office hours overlap ×	0.157	0.242	0.203^{+}	0.217^{+}
Probability of unexpected problems	(0.118)	(0.144)	(0.113)	(0.126)
Log flight time	−0.044	-0.217^{+}	−0.035	−0.191
	(0.065)	(0.114)	(0.065)	(0.119)
Log flight time ×	0.832^{*}	1.166^{*}	0.814^{*}	0.968^{*}
Probability of unexpected problems	(0.345)	(0.452)	(0.322)	(0.365)
Common spoken language	0.005	0.085	0.019	0.100
	(0.168)	(0.166)	(0.176)	(0.171)
Common spoken language ×	−1.049	−1.102	-1.332^{+}	-1.416^{+}
Probability of unexpected problems	(0.765)	(0.784)	(0.733)	(0.802)
Linguistic proximity	−0.019	0.058	0.038	0.114
	(0.220)	(0.343)	(0.229)	(0.365)
Linguistic proximity ×	4.462^{***}	4.272^{**}	4.217^{***}	4.138^{**}
Probability of unexpected problems	(1.026)	(1.328)	(0.980)	(1.265)
Log internet bandwidth	0.034	0.008	0.036	0.009
	(0.024)	(0.029)	(0.024)	(0.027)
Log internet bandwidth ×	0.476^{***}	0.408^{**}	0.506^{***}	0.428^{**}
Probability of unexpected problems	(0.119)	(0.137)	(0.118)	(0.130)
Log GDP	0.255^{***}	0.378^{***}	0.256^{***}	0.380^{***}
	(0.031)	(0.033)	(0.031)	(0.033)
Log GDP per capita	0.144^{***}	−0.038	0.144^{***}	−0.040
	(0.038)	(0.074)	(0.038)	(0.075)
Average years of schooling	0.024	0.064^{***}	0.024	0.065^{***}
	(0.018)	(0.015)	(0.018)	(0.015)
Unit labor cost		-0.850^{*}		-0.856^{*}
		(0.385)		(0.382)
# observations	109,418	69,043	109,418	69,043
# MNEs	3,797	3,030	3,797	3,030
# sectors	38	37	38	37
# countries	101	22	101	22
R-squared (within)	0.202	0.167	0.203	0.167

The sample excludes foreign investments in Austria. Two-way clustered standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign sales per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. *# sectors:* number of parent sectors. *# countries:* number of countries. Columns 1 and 2 use the observed estimates for the probability of unexpected problems in all sectors. Columns 3 and 4 set the probability to zero if the coefficient estimate of the sector dummy is not significantly different from 0 at the 20% level.

Table E.10: Regression results: within-firm differences in log foreign employees across countries

Log foreign sales	1	2	3	4	5	6	7	8	9
Office hours overlap	0.056*	0.057	0.033		0.050	0.049	0.042	0.121*	0.033
	(0.024)	(0.046)	(0.030)		(0.048)	(0.047)	(0.046)	(0.044)	(0.045)
Log flight time	0.027	−0.063	−1.172***		−0.243	−0.249	−0.272	−0.002	−0.430*
	(0.089)	(0.184)	(0.304)		(0.224)	(0.223)	(0.218)	(0.211)	(0.199)
Common spoken	0.070	0.136	0.745+		0.473	0.449	0.572	0.162	−0.227
language	(0.278)	(0.366)	(0.370)		(0.361)	(0.362)	(0.335)	(0.405)	(0.416)
Linguistic proximity	0.318	0.560*	0.330		0.447+	0.459+	0.384	0.536+	0.923**
	(0.209)	(0.259)	(0.247)		(0.245)	(0.246)	(0.248)	(0.275)	(0.308)
Log internet	0.073*	0.083	0.110+		0.023	0.027	0.021	−0.025	0.100+
bandwidth	(0.032)	(0.058)	(0.055)		(0.065)	(0.065)	(0.064)	(0.066)	(0.057)
Log GDP	0.241***	0.319***	0.238***	0.340***	0.385***	0.380***	0.341**	0.438***	0.353**
	(0.038)	(0.075)	(0.050)	(0.058)	(0.092)	(0.092)	(0.092)	(0.101)	(0.094)
Log GDP	−0.377***	−0.649***	−0.715***	−0.454***	−0.774***	−0.766***	−0.854***	−0.888***	−1.066***
per capita	(0.058)	(0.120)	(0.106)	(0.107)	(0.167)	(0.167)	(0.153)	(0.156)	(0.199)
Average years	0.052+	0.076*	0.062+	0.067+	0.115**	0.113**	0.115**	0.079*	0.180**
of schooling	(0.027)	(0.033)	(0.033)	(0.037)	(0.035)	(0.035)	(0.033)	(0.034)	(0.047)
Unit labor cost		−0.924	−1.424*	−0.531	−0.877	−0.891	−0.848	−1.149	1.272
		(0.712)	(0.650)	(0.961)	(0.768)	(0.770)	(0.660)	(0.785)	(1.039)
Log distance			0.914**	−0.212**					
			(0.264)	(0.067)					
Log trade costs					0.255*	0.261*	0.264*	0.155+	0.283**
					(0.103)	(0.099)	(0.095)	(0.086)	(0.079)
Effectively applied						0.007	0.007	0.010	0.002
tariffs						(0.011)	(0.011)	(0.012)	(0.012)
Statutory tax rate							0.018+		
							(0.010)		
Regulatory quality								0.341	
								(0.288)	
Rule of law								−0.039	
								(0.256)	
Government								0.538*	
efficiency								(0.201)	
Corruption								−0.320+	
								(0.175)	
Bilateral trust									0.196
									(0.206)
Estimated effect of	0.081***	0.154***	0.184***		0.196***	0.197***	0.204***	0.159***	0.347***
communic. costs	(0.018)	(0.032)	(0.028)		(0.041)	(0.041)	(0.039)	(0.034)	(0.059)
# observations	77,217	47,537	47,537	49,765	30,075	29,525	29,525	29,252	26,163
# MNEs	3,207	2,528	2,528	2,556	1,623	1,612	1,612	1,612	1,514
# country comb.	6,018	5,732	5,732	5,753	4,001	3,909	3,909	3,909	3,813
R-squared (within)	0.089	0.087	0.093	0.071	0.093	0.091	0.093	0.096	0.109

Sample restricted to 2004-2010. Two-way clustered standard errors in parentheses. $^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign employees per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. *# country comb.:* number of combinations of countries with MNE activity. The estimated effect of the communication costs is computed from the coefficients of the variables office hours overlap, log flight time, common spoken language, linguistic proximity, log internet bandwidth and, in column 3, log distance following Lubotsky and Wittenberg (2006).

Table E.11: Effect of communication costs with the predictability of the production process, log foreign employees

	$\hat{\lambda}$: baseline measure		$\hat{\lambda}$: 0 if insignificant	
Log foreign employees	1	2	3	4
Office hours overlap	0.040	0.024	0.041	0.031
	(0.027)	(0.052)	(0.026)	(0.051)
Office hours overlap ×	0.194	0.440^{+}	0.216	0.390^{+}
Probability of unexpected problems	(0.137)	(0.241)	(0.137)	(0.216)
Log flight time	−0.045	−0.198	−0.029	−0.158
	(0.092)	(0.199)	(0.093)	(0.198)
Log flight time ×	0.938^{*}	1.827^{*}	0.838^{*}	1.487^{*}
Probability of unexpected problems	(0.408)	(0.782)	(0.399)	(0.685)
Common spoken language	0.013	0.068	0.050	0.087
	(0.261)	(0.364)	(0.274)	(0.372)
Common spoken language ×	0.768	0.518	0.367	0.369
Probability of unexpected problems	(0.925)	(1.034)	(0.811)	(0.928)
Linguistic proximity	0.277	0.511^{+}	0.278	0.532^{+}
	(0.201)	(0.256)	(0.207)	(0.258)
Linguistic proximity ×	0.918	1.011	0.988	0.782
Probability of unexpected problems	(0.817)	(0.885)	(0.728)	(0.781)
Log internet bandwidth	0.020	0.041	0.022	0.046
	(0.034)	(0.060)	(0.034)	(0.059)
Log internet bandwidth ×	0.675^{***}	0.577^{**}	0.725^{***}	0.585^{**}
Probability of unexpected problems	(0.161)	(0.202)	(0.148)	(0.174)
Log GDP	0.242^{***}	0.312^{***}	0.243^{***}	0.313^{***}
	(0.039)	(0.075)	(0.039)	(0.075)
Log GDP per capita	-0.372^{***}	-0.626^{***}	-0.373^{***}	-0.628^{***}
	(0.057)	(0.119)	(0.057)	(0.119)
Average years of schooling	0.047^{+}	0.073^{*}	0.047^{+}	0.073^{*}
	(0.026)	(0.031)	(0.026)	(0.032)
Unit labor cost		−0.970		−0.972
		(0.670)		(0.672)
# observations	73,206	45,154	73,206	45,154
# MNEs	3,101	2,452	3,101	2,452
# sectors	38	38	38	38
# countries	99	23	99	23
R-squared	0.093	0.088	0.093	0.087

Sample restricted to 2004-2010. Two-way clustered standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year fixed effects included. *Dependent variable:* log foreign employees per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. *# sectors:* number of parent sectors. *# countries:* number of countries. Columns 1 and 2 use the observed estimates for the probability of unexpected problems in all sectors. Columns 3 and 4 set the probability to zero if the coefficient estimate of the sector dummy is not significantly different from 0 at the 20% level.

F Corporate transferees

F.1 Prediction

Proposition 1 shows that the knowledge level of foreign production workers z_1 increases with the communication costs between the home country and the foreign country θ_{10}. In addition, z_1 depends on the production quantity q_1. As Proposition 1 shows, z_1 generally decreases with q_1. Proposition 3 shows that the optimal production quantity q_1 generally decreases in the communication costs θ_{10}. This implies that the indirect effect of the communication costs θ_{10} on the knowledge level of foreign production workers z_1 through the production quantity q_1 goes in the same direction as the direct effect: Higher communication costs increase the knowledge level z_1.

It is possible to analytically prove that the indirect effect of communication costs on the production workers' knowledge level through the production quantity reinforces the direct effect, except if $w_1 c_1 \theta_{00} < w_0 c_0 \theta_{10}$ and the firm chooses symmetric knowledge levels. In this case, the analytical results are ambiguous. The indirect effect of the communication costs through the production quantity may work against their direct effect. However, overturning the result presupposes that the indirect effect of communication costs through the production quantity is stronger than their direct effect. This is unlikely, as the direct positive effect of the communication costs on the foreign production workers' knowledge level and thus the marginal costs of production drives the indirect effect through the production quantity. In addition, only firms with a lower level of knowledge $\bar{z}$ choose symmetric knowledge levels, so any qualification is unlikely to matter in the aggregate.

Subsection 3.2 shows that higher communication costs θ_{10} increase the cut-off knowledge level $\bar{z}^I$ for investing in a country. This reinforces the positive effect of communication costs on the knowledge level of production workers z_1, as by Proposition 1, z_1 increases in the total level of knowledge.

F.2 Data

Table F.1: Available information on corporate transferees and MNE employment

Source/Host	AU	BH	BE	BR	CA	CN	FR	DE	HK	IN	IT	JP	KW	NL	OM	PL	PT	QA	RU	SA	SG	ZA	KR	ES	SE	CH	TW	AE	GB	US
AU							x	x			x			x										x	x				x	x
BE							x	x			x	x				x	x						x	x	x	x			x	x
BR								x			x			x			x							x						x
CA							x	x			x			x										x	x					x
CN							x	x			x			x										x	x				x	x
FR	x		x		x	x		x	x	x	x	x		x		x	x		x		x		x	x	x	x	x		x	x
DE	x		x	x	x	x	x		x	x	x	x				x	x		x		x	x	x	x	x	x	x		x	x
HK							x	x			x			x										x						x
IN							x	x			x			x										x	x					x
IT	x		x	x	x	x	x	x	x	x		x		x		x	x		x		x	x	x	x	x	x	x		x	x
JP			x				x	x			x			x									x	x	x	x			x	x
NL	x			x	x	x	x		x	x	x	x				x	x		x		x	x	x	x	x	x	x		x	x
PL			x				x	x			x			x			x						x	x	x		x		x	x
PT			x	x			x	x			x			x		x							x	x	x				x	x
RU							x	x			x			x																x
SA																														x
SG							x	x			x			x										x	x				x	x
ZA								x			x			x										x	x					x
KR			x				x	x			x	x		x		x	x							x	x	x			x	x
ES	x		x	x	x	x	x	x	x	x	x	x		x		x	x				x	x	x		x	x			x	x
SE	x		x		x	x	x	x		x	x	x		x		x	x				x	x	x	x		x	x		x	x
CH			x				x	x			x	x		x									x	x	x				x	x
TW							x	x			x			x		x									x				x	x
GB	x		x			x	x	x			x	x		x		x	x				x		x	x	x	x	x			x
US	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	

The source-host country matrix marks the country pairs with non-missing information on MNE employment. The data set also includes flows from Morrocco to France. Countries are denoted with two letter ISO codes.

Corporate transferees. The Finaccord data contain information on corporate transferees

- from the source countries Australia, Belgium, Brazil, Canada, China, France, Germany, Hong Kong, India, Italy, Japan, Netherlands, Poland, Portugal, Russia, Saudi Arabia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, the United Kingdom, the United States

- in the host countries listed as source countries, as well as in Bahrain, Kuwait, Oman, Qatar, and the United Arab Emirates.

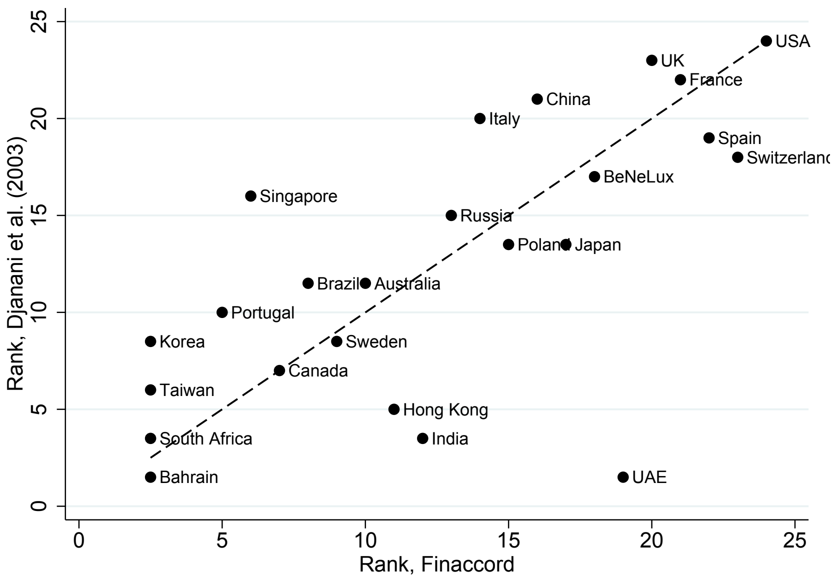
The data also contains information on corporate transferees from Indonesia in Hong Kong, Japan, Qatar, Saudi Arabia, Singapore and Taiwan, from Mexico in Canada and Spain, from Morocco in France and Spain, from the Philippines in Canada, Hong Kong, Japan and Taiwan, from Thailand in Japan and Taiwan, and from Vietnam in China, Japan and Poland.

Employment by MNEs. I use information on the total employment by MNEs from the Organization for Economic Co-operation and Development (OECD). The data contain information reported by the host and the source country. To measure the employment of MNEs from country j in a country k, I use the data on inward employment reported by country k. To measure the employment of MNEs from country k in country j, I use the data on outward employment reported by country k. I only use information reported by the source (host) country j to measure inward (outward) employment of country k if the report from country k is missing.

The employment data are not available for all country pairs with corporate transferees information, predominantly because some countries are not OECD members and/or do not report. Table F.1 displays the country pairs in the final dataset.

F.3 Plausibility check

Figure F.1: Ranking of destinations of German corporate transferees, Finaccord data vs. Djanani et al. (2003)



The Finaccord data contain information on the number of individuals sent as corporate transferees from Germany by destination. Djanani et al. (2003) provide information on the number of firms sending corporate transferees to a destination. The figure compares the ranking of destinations in the two sources. The x-axis plots the rank of a destination in the Finaccord data, and the y-axis plots its rank in Djanani et al. (2003), on a scale of 1=fewest transferees/companies to 24=most transferees/companies. The dashed line is the 45-degrees line. A Wilcoxon test fails to reject the Null that the two rankings are equal ($z = -.501, \Pr > |z|) = 62\%$). The data on Korea, Taiwan, South Africa and Bahrain are censored in the Finaccord data.

F.4 Descriptive statistics

Table F.2: Summary statistics, section 5

(a) Full sample

	N	Mean	SD	Min	Med	Max
Log # corporate transferees	769	5.306	1.175	4.605	4.605	11.717
Indicator: # transferees censored	769	0.546	0.498	0	1	1
Log total # MNE employees	316	10.739	1.750	5.517	10.908	14.678
Share of corporate transferees	316	0.020	0.059	0	0.004	0.402
Log share of corporate transferees	316	−5.393	1.470	−8.698	−5.623	−0.912
Office hours overlap	769	5.397	3.293	0	5.5	10
Flight time in hours	339	8.558	5.547	0.583	9.250	24.167
Common spoken lang.	744	0.192	0.267	0	0.043	1
Linguistic proximity	744	0.823	1.308	0	0	5.838
Log bandwidth (Mbit/s)	695	12.776	2.028	7.448	13.305	15.761

(b) Regression sample

	N	Mean	SD	Min	Med	Max
Log # corporate transferees	316	5.347	1.095	4.605	4.677	11.184
Indicator: # transferees censored	316	0.478	0.500	0	0	1
Log share of corporate transferees	316	−5.393	1.470	−8.698	−5.623	−0.912
Office hours overlap	316	5.446	3.543	0	4	10
Flight time in hours	316	8.577	5.633	0.583	9.542	24.167
Common spoken lang.	316	0.294	0.284	0	0.247	1
Linguistic proximity	316	1.444	1.502	0	1.547	5.838
Log bandwidth (Mbit/s)	293	13.817	1.235	7.448	13.816	15.761

The table displays summary statistics of the corporate transferees data for the full and the regression sample.

F.5 Additional regression results

The data on the corporate transferees is left censored at 100. I assume that the error term is normally distributed $\epsilon_{jk} \sim N(0,1)$ and estimate the Tobit model:

$$\ln(\#\text{ corporate transferees}_{jk}) = \beta_0 + \beta_1 \ln(\text{employment}_{jk} + \text{employment}_{kj}) + \beta_2 \theta_{jk} + \epsilon_{jk} \qquad \text{(F.1)}$$

The Tobit model does not allow to include source and host country fixed effects because they entail an incidental parameters problem: almost 60 fixed effects are estimated off 300 observations. Further, the Tobit model suffers from simultaneity bias because the employment at MNEs has to be included as control variable.

Table F.3: Regression results on the log number of corporate transferees

Log # of transferees	1	2	3	4	5
Log total # of MNE employees	0.766***	0.779***	0.774***	0.761***	0.758***
	(0.075)	(0.078)	(0.078)	(0.079)	(0.078)
Office hours overlap	−0.071**				−0.072°
	(0.024)				(0.051)
Flight time in hours		0.037*			−0.017
		(0.015)			(0.030)
Common spoken lang.			0.141		0.235
			(0.280)		(0.281)
Linguistic proximity			−0.164**		−0.102+
			(0.054)		(0.057)
Log bandwidth (Mbit/s)				0.010	0.015
				(0.104)	(0.097)
Constant	−3.201***	−4.055***	−3.474***	−3.699*	−3.090+
	(0.851)	(0.948)	(0.878)	(1.510)	(1.694)
# observations	316	316	316	315	315
# source countries	26	26	26	26	26
# host countries	30	30	30	29	29
R-squared	0.187	0.183	0.185	0.185	0.197

Robust standard errors in parentheses. ° $p < 0.20$, + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
The table displays the regression results of the Tobit model F.1.

G Sketch of a monitoring based model

An MNE consists of n_h managers in the headquarters in the home country, and n_j production workers in the home country $j = 0$ and the foreign country $j = 1$. Production workers input labor to the production process and the managers supervise them.

As in Qian (1994), output depends on the effort level a_j exerted by the production workers in country j: $q_j = n_j a_j$. Exerting effort is costly. The cost of effort is described by the function $g(a)$ with $g'(a) > 0$. The managers supervise the production workers to ensure that they exert a sufficient amount of effort. I assume that the managers exert full effort $a_h = 1$ in supervision, as in the literature. Production workers know that they are monitored at any point in time with probability P_j.

The workers receive the wage w_j if they are monitored and exert a sufficient amount of effort $a_j \geq a_j^*$ or if they are not monitored, and nothing if they are supervised and found to exert insufficient effort $a_j < a_j^*$. It is necessary to assume that they receive the wage whenever they are not monitored because the firm would otherwise have an incentive to claim that they are never monitored. If workers can prove whether they are monitored or not, the first best solution is implementable (Qian, 1994).

The optimal wage is determined by the incentive compatibility constraint that

$$w_j - g(a_j^*) \geq P_j \cdot 0 + (1 - P_j) \cdot w_j - g(a_j) \quad \forall a_j < a_j^*,$$

so $w_j = \frac{1}{P_j} g(a_j^*)$. Wages increase in the optimal effort level a_j^* and decrease in the monitoring probability P_j.

The firm chooses the country and firm specific optimal monitoring probabilities P_j and the optimal effort levels $a^*_{j,\, j=0,1}$ to minimize the overall costs of production, which are made up of factor input costs and monitoring costs. The costs θ_j to monitor a worker vary by country. It is generally assumed that $\theta_1 \geq \theta_0$, so foreign workers are more costly to monitor. The monitoring costs are influenced by the firm specific monitoring technology ψ, where lower ψ corresponds to a better monitoring technology. The cost minimization problem of an MNE is given by

$$C(q_0, q_1) = \min_{\{P_j, a_j^*\}_{j=0}^1} \sum_{j=0}^{1} n_j (w_j + \psi \theta_j P_j) + n_h$$

$$
\begin{aligned}
\text{s.t.} \quad & n_j a_j^* \geq q_j \qquad \forall j \\
& n_h \geq \sum_{j=0}^{1} n_j P_j \\
& w_j = \frac{1}{P_j} g(a_j^*) \\
& n_h \geq 0, P_j \in [0,1] \quad \forall j \\
& n_j \geq 0, a_j^* \geq 0 \qquad \forall j
\end{aligned}
$$

The remuneration of managers is normalized to 1.

The optimal effort levels are uniform across countries:

$$
a_j^* = \frac{2g(a_j^*)}{g'(a_j^*)}
$$

The optimal monitoring probabilities are given by

$$
P_j = \left(\frac{g(a_j)}{1 + \psi\theta_j} \right)^{\frac{1}{2}}
$$

The optimal monitoring probabilities thus decrease in the monitoring costs θ_j, and increase in better monitoring technologies ψ^{-1}. Within firms, foreign workers consequently receive higher optimal wages, and the marginal costs of production are higher, in countries with higher cross border monitoring costs. The mechanism is therefore suitable for rationalizing the within-firm differences in sales in Table 3.

As foreign marginal costs increase in θ_j, only firms with better monitoring technolgies ψ^{-1} are able to profitably invest abroad. Consequently, the remuneration of domestic production workers of MNEs is lower than the remuneration of production workers of domestic firms, as $P_0 = \left(\frac{g(a_0^*)}{1+\psi\theta_0} \right)^{\frac{1}{2}}$ decreases in ψ and w_0 decreases in P_0. Workers at foreign affiliates of MNEs and workers at domestic firms in the foreign country with the same marginal costs receive the same wages. These implications are at odds with the empirical evidence.