The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

The Organization of Knowledge in Multinational Firms

Anna Gumpert

CESifo Working Paper No. 5401

Category 8: Trade Policy

June 2015

Presented at CESifo Area Conference on Global Economy, May 2014

ISSN 2364-1428

CESifo

Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5401*

The Organization of Knowledge in Multinational Firms

Abstract

This paper provides the first in-depth study of the organization of knowledge in multinational firms. The paper develops a theoretical model that studies how firms optimally split knowledge between their headquarters and their production plants if communication costs impede the access of production plants to headquarter knowledge. The paper assumes that the foreign plants of multinational firms face higher communication costs with headquarters than their domestic plants, and shows that multinational firms therefore systematically assign more knowledge to both their foreign and domestic plants than non-multinationals. This helps explain why multinational firms pay higher wages to their production workers than non-multinational firms, and why their sales and their investment probability decrease across space. Empirical evidence from data on corporate transferees confirms the model predictions for multinationals' organization of knowledge. Data on German multinational firms corroborate the implications of the model in relation to the geography of multinationals' sales and investments.

JEL-Code: D210, D240, F210, F230.

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

Anna Gumpert
University of Munich
Seminar for Comparative Economics
Akademiestrasse 1
Germany – 80799 Munich
anna.gumpert@econ.lmu.de

June 2015
I am grateful to Pol Antrás and Esteban Rossi-Hansberg for helpful discussion and suggestions throughout the development of this project, and to Monika Schnitzer for her constant encouragement and support. I thank Lorenzo Caliendo, Kerem Cosar, Arnaud Costinot, Nadja Dwenger, Carsten Eckel, Stefania Garetto, Inga Heiland, Elhanan Helpman, James R. Hines Jr., Wolfgang Keller, Wilhelm Kohler, Chen Li, Yanping Liu, James Markusen, Marc Melitz, Eduardo Morales, Raffaella Sadun, Heiner Schumacher, Ludwig Straub, Uwe Sunde, Alexander Tarasov, MartinWatzinger, Stephen Yeaple, and the seminar and conference participants at Harvard University, MIT, the University of Munich, Princeton University, the EDGE Jamboree 2013, the 2013 Bundesbank MiDi Workshop, the CESifo Area Conference on Global Economy 2014, the Doctoral Meetings organized by the RIEF network 2014, the ETSG conference 2014, the 2014 annual conference of the Verein für Socialpolitik, the FIW Research Conference 2014, and the European Winter Meetings of the Econometric Society 2014 for their comments. 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 Antrás 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 SFB-Transregio 15 and GRK 801, and through the German National Academic Foundation (Studienstiftung des deutschen Volkes) under the ERP Fellowship Programme.

1 Introduction

In today's economy, knowledge is an essential production factor. Knowledge is typically tacit and embodied in individual employees. Production processes are complex and involve many different employees. The efficient organization of knowledge is therefore a key ingredient for firms' success. It determines in which part of the production process employees specialize, 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.

A growing body of literature studies the organization of knowledge in firms. So far, the literature assumes that the costs of communication between employees 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 in 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 country 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 production network 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 the first in-depth study of the organization of knowledge in multinational firms. The paper develops a theory and confirms its predictions using novel data on the flows of corporate transferees between countries and data on German multinational firms.

In particular, 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). To keep the model analytically tractable and make the special features of multinational firms' organization transparent, I assume that the total knowledge level of firms is exogenously given and fix the number of hierarchical layers. Firms consist of two layers: managers in the domestic corporate headquarters and workers in production plants that can be located in the same country as the headquarters, or a foreign country, or both. Firms endogenously choose the number of managers and workers, as well as the proportion of the total knowledge that they learn. 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). I assume that firms

[1] The share of multinational firms in all firms is low, but they account for a substantial fraction of aggregate output and employment: 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).

are heterogeneous with respect to their total knowledge level. Each firm optimally organizes its knowledge. The organization of knowledge yields endogenous marginal production costs that are heterogeneous across firms. This makes it possible to derive predictions on firms' sales and on the self-selection of firms 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.

Second, multinational firms assign less knowledge to their headquarters than non-multinational firms (i.e., purely domestic firms or exporters). This is because 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 the utilization rate in non-multinationals. Providing knowledge at headquarters is costly, though. 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.

The optimal organization of knowledge yields endogenous marginal production costs. The marginal costs depend on the total knowledge level of the firm and home and foreign country characteristics because these variables affect the organization of knowledge. A special feature of multinational firms with foreign and domestic production plants is that their marginal costs are interdependent across countries. This result arises because several production plants 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. Through the endogenous marginal costs, the organization of knowledge helps explain distinct stylized facts concerning multinational firms.

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). Nonetheless, theoretical explanations are scarce. The organization of knowledge helps explain the residual multinational wage premiums: Multinationals assign more knowledge to their production plants than non-multinationals, 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 foreign sales and investment probability decrease with the distance of a country from a multinational's home country (e.g., Antràs and Yeaple, 2014).[2] 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. 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.

It is difficult to test the model predictions of the organization of knowledge as knowledge is intangible and typically proprietary. Knowledge flows within multinational firms are very hard to observe. To overcome this problem, I use unique data on the flows of corporate transferees between countries. The data are ideally suited to testing the model because 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. I find that the proportion of corporate transferees in the employment of multinationals systematically increases with the communication costs between two countries, as measured by the overlap in office hours, linguistic proximity, communication technology and flight time. This finding is consistent with the prediction that the optimal level of knowledge at a foreign plant increases with the communication costs with the headquarters.

Using firm-level data for German multinational firms, I provide evidence consistent with the interdependence of marginal costs of multinational firms across countries. Furthermore, I confirm the implications of the model for the geography of multinationals' sales and investments. In line with the model predictions, German multinationals have higher sales in countries that are characterized by lower communication costs with Germany, even after controlling for firm heterogeneity. The findings hold for regressions including geographic distance and other trade cost measures, and are robust to adding further determinants of foreign sales, e.g., the quality of the investment climate. Sector-level analyses of the choice between exporting and FDI confirm that higher communication costs discourage foreign investment: The ratio of aggregate affiliate sales to aggregate exports in a sector and foreign country decreases with higher communication costs between Germany and the foreign country, and increases with higher trade costs. The predictions for wages are not testable due to the lack of appropriate data.

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 alsoGaricano and Rossi-Hansberg, forthcoming).[3] 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

[2]The pattern is often called the "gravity of FDI", see, e.g., Irarrazabal et al. (2013), Keller and Yeaple (2013).

[3]For related work on information processing in organizations, see Radner (1993); Bolton and Dewatripont (1994).

work on the organization of exporters by Caliendo and Rossi-Hansberg (2012) and Caliendo et al. (forthcoming). To the best of my knowledge, this paper is the first to study heterogeneity in the communication costs within firms and show that this heterogeneity can be useful to understand the specific features of the behavior of firms with several plants.

Second, the paper provides new insights into the emergence of multinational firm wage premiums (for surveys of the empirical evidence, see Harrison and Rodríguez-Clare, 2010; Malchow-Møller et al., 2013).[4] By focusing on the particular features of the organization of knowledge, the paper proposes an explanation that is specific to multinational firms and distinct from the scale-related arguments used to explain exporter wage premiums.[5] Previous theories on residual MNE wage premiums build on fair wage preferences (Egger and Kreickemeier, 2013).

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.[6] 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 to their mode of internationalization.[7] The paper thus provides a coherent rationale for both the geography of FDI and MNE 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, 2014; Cristea, forthcoming; Defever, 2012; Oldenski, 2012).[8]

Finally, the paper contributes to the literature on the spatial diffusion of knowledge (for a comprehensive survey, see Keller, 2004). Investments by multinational firms are an important channel of international knowledge diffusion.[9] The paper highlights that spatial communication frictions have a substantial impact on multinational firms. Consequently, investment promotion

---

[4]Many papers document that affiliates of multinational firms pay higher wages than domestic firms. The wage premium decreases, but remains significant if regressions control for firm and industry characteristics, and account for the endogeneity of foreign take-overs. It tends to be higher in developing than in developed countries (e.g., Aitken et al., 1996; Hijzen et al., 2013). Although worker heterogeneity accounts for part of the variation, it does not fully explain the multinational wage premium (Malchow-Møller et al., 2013). Likewise, multinational parent companies pay higher wages than domestic firms (Heyman et al., 2007).

[5]In particular, the mechanism differs from that of Caliendo and Rossi-Hansberg (2012), who show that increased demand after trade liberalization leads to adjustments in the organization of knowledge that explain why exporter wage premiums may emerge. The exporter wage premium is thus based on size, whereas the multinational wage premium proposed here results because several plants share common headquarters.

[6]In Keller and Yeaple (2013), firms can either produce intermediaries locally but subject to efficiency losses due to sector-specific communication frictions with the headquarters, or import intermediaries from home, subject to transport costs. Irarrazabal et al. (2013) assume that headquarter inputs are required in the production process of affiliates, but decay due to iceberg-type transport costs. In both papers, the marginal costs of production are independent across countries and do not vary with the production quantity.

[7]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.

[8]Fort (2014) studies the impact of communication technology on the fragmentation of production processes.

[9]Take-overs of domestic firms by multinational firms cause productivity increases in take-over targets through restructuring and the introduction of new technologies (Arnold and Javorcik, 2009). Knowledge spillovers lead to productivity gains for multinationals' suppliers and customers (Harrison and Rodríguez-Clare, 2010).

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 describes why data on corporate transferees provide evidence on the organization of knowledge in multinational firms and presents the regression results. Section 5 contains the empirical evidence concerning the geography of multinational firms' sales, and explains how the analyses allow discrimination between a knowledge based mechanism and a monitoring 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.[10]

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 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. Given $\bar{z}_i$, 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 at home, in the foreign country, or in both countries. He hires employees in the headquarters and the production plant(s) and determines the organization of knowledge. These activities capture non-rival headquarter services similar to those in Markusen (1984) and the subsequent MNE literature. The entrepreneur receives the market wage as well as profits.[11]

To study the differences between the optimal organization of knowledge of domestic and

[10] Fuchs et al. (forthcoming) is an example of recent work on the emergence of hierarchical structures similar to the knowledge hierarchies studied in this paper in a contracting framework.

[11] The entrepreneur receives the market wage because his outside option is to supply time to the labor market.

multinational firms 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 employees in 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(s) are called workers.[12] The term "multinational firm (MNE)" refers to any firm with a foreign production plant. 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 problem arrival rate. A higher value of λ implies that the mass of the probability distribution is concentrated close to zero. Intuitively, 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.

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.[13] Learning knowledge is costly: Employees have to hire teachers to train them. Teachers have to 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).[14]

[12] Managers are hired only at the headquarters. This assumption can be micro-founded by assuming that the entrepreneur provides a moderation service in the headquarters that decreases the communication costs between the headquarters and the plants below the level of communication costs that prevails in the economy. Alternatively, unit constraints at the managerial level provide a micro-foundation for domestic management.

[13] The entrepreneur's time is fully used to provide leadership services.

[14] I follow Caliendo and Rossi-Hansberg (2012) as their approach allows a straightforward general equilibrium analysis. The results of my paper do not depend on these specific assumptions. They merely require that the

The workers and the managers can communicate and leverage the potentially different knowledge levels. Communication is costly. As is standard in the literature (e.g., Bolton and Dewatripont, 1994; Garicano, 2000), the receiver of a message bears the communication costs: he has to spend time listening. The communication costs, i.e., the amount of time that the receiver spends listening, depends on whether the sender is located in the same or another country. The receiver in country j spends $\theta_{kj} \geq 0$ units of time listening to senders 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. 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. 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.[15]

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 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

entrepreneur remunerates the employees for the learning of knowledge. One could alternatively assume that there are multiple time periods and that firms have to hire and remunerate employees during a training period. Very low search costs then suffice to ensure that firms find it optimal to pay trained employees higher wages.

[15]The model describes 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 functionalities of machines, but employ experts for exceptional applications. In service firms, counter personnel deals with common requests of clients, and refers them to specialists for particular needs. As Garicano and Hubbard (2007) show, the framework describes how law firms split tasks between associates and lawyers.

interval of managers ranges from a lower bound Z_h to $\bar{Z}$. z_h denotes the length of this interval $[Z_h, \bar{Z}]$.[16] The entrepreneur chooses the knowledge levels 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 a firm with production plants in the home and the foreign country, and comprises a firm with only a domestic or a foreign plant as special cases. 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.

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

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

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

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

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

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

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 problem arrival rate 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.

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

[16] 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.

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 with production plants in the home and the foreign country. In a setting with only one plant, overlaps cannot be optimal: The overlap of managerial knowledge and workers' knowledge increases costs, but remains unused at the 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.[17]

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.[18]

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.

[17] 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.

[18] Appendix B.1 contains the first order conditions.

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.[19] 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 to the lower communication costs, and in the country with higher wages and learning costs.[20]

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

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

Analogously, if the firm only establishes a production plant abroad, z_1, n_1 and n_h are given by the constraints (2)-(4) with $n_0 = 0$ and managerial knowledge is implicitly defined by

$$\theta_{10} e^{-\lambda(\bar{z} - z_h)} w_0 (c_0 + \lambda(1 + c_0 z_h)) - w_1 c_1 = 0. \tag{11}$$

The first order conditions (9), (10), and (11) 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, $n_j w_j c_j$, or, for a firm with two plants, $\sum_{j=0}^{1} \mathbf{1}(z_j = \bar{z} - z_h) n_j w_j c_j$. The marginal costs are composed of the costs of increasing managerial knowledge, $n_j \theta_{j0} e^{-\lambda(\bar{z} - z_h)} w_0 c_0$, 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, $n_j \theta_{j0} e^{-\lambda(\bar{z} - z_h)} \lambda w_0 (1 + c_0 z_h)$, or $\sum_{j=0}^{1} \mathbf{1}(z_j =$

[19] The parameter restrictions in Appendix A ensure that hiring workers who learn the full set of knowledge and do not communicate with the headquarters is not optimal.

[20] 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 $z_{\bar{j}} = \bar{z} - z_h$.

$\bar{z} - z_h)n_j\theta_{j0}e^{-\lambda(\bar{z}-z_h)}\lambda w_0(1 + c_0z_h)$. The number of workers is omitted from equations (10) and (11), as are domestic wages from equation (10).

A comparison of equations (9), (10) and (11) shows that the optimal organization of knowledge systematically differs between firms with one and two production plants. The knowledge levels of managers and workers in a firm with one production plant depend only on variables that are exogenous to the firm. They are independent of the production quantity. In contrast, a firm with two production plants additionally takes the production quantity into account in allocating knowledge. As is shown in subsection 3.1, the firm organizes in such a way that results in greater cost reduction for a plant the larger its output.

The marginal costs of production are given by

$$\xi_j = \frac{1}{1-e^{-\lambda\bar{z}}}\left[w_j(1 + c_j(\bar{z} - z_h)) + w_0(1 + c_0z_h)\theta_{j0}e^{-\lambda(\bar{z}-z_h)}\right] \qquad \text{for } z_j = \bar{z} - z_h; \qquad (12a)$$

$$= \frac{1}{1-e^{-\lambda\bar{z}}}\left[w_j(1 + c_jz_j) + \frac{1}{\lambda}w_jc_j\right] \qquad \text{for } z_j > \bar{z} - z_h. \qquad (12b)$$

The marginal costs 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.

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 problem arrival rate* λ*.*

Table 1: Comparative statics

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

The table displays the effects of the model parameters on the optimal knowledge levels separately for firms with a domestic production plant, firms with a foreign production plant, and the different cases for firms with plants in both countries. + denotes positive effects, − negative effects, +/− ambiguous effects and 0 no influence. Results denoted * only apply to $j = 1$. Results denoted ** hold if $q_{\bar{j}}\theta_{\bar{j}0}e^{-\lambda(\bar{z}-z_h)}\lambda(1 + c_0z_h) > q_{\hat{j}}\theta_{\hat{j}0}e^{-\lambda z_{\hat{j}}}c_0$, 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 have a production plant in the home country, or in the foreign country, or plants in both countries. 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 usually have similar effects on the organization of knowledge of the firm in the different cases.

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.[21] 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 single 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. Analogously, higher wages w_j decrease the knowledge level of workers and affect managerial knowledge in MNEs.[22]

If a larger quantity q_j is to be produced, more workers need to be hired, each of whom receives $w_j(1 + c_j z_j)$. A firm with two production plants can mitigate this cost increase by adjusting the optimal organization of knowledge within its organization.[23] The production quantity does not affect the workers' optimal knowledge level for firms with only one production plant. An increase in the production quantity leads to a proportional increase in the number of workers, which in turn causes a proportional increase in the number of managers. Similarly, wages scale the total costs of production for a firm with only a domestic production plant. 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.[24]

The knowledge level of the workers and the knowledge level of managers both increase with the total knowledge of the firm $\bar{z}$. The problem arrival rate 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

[21]This prediction is consistent with empirical findings in Bloom et al. (2012) who show, using survey data, that plant managers have fewer responsibilities if the CEO is on site than if he is off site.

[22]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.

[23]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.

[24]The results for domestic firms match the results for single-establishment firms in Bloom et al. (2014).

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.

Taking the first order conditions for managerial knowledge (9), (10), and (11), and the comparative statics together reveals that the optimal level of managerial knowledge in MNEs is systematically different from the optimal managerial knowledge in domestic firms.

Proposition 2. *Multinational firms systematically choose lower levels of managerial knowledge than domestic firms if* $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$.

Proof. See Appendix B.3. □

Intuitively, MNEs choose a lower level of managerial knowledge than domestic firms to ensure an efficient utilization rate of knowledge. Workers in foreign plants have higher levels of knowledge than if they were employed in a domestic plant because of the higher cross-border communication costs. They thus turn to headquarters for help less frequently than workers in domestic firms. This decreases the utilization rate of managerial knowledge. At the same time, managerial knowledge is equally costly for domestic firms and MNEs. MNEs consequently decrease the amount of managerial knowledge to balance its utilization rate and its costs.[25]

In summary, section 2 shows that the optimal organization of knowledge in firms differs with the firm's multinational status. MNEs assign systematically higher levels of knowledge to their workers and systematically lower levels of knowledge to their headquarters to avoid the higher communication costs with the foreign market. MNEs with two production plants may choose asymmetric knowledge levels for their domestic and foreign workers. Their organization of knowledge depends on the foreign and domestic production quantities, whereas the production quantity does not influence the organization of knowledge in single-plant firms.

3 The implications for MNEs' foreign sales and wages

3.1 The self-selection of firms into foreign investment

The analysis of the choice between domestic activity, exporting, and FDI focuses on a firm 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:

$$U(x_j(\bar{z})) = \left(\int_{\Omega_j} x_j(\bar{z}_i)^{\frac{\sigma-1}{\sigma}} M_j \mu(\bar{z}) d\bar{z} \right)^{\frac{\sigma}{\sigma-1}}, \tag{13}$$

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

[25] As will become clearer below, the parameter restriction $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$ is likely to hold for the majority of foreign investments: Foreign investment is only worthwhile if foreign wages and learning costs are sufficiently low to outweigh the comparatively high communication costs between the foreign plant and the headquarters.

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:

$$p_j(\bar{z}_i) = q_j(\bar{z}_i)^{-\frac{1}{\sigma}} Q_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}}, \tag{14}$$

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. Alternatively, the entrepreneur can establish a foreign production plant solely to sell output abroad at fixed costs f^I, and, for additional fixed costs of f^V, export output back to the home country. Such firms are called "vertical MNEs" what follows. To ship output from country k to country $j \neq k$, the firm incurs iceberg transport costs $\tau > 1$.[26] "Horizontal MNEs" serve consumers from two local production plants at fixed costs $f^D + f^I$.

I assume that $f^I > \tau^{\sigma-1} f^X > \frac{Q_1 P_1^{\sigma-1}}{Q_0} f^D$ and $f^V < \tau^{1-\sigma} \frac{Q_0}{Q_1 P_1^{\sigma-1}} f^I$. It is thus never optimal to export but not to serve the domestic market, or to establish only a foreign production plant and not export back home.

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, V for vertical MNEs, and I for horizontal MNEs.[27]

Production quantities and sales. The entrepreneur's profit maximization problem in the case of horizontal FDI is given by

$$\max_{q_0^I, q_1^I \geq 0} \quad \pi^I(\bar{z}_i, w_0, w_1) = \sum_{j=0}^{1} p_j(q_j^I(\bar{z}_i)) q_j^I(\bar{z}_i) - C(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1). \tag{15}$$

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

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

[26] τ units of the good have to be shipped for one unit to arrive.
[27] 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 $q_1 \in \{\tau q_0^V + q_1^V, q_1^I\}$.

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

$$q_j^I(\bar{z}_i) = Q_j P_j^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \xi_j(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1) \right)^{-\sigma}. \tag{16}$$

The entrepreneur analogously maximizes profits for vertical FDI and exporting:

$$\max_{q_0^V, q_1^V \geq 0} \quad \pi^V(\bar{z}_i, w_0, w_1) = \sum_{j=0}^{1} p_j(q_j^V(\bar{z}_i)) q_j^V(\bar{z}_i) - C(\bar{z}_i, w_0, \tau q_0^V(\bar{z}_i) + q_1^V(\bar{z}_i), w_1) \tag{17}$$

$$\max_{q_0^X, q_1^X \geq 0} \quad \pi^X(\bar{z}_i, w_0) = \sum_{j=0}^{1} p_j(q_j^X(\bar{z}_i)) q_j^X(\bar{z}_i) - C(\bar{z}_i, q_0^X(\bar{z}_i) + \tau q_1^X(\bar{z}_i), w_0). \tag{18}$$

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

$$q_0^V(\bar{z}_i) = Q_0 \left(\frac{\sigma}{\sigma-1} \tau \xi_1(\bar{z}_i, w_0, w_1) \right)^{-\sigma} \qquad q_1^V(\bar{z}_i) = Q_1 P_1^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \xi_1(\bar{z}_i, w_0, w_1) \right)^{-\sigma} \tag{19}$$

$$q_0^X(\bar{z}_i) = Q_0 \left(\frac{\sigma}{\sigma-1} \xi_0(\bar{z}_i, w_0) \right)^{-\sigma} \qquad q_1^X(\bar{z}_i) = Q_1 P_1^{\sigma-1} \left(\frac{\sigma}{\sigma-1} \tau \xi_0(\bar{z}_i, w_0) \right)^{-\sigma}. \tag{20}$$

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

Optimal quantities vary by mode. As is well-known, an exporter or vertical MNE sells larger quantities in the country with the production plant than in the country served via trade by $\tau > 1, \sigma > 1$, so concentrating production in one location is more profitable the lower the transport costs τ.

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

$$q_0^D(\bar{z}_i) = q_0^X(\bar{z}_i) \geq q_0^I(\bar{z}_i). \tag{21}$$

This result arises because the entrepreneur cannot tailor the headquarters of a horizontal MNE to domestic needs. Correspondingly, domestic profits are lower in the case of FDI than in the case of exporting or domestic activity. Analogously, quantities sold in the foreign market are higher if the firm conducts vertical FDI than if it conducts horizontal FDI:

$$q_1^V(\bar{z}_i) \geq q_1^I(\bar{z}_i). \tag{22}$$

The higher fixed costs and the sales foregone with two production plants are only worthwhile if the quantities sold in the second country exceed the quantities that could be sold via trade:

$$q_1^I(\bar{z}_i) > q_1^X(\bar{z}_i); \quad q_0^I(\bar{z}_i) > q_0^V(\bar{z}_i), \tag{23}$$

i.e., if foreign production quantities in case of horizontal FDI exceed foreign export quantities, and domestic production quantities in case of horizontal FDI exceed domestic reimports.

Optimal quantities also vary with country characteristics.

Proposition 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. See Appendix C.1. □

As equation (19) shows, the optimal foreign production quantity of a vertical MNE varies negatively with the foreign marginal costs. The foreign marginal costs increase with the communication costs θ_{10}, the foreign learning costs c_1, and the wages w_1, so the foreign quantity decreases with these country characteristics.

Proposition 4. *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 horizontal MNEs increase with the communication costs θ_{10}. They decrease with the foreign production quantity $q_1^I(\bar{z}_i)$ and increase with the domestic production quantity $q_0^I(\bar{z}_i)$. Analogously, domestic marginal costs $\xi_0(\bar{z}_i, q_0^I(\bar{z}_i), w_0, q_1^I(\bar{z}_i), w_1)$ decrease with the domestic production quantity $q_0^I(\bar{z}_i)$ and increase with the foreign production quantity $q_1^I(\bar{z}_i)$. Consequently, the foreign production quantities and sales are generally higher in countries with lower communication costs θ_{10}.*

Proof. See Appendix C.2. □

Examining equation (16) indicates that the optimal production quantities of horizontal MNEs vary negatively with the marginal costs of production. However, the relationship between foreign country characteristics and the optimal production quantities is more complex than in the case of vertical MNEs. The marginal costs depend on the domestic and foreign production quantities due to their influence on the optimal organization of knowledge.

A horizontal MNE chooses the optimal organization of knowledge in such 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, $k \neq j$. The foreign country characteristics thus have a direct and an indirect effect on the production quantities. 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.[28]

It is more difficult to determine the impact of the foreign learning costs and wages on the optimal foreign production quantities of horizontal MNEs because it is not possible to determine their effect on the foreign marginal costs of production in an unambiguous manner.

[28]The indirect adjustments through the production quantities lead to an analytically ambiguous overall effect of the foreign communication costs on the foreign production quantity only if the workers' knowledge levels are symmetric and $w_1 c_1 \theta_{00} < w_0 c_0 \theta_{10}$. The effect is always negative in simulations.

Foreign wages w_1 and learning costs c_1 have a positive direct effect on the foreign marginal costs of production, but changes in wages and learning costs also affect the organization of knowledge. These adjustments work against the direct positive effect, i.e., they decrease the marginal costs.[29] The total effect of foreign wages and learning costs on the marginal costs is thus analytically ambiguous.

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).

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

The different production modes affect the organization of knowledge and thus the marginal production costs. In particular, horizontal FDI leads to the reorganization of knowledge in the firm compared to the case with only one production plant. Unlike previous models of horizontal FDI (e.g., Helpman et al., 2004), the marginal production costs are interdependent across countries. Domestic (or, in the case of vertical FDI, foreign) marginal costs are affected by the decision to set up a second plant, so total net profits—domestic and foreign net profits—with two plants have to exceed the total net profits of exporting or producing only abroad.

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 MNEs' organization of knowledge affects the aggregate foreign investment flows between countries. 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. Serving both countries using a foreign production plant cannot be optimal as the positive effect of communication costs on marginal costs is not outweighed by differences in the other country characteristics. Firms either sell their product only domestically, or export it to the foreign market, or conduct horizontal 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. As indicated at the beginning of the theory section, 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.

[29]The adjustment effect is positive if workers' knowledge levels are symmetric and $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$.

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.3 contains the proofs.

The question of interest is how the frictions in cross-border communication affect the aggregate export and foreign investment flows. 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.

$$wf^D = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\xi_0(\bar{z}^*, w)^{1-\sigma} - w \tag{24}$$

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 $\bar{z}^X$.

$$wf^X = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q(\tau\xi_0(\bar{z}^X, w))^{1-\sigma} - w \tag{25}$$

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*:

$$\begin{aligned}&\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\left(\xi_0(\bar{z}^I, q_0^I(\bar{z}^I), w, q_1^I(\bar{z}^I), w)^{1-\sigma} + \xi_1(\bar{z}^I, q_0^I(\bar{z}^I), w, q_1^I(\bar{z}^I), w)^{1-\sigma}\right) - wf^I = \\ &\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\,\xi_0(\bar{z}^I, w)^{1-\sigma}(1+\tau^{1-\sigma}) - wf^X \end{aligned}\tag{26}$$

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 (24) and (25) 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. Taking wages as given, an increase in τ implies that the exporting cut-off knowledge level $\bar{z}^X$ increases: It is profitable to export to more distant destinations only for firms with lower marginal costs.

$\bar{z}^I > \bar{z}^X$ 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 firms have to have a higher knowledge level to 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 of production, the marginal costs cut-off is thus shifted downwards.

Given wages, the knowledge cut-off $\bar{z}^I$ increases with the communication costs θ_{10} between the host country and the MNE's home country, and decreases with the transportation costs τ between the home and the host country because export profits decrease with τ and profits from FDI decrease with θ_{10}.

To determine the other equilibrium variables, it is necessary to consider the remaining equilibrium conditions. Entrepreneurs enter up to the point at which the net value of entry is zero. The *free entry condition* is given by[30]

$$\begin{aligned} wf = & \int_{\bar{z}^*}^{\bar{z}^I} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\xi_0(\bar{z},w)^{1-\sigma} - w(1+f^D)dG(\bar{z}) \\ & + \int_{\bar{z}^X}^{\bar{z}^I} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q(\tau\xi_0(\bar{z},w))^{1-\sigma} - w(1+f^X)dG(\bar{z}) \\ & + \int_{\bar{z}^I}^{\bar{z}_{max}} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\left(\xi_0(\bar{z},q_0(\bar{z}),w,q_1(\bar{z}),w)^{1-\sigma} + \xi_1(\bar{z},q_0(\bar{z}),w,q_1(\bar{z}),w)^{1-\sigma}\right) \\ & - w(1+f^D+f^I)dG(\bar{z}) \end{aligned} \tag{27}$$

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 $C(\bar{z},\cdot)$.[31]

$$\begin{aligned} N = & \frac{M}{1-G(\bar{z}^*)}\left(f + \left(\int_{\bar{z}^*}^{\bar{z}^I} f^D + C(\bar{z}, q \in \{q_0^D, q_0^X + \tau q_1^X\}, 1)dG(\bar{z}) + \int_{\bar{z}^X}^{\bar{z}^I} f^X dG(\bar{z})\right.\right. \\ & \left.\left. + \int_{\bar{z}^I}^{\bar{z}_{max}} f^D + f^I + C(\bar{z}, q_0, 1, q_1, 1)dG(\bar{z})\right)\right) \end{aligned} \tag{28}$$

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

$$wN = Q \tag{29}$$

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

Inspection of the free entry condition shows that the parameters have an additional effect on the export and FDI knowledge cut-offs through wages. Equilibrium wages decrease 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

[30]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.

[31]By symmetry, the demand for domestic labor from domestic MNEs and MNEs from the foreign country sum to the total demand for labor from a domestic MNE.

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 provides novel insights into MNE wage premiums. Consistent with empirical evidence (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.[32] The wage premiums arise via two different channels: due to firm organization and due to a selection effect.

Firm organization provides a new and MNE-specific explanation for residual MNE wage premiums, i.e., for wage premiums paid by MNEs compared to non-MNEs with the same observable characteristics such as sales. As outlined in section 2, MNEs choose an organization of knowledge with higher levels of worker knowledge than if they were non-MNEs. The higher communication costs involved in foreign production increase the optimal level of knowledge at their foreign plant. MNEs therefore decrease managerial knowledge to balance the utilization rate and costs, and increase the knowledge level of their domestic workers.

Empirical studies typically compare MNEs and non-MNEs with the same observable characteristics. As the marginal costs decrease with total firm knowledge, a non-MNE with the same marginal costs as an MNE has lower knowledge $\bar{z}$ than the MNE. This reinforces the difference in workers' knowledge levels. Thus, MNEs pay higher remuneration to workers than non-MNEs with the same marginal costs of production and the same sales.

Proposition 5 summarizes the results.

Proposition 5. *Vertical MNEs pay higher remuneration to foreign workers than non-MNEs in the foreign country with the same marginal costs and foreign sales. Horizontal MNEs pay higher remuneration to domestic workers than non-MNEs in the home country with the same marginal costs and domestic sales if* $\theta_{00} w_1 c_1 < \theta_{10} w_0 c_0$. *They pay higher remuneration to foreign workers than non-MNEs in the foreign country with the same marginal costs and foreign sales if* $\theta_{00} w_1 c_1 < \theta_{10} w_0 c_0$ *and* $c_1 \geq c_0$. *The parameter restrictions are sufficient, but not necessary conditions.*

Proof. Follows from Proposition 2, see also Appendix B.3. □

Inspection of the parameter restriction $\theta_{00} w_1 c_1 < \theta_{10} w_0 c_0$ shows that the model predicts residual MNE wage premiums both for developed to developed and developed to developing

[32] The model abstracts from contractual imperfections, which are relevant in understanding the evolution of managerial wages (e.g., Marin et al., 2014). This section therefore focuses on predictions for workers' remuneration.

country FDI. Foreign wages and learning costs must not exceed domestic wages and learning costs by more than the friction in cross-border communication: $w_1 c_1 < \frac{\theta_{10}}{\theta_{00}} w_0 c_0$. This includes the case $w_0 c_0 \approx w_1 c_1$, 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 organization of knowledge is an MNE-specific explanation of MNE wage premiums. 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 MNE wage premiums due to the self-selection of firms into FDI. Only firms with a higher knowledge level $\bar{z}$ become MNEs.[33] 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-Møller et al., 2013).

4 Corporate transferees and the organization of knowledge

Proposition 1 describes how the optimal organization of knowledge in MNEs varies with the communication costs between a plant and the headquarters, as well as country and firm characteristics. Its predictions are very difficult to test: Knowledge is intangible in nature, and typically proprietary. Within-MNE knowledge flows are thus very hard to observe.

To overcome this problem, 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

[33] Only higher $\bar{z}$ firms select into vertical FDI whenever $\pi_0^V(\bar{z}^*, w_0, w_1) < (f^I + f^D) w_0$.

of time, for example from the headquarters to a foreign affiliate.[34] 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.).[35] 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 can be used to obtain evidence concerning the predictions of Proposition 1.

Proposition 1 shows that the optimal knowledge level at a foreign plant increases with the communication costs between the plant and the headquarters. MNEs have two options to increase knowledge at a foreign plant through corporate transferees: Either they send knowledgeable employees to the foreign operations, or they train foreign employees at the headquarters. In either case, Proposition 1 predicts that the observed share of corporate transferees in MNEs' employment should increase with the communication costs between the two countries.

Prediction 1. *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.*

4.1 Data

Obtaining evidence for Prediction 1 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.[36] 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 D.1 provides a list of source and host countries.

[34] Many countries provide special visa for corporate transferees, for example the US L-1 visa. Such visas are typically available only for managers, specialists, and trainee employees (e.g., European Union, 2014).

[35] The Canadian Employee Relocation Council (CERC) conducted a survey of Canadian firms in October 2013. Djanani et al. (2003) surveyed all listed stock corporations in Germany in 2003. Both surveys allowed multiple responses. Though firms also use corporate transferees for other purposes—to support the establishment of foreign operations (65.3%) and improve communication within the firm (55.8%) in Germany, and to address talent shortages in the local labor market (61%) and for career development (43%) in Canada—Bonache and Brewster (2001) argue that only knowledge transfer consistently explains the use of corporate transferees by MNEs, in particular why their use has increased despite advances in communication and information technologies.

[36] Back of the envelope calculations using the survey data from Djanani et al. (2003) suggest that more than two thirds of corporate transferees are transferred for periods longer than one year.

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
		Common official/native/spoken language, linguistic proximity	Melitz and Toubal (2014)
		Linguistic proximity to German (only section 5)	Author using CEPII, Ethnologue, Spolaore and Wacziarg (2009)
		Internet bandwidth (Mbit/s)	ITU's ICT Indicators Database
Foreign country characteristics (only section 5)			
Q_{jt}	Market size	GDP, GDP per capita	IMF
c_{jt}	Learning costs	Average years of schooling	Barro and Lee (2013)
w_{jt}	Wage	Unit labor costs	OECD
τ_{jt}	Trade costs	Distance (population weighted)	CEPII
		Effectively applied tariffs by sector	WITS
		Costs of importing/enforcing contracts	World Bank Doing Business
Additional controls (only section 5)			
	Investment climate	Statutory tax rate	IBFD
		Rule of law/ regulatory quality/ government effectiveness/ corruption	World Governance Indicators
	Fixed costs of FDI	Costs/time/# procedures of starting a business	World Bank Doing Business
		Bilateral investment treaty	UNCTAD
	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.

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 D.1 provides details concerning the variable construction.

Communication costs. To approximate the bilateral communication costs θ_{jk}, I employ the overlap of office hours, the flight time between the main economic cities of the source and host countries, several measures for the similarity of languages, and the internet bandwidth as a measure of communication technologies. I refrain from generic measures with various alternative interpretations, such as distance. Table 2 provides an overview of the model parameters, their empirical analogs, and the data sources.

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 higher the share of corporate transferees should be. The variable is computed as $\max\{10 - |\text{time difference in hours}|, 0\}$.

The flight time captures how quickly managers can travel to the foreign operations and address potential issues in the production process on site.[37] Despite technological advances, face-to-face communication is often indispensable to ensure successful production (e.g., UNCTAD, 2004). As higher flight times impede communication, the share of corporate transferees should increase with the flight time.

Linguistic proximity captures difficulties in the direct communication between two individuals. I employ the common official language indicator, the common spoken and common native language measures, as well as the linguistic proximity index from Melitz and Toubal (2014). "Common spoken language" measures the probability $\in [0, 1]$ that two randomly chosen individuals from two countries speak the same language, and "common native language" encompasses the probability that they share their native 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. Undoubtedly, international business communication often takes place in English. Still, non-native English speakers tend to develop their own English dialect, 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.

Summary statistics are provided in Appendix D.2.

4.2 Empirical specification

I specify the following regression equation to provide evidence concerning Prediction 1:

$$\ln\left(\frac{\#\text{ corporate transferees}_{jk}}{\text{Employment}_{jk} + \text{Employment}_{kj}}\right) = \beta_0 + \beta_1\theta_{jk} + \beta_2 d_{\text{cens}} + \boldsymbol{\alpha}_k + \boldsymbol{\alpha}_j + \epsilon_{jk} \tag{30}$$

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. As indicated above, 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.

[37] I use Frankfurt for the flight times from and to Germany for consistency with section 5.

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. Proposition 1 shows that the host country learning costs and wages affect the knowledge level at an affiliate. However, it is difficult to derive their impact on the flow of corporate transferees without information on whether the transferees are sent from headquarters to affiliates or vice versa. 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, resulting in simultaneity bias because the size of foreign operations depends on the organization of knowledge reflected in the corporate transferees. Appendix D.3 reports the results of Tobit regressions without fixed effects that are in line with the main results.

4.3 Descriptive evidence and regression results

Figure 1 provides graphical evidence of the relationship between the share of corporate transferees in total bilateral employment by MNEs and the communication costs. The figure plots the mean share of corporate transferees for the quartiles of the communication cost measures.[38]

Figure 1 lends strong support to Prediction 1: As expected, the share of corporate transferees in the total employment of MNEs increases with higher bilateral communication costs. The mean share of corporate transferees is lower for higher quartiles of the distribution in office hours overlap. Likewise, it is lower for higher quartiles of the probability that two randomly chosen individuals speak the same language, and for higher quartiles of the log internet bandwidth. The mean share of corporate transferees is higher for higher quartiles of the flight time.

The regression results in Table 3 confirm the graphical evidence. 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. The language variables all enter negatively. Only common spoken language and the linguistic proximity index are (marginally) significant, consistent with the ease of communication being decisive for corporate transferee flows.[39] Quantitatively, a higher prob-

[38]Figure 1 provides bar plots as scatter plots are difficult to interpret for three of the four communication cost measures: the office hours overlap is categorical, the internet bandwidth takes on a limited number of values as it is a host country characteristic, and the common spoken language variable has a point mass at 0. Appendix D.3 provides the scatter plot for the flight time.

[39]Jointly included in regressions, common official, native, and spoken languages capture different aspects of language: common native language captures the impact of common ethnicity and trust, common official language the effect of institutions, and common spoken language the ease of communication (Melitz and Toubal, 2014).

Figure 1: Share of corporate transferee flows vs. bilateral communication costs

Figure 1a: Office hours overlap

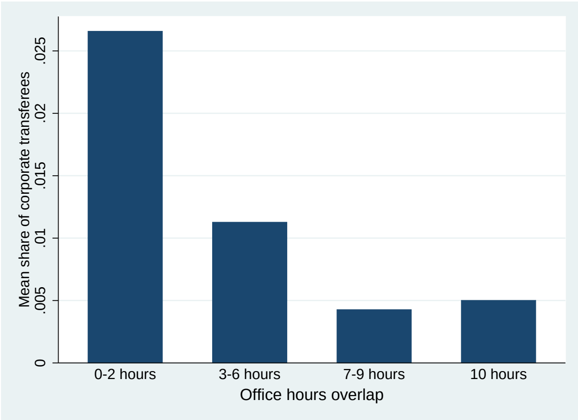


Figure 1b: Flight time

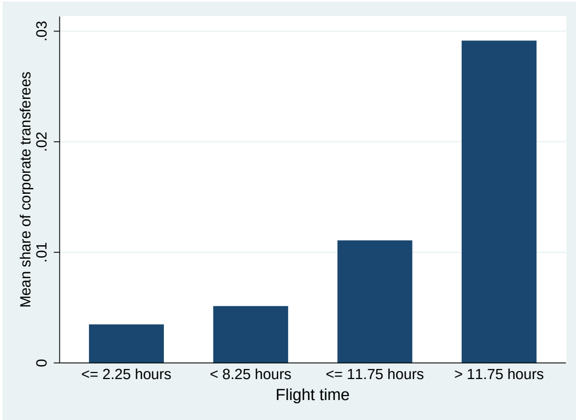


Figure 1c: Common spoken language

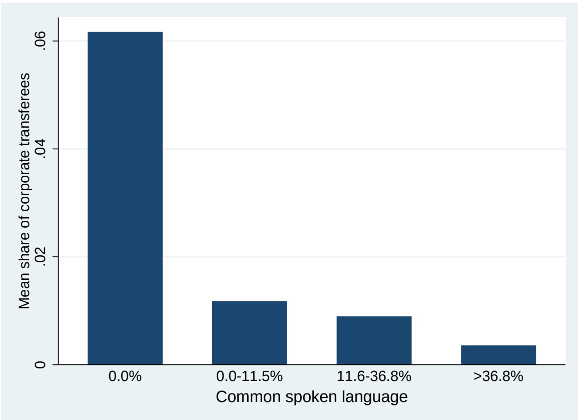
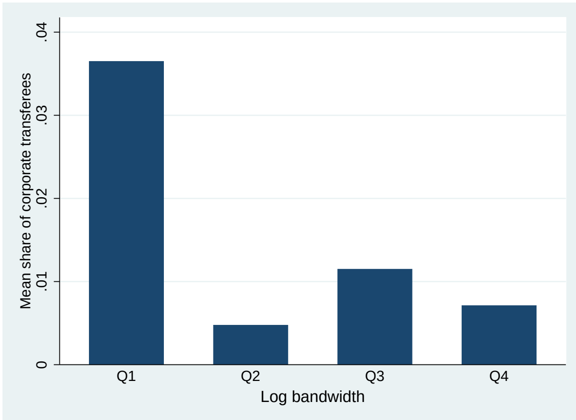


Figure 1d: Log bandwidth



Figures 1a-1d plot the share of corporate transferees in total bilateral employment by MNEs for quartiles of communication cost measures: office hours overlap (1a), flight time (1b), the probability that two randomly chosen individuals speak the same language (1c), and log bandwidth (1d). The figures include only uncensored observations.

ability of 10 percentage points that two randomly chosen individuals speak the same language is associated with decrease in the share of corporate transferees of 5%. 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 significant covariates from columns 1 to 4 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.

Table 3: 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 official lang.			−0.256					
			(0.324)					
Common spoken lang.			−0.514°		−0.540*	0.274	−0.035	−0.069
			(0.359)		(0.246)	(0.314)	(0.324)	(0.385)
Common native lang.			−0.053					
			(0.663)					
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. ° $p < 0.20$, + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Dependent variable:* log share of corporate transferees in MNE employment. *Covariate definitions:* see Table 2.

In summary, the findings confirm Prediction 1: The share of corporate transferees increases with the communication costs. As corporate transferees are predominantly used for knowledge transfer, the findings imply that MNEs assign higher levels of knowledge to foreign affiliates if their communication costs with the headquarters are higher, consistent with Proposition 1.

5 Communication costs and MNEs' foreign sales

5.1 Data

I use detailed firm-level data on German MNEs to obtain empirical evidence concerning the model implications for the distribution of MNEs' foreign sales derived in section 3. I augment the data with sector-level export data. The data are uniquely suited to studying the model predictions. Germany has long been one of the largest source countries of both exports and FDI worldwide: According to WTO and UNCTAD information, Germany was the largest or second largest exporter in the world and held the fourth largest foreign investment position during the 1999-2010 sample period. The communication costs between the home and the foreign country are the main driver of the model predictions, so the empirical analyses put particular emphasis on them. The implications for wages are not testable due to the lack of appropriate data.

Foreign sales. I use the German central bank's Microdatabase Direct investment (MiDi). 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 (Lipponer, 2009). 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. From 2002 onwards, information on the domestic sales and the number of employees is available.

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. The great level of detail in the MNE data makes it possible to obtain evidence concerning the model predictions for the distribution of foreign sales both at the firm and the sector level.

Exports. I use data on German exports from BACI, a trade database provided by CEPII and containing cleaned trade information from the Comtrade database. The data are at the six-digit product level. The database does not contain firm information. I translate the product codes into two-digit sector codes using conversion tables from the United Nations Statistics Division to use the export data and the foreign sales data in sector-level analyses.

Communication costs and other covariates. I employ the same communication cost measures as in section 4, with the exception of linguistic proximity. The linguistic proximity variable in Melitz and Toubal (2014) only takes four values. The classification of languages by Ethnologue provides seven linguistic nodes for German, so I construct a refined 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.[40]

The home country k is Germany for all MNEs in my sample. As the communication cost measures are mostly time-invariant, using host country fixed effects is not possible. I therefore control for the other model parameters and potential omitted variables. Table 2 provides an overview of the model parameters, the empirical analogs, and the data sources.

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[41] and employ information on the unit labor costs to measure wages.[42]

I employ data on the bilateral geographic distance, the average effectively applied tariffs in the parent sector, and the costs of importing and of enforcing contracts to measure trade costs. Taken together, these data cover many of the aspects of trade costs that Anderson and van Wincoop (2004) list in their survey.[43] To account for factors that may influence sales but are

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

[41] 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.

[42] I have experimented with the Occupational Wages around the World Database by Freeman and Oostendoorp. However, its overlap with the MiDi database is limited. Data are available for only 20% of observations.

[43] Data on cif/fib–values are not available for Germany, so the physical transport costs are difficult to capture.

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.

I take the logarithms of the covariates if their distribution in levels is skewed. Appendix E.2 provides summary statistics.

5.2 Communication costs and the within-firm distribution of sales

The high level of detail in the MiDi database makes it possible to provide empirical evidence concerning the MNE-level predictions of the model. Propositions 3 and 4 show that both horizontal and vertical MNEs' foreign sales decrease with the communication costs between the headquarters and the foreign plant. This yields the second empirical prediction.

Prediction 2. *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.*

5.2.1 Empirical specification

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 log-linearize the theory expression for sales $p_j(\bar{z}_i)q_j(\bar{z}_i)$: $\ln\left(p_j(\bar{z}_i)q_j(\bar{z}_i)\right) = (1-\sigma)\ln\left(\frac{\sigma}{\sigma-1}\right) + \ln Q_j + (\sigma-1)\ln P_j + (1-\sigma)\ln \xi_j(\bar{z}_i, q_0, w_0, q_1, w_1)$. I estimate a reduced-form version of the resulting equation. Due to the non-linear nature of the original equation, it is not possible to provide a structural interpretation of the resulting parameter estimates.

$$\ln\left(\text{foreign sales}_{ijt}\right) = \beta_0 + \beta_1\theta_{j0t} + \beta_2 Q_{jt} + \beta_3 c_{jt} + \beta_4 w_{jt} + \delta\mathbf{X}_{jt} + \alpha_{it} + \epsilon_{ijt} \tag{31}$$

The dependent variable is the natural log of the foreign sales of firm 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. β_1 is expected to be positive unless the communication costs are measured with the flight time. I control for 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, 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 absorb the effect of $\bar{z}_i$ and, more generally, of any MNE characteristics that may influence performance across destinations. To account for correlations of sales across countries and over time, the standard errors are clustered by MNE.

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. Subsection 5.3 addresses the issue of self-selection. It is necessary to keep it in mind when interpreting the regression results in this subsection.[44]

5.2.2 Regression results

Table 4 presents the regression results for Prediction 2. The table displays seven specifications. Columns 1 and 2 contain the model parameters: the communication and learning costs, the wages, and the market size. The specifications are displayed separately because the wage data are only available for OECD countries, so the sample size decreases once wages are included. Columns 3 to 5 add the trade costs. Columns 6 and 7 additionally include measures for the quality of the investment climate.

Table 4 displays the number of MNEs, the number of country combinations, and the number of countries in the regression sample, 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.[45] 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 2. The office hours overlap has a positive effect on foreign sales typically significant at the 5% or 1% level. The flight time generally has a highly significant negative effect on foreign sales. The two measures are not always both significant, as they are strongly correlated, similar to the regression results in section 4. At least one of the language variables has a significantly positive effect through specifications, except for column 5, where the sample size is low. Higher internet bandwidth significantly increases foreign sales. The coefficient is positive and of a similar size, but insignificant in column 5, which probably also stems from the smaller number of observations. Consistent with the model, the regressions show that communication costs affect MNEs' foreign sales, even if other determinants of foreign sales are controlled for.

The other covariates have plausible effects. Foreign sales tend to be higher in larger countries as measured by log GDP. Log GDP per capita has a negative effect on foreign sales once labor costs are controlled for. Given the market size and the labor costs, MNEs tend to sell less in less populous countries. Lower learning costs, as measured by higher average years of schooling, significantly increase foreign sales, as do lower unit labor costs. 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 trade costs—as reflected by higher distance and higher tariffs—increase foreign sales, consistent with a horizontal motive of foreign investment. The positive effect of distance is notable: Once communication costs are controlled for, distance does not decrease foreign sales.

[44] A natural experiment, i.e., an exogenous change in the communication costs with a subgroup of countries, would help to address this problem. Estimating a Heckman selection model, as in Keller and Yeaple (2013), is an alternative option, but would not permit a focus on within-MNE variation across countries as using firm–year fixed effects is not possible.

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

Table 4: Regression results: within-firm differences in log foreign sales across countries

Log foreign sales	1	2	3	4	5	6	7
Office hours overlap	0.047***	0.034**	0.028*	0.033*	0.022	0.028+	0.030+
	(0.008)	(0.013)	(0.013)	(0.015)	(0.018)	(0.015)	(0.016)
Log flight time	0.020	−0.166***	−0.541***	−0.530***	−0.496*	−0.556***	−0.597***
	(0.027)	(0.046)	(0.125)	(0.146)	(0.252)	(0.144)	(0.174)
Common official language	−0.032	−0.164	−0.172+	0.092	0.012	0.009	0.016
	(0.055)	(0.103)	(0.103)	(0.120)	(0.150)	(0.123)	(0.146)
Common spoken language	−0.091	−0.137	0.067	0.279+	0.227	0.276+	0.273
	(0.078)	(0.137)	(0.144)	(0.167)	(0.196)	(0.167)	(0.208)
Common native language	0.208*	0.083	−0.002	0.506**	0.374	0.492**	0.596**
	(0.092)	(0.168)	(0.173)	(0.186)	(0.292)	(0.186)	(0.216)
Linguistic proximity	0.308*	0.808**	0.818**	−0.062	0.141	0.036	−0.042
	(0.134)	(0.308)	(0.307)	(0.341)	(0.464)	(0.341)	(0.407)
Log internet bandwidth	0.074***	0.033*	0.033*	0.029+	0.015	0.027+	0.005
	(0.010)	(0.014)	(0.014)	(0.016)	(0.026)	(0.016)	(0.016)
Log GDP	0.253***	0.375***	0.357***	0.373***	0.375***	0.355***	0.385***
	(0.014)	(0.017)	(0.018)	(0.021)	(0.028)	(0.022)	(0.026)
Log GDP per capita	0.139***	−0.043	−0.072*	−0.128***	−0.168+	−0.140***	−0.146**
	(0.014)	(0.031)	(0.031)	(0.038)	(0.089)	(0.039)	(0.048)
Average years of schooling	0.027***	0.070***	0.066***	0.070***	0.084***	0.072***	0.077***
	(0.006)	(0.009)	(0.009)	(0.010)	(0.012)	(0.010)	(0.013)
Unit labor cost		−0.771***	−0.807***	−0.629**	−0.689*	−0.669**	−0.401
		(0.191)	(0.191)	(0.228)	(0.337)	(0.227)	(0.273)
Log distance			0.312***	0.336**	0.251	0.341**	0.372**
			(0.094)	(0.109)	(0.205)	(0.109)	(0.130)
Effectively applied tariffs				0.007***	0.005	0.007***	0.006**
				(0.002)	(0.008)	(0.002)	(0.002)
Log costs of importing					0.050		
					(0.084)		
Log costs of enforcing contracts					−0.033		
					(0.056)		
Statutory tax rate						0.007*	
						(0.003)	
Regulatory quality							0.154**
							(0.069)
Rule of law							−0.249**
							(0.076)
Government efficiency							0.216***
							(0.054)
Corruption							−0.047
							(0.050)
# observations	148,426	103,167	103,167	74,789	36,972	74,789	61,741
# MNEs	8,524	7,363	7,363	5,396	3,587	5,396	5,025
# country combinations	8,802	8,453	8,453	6,413	3,763	6,413	5,646
R-squared	0.193	0.155	0.156	0.161	0.135	0.161	0.155

Clustered standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant and parent–year dummies included. *Dependent variable:* log foreign sales per MNE, country, and year. *Covariate definitions:* see Table 2. *# MNEs:* number of MNEs. *# country combinations:* number of combinations of countries with MNE activity. Regressions include 105 countries in column 1 and 23 countries in columns 2 to 7.

The trade cost measures are all insignificant in column 5, which is most probably driven by the low number of observations as the costs of importing and enforcing contracts are only available for a relatively small number of countries and years.

MNEs tend to sell more in countries with a higher statutory tax rate. This probably reflects the fact that higher sales are necessary to produce profitably despite the higher tax rates. MNEs also tend to sell more in countries with better governance. One coefficient is surprising: The rule of law measure enters as significantly negative. Taken together, though, the investment climate measures still speak in favor of a better investment climate increasing foreign sales.

5.2.3 Evidence of the interdependence of MNEs' sales

The comprehensiveness of the MiDi database also makes it possible to provide evidence concerning the prediction that MNEs' sales are interdependent across countries.

Prediction 3. *An MNE's foreign sales in a host country j decrease with the production quantities at other investment destinations $q_k, k \neq j$ of the MNE.*

Prediction 3 is based on the second implication of Proposition 4: The sales of an MNE in different countries are interdependent because the affiliates share common headquarters. The MNE organizes knowledge such that the marginal costs of production are lower the larger an affiliate, at the expense of higher marginal costs in other locations. Consequently, output in country $k \neq j$, ceteris paribus, has a negative impact on the sales in country j.

Providing evidence regarding this prediction entails several difficulties. First, regressing sales in j on sales in k is subject to endogeneity bias due to simultaneity. Second, firms non-randomly establish and shut down affiliates. Finally, an MNE's output and sales depend on total firm knowledge. The model takes knowledge as given, but in practice, MNEs may have an incentive to innovate if their affiliates grow. Thus, high sales in k may lead to high sales in j through adjustments in firm knowledge.

To address these issues, I focus on foreign affiliates of German MNEs in one foreign country. I choose France, because it yields the highest number of observations for the following analyses. I follow the French affiliates from the first year in my sample and study whether their sales growth is negatively correlated with the sales growth of their parents' affiliates in other countries. I focus on sales growth rather than sales in levels to take account of the fact that high foreign sales may lead to adjustments in firm knowledge that increase sales across countries. To address the simultaneity bias, I apply an identification strategy proposed by Desai et al. (2009) and use GDP growth as an exogenous determinant of the sales growth in instrumental variable (IV) regressions. I instrument the observed sales growth of an MNE in other countries with the GDP growth at the 1999 locations of the firm. Thus, the estimation is robust to non-random entry and exit.

Specifically, I estimate the following regression:

$$\text{Sales growth}_{i\,\text{France}\,t} = \beta_0 + \beta_1 \text{Sales growth}_{i\,j \neq \text{France}\,t} + \boldsymbol{\alpha}_s + \boldsymbol{\alpha}_t + u_{it} \qquad (32)$$

i.e., I regress sales growth in France on sales growth at the MNE's other locations, and a set of sector and year dummies $\boldsymbol{\alpha}_{s,t}$. I estimate two variants of the equation. First, I use the difference in log sales in France as the dependent variable and the log difference of average sales at the MNE's other locations as the independent variable. I instrument the latter with the

Table 5: Regression results: within-firm interdependence

	Δ log sales		Sales growth rate	
	1	2	3	4
Δ log average other sales	0.199***	–0.758°		
	(0.028)	(0.463)		
Average other sales growth rate			–0.889°	–0.827°
			(0.569)	(0.515)
# observations	4,405	4,152	4,082	4,082
# MNEs	685	621	580	580
Specification	OLS	IV	IV	IV
Instrument	–	GDP	GDP	GDP per capita

Clustered standard errors in parentheses. ° $p < 0.20$, + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. P-values: 10.1% in column 2, 11.8% in column 3, and 10.8% in column 4. Constant, year, and sector dummies included. *Dependent variable:* log difference of foreign sales in France per MNE and year in columns 1 and 2, sales growth rate in France per MNE and year in column 3 and 4. *Covariate definitions: Δ log average other sales:* log difference of average sales at other foreign locations of MNE; *Average other sales growth rate:* average growth rate of sales at other foreign locations of the MNE. The sales growth rates in France and other countries are the ratios of changes in sales between years to averages of the beginning and end of period values.

log difference of average GDP at the MNE's 1999 locations. Second, I calculate sales growth rates following Desai et al. (2009), and instrument the sales growth rates at the MNE's other locations using GDP growth rates and, to make them directly comparable, the growth rates of GDP per capita.

Table 5 displays the regression results. Higher sales growth at the other locations of the firm is associated with higher sales growth in France according to the OLS regression in column 1. The magnitude of the coefficient is similar to the coefficients found in Desai et al. (2009). Column 2 presents the results of the IV specification that takes the endogeneity of the sales growth at the other locations due to simultaneity and due to innovations in the knowledge level into account. Consistent with the predictions of the model, I find that higher sales growth at the other locations of an MNE is associated with *lower* sales growth in France. The coefficient is marginally significant with a P-value of 10.1%. Columns 3 and 4 confirm this result.

The results support Prediction 3: Higher growth at an MNE's other foreign locations is associated with lower growth in France. This finding may seem at odds with the finding of Desai et al. (2009), who show that higher foreign employment and asset growth are associated with higher domestic growth in these variables. In fact, though, the different findings are easily reconcilable within the model of the organization of knowledge in MNEs. On the one hand, Proposition 4 predicts that the sales of an MNE decrease with the production quantities at the MNE's other location, consistent with Table 5. On the other hand, foreign activity may lead to employment growth in the home country—even if the domestic production quantity decreases—because the MNE hires additional managers at headquarters. In addition, multinational parents may expand domestic assets and employment because foreign growth entails an incentive to innovate, as also pointed out by Desai et al. (2009).

Although Table 5 thus supports the predictions of the model in this paper, it is important to note that various forces may drive the results. Other than the organization of knowledge, financial or managerial constraints may lead to a negative association in sales across countries (see, e.g., Manova et al., forthcoming; Yeaple, 2013). Unfortunately, disentangling the sources is not possible as it requires more detailed information on MNEs' organization and activities.

5.3 Communication costs, aggregate exports and affiliate sales

The effect of the bilateral communication costs on single MNEs, together with the model predictions on the self-selection of firms into foreign destinations, implies that communication costs have an important impact on aggregate foreign investment flows.

Prediction 4. *The ratio of aggregate affiliate sales to aggregate exports in a host country j decreases with the communication costs between country j and the home country k, θ_{jk}. It increases with the trade costs between country j and the home country k, τ_{jk}.*

5.3.1 Empirical specification

To test Prediction 4, I aggregate the affiliate-level information on foreign sales in the MiDi data at the country–affiliate sector–year–level. I aggregate across affiliate sectors because the model predicts that foreign affiliate sales and exports are substitutable. I merge the aggregate affiliate sales data with the sector level export data from the BACI database.

I specify a regression equation similar to Helpman et al. (2004) and Oldenski (2012):

$$\ln\left(\frac{\text{Affiliate sales}_{sjt}}{\text{Exports}_{sjt}}\right) = \beta_0 + \beta_1\theta_{j0t} + \beta_2\tau_{j0t} + \beta_3 Q_{jt} + \beta_4 c_{jt} + \beta_5 w_{jt} + \beta_6 \mathbf{X}_{jt} + \boldsymbol{\alpha}_s + \boldsymbol{\alpha}_t + \epsilon_{sjt} \quad (33)$$

The dependent variable is the natural logarithm of the ratio of total affiliate sales to total export sales in a sector s in country j and year t. The covariates of interest are the bilateral communication costs θ_{j0t} and the trade costs τ_{j0t} between a foreign country j and Germany, country 0. The other covariates are defined as above. $\boldsymbol{\alpha}_s$ is a vector of sector fixed effects that capture sectoral differences in the fixed costs of exporting and FDI, differences in the size dispersion of firms across sectors, and other sectoral characteristics that may affect the choice between exporting and FDI (for the impact of sectoral characteristics, see Helpman et al., 2004; Oldenski, 2012). $\boldsymbol{\alpha}_t$ is a vector of year fixed effects. ϵ_{sjt} is an error term.

The model predicts that higher communication costs have a negative effect and higher trade costs have a positive effect on the ratio of affiliate sales to exports. One component of the trade costs are information costs (Anderson and van Wincoop, 2004). The measures for the bilateral communication costs are likely to be correlated with the information costs. In fact, both language and flight time empirically affect trade flows (e.g., Melitz and Toubal, 2014; Hummels and Schaur, 2013), although Helpman et al. (2008) have shown that language predominantly affects the self-selection of firms into exporting. This introduces a bias against finding an effect from the communication costs because the negative effect of the communication costs on the aggregate foreign sales has to outweigh the positive influence of information costs on exports.

The sector-level regressions mitigate the bias due to the self-selection of MNEs into foreign countries that may be present in the MNE-level regressions. Aggregating over firms integrates out random unobservables that affect both the performance and the investment decision of firms. At the same time, the sector level regressions may still be biased if exports or foreign sales are zero for some sector–country combinations: Observed zeros are non-random and result if no firm in a sector is sufficiently productive to invest in or export to a certain country (Helpman et al., 2008). I observe positive exports in all sectors to all foreign countries with a full set of covariates, so I do not need to account for self-selection into exporting. However, I do not observe positive foreign sales in all sector–foreign country pairs. I therefore adapt the two-stage estimation procedure in Helpman et al. (2008) and estimate a selection equation for foreign investment using the costs, time, and number of procedures of starting a business and the existence of a bilateral investment treaty as exclusion restrictions (see Appendix E.3 for details). I then insert non-linear transformations of the predicted probability to conduct FDI in estimation equation (33). Specifically, I insert the Mill's ratio and polynomials of the sum of the Mill's ratio and the inverse predicted probability of FDI, as Helpman et al. (2008).

5.3.2 Regression results

Table 6 displays the regression results. Columns 1 to 3 contain the measures for the trade costs, market size, learning cost, and wages, i.e., only known determinants of the export–FDI trade-off. Column 4 adds the communication costs. Columns 5 and 6 additionally include investment climate measures. Column 7 takes selection into foreign investment into account.[46]

Consistent with the literature, higher trade costs—as reflected in greater geographic distance and higher tariffs—shift the decision between exporting and affiliate sales towards affiliate sales. The effects are significant, with the exception of column 3. Column 3 includes the costs of importing and enforcing contracts as additional trade cost measures. The costs of importing take up the effect of distance and tariffs. Higher costs of enforcing contracts, although listed as trade costs in Anderson and van Wincoop (2004), have a stronger negative effect on affiliate sales than exports, most likely reflecting their correlation with the quality of the investment climate. Notably, the effect of the average tariffs is of similar size throughout the specifications, whereas the coefficient of distance increases substantially once the communication costs measures are included. This finding is consistent with the interpretation that distance picks up the effect of omitted communication cost measures in gravity regressions of foreign affiliate sales, which leads to a negative coefficient of distance in the gravity regressions and a downward bias in the coefficients in columns 1 to 3.

The results for the impact of the communication costs strongly support the model predictions. A larger overlap in office hours increases the ratio of affiliate sales to exports. The effect is highly significant and of similar magnitude throughout the specifications. A longer flight time has the predicted negative effect on the dependent variable. The effect is insignificant if the specification accounts for selection into foreign investment, although this is probably driven by a loss of precision reflected in the much larger standard error. A higher probability of a common

[46] The number of observations drops in column 7 because the costs of starting a business used as an exclusion restriction are only available from 2003, and all MNEs invest in all countries in three sectors, so it is not possible to estimate a selection equation including sector dummies for these sectors.

Table 6: Regression results on the aggregate affiliate sales vs. exports by sector

Log ratio of affiliate sales to exports	1	2	3	4	5	6	7
Log distance	0.309***	0.100**	−0.033	1.684***	1.689***	1.555***	1.292*
	(0.023)	(0.031)	(0.039)	(0.192)	(0.175)	(0.308)	(0.534)
Effectively applied tariffs	0.013***	0.007	−0.003	0.008**	0.008**	0.007**	0.010+
	(0.004)	(0.012)	(0.026)	(0.002)	(0.002)	(0.002)	(0.006)
Log costs of importing			0.602***				
			(0.139)				
Log costs of enforcing contracts			−0.835***				
			(0.109)				
Office hours overlap				0.228***	0.226***	0.241***	0.196***
				(0.030)	(0.030)	(0.037)	(0.059)
Log flight time				−1.009***	−1.025***	−0.725+	−0.737
				(0.287)	(0.259)	(0.437)	(0.659)
Common official language				0.063	0.034	0.569+	−0.588
				(0.214)	(0.230)	(0.311)	(0.494)
Common spoken language				1.013**	1.001**	1.017*	−0.521
				(0.342)	(0.320)	(0.403)	(0.651)
Common native language				0.466	0.468	0.924*	−0.892
				(0.343)	(0.395)	(0.415)	(0.667)
Linguistic proximity				0.365	0.395	−0.580	2.626*
				(0.654)	(0.713)	(0.841)	(1.096)
Log internet bandwidth				0.158***	0.157***	0.174***	−0.024
				(0.038)	(0.036)	(0.051)	(0.112)
Log GDP	0.097***	0.382***	0.483***	0.298***	0.291***	0.356***	0.254*
	(0.027)	(0.030)	(0.055)	(0.042)	(0.042)	(0.067)	(0.100)
Log GDP per capita	−0.009	−0.255***	0.573***	−0.650***	−0.646***	−0.943***	−0.440
	(0.026)	(0.066)	(0.144)	(0.092)	(0.090)	(0.143)	(0.442)
Average years of schooling	0.026*	0.051***	0.127***	0.076***	0.078***	0.021	0.100***
	(0.011)	(0.015)	(0.022)	(0.018)	(0.022)	(0.029)	(0.035)
Unit labor cost		−1.115**	−3.383***	−1.166**	−1.187**	−1.634**	−2.927***
		(0.384)	(0.537)	(0.399)	(0.376)	(0.565)	(0.805)
Statutory tax rate					0.002		
					(0.006)		
Regulatory quality						1.176***	0.737
						(0.199)	(0.042)
Rule of law						0.175	−0.326
						(0.238)	(0.281)
Government effectiveness						0.274	0.281
						(0.179)	(0.239)
Corruption						−0.688***	−0.099
						(0.168)	(0.275)
# observations	8,921	4,254	2,071	4,166	4,166	3,439	2,209
# countries	98	22	22	22	22	22	22
# sectors	22	22	22	22	22	22	19
R-squared	0.235	0.335	0.393	0.371	0.371	0.394	0.433

Bootstrapped standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant, year, and sector dummies included. *Dependent variable:* log ratio of aggregate affiliate sales to aggregate exports in a sector, foreign country and year. *Covariate definitions:* see Table 2. Selection correction in column 7.

spoken language or higher linguistic proximity increases the ratio of affiliate sales to exports. A common native and official language are mostly insignificant, consistent with communication being the decisive driver of the effect of language on foreign affiliate sales (see also the discussion in section 4). The positive effect of linguistic proximity in column 7 is particularly noteworthy: Helpman et al. (2008) find that language does not affect exports once the selection of firms into exporting is taken into account. In contrast, higher linguistic proximity increases affiliate sales even if the regressions account for the self-selection into FDI. This underlines the relevance of communication costs for MNEs. Higher internet bandwidth has a positive effect on the dependent variable. Internet bandwidth has a positive effect on firms' self-selection into foreign countries (see Table E.2 in Appendix E.3), but does not affect the ratio of affiliate sales and exports once the self-selection is accounted for. The regression results thus confirm that communication costs have a substantial effect on the aggregate foreign investment flows.

Concerning the other covariates, GDP has a positive effect, as has the average years of schooling. GDP per capita and foreign wages have negative effects, all as in the MNE-level regressions. GDP is less significant and GDP per capita is insignificant in column 7, indicating that much of their effect in columns 1 to 6 is driven by the self-selection of firms into exporting and FDI. Higher regulatory quality and lower corruption increase foreign sales relative to exports, which is plausible, but they are not robustly significant.

In summary, the estimates strongly support that communication costs are an important determinant of aggregate foreign sales and make a convincing case in support of the general equilibrium implications of the model of the organization of knowledge in MNEs.

5.4 The organization of knowledge vs. monitoring

One could be concerned that the empirical results capture the effect of monitoring and not the effect of the organization of knowledge on MNEs' behavior, for example because the communication costs measures are correlated with monitoring costs. 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 headquarters is associated with shorter establishment survival (Kalnins and Lafontaine, 2013). Both articles attribute their findings to monitoring problems and information asymmetries between firm headquarters and establishments. While I do not deny that monitoring is an important factor in firm efficiency, I would like to stress that monitoring does not fully explain the behavior of MNEs for three reasons.

First, the data on corporate transferees provides direct evidence that higher communication costs increase within-MNE knowledge transfers. Monitoring is not even mentioned as a motivation for the use of corporate transferees in the surveys of German and Canadian firms.

Second, a monitoring model does not explain why MNE wage premiums emerge. Instead, the predictions of monitoring-based models are at odds with the empirical evidence. Assume that the costs of monitoring vary across countries and that the cross-border monitoring costs exceed the within-country monitoring costs. Then, 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

Table 7: Regression results on the impact of bilateral trust

	1 MNE level	2 Sector level
Bilateral trust	0.056	−0.129
	(0.109)	(0.234)
Office hours overlap	0.020	0.242^{***}
	(0.023)	(0.048)
Log flight time	-0.651^{***}	-1.689^{***}
	(0.192)	(0.486)
Common official language	-0.390^{+}	−0.729
	(0.211)	(0.487)
Common spoken language	−0.420	−0.747
	(0.338)	(0.527)
Common native language	0.181	−0.934
	(0.293)	(0.636)
Linguistic proximity	1.250^{*}	3.790^{**}
	(0.620)	(1.386)
Log internet bandwidth	0.008	0.108^{+}
	(0.016)	(0.056)
Log GDP	0.394^{***}	0.400^{***}
	(0.032)	(0.065)
Log GDP per capita	-0.269^{***}	-1.109^{***}
	(0.073)	(0.156)
Average years of schooling	0.120^{***}	0.169^{***}
	(0.015)	(0.039)
Unit labor cost	0.517	−0.623
	(0.375)	(0.818)
Log distance	0.289^{+}	1.911^{***}
	(0.164)	(0.334)
Effectively applied tariffs	0.006^{***}	0.003
	(0.001)	(0.028)
Regulatory quality	0.013	0.319
	(0.071)	(0.248)
Rule of law	-0.352^{***}	0.319
	(0.085)	(0.385)
Government efficiency	0.236^{***}	−0.019
	(0.055)	(0.226)
Corruption	0.067	−0.158
	(0.053)	(0.227)
Observations	56,204	2,777
R-squared	0.157	0.446

Clustered (robust) standard errors in parentheses in column 1 (column 2). $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column 1 includes constant and parent-year dummies; column 2 includes constant, year, and sector dummies. *Dependent variables:* log foreign sales per MNE, country and year in column 1; log ratio of aggregate affiliate sales to exports in a sector, country, and year in column 2. *Covariate definitions:* see Table 2.

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 F sketches a formal analysis of this argument.

Third and finally, the bilateral communication costs have an effect on foreign sales even if the monitoring costs are controlled for by including bilateral trust as a proxy variable in the regressions. Higher bilateral trust decreases monitoring costs and allows firms to decentralize more easily (Bloom et al., 2012). Table 7 displays regression results including bilateral trust between countries from the Eurobarometer survey as a covariate. Bilateral trust does not have a significant effect on the MNE-level distribution of sales or on the ratio of aggregate affiliate sales to exports. I obtain similar estimates for the communication costs, even though the number of observations is substantially reduced. Thus, monitoring cannot explain the communication cost estimates in subsections 5.2 and 5.3.

In summary, the empirical analyses provide convincing evidence in favor of the predictions in the model of the organization of knowledge in MNEs. Evidence concerning the flows of corporate transferees between countries confirms the prediction of the effect of the bilateral communication costs on the organization of knowledge. MNE-level analyses support the predictions regarding the relationship between the communication costs and MNEs' sales, and the interdependence of sales across countries. The effect of the bilateral communication costs on aggregate foreign sales and exports is likewise consistent with the model. Finally, the analyses confirm that alternative possible explanations, such as monitoring, cannot explain the estimation results.

6 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 than if they were non-multinational. The knowledge levels of production workers increase with the communication costs between the plant and the headquarters, and they are interdependent across countries. 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 multinational firm. 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 comprehensive empirical evidence in support of the different aspects of the model using novel data on the flows of corporate transferees between countries and data on German multinational firms.

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

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.

Anderson, J. E. and E. van Wincoop (2004). Trade Costs. *Journal of Economic Literature 42*(3), 691–751.

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.

Bahar, D. (2014). Heavier than Air? Knowledge transmission within the multinational firm. mimeo, Harvard University.

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 (forthcoming). The anatomy of French production hierarchies. *The Journal of Political Economy*.

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. (forthcoming). The effect of communication costs on trade in headquarter services. *Review of World Economics*.

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

Desai, M. A., C. F. Foley, and J. R. Hines (2009). Domestic effects of the foreign activities of US multinationals. *American Economic Journal: Economic Policy 1*(1), 181–203.

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.

European Union (2014). Directive 2014/66/EU of the European Parliament and of the Council of 15 May 2014 on the conditions of entry and residence of third-country nationals in the framework of an intra- corporate transfer. Official Journal of the European Union.

Fort, T. (2014). Technology and production fragmentation: Domestic versus foreign sourcing. mimeo, Dartmouth.

Fuchs, W., L. Garicano, and L. Rayo (forthcoming). Optimal Contracting and the Organization of Knowledge. *The Review of Economic Studies*.

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*.

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.

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

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *The Quarterly Journal of Economics 123*(2), 441–487.

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.

Hummels, D. L. and G. Schaur (2013). Time as a Trade Barrier. *The American Economic Review 103*(7), 2935–2959.

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.

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

Manova, K., S.-J. Wei, and Z. Zhang (forthcoming). Firm exports and multinational activity under credit constraints. *Review of Economics and Statistics*.

Marin, D., J. Schymik, and A. Tarasov (2014). Trade in Tasks and the Organization of Firms. mimeo, University of Munich.

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.

Oostendorp, R. H. (2013). The Occupational Wages around the World (OWW) Database: Update for 1983-2008. Background Paper for the World Development Report 2013, World Bank.

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

Radner, R. (1993). The organization of decentralized information processing. *Econometrica 61*(5), 1109–1146.

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

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 No. 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 equation (11).*

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 bounds).

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 a 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 bounds, 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)

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 production only	+	+	0	+	-	-
# managers n_h, foreign production only	+	+	+	+	-	-
# managers n_h, both, $z_0 = z_1 = \bar{z} - z_h$	+	+*	+*	+	-	-
# managers n_h, both, $z_j = \bar{z} - z_h$	+**	+	+	+	-	-
# managers n_h, both, $z_j > \bar{z} - z_h$	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.

$$\frac{\partial n_h}{\partial \theta_{j0}} = \frac{q_j e^{-\lambda z_j} - \lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial \theta_{j0}}}{1-e^{-\lambda\bar{z}}}$$

$$\frac{\partial n_h}{\partial c_j} = -\frac{\lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial c_j}}{1-e^{-\lambda\bar{z}}}$$

$$\frac{\partial n_h}{\partial w_j} = -\frac{\lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial w_j}}{1-e^{-\lambda\bar{z}}}$$

$$\frac{\partial n_h}{\partial q_j} = \frac{\theta_{j0} e^{-\lambda z_j} - \lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial q_j}}{1-e^{-\lambda\bar{z}}}$$

$$\frac{\partial n_h}{\partial \bar{z}} = -\frac{\lambda e^{-\lambda\bar{z}}}{(1-e^{-\lambda\bar{z}})^2} \sum_j q_j \theta_{j0} e^{-\lambda z_j} - \frac{1}{1-e^{-\lambda\bar{z}}} \sum_j \lambda e^{-\lambda z_j} q_j \theta_{j0} \frac{\partial z_j}{\partial \bar{z}}$$

$$\frac{\partial n_h}{\partial \lambda} = \sum_{j=0}^{1} \frac{(-z_j q_j \theta_{j0} e^{-\lambda z_j} - \lambda q_j \theta_{j0} e^{-\lambda z_j} \frac{dz_j}{d\lambda})(1-e^{-\lambda\bar{z}}) - \bar{z} e^{-\lambda\bar{z}} q_j \theta_{j0} e^{-\lambda z_j}}{1-e^{-\lambda\bar{z}}}$$

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

B.2.1 Only domestic production.

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 $\frac{d(10)}{dz_h} = \lambda\theta_{00} e^{-\lambda(\bar{z}-z_h)}(2c_0 + \lambda(1+c_0 z_h)) > 0$.

$$\frac{d(10)}{d\theta_{00}} = e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1+c_0 z_h)) > 0 \qquad \Rightarrow \frac{dz_h}{d\theta_{00}} < 0$$

$$\frac{d(10)}{dc_0} = \theta_{00} e^{-\lambda(\bar{z}-z_h)}(1+\lambda z_h) - 1 < 0 \qquad \Rightarrow \frac{dz_h}{dc_0} > 0$$

$$\frac{d(10)}{dw_0} = \frac{d(10)}{dq_0} = 0 \qquad \Rightarrow \frac{dz_h}{dw_0} = \frac{dz_h}{dq_0} = 0$$

$$\frac{d(10)}{d\bar{z}} = -\lambda\theta_{00} e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1+c_0 z_h)) < 0 \qquad \Rightarrow \frac{dz_h}{d\bar{z}} > 0$$

$$\frac{d(10)}{d\lambda} = \theta_{00} e^{-\lambda(\bar{z}-z_h)}((1+c_0 z_h) - (\bar{z}-z_h)(c_0 + \lambda(1+c_0 z_h))) \qquad \Rightarrow \frac{dz_h}{d\lambda} \lesseqgtr 0$$

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 Only foreign production.

By the implicit function theorem, $\frac{dz_h}{dx_1} = -\frac{\frac{d(11)}{dx_1}}{\frac{d(11)}{dz_h}}$.

The sign of $\frac{dz_h}{dx_1}$ is given by $-\frac{d(11)}{dx_1}$ because $\frac{d(11)}{dz_h} = \lambda\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(2c_0 + \lambda(1 + c_0 z_h)) > 0$.

$$\frac{d(11)}{d\theta_{10}} = e^{-\lambda(\bar{z}-z_h)}w_0(c_0 + \lambda(1 + c_0 z_h)) > 0 \qquad \Rightarrow \frac{dz_h}{d\theta_{10}} < 0$$
$$\frac{d(11)}{dc_1} = -w_1 < 0 \qquad \Rightarrow \frac{dz_h}{dc_1} > 0$$
$$\frac{d(11)}{dw_1} = -c_1 < 0 \qquad \Rightarrow \frac{dz_h}{dw_1} > 0$$
$$\frac{d(11)}{dq_1} = 0 \qquad \Rightarrow \frac{dz_h}{dq_1} = 0$$
$$\frac{d(11)}{d\bar{z}} = -\lambda\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0 + \lambda(1 + c_0 z_h)) < 0 \qquad \Rightarrow \frac{dz_h}{d\bar{z}} > 0$$
$$\frac{d(11)}{d\lambda} = \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0((1 + c_0 z_h) - (\bar{z} - z_h)(c_0 + \lambda(1 + c_0 z_h))) \qquad \Rightarrow \frac{dz_h}{d\lambda} \lesseqgtr 0$$

$\frac{dz_1}{d\theta_{10}} > 0$, $\frac{dz_1}{dc_1} < 0$, $\frac{dz_1}{dw_1} < 0$, $\frac{dz_1}{dq_1} = 0$, $\frac{dz_1}{d\lambda} \lesseqgtr 0$ 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.

B.2.3 Domestic and foreign production

$\mathbf{z_0 = z_1 = \bar{z} - z_h}$. 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 $\frac{d(9)}{dz_h} = \lambda e^{-\lambda(\bar{z}-z_h)}w_0(2c_0 + \lambda(1 + c_0 z_h))\sum_{j=0}^{1} q_j\theta_{j0} > 0$.

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

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.

$\mathbf{z_{\hat{j}} > \bar{z} - z_h, z_{\bar{j}} = \bar{z} - z_h}$**.** 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 $j = 0$.

$$
\begin{aligned}
\frac{\partial z_h}{\partial c_0} &= -d^{-1}\left(\left(q_1\theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(1+\lambda z_h) + q_0\theta_{00}e^{-\lambda z_0}w_0\right)(1+c_0z_h)\right.\\
&\quad \left. - \frac{1}{\lambda}q_1c_0\left(w_0 - \theta_{00}e^{-\lambda z_0}\lambda w_0 z_h\right)\right) \qquad < 0\\
\frac{\partial z_h}{\partial c_1} &= d^{-1}q_1w_1(1+c_0z_h) \qquad > 0\\
\frac{\partial z_h}{\partial w_0} &= -d^{-1}(1+c_0z_h)\left(q_1\theta_{10}e^{-\lambda(\bar{z}-z_h)}(c_0+\lambda(1+c_0z_h)) + q_0\theta_{00}e^{-\lambda z_0}c_0\right) \qquad < 0\\
\frac{\partial z_h}{\partial w_1} &= d^{-1}q_1c_1(1+c_0z_h) \qquad > 0
\end{aligned}
$$

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 $\hat{j}$,

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

For $\hat{j} = 1$,

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

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 $\hat{j} = 0$,

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

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.3 Managerial knowledge (Proposition 2)

Horizontal MNEs. $\theta_{00}w_1c_1 < \theta_{10}w_0c_0$, so the knowledge constraint is binding in the home country: $z_1 \geq z_0 = \bar{z} - z_h$. Take a domestic firm and a horizontal 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, 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.

Vertical MNEs. $w_1c_1 < w_0c_0$, 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 $\theta_{00}e^{-\lambda(\bar{z}-z_h^D)}w_0(c_0 + \lambda(1 + c_0z_h^D) - w_0c_0 = 0$. Vertical MNEs determine z_h^V via $\theta_{10}e^{-\lambda(\bar{z}-z_h^V)}w_0(c_0 + \lambda(1 + c_0z_h^V)) - w_1c_1 = 0$. 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 $\theta_{11}e^{-\lambda(\bar{z}-z_h^D)}w_1(c_1 + \lambda(1 + c_1z_h^D) - w_1c_1 = 0$. Vertical MNEs determine z_h^V via $\theta_{10}e^{-\lambda(\bar{z}-z_h^V)}w_0(c_0 + \lambda(1 + c_0z_h^V)) - w_1c_1 = 0$. The equations are increasing in z_h. As $\theta_{10} > \theta_{11}$ and $w_0c_0 > w_1c_1$, $z_h^V < z_h^D$.

B.4 Knowledge gaps

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

$$
\begin{aligned}
C(\bar{z}, q_0, w_0, q_1, w_1) = \min_{\{n_j,z_j\}_{j=0}^1, n_h, z_h} & \sum_{j=0}^{1} n_jw_j(1 + c_jz_j) + n_hw_0(1 + c_0z_h) + w_0 \\
\text{s.t.} \quad n_j(1 - e^{-\lambda\bar{z}}) \geq q_j & \qquad \forall j \text{ s.t. } z_j \geq \bar{z} - z_h \\
n_j(1 - e^{-\lambda z_j} + e^{-\lambda(\bar{z}-z_h)} - e^{-\lambda\bar{z}}) \geq q_j & \qquad \forall j \text{ s.t. } z_j < \bar{z} - z_h \\
n_h \geq \sum_{j=0}^{1} n_j\theta_{j0}e^{-\lambda z_j} & \\
n_h \geq 0,\ z_h \geq 0, z_h \leq \bar{z}; \quad n_j \geq 0,\ z_j \geq 0, z_j \leq \bar{z} & \qquad \forall j
\end{aligned}
$$

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.

$$
\begin{aligned}
\mathcal{L} = & \sum_{j=0}^{1} n_jw_j(1 + c_jz_j) + n_hw_0(1 + c_0z_h) + w_0 \\
& + \sum_{j=0}^{1} \xi_j \left[q_j - n_j(1 - e^{-\lambda\bar{z}} + \mathbf{1}(z_j < \bar{z} - z_h)(e^{-\lambda(\bar{z}-z_h)} - e^{-\lambda z_j}))\right] \\
& + \kappa \left[\sum_{j=0}^{1} n_j\theta_{j0}e^{-\lambda z_j} - n_h\right] - \sum_{j=0}^{1} \upsilon_jn_j - \upsilon_hn_h - \sum_{j=0}^{1} \nu_jz_j - \nu_hz_h + \sum_{j=0}^{1} \bar{\nu}_j(z_j - \bar{z}) + \bar{\nu}_h(z_h - \bar{z})
\end{aligned}
$$

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

Insights. The knowledge level of production workers is determined by

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

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

C.1 Profit maximization of vertical MNEs (Proposition 3)

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

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

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

$$\frac{\partial \pi(\cdot)}{\partial q_1(\bar{z}_i)} = \frac{\sigma-1}{\sigma} q_1(\bar{z}_i)^{-\frac{1}{\sigma}} Q_1^{\frac{1}{\sigma}} P_1^{\frac{\sigma-1}{\sigma}} - \xi_1(\bar{z}, w_0, w_1) = 0$$

$$\Rightarrow \quad \frac{\partial q_1(\bar{z}_i)}{\partial x_1} = -\frac{-\frac{\partial \xi_1}{\partial x_1}}{-\frac{1}{\sigma}\frac{\sigma-1}{\sigma} q_1(\bar{z}_i)^{-\frac{1}{\sigma}-1} Q_1^{\frac{1}{\sigma}} P_1^{\frac{\sigma-1}{\sigma}}}$$

$$\Rightarrow \quad \operatorname{sgn}\left(\frac{\partial q_1(\bar{z}_i)}{\partial x_1}\right) = -\operatorname{sgn}\left(\frac{\xi_1}{x_1}\right)$$

Results on sales follow by sales increasing in the output.

C.2 Profit maximization of horizontal MNEs (Proposition 4)

Foreign marginal costs of production $\xi_1(\bar{z}, q_0, w_0, q_1, w_1)$.

Symmetric knowledge levels. $\xi_1 = \frac{1}{1-e^{-\lambda \bar{z}}} \left(w_1(1 + c_1(\bar{z} - z_h)) + w_0(1 + c_0 z_h)\theta_{10} e^{-\lambda(\bar{z} - z_h)}\right)$.

$$\frac{\partial \xi_1}{\partial q_1} = \frac{1}{1-e^{-\lambda \bar{z}}}(-w_1 c_1 + \theta_{10} e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1 + c_0 z_h)))\frac{\partial z_h}{\partial q_1} \quad < 0$$

$$\frac{\partial \xi_1}{\partial q_0} = \frac{1}{1-e^{-\lambda \bar{z}}}(-w_1 c_1 + \theta_{10} e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1 + c_0 z_h)))\frac{\partial z_h}{\partial q_0} \quad > 0$$

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

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

$$> 0 \text{ if } w_1 c_1 < \frac{\theta_{10}}{\theta_{00}} w_0 c_0, \gtrless 0 \text{ otherwise}$$

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

$$> 0 \text{ if } w_1 c_1 < \frac{\theta_{10}}{\theta_{00}} w_0 c_0, \gtrless 0 \text{ otherwise}$$

Asymmetric knowledge levels. Two cases:

1. $\phi_1 = 0$, $\xi_1(\bar{z}, q_0, w_0, q_1, w_1) = \frac{1}{1-e^{-\lambda \bar{z}}} \left(w_1(1 + c_1 z_1) + \frac{1}{\lambda} w_1 c_1\right)$

$$\frac{\partial \xi_1}{\partial q_1} = \frac{1}{1-e^{-\lambda \bar{z}}} w_1 c_1 \frac{\partial z_1}{\partial q_1} \quad < 0$$

$$\frac{\partial \xi_1}{\partial q_0} = \frac{1}{1-e^{-\lambda \bar{z}}} w_1 c_1 \frac{\partial z_1}{\partial q_0} \quad > 0$$

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

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

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

2. $\phi_1 \neq 0$, $\xi_1 = \frac{1}{1-e^{-\lambda\bar{z}}}\left(w_1(1+c_1(\bar{z}-z_h)) + w_0(1+c_0 z_h)\theta_{10}e^{-\lambda(\bar{z}-z_h)}\right)$

$$
\begin{aligned}
\frac{\partial \xi_1}{\partial q_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}(-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\frac{\partial z_h}{\partial q_1} \quad < 0 \\
\frac{\partial \xi_1}{\partial q_0} &= \frac{1}{1-e^{-\lambda\bar{z}}}(-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\frac{\partial z_h}{\partial q_0} \quad > 0 \\
\frac{\partial \xi_1}{\partial \theta_{10}} &= \frac{1}{1-e^{-\lambda\bar{z}}}\Bigg((-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\frac{\partial z_h}{\partial \theta_{10}} \\
&\quad + w_0(1+c_0z_h)e^{-\lambda(\bar{z}-z_h)}\Bigg) \quad > 0 \\
\frac{\partial \xi_1}{\partial c_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}\left(w_1(\bar{z}-z_h) + \frac{\partial z_h}{\partial c_1}(-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\right) \quad \gtreqless 0 \\
\frac{\partial \xi_1}{\partial w_1} &= \frac{1}{1-e^{-\lambda\bar{z}}}\left(1+c_1(\bar{z}-z_h) + \frac{\partial z_h}{\partial w_1}(-w_1c_1 + \theta_{10}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\right) \quad \gtreqless 0
\end{aligned}
$$

Domestic marginal costs of production $\xi_0(\bar{z}, q_0, w_0, q_1, w_1)$.

$$
\begin{aligned}
\frac{\partial \xi_0}{\partial q_j} &= \frac{1}{1-e^{-\lambda\bar{z}}}(-w_0c_0 + \theta_{00}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0z_h)))\frac{\partial z_h}{\partial q_j} \\
&\quad > 0 \text{ if } j=1, < 0 \text{ if } j=0 \quad \text{for } \phi_0 \neq 0 \\
\frac{\partial \xi_0}{\partial q_j} &= \frac{1}{1-e^{-\lambda\bar{z}}}w_0c_0\frac{\partial z_0}{\partial q_j} \quad > 0 \text{ if } j=1, < 0 \text{ if } j=0 \quad \text{for } \phi_0 = 0
\end{aligned}
$$

Foreign output q_1 and sales p_1q_1. 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, I totally differentiate the system of first order conditions:

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

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

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

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

$$
\begin{aligned}
\operatorname{sgn}\left(\frac{dq_1}{d\theta_{10}}\right) &= -\operatorname{sgn}\left(\frac{d\xi_1}{d\theta_{10}}\right) \quad < 0 \\
&\quad \text{for } z_1 > \bar{z} - z_h
\end{aligned}
$$

$$\text{by } \frac{d\xi_0}{d\theta_{10}} = 0 \text{ and } \left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} < 0$$

$$\text{sgn}\left(\frac{dq_1}{d\theta_{10}}\right) = \text{sgn}\left(\frac{d\xi_1}{d\theta_{10}}\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} + \frac{d\xi_0}{d\theta_{10}}\right) < 0$$

$$\text{for } z_1 = \bar{z} - z_h,\ z_0 > \bar{z} - z_h, \text{ and } z_1 = z_0 = \bar{z} - z_h \text{ with } w_1c_1\theta_{00} > w_0c_0\theta_{10}$$

$$\text{by } \frac{d\xi_0}{d\theta_{10}} < 0 \text{ and } \left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} < 0$$

$$\text{sgn}\left(\frac{dq_1}{d\theta_{10}}\right) = \text{sgn}\left(\frac{d\xi_1}{d\theta_{10}}\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} + \frac{d\xi_0}{d\theta_{10}}\right) \gtrless 0$$

$$\text{for } z_1 = z_0 = \bar{z} - z_h \text{ with } w_1c_1\theta_{00} < w_0c_0\theta_{10}$$

$$\text{by } \frac{d\xi_0}{d\theta_{10}} > 0 \text{ and } \left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\frac{1}{\frac{d\xi_1}{dq_0}} < 0$$

$$\text{sgn}\left(\frac{dq_1}{dc_1}\right) = \text{sgn}\left(\frac{1}{1-e^{-\lambda\bar{z}}}w_1\left(z_1\left(-\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}} - \frac{d\xi_0}{dq_0}\right)\right.\right.$$
$$\left.\left. - c_1\frac{c_0}{\lambda(1+c_0z_h)}\frac{dz_h}{dc_1}\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_0^{-\frac{1}{\sigma}-1}Q_0^{\frac{1}{\sigma}}\right)\right) \gtrless 0 \quad \text{for } z_1 > \bar{z} - z_h$$

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

The effect of wages w_1 is analogous to the effect of c_1. Results on sales follow by sales increasing in the output.

C.3 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 (24), (25), (26) and (27). 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 $\bar{z}^X$.

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

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

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

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

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

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

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

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.

D Corporate Transferees

D.1 Data

Table D.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.

In addition, the data contains information on corporate transferees from the source country Indonesia in the host countries Hong Kong, Japan, Qatar, Saudi Arabia, Singapore and Taiwan, from the source country Mexico in the host countries Canada and Spain, from the source country Morocco in the host countries France and Spain, from the source country the Philippines in the host countries Canada, Hong Kong, Japan and Taiwan, from the source country Thailand in the host countries Japan and Taiwan, and from the source country Vietnam in in the host countries 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 D.1 displays the country pairs in the final dataset.

D.2 Descriptive statistics

Table D.2: Summary statistics, section 4

(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 official lang.	744	0.133	0.340	0	0	1
Common spoken lang.	744	0.192	0.267	0	0.043	1
Common native lang.	744	0.038	0.142	0	0	0.990
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 official lang.	316	0.111	0.314	0	0	1
Common spoken lang.	316	0.294	0.284	0	0.247	1
Common native lang.	316	0.036	0.135	0	0	0.990
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.

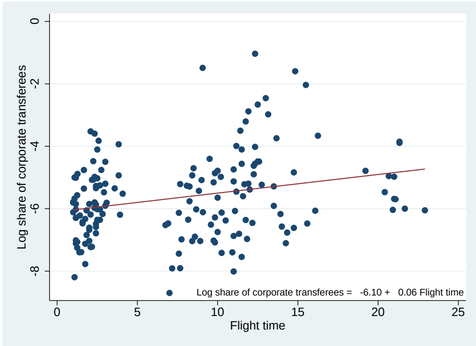
D.3 Additional regression results

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:

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

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.

Figure D.1: Scatter plot: share of corporate transferees vs. flight time



The figure scatters the log share of corporate transferees in total employment by MNEs against the flight time. It includes only uncensored observations.

Table D.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***	0.740***	0.761***	0.758***
	(0.075)	(0.078)	(0.077)	(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 official lang.			0.518°		
			(0.401)		
Common spoken lang.			−0.686+		0.235
			(0.404)		(0.281)
Common native lang.			1.131°		
			(0.712)		
Linguistic proximity			−0.049		−0.102+
			(0.060)		(0.057)
Log bandwidth (Mbit/s)				0.010	0.015
				(0.104)	(0.097)
Constant	−3.201***	−4.055***	−3.130***	−3.699*	−3.090+
	(0.851)	(0.948)	(0.874)	(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	0.192	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 D.1.

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 information on their investments to the central bank once their investments meet the reporting requirements. The reporting requirements vary across years. Until 2002, information 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, information 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 information 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 5

	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
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 official language	163,989	0.140	0.347	0.000	0.000	0.000	0.000	1.000
Common spoken language	163,989	0.431	0.276	0.006	0.219	0.389	0.612	0.993
Common native language	163,989	0.073	0.221	0.000	0.000	0.000	0.000	0.864
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
∅ effectively applied tariffs	116,186	0.536	2.740	0.000	0.000	0.000	0.260	2.648
Log costs of importing	89,213	6.961	0.352	6.190	6.817	7.046	7.144	7.482
Log costs of enforcing contracts	113,474	8.903	1.076	6.567	8.267	9.279	9.680	10.121
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. The number of observations varies due to differences in country coverage. Maximum possible number of observations: 164,604.

E.3 Selection correction in sector-level regressions

The dependent variable in the sector level regressions is the log of the ratio of aggregate affiliate sales to aggregate exports in an affiliate sector, country and year. The model predicts that

$$\begin{aligned}
\ln\left(\frac{\text{Affiliate sales}_{sjt}}{\text{Exports}_{sjt}}\right) &= \ln\left(\int_{\bar{z}^I}^{\bar{z}_{max}} Q\left(\frac{\sigma}{\sigma-1}\xi_j(\bar{z}, q_0(\bar{z}), w, q_1(\bar{z}), w)\right)^{1-\sigma} dG(\bar{z})\right)\\
&- \ln\left(\int_{\bar{z}^X}^{\bar{z}^I} Q\left(\frac{\sigma}{\sigma-1}\tau\xi_0(\bar{z}, w)\right)^{1-\sigma} dG(\bar{z})\right)\\
&= \ln Q + \ln\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} + \ln\left(\int_{\bar{z}^I}^{\bar{z}_{max}} \xi_j(\bar{z}, q_0(\bar{z}), w, q_1(\bar{z}), w)^{1-\sigma} dG(\bar{z})\right)\\
&- \ln Q - \ln\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} - \ln\left(\int_{\bar{z}^X}^{\bar{z}^I} (\tau\xi_0(\bar{z}, w))^{1-\sigma} dG(\bar{z})\right)\\
&= \ln\left(\int_{\bar{z}^I}^{\bar{z}_{max}} \xi_j(\bar{z}, q_0(\bar{z}), w, q_1(\bar{z}), w)^{1-\sigma} dG(\bar{z})\right)\\
&- \ln\left(\int_{\bar{z}^X}^{\bar{z}^I} (\tau\xi_0(\bar{z}, w))^{1-\sigma} dG(\bar{z})\right)\\
&\equiv V_j^I - V_j^X,
\end{aligned}$$

assuming symmetry of the foreign country j and the home country $j = 0$. The market size $\ln Q$ drops from the equation. It is straightforward to show that the equation can be generalized to several and potentially asymmetric countries. The foreign market size then influences the trade-off because the foreign marginal costs $\xi_1(\cdot)$ are a function of the foreign production quantity and thus of market size.

Helpman et al. (2008) show that selection bias arises if $V_j^X = 0$ because $\bar{z}^X > \bar{z}_{\max}$, i.e., because no firm has sufficient knowledge to export profitably. As there are positive exports in all sector-country-year pairs with a full set of covariates, such a selection problem does not arise. Analogously however, selection bias arises if $\bar{z}^I > \bar{z}_{\max}$, i.e., if no firm has sufficient knowledge to invest abroad profitably.

Following Helpman et al. (2008), it is possible to address this bias by estimating a selection equation for the self-selection of firms in sector s into FDI in country j in year t. I use the costs, time and number of procedures of starting a business and the existence of a bilateral investment treaty, denoted by the vector $\mathbf{Z}_{jt}$. I specify the selection equation

$$\begin{aligned}
\rho_{sjt} &= \Pr(\text{Affiliate sales}_{sjt} > 0|\text{observed variables})\\
&= \Phi(\gamma_0 + \gamma_1\theta_{j0t} + \gamma_2\tau_{j0t} + \gamma_3 Q_{jt} + \gamma_4 c_{jt} + \gamma_5 w_{jt} + \gamma_6\mathbf{X}_{jt} + \gamma_7\mathbf{Z}_{jt} + \alpha_s + \alpha_t)
\end{aligned}$$

and, as Helpman et al. (2008), include the following transformations of the predicted probability $\hat{\rho}_{sjt}$ in the regression equation (33):

$$\begin{aligned}
\hat{\eta}_{sjt} &= \frac{\phi(\Phi^{-1}(\hat{\rho}_{sjt}))}{\hat{\rho}_{sjt}}\\
\hat{\nu}^1_{sjt} &= \Phi^{-1}(\hat{\rho}_{sjt}) + \hat{\eta}_{sjt}\\
\hat{\nu}^2_{sjt} &= \left(\Phi^{-1}(\hat{\rho}_{sjt}) + \hat{\eta}_{sjt}\right)^2\\
\hat{\nu}^3_{sjt} &= \left(\Phi^{-1}(\hat{\rho}_{sjt}) + \hat{\eta}_{sjt}\right)^3.
\end{aligned}$$

Table E.2: Regression results, selection equation

$\Pr(\text{Affiliate sales}_{sjt} \geq 0)$	
Log distance	1.428***
	(0.333)
Effectively applied tariffs	0.002
	(0.003)
Office hours overlap	0.195***
	(0.057)
Log flighttime	–1.184+
	(0.486)
Common official language	1.838***
	(0.354)
Common spoken language	1.004
	(0.584)
Common native language	1.490*
	(0.594)
Linguistic proximity	–2.184*
	(1.067)
Log bandwidth	0.229**
	(0.077)
Log GDP	0.330***
	(0.102)
Log GDP per capita	–1.692***
	(0.178)
Average years of schooling	0.016
	(0.041)
Unit labor cost	–1.286
	(0.659)
Regulatory quality	2.458***
	(0.290)
Rule of law	0.327
	(0.309)
Government effectiveness	–0.347
	(0.211)
Corruption	–0.514*
	(0.250)
∃ investment treaty	0.251
	(0.229)
Log costs of starting a business	–0.204*
	(0.082)
Log time of starting a business	–0.107**
	(0.037)
# observations	3,154
Pseudo R-squared	0.384

Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Constant, year and sector dummies included. *Dependent variable:* indicator variable on positive aggregate affiliate sales a sector, foreign country and year. *Covariate definitions:* see Table 2.

F Sketch of a monitoring based model

A 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 a MNE is given by

$$\begin{aligned}
C(q_0, q_1) = &\min_{\{P_j, a_j^*\}_{j=0}^1} \sum_{j=0}^{1} n_j(w_j + \psi\theta_j P_j) + n_h \\
\text{s.t.} \quad & n_j a_j^* \geq q_j \qquad \forall j \\
& n_h \geq \sum_{j=0}^{1} n_j P_j \\
& w_j = \frac{1}{P_j} g(a_j^*) \\
& n_h \geq 0, P_j \in [0, 1] \quad \forall j \\
& n_j \geq 0, a_j^* \geq 0 \qquad \forall j
\end{aligned}$$

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 revealed in subsections 5.2 and 5.3.

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.