Knowledge, organization and taxation

An analysis of the driving forces behind multinational enterprises' investment patterns

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2013

vorgelegt von

Anna Gumpert

Referentin: Prof. Dr. Monika Schnitzer
Korreferent: Prof. Dr. Joachim Winter
Promotionsabschlussberatung: 14. Mai 2014

Datum der mündlichen Prüfung: 7. Mai 2014

Namen der Berichterstatter: Monika Schnitzer, Joachim Winter, Carsten Eckel

Danksagung

Mein Dank gilt an erster Stelle meiner Erstbetreuerin Monika Schnitzer, die in vielfältiger Weise zum Erfolg meines Promotionsvorhabens beigetragen hat. Sie hat mich seit dem Beginn meiner Tätigkeit als studentische Hilfskraft an ihrem Lehrstuhl stets fachlich gefördert und in meiner persönlichen Entwicklung unterstützt. Ihr Engagement hat mir zahlreiche Möglichkeiten eröffnet. Ihre Ermutigung hat maßgeblich dazu beigetragen, dass ich mich für eine wissenschaftliche Laufbahn entschieden habe. Ganz herzlich bedanke ich mich auch bei meinem Zweitbetreuer Joachim Winter, der mich bei meinem Promotionsvorhaben in wichtigen Fragen beraten und unterstützt hat. Carsten Eckel danke ich für zahlreiche wertvolle Hinweise und für sein Engagement als Drittgutachter.

Mein Dank gilt auch Pol Antràs, der mich zu einem einjährigen Forschungsaufenthalt an die Harvard University eingeladen und mir dadurch viele neue Perspektiven eröffnet hat. Ihm und Esteban Rossi-Hansberg danke ich zudem für ihre konstruktive Unterstützung meiner Forschung und intensive wissenschaftliche Diskussionen. Meinen Koautoren James Hines, Jr. und Manfred Antoni danke ich für eine sehr angenehme und produktive Zusammenarbeit.

Während meiner Zeit an der LMU habe ich es sehr genossen und sehr davon profitiert, mich mit meinen Kolleginnen und Kollegen Katharina Eck, Lisandra Flach, Inga Heiland, Iris Kesternich, Henrike Lindemann, Michal Masika, Tatjana Nabokin, Katrin Peters und Dominik Ruderer wissenschaftlich austauschen zu können. Ich danke ihnen sehr dafür, dass sie stets ein offenes Ohr für meine Anliegen hatten. Dies gilt in besonderer Weise für meine Kollegen Nadja Dwenger, Martina Engemann und Martin Watzinger, die mir bei schwierigen Problemen oft mit großem Engagement weiter geholfen haben.

Karin Fritsch und vielen fleißigen studentischen Hilfskräften danke ich dafür, dass sie mir durch ihre tatkräftige Unterstützung das Leben leichter gemacht haben.

Meine Dissertation wurde durch ein Promotionsstipendium der Deutschen Forschungsgemeinschaft (GRK 801) und von der Studienstiftung des deutschen Volkes vergebenes ERP-Stipendium finanziell gefördert. Dafür möchte ich mich herzlich bedanken. Der Bundesbank und dem Institut für Arbeitsmarkt- und Berufsforschung danke ich dafür, dass sie mir Zugang zu den dieser Dissertation zugrundeliegenden Mikrodaten gewährt haben.

Schließlich möchte ich mich bei meinem Vater Edmund Gumpert, meiner Familie und meinen Freunden herzlich für ihre vielfältige Unterstützung bedanken. Meinem Mann Erik Röder danke ich, dass er während der gesamten Promotionszeit an mich geglaubt hat und mir zur Seite stand.

Ich widme diese Arbeit meiner Mutter Maria Gumpert und meiner Großmutter Isabella Haaf.

Anna Gumpert, Mai 2014

Contents

Introduction

Multinational enterprises (MNEs) are crucial players for economic growth and prosperity. In 2012, for instance, the ten largest German MNEs alone generated total revenues of 922.3 billion Euros and employed 2.6 million people in—on average—46 countries (FAZ, 2013).[1] MNEs are not only big in absolute terms, they are typically also more productive than exporters or internationally inactive firms (Tomiura, 2007) and account for disproportionate shares of domestic output and employment. The mere 1% of U.S. manufacturing firms that are multinationals generate 34% of total U.S. output and employ 26% of the U.S. workforce (Bernard and Jensen, 2007). As these exemplary figures clearly illustrate, the economic significance of MNEs cannot be overestimated. Understanding the behavior of MNEs is thus essential to comprehending economic fluctuations and to drafting sound economic policies.

A vast literature analyzes the characteristics of MNEs, the incentive to conduct foreign direct investment (FDI) and the impact of activities of multinational firms on home and host countries (e.g. on the incentives for FDI Helpman, Melitz, and Yeaple, 2004; Helpman, 1984; Markusen, 1984; Chen and Moore, 2010; on investment promotion Harding and Javorcik, 2011; on the effects of FDI on host country technology adoption Javorcik, 2004; on the labor market effects of FDI Muendler and Becker, 2010). Still, some of the most basic aspects of MNEs' behavior are not well understood. Though MNE affiliate networks are commonly believed to span the globe, MNEs are essentially most likely to invest in host countries that are proximate to their home country, and their revenues tend to decrease in distance from the home country (e.g. Keller and Yeaple, 2013). This is the case even though serving proximate markets via exports is relatively cheap, so

[1] Companies are ranked according to their revenues.

common theories predict that exporting should be preferred to FDI (e.g. Brainard, 1997; Helpman, Melitz, and Yeaple, 2004), and many attractive investment destinations with low factor costs are located far from typical FDI source countries. Despite the surge of innovation in communication technologies and continual upgrading of transportation infrastructures, distance has thus remained a relevant factor for MNEs. Further, MNEs tend to pay higher wages than only nationally active firms or competitors that only export in both their domestic and foreign markets (Girma and Görg, 2007; Heyman, Sjöholm, and Tingvall, 2007; Hijzen, Martins, Schank, and Upward, 2013). It is unclear why that should be the case. After all, MNEs have the option to relocate expensive activities to countries with lower factor input costs. Most explanations of these multinational wage premiums so far are based on anecdotal evidence or informal arguments, and relate to factors such as international rent sharing agreements or compensation for restructurings after multinational take over (e.g. Girma and Görg, 2007). Finally, MNEs have repeatedly been exposed to sharp criticism as they increasingly channel their investments through offshore financial centers or tax havens (e.g. The Economist, 2012). Policy makers have raised severe concerns about the consequences and the legitimacy of this behavior (e.g. the G20 summit in 2013, the OECD, 2013's Action Plan on Base Erosion and Profit Shifting). After all, MNEs tend to be the largest players in the economy, so the resulting losses to public finances may be sizeable. However, the only available systematic evidence so far indicates, counterintuitively, that MNEs that face lower tax rates are more likely to invest in tax havens (Desai, Foley, and Hines, 2006). This seems inconsistent with tax haven investment motivated by tax avoidance.

This thesis contributes to the understanding of these aspects of MNEs' behavior. Chapters 1 and 2 propose and test a mechanism that rationalizes the importance of distance for MNEs' location decisions and their pattern of revenues as well as the emergence of multinational wage premiums.

Both chapters consider a set-up where knowledge and labor are required for production, building on Garicano (2000). Knowledge refers to the competencies required to take decisions in the production process, including knowledge about the firm's production and management technologies. Knowledge is only useful in production if a firm's employees learn it. Knowledge is thus not just a Hicks shifter in the production function that enables a firm to turn labor input more efficiently into output. Instead, a firm's employees have to spend time to acquire knowledge. Finally, knowledge can be communicated among employees, so employees can leverage their potentially different knowledge sets. Communication, however, is costly. Employees are remunerated for the time that they contribute to the production process and for the knowledge that they acquire.

Whereas previous papers restrict their analyses to firms that are active at exactly one location (e.g. Garicano, 2000; Caliendo and Rossi-Hansberg, 2012), firms are assumed to maintain production establishments at multiple locations in our work. They face a trade-off between costly local knowledge acquisition and costly communication across space. This basic trade-off is identical in Chapters 1 and 2, but the structure of the two models is distinct: the model elements in each chapter are tailored to the data that are used in the respective chapter to test model insights.

Chapter 1 develops and tests a stylized model of the organization of a multinational firm. The overall knowledge level of the firm is given exogenously. The firm employs managers in its headquarters in the home country and production workers both in the home country and in the foreign country. Countries vary with respect to the local wage level, the knowledge acquisition costs, the labor productivity and the bilateral communication costs with the home country. Both domestic and foreign workers can communicate with the managers. The firm may thus choose to have its knowledge acquired locally—at home or abroad—by production workers or centrally by managers.

In optimum, firms allocate a higher amount of knowledge to locations with lower knowledge acquisition costs and wages, higher labor productivity and higher bilateral communication costs between the production workers and the headquarters. This behavior implies that the marginal costs of production are increasing in the bilateral communication costs with the headquarters. Foreign sales and the probability of entry are correspondingly decreasing in the bilateral communication costs. As communication costs and distance are correlated, the mechanism proposed in Chapter 1 thus explains the importance of distance for MNEs' investment patterns in an equally simple and elegant way. The allocation of knowledge within the firm additionally affects wage payments in a way that is consistent with multinational wage premiums in the home and host market. The higher the bilateral communication costs, the more knowledge is allocated to the foreign country, so wages increase in the host country (consistent with Girma and Görg, 2007). Due to non-random selection of firms into FDI and restructuring of firms after FDI, domestic production workers earn higher wages in multinational than non-multinational firms (consistent with Heyman, Sjöholm, and Tingvall, 2007).

We test the insights on the relation between factor input cost and communication costs and the marginal costs of production, as well as additional predictions on the cut-off productivity levels required for foreign market entry using the Microdatabase Direct investment (MiDi), a firm-level database on German MNEs provided by the German central bank.[2] All predictions are strongly supported by the data.

Chapter 1 contributes to the literature by proposing a novel mechanism that rationalizes both the relevance of distance for MNEs' behavior and multinational wage premiums based on one simple insight. Most previous papers on gravity in FDI adopt the standard model by Helpman, Melitz, and Yeaple (2004) and are silent on implications of FDI for firms' wage setting behavior (e.g. Bahar, 2013; Keller and Yeaple, 2013). To our knowledge, Chapter 1 is further the fist paper that extends the framework by Garicano (2000) to a set-up with more than one establishment.

The second chapter studies the proposed model mechanism in greater detail and uses national data on multi-establishment firms to test the resulting hypotheses. The theory section endogenizes the overall knowledge of the firm that is taken as given in the first chapter. In addition, each of the production establishments may have a more complex structure with employees assigned to several hierarchical layers instead of just a group of production workers. The number of layers is endogenously determined.

The knowledge allocated to an establishment increases in the bilateral communication costs and decreases in local factor input costs for a given number of layers, as in Chapter 1. We additionally demonstrate that the increase in knowledge, and correspondingly the increase in employees' remuneration, is not linear: lower layers are assigned relatively less

[2] To facilitate readability, the pronoun "we" is used throughout this thesis to refer to the author or the authors of the respective chapter.

additional knowledge than higher layers. This implies that higher layers benefit relatively more in terms of higher remuneration. To test the model, we construct a novel data set that contains linked employer-employee data at the establishment level from the German Social Security records as well as information on the affiliation of establishments to firms. We use record-linkage techniques to combine the data sets. The available information is very reliable, as it is relevant for social security contributions. We construct variables that allow assigning employees to hierarchical layers in each establishment and measuring the average remuneration per layer as well as the span of control, i.e., the average number of lower layer employees that employees in a layer are responsible for.

We find that the average remuneration per layer increases in the bilateral communication costs between the establishments and the headquarters in the cross section, as predicted by the model. Chapter 2 thus contributes to the literature by furthering our understanding of the organization of multi-establishment firms. In addition, chapter 2 demonstrates that firm organization is an important element to understand employer wage premiums also in the national context. Differences in remuneration of observationally equivalent workers may be driven by the responsibilities that they are assigned within the company, and these vary systematically with the position of the employee within the firm.

The third chapter develops and tests a stylized model of the decision to invest in tax havens to explore whether tax avoidance is a relevant factor in MNEs' use of tax havens. The chapter analyzes the decision to invest in a tax haven for MNEs whose foreign income is exempt from domestic taxation, unlike earlier studies such as Desai, Foley, and Hines (2006). The chapter contributes to the literature by informing the debate on tax haven investment by MNEs as the incentives faced by the firms in our sample correspond to incentives faced by MNEs in most economies in the world.

We assume that MNEs face heterogeneous tax rates at their different foreign locations and have the option to engage in costly profit reallocation to a tax haven. Higher levels of taxation are conducive to tax haven investment, but fixed costs of investment in tax havens as well as marginal costs of profit shifting hamper tax haven investment. Further, the decision to invest in a tax haven has repercussions on the optimal investment decisions in non-haven countries. We test the model using the Microdatabase Direct investment mentioned above. We specify a linear probability model with observed tax haven investment as dependent variable and the average foreign non-haven tax rate as well as other firm characteristics as covariates. To take the effect of tax haven investment on optimal non-haven investments into account, we use a novel instrumentation strategy. We instrument the average foreign non-haven tax rate using the counterfactual average foreign non-haven tax rate had the firm refrained from any adjustments in its set of foreign non-haven destinations after the first period. We further control for heterogeneity between firms by employing firm fixed effects. Among German manufacturing firms, a one-percentage point higher foreign tax rate is associated with a 0.3 percentage point greater likelihood of owning a tax haven affiliate. This is consistent with tax avoidance incentives, and contrasts with earlier evidence for U.S. firms. Tax haven investment of firms in sectors with higher R&D intensity is more sensitive to higher taxation than the investment of firms in less R&D intensive sectors. The relationship seems less strong for firms in service industries, possibly reflecting the difficulty of reallocating taxable service income. The scope of profit reallocation and thus the incentive for tax haven investment varies with firm and sector characteristics.

Chapter 1

The organization of knowledge in multinational firms

1.1 Introduction

Firm organization has long been considered as one of the decisive determinants of productivity (see Bloom, Sadun, and Van Reenen, 2010, for a recent survey). Multinational enterprises are particularly challenged: they have to manage production processes that span different countries and involve employees from diverse economic and cultural contexts. Spatial and intercultural frictions often hamper communication and information flows within multinational firms (Ambos and Ambos, 2009; De Long and Fahey, 2000; UNCTAD, 2004, ch. IV A. 2).[1] Organizational design is thus of particular importance for multinational enterprises. It helps to ensure that a firm's knowledge is productively allocated within the firm, and that communication across the organization takes place smoothly.

The first chapter of this thesis analyzes the organization of knowledge within multinational firms and derives its consequences for foreign direct investment patterns. As in much of the previous literature (e.g. Irarrazabal, Moxnes, and Opromolla, 2013; Markusen

[1]The business economics and management literature discusses and evaluates various strategies to address the frictions in within-firm communication faced by internationally active firms (e.g. Foss and Pedersen, 2002; Ghemawat, 2007; Hansen and Løvås, 2004; Lagerström and Andersson, 2003).

and Maskus, 2001), knowledge is a non-rival factor within the firm. The chapter additionally assumes that knowledge can only be fruitfully employed in production if it is learned by a worker. This assumption captures the fact that technologies are typically only useful if someone knows how to utilize them. Based on a stylized model of the organizational structure of a multinational firm, the chapter determines the optimal allocation of knowledge within the firm given internationally heterogeneous market sizes, wages, labor productivities, knowledge acquisition costs and communication costs between the employees at the production locations and the headquarters. The resulting predictions on the within-firm distribution of productivities across countries and the cut-off productivity levels for investment in a country are consistent with the investment patterns of German multinational firms.

The chapter shows that the optimal amount of knowledge allocated by the firm to foreign employees in the host country increases with the communication costs between managers in the multinational's headquarters in the home country and the employees in the host country. Lower foreign labor productivities, higher foreign wages, higher knowledge acquisition costs and larger market size discourage the allotment of expertise. Correspondingly, the marginal costs of foreign production increase in the communication costs between the home country and the host country. At the same time, firms mitigate cost increases due to higher wages and knowledge acquisition costs or lower labor productivities by reallocating knowledge.

The implications of the model are consistent with the stylized facts of multinational firms' investments: both the probability of entry and MNEs' sales are predicted to be geographically concentrated and decreasing in distance, i.e., the investment patterns of MNEs are predicted to exhibit gravity. The concentration is higher in sectors with more complex production processes, as found in Bahar (2013). In line with empirical evidence in the literature (Girma and Görg, 2007; Heyman, Sjöholm, and Tingvall, 2007), the model also explains why the remuneration of employees of multinational firms is higher than the remuneration of employees of domestic firms both in the home and the host country, and that multinational wage premiums vary with distance between the domestic and the foreign country.

To evaluate the model systematically, the empirical analysis derives and tests predictions on within-firm differences in performance across different foreign countries and the selection of firms into foreign investment. Consistent with the theoretical hypotheses, German multinational firms tend to be relatively more productive in countries that are larger, that are characterized by lower bilateral communication costs with Germany, higher labor productivities, lower knowledge acquisition costs and lower wages. The findings are robust to the inclusion of additional determinants of foreign performance as well as alternative ways of measuring the model parameters. The cut-off productivity levels for entry are found to decrease in country size and increase in bilateral communication costs.

The results of the chapter have important implications for investment promotion purposes. The chapter recommends that investment promotion policies should not only focus on improving the investment climate inside a country, but also on ameliorating bilateral communication facilities with targeted source countries of FDI, for example by improving language training and communication infrastructures.

The chapter builds on two distinct strands of the literature. In developing the model of the optimal organizational structure, the chapter uses ingredients from the literature of firms as communication networks and knowledge hierarchies (e.g. Bolton and Dewatripont, 1994; Garicano, 2000; Radner, 1993). This literature studies how firms organize in response to differences in knowledge acquisition costs and in the cost of communication between agents, how hierarchical structures emerge, and how knowledge is efficiently allocated within a firm. Garicano (2000)'s knowledge hierarchies framework has been applied to various settings, and its main predictions have been confirmed empirically (e.g. Bloom, Garicano, Sadun, and Van Reenen, 2009; Garicano and Hubbard, 2009). Caliendo and Rossi-Hansberg (2012) and Caliendo, Monte, and Rossi-Hansberg (2012) demonstrate theoretically and empirically that the model framework is useful to understand the labor market consequences of international trade. To the best of our knowledge, this chapter is the first paper to apply the knowledge hierarchy framework to a context with several establishments.

Concerning the international investment aspects, the chapter draws on the large literature on multinational firms. Much of this literature has associated multinational activity predominantly with its advantages for a firm, such as savings of trade costs with the host market (e.g. Markusen, 1984; Helpman, Melitz, and Yeaple, 2004), savings of trade costs with third markets (e.g. Ekholm, Forslid, and Markusen, 2007; Tintelnot, 2012), savings of factor input costs (e.g. Helpman, 1984; Antràs, 2003; Antràs and Helpman, 2004), or a combination of these (e.g. Carr, Markusen, and Maskus, 2001; Grossman, Helpman, and Szeidl, 2006; Yeaple, 2003). In line with recent evidence on U.S. multinational firms indicating that most foreign affiliates are market-seeking, the theory section of this chapter assumes a horizontal motive of foreign activities (Ramondo, Rappoport, and Ruhl, 2013).

Within the literature on multinational firms, the strand most relevant to this chapter is concerned with the role of headquarter inputs for local affiliate production. Keller and Yeaple (2013) estimate the cost of transferring knowledge within multinational firms using data on U.S. manufacturing FDI. In their model, firms face a trade-off between the costs associated with disembodied and embodied knowledge transfer from their headquarters. Firms can either produce intermediaries locally but subject to efficiency losses due to communication frictions with the headquarters, or import intermediaries from home subject to transport costs. This chapter analyzes a complementary margin of multinational firms' behavior: it focuses on the choice between employing domestic expertise from the headquarters or hiring foreign expertise in the host market. The mechanism explains the gravity of foreign direct investment in sales and the probability of entry, as does the model in Keller and Yeaple (2013), but is applicable across sectors and also generates multinational wage premiums consistent with empirical evidence in the literature. Yeaple (2013) develops a model featuring managerial expertise as a scarce input within a multinational firm to explain empirical regularities in the expansion patterns of U.S. multinational multi-product firms. Similarly, Irarrazabal, Moxnes, and Opromolla (2013) assume that headquarter inputs, such as managerial oversight or marketing services, are required in the production process of affiliates and demonstrate that this assumption helps to explain the gravity of FDI in sales and the probability of entry. Defever (2012) assumes that smooth communication between foreign affiliates and the headquarters acts as a productivity shifter in a Cobb–Douglas production function and finds that geographical proximity to existing affiliates is a relevant consideration for the location decision of a multinational firm, above and beyond other locational characteristics. Bahar (2013) demonstrates that the probability of firms' conducting horizontal FDI decreases in their knowledge intensity, and that firms tend to prefer proximate destinations for the location of knowledge intensive FDI. He argues that iceberg type knowledge transfer costs that increase in the knowledge intensity of activities are appropriate for explaining this finding. All papers share their focus on the gravity of FDI and their theoretical approach, extending the basic framework in Helpman, Melitz, and Yeaple (2004) with productivity-shifting mechanisms. This chapter is distinct in modeling the organizational structure of multinational firms and contributes to the literature by providing a coherent rationale for both the gravity and multinational wage premiums.

The following section develops the theoretical model. It is divided into four subsections. The first subsection analyzes the cost-minimization problem at the firm level. The second subsection investigates the internationalization decision of the firm in a partial equilibrium framework. The third subsection outlines the general equilibrium implications. The last subsection discusses the robustness of the theoretical insights to specific assumptions. Section 1.3 discusses the empirical strategy and Section 1.4 describes the data used in the empirical analysis. Section 1.5 presents the empirical results. Section 1.6 examines how far the empirical results discriminate between the mechanism proposed in this chapter and alternative models that may generate the observed empirical patterns. The last section concludes.

1.2 Theory

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 that are each endowed with one unit of time.

There is an unbounded mass of potential firms in each country. To enter, firms have to pay a fixed entry cost of f^e units of domestic labor which is thereafter sunk. Upon entry, each entrant i draws a firm-specific knowledge level $\bar{z}_i$ from a known distribution $G(\bar{z})$ which is symmetric in the two countries. Mathematically, knowledge is an interval that ranges from zero to a firm-specific upper bound $\bar{Z}_i$.$\bar{z}_i$ denotes the length of a firm's knowledge interval $[0, \bar{Z}_i]$ (i.e. its Lebesgue measure). $\bar{z}_i$ captures the competencies required to produce a specific product and empirically corresponds to the state of a firm's technology.

The problem of each firm is to determine the set of countries it would like to serve and the set of production locations given its knowledge level $\bar{z}_i$. In addition, firms choose the cost-minimizing allocation of knowledge within the firm.

The following section first analyzes the cost-minimization problem for different sets of production locations, then determines the optimal set of countries and production locations, and finally investigates the entry decision.

1.2.1 Cost minimization

1.2.1.1 Assumptions

To simplify the exposition, subsection 1.2.1 focuses on the cost-minimization problem of a single firm located in the home country and characterized by the knowledge level $\bar{z}$. The firm consists of a headquarters in country $j = 0$ and production plants in country $j = 0$ and, potentially, in $j = 1$ as well. The headquarters are composed of a number n_h of managers, and the production plants are made up of a number n_j of production workers.

Production is assumed to be a problem solving process that is based on labor and knowledge (as in Garicano, 2000). The firm can employ labor for production in both countries $j = 0$ and $j = 1$. For each unit of labor employed in production, a mass $A_j \geq 1$ of problems is realized. Problems entail production possibilities. Transforming labor input into output requires that the problems be solved. A_j therefore corresponds to the labor productivity: higher values of A_j imply that more production possibilities, and thus potentially more output, can be generated with a given amount of labor. A_j varies across countries. The problems are distributed according to a problem probability distribution function from the exponential family:

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

where $z \in [0, \infty)$ refers to the domain of possible problems and $\lambda > 0$ denotes the problem arrival rate. Higher 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.

A firm is able to solve a problem if it is realized within the firm's knowledge interval. Consistent with the previous literature (e.g. Irarrazabal, Moxnes, and Opromolla, 2013), knowledge is assumed to be non-rival within the firm and can thus be used at all locations. Unlike previous papers, knowledge is only useful if it has been acquired by an employee. The underlying idea is that firm knowledge by itself is insufficient for production. Knowledge is only useful if there is someone who knows how to employ it.

Knowledge can either be acquired locally by production workers, or centrally by managers. Employees can communicate with each other and leverage the potentially different knowledge sets. Communication is costly. As is standard in the literature (e.g. Bolton and Dewatripont, 1994; Garicano, 2000; Bloom, Garicano, Sadun, and Van Reenen, 2009), communication costs are borne by the employee who is asked a question: the employee has to spend time listening to the problems. Garicano (2000) demonstrates that at optimum, only production workers spend their time in production and managers use their time solely for communication. Both production workers and managers are optimally characterized by knowledge levels that are uniform within each group and different between the two groups. The knowledge level of production workers covers the solution to the more frequent problems, whereas managers know the solution to problems that occur more rarely.

The production process thus works as follows. During each unit of time that they spend in production, the production workers scrutinize which of the problems are covered by their knowledge interval. Problems covered by the knowledge interval are solved immediately and output is produced. The production workers communicate all problems that are not covered by their knowledge interval to the managers. The managers spend a production-country specific amount θ_j, $1 > \theta_j > 0$ of time listening to the problems delegated by the production workers and solve all problems covered by their knowledge interval. It is generally assumed that $\theta_1 \geq \theta_0$ capture the fact that frictions in international communication may be more severe than frictions in communication within a country. Any problems that are not covered by the knowledge intervals of either the workers or the managers remain unsolved.

The output q_j of one unit of labor input can be calculated as the value of the cumulative distribution function:

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

n_j workers are correspondingly able to produce $q_j = n_j A_j(1 - e^{-\lambda \bar{z}})$ units of output. A higher value of $\bar{z}$ implies that more infrequent problems can be solved, so the resulting product is more sophisticated.

This theory of production is similar to Kremer (1993)'s O-ring theory of production. According to the O-ring theory of production, a single error in the production process implies that the product is wasted. In the present set-up, a single problem that the firm cannot solve implies that the production process cannot continue. As the firm is assumed to be able to distinguish between problems that it can solve and those it cannot, the product is not wasted. Instead, the firm finishes a less knowledge intensive product based on the knowledge it has available.

Products are heterogeneous with respect to their knowledge intensity, defined as the ratio of knowledge to labor input per unit of output. The knowledge intensity is given by

$$\frac{\bar{z}}{\frac{1}{A_j(1-e^{-\lambda \bar{z}})}} = \bar{z} A_j(1 - e^{-\lambda \bar{z}})$$

and is increasing in $\bar{z}$ and A_j. Intuitively, both the production of products by more knowledge-intensive firms and in countries with higher labor productivity involve lower labor inputs.

Knowledge acquisition is costly: employees have to hire teachers to teach them (as in Caliendo, Monte, and Rossi-Hansberg, 2012). Similar to the modelling of communication costs, teaching costs are fully borne by the teachers. TEachers have to spend $c_j x$ units of time to teach a knowledge interval of length x. All agents receive wage w_j per unit of time they spend working. Correspondingly, teachers receive remuneration $w_j c_j x$. The firm in turn has to remunerate its employees for the time they spend in production and for their knowledge acquisition expenses.

The problem of the firm is to design the optimal knowledge acquisition process. As outlined above, Garicano (2000) shows that only production workers spend their time in production and that they optimally learn the solutions to problems that occur more frequently in the production process. The knowledge interval of workers correspondingly starts at 0, where the mass of the problem density is highest, and ranges to a country specific upper bound Z_j, $j = 0, 1$. z_j denotes the length of the knowledge interval of production workers $[0, Z_j]$. The managers learn to solve infrequent problems. It is never optimal for the firm not to learn part of its knowledge interval $[0, \bar{Z}]$. More knowledge enables the firm to produce more output with a given amount of labor input and, as will be shown below, knowledge decreases marginal costs. The upper bound of managerial knowledge and the upper bound of the knowledge interval of the firm thus coincide. The knowledge interval of managers ranges from a lower bound Z_h to $\bar{Z}$. z_h denotes the length of this interval $[Z_h, \bar{Z}]$. The firm chooses the knowledge levels z_j and z_h as well as the number of production workers n_j and the number of 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 the managerial knowledge in the headquarters Z_h.

1.2.1.2 The optimization problem

The firm strives to minimize the costs of producing a given quantity q_j in country j. The costs are composed of the cost for personnel at the headquarters and at the different production locations. Each employee is remunerated with the country specific wage w_j for each unit of time spent working for the firm and for the knowledge acquisition costs $w_j c_j z_k$, $k = h, j$.

The firm optimally chooses the number of production workers $\{n_j\}_{j=0}^1$ employed at the different production locations, the country specific knowledge level of the production knowledge $\{z_j\}_{j=0}^1$, the number of managers employed in the headquarters n_h, and the knowledge level of those managers z_h.

$$C(q_0, q_1, w_0, w_1) = \min_{\{n_j, z_j\}_{j=0}^1, n_h, z_h} \sum_{j=0}^{1} n_j w_j (1 + c_j z_j) + n_h w_0 (1 + c_0 z_h) \tag{1.1}$$

$$\text{s.t.} \quad n_j A_j (1 - e^{-\lambda \bar{z}}) \geq q_j \quad \forall j \tag{1.2}$$

$$z_j \geq \bar{z} - z_h \quad \forall j \tag{1.3}$$

$$n_h \geq \sum_{j=0}^{1} n_j A_j \theta_j e^{-\lambda z_j} \tag{1.4}$$

$$n_h \geq 0, z_h \geq 0, z_h \leq \bar{z} \tag{1.5}$$

$$n_j \geq 0, z_j \geq 0, z_j \leq \bar{z} \quad \forall j \tag{1.6}$$

In the cost minimization problem, the production quantities $\{q_j\}_{j=0}^1$ are taken as given, but they will be endogenized in subsection 1.2.2, as will be the set of production locations of the firm. Wages $\{w_j\}_{j=0}^1$ will be endogenized in subsection 1.2.3. The problem arrival rate λ, communication costs $\{\theta_j\}_{j=0}^1$, labor productivities $\{A_j\}_{j=0}^1 > 0$ and knowledge acquisition costs $\{c_j\}_{j=0}^1 > 0$ are exogenous parameters determined by the state of a country's economic development and institutions.

In choosing the optimal values of $\{n_j\}_{j=0}^1$, n_h, $\{z_j\}_{j=0}^1$ and z_h the firm has to take four types of constraints into account:

Eq. (1.2): The firm has to produce a total output $n_j A_j (1 - e^{-\lambda \bar{z}})$ of at least q_j units.

Eq. (1.3): The knowledge intervals of the production workers $z_j = [0, Z_j]$ and the managers $z_h = [Z_h, \bar{Z}]$ have to jointly cover the full knowledge interval $\bar{z} = [0, \bar{Z}]$ of the firm.

Eq. (1.4): The firm has to hire a sufficient number of managers such that the managers are able to listen to all problems brought up to them. The amount of problems is calculated as the sum of the country specific masses of problems $n_j A_j$ times the probability that the solution is *not* found by the production workers in j $e^{-\lambda z_j}$. This term is multiplied by the communication costs θ_j to obtain the required number of managers.

Eq. (1.5, 1.6): All choice variables are restricted to be positive. Employees' knowledge cannot exceed the total knowledge of the firm.

Assumption 1. The exogenous parameters $\bar{z}, \lambda, c_j, A_j$ and θ_j, 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 restriction:

$$e^{\lambda \bar{z}} \leq \frac{\frac{q_0 w_0}{A_0}(\lambda + c_0) + \frac{q_1 w_1}{A_1}(\lambda + c_1)}{c_0\left(\frac{q_0 w_0 c_0}{A_0} + \frac{q_1 w_1 c_1}{A_1}\right)}(c_0 + \lambda(1 + c_0 \bar{z})) \tag{1.7}$$

The parameter restriction ensures that constraints (1.3) and (1.4) are innocuous in the sense that the firm is never forced to adopt a knowledge level $\bar{z}$ that exceeds the level it would adopt if it were free to choose its total knowledge level, and that the firm is never forced to hire managers even though it would prefer to produce a product based on a lower knowledge level allocated to the production workers. Given Assumption 1, the marginal costs of production are strictly decreasing in the overall knowledge level, so covering the full knowledge interval is always optimal for the firm. The parameter restriction is formally derived in Appendix A.1.1.1.

The Lagrangian equation is given by

$$\begin{aligned}\mathcal{L} &= \sum_{j=0}^{1} n_j w_j (1 + c_j z_j) + n_h w_0 (1 + c_0 z_h) \\ &+ \sum_{j=0}^{1} \xi_j \left[q_j - n_j A_j (1 - e^{-\lambda \bar{z}})\right] + \sum_{j=0}^{1} \phi_j \left[\bar{z} - z_h - z_j\right] + \kappa \left[\sum_{j=0}^{1} n_j A_j \theta_j e^{-\lambda z_j} - n_h\right].\end{aligned}$$

The Lagrangian multiplier ξ_j denotes the marginal costs of production. The other Lagrangian multipliers do not have intuitive interpretations.

The first order conditions are detailed in Appendix A.1.1.2. The optimal number of production workers is determined by the quantity constraint:

$$n_j = \frac{q_j}{A_j(1 - e^{-\lambda \bar{z}})}. \tag{1.8}$$

The optimal number of managers results from the constraint on the number of managers:

$$n_h = \sum_{j=0}^{1} n_j A_j \theta_j e^{-\lambda z_j} = \sum_{j=0}^{1} \frac{q_j \theta_j e^{-\lambda z_j}}{1 - e^{-\lambda \bar{z}}}. \tag{1.9}$$

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

$$z_j = \bar{z} - z_h. \tag{1.10}$$

Intuitively, if the knowledge constraint were non-binding for both countries, the overlap of knowledge at the headquarters and all production locations would remain unused. This cannot be optimal because any superfluous knowledge is saved in optimum.

If the knowledge constraint is non-binding, the knowledge level of the production workers is determined by

$$e^{-\lambda z_j} = \frac{w_j c_j}{A_j \theta_j \lambda w_0 (1 + c_0 z_h)}. \tag{1.11}$$

The knowledge constraint is binding whenever the marginal costs of additional local knowledge exceed the marginal benefit of using fewer services from the headquarters, i.e. whenever the firm would actually like to choose knowledge level $\frac{1}{\lambda} \ln \left[\frac{\lambda A_j \theta_j w_0 (1 + c_0 z_h)}{w_j c_j} \right] < \bar{z} - z_h$. From $\frac{1}{\lambda} \ln \left[\frac{\lambda A_{\bar{j}} \theta_{\bar{j}} w_0 (1 + c_0 z_h)}{w_{\bar{j}} c_{\bar{j}}} \right] < \bar{z} - z_h < \frac{1}{\lambda} \ln \left[\frac{\lambda A_j \theta_j w_0 (1 + c_0 z_h)}{w_j c_j} \right]$, it follows that the country characteristics of the country with binding constraint $\bar{j}$ and non-binding constraint j are related by

$$\theta_{\bar{j}} A_{\bar{j}} w_j c_j < \theta_j A_j w_{\bar{j}} c_{\bar{j}}.$$

The knowledge constraint is more likely to be binding in countries with ceteris paribus lower communication costs, lower labor productivity, higher wages and higher knowledge acquisition costs.

Managerial knowledge is implicitly determined by

$$\sum_{j=0}^{1} \Bigg[\mathbf{1}(z_j = \bar{z} - z_h) \lambda (1 + c_0 z_h) \left(q_j \theta_j e^{-\lambda (\bar{z} - z_h)} w_0 (c_0 + \lambda (1 + c_0 z_h)) - \frac{q_j w_j c_j}{A_j} \right) \\ + \mathbf{1}(z_j > \bar{z} - z_h) c_0 \frac{q_j w_j c_j}{A_j} \Bigg] = 0. \tag{1.12}$$

The indicator function $\mathbf{1}(\cdot)$ determines whether the constraint $z_j = \bar{z} - z_h$ is binding or not.

If the firm only produces in the domestic country, z_0, n_0 and n_h are determined by the constraints (1.2)-(1.4). Managerial knowledge is implicitly defined by the simpler condition

$$\theta_0 e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1 + c_0 z_h)) - \frac{c_0}{A_0} = 0. \tag{1.13}$$

The marginal costs of production are given by

$$\xi_j = \frac{1}{A_j(1-e^{-\lambda\bar{z}})}\left[w_j(1+c_j z_j) + \frac{1}{\lambda}w_j c_j\right] \qquad \text{for } z_j > \bar{z} - z_h; \tag{1.14}$$

$$= \frac{1}{A_j(1-e^{-\lambda\bar{z}})}\left[w_j(1+c_j(\bar{z}-z_h)) + w_0(1+c_0 z_h)A_j\theta_j e^{-\lambda(\bar{z}-z_h)}\right] \quad \text{for } z_j = \bar{z} - z_h. \tag{1.15}$$

As $\xi_j > 0 \quad \forall q_j$, the cost function is strictly increasing in q_j.

1.2.1.3 Comparative statics

Proposition 1. *The optimal choices of the firm vary with country and firm characteristics.*

1. *The number of the production workers n_j is increasing in the production quantity q_j and decreasing in labor productivity A_j and the firm's knowledge level $\bar{z}$.*
2. *The knowledge level of production workers z_j is increasing in communication costs θ_j and in labor productivity A_j, and decreasing in wages w_j, knowledge acquisition costs c_j and the production quantity q_j. If the firm is active only domestically, the workers' knowledge level z_0 is independent of wages w_0 and production quantity q_0.*
3. *The number of managers n_h is increasing in the production quantity q_j, the knowledge acquisition costs c_j and wages w_j. It is decreasing in labor productivity A_j and the firm's knowledge level $\bar{z}$. The effect of communication costs is ambiguous.*

4. *Managerial knowledge z_h increases in firm knowledge $\bar{z}$. The effect of the country characteristics depends on whether the knowledge constraint is binding or not. If the knowledge constraint is binding, managerial knowledge decreases in labor productivity A_j and communication costs θ_j, and increases in the production quantity q_j, knowledge acquisition costs c_j and wages w_j. If the knowledge constraint is non-binding, country characteristics have opposite effects on managerial knowledge, except for communication costs. Managerial knowledge is independent of communication costs θ_j if the knowledge constraint is non-binding.*

Proof. See Appendix A.1.1.3. □

Paragraph 1 is intuitive. Concerning the knowledge level of the production workers, higher communication costs entail an incentive to assign workers more knowledge to increase the amount of problems that do not need to be communicated to the headquarters and thus save on communication costs. Higher wages w_j and higher learning costs c_j increase the remuneration for every single worker, so it is optimal to decrease their knowledge to mitigate cost increases. If labor productivity is lower, or a larger quantity q_j is to be produced, more workers need to be hired, each of whom receives $w_j(1 + c_j z_j)$. It is optimal to decrease their knowledge level to decrease total labor costs.

Wages and the production quantity do not affect the production workers' optimal knowledge level for domestic firms. The production quantity scales the "breadth" of the organization. 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. Changing the allocation of knowledge within the firm does not affect this requirement. Similarly, wages scale the total costs of production. The effect of knowledge acquisition costs and communication costs is different. The firm faces a trade-off: allocating more (less) knowledge to the production workers increases (decreases) the total costs at the production level, but decreases (increases) the costs that accrue due to communication between production workers and managers. The optimal amount of knowledge allocated to the production workers therefore decreases in the knowledge acquisition costs c_0 and increases in the communication costs with the headquarters θ_0 also in a firm that produces only domestically. All the results for domestic firms are in line with the findings for single-establishment firms in Bloom, Garicano, Sadun, and Van Reenen (2009).

The effect of the bilateral communication costs on the knowledge level of production workers is interesting in view of the empirical findings on the labor market implications of foreign direct investment. Girma and Görg (2007) analyze establishment level data from the UK and find that despite controlling for establishment characteristics, establishments acquired by U.S. firms pay higher post-acquisition wages than comparable domestic establishments or establishments acquired by investors from Europe. They discuss different potential reasons for this finding, including a more efficient use of firm specific assets by U.S. firms, international rent-sharing agreements within MNEs, and non-random selection of workers to firms. The present thesis offers an additional explanation. The difference in post-acquisition wages may be driven by U.S. multinational firms' allocating more competencies to the U.K. Bilateral communication costs between the U.K. and the U.S. may be higher than between the U.K. and European firms even though English is officially spoken in both countries, for example as it takes longer for management to reach the U.K. from the U.S. or due to time zone differences.

The effect of the production quantity and labor productivity on the number of managers is straightforward. The number of managers increases in wages and knowledge acquisition costs because these lead to a decrease in production workers' knowledge. This implies that the number of problems sent to headquarters increases, so more managers have to be hired to deal with them. Communication costs have an ambiguous effect on the optimal number of managers. On the one hand, an increase in communication costs implies that managers have to spend more time to accomodate a given number of problems, so the number of managers increases. On the other hand, higher communication costs lead to higher knowledge levels at the production locations. The number of problems referred to the headquarters decreases, as does the number of managers.

If the knowledge constraint is binding, country characteristics have exactly the opposite effect on managerial knowledge as they have on production knowledge, because the sum of the two knowledge levels has to cover the full knowledge interval. Country characteristics (except communication costs) have an identical effect on managerial knowledge as on production knowledge if the knowledge constraint is not binding. This result may seem counterintuitive at first. Take the effect of wages as an example. Higher wages imply that the firm decreases the level of knowledge of production workers. Consequently, the number of problems sent to headquarters increases. This entails an incentive to decrease the marginal costs of using the headquarters, which is achieved by decreasing heaquarter knowledge. This is possible, as the knowledge constraint is not binding. The effect of the other characteristics can be derived analogously.

1.2.2 Profit maximization

There are many firms in each country $j = 0$ and $j = 1$. Each firm i is characterized by a firm-specific knowledge level $\bar{z}_i$ and solves the cost minimization problem analyzed in the previous subsection. The analysis is conducted for a firm in the domestic country $j = 0$, and the analogous results apply to firms in the foreign country $j = 1$.

Consumers in country j have CES preferences:

$$U(x_j(\bar{z})) = \left(\int_{\zeta_j} x_j(\bar{z}_i)^{\frac{\sigma-1}{\sigma}} d\bar{z} \right)^{\frac{\sigma}{\sigma-1}} \tag{1.16}$$

where ζ_j is the set of varieties available in country j, $\sigma > 1$ is the elasticity of substitution and $x_j(\bar{z}_i)$ is the individual consumption level of the variety produced with knowledge input $\bar{z}_i$ in country j. Preferences are assumed to be symmetric across countries. The set of varieties ζ_j is determined by the set of firms that are active in the economy, which is endogenized in the next subsection.

Utility maximization subject to the individual's budget constraint implies that the demand function for product i is given by

$$p(\bar{z}_i) = q_j(\bar{z}_i)^{-\frac{1}{\sigma}} Q_j^{\frac{1}{\sigma}} P_j^{\frac{\sigma-1}{\sigma}} \tag{1.17}$$

where Q_j is the consumption basket in country j and P_j denotes the price index. We normalize the price index to 1.

The total demand is given by the product of the number of customers N_j and the individual demands:

$$q_j(\bar{z}_i) = N_j x_j(\bar{z}_i). \tag{1.18}$$

Firms maximize profits. Each firm chooses whether and how to serve the foreign country, and determines the profit maximizing quantities q_0 and q_1. Firms can access the foreign country via exporting or foreign direct investment. They incur fixed costs f_X for the former and f_I for the latter option, where $f_I > f_X$.

In the following, optimal quantities are characterized by the mode superscripts D for domestic firms, X for exporters, and I for multinational firms.

The profit maximization problem of a multinationally active firm is given by

$$\max_{q_0^I, q_1^I} \quad \pi^I(\bar{z}_i, w_0, w_1) = \sum_{j=0}^{1} p_j(q_j^I(\bar{z}_i))q_j^I(\bar{z}_i) - C(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1). \tag{1.19}$$

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

$$p_j(\bar{z}_i) = \frac{\sigma}{\sigma - 1}\xi_j(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1). \tag{1.20}$$

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

$$q_j^I(\bar{z}_i) = Q_j \left(\frac{\sigma}{\sigma - 1}\xi_j(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1)\right)^{-\sigma}. \tag{1.21}$$

An exporting firm solves an analogously defined profit maximization problem:

$$\max_{q_0^X, q_1^X} \quad \pi^X(\bar{z}_i, w_0, w_1) = \sum_{j=0}^{1} p_j(q_j^X(\bar{z}_i))q_j^X(\bar{z}_i) - C(\bar{z}_i, q_0^X(\bar{z}_i), q_1^X(\bar{z}_i), w_0, \tau_1). \tag{1.22}$$

Optimal prices are a constant mark-up over marginal costs including transport costs τ_1 if applicable. The marginal costs are constant. The optimal quantities are given by

$$q_0^X(\bar{z}_i) = Q_0 \left(\frac{\sigma}{\sigma - 1}\xi_0(\bar{z}_i, w_0)\right)^{-\sigma}; \qquad q_1^X(\bar{z}_i) = Q_1 \left(\frac{\sigma}{\sigma - 1}\tau_1\xi_0(\bar{z}_i, w_0)\right)^{-\sigma}. \tag{1.23}$$

The optimal production quantity of a purely domestically active firm is determined by similar considerations. As the marginal costs are constant in $q = \{q_0^D(\bar{z}_i), q_0^X(\bar{z}_i) + \tau_1 q_1^X(\bar{z}_i)\}$ if the firm only produces domestically,

$$q_0^D(\bar{z}_i) = q_0^X(\bar{z}_i). \tag{1.24}$$

Firms can either be active purely domestically, produce domestically and export to the foreign country or produce both at home and abroad. The profits of these three options are given by

$$\pi_0^D(\bar{z}_i) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q_0 \xi_0(\bar{z}_i, w_0)^{1-\sigma}; \tag{1.25}$$

$$\pi_0^X(\bar{z}_i) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left(Q_0 + Q_1\tau_1^{1-\sigma}\right) \xi_0(\bar{z}_i, w_0)^{1-\sigma}; \tag{1.26}$$

$$\begin{aligned}\pi_0^I(\bar{z}_i) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} &\Big(Q_0 \xi_0(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1)^{1-\sigma} \\ &+ Q_1 \xi_1(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1)^{1-\sigma}\Big). \end{aligned} \tag{1.27}$$

Firms start exporting if

$$p_1(q_1^X(\bar{z}_i)) q_1^X(\bar{z}_i) - \tau_1 \xi_0(\bar{z}_i, w_0) q_1^X(\bar{z}_i) \geq w_0 f^X \tag{1.28}$$

and become multinationals if

$$\pi^I(\bar{z}_i, w_0, w_1) - w_0 f^I \geq \pi^X(\bar{z}_i, w_0, w_1) - w_0 f^X. \tag{1.29}$$

The exporting decision only depends on foreign profits because the firm is able to produce additional output without adjusting its organizational structure. Domestic profits are unaffected by the exporting decision. Foreign investment entails reallocations to the optimal allotment of knowledge inside the firm because the firm has to balance domestic and foreign requirements. This implies that domestic performance is affected by the FDI decision, so total profits—domestic and foreign profits—in case of FDI have to exceed the total profits in case of exporting. Unlike conventional models of FDI, the model thus features an interdependence between the marginal costs of production and the profits across countries.

1.2.2.1 Firm performance across different modes

The quantities sold in the domestic country in case of exporting and only domestic activity are equal due to the constancy of the marginal costs with a single production location. Quantities sold domestically by a multinational firm are lower than domestically sold quantities if the firm produced only domestically:

$$q_0^D(\bar{z}_i) = q_0^X(\bar{z}_i) \geq q_0^I(\bar{z}_i). \tag{1.30}$$

This result arises because a multinational firm cannot tailor its headquarters to domestic needs (unless the knowledge constraint is binding at both locations). Instead, the knowledge level chosen at the headquarters balances domestic and foreign requirements. Correspondingly, domestic marginal costs increase in the quantity of foreign production and domestic output, and profits are lower in the case of FDI than in the case of exporting or domestic activity.

In the foreign country, quantities sold in the case of FDI exceed export quantities.

$$q_1^I(\bar{z}_i) > q_1^X(\bar{z}_i) \tag{1.31}$$

Otherwise, the difference in fixed costs and the sales foregone in the domestic country would not be worthwhile. FDI therefore only takes place if $\xi_1(\bar{z}_i, q_0^I, q_1^I, w_0, w_1) < \tau_1\xi_0(\bar{z}_i, w_0)$. This is more likely to be the case for high τ_1 and low θ_1, low w_1, low c_1, high A_1 and a large market size (see Proposition 2 below).

1.2.2.2 Firm performance across different countries

An exporting firm sells larger quantities in the domestic country than in the foreign country by $\tau_1 > 1$ and $\sigma > 1$.

$$q_0^X(\bar{z}_i) > q_1^X(\bar{z}_i) = \tau_1^{-\sigma} q_0^X(\bar{z}_i)$$

Foreign sales decrease in transport costs, so country attractiveness as export destination decreases with τ_1.

A multinational firm's foreign performance varies with country characteristics.

Proposition 2. *Foreign marginal costs ξ_1 are increasing in the communication costs with the headquarters θ_1, with foreign wages w_1 and foreign learning costs c_1. Foreign marginal costs ξ_1 are decreasing in foreign labor productivity A_1 and the production quantity q_1.*

Proof. Proof see Appendix A.1.2. □

It is important to note that the marginal costs of production ξ_1 are decreasing in the foreign production quantity q_1, but increasing in the domestic production quantity q_0 (see Appendix A.1.2). Intuitively, the firm chooses the optimal allocation of knowledge in a way that benefits affiliates with larger output. Equation (1.12) foreshadows this implication: the firm takes the size of its output at the different affiliates into account in choosing the optimal headquarter knowledge level z_h.

Proposition 2 implies that foreign output and foreign sales are higher in countries with lower communication costs θ_1, lower wages w_1 and lower knowledge acquisition costs c_1 as well as higher labor productivity A_1 (see Appendix A.1.2). The effects of the communication costs, wages, the knowledge acquisition costs and the labor productivity work through two channels. Take the communication costs as an example. Higher communication costs have a direct, positive effect on the marginal costs of production, which exerts a negative effect on foreign output and foreign sales. Higher communiation costs also have an indirect, positive effect on the marginal costs of production as the firm favors affiliates with larger output. Due to the lower production quantity, the affiliate is relatively less important for the firm. The firm adjusts the allocation of knowledge. A lower production quantity due to higher communication costs thus implies that the marginal costs of production increase even further, depressing foreign output and foreign sales. In summary, the foreign country is more attractive for FDI if the communication costs with the home country, wages and the knowledge acquisition costs are lower and the labor productivity is higher due to the positive effect of these variables on foreign output and sales.

1.2.3 General equilibrium

1.2.3.1 Closed economy

Firms draw a firm-specific knowledge level $\bar{z}_i$ upon paying the fixed entry cost f^e in units of domestic labor. The distribution of potential knowledge levels $G(\bar{z})$ is known and defined on the interval $[0, \bar{z}_{max}]$, where $\bar{z}_{max}$ is the highest knowledge level that fulfils Assumption 1. Production additionally entails a fixed cost of production f paid in domestic labor. Firms with adverse knowledge draws cannot profitably cover these fixed costs, and so exit immediately.

The general equilibrium conditions determine the cut-off knowledge level for entry, $\bar{z}^*$, the number of firms M, wage w and total income Q. Recall that the price index P is normalized to unity. The parameters λ, A, c, θ and N are exogenous. Country subscripts are suppressed for simplicity of exposition.

A single firm's variable profits are given by

$$\pi^D(\bar{z}_i) = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\xi(\bar{z}_i, w)^{1-\sigma}.$$

The marginal costs of production ξ are constant in the production quantity q and strictly decreasing in the knowledge level $\bar{z}_i$ (for a proof, see Appendix A.1.3.1). Profits are thus strictly increasing in the knowledge level $\bar{z}_i$.

The least productive active firm is indifferent between entering and remaining inactive: its variable profits are equal to the fixed costs of production. The *zero cut-off profit* *condition* determines the knowledge level $\bar{z}^*$ of the marginal entrant.

$$wf = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\xi(\bar{z}_i^*, w)^{1-\sigma} \tag{1.32}$$

As profits are strictly increasing in knowledge, the cut-off is unique. The cut-off increases with the knowledge acquistion costs c, wages w and communication costs θ and decreases in labor productivity A (for a proof, see Appendix A.1.3.1).

Potential entrants enter up to the point where the expected net value of entry is zero: entry occurs until the expected value of profits is equal to the sunk cost of entry. The *free entry condition* determines wages w:

$$wf^e = \int_{\bar{z}^*}^{\bar{z}_{max}} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q\xi(\bar{z}, w)^{1-\sigma} dG(\bar{z}) \tag{1.33}$$

A unique equilibrium exists as the zero cut-off profit condition is strictly increasing in $\bar{z}^*$ and the free entry condition is strictly decreasing in $\bar{z}^*$.

In equilibrium, labor and goods markets clear. Labor supply N equals labor demand. Labor is used to cover the fixed cost of entry, the fixed costs of investment, and the demands for labor in production, for management, and for teaching. Labor demand for production, management and teaching can be calculated by setting wages equal to 1 in the cost function $C(q, w)$. The *labor market clearing condition* determines the number of firms M.

$$N = M\left(f^e + \frac{1}{1-G(\bar{z}^*)}\int_{\bar{z}^*}^{\bar{z}_{max}} (f + C(q,1))dG(\bar{z})\right) \tag{1.34}$$

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

$$wN = Q \tag{1.35}$$

1.2.3.2 Open economy

Firms can either export or conduct foreign direct investment. Both activities entail fixed costs f^X and f^I to be paid in units of domestic labor, with $f^I > f^X > f$.

The general equilibrium conditions determine the cut-off knowledge level for activity $\bar{z}^*$, the cut-off knowledge level for exporting $\bar{z}^X$ and for FDI $\bar{z}^I$, the number of firms $\{M_j\}_{l=0}^1$, wages $\{w_j\}_{l=0}^1$ and total income $\{Q_j\}_{l=0}^1$.

As above, the least productive active firm is indifferent between entering and remaining inactive. The variable domestic profit is equal to the fixed costs of production. The first *zero cut-off profit condition* determines the knowledge level $\bar{z}^*$ of the marginal entrant.

$$w_0 f = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q_0 \xi_0(\bar{z}_i^*, w_0)^{1-\sigma} \tag{1.36}$$

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 $\bar{Z}^X$.

$$w_0 f^X = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q_1 (\tau_1 \xi_0(\bar{z}_i^X, w_0))^{1-\sigma} \tag{1.37}$$

Domestic and export profits are strictly increasing in $\bar{z}_i$ as the marginal costs ξ_0 are strictly decreasing in $\bar{z}_i$. The marginal costs of the marginal exporter are lower than the marginal costs of the marginal entrant if the market sizes are similar, as in Melitz (2003):

$$\xi_0(\bar{z}_i^X, w_0) = \left(\frac{Q_1 f}{Q_0 f^X}\right)^{\frac{1}{\sigma-1}} \frac{1}{\tau_1}\xi_0(\bar{z}_i^*, w_0) \quad < \xi_0(\bar{z}_i^*, w_0) \quad \text{for } Q_0 \approx Q_1.$$

Consequently, the exporting cut-off knowledge level is higher than the zero cut-off knowledge level: $\bar{z}^X > \bar{z}^*$. An increase in τ_1 or a decrease in Q_1 imply that the exporting cut-off knowledge level $\bar{z}^X$ increases: it is profitable to export to more distant destinations and to smaller markets only for firms with lower marginal costs.

The marginal multinational firm is indifferent between exporting and foreign direct investment. The net total export profits 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.*

$$\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\xi_0(\bar{z}^I,w_0)^{1-\sigma}(Q_0+Q_1\tau_1^{1-\sigma})-w_0f^X=$$
$$\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\Big(Q_0\xi_0(\bar{z}^I,q_0^I(\bar{z}^I),q_1^I(\bar{z}^I),w_0,w_1)^{1-\sigma}+$$
$$Q_1\xi_1(\bar{z}^I,q_0^I(\bar{z}^I),q_1^I(\bar{z}^I),w_0,w_1)^{1-\sigma}\Big)-w_0f^I \tag{1.38}$$

The total profits from FDI have to exceed the total profits from exporting plus the difference in the fixed costs. Both total FDI and total export profits increase in the knowledge level $\bar{z}$. As the fixed costs of foreign direct investment are higher than the fixed costs of exporting, firms have to have a higher knowledge level to profitably carry out foreign investment: $\bar{z}^I > \bar{z}^X$. Domestic profits decrease in 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 productivity cut-off is thus shifted upwards. Export profits decrease with τ_1 and profits from FDI decrease with w_1, c_1 and θ_1, and increase with A_1. Consequently, the productivity cut-off $\bar{z}^I$ varies with the characteristics of the host country.

Proposition 3. *The productivity cut-off $\bar{z}^I$ increases in the bilateral communication costs θ_1 between the host country and the multinational's home country, in the foreign wages w_1 and the foreign knowledge acquisition costs c_1, and decreases in the foreign labor productivity A_1 as well as the bilateral transportation costs τ_1 between the home and the host country.*

Proof. See Appendix A.1.3.2. □

The depressing effect of the bilateral communication costs θ_1 on the tendency to invest decreases in λ. A higher value of λ is associated with a more predictable production process. The productivity cut-off $\bar{z}^I$ is thus even higher in sectors with less predictable, more complex production processes. This result is consistent with the findings in Bahar (2013) who shows that firms in knowledge intensive sectors are less likely to conduct foreign direct investment, and that given investment, high communication costs have a negative effect on the location probability of knowledge intensive activities.

Firms enter up to the point where the net value of entry is zero. The expected value of the profits has to be adjusted for export and multinational profits. The *free entry condition* is given by

$$\begin{aligned} w_0 f^e = & \int_{\bar{z}^*}^{\bar{z}^I} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q_0 \xi_0(\bar{z}, w_0)^{1-\sigma} dG(\bar{z}) \\ & + \int_{\bar{z}^X}^{\bar{z}^I} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} Q_1(\tau_1 \xi_0(\bar{z}, w_0))^{1-\sigma} dG(\bar{z}) \\ & + \int_{\bar{z}^I}^{\bar{z}_{max}} \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \Big(Q_0 \xi_0(\bar{z}, q_0(\bar{z}), q_1(\bar{z}), w_0, w_1)^{1-\sigma} \\ & \quad + Q_1 \xi_1(\bar{z}, q_0(\bar{z}), q_1(\bar{z}), w_0, w_1)^{1-\sigma}\Big) dG(\bar{z}) \end{aligned} \tag{1.39}$$

The *labor market clearing condition* is modified to include the demand for labor to cover the fixed costs of foreign activity and the demand for labor by foreign investors.

$$\begin{aligned} N_0 = & M_0 \left(f^e + \frac{1}{1-G(\bar{z}_0^*)} \left(\int_{\bar{z}_0^*}^{\bar{z}_{0,max}} f + C_0^0(q_0, q_1, 1, w_1) dG(\bar{z}) \right.\right. \\ & \left.\left. + \int_{\bar{z}_0^X}^{\bar{z}_0^I} f^X dG(\bar{z}) + \int_{\bar{z}_0^I}^{\bar{z}_{0,max}} f^I dG(\bar{z}) \right) \right) \\ & + M_1 \frac{1}{1-G(\bar{z}_1^*)} \int_{\bar{z}_1^I}^{\bar{z}_{1,max}} C_1^0(q_1, q_0, w_1, 1) dG(\bar{z}) \end{aligned} \tag{1.40}$$

where $C_0^0(q_0, q_1, 1, w_1)$ refers to the fraction of the costs of MNEs from country $j = 0$ that accrue in their home market and $C_1^0(q_1, q_0, w_1, 1)$ denotes the fraction of the costs of foreign MNEs that accrue in country $j = 0$. That is, $C_0^0(q_0, q_1, 1, w_1)$ counts the number of domestic employees of domestic MNEs, $C_0^0(q_0, q_1, 1, w_1) = n_0^*(1 + c_0 z_0^*) + n_h^*(1 + c_0 z_h^*)$ (with the * referring to optimal choices as derived in subsection 1.2.1). $C_1^0(q_1, q_0, w_1, 1)$ analogously counts the number of employees in country $j = 0$ of foreign MNEs.

The *goods market clearing condition* corresponds to the above equation, with the consumption basket adjusted for the available foreign goods.

1.2.4 Robustness

The model assumes that all managers have the same knowledge level, so every manager can be approached with any problem from any location. Large multinational corporations may afford specialized divisions at their headquarters that are responsible for certain countries only and that possess only specialized knowledge.

It is possible to accomodate this consideration in the model by allowing the firm to choose eight endogenous variables: $\{z_j, n_j, z_h^j, n_h^j\}_{l=0}^1$, where n_h^j is the number of managers responsible for production in country j and z_h^j is their knowledge level. A formal analysis is provided in Appendix A.1.4.

Such a modification does not affect the main results: as in Proposition 1, production knowledge increases in the communication costs with the headquarters θ_j and labor productivity A_j and decreases in the knowledge acquisition costs c_j. Domestic production knowledge is independent of wages w_0, and foreign production knowledge decreases in wages w_1. Unlike the above, the knowledge levels are independent of the production quantity. The marginal costs increase in θ_j, w_j and c_j, and decrease in A_j.

If the headquarter divisions are fully specialized, the marginal costs of production are constant and independent across countries, in line with the assumptions made in previous papers (e.g. Helpman, Melitz, and Yeaple, 2004; Tintelnot, 2012). Multinational firms probably have specialized departments in their headquarters. Still, upper management is responsible for worldwide operations. At least at some level of seniority, managers thus have to be able to address issues brought up from anywhere in the corporation, as captured by the assumption of non-tailored managerial knowledge.

1.3 Empirical strategy

The empirical analysis is based on the Microdatabase Direct investment (MiDi), a data set with balance sheet information on virtually the universe of foreign affiliates of German multinational firms from 1999 until 2010. The data set contains detailed balance sheet information on every affiliate, including its sales, its number of employees and its financial structure. Information on parent and affiliate sectors, mostly at the two digit level, is also provided. From 2002 onwards, information on the sales and the number of employees of the German investor is available.

The data allow testing the following predictions of the model:

Testable hypotheses.

1. *The foreign productivity of a given multinational firm increases in the size of the foreign market $\tilde{Q}_j$ and the labor productivity $\tilde{A}_j$ and decreases in the bilateral communication costs with Germany $\tilde{\theta}_j$, foreign wages $\tilde{w}_j$ and knowledge acquisition costs $\tilde{c}_j$.*
2. *The cut-off productivity level for investment in a foreign country increases with the bilateral communication costs with Germany $\tilde{\theta}_j$, foreign wages $\tilde{w}_j$ and knowledge acquisition costs $\tilde{c}_j$, and decreases in market size $\tilde{Q}_j$ and foreign labor productivity $\tilde{A}_j$.*

The first hypothesis follows from Proposition 2 and the second hypothesis results from Proposition 3 by associating productivity with inverse marginal costs $\frac{1}{\xi_j}$, as is standard in the literature (e.g. Melitz, 2003; Helpman, Melitz, and Yeaple, 2004). A "$\sim$" is used to denote the empirical analog of a model parameter. The production quantity q_j is an endogenous model outcome, so market size is used to take into account the fact that demand and thus q_j will be higher in larger markets.

We use the natural log of sales over employees $\ln(sales/employees)$ to measure productivity. To be consistent with the level of analysis of the model, we aggregate the available information on the parent-country level, taking the degree of participation of the parent in the affiliate into account where applicable.

Recall that the model demonstrates that a firm's marginal cost depends on $\bar{z}_i$, the knowledge level of the firm, which also determines firm selection across countries. To be consistent with the model predictions, the empirical test of Hypothesis 1 focuses on within-parent variation in performance across countries, thus taking into account the fact that differences in $\bar{z}_i$ may affect foreign performance. We estimate the following regression equation:

$$y_{ijt} = \beta_0 + \beta_1 \tilde{Q}_{jt} + \beta_2 \tilde{\theta}_{jt} + \beta_3 \tilde{c}_{jt} + \beta_4 \tilde{w}_{jt} + \beta_5 \tilde{A}_{jt} + \alpha_{it} + \epsilon_{ijt} \tag{1.41}$$

where y_{ijt} denotes the foreign labor productivity of firm i in country j and period t, α_{it} is a parent–year fixed effect, and ϵ_{ijt} is a firm–country–year specific error term. The parent–year fixed effects absorb the effect of $\bar{z}_i$, i.e., of parent characteristics that may influence performance across destinations. The regression results uncover the relation between deviations in the performance measure y_{ijt} from its parent–year specific mean and deviations in the country characteristics $\tilde{x}_{jt}$ from their respective parent–year specific means, abstracting from parent-specific differences. To account for correlations across time, the standard errors are clustered at the level of the parent.

This approach helps to distinguish the predictions of the model from bias due to the self-selection of firms across countries. Large markets with high labor productivity and low factor costs are attractive investment destinations. The higher average productivity of firms in such markets could result from more productive multinationals selecting into those destinations. The empirical specification focuses on the performance of the same multinational firm across different countries and thus tests the model in a clean fashion.

It is still necessary to interpret the resulting estimates $\hat{\beta}$ with caution. The set of locations is a choice variable on the part of the firm, and does not vary exogenously. It is difficult to guarantee that our estimation conditions on all information available to the investor, so the results may still be biased due to unobservable firm–country-specific variables. One possibility for addressing this problem would be to analyze the effect of a large and unexpected shock in one of the regressors in a country on the outcome variables of firms which had already been active in that country before the shock occurred. We have unfortunately not been able to find such a shock, and thus have to stick with more suggestive empirical evidence. Estimating a Heckman selection model is an alternative option. However, such a specification does not allow maintaining a strict focus on within-parent variation across countries, as using firm–year fixed effects is not possible. In addition, the strong distributional assumptions of the Heckman selection model are not fulfilled in the data. These aspects are discussed in more detail in Appendix A.3.

To learn whether the selection pattern across countries is consistent with the predictions of the model, we construct the cut-off productivity level by country for each two digit parent sector group. The model predicts that firm performance across countries is interdependent and that foreign activity negatively affects domestic performance. We therefore employ two different strategies to approximate the cut-off productivity level. We employ the minimum domestic productivity of investors active in a country, which is subject to the caveat just mentioned. We additionally construct the global productivity of a firm as the log of global sales over the global number of employees and use the minimum worldwide productivity of investors in a country as measure for the cut-off productivity level. The estimation equation is:

$$y_{sjt} = \gamma_0 + \gamma_1 \tilde{Q}_{jt} + \gamma_2 \tilde{\theta}_{jt} + \gamma_3 \tilde{c}_{jt} + \gamma_4 \tilde{w}_{jt} + \gamma_5 \tilde{A}_{jt} + \delta_t + \delta_s + \upsilon_{sjt} \tag{1.42}$$

where y_{sjt} denotes the cut-off productivity level of firms in sector s in country j and period t, δ_t, δ_s are a year and sector fixed effect, and ϵ_{sjt} is a sector–country–year specific error term. Standard errors are clustered at the country level.

1.4 Data and descriptive statistics

The Microdatabase Direct investment (MiDi) is provided by the Bundesbank, the German central bank. We use the information on outward foreign direct investment by German companies. The database consists of a panel of yearly information on virtually the universe of foreign affiliates of German firms from 1999 until 2010. By the German Foreign Trade and Payment Regulation (Außenwirtschaftsverordnung), any resident who holds shares or voting rights of at least 10% in a company with a balance sheet total of more than 3 million euro is obliged to report information on the financial characteristics of these affiliates to the Bundesbank. Until 2002, information on stakes of at least 10% in a company with a balance sheet total of more than 5 million euro and stakes of at least 50% in a company with a balance sheet total of more than 0.5 million euro had to be reported (Lipponer, 2009). The same information has to be provided on branches or permanent establishments abroad if their operating assets exceed the reporting threshold. We clean the data so that all observations meet a uniform threshold: we keep 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%.[2]

We drop observations on 20,016 affiliates of investors that are government institutions or private households, agriculture or mining companies and housing enterprises.[3] The resulting data set contains 271,178 affiliate–year observations that correspond to 179,658 parent–country–year observations. Some affiliates are reported several times, because multiple investors hold participating interests in them.

We augment the MiDi with information on country characteristics used as proxies for the model parameters. An overview of the model parameters, the empirical analogs, and the corresponding data source is given in Table 1.1.

[2] 37,251 observations are thus dropped from the sample.

[3] Government institutions and private households are dropped because they are not multinational firms. Agriculture and mining companies are dropped because natural resources are decisive factors for their investments, but ignored in the theoretical and empirical analyses. Housing enterprises are dropped because they often report sales of zero, even though they are not small, which would lead to measurement error in our analysis.

Table 1.1: Overview of model parameters and their empirical analogs

Variables	Definition	Empirical analog	Data source
Endogenous outcomes			
ξ_j	Marginal costs of production	Inverse foreign labor productivity	MiDi
$\xi_0(\bar{z}^I, w_0)$	Cut-off productivity level	Minimum inverse domestic/ global labor productivity	MiDi
Firm characteristics			
$\bar{z}_i$	Knowledge level	Parent-year fixed effect α_{it}	n/a
Country characteristics			
Q_j	Market size	GDP	IMF
		Market potential (sum of GDP weighted by distance)	Own calculation/ CEPII, IMF
θ_j	Bilateral communication costs	Flight time from Frankfurt to main city	www.weltinfo.com/service/flugzeiten.html, www.meine-flugzeit.de, main city: CEPII
		Linguistic distance	Own calculation/ CEPII, Ethnologue, Spolaore and Wacziarg (2009)
		Share of population speaking English	Eurobarometer 52, 55, 237, 243, 386 Crystal (2003)
		Common official language indicator	CEPII
		Time zone differences in minutes	www.timeanddate.com
c_j	Knowledge acquisition costs	Per pupil public expenditure on education as % share of GDP per capita	World Bank (World DataBank)
		PISA maths, science scores	World Bank (World DataBank)
A_j	Labor productivity	GDP per employee	World Bank (World DataBank)
		Labor productivity per employee	OECD
w_j	Wage	Compensation per employee	OECD
τ_j	Transport costs	Distance	CEPII
Additional controls			
	Social capital	Bilateral trust	Eurobarometer 46
	Investment climate	Statutory tax rate	IBFD
		Cost and time to start a business	World Bank (Doing Business)
		Cost, time and # of procedures to enforce contracts	World Bank (Doing Business)
		Cost, time and # of procedures to register property	World Bank (Doing Business)
		Rule of law, government effectiveness	World Governance Indicators

We use GDP data from the IMF to measure differences in demand across countries. We alternatively employ market potential calculated as the sum of GDP of all countries weighted by their distance to the host country (e.g. Chen and Moore, 2010). The bilateral communication costs between headquarter managers and foreign production workers are difficult to capture. We employ a number of different measures to approximate them. We focus on measures that approximate frictions in information and communication flows between people and refrain as far as possible from measures which could also capture differences in the flow of goods such as distance. The main measures are the duration of a flight from Frankfurt to the main city of the host country, the linguistic distance between German and the language(s) spoken in the host country, and the difference in time of day in minutes. We calculate the linguistic distance as a function of the number of common nodes of German and each language spoken by at least 20% of people in the host country according to the classification of languages provided by Ethnologue (Spolaore and Wacziarg, 2009) and take the simple average in case of several languages.[4] The flight time is used to capture how quickly managers can travel to the affiliate and address potential issues in the production process on site. Despite technological advances, face-to-face communication is often indispensable to ensure successful production (e.g. UNCTAD, 2004). In principle, the flight time is correlated with geographic distance. However, there are no direct flights for a number of destinations, so the flight time contains important information on the accessibility of foreign countries beyond their geographic distance. The linguistic distance is based on the intuition that the closer a language is to one's mother tongue, the easier it is for most people to learn it and to express themselves precisely. Undoubtedly, international business communication likely often takes place in English (and we thus use the share of the population speaking English as measure for the communication costs in robustness checks below). Still, as a recent documentation by the Secretariat General Translation Directorate of the European Court of Auditors shows, non-native English speakers tend to develop their own English dialect that is strongly influenced by their native languages and often difficult to understand by native speakers (Gardner, 2013). The linguistic distance is therefore apt to capture frictions in

[4]Specifically, we use $\sqrt{\frac{7-\#\text{ common nodes}}{7}}$, slightly modifying the formula used in Spolaore and Wacziarg (2009), as 7 is the number of linguistic nodes of German and thus the maximum number of common nodes possible.

communication despite the use of English in business contexts. The difference in time of day captures the fact that personnel at either location may have to work unconventional hours to facilitate communication if the affiliate is located in a different time zone. This may increase costs for the firm. We use the difference in the time of day in minutes and not the abosolute value of the deviation in time of day, because the time difference may have an asymmetric effect. The working day at the affiliate location starts before the parent's working day if the time difference is negative, but ends after the parent's working day if the time difference is positive. This may affect affiliate performance because in the former case, parents can be approached with problems at the same day, whereas affiliates may have to wait until the next day to have their problems addressed in the latter case. As robustness checks, we use the share of people speaking English and an indicator that German is among the official languages of the host country. The share of people speaking English is only available for a subset of countries and for a subset of years, but does not vary greatly over time. We therefore assign the available information for each country to all years and use the nearest value for each period.

Knowledge acquisition costs are measured using the expenditure per pupil as percentage share of GDP per capita and PISA maths and science scores. The former is an input measure based on the idea that more public investment in education decreases the marginal costs of schooling. The latter are output measures that take into account the fact that large amounts of public funds per se do not imply that the educational system works efficiently. PISA scores are internationally comparable measures of how well the educational system is able to teach abilities to students. As they are only available for a relatively small subset of countries and years, we prefer the former measure. We use GDP per employee to measure labor productivity. This measure is imperfect, as it is does not reflect differences in hours worked, but is available for a large number of countries. We use labor productivity per employee for OECD countries as robustness check. To measure wages, we employ the compensation per employee provided by the OECD. However, these measures are only available for the majority of OECD countries, and the OECD countries are a non-representative sub-sample of all investment destinations of German firms. We therefore stick to GDP per employee as labor productivity measure and omit wages in most regressions below.[5] The omission of wages is defendable, as wages are endogenously determined in general equilibrium and thus a function of the other covariates. Finally, transport costs are measured using data on bilateral distance. We take the logarithms of the covariates if their distribution in levels is skewed.

To take into account factors which may influence the variation in productivity across countries but are not explicitly included in the model, we also include data on bilateral trust and investment climate measures. We use data on bilateral trust from Germany towards other countries from the Eurobarometer. As investment climate measures, we use statutory tax rates as well as the indicators on the ease of enforcing contracts and of registering property from the World Bank Doing Business Indicators and indicators on the rule of law and government effectiveness from the World Governance Indicators. The cost and time to start a business are used as additional controls in the regressions on the cut-off productivity level.

Table 1.2 provides summary statistics of the variables used in the regression analysis.

[5] We have experimented with the Occupational Wages around the World Database by Freeman and Oostendoorp provided via the NBER (Oostendorp, 2013), a standardized database on wages by sector based on the ILO October Inquiries. However, its overlap with the MiDi database is even worse. Data are available for only 20% of observations.

Table 1.2: Summary statistics

Variable	N	Mean	SD
Foreign productivity	150, 570	5.583	1.189
Domestic productivity (firm level)	24, 357	5.803	1.150
Number of countries per parent (firm level)	54, 961	3.016	4.782
Log GDP	163, 890	13.269	1.533
Log market potential	163, 686	6.457	1.431
Log flight time	165, 349	5.154	.959
Time difference to Germany, minutes	165, 662	0.198	3.588
Linguistic distance	163, 805	0.803	0.232
Official language German	164, 992	0.139	0.346
Per pupil public expenditure on education, % of GDP p.c.	128, 653	22.978	3.982
PISA maths score	46, 329	492.354	38.925
PISA science score	46, 329	495.513	32.579
GDP per employee (in 1990 $ 1,000)	163, 071	3.806	1.581
Labor productivity per employee, OECD	123, 239	11.532	1.390
Log distance	164, 992	7.320	1.298
Log compensation per employee	113, 157	11.087	1.432
Bilateral trust from Germany, survey 1996	120, 875	2.548	0.421
Log costs to enforce a contract	114, 971	3.009	0.363
Log time to enforce a contract	114, 971	6.162	0.439
# of procedures to enforce a contract	114, 971	32.941	5.680
Log costs to register property	102, 230	1.074	1.010
Log time to register property	102, 645	3.416	1.050
# of procedures to register property	102, 645	5.323	2.201
Costs to start a business	114, 980	10.080	37.595
Time to start a business	114, 980	24.525	26.056
Statutory tax rate	164, 110	28.933	7.672

Summary statistics for regression sample. *Variable definitions:* see Table 1.1. The number of observations varies due to differences in country coverage. Maximum possible number of observations: 165,760.

Figures 1.1 to 1.4 provide graphical evidence about the basic relations in the data. Each dot refers to one country. The upper panel scatters the cut-off productivity level by country, where the cut-off productivity level is measured using domestic productivity. The lower panel scatters the average deviation of foreign productivities from the parent–year specific mean. The average deviation of foreign productivities from the parent–year specific mean is calculated as follows. We calculate the mean of foreign productivities for each parent and year and substract it from the observed foreign productivities. We obtain the deviation in the foreign productivity per country from the parent–year specific mean. For every country, we calculate the average of these deviations. This implies that the scatter plots indicate the relationship between country characteristics and the relative performance of firms, i.e. whether firms tend to be more or less productive than average in countries with certain characteristics. The left figures scatter the respective variables against log GDP, the right figures scatter the residuals from a regression of the variables on GDP against log flight time. The relationship between performance and flight time are therefore robust to GDP differences.

Figure 1.1: Cut-off parent productivity vs. GDP, 2005

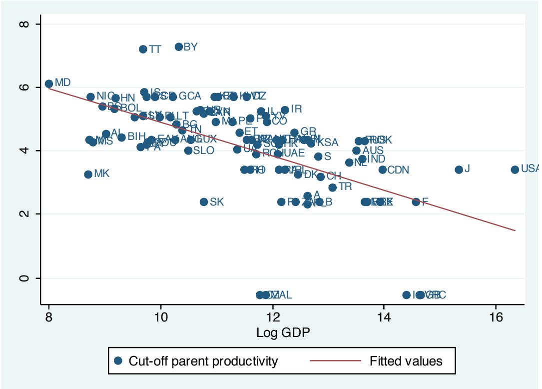


Figure 1.2: Cut-off parent productivity vs. flight time, 2005

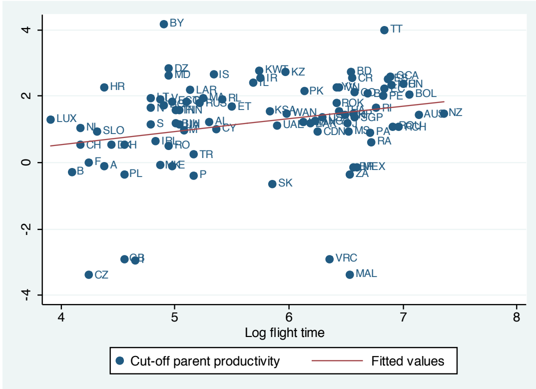


Figure 1.1 scatters the minimum domestic productivity of investors active in a country versus the natural log of country GDP. Figure 1.2 scatters the residuals from a regression of the minimum domestic productivity on log GDP versus the natural log of the flight time from the main city in the host country to Germany. For data confidentiality, only countries where at least three investors are active are included.

Figure 1.3: Foreign productivity vs. GDP, 2005, deviations from parent mean

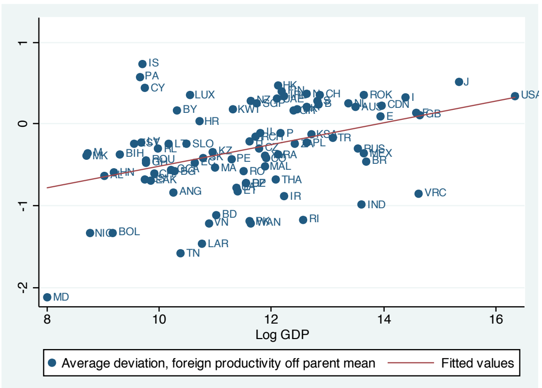


Figure 1.4: Foreign productivity vs. flight time, 2005, deviations from parent mean

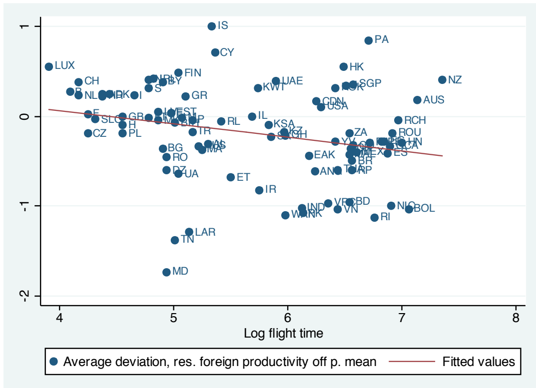


Figure 1.3 scatters the average of the deviations of foreign productivity (ln foreign sales/foreign employees) in a country from the parent specific mean versus the log of host country GDP. Figure 1.4 scatters the residual from a regression of the average deviation of foreign productivity on log GDP versus the log of flight time between the host country and Germany. For data confidentiality, only countries where at least three investors are active are included. Two outliers, the Bahamas and Mauritius, commonly considered as tax havens, are dropped.

The model predicts that only more knowledge intensive and thus more productive firms invest in more challenging destinations (cf. Proposition 3). Figures 1.1 and 1.2 corroborate the conjecture that firm selection is non-random across countries: the cut-off productivity level as measured by the minimum domestic productivity level tends to be higher in smaller markets, and lower in countries that are quicker to reach.

Figures 1.3 and 1.4 show that firms tend to be relatively more productive in countries that are larger and that are more quickly accessible. The relationships are reversed compared to the patterns displayed in Figures 1.1 and 1.2. Larger and less remote countries attract investors that are less productive in the home country, but foreign productivity tends to be higher in larger and less remote countries given domestic productivity. The pattern is consistent with higher $\bar{z}_i$ firms' being able to invest in more difficult destinations, but all firms being more productive in larger markets with lower bilateral communication costs with the home country.

To run the regressions in the next section, we aggregate direct and indirect participation interests per affiliate and further restrict the sample to majority owned affiliates.[6] The relationships predicted by the model hold whenever the parent is actively involved in the management of local production. This may not be the case when the majority of the affiliate is owned by other shareholders. The results are largely robust if we abandon this restriction. We report the regression results for the full sample in Appendix A.2.

1.5 Results

1.5.1 Within-firm differences in performance across countries

Table 1.3 presents the main results on Hypothesis 1. The table displays the number of parents and the number of country combinations contained in the regression sample along with the coefficients. The number of country combinations denotes the number of distinct combinations of countries that the investors included in the regression sample are active in. The number is decisive because the regression results are driven by the variation within firms across countries, so it is important to ensure that the number is sufficiently

[6] 24,060 observations are thus dropped from the sample.

high. The number of country combinations often exceeds the number of parents, which implies that parents change their set of investment destinations over time.

Table 1.3 shows nine specifications. The first specification (column 1) includes log GDP and the measures of communication costs: log flight time in minutes, linguistic distance and the time difference from Germany in minutes. The second specification (column 2) adds knowledge acquisition costs measured by public expenditures on education per pupil as share of GDP per capita and labor productivity measured as GDP per employee. The third specification (column 3) also includes a measure for wages: the log compensation per employee. We separate the first from the other specifications to make transparent whether the measures of communication costs take up omitted country characteristics. The second and third specification are separately displayed because the wage data are only available for OECD countries, so the sample size decreases non-randomly once wages are included. Due to the selectivity of the sample including wages and the associated decrease in the number of observations, we omit wages from further regressions. Columns 4 to 9 include other potential determinants of cross-country differences in productivity that might be omitted variables in the main specification. These specifications will be detailed after discussing the main regression results. Table A.1 presents the regression results if all affiliates, including non-majority owned affiliates, are included in the regression sample. The regression coefficients are quantitatively similar, but less significant, which indicates that non-majority owned affiliates add noise to the sample. Table A.2 displays regression results for columns 4-9 including wages. The results are largely robust, but the number of observations is considerably lower. Recall that these regressions are based on a non-random subsample of countries.

We expect that parents are more productive in larger and more proximate countries with higher labor productivities and lower knowledge acquisition costs and wages.

Table 1.3: Regression results for the within-firm differences in productivity across countries

	1	2	3	4	5	6	7	8	9
Log GDP	0.140***	0.057***	0.081***	0.098***	0.072***	0.035***	0.049***	0.102***	0.053***
	(0.006)	(0.008)	(0.011)	(0.011)	(0.008)	(0.009)	(0.009)	(0.012)	(0.008)
Log flight time	−0.155***	−0.044***	−0.005	−0.028*	−0.034**	−0.057***	−0.080***	−0.055**	−0.105***
	(0.009)	(0.011)	(0.013)	(0.013)	(0.011)	(0.013)	(0.014)	(0.018)	(0.026)
Linguistic distance	−0.631***	−0.071*	−0.064+	−0.084**	−0.059*	−0.147***	−0.059+	−0.069+	−0.067*
	(0.029)	(0.030)	(0.034)	(0.032)	(0.030)	(0.036)	(0.033)	(0.035)	(0.030)
Time difference to Germany	0.002	0.020***	0.036***	0.042***	0.019***	0.024***	0.019***	0.046***	0.019***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)
Public expenditure per pupil,		0.016***	0.018***	0.013***	0.017***	0.018***	0.019***	0.017***	0.015***
% GDP per capita		(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
GDP per employee		0.207***	0.151***	0.108***	0.211***	0.235***	0.189***	0.144***	0.215***
		(0.008)	(0.012)	(0.015)	(0.008)	(0.009)	(0.009)	(0.017)	(0.009)
Log compensation per employee			−0.013*						
			(0.006)						
Trust, survey 1996				0.231***				0.194***	
				(0.032)				(0.038)	
Statutory tax rate					−0.008***			−0.008***	
					(0.001)			(0.002)	
Log costs to enforce contracts						0.029			
						(0.025)			
Log time to enforce contracts						−0.004		0.177***	
						(0.022)		(0.019)	
# of procedures to enforce contracts						0.012***			
						(0.002)			
Log costs to register property							−0.002		
							(0.007)		
Log time to register property							−0.079***	−0.077***	
							(0.008)	(0.008)	
# of procedures to register property							0.010**		
							(0.003)		
Log distance									0.046**
									(0.018)
Constant	5.026***	3.960***	3.767***	3.264***	3.895***	3.775***	4.495***	2.577***	4.002***
	(0.078)	(0.120)	(0.159)	(0.159)	(0.119)	(0.195)	(0.151)	(0.259)	(0.121)
R^2	0.092	0.194	0.136	0.138	0.196	0.204	0.208	0.159	0.195
Observations	149,223	117,519	91,600	99,061	117,315	82,567	72,792	59,728	117,519
# parents	8,331	7,732	6,867	7,239	7,724	6,253	5,903	5,489	7,732
# country combinations	9,249	9,042	8,628	8,638	9,040	6,747	6,140	5,826	9,042

Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Dependent variable:* foreign productivity defined as ln(foreign sales/foreign employees). *Covariate definitions:* see Table 1.1. Sample based on all foreign affiliates with degree of participation of at least 50%.

All signs of the coefficients are in line with the theoretical predictions. The estimated coefficients are generally highly significant, often at the 0.1% level. Parents tend to be relatively more productive in relatively larger countries. Higher communication costs—as measured by longer flight times and higher linguistic distance—generally have a negative impact on the outcome variables, though the effect is less significant once the sample is reduced to OECD countries. Positive time differences increase productivity. It is plausible to interpret this finding as an indication that a longer time lag between the incidence of a problem at a foreign location and the reaction of the parent is productivity decreasing. If affiliates work ahead of the parent, it is possible for them to send the issues to the headquarters at the beginning of the headquarters' office hours and have them addressed the same day; if the working day at affiliate locations begins after the parent's, problems that arise later during the day have to be postponed until the next day. In unreported regressions, we explore whether the relationship is non-linear, but do not find robust evidence for a U-shaped relationship. Relatively lower knowledge acquisition costs due to higher public investments in education and relatively higher labor productivity increases foreign performance, whereas higher wages decrease foreign productivity.

These findings are robust to the inclusion of potential omitted variables. The coefficients remain highly significant. Column 4 includes a measure of bilateral trust between Germany and the host countries calculated based on the survey question in the Eurobarometer from the year 1996 (see Bloom, Sadun, and Van Reenen, 2012; Guiso, Sapienza, and Zingales, 2009). Bloom, Sadun, and Van Reenen (2012) consider trust as a form of social capital and argue that higher levels of bilateral trust between a multinational firm's headquarters and the affiliate increase the firm's productivity by allowing the firm to decentralize more easily. The regression results confirm that higher bilateral trust increases firm productivity. The impact of the communication cost measures is unchanged, even though Guiso, Sapienza, and Zingales (2009) find that trust is positively associated with common linguistic roots. Communication costs and trust have distinct effects on firm performance.

Columns 5 to 8 include statutory tax rates, the cost, time and number of procedures required to register property and to enforce contracts as measures for differences in the investment climate across countries. The first three columns include the measures separately, column 8 includes trust, tax rates and the time to enforce contracts and register property. We choose time to enforce contracts and register property as investment climate measures because these indicators seem most comparable across countries and least likely to pick up differences in the income per capita, which is used as reference to obtain comparable costs measures in the Doing Business Indicators. Our findings are robust throughout specifications, and the investment climate measures affect productivity in intuitive ways. Higher taxation negatively affects foreign productivity. The cost and time to enforce contracts does not have a significant effect, but the number of procedures is positively associated with productivity. This finding is contrary to expectations, but may reflect the fact that given the time and cost of enforcing contracts, a higher number of steps in judicial proceedings need not reflect red tape, but may be attributable to the demand for carefully finding a fair solution. This effect may be taken up by the coefficient of the time to enforce contracts in column 8. The number of procedures required to register property has a similar effect. When it takes longer to register property, a firm's productivity in a country is negatively affected. Column 9 includes geographic distance in the main specification, a common regressor in the gravity literature. The main regression results are robust. Relatively more distant affiliates are predicted to be relatively more productive.

Table 1.4: Regression results for the within-firm differences in productivity across countries, alternative regressors

	1	2	3	4	5	6	7	8
Log GDP		0.034^{***}	0.064^{***}	0.058^{***}	0.053^{***}	0.199^{***}	0.057^{***}	0.058^{***}
		(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Log flight time	-0.043^{***}	-0.090^{***}	-0.038^{***}	-0.051^{***}	-0.057^{***}	-0.015	-0.041^{***}	-0.042^{***}
	(0.011)	(0.012)	(0.011)	(0.010)	(0.009)	(0.013)	(0.012)	(0.012)
Linguistic distance	-0.068^{*}			-0.143^{***}	-0.151^{***}	-0.286^{***}	-0.045	-0.051^{+}
	(0.030)			(0.032)	(0.032)	(0.035)	(0.031)	(0.031)
Time difference to Germany	0.020^{***}	0.025^{***}	0.020^{***}	0.014^{***}	0.016^{***}	0.041^{***}	0.019^{***}	0.019^{***}
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
Public expenditure per pupil,	0.016^{***}	0.015^{***}	0.015^{***}			0.037^{***}	0.017^{***}	0.017^{***}
% GDP per capita	(0.002)	(0.002)	(0.002)			(0.002)	(0.002)	(0.002)
GDP per employee	0.209^{***}	0.234^{***}	0.202^{***}	0.188^{***}	0.198^{***}		0.190^{***}	0.191^{***}
	(0.008)	(0.010)	(0.008)	(0.009)	(0.008)		(0.010)	(0.010)
Log market potential	0.056^{***}							
	(0.008)							
Share speaking English		0.001^{*}						
		(0.000)						
Official language German			0.126^{***}					
			(0.022)					
PISA Maths score				0.002^{***}				
				(0.000)				
PISA Science score					0.001^{***}			
					(0.000)			
Labor productivity per employee, OECD						-0.047		
						(0.062)		
Log compensation per employee						0.031		
						(0.061)		
Rule of law							0.047^{**}	
							(0.017)	
Government effectiveness								0.042^{*}
								(0.017)
Constant	4.335^{***}	4.304^{***}	3.806^{***}	3.752^{***}	3.961^{***}	2.635^{***}	3.941^{***}	3.923^{***}
	(0.096)	(0.153)	(0.126)	(0.163)	(0.166)	(0.156)	(0.128)	(0.130)
R^2	0.194	0.169	0.196	0.171	0.170	0.124	0.199	0.199
Observations	117,519	103,170	117,557	42,079	42,079	91,600	100,164	100,164
# parents	7,732	7,433	7,735	6,753	6,753	6,867	7,255	7,255
# country combinations	9,042	8,811	9,047	4,764	4,764	8,628	8,084	8,084

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ *Dependent variable:* see Table 1.3. *Covariate definitions:* see Table 1.1.

1.5.1.1 Robustness checks

Table 1.4 replicates the regression results when alternative variables are used as measures for the model parameters.

We use the market potential instead of GDP (column 1), and the share of the population speaking English (column 2) and an indicator for German as an official language of a country (column 3) as alternative measures for linguistic distance. The baseline results are robust and significant at least at the 5% level. The alternative measures have the expected effects and are significant.

The results are likewise robust when PISA maths and science scores are employed as an alternative measure for the knowledge acquisition costs (columns 4, 5). As predicted by the model, lower knowledge acquisition costs, here measured using the output of the education system, increase productivity. The sample size is considerably smaller as the measures are only available for a subset of years.

We use labor productivity measures from the OECD as an alternative for GDP per employee and reinclude wages as the sample is restricted to OECD countries anyways (column 6). Labor productivity is calculated as the gross value added divided by the number of employees. As in the main regressions, flight time is insignificant once the sample is restricted to OECD countries. The size of the coefficient of linguistic distance increases by almost 100%. Both labor productivity and wages are insignificant. This finding is surprising, given that productivity is robustly significant in all other specifications. This may result from a combination of factors: on the one hand, GDP per employee is a coarse measure and may pick up effects of income per capita differences. On the other hand, the labor productivity measure is only available for OECD countries, which are relatively homogeneous. Thus, there may just be too little variation to estimate a meaningful effect.

Finally, we use alternative measures for the quality of the investment climate from the World Governance Indicators (columns 7, 8). We include the rule of law, a measure that is also used in Nunn (2007) and government effectiveness. The baseline results are robust. Linguistic distance is significant at the 15% and 10% level respectively. As Kaufmann, Kraay, and Mastruzzi (2010) point out, the World Governance Indicators are survey measures, so the precision of the individual measures should not be overestimated. To take this issue into consideration, we additionally run unreported regressions including the upper and lower bounds of the confidence interval for the measures instead of the point estimates. Results are robust.

1.5.1.2 Relevance

To assess the relevance of our findings, Table 1.5 presents the effect of an increase in the independent variables by one standard deviation on foreign labor productivity expressed in standard deviations. For comparison, it also displays the coefficient from the main regression (Table 1.3, column 2) and the coefficient range. The coefficient range is taken as being from the second lowest to the second largest coefficient, to remove outlying coefficients.

Table 1.5: Effect of covariates in standard deviations of dependent variable

			Coefficient range	
Variables	Main coefficient	Effect in SD	Lower	Upper
Log GDP	0.057	0.073	0.035	0.140
Log flight time	-0.044	−0.035	−0.105	−0.015
Linguistic distance	-0.071	−0.014	−0.286	−0.051
Time difference	0.020	0.060	0.014	0.042
Per pupil public expenditure	0.016	0.054	0.015	0.019
GDP per employee	0.207	0.275	0.144	0.234

Table lists coefficient from Table 1.3, column 2, the effect of an increase in the variable value by one standard deviation on the dependent variable expressed in standard deviations and the coefficient range. The coefficient range is the second lowest to the second largest coefficient to remove outliers. All coefficients are significant.

The labor productivity level in a country as measured by GDP per employee is the quantitatively most important influencing factor of productivity at the firm level. The impact of market size, communication costs and knowledge acquisition costs on productivity differences across countries is similar in magnitude. An increase in market size, or a decrease in communication or knowledge acquisition costs by one standard deviation leads to an increase in foreign productivity by around 0.05 standard deviations.

1.5.2 Cut-off productivity levels

The regression results for the cut-off productivity level are shown in Table 1.6. The minimum domestic productivity of investors per sector in a country and year is used as cut-off productivity measure in the upper panel, and the minimum global productivity of investors per sector, country and year is used in the lower panel. Consistent with Hypothesis 2, the domestic productivity of the least productive investor in a country decreases in market size as measured by log GDP. The cut-off productivity level increases in the communication costs between the host country and the home country as measured by log flight time and linguistic distance. These results confirm the predictions of the model. The generally high significance levels, usually 0.1%, are reassuring, especially given that the number of observations is considerably smaller than in the regressions at the firm level. The signs and significance levels are robust to the inclusion of measures for the fixed costs of investment. The effect of flight time is still positive, but turns insignificant if distance is included. The effect is significant at the 0.1% level once the costs of starting a business are controlled for. The other covariates do not have robust effects. Tables A.3 and A.4 in Appendix A.2 demonstrate that the regression results are robust to the inclusion of additional regressors and largely robust to using alternative measures for the covariates, similar to the robustness checks in the previous subsection.

Table 1.6: Regression results for the cut-off productivity level

Domestic productivity	1	2	3	4	5	6
Log GDP	-0.161^{***}	-0.157^{***}	-0.168^{*}	-0.165^{***}	-0.163^{***}	-0.182^{**}
	(0.017)	(0.028)	(0.059)	(0.031)	(0.028)	(0.046)
Log flight time	0.155^{***}	0.210^{***}	0.321^{***}	0.214^{***}	0.032	0.442^{***}
	(0.034)	(0.056)	(0.058)	(0.061)	(0.114)	(0.088)
Linguistic distance	0.443^{***}	0.404^{***}	0.404^{***}	0.389^{***}	0.404^{***}	0.485^{***}
	(0.105)	(0.097)	(0.092)	(0.106)	(0.111)	(0.109)
Time difference to Germany	0.004	0.009	0.005	0.008	0.007	0.001
	(0.007)	(0.009)	(0.012)	(0.008)	(0.008)	(0.009)
GDP per employee		-0.003	0.048	-0.015	0.015	0.111^{*}
		(0.035)	(0.056)	(0.040)	(0.039)	(0.051)
Per pupil public expenditure,		0.007	0.031^{*}	0.006	0.002	0.035^{*}
% GDP p.c.		(0.011)	(0.014)	(0.012)	(0.011)	(0.014)
Log compensation per employee			0.027			0.027^{+}
			(0.025)			(0.018)
Log costs of starting a business				-0.027		0.066^{**}
				(0.030)		(0.022)
Log time to start a business				0.010		0.049^{+}
				(0.034)		(0.030)
Log distance					0.138	-0.066
					(0.071)	(0.053)
Constant	7.728^{***}	6.410^{***}	4.961^{***}	6.574^{***}	6.438^{***}	4.318^{***}
	(0.224)	(0.558)	(0.907)	(0.637)	(0.540)	(0.734)
R-squared	0.308	0.325	0.394	0.330	0.327	0.394
Observations	11,469	8,280	4,708	8,000	8,280	4,527
Sector dummies	Y	Y	Y	Y	Y	Y
Year dummies	Y	Y	Y	Y	Y	Y
Global productivity	**1**	**2**	**3**	**4**	**5**	**6**
Log GDP	-0.113^{***}	-0.151^{***}	-0.144^{*}	-0.149^{***}	-0.155^{***}	-0.139^{*}
	(0.016)	(0.031)	(0.054)	(0.035)	(0.031)	(0.053)
Log flight time	0.154^{***}	0.233^{***}	0.294^{***}	0.216^{***}	0.098	0.547^{***}
	(0.036)	(0.053)	(0.058)	(0.055)	(0.154)	(0.125)
Linguistic distance	0.202^{+}	0.414^{***}	0.491^{***}	0.388^{**}	0.409^{**}	0.492^{***}
	(0.153)	(0.118)	(0.101)	(0.125)	(0.134)	(0.123)
Time difference to Germany	-0.001	0.003	0.023^{+}	0.005	0.002	0.025
	(0.007)	(0.009)	(0.016)	(0.009)	(0.009)	(0.013)
GDP per employee		0.083	0.136^{*}	0.078^{+}	0.096	0.143
		(0.044)	(0.063)	(0.052)	(0.049)	(0.068)
Public expenditure per pupil,		0.013^{+}	0.035^{*}	0.006	0.010	0.034^{*}
% GDP p.c.		(0.010)	(0.013)	(0.010)	(0.010)	(0.014)
Log compensation per employee			-0.004			-0.007
			(0.025)			(0.019)
Log costs of starting business				-0.028		0.021
				(0.027)		(0.025)
Log time to start business				0.023		0.070^{*}
				(0.033)		(0.032)
Log distance					0.106	-0.175^{*}
					(0.114)	(0.080)
Constant	2.347^{***}	4.191^{***}	3.166^{**}	4.538^{***}	4.172^{***}	3.365^{***}
	(0.256)	(0.527)	(0.815)	(0.591)	(0.518)	(0.784)
R^2	0.316	0.356	0.460	0.361	0.357	0.466
Observations	14,476	10,309	5,512	9,939	10,309	5,291
Sector dummies	Y	Y	Y	Y	Y	Y
Year dummies	Y	Y	Y	Y	Y	Y

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Dependent variable: cut-off productivity - min(ln (domestic sales/domestic employees)) by country and sector group.
Covariate definitions: see Table 1.1.

1.6 Discussion

The results on within-parent across-country differences in firm performance and the cut-off productivity levels of foreign destinations obtained in the empirical analysis are consistent with the model predictions. A given investor exhibits superior performance in large markets characterized by relatively low bilateral communication costs between the home country and the host country, comparatively low knowledge acquisition costs, and relatively high labor productivity. The cut-off productivity levels decrease in market size and increase in bilateral communication costs. This evidence is indirect: the data do not contain information on the allocation of knowledge across countries which are the focus of the model mechanism. It is therefore a key concern with respect to the empirical results whether they can really be attributed to the organization of knowledge in multinational firms.

Two recent papers study the relation between the geographical distribution of establishments and firm performance using U.S. data on national multi-establishment firms. Giroud (2013) finds that investment in a plant increases after a new airline route between the firm's headquarter and the plant location is introduced. Kalnins and Lafontaine (2013) demonstrate that greater distance of the establishment from the headquarters is associated with shorter establishment survival. Both articles attribute their findings to monitoring problems and information asymmetries between firm headquarters and establishments.

This chapter proposes an alternative explanation to rationalize these empirical findings: new airline routes and lower distance decrease communication costs between headquarters and plants, which renders plants more productive and increases firms' investment incentives. At the same time, one could argue in the vein of Giroud (2013) that the empirical results in this chapter are driven by monitoring problems within multinational firms. The effect of flight time can be rationalized using a model that features monitoring problems between production workers and headquarter managers, variation in cross-border monitoring costs, and heterogeneity in firms' monitoring technology, for example along the lines of Qian (1994). If monitoring costs are higher, a firm has to pay higher wages to implement the optimal effort level, so marginal costs of production are higher. Thus, such a model would likewise generate lower within-firm productivity in countries with higher cross-border monitoring costs. (A formal analysis of such a model is sketched in Appendix A.4.)

Neither of the two mechanisms is susceptible to a direct test with the data available. While we do not deny that monitoring is an important factor for efficient production, we are convinced of the empirical relevance of the mechanism based on the organization of knowledge proposed in this chapter. Higher bilateral trust is likely to decrease monitoring costs. We find that communication costs are relevant even if bilateral trust is controlled for. In addition, we approximate communication costs not only by flight time, but also using linguistic distance. At least with respect to routine tasks, it is difficult to claim that monitoring problems are less easily mitigated if the linguistic distance between the home country and the host country is larger. A supervisor's assessment of a worker's performance depends on observing what the worker does, not what the worker claims to do.

Finally, our knowledge-based mechanism is consistent with empirical evidence on sectoral differences in the geographical concentration of foreign direct investment (Bahar, 2013, c.f. section 1.2.3.2) and explains *home* country labor market effects of multinational activity. Heyman, Sjöholm, and Tingvall (2007) provide evidence that both foreign owned firms and domestic multinational firms tend to pay higher wages than domestic non-multinational firms using Swedish establishment level data. As domestic knowledge increases in overall firm knowledge ($\frac{\partial z_0}{\partial \bar{z}} > 0$, cf. Appendix A.4) and multinational firms exhibit higher knowledge levels than non-multinational firms, our model is consistent with these findings. If monitoring were the only driver behind the empirical patterns presented in the previous section, the opposite should be the case: to overcome higher cross-border monitoring costs, firms would have to dispose of a better monitoring technology. Firms with better monitoring technology are able to implement optimal effort levels with lower wage payments. Multinational parents are therefore unambigously predicted to pay lower wages than their domestic counterparts in the home country according to a monitoring based model, which is at odds with the empirical evidence.

1.7 Conclusion

This chapter studies the organization of knowledge in multinational firms. Multinational firms are predicted to optimally allocate more knowledge to foreign countries characterized by higher bilateral communication costs with the home country, higher labor productivity, lower wages and lower knowledge acquisition costs. The distribution of productivities across countries within German multinational firms is shown to be consistent with these predictions, as is the pattern of cut-off productivity levels. In addition, both the home and host country multinational wage premiums generated by the model are in line with the available empirical evidence.

The chapter offers relevant insights for the design of investment promotion policies. Creating well-paid, relatively knowledge intensive new jobs is one of the main targets of investment promotion efforts (Javorcik, 2012). The results of this chapter generally support the presumption that employment in multinational affiliates is likely to be more knowledge intensive and better paid than employment in domestic firms. In their efforts to reap these benefits, countries tend 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 chapter demonstrates, targeted foreign language training and good communication infrastructures may be equally relevant to foster FDI inflows, as they facilitate multinational enterprises' task of efficiently organizing across countries.

Chapter 2

Knowledge transmission costs and the organization of multi-establishment firms*

2.1 Introduction

Knowledge is a crucial input in many production processes and an important determinant of productivity. A large literature studies the diffusion of knowledge across space and its impact on macroeconomic growth and performance (for a survey, see Keller, 2004). Despite the surge of innovation in the communication sector, the spatial transmission of knowledge is still limited, suggesting that a sizeable fraction of knowledge is tacit in nature (Keller, 2004). Notwithstanding, many microeconomic studies treat knowledge as a public good within the firm that is available at negligible costs in all divisions, including spatially separate ones (see e.g. Markusen, 1984; Markusen and Maskus, 2002, for a survey). Comparatively few papers analyze costly communication within firm boundaries to date, but these papers consistently find that communication costs matter: for example, within-firm communication costs rationalize the development of firm-specific expressions and languages (Cremer, Garicano, and Prat, 2007), and are an important influencing

*This chapter is based on joint work with Manfred Antoni.

factor of the international expansion of multinational firms (Bahar, 2013; Keller and Yeaple, 2013). Disregarding communication costs inside the firm may thus hamper the understanding of firm organization and growth.

The second chapter of this thesis studies the impact of knowledge transmission costs on the organization of multi-establishment firms, central players in today's economies. Multi-establishment firms make up only a small share of the population of firms, but account for disproportionate shares of output and employment (Bernard and Jensen, 2007). In the data from German Social Security Records used in this chapter, only 10% of firms consist of several establishments, but almost half of observed employees work for them. Given their economic importance, understanding the organization, investment and remuneration decisions of multi-establishment firms is crucial to correctly assess the effect of economic and labor market policies.

For this purpose, this chapter constructs a model of the organization of multi-establishment firms. The model is based on three assumptions. First, both labor and knowledge are required for production. Knowledge is defined in a broad sense and refers to all competencies necessary to take decisions on the production process, including knowledge about technology and about management. Second, knowledge is only useful in production if it is learned by a firm's employees, which is costly. Third, communication between employees within the firm is possible, but also costly. Employees can thus leverage on potentially different knowledge sets. Multi-establishment firms consist of headquarters and production establishments that may be located at the same place as the headquarters or at a different location. Multi-establishment firms can allocate firm knowledge centrally at the headquarters or decentrally at the different production establishments. In choosing among these alternatives, they face a trade-off between communicating across space or building costly local knowledge capacities. The optimal amount of knowledge allocated to the single establishments within a company is heterogeneous and increases with higher bilateral cost of communication with the company headquarters and lower local factor input costs.

We demonstrate that the allocation of knowledge within multi-establishment firms affects the remuneration and the span of control of their employees, i.e. the number of employees at lower hierarchical layers that an employee is responsible for.[1] Both the remuneration and the span of control increase in knowledge. To test the model implications empirically, we hypothesize that the observed remuneration and the observed span of control of a firm's employees increase in the bilateral costs of communication between an establishment and the headquarters. As factor input costs have a direct effect on the remuneration in addition to the indirect effect via knowledge, the effect of factor input costs on knowledge is only testable via the span of control. The span of control is predicted to decrease in wages and knowledge acquisition costs.

We construct a novel data set with linked employer-employee data from the German Social Security Records and information on the affiliation of establishments to firms using record-linkage techniques. The data are particularly apt to test the question of interest. Germany is the forth largest economy in the world and has multiple economic centres.[2] The data contain comprehensive and reliable information from administrative sources on the location, sector and employees' characteristics of a large sample of multi-establishment firms with varying location patterns. We find evidence that the organization of German multi-establishment firms is in line with model predictions. In particular, the remuneration increases in the costs of communication between an establishment and the headquarters, as predicted by the model.

The chapter is based on the literature of firms as communication networks and knowledge hierarchies (e.g. Bolton and Dewatripont, 1994; Garicano, 2000; Radner, 1993). This literature analyzes how a firm's organizational structure adjusts to knowledge acquisition costs and communication costs between agents. More specifically, we build on the framework proposed in Garicano (2000). Garicano (2000) characterizes firms as knowledge hierarchies that endogenously emerge in a set-up with knowledge and labor as production inputs, and demonstrates that the allocation of knowledge within an organization is affected by within-firm communication costs, knowledge acquisition costs and sector

[1] The term "remuneration" refers to the wages earned by a firm's employees. We use the term remuneration to distinguish firm level wages from the level of wages in the local labor market that is exogenous to the firm.

[2] To illustrate, 18% of the population of France and 13% of the population of the UK lived in their largest cities in 2012. In Germany, the 12 largest cities need to be taken together to account for 15% of the population (Eurostat, 2013).

characteristics. The model framework has been applied to various settings, for example labor economics (Garicano and Rossi-Hansberg, 2006) and international economics (e.g. Antràs, Garicano, and Rossi-Hansberg, 2006). Many implications have been confirmed empirically (Garicano and Hubbard, 2007; Bloom, Sadun, and Van Reenen, 2012). Caliendo and Rossi-Hansberg (2012) and Caliendo, Monte, and Rossi-Hansberg (2012) add unit constraints in management input to the model framework in order to explain labor market effects of exporting.

We extend the knowledge hierarchy framework to a setting where firms are active at several locations and consist of multiple establishments and headquarters. In our set-up, the firm as a whole as well as each establishment are structured as Garicano-type knowledge hierarchies: employees are assigned to different hierarchical layers with different knowledge levels that imply different remuneration levels. The model features within-establishment heterogeneity in remuneration due to the endogenous emergence of hierarchical layers. Knowledge allocation is also heterogeneous between establishments: the firm allocates more knowledge to an establishment if communication costs between the establishment and firm headquarters are higher, or if local wages and knowledge acquisition costs are lower. Observationally equivalent employees at the same hierarchical layer but distinct locations are thus endowed with different competencies, so their remuneration varies.

The chapter contributes to different discussions in the literature. A few recent studies provide evidence on the importance of geography for the organization of multi-establishment firms. Giroud (2013) demonstrates that new airline routes between headquarters and establishments increase establishment productivity and investment. Kalnins and Lafontaine (2013) show that headquarter proximity increases establishment survival rates. This chapter relates to the literature by demonstrating that spatial proximity affects the organizational structure of multi-establishment firms through its impact on communication costs inside the firm.

The chapter also adds to the discussion on employer wage premiums. A large literature has demonstrated that observationally equivalent workers receive different remuneration at different employers, and that the levels of remuneration tend to increase in employer size (e.g. Bayard and Troske, 1999; Troske, 1999). Card, Heining, and Kline (2013) highlight the increasing relevance of establishment heterogeneity for explaining the rise in aggregate wage inequality. This chapter features endogenously arising inequality in remuneration of observationally equivalent workers within firms. Firm organization helps interpret employer wage premiums. Similarly, Caliendo and Rossi-Hansberg (2012) demonstrate that a single-firm knowledge hierarchy model is useful to understand the effect of exporting on wages.

The chapter is structured as follows. The next section analytically studies the problem of firm organization and derives comparative statics results on the impact of communication costs and factor input costs on firm organization. Section 2.3 describes the data used in the empirical analysis and presents descriptive statistics. Section 2.4 outlines the empirical strategy and section 2.5 presents the results. Section 2.6 discusses how the insights in this chapter help to understand the role of establishment heterogeneity for wage inequality. The last section concludes.

2.2 Theory

2.2.1 Set-up

The theory section studies the optimal organization of an exemplary multi-establishment firm. Demand for establishment output and establishment locations are taken as given. The analysis determines the optimal organizational structure of each establishment and of the headquarters as well as the impact of firm organization on employees' remuneration and their span of control across locations.

The firm is assumed to consist of headquarters located at $j = 0$ and two production establishments. We assume that one of these two establishments is situated at the same location as the headquarter, and the other at a geographically distinct location $j = 1$. Locations are heterogeneous with respect to local wages w_j, knowledge acquisition costs c_j, labor productivities A_j and the bilateral communication costs with the headquarters θ_j.

There is a sufficient number of agents that the firm can hire and that are ex-ante identical. Each agent is endowed with one unit of time. Production is assumed to be a problem solving process based on labor and knowledge. Knowledge captures any competencies required to produce, including both expertise about technology and the capability to take decisions on the production process. For each unit of labor employed in production, a mass 1 of problems are realized. The problems are distributed according to a distribution function with exponential density:

$$f(z) = \lambda e^{-\lambda z} \tag{2.1}$$

where $z = [0, \infty)$ is the domain of possible problems and $\lambda > 0$ denotes the problem arrival rate. Higher λ is associated with a more predictable production process: the mass of the problem probability distribution is concentrated close to zero. This density is chosen for analytical tractability. To transform labor into output, the problems that arise in production have to be solved.

The firm endogenously chooses the amount of knowledge it would like to acquire. Knowledge is modelled as an interval. The firm is able to solve all problems that are realized within its knowledge interval. As the mass of the problem probability distribution is largest close to zero, it is optimal for each firm to choose zero as lower bound of the knowledge interval. The firm additionally determines the upper bound $\bar{Z}$ of the knowledge interval. Labor productivity A_j denotes the maximum amount of output that can be produced using one unit of labor input. The output q_j from one unit of labor input is calculated as the value of the cumulative distribution function:

$$q_j = A_j(1 - e^{-\lambda \bar{Z}}) \tag{2.2}$$

Knowledge can be communicated among the firm's employees, but communication is costly. As is standard in the literature (Bolton and Dewatripont, 1994; Garicano, 2000, e.g.), the costs of communication are born by the receiver of the message. The receiver of the message spends η units of time listening to the sender if both are located at the same place, where $1 > \eta > 0$. It is possible to communicate across space. This possibility comes at an increase of the cost of communication by θ_j, so the receiver of a message at location 0 from location 1 has to spend $\theta_1\eta < 1$ units of time listening to the sender.[3] To capture frictions in communication across space, it is assumed that $\theta_1 \geq \theta_0 = 1$.

Each employee is remunerated for working for the firm. The employees receive a location specific wage w_j for each unit of time. They are also compensated for their knowledge z. To acquire knowledge they have to spend a location specific amount $c_j z$ of time for studying. Total remuneration per employee is thus $w_j(1 + c_j z)$. For expositional clarity, the term "remuneration" always refers to the payments that employees receive from their employer, whereas wages denote the local equilibrium wage that is exogenous to the firm.

Garicano (2000) demonstrates that the optimal organizational structure in the described setup is an endogenously emerging hierarchy. The organization consists of a number of hierarchical layers L with n_L^ℓ employees per layer. Knowledge levels z_L^ℓ are homogeneous within layers, but distinct between layers. Only the employees at the lowest layer $\ell = 0$ spend their time in production. The employees at higher layers $\ell = 1, ..., L-1$ spend their time listening to and solving problems sent by lower layers of the hierarchy. The employees only communicate with other employees in adjacent hierarchical levels: the employees from layer $\ell - 1$ send unsolved problems to employees in layer ℓ, who solve those problems realized within their knowledge interval and send the remaining problems to employees at layer $\ell + 1$ (or, if $\ell = L - 1$, the headquarters). The employees at lower layers learn the solutions to more common problems and the employees at higher layers learn solutions to problems that are less likely to occur.

In Garicano (2000), neither knowledge gaps nor knowledge overlaps occur between the knowledge intervals at two different layers. Knowledge gaps are sub-optimal because it is always rational to learn the solution to more frequent problems first. Knowledge overlaps are also not optimal as identical knowledge at two layers in the organization increases costs, while the knowledge remains unused by one of the two layers. In our set-up, it can be optimal to have knowledge overlaps between the headquarters and one of the two establishments due to asymmetries in communication costs between the

[3] Intuitively, the communication costs η and $\theta_1\eta$ are required to be smaller than one because it is not optimal to communicate if communication requires more time than the generation of new problems.

establishment and the headquarters as well as local factor input costs, as long as the redundant headquarter knowledge is used to address problems that arise in the other establishment. In principle, knowledge gaps between one of the two establishments and headquarters emerge whenever the marginal costs of closing the knowledge gap exceed the marginal benefit of doing so. To increase analytical tractability, we assume that knowledge gaps must not occur. This assumption does not affect the insights that we obtain. The optimal organizational structure including knowledge gaps is analyzed in Appendix B.1.4.

The problem of the firm is to determine the optimal amount of overall knowledge $\bar{Z}$ as well as the optimal design of the firms' organizational structure. The organizational structure consists of the number of hierarchical layers L_j at each of its locations, the number of employees $n^{\ell}_{L,j}$ in each of the layers $\ell = 0, ..., L-1$ and the knowledge level of the employees $z^{\ell}_{L,j}$, as well as the knowledge assigned to the headquarters $[\underline{Z}_H, \bar{Z}_H]$. Given the insight derived in Garicano (2000) that higher layers learn solutions to problems that occur less frequently in the production process, it is evident that $\bar{Z}_H = \bar{Z}$, i.e. the knowledge of the headquarters is most specialized.

We assume that the firm is headed by a single CEO endowed with one unit of time (as in Caliendo and Rossi-Hansberg, 2012). As argued by Kaldor (1934), the assumption of a fixed amount of managerial input helps to rationalize equilibrium heterogeneity in the optimal size of establishments. The assumption seems appropriate with view to real world firms. Most firms are led either by a single CEO, managing director or owner-entrepreneur, and firms exhibit heterogeneity in their organizational structure. Without this assumption, all firms would choose the same optimal organizational structure independent of their size: given the knowledge acquisition costs and the communication costs, there is a unique cost-minimizing design of the firm's organization that can be replicated for firms of any arbitrary size by choosing the appropriate number of headquarter managers (that may infinitesimally small, if necessary).[4]

[4]Heterogeneity in the optimal organization emerges in the absence of above constraint if the overall knowledge of the firm is not endogenized and the heterogeneity in the level of overall knowledge across firms is assumed ex ante, as in Chapter 1.

For simplicity, we further assume that the headquarters consist only of the CEO. This assumption is not crucial for the main implications of the chapter, but simplifies the analysis, as demonstrated in Appendix B.1.3. The number of employees hired in the production establishments is endogenously determined.

Finally, Assumption 1 ensures that the knowledge levels $z^{\ell}_{L,j}$ are positive $\forall \ell = 1, ..., L-1$.

Assumption 1. The parameters η, λ and c_j are such that $\frac{c_j}{\lambda} \leq \frac{\eta}{1-\eta} \quad \forall j$.

2.2.2 Cost minimization problem

The cost minimization problem of the firm is analyzed in two steps. First, the optimal number of layers of the establishments, the optimal number of employees per layer and the knowledge level of employees at the establishments are determined. In doing so, the knowledge level of the headquarters $[\underline{Z}_H, \bar{Z}_H]$, the cost of headquarter services p_j and demand for output q_j are taken as given. Afterwards, the knowledge level of the headquarters is endogenized.

The firm faces the following cost minimization problem at each production location j. The decision on the number of layers of each establishment is a discrete decision.

$$C_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) = \min_{L_j \geq 0} C_{L,j}(q_j, \underline{Z}_H, \bar{Z}_H, p_j) \tag{2.3}$$

The firm chooses the number of layers L_j with the minimum costs to produce a given amount of output at the production location. That is, the optimal number of layers yields minimal average costs with respect to total establishment output.

Given the number of layers, the firm optimally chooses the number of employees per layer and their knowledge level.

$$C_{L,j}(q_j, \underline{Z}_H, \bar{Z}_H, p_j) = \min_{\{n^l_{L,j} \geq 0, z^l_{L,j} \geq 0\}_{l=0}^{L-1}, d_j \geq 0} \left(\sum_{l=0}^{L-1} n^l_{L,j} w_j (1 + c_j z^l_{L,j}) + p_j d_j \right) \tag{2.4}$$

$$\text{s.t.} \quad n^0_{L,j} A_j (1 - e^{-\lambda \bar{Z}_H}) \geq q_j \tag{2.5}$$

$$\sum_{\ell=0}^{L-1} z^\ell_{L,j} \geq \underline{Z}_H \tag{2.6}$$

$$\sum_{\ell=0}^{L-1} z^\ell_{L,j} \leq \bar{Z}_H \tag{2.7}$$

$$n^l_{L,j} = n^0_{L,j} \eta e^{-\lambda \sum_{k=0}^{\ell-1} z^k_{L,j}} \tag{2.8}$$

$$d_j = n^0_{L,j} \eta \theta_j e^{-\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} \tag{2.9}$$

where d_j denotes the amount of CEO time employed in production at location j.

The firm minimizes the total remuneration per layer and the costs of headquarter inputs. It is constrained to produce at least q_j units of output per establishment (Equation 2.5) and to assign at least a total knowledge of $\underline{Z}_H$ to the establishment (Equation 2.6) in order to rule out knowledge gaps between the establishment and the headquarters. In addition, establishment knowledge is limited to overall firm knowledge $\bar{Z}_H$ (Equation 2.7). The number of employees per layer has to be sufficient to answer all problems that are sent (Equation 2.8). The number of employees at layer ℓ, $n^\ell_{L,j}$ is calculated as the mass of problems generated in the lowest layer $n^0_{L,j}$ times the probability that the solution to the problem is not contained in the knowledge sets up to layer ℓ, $e^{-\lambda \sum_{k=0}^{\ell-1} z^k_{L,j}}$, times the communication costs η. If problems are sent to the headquarters, this calculation has to be adjusted for the possibility that headquarters are not at the same location by inflating communication costs with θ_j (Equation 2.9).Finally, all choice variables are restricted to be positive.

Using (2.8) and (2.9), the Lagrangian equation is given by:

$$
\begin{aligned}
\mathcal{L}_j = &\frac{d_j}{\eta\theta_j} e^{\lambda \sum_{l=0}^{L-1} z_{L,j}^l} w_j(1 + c_j z_{L,j}^0) + \sum_{l=1}^{L-1} \frac{d_j}{\theta_j} e^{\lambda \sum_{k=l}^{L-1} z_{L,j}^k} w_j(1 + c_j z_{L,j}^l) + p_j d_j \\
&+ \phi_j \left[\underline{Z}_H - \sum_{l=0}^{L-1} z_{L,j}^l \right] + \chi_j \left[\sum_{l=0}^{L-1} z_{L,j}^l - \bar{Z}_H \right] + \xi_j \left[q_j - A_j(1 - e^{-\lambda \bar{Z}_H}) \frac{d_j}{\eta\theta_j} e^{\lambda \sum_{l=0}^{L-1} z_{L,j}^l} \right].
\end{aligned}
$$

The multiplier ξ_j denotes the marginal costs of production. ϕ_j denotes the marginal costs of increasing $\underline{Z}_H$. ψ_j does not have an intuitive interpretation.

To endogenize $[\underline{Z}_H, \bar{Z}_H]$, the firm solves

$$
\begin{aligned}
\min_{\underline{Z}_H, \bar{Z}_H, \{p_j\}_{j=0}^J} \quad & w_0(1 + c_0(\bar{Z}_H - \underline{Z}_H)) \\
& + \sum_{j=0}^{1} \Big(C_j(d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j), \underline{Z}_H, \bar{Z}_H, p_j) - p_j d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) \Big) \qquad (2.10) \\
\text{s.t.} \quad & \bar{Z}_H \geq \underline{Z}_H \qquad (2.11) \\
& 1 = \sum_{j=0}^{1} d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) \qquad (2.12)
\end{aligned}
$$

The first summand corresponds to the costs that arise for remunerating the CEO. The second summand denotes the sum of the costs that accrue at the production establishments, where the costs for headquarter knowledge employed in local production are deducted because they are an intra-firm transfer. The firm chooses the cost minimizing amount of overall and CEO knowledge subject to the constraints that CEO knowledge has to be positive and CEO time has to be used. The former constraint has to be slack in optimum: if CEO knowledge is zero, the CEO is useless. The latter constraint has to be binding, as otherwise costly CEO capacity would remain unused.

Appendix B.1.1 provides a formal derivation of the optimal organizational structure.

2.2.3 Optimal organizational structure

The optimal amount of headquarter time employed at each establishment is determined by the quantity constraint (2.5):

$$d_j = \frac{\eta \theta_j q_j e^{-\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}}}{A_j(1 - e^{-\lambda \bar{Z}_H})} \tag{2.13}$$

The number of employees at layers $\ell = 0, ..., L-1$ is determined via Equations (2.8) and (2.9):

$$n_{L,j}^{0} = \frac{q_j}{A_j(1 - e^{-\lambda \bar{Z}_H})}$$
$$n_{L,j}^{\ell} = \frac{q_j \eta e^{-\lambda \sum_{k=0}^{\ell-1} z_{L,j}^{k}}}{A_j(1 - e^{-\lambda \bar{Z}_H})}$$

The span of control $s_{L,j}^{\ell}$ denotes the number of employees in layer $\ell - 1$, $\ell \geq 1$ that an employee in layer ℓ is responsible for. It is determined by the knowledge levels:

$$s_{L,j}^{\ell} = \frac{n_{L,j}^{\ell-1}}{n_{L,j}^{\ell}} = e^{\lambda z_{L,j}^{\ell-1}} \tag{2.14}$$

To determine establishment knowledge, it is necessary to distinguish whether the knowledge constraint (2.6) is binding or slack. The knowledge constraint is binding for at least one of the establishments. Otherwise, the overlap of knowledge at the two production locations and the headquarters would never be used in production and is thus superfluous. If the knowledge constraint is binding, establishment knowledge is determined as $\sum_{\ell=0}^{L-1} z_{L,j}^{\ell} = \underline{Z}_H$. If the constraint is slack, the knowledge level of the establishment is determined by the following equi-marginality condition:

$$p_j = \frac{c_j w_j}{\lambda \theta_j} e^{\lambda z_{L,j}^{L-1}}. \tag{2.15}$$

The marginal cost of using headquarter services is equal to the marginal costs of saving headquarter services and employing more knowledge locally in the establishment.

The knowledge levels in the different layers relate recursively as

$$z_{L,j}^{\ell-1} = \frac{1}{\lambda}\ln\left(\frac{\lambda}{c_j}\left(1 + c_j z_{L,j}^{\ell} + \frac{c_j}{\lambda}\right)\right) \quad \forall \ell \in (2, ..., L-1); \tag{2.16}$$

$$z_{L,j}^{0} = \frac{1}{\lambda}\ln\left(\frac{\eta\lambda}{c_j}\left(1 + c_j z_{L,j}^{1} + \frac{c_j}{\lambda}\right)\right) \tag{2.17}$$

Finally, the marginal costs of production are given by

$$\xi_j = \frac{w_j(1 + c_j z_{L,j}^0 + \frac{1}{\lambda}c_j)}{A_j(1 - e^{-\lambda \bar{Z}_H})} \tag{2.18}$$

if the knowledge constraint is slack, and by

$$\xi_j = \frac{w_j(1 + c_j z_{L,j}^0 + \frac{1}{\lambda}c_j) - \frac{\eta}{\lambda}e^{-\lambda \sum_{l=0}^{L-2} z_{L,j}^l} w_j c_j + \eta\theta_j p_j e^{-\lambda \underline{Z}_H}}{A_j(1 - e^{-\lambda \bar{Z}_H})} \tag{2.19}$$

if it is binding.

The upper bound of headquarter knowledge is determined by

$$e^{\lambda \bar{Z}_H} = 1 + \sum_{j=0}^{1} \frac{\lambda q_j w_j}{A_j w_0 c_0}\left(1 + c_j z_{L,j}^0 + \frac{c_j}{\lambda}\right). \tag{2.20}$$

The lower bound of headquarter knowledge $\underline{Z}_H$ is determined by the constraint

$$\sum_{j=0}^{1} d_j = 1. \tag{2.21}$$

2.2.4 Comparative statics

Unless location characteristics are homogeneous, the optimal organizational structure is heterogeneous across establishments.

Local wages and knowledge acquisition costs have ambiguous effects on the optimal amount of knowledge at location $j = 0$ because they affect both the cost of production in the establishment and the cost of using the headquarters. θ_0 is fixed to 1 by definition: as the headquarters is located at $j = 0$, employees in the establishment at

$j = 0$ can communicate with the headquarters at cost η. We therefore focus on the establishment at $j = 1$ in the following.

Proposition 1. *Given the number of layers L_1, the optimal knowledge level of establishment $j = 1$ is increasing in the mark-up on bilateral communication costs between the headquarters and the establishment θ_1 and decreasing in wages w_1 and knowledge acquisition cost c_1.*

Proof. See Appendix B.1.2. □

Intuitively, higher bilateral communication costs between the headquarters and the establishment θ_1 entail an incentive to assign more knowledge to the establishment to save on headquarter time. Local knowledge acquisition is more costly if local wages w_1 or knowledge acquisition costs c_1 are higher, so the optimal level of knowledge decreases.

Proposition 2. *Given the number of layers L_1, the difference in knowledge levels $z^{\ell}_{L,1} - z^{\ell-1}_{L,1}$ is increasing in the mark-up on bilateral communication costs between the headquarters and the establishment θ_1 and decreasing in wages w_1. The effect of knowledge acquisition cost c_1 is ambiguous.*

Proof. See Appendix B.1.2. □

The intuition for this result works as follows. Allocating more knowledge to a group of employees implies that their remuneration increases. The number of employees per layer is lower at higher hierarchical layers. It is therefore optimal to increase the knowledge levels at higher layers overproportionally, as this cushions the increase in production costs. Factors that increase the overall knowledge level thus increase differences in the remuneration of employees within the establishment, whereas factors that dampen the overall knowledge level dampen differences in the remuneration.

Output is the main determinant of the number of hierarchical layers. By analogy to Proposition 2 in Caliendo and Rossi-Hansberg (2012), the number of hierarchical layers is increasing in output. Given output, the number of layers is weakly increasing in the knowledge acquisition costs, but does not depend on wages, labor productivity or bilateral communication costs, as these shift total costs, but do not affect the allocation of knowledge among layers.

As the implications on the number of layers are not testable with the data at hand due to the lack of information on output, we neglect considerations concerning the number of layers in the following. Instead, we test the predictions on the optimal allocation of knowledge through their effect on the remuneration and the span of control of employees, two observables in our data set.

The remuneration of an employee is $w_j(1 + c_j z^{\ell}_{L,j})$. It is increasing in the bilateral communication costs as the knowledge level is increasing in the communication costs. Wages and knowledge acquisition costs exert a direct positive effect on remuneration and an indirect negative effect via knowledge. The effect of wages and knowledge acquisition costs on knowledge is therefore not testable via the observed remuneration of an employee, whereas the effect of communication costs is testable.

The effect of factor input costs are testable via the span of control: the span of control $s_j = e^{\lambda z^{\ell-1}_{L,j}}$ is determined by the knowledge levels. It is increasing in the communication costs and decreasing in wages and knowledge acquisition costs due to their corresponding effects on knowledge levels.

2.3 Data and descriptive statistics

2.3.1 Data construction

We construct a novel data set that combines linked employer-employee data at the establishment level from the German Social Security System with information on the affiliation of establishments to firms. We use a cross-section for the year 2007. The data set contains information on the daily earnings, occupation, part-time or full time status and sociodemographic characteristics of each employee, and the sector, district and total number of employees of each establishment at the reference date 30 June. The data additionally identify establishments that belong to the same firm and distinguish between establishments with and without headquarter function.

We assemble the data from three distinct sources: the Integrated Employment Biographies (IEB) and the Establishment History Panel (BHP) provided by the Institute for Employment Research (IAB), and the 2008 wave of the Dafne data base provided by Bureau van Dijk. As the Dafne data base builds on balance sheet data, the information contained is subject to a small time lag. We choose to use the cross-section for the year 2007 because it therefore has the highest overlap between the Dafne and the establishment level data.

The Integrated Employment Biographies (IEB) are based on records from the German Social Security System. They contain information on all employees subject to social insurance contributions since 1975 and are updated annually. That is, the data cover nearly all private sector employees in Germany, but do not cover civil servants and self-employed workers. The IEB contain information on birth year, gender, nationality, education, occupation, full time or part-time status and daily earnings of each employee. Jacobebbinghaus and Seth (2007) and vom Berge, König, and Seth (2013) provide a detailed description of the structure of the data.[5] The data set, or sub-samples from it, have been used in many recent labor market studies, amongst others by Card, Heining, and Kline (2013), Schmieder, Von Wachter, and Bender (2012), Dustmann, Ludsteck, and Schönberg (2009) and von Wachter and Bender (2006).

The earnings information is very accurate because it is relevant for social security contributions, but earnings are right censored at the highest level that is subject to social security contributions.[6] We impute the earnings information for censored observations following Dustmann, Ludsteck, and Schönberg (2009) and Card, Heining, and Kline (2013). A detailed description of the imputation procedure is provided in Appendix B.2.1. Information on education is not reported for all periods for every individual, but can be inferred from other observations on the same individual. We follow imputation procedure 1 in Fitzenberger, Osikominu, and Völter (2005) and impute missing values for the education variable based on past and future information. We restrict the sample to regular

[5] The paper by vom Berge, König, and Seth (2013) focuses on the Sample of Integrated Labor Market Biographies (SIAB), a 2% random sample drawn from the IEB.

[6] The limit varies from year to year. Values are listed on the homepage of the Research Data Centre of the Institute for Employment Research: `http://fdz.iab.de/de/FDZ_Individual_Data/integrated_labour_market_biographies/SIAB_Working_Tools.aspx`.

full time employees to ensure comparability across employees, i.e., we drop employees in apprenticeship training, part-time employment, or with marginal employment.

The Establishment History Panel (BHP) is a panel data set that contains information on the sector, number of employees and location of all establishments with at least one dependent employee on 30 June of each year. The data has been collected since 1975. Following the regulations of the German Federal Employment Agency (BA), an establishment is defined as the aggregation of all employees in a municipality that are working for the same company in the same sector.[7] Sectors are defined based on the Classification of Economic Activities of the German Statistical Office for the year 2008. The location of establishments is provided at the district level. Germany is divided into 412 districts with around 200,000 inhabitants on average. German districts are roughly comparable to counties in the US. A detailed description of the data set is provided by Gruhl, Schmucker, and Seth (2012).

For data confidentiality reasons, we are only allowed to use information on establishment location at the district level in our empirical analysis, but have restricted access to the full address of establishments reported to the BA for data linkage. We link the Establishment History Panel and the 2008 wave of the Dafne data base using record-linkage techniques (Herzog, Scheuren, and Winkler, 2007). We use the company name and the legal form in the matching process. We exploit a German regulation that requires establishments to register the name of the company that they belong to in their own name when they first apply for an establishment identifier.[8] With this procedure, we are able to identify at least one establishment for about 50% of all firms in the 2008 wave of the Dafne data base. We assume that the establishment located at the firm's official address is the headquarter establishment. In the model, the headquarters decides on problems arising at any of the production locations, so it empirically corresponds to the highest management unit. To identify the establishment where also the headquarters is located, we use the address information at the BA through the Institute for Employment

[7]That is, if a company has several plants in one and the same municipality, all plants in the same sector are assigned the same establishment identifier. Plants in different sectors have distinct identifiers.

[8]Establishment identifiers are issued centrally by the BA and used for the social security notifications of German employers.

Research and match establishment zip code information with the zip code in the firms' official address provided in the Dafne data base. For a subset of firms, more than one establishment is located within the zip code area of the firm's official address. We drop these firms, because we otherwise would have to choose the headquarters discretionarily among the establishments.

Finally, we link the matched establishment-firm data and the IEB based on unique establishment identifiers provided in the IEB and the BHP.

2.3.2 Variable definition

Hierarchical layers. As demonstrated by Caliendo and Rossi-Hansberg (2012), a change in the number of layers leads to non-linear adjustments in the knowledge levels and thus in the remuneration per layer. This implies that only workers in establishments with the same number of layers are comparable. The layers are crucial to map and interpret the data. We use the occupation information in our data to construct hierarchical layers (similar to Caliendo, Monte, and Rossi-Hansberg, 2012). Higher layers acquire the knowledge to solve problems that occur less frequently in the production process: they are characterized by higher levels of expertise or higher management responsibilities. The International Standard Classification of Occupations (ISCO 2008) from the International Labor Organization (ILO) reflects these considerations and classifies occupations according to the level of skill that they require. We use the "major groups" of the ISCO 2008 as basis for the hierarchical layers and assign them as follows:

Layer 3	1 "Managers"
Layer 2	2 "Professionals"
Layer 1	3 "Technicians and associate professionals"
Layer 0	4 "Clerical support workers"
	5 "Service and sales workers"
	7 "Craft and related trade workers"
	8 "Plant and machine operators, and assemblers"
	9 "Elementary occupations"

We drop groups 6 and 0 that refer to agricultural and armed forces occupations.

We map the IAB classification used in our data into the ISCO 2008 classification using official correspondence tables from the German Federal Employment Agency (BA). The assignment of occupations to layers is often unique. Whenever more than one layer is assigned to an occupation, we exploit information on the core area of the occupation provided by the BA. The layer assigned to the core occupation and the mode layer for the occupation coincide for more than 85% of the relevant occupations. In the remaining cases, we match occupations based on the IAB description of occupations and the correspondence tables. We provide a list of these cases and the respective criteria for the assignment of these cases to the different layers in Appendix B.2.2.

The resulting classification is similar to the classification used by Caliendo, Monte, and Rossi-Hansberg (2012). As plausibility check, we replicate their Tables 1 and 2. The results are reported in Appendix B.2.3.

Covariates. We augment the data set using district level information from various sources as empirical analogs of our model parameters. We approximate communication costs using geographic distance and the dissimilarity of dialects. We use the average of the great-circle distances between all municipalities in each district to all municipalities in the respective other district. This way of measuring distance is convenient because it assigns a positive distance to all non-headquarter establishments that are located in the same district as the headquarter establishment. To measure dissimilarity of dialects, we use the index on dialect similarity developed by Falck, Heblich, Lameli, and Südekum (2012) multiplied by (-1).[9] The measure is based on Georg Wenker's famous linguistic atlas. The atlas maps the pronunciation of 66 significant characteristics of the German language. The dissimilarity of dialects within Germany is sizeable: some of the dialects share as few as 11 traits. As Falck, Heblich, Lameli, and Südekum (2012) discuss in detail, the similarity of dialects captures far more than the mere effect of geographic proximity and is a meaningful measure of cultural similarity that facilitates economic

[9]The dissimilarity of dialects is measured at the district level, the same level of aggregation as our data. The geographic delineation of districts in the land *Sachsen* and the land *Mecklenburg-Vorpommern* has changed since the publication of their paper. We map old to new districts and use a simple average of the measure of dialect similarity for these districts.

exchange. Dialect dissimilarity is thus perfectly apt to approximate communication costs in our framework.

We use information on labor productivities and wages per hour at the district level from the German national accounts. We approximate knowledge acquisition costs with the average distance of the largest municipality to universities, including universities of applied sciences. We obtain information on the location and establishment date of all German universities from the Hochschulrektorenkonferenz, the association of universities and other higher education institutions in Germany. Jäger (2013) demonstrates that proximity to universities causally increases the probability that people decide to pursue academic education in Germany. In addition, proximity to universities may help firms train their employees by cooperating on targeted education measures.

2.3.3 Descriptive statistics

Figure 2.1: Distribution of log number of employees, firm-level

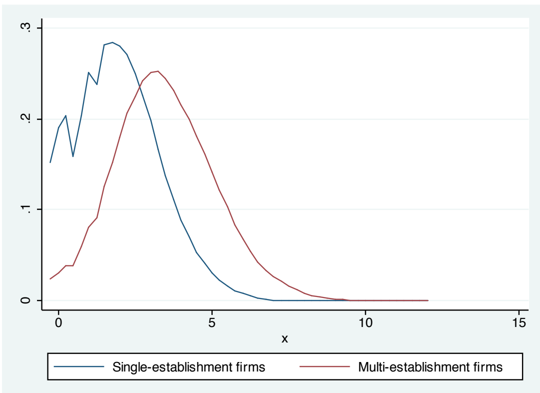
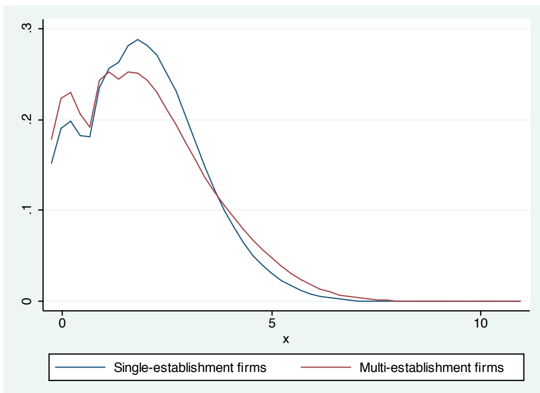


Figure 2.2: Distribution of log number of employees, establishment-level



Kernel density estimates of the log number of full-time employees at the firm and establishment level, separately for multi-establishment and single-establishment firms. X-axis: log number of employees, y-axis: density. Data source: Match of BHP and Dafne, 2007.

11% of the firms in our sample are multi-establishment firms, but they account for 50% of employment.[10] This phenomenon is reflected in Figure 2.1 which displays the distribu-

[10]The nature of the data does not allow to track the share of multi-establishment firms in the population of firms over time. IEB and BHP are panel data sets, but they are matched to a cross-section of firm level data from the Dafne data base. This implies that firms that exit before 2008 or enter after 2008 are not observed, so information on the developments over time is subject to attrition bias.

tion of size, as measured by the number of employees of multi-establishment and single-establishment firms. The distributions are statistically different (Kolmogorov-Smirnov test).

At the establishment level shown in Figure 2.2, the distribution of sizes is much more similar. Distributions are still statistically distinct. Both the mean number of employees per establishment and the variance of establishment size are higher within the group of multi-establishment firms than within the group of single-establishment firms, but differences are considerably smaller than differences at the firm level.

Figure 2.3: Distribution of log remuneration, layer 0

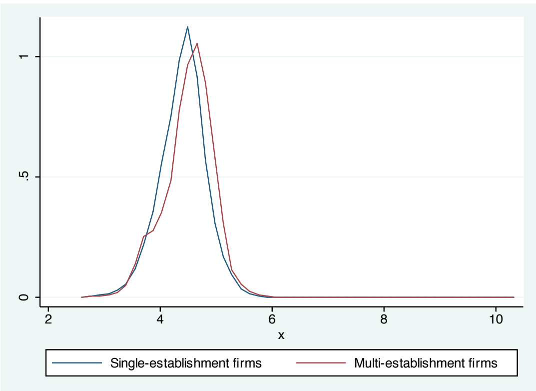


Figure 2.4: Distribution of log remuneration, layer 1

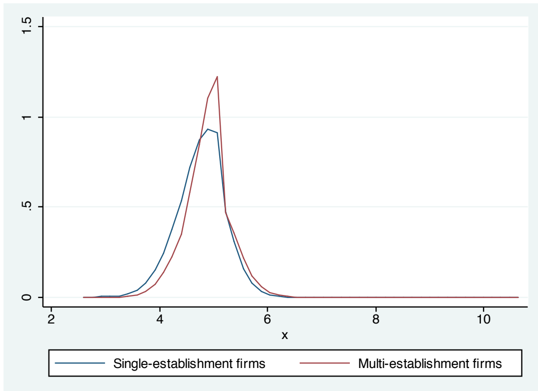


Figure 2.5: Distribution of log remuneration, layer 2

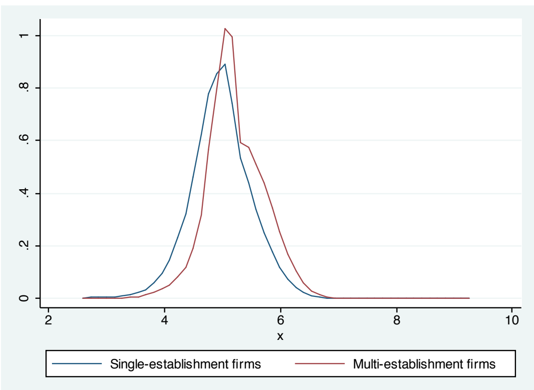
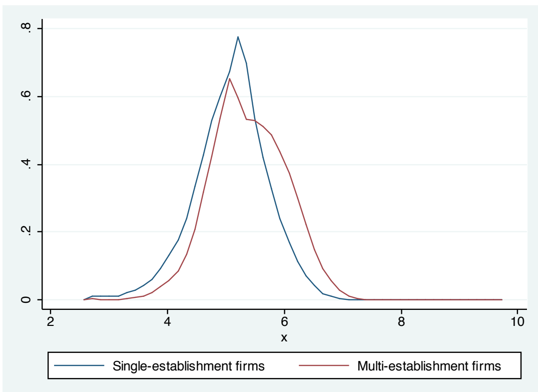


Figure 2.6: Distribution of log remuneration, layer 3



Kernel density estimates of log imputed daily remuneration per employee by layer, separately for multi-establishment and single-establishment firms. Layer 0 comprises employees with job classification in ISCO groups 4, 5, 7-9; layer 1 comprises employees with job classification in ISCO group 3; layer 2 comprises employees with job classification in ISCO group 2; layer 3 comprises employees with job classification in ISCO group 1. Data source: Match of IEB, BHP and Dafne.

The average log remuneration paid by multi-establishment firms is significantly higher than the average log remuneration paid by single-establishment firms. The average log remuneration at single-establishment firms is 4.508, the corresponding average at multi-establishment firms is 4.637. The difference seems small, but corresponds to a disparity of 12.50 Euros per day (4,500 Euro per year), around 12% of average daily remuneration in our data. The distribution of the remuneration across layers by firm type corresponds to this finding: as Figures 2.3 to 2.6 show, the distributions are pretty similar, though the distribution of the remuneration is shifted rightwards for multi-establishment firms compared to single-establishment firms. Distributions are statistically different (Kolmogorov-Smirnov test).

Table 2.1: Distribution of organizational structures, multi-establishment firms

	Layer 0	Layers 0 & 1	Layers 0, 1 & 2	All layers	Layers 0 & 3	Other
# of establishments	52,652	11,545	6,131	8,618	5,867	17,333
Share in total #	51.55%	11.30%	6.00%	8.44%	5.74%	16.97%

The table displays the distribution of establishment organizational structure within multi-establishment firms. Establishments are classified as having layer 0 if they employ workers with job classification in ISCO groups 4, 5, 7-9; as having layer 1 if they employ workers with job classification in ISCO group 3; as having layer 2 if they employ workers with job classification in ISCO group 2; and as having layer 3 if they employ workers with job classification in ISCO group 1. Organizational structures that are chosen by at least 5% of establishments are separately displayed.

The distribution of organizational structures across establishments within multi-establishment firms is displayed in Table 2.1. Most establishments of multi-establishment firms exhibit "regular" organizational structures in a sense that they employ a layer of production workers and one to three layers of managers. Less than a quarter of establishments choose different subsets of layers, such as workers in layer 0 and managers in layer 3.

44.8% of multi-establishment firms are active in only one sector, whereas 55.2% of multi-establishment firms maintain establishments in different sectors.

Table 2.2 presents the average geographic distance between establishments and the headquarter establishment for different numbers of establishments per firm. Firms with fewer establishments tend to have more proximate establishment networks: average distance increases in the number of establishments.

Table 2.2: Average distance of establishments by number of establishments

Firm size, # establishment	2	3	4	5
Mean distance	101.51	137.60	156.31	172.98
# establishments	16,829	6,683	3,805	2,559

The table displays the average geographic distance between establishments and the headquarters at district level for different numbers of establishments per firm.

Table 2.3 presents summary statistics. The upper part of the table presents statistics at the employee level. The lower part of the table displays statistics at the establishment-layer level, the level of the regression analysis. Only multi-establishment firms are contained in the regression sample. Following Card, Heining, and Kline (2013), we exclude employess younger than age 20 and older than age 60 in the regressions.

Table 2.3: Summary statistics, multi-establishment firms

Employee level	N	Mean	SD	Min	Med	Max
Daily remuneration	3,681,744	111.891	44.417	13.694	107.232	2707.596
Log daily imputed remuneration	3,465,320	4.631	0.525	2.617	4.641	10.308
Age	3,681,744	41.054	10.207	14	41	76
Gender (1=female)	3,681,744	0.237	0.425	0	0	1
Education cat. 1	3,582,276	0.100	0.300	0	0	1
Education cat. 2	3,582,276	0.658	0.474	0	1	1
Education cat. 3	3,582,276	0.080	0.271	0	0	1
Education cat. 4	3,582,276	0.162	0.369	0	0	1
Indicator foreign	3,681,219	0.077	0.267	0	0	1
Tenure in days	3,681,744	3192.406	3051.591	0	2251	11868
Distance, district	2,266,412	106.089	149.252	0	20.372	811.698
Log distance , district	2,266,412	3.170	2.126	0	3.062	6.700
Dissimilarity of dialects, district	2,266,412	−53.047	16.409	−66	−66	−11
Average wages, district	3,681,744	21.620	2.805	13.663	21.655	28.574
Labor productivity, district	3,681,744	46.501	7.970	28.720	45.442	81.322
Mean distance to university, district	3,681,744	301.141	50.283	223.787	287.038	496.960
Number of establishments per firm	3,681,744	131.792	546.842	2	7	3352
Establishment-layer level	**N**	**Mean**	**SD**	**Min**	**Med**	**Max**
Average log daily imputed remuneration	167,269	4.557	0.483	2.617	4.547	7.010
Span of control	70,450	9.539	24.687	0.001	3	1197
Distance, district	99,256	159.761	168.962	0	90.483	811.698
Log distance, district	99,256	3.962	1.9854	0	4.516	6.700
Dissimilarity of dialects, district	99,256	−46.933	17.286	−66	−47	−11
Average wages, district	171,657	20.671	2.733	13.663	20.947	28.574
Labor productivity, district	171,657	44.421	7.862	28.720	43.954	81.322
Mean distance to university, district	171,657	304.652	53.474	223.787	288.715	496.960
Log mean distance to university, district	171,657	5.704	0.170	5.411	5.665	6.209
Number of establishments per firm	171,657	239.776	749.907	2	6	3352

The table displays summary statistics at the employee and the establishment-layer level. The sample is restricted to multi-establishment firms. The number of observations varies because the remuneration is not imputed for employees with job classification in ISCO 0 or 6, because education and nationality are missing for some observations, because dialect and distance can only be assigned when the location of the firm's headquarters is known and unambiguous and because the span of control is only defined for higher layers. Information marked as "district" information is matched using information on the location of the establishment at district level.

Comparing the statistics in the two different panels, we find that aggregating the individual remuneration at the establishment layer level decreases the variance in the log imputed daily remuneration. This is intuitive: the effect of extreme values is dampened by taking the average. The average and median of the geographic distance and the dissimilarity of dialects is higher in the lower panel, which indicates that establishments that are proximate to the headquarters tend to be larger.

2.4 Empirical strategy

The data allow to test two hypotheses derived from the model:

Testable hypotheses.

1. *For a given establishment organizational structure, the remuneration of employees in layer ℓ of establishment j increases in the communication costs $\tilde{\theta}_{ij}$ between establishment j and the headquarters 0.*
2. *For a given establishment organizational structure, the span of control of employees in establishment j increases in the communication costs $\tilde{\theta}_{ij}$ between establishment j and the headquarters 0 and decreases in wages $\tilde{w}_j$ and knowledge acquisition costs $\tilde{c}_j$.*

The tilde "$\sim$" denotes the empirical analog of a model parameter.

The comparative statics only hold for a given number of layers. It is necessary to ensure that all establishments within the regression sample have the same organizational structure. The downside of this requirement is that many observations are lost in the empirical analysis. Furthermore, the comparative statics hold given the problem arrival rate λ. This implies that they are most likely to hold within homogeneous groups of establishments with similar production processes. For the purpose of our analysis, we therefore include sector dummies in the pooled regressions.

To test Hypothesis 1, we study the effect of communication costs on the average remuneration of employees per layer, which theory predicts to be unambiguously positive. We use district fixed effects in this set of regressions to capture the effect of wages and knowledge acquisition costs. As explained above, these variables affect remuneration. As they exert both a direct effect on the remuneration and an indirect effect via knowledge, it is not possible to test their effect on knowledge in the regressions using remuneration as dependent variable. Using district dummies implies that the effect of labor productivity on remuneration is not identified either.

We specify the following regression equations:

$$\tilde{y}_{\ell ij} = \beta_0 + \beta_1 \tilde{\theta}_{ij} + \alpha_s + \alpha_j + u_{ij} \tag{2.22}$$

$$\tilde{y}_{\ell ij} = \beta_0 + \beta_1 \tilde{\theta}_{ij} + \alpha_i + \alpha_j + u_{ij} \tag{2.23}$$

$\tilde{y}_{\ell ij}$ denotes the average remuneration of employees in layer ℓ of firm i in the establishment at j. As outlined above, we use bilateral distance and the similarity of dialects as measures for observed communication costs $\tilde{\theta}_{ij}$. α_j is a district fixed effect which captures differences in wages, knowledge acquisition costs, but also labor productivities as well as other location-specific characteristics that may influence remuneration, but are omitted in our theoretical analysis. α_s is a sector dummy to capture differences across production processes. We additionally run regressions with firm fixed effects α_i. Firms may be heterogeneous with respect to their overall knowledge level $\bar{Z}_H$ that may influence knowledge allocation at all locations. The firm fixed effect isolates within firm differences in the dependent variable and the covariates across locations, so results are robust to heterogeneity across firms.

In a second set of regressions, we focus on the span of control as outcome variable. The span of control allows to test the effect of all model parameters on knowledge levels. We analyze the impact of the model parameters on the span of control using

$$\tilde{y}_{\ell ij} = \gamma_0 + \gamma_1 \tilde{\theta}_{ij} + \gamma_2 \tilde{w}_{ij} + \gamma_3 \tilde{c}_{ij} + \gamma_1 \tilde{A}_{ij} + \alpha_s + u_{ij} \tag{2.24}$$

$$\tilde{y}_{\ell ij} = \gamma_0 + \gamma_1 \tilde{\theta}_{ij} + \gamma_2 \tilde{w}_{ij} + \gamma_3 \tilde{c}_{ij} + \gamma_1 \tilde{A}_{ij} + \alpha_i + u_{ij} \tag{2.25}$$

$\tilde{y}_{\ell ij}$ is the span of control of employees in layer ℓ in firm i at establishment j. $\tilde{\theta}_{ij}$, $\tilde{w}_{\ell ij}$, $\tilde{c}_{\ell ij}$ and $\tilde{A}_{\ell ij}$ are defined as above. As before, we run both pooled regressions with sector fixed effects α_s and regressions including firm fixed effects α_i to take heterogeneity across firms into account.

We exclude establishments with headquarter function from the analysis. We cannot distinguish whether an employee of these establishments does or does not have responsibilities for the whole firm. Results including establishments with headquarter functions may thus be biased, as employees with establishment-level responsibilities should be remunerated differently than employees having firm-level responsibilities.

We run regressions separately for each layer. The theoretical model recommends this strategy, as the effect of communication costs on knowledge levels at different layers is not linear. Empirically, we have tested whether pooling layers within establishments is admissible, i.e. we have tested whether coefficients are statistically different across layers. The null hypotheses of equal coefficients were rejected at very high significance levels.

It is important to note that the regression results are conditional correlations and do not suffice to establish causality between the location characteristics and the organization of multi-establishment firms. Though the regressions control for heterogeneity between firms, they do not take into account that both location and organizational structure of the establishments are self-selected. Omitted variables that are relevant for the firm, but unobservable to us may thus bias the results.

2.5 Results

2.5.1 Remuneration

2.5.1.1 Bivariate correlations at employee level

Table 2.4 displays bivariate correlations between log observed remuneration per employee and communication costs by layer. Communication costs are measured with bilateral distance and the dissimilarity of dialects. We calculate log distance as the log of observed distance plus 1 to be able to include single-establishment firms in Panel A. The model predicts that higher communication costs exert a positive effect on knowledge in an establishment given local wages and knowledge acquisition costs. The two covariates are thus expected to have a positive influence on wages.

The left part of the table is based on observed log remuneration. The right part is based on the residuals of a regression of log remuneration on district dummies, so it takes regional differences in wages and knowledge acquisition costs into account.

Table 2.4: Bivariate correlations, log individual remuneration and communication costs

Observed remuneration					Residual remuneration				
A. All establishments									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	0.065***	0.110***	0.140***	0.074***	Log distance	0.070***	0.125***	0.139***	0.069***
Dialect	0.018***	0.055***	0.078***	0.014***	Dialect	0.031***	0.076***	0.094***	0.032***
B. Establishments of multi-establishment firms									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	-0.148***	-0.075***	-0.071***	-0.176***	Log distance	-0.041***	0.046***	0.064***	-0.005
Dialect	-0.102***	-0.053***	-0.051***	-0.148***	Dialect	-0.049***	0.014***	0.040***	-0.036
C. Establishments of multi-establishment firms without headquarter function									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	-0.132***	-0.098***	-0.074***	-0.177***	Log distance	-0.032***	0.025***	0.061***	-0.005
Dialect	-0.095***	-0.064***	-0.055***	-0.154***	Dialect	-0.044***	0.005*	0.034***	-0.036***
D. Manufacturing establishments of multi-establishment firms without headquarter function									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	-0.139***	-0.099***	-0.067***	-0.139***	Log distance	0.035***	0.060***	0.188***	0.211***
Dialect	-0.097***	-0.061***	-0.043***	-0.126***	Dialect	0.024***	0.037***	0.147***	0.153***

The table displays bivariate correlations between log observed remuneration at the individual level per layer and communication costs, measured using log distance and the dissimilarity of dialects. We calculate log distance as the log of observed distance plus 1 to be able to include single-establishment firms in Panel A. The left part of the table uses log observed remuneration in levels. The right part of the table uses the residual from a linear regression of log observed remuneration on district dummies. Panel A includes all establishments. Panel B includes only establishments of multi-establishment firms. Panel C drops establishments with headquarter function. Panel D only uses establishments in the manufacturing sector. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The first panel (A) displays correlations between the remuneration and communication costs for all establishments, whether they belong to single- or multi-establishment firms. Both distance and the dissimilarity of dialects exert a positive effect on observed and residual remuneration that is significant at the 0.1% level. Panels B to D gradually restrict the sample, first to multi-establishment firms, then only to establishments without headquarter function and finally to establishments without headquarter function in the manufacturing sector. For these panels, results differ between the two dependent variables. For observed remuneration, distance and dialect have a significantly negative effect once only establishments of multi-establishment firms are included in the sample. For residual remuneration, effects are significantly positive—consistent with expectations—for layer 1 and 2 throughout panels and for all layers in the lowest panel. The correlations for imputed remuneration are similar and are displayed in Table B.6 in section B.2.4 of the Appendix.

These results highlight important patterns in the data. Bivariate correlations flip signs in the left part of the table once only multi-establishment firms are used in the analysis.

This observation is a strong indication that multi-establishment firms are a non-random sub-sample of all firms and that they tend to pay higher remuneration, consistent with what we saw in Figures 2.3 to 2.6. The positive coefficient of distance in Panel A is likely to be upward biased: distance is zero for single-establishment firms and is likely to take up the (omitted) general difference in remuneration between single and multi-establishment firms. The differences in coefficients estimated on observed and residual remuneration in Panels B to D are consistent with non-random location of multi-establishment firms. Distant locations seem to have relatively low wages and knowledge acquisition costs levels, but given wages and knowledge acquisition costs, employees tend to earn a higher remuneration. Last, the results in the right part of panel D are most consistent with model predictions. The size of the effects for layers 1 to 2 increases between the panels C and D. This implies that differences in sector characteristics as captured by λ in the model may affect patterns of remuneration: results are stronger if a homogeneous sample of establishments is chosen.

2.5.1.2 Regressions at the layer level

Table 2.5 displays results of regressions of the average log imputed remuneration per layer on measures of communication costs and district dummies for different organizational structures. Exemplarily for all layers, results on the highest layer within each organizational structure are reported. The table displays two types of results: results without firm fixed effects, but including sector dummies, and results including firm fixed effects, but not sector dummies, as prescribed by the model.

We find that results are consistent with model predictions in regressions without firm fixed effects, except for organizations with all layers where coefficients are insignificantly different from zero. Both distance and dissimilarity of dialects exert a positive effect that is significant at the 1% level for organizations with layer 0 and layers 0 and 1 and at the 5% level for organizations with layers 0 to 2. Quantitatively, an increase in distance by one standard deviation is associated with an increase in remuneration by 0.126-0.158 standard deviations. The effect of an increase in the dissimilarity of dialects by one standard deviation is similar despite the difference in coefficient sizes, as

Table 2.5: Regression results for the average remuneration per layer

	Layer 0		Layer 0 & 1		Layer 0, 1 & 2		All layers	
Highest layer	0	0	1	1	2	2	3	3
Log distance	0.033***	0.001	0.025**	−0.001	0.034*	0.024+	−0.008	−0.001
	(0.008)	(0.003)	(0.009)	(0.008)	(0.016)	(0.013)	(0.017)	(0.012)
R-squared	0.345	0.047	0.568	0.211	0.641	0.352	0.473	0.331
Firm fixed effects	N	Y	N	Y	N	Y	N	Y
District dummies	Y	Y	Y	Y	Y	Y	Y	Y
Sector dummies	Y	N	Y	N	Y	N	Y	N
# firms	2,243	2,243	518	518	319	319	364	364
# observations	17,138	17,138	2,499	2,499	1,308	1,308	1,328	1,328
Dissimilarity of dialects	0.002**	−0.000	0.002**	−0.001	0.002*	0.001	0.000	0.000
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.345	0.047	0.568	0.212	0.641	0.350	0.472	0.331
Firm fixed effects	N	Y	N	Y	N	Y	N	Y
District dummies	Y	Y	Y	Y	Y	Y	Y	Y
Sector dummies	Y	N	Y	N	Y	N	Y	N
# firms	2,243	2,243	518	518	319	319	364	364
# observations	17,138	17,138	2,499	2,499	1,308	1,308	1,328	1,328

The table displays the results of regressions of average log imputed daily remuneration per layer on measures of communication costs, district dummies and sector dummies or firm fixed effects by establishments grouped according to their organizational structure. Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the variation in dissimilarity of dialects is larger than the variation in log distance by one order of magnitude. Once firm fixed effects are included, coefficients are generally insignificantly different from zero, except for establishments with layers 0 to 2. For these establishments, the effect of distance is significantly positive at the 10% level and similar in magnitude as the coefficient in the pooled regression. This finding is consistent with ex-ante expectations.

Restricting the sample to manufacturing firms leaves less than 700 observations per group. Results are generally insignificant, so we do not report them here.

The pattern of coefficients demonstrates that firms which invest in locations with higher bilateral communication costs tend to pay higher remuneration. The effect is not robust to the inclusion of firm fixed effects, except for establishments with layers 0 to 2, so it is difficult to distinguish between communication costs and firm heterogeneity as driving force of the positive coefficients in pooled regressions. The lack of significance in the fixed effects regressions may be attributable to the small number of establishments that are used in the analysis. On average, firms included in the sample have only 4 to 8 establishments, so a high degree of within-firm variation in remuneration between establishments is required to obtain a strong effect in the fixed effects regression. Our current technique to identify the headquarter establishment excludes all firms with several establishments in the headquarter zip code area. This implies that particularly those firms with high degrees of variation, i.e. with establishments that are very proximate and establishments at more distant locations, are excluded from the sample. We therefore intend to improve the procedure in future work. Alternatively, it is possible that firms are not able to support high degrees of within-firm heterogeneity in remuneration for similar activities, either due to institutional constraints such as union agreements or due to equity concerns of their employees.

We next analyze the second establishment characteristic, the span of control.

2.5.2 Span of control

Analyzing the span of control allows to assess both the effect of differences in communication costs and the impact of other location characteristics such as local wages or knowledge acquisition costs.

Table 2.6 presents the regression results. The table is structured similarly to Table 2.5. Exemplarily, the regression results for the highest layer are displayed for each organizational structure. We distinguish organizational structures by the number of layers to obtain a larger sample for each regression. We tested whether pooling establishments with the same number of layers, but a different set of layers (e.g. layers 0 and 1 or layers 0 and 2) is admissible, and failed to reject the null of equal coefficients.

The regression results draw an inconclusive picture. The effect of distance and the dissimilarity of dialects is significantly negative for establishments with two layers, but significantly positive for establishments with three layers. It is insignificant in regressions for establishments with all layers. Wages and labor productivity do not have a significant effect. The knowledge acquisition costs are mostly insignificant, but exert a significantly positive effect on the span of control of employees in the highest layer of establishments with all layers, inconsistent with model predictions.

Table 2.6: Regression results for the span of control

	2 Layers		3 Layers		All layers	
	1	1	2	2	3	3
Log distance to HQ	-0.822^{+}	-0.783^{*}	0.749^{*}	0.547	−0.923	−1.200
	(0.427)	(0.390)	(0.364)	(0.519)	(1.119)	(1.425)
Hourly wages	0.158	0.232	0.181	−0.022	−0.249	0.140
	(0.167)	(0.179)	(0.145)	(0.081)	(0.589)	(0.633)
Hourly labor productivity	0.016	0.020	0.073	0.046	0.204	0.067
	(0.055)	(0.049)	(0.050)	(0.044)	(0.239)	(0.277)
Log average distance to university	1.592	3.136	1.689	1.691	17.124^{*}	15.492^{*}
	(1.802)	(2.104)	(1.530)	(1.049)	(8.050)	(7.715)
Constant	5.981	−8.555	-21.340^{+}	−10.250	-97.961^{*}	-78.202^{+}
	(10.804)	(12.623)	(11.733)	(8.544)	(46.293)	(40.797)
R-squared	0.245	0.003	0.057	0.003	0.127	0.010
Firm fixed effects	N	Y	N	Y	N	Y
Sector dummies	Y	N	Y	N	Y	N
# firms	937	937	525	525	364	364
# observations	4,838	4,838	2,305	2,305	1,328	1,328
Dissimilarity of dialects	-0.052^{+}	-0.054^{*}	0.052^{+}	0.036	−0.084	−0.156
	(0.029)	(0.026)	(0.029)	(0.028)	(0.085)	(0.101)
Hourly wages	0.192	0.270	0.159	−0.038	−0.227	0.154
	(0.168)	(0.185)	(0.139)	(0.086)	(0.566)	(0.604)
Hourly labor productivity	0.016	0.020	0.072	0.044	0.208	0.068
	(0.055)	(0.049)	(0.050)	(0.043)	(0.245)	(0.283)
Log average distance to university	0.534	2.216	2.768	2.423^{+}	15.567^{*}	13.238^{*}
	(1.863)	(2.085)	(1.765)	(1.463)	(7.675)	(6.475)
Constant	−5.709	−10.119	-21.780^{+}	−9.879	-93.827^{*}	-77.426^{+}
	(11.510)	(12.860)	(12.069)	(8.264)	(46.702)	(40.175)
R-squared	0.245	0.003	0.057	0.003	0.127	0.013
Firm fixed effects	N	Y	N	Y	N	Y
Sector dummies	Y	N	Y	N	Y	N
# firms	937	937	525	525	364	364
# observations	4,838	4,838	2,305	2,305	1,328	1,328

Dependent variable: span of control, defined as $s_j = n^{\ell-1}_{L,j}/n^{\ell}_{L,j}$. Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Overall, it is thus difficult to draw conclusions based on the regressions on the span of control. We have explored restricting the sample to manufacturing establishments and thus to a more homogeneous group of firms, and report the results of these regressions in Table B.7 in section B.2.4 of the Appendix. A larger average distance to universities is associated with a lower span of control for establishments with two layers in these regressions, as predicted by the model. The sample size is very small however, with less than 150 firms per group. Significance levels are thus even lower than in the main regressions.

2.6 Discussion

The results in the previous section demonstrate that knowledge transmission costs matter not only at the macro level and internationally, but also at the firm level and within countries. The evidence indicates that multi-establishment firms adapt the organizational structure of each single establishments to the communication costs with the headquarters as well as to local factor input costs. These findings help to understand the structure of multi-establishment firms' investment, which is of major importance as these firms account for a significant share of the economy, as argued in subsection 2.3.3 and by Bernard and Jensen (2007).

In addition, the results in this chapter contribute to a better understanding of the heterogeneity in remuneration of observationally equivalent workers at different employers, so-called employer wage premiums. Despite a large body of literature, the causes and mechanisms behind these premiums are still little understood. At the same time, the well-documented increase in wage inequality across countries and long-term consequences of inequality call for research on the wage setting process to assess the need and scope of policy interventions (e.g. Dustmann, Ludsteck, and Schönberg, 2009; Card, Heining, and Kline, 2013). In line with Caliendo and Rossi-Hansberg (2012), this chapter argues that understanding the organization of firms and the function of individual employees within firms is a fruitful avenue to comprehend some factors underlying the empirical evidence.

Table 2.7 displays summary statistics on the observed log imputed remuneration (Panel A) and the residual remuneration (Panel B) from a regression of log imputed remuneration on district dummies. Three interesting facts emerge, which underline that understanding firms' organization may help understand variation in remuneration: first, multi-establishment firms pay higher remuneration on average than single-establishment firms (Panel A). Second, the between-variation in residual log remuneration (Panel B.1, column 4) is larger than the within variation for all firms (Panel B.1, column 5), but smaller than the within variation for multi-establishment firms. To put it differently, the remuneration varies more strongly within multi-establishment firms than between them. Third, this effect is reversed at the establishment level (Panel B.2). The variation between

Table 2.7: Within and between variation in remuneration across firm groups

	(1)	(2)	(3)	(4)	(5)
Firm group	N	Mean	SD overall	SD between	SD within
A. Observed log imputed remuneration					
Overall	7,142,649	4.562	0.511	0.405	0.333
Multi-establishment firms	3,483,767	4.630	0.526	0.370	0.345
B. Residual log imputed remuneration (covariates: district dummies)					
B.1. Firm level					
Overall	6,869,943	0	0.391	0.354	0.301
Multi-establishment firms	3,386,088	0	0.390	0.292	0.313
B.2. Establishment level					
Overall	6,869,943	0	0.391	0.369	0.290
Multi-establishment firms	3,386,088	0	0.390	0.371	0.291

The table displays summary statistics on variation between and within firms and establishments by firm groups. Panel A displays the mean of the observed log imputed remuneration for all and multi-establishment firms as well as the measures on the overall variance, the variance between firms and the variance within firms. Panel B displays the same measures for the residuals from a regression of log imputed remuneration on district dummies. Panel B distinguishes variation between and within firms from variation between and within establishments.

and within establishments is virtually identical for both the overall group of firms and the group of multi-establishment firms. Variation between establishments plays a relatively larger role than variation within establishments. We expect that a better understanding of the organizational structure of firms helps to interpret these patterns.

2.7 Conclusion

This chapter develops a model of the optimal organizational structure of multi-establishment firms. The model postulates that the optimal organizational structure of establishments within multi-establishment firms depends on the bilateral costs of communication between an establishment and the headquarters of the firm as well as local factor input costs. Using a novel data set with linked employer-employee data on German multi-establishment firms, we provide empirical evidence that is in line with theoretical predictions.

Notwithstanding technological progress in the communication sector, communication costs matter. With the increasing importance of knowledge for today's economies and the ongoing innovations in communication technologies, knowledge transmission within and across countries and firms thus remains a fruitful avenue for future research.

Chapter 3

Multinational firms and tax havens*

3.1 Introduction

Tax havens are typically small, well-governed states that impose low or zero tax rates on foreign investors (Dharmapala and Hines, 2009). Multinational enterprises (MNEs) are widely believed to use tax havens to avoid taxation. Sophisticated tax planning strategies involving tax havens have received considerable attention in the media (e.g. Drucker, 2010; The Economist, 2012; Lucas, Jopson, and Houlder, 2012), and tax havens have repeatedly been in the focus of national and international policy measures. To name a few examples, the OECD launched the "Initiative on Harmful Tax Competition" in 1998 to pressure tax havens to abolish harmful tax provisions and practices. France announced plans to introduce a 50% tax on income earned by French affiliates in tax havens in February 2010. The U.S. House Committee on Ways and Means held a background hearing on the transfer pricing practices of U.S. taxpayers, with an emphasis on income reallocation to offshore tax havens.[1] Most recently, the United Kingdom Parliament Government Accounts Committee held widely publicized hearings in 2012 concerning the use of tax havens by foreign multinational firms operating in Britain, and partly as a consequence

*This chapter is based on joint work with James R. Hines Jr. and Monika Schnitzer.

[1]The Staff of the Joint Committee on Taxation prepared a detailed report including six case studies of the tax avoidance practices of large US firms (Joint Committee on Taxation, 2010).

the G-20 finance ministers in their February 2013 meeting pledged to take collective action to develop measures to address tax base erosion and profit shifting.

What drives the policy interest is the concern that tax havens are widely used to avoid tax obligations, particularly those due to high-tax countries. It is clear, however, that corporations cannot use tax haven operations to avoid taxes easily and comprehensively, as reflected in persistent significant corporate tax collections by high-tax countries and the fact that many multinational firms choose not to have tax haven affiliates. Consequently it may be the case that substantial tax avoidance opportunities through the use of tax haven operations are selectively available only to certain firms, industries, or activities.

The purpose of the final chapter of this thesis is to identify factors associated with a firm's demand for tax haven operations, using data on German multinational firms. The evidence indicates that, among manufacturing firms, those that are larger and more productive are the most likely to have tax haven affiliates. Notably, manufacturing firms whose non-haven foreign operations are located in high-tax countries are more likely than others to have tax haven affiliates. These patterns are consistent with a simple model of profit reallocation, in which some firms find that the benefits of being able to reallocate income from high-tax jurisdictions exceeds the cost of establishing tax haven affiliates - whereas other multinational firms, with fewer profits to reallocate or fewer taxes to save, do not. Among firms in service industries tax haven use is less closely associated with high foreign tax rates, quite possibly reflecting the more limited scope for taxable income reallocation among firms in service industries. Much of the available evidence on the determinants of tax haven use by multinational firms comes from studies of U.S. firms, whose tax treatment differs from those of firms based in almost any other major capital exporting country. Income earned by foreign subsidiaries of U.S. firms is subject to U.S. taxation when repatriated, at which time U.S. taxpayers can claim credits for income taxes paid to foreign governments. This system reduces the incentive to use tax haven operations to avoid foreign taxes, since foreign tax reductions entail fewer foreign tax credits, and therefore greater U.S. tax liability, when income is ultimately repatriated. The system does, however, encourage the use of tax havens to the extent that they facilitate deferral of home country taxes, such as by serving as coordination centers to direct foreign profits to new foreign investments.

Germany taxes only 5 percent of the active foreign business profits of its resident corporations.[2] In this respect the German tax system is similar to those in the United Kingdom, Japan, France, Canada, Italy, and most OECD countries, particularly the major capital exporters. German firms consequently have strong incentives to avoid foreign taxes, since foreign tax savings do not entail greater home country taxes when income is repatriated. Furthermore, German firms do not have incentives to structure their foreign operations in ways that avoid repatriating income. Therefore, the tax incentives for German firms to establish tax haven affiliates are likely to differ from those of U.S. firms, and bear strong similarities to those of other G-7 and OECD firms. To gauge the impact of foreign taxation on tax haven investment by German firms, the chapter estimates a linear probability model of tax haven investment using the sizes of a firm's foreign and domestic activities, and the firm's R&D intensity, as additional control variables. The empirical strategy uses the panel nature of the data to account for the fact that the tax rates a firm faces at its foreign locations may be endogenous to its decision to invest in a tax haven: specifically, the estimation relies on statutory foreign tax changes subsequent to a base year. The estimates indicate that a one percentage point higher foreign tax rate is associated with a 0.3 percentage point greater likelihood of having a tax haven affiliate, presumably reflecting the greater benefit of reallocating taxable income from high-tax countries.

This chapter is related to two strands of the literature, one on the use of tax havens by multinational firms, and the other on profit reallocation. Harris, Morck, Slemrod, and Yeung (1993) use a five-year panel on 200 large U.S. manufacturing firms and find that U.S. tax liabilities of U.S. firms holding affiliates in Ireland or one of the four low tax "dragon" Asian countries are systematically lower than those of U.S. firms without such activities. Hines and Rice (1994) analyze a cross-section of country level data on the activities of U.S. multinational firms, finding that U.S. multinationals report disproportionate shares of profits in tax havens, which suggests that income may be

[2] For details on the taxation of foreign profits of German multinational firms, please refer to Appendix C.2.

reallocated for tax purposes. Grubert and Slemrod (1998) use a cross-section of data and estimate a joint model of the investment and profit shifting decision of U.S. multinationals in Puerto Rico which, due to its special status, can serve as a tax haven for U.S. firms. They find that firms with intangible assets are more likely than others to invest in Puerto Rico.

Desai, Foley, and Hines (2006), who are closest to our analysis, use an affiliate-level data set on U.S. multinationals' foreign activities in four years between 1982 and 1999. They estimate a logit model of tax haven investment given parent characteristics and take into account the endogeneity of the foreign non-haven tax rate due to simultaneity of a parent's location decisions. They find a negative effect of the average foreign non-haven tax rate on the probability of investing in a tax haven, interpreting their finding as evidence of the impact of incentives induced by the ability to defer home country taxation of unrepatriated foreign profits. Thus, it is particularly interesting to compare the U.S. evidence with the tax haven investment behavior of firms that are subject to a tax exemption regime, as German firms are, which have clear incentives to use tax haven operations to reallocate taxable income.

The literature on international profit shifting is vast, and for brevity, we will restrict our review to a few recent examples for the different strands of the literature. That taxes matter for profit shifting of international firms has been documented by Huizinga and Laeven (2008), amongst others. They use a cross-section of European MNEs and find evidence for substantial profit shifting between different countries in Europe, which fits international profit shifting incentives that arise from tax differences both between the parent and host country and among different affiliate locations. Weichenrieder (2009) analyzes a panel data set of German inbound and outbound FDI and identifies empirical patterns that are consistent with profit shifting in both cases.

With respect to different profit shifting strategies, Clausing (2001, 2003, 2006) provides empirical evidence that taxes exert a substantial impact on transfer prices and intra-firm trade flows of U.S. firms. Dischinger and Riedel (2011) offer evidence from a panel data set of European firms that MNEs prefer locating intangible assets at low-tax locations, arguably doing so because they are able to choose favorable transfer prices for intangible assets. Karkinsky and Riedel (2012) report similar findings with respect to patent location within MNEs. Egger, Eggert, Keuschnigg, and Winner (2010) compare the debt-to-asset ratios of domestically and foreign owned European firms and identify a gap in the ratios which is systematically related to corporate tax rates. Buettner, Overesch, Schreiber, and Wamser (2009) provide further evidence on tax-motivated choice of capital structure using a panel data set of German MNEs. Using the same data, Buettner and Wamser (2013) analyse the use of intrafirm-loans for profit shifting, but find that they have rather small tax revenue effects. Weichenrieder and Mintz (2010) as well as Wamser (2011) show, using data on German MNEs, that firms tactically locate their direct and indirect affiliates and strategically use ownership chains in a way that facilitates tax avoidance.

The chapter is structured as follows. First, we present our theoretical model and derive the hypotheses for our empirical analysis. Section 3.3 describes the data used in the analysis and provides descriptive statistics; section 3.4 outlines our empirical approach. Section 3.5 summarizes our regression results, section 3.6 provides some robustness checks and section 3.7 discusses the implications. The last section concludes.

3.2 Incentives to establish tax haven operations

In this section we lay out a stylized theoretical framework to describe the incentives of a multinational firm to invest in a tax haven and to derive the empirical predictions to be tested later on. For this purpose, we study a multinational firm that can invest in a range of countries $i = 0, ..., n$, including a tax haven, which is denoted as country 0. Starting a foreign affiliate involves fixed set-up cost c_i. Let ρ_i denote before-tax profits earned in country i by the affiliate once it is installed. Reported profits are taxed at rate τ_i in country i. Without loss of generality we assume that $\tau_0 = 0$, i.e. there is no taxation in the tax haven.

Firms can reallocate an amount ψ_i of their actual profits in country i to a country that taxes reported profits at a lower tax rate, most notably to the tax haven country, for example by adjusting their transfer prices. This is possible only at some cost. Firms may need to set up additional facilities to make transfer prices seem plausible, inefficient relocation of production and intra-firm trade may be needed to arrange income reallocation, and transaction costs are incurred, like legal expenses. We assume that income reallocation gets increasingly expensive as the amount reallocated increases relative to income earned in country i. Following Hines and Rice (1994), these income reallocation costs are assumed to be $(a/2)({\psi_i}^2/\rho_i)$.[3] Parameter a captures how much the cost of income reallocation increases with the amount reallocated. Note that a is a firm-specific parameter because income reallocation costs vary with firm-specific characteristics such as the R&D intensity of a firm. As indicated above, firms with more R&D activities and larger intangible assets have been shown to be more easily able to reallocate income due to the lack of comparable market prices. The reported profit in country i, π_i, after incurring fixed cost c_i, is thus

$$\pi_i = \rho_i - \psi_i - \frac{a}{2}\frac{{\psi_i}^2}{\rho_i} \quad . \tag{3.1}$$

Consider now the option of setting up an affiliate in a tax haven at cost c_0. To save on notation, we set $\rho_0 = 0$ and let c_0 capture the net cost of investing in a tax haven, after deducting any profits that arise genuinely in this country. Note that due to anti-deferral regulations, firms may have to incur considerable fixed costs to set up a tax haven affiliate that may be used for profit reallocation purposes.[4] For $c_0 < 0$, the multinational has an interest in investing in a tax haven country, and does so, independent of investments in other countries. This interest could arise from plans to reallocate income from the home country. However, since our data set contains information on parent firms from only one home country it is not possible to gauge the impact of this tax incentive empirically. Thus, we focus on multinationals that invest in non-haven countries as well.

In order to evaluate the incentive to invest in a tax haven, consider first the situation of a multinational with a tax haven affiliate. The firm chooses in which other countries to locate affiliates and how much of their profits to reallocate to the tax haven. Thus, the investor's maximization problem, given that it has a tax haven affiliate, is

[3]For simplicity, we assume that the cost of reallocating income to a tax haven and to another non-haven country are equal. This assumption does not affect the main intuition of the model, but renders notation far more tractable.

[4]For details on the anti-deferral regulations, please refer to Appendix C.2.

$$\max_{d_i,\psi_i} \sum_{i=1}^{n} d_i \left[\psi_i + (1-\tau_i)(\rho_i - \psi_i - \frac{a}{2}\frac{\psi_i{}^2}{\rho_i}) - c_i \right] \tag{3.2}$$

with $d_i \in \{0,1\}$, s.t.

$$\rho_i - \psi_i - \frac{a}{2}\frac{\psi_i{}^2}{\rho_i} \geq 0 \ \forall i = 1, ..., n \ . \tag{3.3}$$

The following proposition describes the solution to this maximization problem.

Proposition 1. *Suppose the investor has a tax haven affiliate. Then the optimal amount of profit shifting is*

$$\psi_i^{*th} = \frac{\tau_i \rho_i}{a(1-\tau_i)} \quad if \quad \tau_i \leq 1 - \sqrt{\frac{1}{2a+1}} \tag{3.4}$$

and

$$\bar{\psi}_i = \frac{\rho_i}{a}\left(\sqrt{2a+1} - 1\right) \quad otherwise. \tag{3.5}$$

Proof. See Appendix C.3. □

For ease of presentation, in the following we restrict consideration to cases in which the parameter condition in Equation (3.4) is satisfied such that condition (3.3) holds for the optimal ψ_i^{*th}, and discuss deviations only when necessary for the results.

Let countries be numbered such that country $i = 1$ yields the highest after-tax profit, including the fixed cost of setting up the affiliate, and country $i = n$ yields the lowest profit. Then the multinational chooses $d_i = 1$ for all countries $i = 1, ..., \tilde{n}$, where $\tilde{n}$ is determined by the condition

$$\begin{aligned} &\psi_{\tilde{n}} + (1-\tau_{\tilde{n}})(\rho_{\tilde{n}} - \psi_{\tilde{n}} - \frac{a}{2}\frac{\psi_{\tilde{n}}^2}{\rho_{\tilde{n}}}) - c_{\tilde{n}} \geq 0 \\ &\qquad > \psi_{\tilde{n}+1} + (1-\tau_{\tilde{n}+1})(\rho_{\tilde{n}+1} - \psi_{\tilde{n}+1} - \frac{a}{2}\frac{\psi_{\tilde{n}+1}^2}{\rho_{\tilde{n}+1}}) - c_{\tilde{n}+1} \ . \end{aligned} \tag{3.6}$$

Consider now the multinational's situation if it has no tax haven affiliate. In this case, profit-shifting has to be directed to the country charging the lowest tax rate among those in which the multinational holds an affiliate.[5] Let $\underline{\tau}$ denote the minimum of all tax rates

[5] We assume for simplicity that the multinational shifts profits to one country only. Giving up this assumption would yield computation far more complicated, but would not affect our results qualitatively.

charged in countries in which the multinational invests. In Appendix C.3, we derive the optimal amount of profit shifting ψ_i^{*nth} and we describe the optimal number of countries $\hat{n}$ in which to set up a foreign affiliate. It is straightforward to show that $\tilde{n} \geq \hat{n}$, since the profits realized from each country are potentially larger if it is possible to reduce taxes by reallocating income to a tax haven.

For the multinational, investing in a tax haven is worth the set up cost c_0 if and only if it results in higher net profits. This incentive is captured in the following Proposition:

Proposition 2. *Let Inc_{th} denote the net benefit ("Incentive") from investing in a tax haven. Then investing in a tax haven is optimal if and only if this incentive is positive, i.e.*

$$Inc_{th} \equiv \sum_{i=1}^{\hat{n}} \frac{\rho_i \underline{\tau}(2\tau_i - \underline{\tau})}{2a(1-\tau_i)} + \sum_{i=\hat{n}+1}^{\tilde{n}} \left[(1-\tau_i)\rho_i + \frac{{\tau_i}^2 \rho_i}{2a(1-\tau_i)} - c_i \right] - c_0 > 0. \tag{3.7}$$

Proof. See Appendix C.3. □

To determine the impact of tax rates we have to distinguish the tax rates in countries in which the multinational is active independent of a tax haven investment versus those that only are attractive with a tax haven. Simple inspection of Inc_{th} yields the following comparative statics.

$$\frac{dInc_{th}}{d\tau_i} = \frac{\rho_i \underline{\tau}(2-\underline{\tau})}{2a(1-\tau_i)^2} > 0 \ \forall i = 1, ..., \hat{n} \ . \tag{3.8}$$

Thus, the higher are the tax rates in countries in which the multinational is active even without a tax haven investment, the more profitable it becomes to invest in a tax haven. Furthermore,

$$\frac{d^2 Inc_{th}}{d\tau_i d\rho_i} = \frac{\underline{\tau}(2-\underline{\tau})}{2a(1-\tau_i)^2} > 0 \ \forall i = 1, ..., \hat{n} \ . \tag{3.9}$$

This shows that the effect of a foreign tax rate is sensitive to the profitability of the respective affiliate, with higher profits increasing the effect of the foreign tax rate. In addition,

$$\frac{d^2 Inc_{th}}{d\tau_i da} = -\frac{\rho_i \underline{\tau}(2-\underline{\tau})}{2a^2(1-\tau_i)^2} < 0 \ \forall i = 1, ..., \hat{n} \ . \tag{3.10}$$

Thus, the more difficult profit-shifting is for the multinational, the less sensitive will be its reaction to foreign tax rate changes.

It is instructive to evaluate the effect of tax changes in countries in which the multinational is active only in case of a tax haven investment.

$$\frac{dInc_{th}}{d\tau_i} = -\rho_i + \tfrac{\tau_i(2-\tau_i)\rho_i}{2a(1-\tau_i)^2} < 0 \quad \forall i = \hat{n}+1,...,\tilde{n} \quad \forall \tau_i \leq 1 - \sqrt{\frac{1}{2a+1}}\,, \tag{3.11}$$

$$\frac{dInc_{th}}{d\tau_i} = 0 \quad \forall i = \hat{n}+1,...,\tilde{n} \quad \forall \tau_i > 1 - \sqrt{\frac{1}{2a+1}}\,. \tag{3.12}$$

This result has the notable implication that a multinational may in fact be tempted to invest in a tax haven following a tax reduction in a country in which it has not been present so far. This counterintuitive situation can arise if this tax reduction makes an investment in this country attractive and hence adds to the potential base for profit shifting.

Inspection of Equation (3.7) further shows that firms in industries with lower fixed costs c_0 of establishing tax haven affiliates are more likely than others to have haven affiliates. Note that the fixed cost c_0 should be interpreted as the net cost of establishing a tax haven affiliate to use for tax avoidance purposes. If a tax haven affiliate would be profitable in the course of ordinary business activity that does not include any tax-motivated income reallocation, then c_0 would be negative. Hence firms in industries in which tax haven operations can serve the dual function of facilitating profit reallocation and generating ordinary business returns effectively face lower costs of engaging in profit reallocation through havens, and are therefore likely to do more of it.

We can summarize these results in the following two empirical predictions, based on Equations (3.8) and (3.10), respectively: (i) The larger the tax rate in a foreign non-haven country in which the multinational holds an affiliate, the stronger is the incentive and hence the more likely is it for a multinational to invest in a tax haven. (ii) The less costly it is to shift profits to a tax haven country, the stronger is the influence of foreign taxation on a multinational firm's tax haven investment.

Testing Prediction (i): When attempting to identify the effect of foreign tax rates on the tax haven decision, we need to take into account that the multinational is potentially engaged in several countries and that therefore the tax rates of all these countries matter. As Equation (3.9) shows, they do so to a different extent, however, depending on the profitability of the individual affiliates. We capture this by investigating the impact of the *average* non-haven tax rate, where all the foreign tax rates are weighted by the profitability of the individual affiliate. If the multinational has not invested in a tax haven, this average foreign tax rate is determined by the tax rates in countries $i = 1, ..., \hat{n}$ and is given by

$$\frac{\sum_{i=1}^{\hat{n}} \tau_i \rho_i}{\sum_{i=1}^{\hat{n}} \rho_i}. \tag{3.13}$$

In our empirical analysis we encounter the difficulty that we are not able to observe the actual profits ρ_i in country i, only reported after-tax profits $(1-\tau_i)\pi_i$. These reported profits are distorted due to taxation and income reallocation. In case of a tax haven investment they are given by

$$(1-\tau_i)\pi_i = (1-\tau_i)(\rho_i - \psi_i - \frac{a}{2}\frac{{\psi_i}^2}{\rho_i}) = (1-\tau_i)\left[1 - \frac{\tau_i(2-\tau_i)}{2a(1-\tau_i)^2}\right]\rho_i\ . \tag{3.14}$$

Inspection shows that this distortion is the higher the higher the country's tax rate τ_i. Thus, we require appropriate proxies to capture the effect of an affiliate's profitability on the decision to invest in a tax haven. We will proxy the affiliate's profitability by the number of employees, as discussed in more detail in Section 3.4.

Furthermore, we need to account for the fact that the average foreign tax rate we observe is potentially affected by the very fact whether the multinational holds a tax haven affiliate. The tax haven investment may make it profitable to invest in foreign countries $i = \hat{n}+1, ..., \tilde{n}$, that would not have been attractive destinations for investments without the income reallocation opportunities created by the tax haven investment. Consider a change in tax rates $\Delta_i \geq 0$ in countries $i = 1, ..., \tilde{n}$ such that the investor chooses to invest in a tax haven after this change in tax rates, but would not do so before. Both an increase in the tax rates at locations $i = 1, ..., \hat{n}$ where the multinational already holds an affiliate and a decrease in the tax rates at locations $i = \hat{n}+1, ..., \tilde{n}$ which become attractive only after tax haven investment could render tax haven investment optimal. The average non-haven tax rate for the investor changes from the status quo described in Equation (3.13) to the new average non-haven tax rate

$$\frac{\sum_{i=1}^{\tilde{n}}(\tau_i + \Delta_i)\rho_i}{\sum_{i=1}^{\tilde{n}} \rho_i} . \tag{3.15}$$

Taking the difference between (3.15) and (3.13) yields the *observed* change in the non-haven average tax rate which can be rewritten as

$$\frac{\sum_{i=1}^{\hat{n}} \Delta_i \rho_i}{\sum_{i=1}^{\tilde{n}} \rho_i} + \frac{\sum_{i=\hat{n}+1}^{\tilde{n}} \rho_i \left(\frac{\sum_{i=\hat{n}+1}^{\tilde{n}}(\tau_i\rho_i + \Delta_i\rho_i)}{\sum_{i=\hat{n}+1}^{\tilde{n}} \rho_i} - \frac{\sum_{i=1}^{\hat{n}} \tau_i\rho_i}{\sum_{i=1}^{\hat{n}} \rho_i} \right)}{\sum_{i=1}^{\tilde{n}} \rho_i} . \tag{3.16}$$

In our empirical analysis we are interested in identifying the effect of exogenous changes in tax rates in countries $i = 1, ..., \hat{n}$, i.e. countries in which the multinational would invest even without a tax haven affiliate. This exogenous change in tax rates is captured by the first term. As shown in Equation (3.8) above, investing in a tax haven is positively influenced by an increase in the tax rates of the countries in which the multinational already holds affiliates. Thus, when estimating the impact of foreign tax rates, higher tax rates in countries in which a multinational firm would invest under any circumstances should stimulate greater demand for tax haven affiliates.

The second term captures the change in the observed non-haven tax rate that is due to endogeneity of the multinational's investment decision. Evaluating the numerator of the second term we find that the observed change in the average non-haven tax rate exceeds the change of interest if the new affiliates the multinational opens due to the tax haven investment are located in countries that exhibit on average higher tax rates than the previous average tax rate, and conversely.

This has important implications for the interpretation of the causal effects of tax changes. In particular, OLS results overestimate the true effects, as captured by the IV estimates, if the tax rates at the firm's new locations increase the firm's average foreign non-haven tax rate, and underestimate the true effects if the tax rates faced at the new locations are lower than the previous average foreign non-haven tax rate.

Analogously, if the multinational firm decides to stop being active in the tax haven, the observed change in the average foreign non-haven tax rate may differ from the causal change because of subsequent adjustments in its non-haven locations. Leaving the tax haven after changes in foreign non-haven tax rates may be optimal if the fixed costs of investing there c_0 have to be paid in every period, or else if part of the fixed costs can be recovered upon exit.

To account for this potential endogeneity of the observed tax rate, we employ an instrumentation strategy that holds the locations of the multinationals' foreign non-haven affiliates fixed to the ones of the starting period of our sample and restricts attention to changes in observed tax rates for these locations only.

Testing Prediction (ii): The second prediction relates the impact of foreign taxation on tax haven investment to the cost of shifting profits. Average foreign tax rates and values of the shifting cost parameter are likely to differ between firms, and may vary systematically between industries. Industries may differ in average values of the shifting cost parameter a, reflecting differences in the importance of intangible assets and other business features that facilitate profit reallocation; and industries may also differ in the extent to which a varies among firms in the industry. Differentiating (3.10) with respect to a indicates that:

$$\frac{d^3 Inc_{th}}{d\tau_i d^2 a} = \frac{\rho_i \underline{\tau}(2-\underline{\tau})}{a^3(1-\tau_i)^2} > 0 \tag{3.17}$$

Since the expression in (3.17) is positive, it follows that the effect of a on $\frac{dInc}{d\tau_i}$ is nonlinear, and more specifically, that a mean-preserving spread in the distribution of a produces a greater average value of $\frac{dInc}{d\tau_i}$. Consequently, industries in which firms have very different costs of profit reallocation should be expected to display greater average sensitivity of tax haven demand to non-haven tax rates than do other industries, even though average costs of profit reallocation do not differ.

The second prediction cannot be tested directly with the available data, since it is not possible to measure firm-specific income reallocation costs. Instead, we use firm-fixed effects in the baseline econometric analysis to control for differences in marginal costs of income reallocation and distinguish firms by industrial sectors in an attempt to proxy for cost differences that vary with industry.

3.3 Data and descriptive statistics

Our analysis is based on the Microdatabase Direct investment (MiDi) provided by the Bundesbank, the German central bank. We use the information on outward foreign direct investment by German companies. The database consists of a panel of yearly information on the foreign affiliates of German firms for the period from 1996 until 2008. By the German Foreign Trade and Payment Regulation (Außenwirtschaftsverordnung), any resident who holds shares or voting rights of at least 10% in a company with a balance sheet total of more than 3 million euro is obliged to report information on the financial characteristics of these affiliates to the Bundesbank (Lipponer, 2009).[6] The same information has to be provided on branches or permanent establishments abroad if their operating assets exceed 3 million euro. The comprehensiveness of these data suggests that they can be used to draw a very reliable picture of the foreign investment of German companies.

The MiDi data include information on parent companies only for the years 2002 to 2008, so the analysis is restricted to these years. During the 2002-2008 period, the MiDi contains 173,312 affiliate-year observations. Some affiliates are reported several times, because multiple investors hold participating interests in them. We focus our analysis on directly held foreign affiliates and thus abstract from more complex incentive structures that may exist in multi-level holding chains.[7] This limits the analysis to 117,585 affiliate-year observations.

For consistency across parents, we delete 218 observations for which the degree of participation of the parent is smaller than the reporting requirement of 10%. In addition, we drop observations on parents in a number of sectors, including government institutions

[6] The reporting thresholds have changed several times in the past. We only refer to the reporting threshold as of 2002 that is relevant to us.

[7] For an in-depth discussion of the complex determinants of ownership chains, see Weichenrieder and Mintz (2010).

and private households. We drop observations on parents in the financial sector, because they are subject to special balancing requirements, and the reporting requirements for these companies changed during the period of analysis. We delete the sectors housing enterprises and other real estate activities, as they report neither sales nor employees, which we will use as size measure in our analysis. Similarly, we drop the sector holding companies as reported sales and employees are very often zero, even though these companies are not small.[8] We later remove this restriction as a robustness check and find that our results are unaffected.

We finally obtain a sample of 54,367 affiliate-year observations that correspond to 19,165 parent-year observations. The observations are distributed evenly across years with a minimum of 2,639 observations and a maximum of 2,875 observations.

We augment the MiDi with information on statutory tax rates mainly from the International Bureau of Fiscal Documentation (IBFD) and information on GDP from the International Monetary Fund (IMF). We use the definition of tax havens derived by Hines and Rice (1994) which is widely accepted in the literature and was only recently used by Dharmapala and Hines (2009).[9] Alternatively, we could have used the definition propagated by the OECD (OECD, 2000). We chose Hines and Rice (1994)'s tax haven definition to derive results which are comparable to the literature, in particular the study by Desai, Foley, and Hines (2006). Further, no OECD member countries appear on the OECD's tax haven list, which thereby omits a number of tax havens popular with German firms, such as Switzerland. Very few investors in the MiDi data hold branches or affiliates in the island states on the OECD tax haven list. Using the OECD's tax haven definition would also preclude using a linear probability model, because the model may not yield accurate coefficients given the low incidence of investment in those tax havens (see also Durlauf, Navarro, and Rivers, 2010).

Table 3.1 presents descriptive evidence on the use of tax havens by sectoral group. For comparative purposes, information on financial firms is provided in addition to informa-

[8] In addition, we delete 331 affiliate-year observations for parents which are not classified holdings, but are de facto holdings after consultations with the statistical department of the Bundesbank.

[9] For a list of tax havens, please refer to Appendix C.1.

Table 3.1: Choice of tax havens, by sectoral group

Parent sector	Manufacturing		Service		Financial	
Total number of parent years	11,603		6,733		2,506	
of which with tax haven affiliate	1,976		1,337		932	
of which						
internationally active parents	75.81%		52.43%		57.19%	
with more than one t.h. affiliate	22.87%		18.03%			
	non haven	tax haven	non haven	tax haven	non haven	tax haven
Number of affiliate years	33,203	2,829	14,427	1,768	7,897	2,294
of which						
in manufacturing sector	51.19%	32.63%	12.08%	4.81%	3.89%	0.74%
in service sector	46.69%	63.56%	82.20%	90.16%	15.35%	18.09%
in financial sector	1.38%	3.39%	4.76%	4.58%	79.84%	81.17%
other	0.75%	0.42%	0.96%	0.45%	0.92%	–
Mean number of affiliates per parent	3.77	1.43	4.07	1.32	3.75	2.46

Choice of haven	Manufacturing aff. years	Manufacturing par. years	Service aff. years	Service par. years	Financial aff. years	Financial par. years
Big havens: more than 1 million inhabitants						
Hong Kong	459	410	233	219	164	104
Ireland	226	215	78	61	252	188
Lebanon	12	12	.	.	8	8
Liberia	.	.	16	16	.	.
Panama	19	19	20	20	3	3
Singapore	517	467	204	185	203	127
Switzerland	1,368	1,242	880	814	359	312
Small havens: less than 1 million inhabitants						
Bermuda	.	.	13	13	23	19
British Virgin Islands	21	17	11	11	22	20
Cayman Islands	.	.	3	3	233	127
Cyprus	22	22	60	17	8	8
Channel Islands	.	.	19	19	89	28
Luxembourg	124	114	163	151	864	587
Malta	38	38	39	30	16	10
Other	23	21	29	27	50	32
Total	2,829	2,577	1,768	1,586	2,294	1,573

. denotes tax havens where fewer than three affiliate-years or parent-years are observed, so the exact number of investments must not be reported for confidentiality reasons.
Manufacturing firms: firms classified NACE 1500-3700, *service firms:* firms classified NACE 5000-9300, with the before mentioned sample restrictions, *financial firms:* firms classified NACE 6500-7000.
If a parent invests in several tax havens, it is counted multiple times (once per tax haven).

tion on firms in the manufacturing and service sector which are analyzed later on. On average, a tax haven affiliate is held in 20.4% of parent-years (17.9% excluding financial companies). This figure seems low by international standards: Desai, Foley, and Hines (2006) report that tax haven investment is observed for 37.8% of parent-years in their sample of U.S. multinationals. This difference reflects, in part, the inclusiveness of the MiDi data, in that the size thresholds for reporting are much lower than in the U.S. data analyzed by Desai, Foley, and Hines (2006), resulting in a higher proportion of small firms and those with relatively small foreign operations.[10]

The proportion of firms owning tax haven affiliates is higher for service firms (19.9%) than for manufacturing firms (17.0%), and a larger proportion of service firms own a tax haven affiliate but are not internationally active in non-haven countries. About a fifth of both manufacturing and service firms that are present in tax havens own more than one tax haven affiliate, and the mean number of tax haven affiliates is also approximately equal. In contrast, 37.2% of financial firms hold affiliates in tax havens, and they own on average twice as many tax haven affiliates as do manufacturing and service firms.

The share of affiliates in tax havens that are in the service sector is disproportionately high. For manufacturing firms, the share of service affiliates in tax havens is about 17 percentage points higher than their overall share of affiliates in the service sector, and for service firms, it is eight percentage points higher. Also for financial companies, investment in service affiliates is more common in tax havens than in non-haven countries.

The lower panel of Table 3.1 reports the number of affiliate-year and parent-year observations by tax haven and sectoral group of the parent firm. It shows that the preferred tax haven destination varies by sectoral group. Manufacturing firms clearly prefer the big tax havens. More than 90% of observations are accumulated there; about 48% in Switzerland alone. The island tax havens, in particular Bermuda, the Cayman Islands and the Channel Islands, are very rare investment destinations. Switzerland is similarly popular among service firms; about half of their tax haven affiliates are located there.

[10]The summary statistics provided by Desai, Foley, and Hines (2006) are based on 81,604 affiliate-years and their regressions use 8,435 parent-years, so crude calculations imply a mean of 9.7 affiliates per parent. Desai, Foley, and Hines (2004) use the same data set and report that U.S. parents own between 7.5 and 7.8 affiliates on average in the years 1982, 1989 and 1994. In contrast, parents in our sample average only 2.8 foreign affiliates (4.0 affiliates if indirectly held affiliates are included).

Service firms more extensively use the small havens, where almost a fifth of tax haven affiliates are located, most prominently 9% in Luxembourg. For financial companies, Luxembourg is distinctly the most popular destination with 38% of affiliate-year observations in tax havens. The Cayman Islands are their fourth most important tax haven destination: 10% of affiliate-year observations in tax havens are located there. Evidently, the attractiveness of tax havens strongly depends on sector characteristics.

Figures 3.1 and 3.2 provide local polynomial plots of the relationship of tax haven investment and foreign non-haven taxation for manufacturing and service firms in the pooled sample.[11] The x-axis shows the average of the statutory tax rates faced by a parent firm's foreign affiliates in non-haven countries weighted by the number of employees of the respective affiliate, which is the measure of foreign non-haven taxation that we will use in our empirical analysis.[12] The variable on the y-axis is tax haven use, which is a binary variable that is equal to one whenever a parent invested in at least one tax haven in a given year. Both for service firms and for manufacturing firms, the cross-sectional relationship of taxation and tax haven use is increasing up to a level of around 30%. The increase for manufacturing firms is steeper, and the maximium tax haven use is higher than for service firms. Counterintuitively, the relationship seems to be decreasing above the level of 30% for manufacturing firms, with a resurge for high value of taxation. This picture may stem from the cross-sectional nature of the data, or else from the investment pattern of newly entering multinational firms. For the tax rates of the U.S. and France, the two most popular investment destinations for German firms among entrants, fall exactly into this range, so the apparent decrease in tax haven use for high tax rates may result from emerging multinational firms with few foreign locations.[13]

[11] Observations in the lowest and highest percentile of the average foregin non-haven tax rate are excluded.

[12] The number of employees of the affiliate is adjusted by the degree of participation of the parent in the affiliate.

[13] We explored an empirical specification including a quadratic term of the foreign non-haven tax rate, but a quadratic relationship is not robustly supported by the data.

Figure 3.1: Manufacturing firms

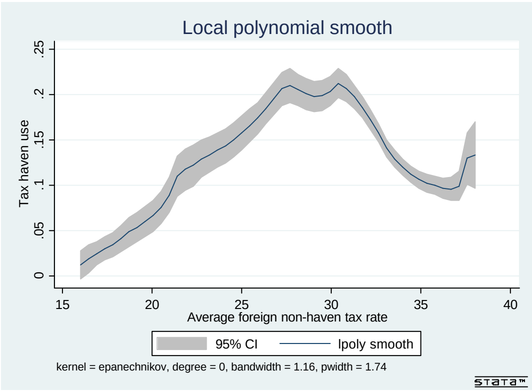


Figure 3.2: Service firms

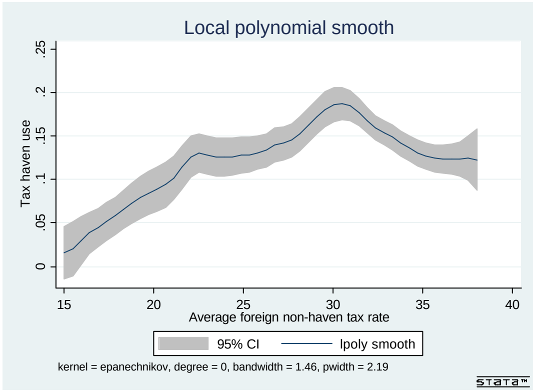


Table 3.2 provides an overview of the main variables used in our regression analysis for the full sample and the two subgroups we are going to consider. The variables will be explained in detail in the next section. The proportion of firms investing in a tax haven is lower (around 14%) than for the full data and equal across sectoral groups, because firms investing only in a tax haven drop from the regression sample. As firms with zero sales or employees drop, the average size of the firms used in our regressions is slightly higher than the average size of all firms in the sample. The statistics of the average foreign non-haven tax rate and the instruments for the regression sample are similar.[14] The third columns for every group report mean difference tests of the main regressors by the dependent variable. Firms that invest in a tax haven are on average significantly larger, both domestically and internationally. In addition, they face significantly higher average foreign tax rates, which is consistent with the incentives discussed earlier.

[14] 4.1% of manufacturing firms and 8.6% of service firms drop because only investment in tax havens is observed. 4.0% of the remaining manufacturing firms and 18.0% of service firms drop due to their zero number of employees. Table C.1 in Appendix C.4 presents the corresponding summary statistics for the full data.

Table 3.2: Summary statistics, regression sample

	Full sample			Manufacturing firms			Service firms		
	Mean	SD	Mean Diff.	Mean	SD	Mean Diff.	Mean	SD	Mean Diff.
Have haven	.140	.347	—	.140	.348	—	.139	.346	—
[within variation]		.132	—		.127	—		.129	—
# parent employees (in 1,000)	1.252	7.255	3.660***	1.268	6.471	3.790***	1.258	9.058	3.602***
Ln (# parent employees)	5.512	1.784	1.217***	5.932	1.407	1.094***	4.618	2.128	1.409***
# non-haven employees (in 1,000)	.463	2.306	1.418***	.525	2.677	1.761***	.340	1.375	.757***
Ln (# non-haven employees)	4.484	1.737	1.369***	4.679	1.666	1.504***	4.097	1.790	1.085***
Ave. foreign non-haven tax rate	.295	.058	0.006***	.299	0.056	0.005***	.293	0.058	0.007***
Ave. tax rate at 2001 non-h. loc.	.315	.058	0.016***	.320	0.055	0.015**	.307	0.060	0.017***
Observations (for instrument)	16410	(12755)		10661	(8533)		5052	(3751)	

Regression sample if number of employees used as size measure. Number of observations is lower for instrument, because firms that enter after 2001 drop from sample. Full sample additionally includes parent firms in the sectors agriculture, mining, electricity and water supply, and construction. Mean difference by haven status; base category: does not hold an affiliate in tax haven. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Manufacturing firms: firms classified NACE 1500-3700, *service firms:* firms classified NACE 5000-9300, with the before mentioned sample restrictions. *Have haven:* indicator variable; 1 if parent firm owns at least one affiliate in at least one tax haven in a given year, 0 otherwise. *# parent employees:* number of employees, parent firm. *Parent productivity:* natural logarithm of parent sales over parent employees. *# non-haven employees:* sum of number of employees in affiliates which are not located in tax havens, reduced according to share of participating interests. *Ave. foreign non-haven tax rate:* Average of the statutory tax rates faced by foreign affiliates in non-haven locations weighted by affiliate number of employees, adjusted by share of participation where appropriate. *Ave. tax rate at 2001 non-h. locations:* Average of the statutory tax rates faced by foreign affiliates in non-haven locations of 2001 weighted by GDP.

3.4 Empirical strategy

As outlined in section 3.2, a multinational firm's decision to invest in a tax haven depends on the taxation it faces at its foreign non-haven locations, its marginal cost of reallocating taxable income and the fixed cost of tax haven investment. The probability of tax haven investment should increase as foreign non-haven tax rates rise, with this effect being strongest for firms with low costs of reallocating profits.

We specify the following linear probability model:

$$y_{jt} = \beta_0 + \beta_1 \tau_{jt} + \beta_2 p_{jt} + \beta_3 p_{jt}^2 + \beta_4 nh_{jt} + \beta_5 nh_{jt}^2 + \alpha_i + \gamma_t + u_{jt} \tag{3.18}$$

The dependent variable y_{jt} is a dummy which is equal to one if a firm j holds at least one affiliate in at least one tax haven in a year t. Our independent variables are τ_{jt}, the average of the statutory tax rates faced by j's non-haven affiliates in t weighted by the number of affiliate employees, p_{jt}, the natural log of the size of company j in period t and its square, p_{jt}^2, nh_{jt}, the natural log of the size of j's foreign non-haven activities in t and its square, nh_{jt}^2, α_i, the firm-specific costs of reallocating profits across countries, and γ_t, a year fixed effect.

The coefficient of main interest is β_1. It captures the effect of the taxes levied on profits of a multinational's foreign non-haven affiliates on the probability that it invests in a tax haven. Equation (3.9) implies that greater firm profitability increases the impact of non-haven tax rates on the likelihood of investing in a tax haven affiliate. Thus, we use a weighted, not a simple average of the foreign non-haven tax rates. We cannot use before-tax profits as weights, as our data contain only after-tax profits which are doubly distorted due to taxation and profit reallocation. Instead, we use the number of employees of the affiliate, adjusted by the degree of participation of the parent, to approximate the relative importance of an affiliate for a multinational group.[15] As indicated above, we expect $\beta_1 > 0$.

In principle, the probability of tax haven investment is also influenced by taxation in the multinational's home country. As we use a panel data set of German multinational firms, this effect cannot be gauged explicitly due to lack of sufficient variation. Still, changes in home country taxation are indirectly taken into account through the year fixed effect.

Other independent variables include parent size and the size of the parent's non-haven activities, capturing the impact of size on profitability. Recent literature on foreign direct investment suggests that larger firms with bigger international activities can be expected to be more productive than their smaller competitors (e.g. Helpman, Melitz, and Yeaple, 2004; Tomiura, 2007; Yeaple, 2009; Chen and Moore, 2010). Consequently, these firms are better able to overcome the fixed and variable costs associated with setting up an affiliate in a tax haven and its subsequent use for income reallocation.

We use the number of employees to measure parent size and the size of the company's foreign non-haven activities, reduced according to the share of participation interests where applicable. The advantage of this size measure is that it is less likely to be affected by profit reallocation activities than are alternatives such as sales. For example, foreign affiliates may be permitted to use the distribution network of the parent company in exchange for a small fee to sell their products directly to customers, so sales and profits accrue abroad.[16] As the distribution of the size variables is strongly skewed to the right, the regressions use the natural log of employees as a size measure. Thus, observations for which the size variables are zero drop from our regression sample. Following Desai, Foley, and Hines (2006), regressions include the size measures both linearly and squared.

[15] As an alternative to the number of employees, we considered using assets or sales as firm level weights. None of these measures is similarly satisfactory however. We observe only fixed *and* intangible assets, not fixed assets separately, so this variable is very likely influenced by tax-avoidance behavior. As pointed out below, a similar concern can be raised against the use of sales. We consider it least likely that the number of employees will be systematically distorted due to taxation. Nonetheless, we will use GDP as weights in our instrument.

[16] For an illustrative example, see the case study "Alpha Company" in the report prepared by the Joint Committee on Taxation for the public background hearing by the House Committee on Ways and Means in July 2010 (Joint Committee on Taxation, 2010).

The variable cost of using a tax haven should vary with firm-specific characteristics such as the R&D intensity of a firm. The location of intangible assets, licence arrangements and royalty payments have been shown to be used as income reallocation tools (e.g. Dischinger and Riedel, 2011; Karkinsky and Riedel, 2012).[17] A firm with larger intangible assets should have greater discretion in choosing transfer prices due to the lack of comparable market transactions. Thus, and as also shown in Equation (3.10), the response to changes in foreign taxation should vary across firms depending on their marginal cost of income reallocation. These firm specific characteristics are, however, unobservable.

We take two measures to address this issue. On the one hand, we conduct our analysis separately for the group of manufacturing firms (NACE 1500-3700) and for the group of service firms (NACE 5000-9300, with the before mentioned sample restrictions),[18] because the latter have a systematically smaller R&D intensity than the former. Using sector-level data from the Innovation Survey of the Center for European Economic Research (Zentrum für Europäische Wirtschaftsforschung, ZEW) for the years 1996-2008, we find that the average R&D intensity for the manufacturing sector is twice as high as the R&D intensity for service sectors. At the same time, the descriptive evidence provided in section 3.3 shows that the proportion of service firms owning tax haven affiliates and the share of service firms' affiliate-years observed in tax havens are higher than the corresponding statistics for manufacturing firms. This could point to lower fixed costs of tax haven investment for service firms, so not only taxation, but also parent size could have a differential impact on tax haven investment by manufacturing and service firms. In addition, the share of service affiliates of manufacturing parents located in tax havens is disproportionately higher than the share located in non-haven countries. Overall, there is thus reason to believe that the processes governing tax haven investment by service firms and by manufacturing firms may differ significantly, which suggests that they should be analyzed separately. We report the results of a full-sample analysis in Appendix C.4 in Table C.2.

[17] In addition, a variant of this type of strategy is part of all six case studies of the report by the Joint Committee on Taxation prepared for the public hearing before the House Committee on Ways and Means (Joint Committee on Taxation, 2010).

[18] This implies that we do not consider parent firms in agriculture, mining, electricity and water supply, and construction in our analysis.

On the other hand, the empirical analysis uses firm-fixed effects to capture the influence of firm-specific differences in the marginal cost of income reallocation, at least to the extent that they are approximately constant over the sample period. Fixed effects also account for unobserved firm-specific characteristics such as the degree of tax sensitivity, that is, the importance that a firm assigns to the amount of its tax payments, which may render firms ex ante more or less likely to invest in tax havens.[19] Likewise, the data provide information on the sector of the affiliates mostly at the two-digit NACE Rev. 1 level, so particular incentive schemes for firms in sub-sectors cannot be taken into account, and we do not have information on the sub-national location of firms, so we cannot account for local taxation. The use of firm-fixed effects controls for time-invariant aspects of these firm attributes.

In estimating Equation (3.18), it is necessary to take into account that the average foreign non-haven tax rate is endogenous because entry in a tax haven may have a feedback effect on the optimal location decisions of a firm, as shown in section 3.2. To address this issue, we use an instrumentation strategy based on the initial location decisions of the firm. We fix the location decisions of the firm in the year 2001, the year prior to our analysis period and instrument the observed average foreign non-haven tax rate with a hypothetical average foreign non-haven tax rate had the firm sticked to these location decisions. The instrument thus only captures changes in firms' foreign non-haven taxation that result from changes in tax rates, not from changes in location decisions. To counter any endogeneity that may be left on the intensive margin, we use GDP instead of the number of employees as weights when calculating our instrument.[20] Our instrumentation strategy thus eliminates the endogeneity in the observed average foreign non-haven tax rate by leveraging the insights drawn from our theoretical analysis: with the endogeneity stemming from firms adjusting their non-haven investment due to tax haven investment, our instrument abstracts from any changes in the average foreign non-haven tax rate that

[19] This issue has already been raised, but not addressed in Desai, Foley, and Hines (2006, p. 514).

[20] Earlier studies indicate that GDP correlates very closely with foreign investment and foreign profitability, both in an aggregate cross section (e.g. Hines and Rice, 1994) and in a firm-level panel (e.g. Desai, Foley, and Hines, 2006).

may be induced due to firms investing or ceasing to invest in a tax haven after 2001.[21] In sum, we estimate our regression equation in four different ways:

- pooled linear probability model,
- pooled linear instrumental variables model,
- linear fixed effects model,
- linear fixed effects model with instrumental variables.

The linear fixed effects model with instrumental variables is our preferred specification because it takes all sources of endogeneity into account. Nevertheless, we report the results of all four specifications, because they offer evidence of the factors that drive a firm's decision to invest in a tax haven beyond that available from only the fixed effects IV regression. By comparing pooled and instrumental variables estimates, it is possible to assess the bias due to endogeneity of the foreign location decisions. Juxtaposing the results of pooled and fixed effects specifications facilitates a balanced assessment of the influence of taxation, abstracting from unobservable differences in costs of using tax haven operations.

We use a linear probability model because otherwise it would be difficult to address the endogeneity issues satisfactorily in a limited dependent variables framework.[22] Using logit or probit would yield more accurate marginal effects at different points of the distribution of the covariates. In the logit framework, using firm fixed effects would be possible, but it is more problematic to use instruments.[23] In the probit framework, we could conduct an

[21]Desai, Foley, and Hines (2006) use the competitors' average foreign non-haven tax rate as instrument for the average foreign non-haven tax rate of a firm. Results based on an instrumentation strategy similar to theirs are reported in Appendix C.6.

[22]As far as possible, we have replicated our results using logit, probit, fixed effects logit and IV probit, see section 3.6.1 below.

[23]Purely practically, one could construct an IV variant of fixed effects logit by plugging in the predicted values from an OLS first stage regression in place of the endogenous variable and run a fixed effects logit second stage regression. We refrain from doing so because this approach may not produce consistent estimates, as conditional expectations do not pass through non-linear functions (see Wooldridge (2002, p. 235-237) and Angrist and Pischke (2009, p. 190-192)). Alternatively, a control function approach could be employed. This type of estimation strategy is subject to strong assumptions however, in particular concerning the independence of the full distribution of the error and the instrument.

instrumental variables analysis (though under very strong distributional assumptions), but would not be able to use firm fixed effects.

We generally use standard errors clustered at the level of the parent. For the fixed effects instrumental variables regression, we use bootstrapped standard errors, as clustered standard errors cannot be estimated. As recommended by Efron and Tibshirani (1998), the bootstrap estimates are based on 200 replications.

3.5 Results

Table 3.3 presents the regression results. The first four columns in the table present results for the sample of manufacturing firms and columns 5-8 present results for the sample of service firms. The odd-numbered columns present the results if no instrument is used, and the even-numbered columns contain the IV estimates.[24] The F-test for exclusion of the instrument in the first stage regression is rejected at high significance levels in all cases. The absolute value of the F-statistic is higher than the threshold of ten recommended by Staiger and Stock (1997) to obtain unbiased estimates, and higher than the thresholds proposed by Stock and Yogo (2002) to assume standard errors are unbiased.

The average foreign non-haven tax rate is estimated to have a significantly positive effect on the probability of a manufacturing firm's investing in a tax haven throughout regressions. The coefficient in the 2SLS regressions is about twice as high, and significantly so, as the coefficient in the pooled OLS regression, which does not take the endogeneity of the average foreign non-haven tax rate into account. Likewise, the coefficient in the fixed effects IV specification is slightly higher than in the simple fixed effects regressions, though the difference is not significant. This suggests that the true effect of the average foreign non-haven tax rate is underestimated if its endogeneity due to simultaneity is not taken into account. As explained in section 3.2, this finding indicates that multi-

[24]Regression results for the full sample are reported in Appendix C.4 in Table C.2. The effect of the tax rate is estimated to be significantly positive both in the OLS and the 2SLS specification, and insignificant if fixed effects are included.

Table 3.3: Regression results

	Manufacturing firms				Service firms			
	1	2	3	4	5	6	7	8
	OLS	2SLS	FE	FE IV	OLS	2SLS	FE	FE IV
Ave. foreign	0.566***	0.999***	0.175*	0.347**	0.508***	1.018***	-0.085	0.171
n.h. tax rate	(0.096)	(0.156)	(0.094)	(0.172)	(0.162)	(0.211)	(0.124)	(0.202)
Parent size	-0.061***	-0.058**	-0.013	-0.014	0.024	0.016	-0.002	-0.010
	(0.021)	(0.025)	(0.003)	(0.027)	(0.016)	(0.018)	(0.013)	(0.014)
Parent size,	0.008***	0.008***	0.003	0.004	-0.000	0.001	0.001	0.001
squared	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Foreign non-	-0.051***	-0.049***	-0.052***	-0.044**	-0.064***	-0.065**	-0.037*	-0.051**
haven size	(0.015)	(0.019)	(0.014)	(0.018)	(0.023)	(0.030)	(0.020)	(0.024)
Foreign n.h. size,	0.011***	0.011***	0.009***	0.008***	0.011***	0.012***	0.009***	0.010***
squared	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.004)
Constant	0.022	-0.117	0.070	0.007	-0.103	-0.176*	0.141***	0.118
	(0.062)	(0.086)	(0.069)	(0.080)	(0.075)	(0.093)	(0.053)	(0.074)
# of observations	10662	8533	10662	8533	5052	3751	5052	3751
# of parents	2321	1696	2321	1696	1270	832	1270	832
R-squared overall	0.16	0.17	0.16	0.17	0.10	0.10	0.008	0.09
R-squared within	–	–	0.02	0.02	–	–	0.02	0.03
Year Dummies	Y	Y	Y	Y	Y	Y	Y	Y
Instrument	N	Y	N	Y	N	Y	N	Y
F-Statistics	–	4064.39***	–	465.73***	–	1872.06***	–	717.80***
Standard errors	Cluster	Cluster	Cluster	Bootstrap	Cluster	Cluster	Cluster	Bootstrap

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Standard errors are clustered at the level of the parent firm or bootstrapped with 200 replications.
Regression sample: column 1-4 manufacturing firms, i.e. firms classified NACE 1500-3700; column 5-8 service firms, i.e. firms classified NACE 500-9300, except NACE 65xx, 70xx, 7490, 75xx, 95xx. Dependent variable: dummy variable which denotes whether a parent owns at least one affiliate in at least one tax haven in a given year. Independent variables: see Table 3.2.

nationals expand their activities in a way such that their average foreign non-haven tax rate (weakly) decreases following tax haven investment. Given that some of the largest and most popular investment destinations of German firms, such as the United States or France, also have the highest statutory tax rates, this is consistent with multinationals' investing in more sizeable and profitable markets first.

The estimates suggest that the magnitude of the tax effect is rather modest. The 0.347 coefficient estimate in column 4 of Table 3.3 implies that increasing a firm's average foreign tax rate by one percentage point is associated with a 0.3 percent greater likelihood of holding a tax haven affiliate. Greater domestic and foreign activities are associated with higher likelihood of tax haven investment, and the estimated coefficients on the size variables are largely unaffected by the use of the instruments. The effect of parent size turns insignificant in specifications that include fixed effects.

As the regression results in the columns 5-8 of Table 3.3 show, the estimated effect of the average foreign non-haven tax rate on tax haven investment by service firms varies with the estimation approach. It is positive and significant in the pooled OLS specification; larger and significant when estimated with 2SLS; and the point estimates of the tax coefficients in the regressions for service firms in columns 5 and 6 are almost identical to the point estimates of the tax coefficients in the corresponding regressions for manufacturing firms presented in columns 1 and 2. The tax coefficients in the fixed effects regressions for service firms reported in columns 7 and 8 are statistically zero.

The probability of tax haven investment by service firms is generally unaffected by the size of the parent company. Levels of foreign activity outside of tax havens are estimated to have nonlinear and significant effects on tax haven operations, similar to the effects found for manufacturing firms.

Our preferred estimates come from the fixed effects IV specification. They show that manufacturing firms are more likely to invest in tax havens if they face higher tax rates in their foreign non-haven locations, even if unobservable differences in the cost of income reallocation are taken into account. In contrast, we do not find a clear tax effect for service firms in our preferred framework. This difference may reflect a combination of factors: that service firms have higher costs of reallocating profits, or exhibit lower variability in these costs. Manufacturing firms may rely to a greater average extent than service firms on the returns to intangible property, the location of which may be more readily reallocated for tax purposes than are other forms of income, but the distribution of which is highly skewed, with some manufacturing firms earning significant fractions of their profits from intangible property, and others very little. These factors would imply that foreign non-haven tax rates should affect tax haven demand by manufacturing firms more heavily than service firms, even though service firms are at least as likely as manufacturing firms to establish tax haven operations in the first place. The 2SLS estimates appear to pick up the effect of omitted differences in the costs of income reallocation. A further indication in this direction is that the sector-level R&D intensities that are used in subsequent regressions to proxy for the costs of reallocating taxable income in section 3.6.3 are positively correlated with average foreign non-haven tax rates, with a slightly higher correlation coefficient for service firms than for manufacturing firms (0.129 for service firms vs. 0.102 for manufacturing firms).

A further interesting finding is that parent size is a significant determinant of tax haven use by manufacturing firms only. Together with the relatively higher prevalence of tax haven investment by service firms described in section 3.3, this evidence is in accord with a much lower fixed cost of tax haven investment by service firms. The German anti-deferral provisions imply that a multinational does not only need to formally establish a company in a tax haven (often referred to as "letterbox company"), but has to locate some productive activity there.[25] Our evidence suggests that this could be more costly on net for manufacturing than for service firms.

3.6 Robustness checks

The following two subsections present estimated coefficients from specifications that modify the basic econometric framework; from this evidence it appears that the results reported in section 3.5 are largely robust. The estimates in subsection 3.6.3 attempt to shed light on the firm characteristics captured by the firm fixed effects. Subsection 3.6.4 summarizes results of robustness checks that address endogeneity of non-haven size and non-random sample selection of service firms as potential purely econometric interpretations of our results.[26] On closer inspection, none of these issues appears to explain our results.

[25] For details on the German anti-deferral regulations, please refer to Appendix C.2.

[26] Overall, only 75 percent of service firms are used in the regression. 8.6 percent of firms drop because they only invest in a tax haven. 18.0 percent of the remaining observations on service firms drop as the observed size measure is zero. In contrast, only 5 percent of observations on manufacturing firms with non-haven investment drop for this reason.

3.6.1 Simple modifications of baseline specification

To test whether our results are driven by the specific setup of our baseline econometric analysis, we explore a number of variations, none of which substantially affects our main findings. The results of these robustness checks are tabled in Appendix C.4.

First, we re-estimate the linear probability model using the GDP weighted average foreign non-haven tax rate at the locations in the period a firm enters the sample as instrument instead of the locations in 2001.[27] The advantage of our baseline instrumentation strategy is that the locations as of 2001 are exogenous to any changes in location decisions, be they haven or non-haven, that occur during the period of analysis from 2002 to 2008. At the same time, we only use firms that invest abroad in 2001 already, which may induce sample selection bias. Such issues do not seem to affect our results, though (see Table C.3 in Appendix C.4).

In addition, we re-estimate the linear probability model after restricting our sample to those firms that do not hold a tax haven affiliate in 2001. This implies that our sample is selected endogenously based on the realization of the dependent variable, so the behavior may not be representative for the full sample. We nevertheless do so to exclude that our results are driven by firms that hold a tax haven affiliate in 2001 already. For manufacturing firms, the sign pattern of the coefficient remains robust, but the tax coefficient decreases in absolute value and is insignificant in the fixed effects specification. Parent size is insignificant throughout specifications, and non-haven size is less significant than before. For service firms, coefficients are mostly insignificant throughout specifications (see Table C.4 in Appendix C.4). Whereas our model helps explain manufacturing firms decision to invest in a tax haven, the variables seem not to affect service firms' tax haven investment. Note though, that the restriction imposed on the data implies that only part of the variation in the dependent variable is used, because the effect of covariates on firms' decision to disinvest from a tax haven is not analyzed.

Further, we re-estimate our baseline equation using limited dependent variable models as far as possible: probit, logit, IV probit and fixed effects logit. The IV probit model

[27] We use locations as of 2001 if the firm enters the sample before 2002.

rests on the assumption that the endogenous variable is normally distributed conditional on the instrument and parameters are only consistent if the error term is homoskedastic. As it is uncertain whether these assumptions are valid in our case, the results have to be interpreted with caution. The sign pattern of the coefficients is largely robust. The average marginal effects of the average foreign non-haven tax rate are almost identical to the marginal effects estimated using the linear probability model, except for the fixed effects logit average marginal effect being insignificant for manufacturing firms. Thus, the findings largely confirm our previous results (see Tables C.5 and C.6 in Appendix C.4).

Next, we re-include the sector holding companies in our analysis. We dropped this sector before because holding companies usually report zero sales and employees even though the actual companies are not small, and as the sector comprises firms with the same administrative structure, but activities that actually belong to various other sectors. To address the diversity of the sector, we assign parents the sector of the corporate group, using a variable specifically created to address this issue in our data set. After dropping financial companies, government institutions and private households, we obtain a sample of 21,104 parent-year observations in the manufacturing sector and 11,783 parent-year observations in the service sector.[28] Our findings both for manufacturing firms and for service firms are robust, though the significance level of the tax coefficient in the fixed effects IV specification for manufacturing firms is only 10.1% (see Table C.7 in Appendix C.4).

3.6.2 Sensitivity to choice of tax havens and tax rates

The descriptive statistics in section 3.3 show that Switzerland is by far the most important tax haven for German firms. This is not surprising given its geographical, linguistic and cultural proximity to Germany. It is nonetheless potentially worrisome that the findings could be driven by the dominance of this single tax haven, but as it happens, the results are largely robust to dropping all affiliates located in Switzerland (see Table C.8 in Appendix C.4).

[28]Note that only two thirds of observations on manufacturing firms and only half of observations on service firms are usable in our regressions because the reported number of employees is zero.

The main analysis uses statutory corporate tax rates to capture incentives for income reallocation and thus tax haven use. This may not correctly capture tax differences because profits may not be taxed in full at this rate. To address this concern, we alternatively use the effective tax rates collected by Djankov, Ganser, McLiesh, Ramalho, and Shleifer (2010). As these data are available for the year 2004 only, we can only test the robustness of the pooled specifications. We obtain coefficients of the same sign and significance and similar magnitude as before, though the difference between the OLS and 2SLS tax coefficient does not persist for manufacturing firms (see Table C.9 in Appendix C.4).

3.6.3 Sector-level R&D intensities as proxy for the marginal cost of profit shifting

A firm's R&D intensity is a factor that is particularly likely to influence the marginal cost of income reallocation. We use firm-fixed effects in our main econometric analysis to capture firm-specific differences in the marginal costs of income reallocation, because we do not have a firm level measure of R&D activities.

In order to shed light on the question of the extent to which the marginal costs of income reallocation play a role for tax haven investment, we use sectoral data on R&D intensity, which are provided by the Center for European Economic Research (Zentrum für Europäische Wirtschaftsforschung, ZEW) based on its annual Innovation Survey. We include the industry R&D intensities as well as an interaction term of industry R&D intensity and average foreign non-haven tax rate in our pooled regressions. We refrain from doing so in the regressions with firm fixed effects because the firm fixed effects capture firm level heterogeneity with regard to the R&D intensity, so the firm fixed effects and the sectoral R&D data are collinear.

We cluster the standard errors on industry level and drop firms assigned to different sectors in different years to avoid artificial variation.[29] Our findings are presented in Table 3.4.

[29] 639 manufacturing firms and 322 service firms are dropped for this reason.

Table 3.4: Regression results including sector R&D intensity

	Manufacturing				Services			
	1	2	3	4	5	6	7	8
	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
Average foreign nh.	0.561***	0.445**	1.028***	0.564***	0.361***	0.388***	0.971***	0.905***
tax rate	(0.080)	(0.200)	(0.188)	(0.214)	(0.106)	(0.078)	(0.164)	(0.136)
Parent size	-0.058**	-0.058**	-0.056**	-0.053*	0.014	0.014	0.010	0.010
	(0.022)	(0.023)	(0.027)	(0.028)	(0.015)	(0.015)	(0.017)	(0.017)
Parent size, squared	0.008***	0.008***	0.008***	0.008**	0.002	0.002	0.002	0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Non-haven size	-0.051**	-0.052**	-0.048*	-0.051*	-0.077***	-0.077***	-0.082***	0.082***
	(0.021)	(0.021)	(0.028)	(0.027)	(0.022)	(0.022)	(0.020)	(0.020)
Non-haven size,	0.011***	0.011***	0.011***	0.011***	0.014***	0.014***	0.015***	0.015***
squared	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
Sector R&D intensity	-0.001	-0.008	-0.001	-0.033**	0.010	0.017	0.008	-0.009
	(0.006)	(0.013)	(0.007)	(0.016)	(0.006)	(0.011)	(0.007)	(0.017)
Interaction tax R&D		0.027		0.107**		-0.021		0.054
		(0.036)		(0.044)		(0.031)		(0.062)
Constant	0.004	0.038	-0.140	-0.006	-0.034	-0.043	-0.190***	-0.169***
	(0.070)	(0.084)	(0.121)	(0.108)	(0.052)	(0.045)	(0.056)	(0.055)
# of observations	9916	9916	7962	7962	3860	3860	2858	2858
# of sectoral groups	23	23	23	23	14	14	14	14
R-squared	0.17	0.17	0.17	0.17	0.15	0.15	0.16	0.16
Year Dummies	Y	Y	Y	Y	Y	Y	Y	Y
Instrument	N	N	Y	Y	N	N	Y	Y
F-Statistics tax	–	–	3121.24***	1595.31***	–	–	1463.28***	1911.39***
F-Statistics int.	–	–	–	3093.64***	–	–	–	915.84***
Standard errors	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster	Cluster

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Standard errors are clustered at the level of parent industry. P-values of R&D intensity for service firms in columns 5-8 are 11.9%, 15.2%, 22.7% and 59.5%.
Dependent variable: dummy variable which denotes whether a parent owns at least one affiliate in at least one tax haven in a given year. Independent variables: see Table 3.2; *sector R&D intensity:* calculated as expenditures on innovation over total sales by sector; *Interaction tax R&D:* interaction of sector R&D intensity and firm-level average foreign non-haven tax rate.

The estimates for manufacturing firms are largely unaffected by the inclusion of the R&D intensity, which has an insignificantly negative coefficient in the OLS and 2SLS specifications in columns 1 and 3. The interaction of R&D intensity and the average foreign non-haven tax rate is positive, and, in the 2SLS specification, significantly so. Given the same average foreign non-haven tax rate, firms in sectors with higher R&D intensity and thus lower costs of profit reallocation are more likely to invest in tax havens. The base effect of the sector R&D intensity is negative. This implies that lower cost of profit reallocation do not increase tax haven investment per se, but only if taxation renders profit reallocation sufficiently lucrative.

For service firms, neither the base effect nor the interaction of R&D intensity with the tax rate are significant, and the other coefficients are largely unaffected by their inclusion.

Taken together, these observations are evidence in favor of the interpretation that fixed costs of tax haven investment are higher for manufacturing firms than for service firms, so the former react more strongly to differences in taxation and the marginal costs of profit shifting. Service firms' tax haven investment in contrast is largely unaffected by (sectoral) differences in the marginal cost of profit shifting; together with the observation that the tax coefficient is insignificant in the fixed effects specification, this points to low fixed costs of tax haven investment, or else to business opportunities of certain service firms in tax havens.

3.6.4 Further robustness checks

A comparison of Equations (3.6) and (C.10) in Appendix C.3 shows that investing in a tax haven renders investing at other foreign non-haven locations more attractive for a multinational firm, because it can thereby enlarge its base of income available to reallocate. Therefore, not only the average foreign non-haven tax rate, but also non-haven size, may be endogenous. We use a modified version of our instrumentation strategy to address this issue and report the corresponding regression results along with a more detailed description of the approach in Appendix C.5. The estimation results for manufacturing firms are largely consistent with estimated tax effects obtained earlier. Results for service firms are generally consistent with previous insights, but more difficult to interpret, largely due to the fact that the sample becomes very small (32 firms), so results obtained from instrumental variable estimations may suffer from bias.

To ease comparison of our findings to the earlier findings by Desai, Foley, and Hines (2006), we replicate Tables 3.3 and 3.4 using an instrumentation strategy similar to theirs, based on the competitors' average foreign non-haven tax rate. We report the regression results in Tables C.11 and C.12 in Appendix C.6.

Finally, we explore whether and in how far regression results for service firms may be driven by sample selection. Almost a fifth of observations on parent firms in services drop from the regression sample because the number of parent or non-haven employees is zero, so size is not defined. The regression results for service firms may consequently reflect this aspect of sample selection, and possibly produce estimates of behavioral parameters that are not representative of the entire sample.

The reason for observed zero employees is not clear. We exclude the sector holding companies and housing and real estate where such a figure may (and does frequently) occur. Zero observed sales are a result of the reporting requirements. Sales are surveyed in million euros, so they are zero for any firm with a turnover of strictly less than 500,000 euros. This implies that a reported figure x disguises possible true sales values ranging from $1,000,000x - 500,000$ to $1,000,000x + 500,000$, except for 0 which disguises the interval $[0; 499,999]$.

We exploit this insight to impute the sales variables based on a model for grouped data developed in the statistical literature by Heitjan and Rubin (1990) in order to use all the observations. The use of imputed data does not appear to produce major changes in the coefficient and significance patterns. A detailed description of the procedure and the results (based on the instrumentation strategy similar to Desai, Foley, and Hines (2006)) are reported in Appendix C.6.

3.7 Discussion

Table 3.5 lists the effect of an increase of the independent variables by one standard deviation according to our estimates on the probability of investment in a tax haven, expressed in standard deviations of the dependent variable.

The left part of the table shows the implied effects for manufacturing firms and the right part for service firms.

The implied economic effect of a change in the average foreign non-haven tax rate by one standard deviation is equal for manufacturing and service firms according to the pooled estimation. If unobservable firm-specific differences are taken into account in the fixed effects specification, the effect becomes smaller in both cases, but is still larger and significant for manufacturing firms. The tax coefficient in the sample of service firms is insignificant. It appears likely that the estimates of tax effects in the pooled

Table 3.5: Economic significance

	Manufacturing firms		Service firms	
	IV	FE IV	IV	FE IV
Ave. foreign n.h. tax rate	**0.16**	**0.05**	**0.17**	0.03
Parent employees	**0.19**	0.16	0.17	0.01
Foreign non-haven employees	**0.35**	**0.21**	**0.28**	**0.25**

The table contains the effect of an increase in one of the independent variables by one standard deviation at the mean of all independent variables on the probability of tax haven investment, expressed in standard deviations. The mean probability of tax haven investment for manufacturing firms is 0.140, for service firms 0.139, and the corresponding standard deviations are 0.348 and 0.346 respectively. Manufacturing firms' mean average foreign non-haven tax rate is 0.299, service firms' mean average foreign non-haven tax rate is 0.293, and the corresponding standard deviations are 0.056 and 0.058. Mean parent size is 5.932 for manufacturing firms and 4.618 for service firms, with standard deviations of 1.407 and 2.128. Mean non-haven size is 4.679 for manufacturing firms and 4.097 for service firms, with standard deviations 1.666 and 1.790 respectively.
Statistically significant effects are printed in bold.

specification reflect the impact of omitted firm-specific characteristics, which are relatively more important for service firms than for manufacturing firms. Our findings are consistent with either high marginal costs of income reallocation by service firms, or - compared to manufacturing firms - relatively little variability in the cost of income reallocation, either of which would be consistent with small effects of foreign tax rates on tax haven investment. These influences may also drive that we do not find a clear effect of taxation in the full sample (see Table C.2 in Appendix C.4).

Similarly, parent size plays a smaller role for service firms than for manufacturing firms: the coefficients are smaller and insignificant throughout. This may point to smaller costs of setting up a tax haven affiliate for service firms, so the productivity thresholds are lower. The effect of non-haven size is comparable across sectoral groups.

How do these findings compare to earlier research? Desai, Foley, and Hines (2006) show that for U.S. firms, higher foreign non-haven taxation is associated with a lower probability of tax haven investment. Our results for German manufacturing firms contrast with their findings, as the decision of German manufacturing firms to invest in a tax haven is positively influenced by higher foreign non-haven taxation. As Desai, Foley, and Hines (2006) use a different instrumentation strategy based on the location decisions of a firm's competitors in the same industry as a given firm, one may object that our findings are not directly comparable. We report the results of estimations based on an instrumentation strategy similar to their instrumentation strategy in Appendix C.6, and obtain qualitatively similar results.

3.8 Conclusion

Tax haven operations can facilitate tax avoidance by multinational firms, but the evidence suggests that these opportunities are limited. Less than 20 percent of German non-financial multinational firms do have affiliates in tax havens, reflecting that the available tax savings are less than the costs of establishing the affiliates. Furthermore, even among the firms with tax haven operations, some undoubtedly choose their locations for normal business reasons unrelated to tax savings. Despite the apparently limited appeal of using tax haven operations to avoid taxes, the evidence is consistent with tax motivations of some firms, particularly those in manufacturing industries. Larger firms, and those with other foreign operations located in high-tax countries, are more likely than others to have affiliates in tax havens. Research-intensive firms that may have greater tax avoidance opportunities than others are the most likely to have tax haven operations.

The model implies that high foreign tax rates encourage tax haven investment, but that this effect is dampened by firm-specific marginal costs of income reallocation. Further, the model indicates that the relationship between non-haven taxation and the incentive to invest in a tax haven is complex and composed of two opposite effects. Higher tax rates at the locations where a firm is already present before investing in a tax haven increase the probability of investing in a tax haven, as expected. In contrast, the opposite relationship holds for tax rates at locations that become attractive investment venues only for firms that also have tax haven investments: the attractiveness of tax havens increases as tax rates fall in these potential investment locations. This mechanism may in part explain the persistence of tax haven investment despite falling tax rates elsewhere.

There appear to be significant differences between the tax haven investment patterns of service and manufacturing firms. High foreign tax rates are associated with tax haven investments of manufacturing firms, which is consistent with tax havens being used to reallocate taxable income from jurisdictions in which it is taxed more heavily. At the mean, an increase in the average foreign non-haven tax rate of one percentage point increases the probability that a manufacturing firm invests in a tax haven by 0.3 percentage points. This effect is robust to controlling for unobservable firm-specific differences. Tax haven investment by service firms is not significantly influenced by taxation if unobservable firm-specific characteristics are taken into account. This evidence is consistent with service firms facing high marginal costs of income reallocation, and relatively little variability in these costs, which together depress the effects of foreign tax rate differences. Still, tax haven investment is relatively more common among service firms than among manufacturing firms, reflecting the attractiveness of tax haven locations for ordinary business activities in service industries. This suggests that policy measures that raise the cost of income reallocation may discourage tax haven investment. At the same time, such policy measures may encourage firms to shift real activities to tax havens.

Given the increasing share of service industries in Western economies, the tax avoidance activities of service firms, and their consequences, offers a fruitful area for further research.

Appendices

Appendix A

Appendix to Chapter 1

A.1 Theory

A.1.1 Cost minimization

A.1.1.1 Parameter restriction (Assumption 1)

For simplicity of exposition, we focus on the case of a domestic firm. Assume that the firm chooses not only the allocation of knowledge within the firm, but also the overall knowledge level z^* subject to the constraint that its overall knowledge cannot exceed the knowledge draw, $z^* \leq \bar{z}$.

The firm correspondingly solves the following modified cost-minimization problem:

$$\begin{aligned} \min_{z_h, z^*} \quad & \frac{q_0 w_0}{A_0(1-e^{-\lambda z^*})}\Big((1+c_0(z^*-z_h)) + A_0\theta_0 e^{-\lambda(z^*-z_h)}(1+c_0 z_h)\Big) \\ \text{s.t.} \quad & z^* \leq \bar{z} \end{aligned}$$

where the number of production workers and headquarter managers as well as the production knowledge are determined by the constraints (1.2) to (1.4).

We need to derive a restriction which ensures that $z^* = \bar{z}$.

$$\mathcal{L} = \frac{q_0 w_0}{A_0(1-e^{-\lambda z^*})}\Big((1+c_0(z^*-z_h)) + A_0\theta_0 e^{-\lambda(z^*-z_h)}(1+c_0 z_h)\Big) + \phi(z^*-\bar{z})$$

A sufficient condition for $z^* = \bar{z}$ is $\phi > 0$; $\phi \geq 0$ is necessary. $\phi \geq 0$ whenever

$$\lambda(1+c_0(\bar{z}-z_h)) + c_0 \geq A_0\theta_0 c_0 e^{\lambda z_h}.$$

Equation (1.13) implies

$$e^{\lambda z_h} = \frac{c_0 e^{\lambda \bar{z}}}{A_0\theta_0(c_0+\lambda(1+c_0 z_h))},$$

so

$$e^{\lambda\bar{z}} \leq \frac{1}{c_0^2}(c_0+\lambda(1+c_0 z_h))(c_0+\lambda(1+c_0(\bar{z}-z_h))).$$

Here, z_h is restricted to be in the interval $[0,\bar{z}]$. In either case,

$$e^{\lambda\bar{z}} \leq \frac{1}{c_0^2}(c_0+\lambda)(c_0+\lambda(1+c_0\bar{z})).$$

An analogous argument for a multinationally active firm yields that $\phi \geq 0$ whenever

$$\begin{aligned} & e^{-\lambda\bar{z}}\left(\frac{q_0 w_0}{A_0}(\lambda(1+c_0(\bar{z}-z_h))+c_0) + \frac{q_1 w_1}{A_1}(\lambda(1+c_1(\bar{z}-z_h))+c_1)\right) \\ \geq & \frac{c_0}{\lambda(1+c_0 z_h)}\frac{q_j w_j c_j}{A_j} + q_{\bar{j}}\theta_{\bar{j}} e^{-\lambda(\bar{z}-z_h)} w_0 c_0 \end{aligned}$$

where $\bar{j}$ denotes the country with $z_{\bar{j}} = \bar{z} - z_h$ and j denotes the country with $z_j \geq \bar{z} - z_h$, and the left hand side takes into account that $z_j \geq \bar{z} - z_h$.

Using Equation (1.12), this equation can be simplified to

$$e^{\lambda\bar{z}} \leq \frac{\frac{q_0 w_0}{A_0}(\lambda(1+c_0(\bar{z}-z_h))+c_0) + \frac{q_1 w_1}{A_1}(\lambda(1+c_1(\bar{z}-z_h))+c_1)}{c_0\left(\frac{q_0 w_0 c_0}{A_0} + \frac{q_1 w_1 c_1}{A_1}\right)}(c_0+\lambda(1+c_0 z_h)),$$

analogous to the condition for the domestic firm.

Here, z_h is restricted to be in the interval $[0, \bar{z}]$. The condition yields

$$e^{\lambda \bar{z}} \leq \frac{\frac{q_0 w_0}{A_0}(\lambda + c_0) + \frac{q_1 w_1}{A_1}(\lambda + c_1)}{c_0 \left(\frac{q_0 w_0 c_0}{A_0} + \frac{q_1 w_1 c_1}{A_1} \right)} (c_0 + \lambda(1 + c_0 \bar{z})) \qquad \text{for } z_h = 0$$

$$e^{\lambda \bar{z}} \leq \frac{\frac{q_0 w_0}{A_0}(\lambda(1 + c_0 \bar{z}) + c_0) + \frac{q_1 w_1}{A_1}(\lambda(1 + c_1 \bar{z}) + c_1)}{c_0 \left(\frac{q_0 w_0 c_0}{A_0} + \frac{q_1 w_1 c_1}{A_1} \right)} (c_0 + \lambda) \qquad \text{for } z_h = \bar{z}$$

The former of the two conditions is binding whenever $c_0 < c_1$.

A.1.1.2 Lagrangian equation and first order conditions

$$\begin{aligned}
\mathcal{L} &= \sum_{j=0}^{1} n_j w_j (1 + c_j z_j) + n_h w_0 (1 + c_0 z_h) \\
&\quad + \sum_{j=0}^{1} \xi_j \left[q_j - n_j A_j (1 - e^{-\lambda \bar{z}}) \right] + \sum_{j=0}^{1} \phi_j \left[\bar{z} - z_h - z_j \right] + \kappa \left[\sum_{j=0}^{1} n_j A_j \theta_j e^{-\lambda z_j} - n_h \right] \\
\frac{\partial \mathcal{L}}{\partial n_j} &= w_j (1 + c_j z_j) - \xi_j A_j (1 - e^{-\lambda \bar{z}}) + \kappa A_j \theta_j e^{-\lambda z_j} = 0 \\
\frac{\partial \mathcal{L}}{\partial z_j} &= n_j w_j c_j - \phi_j - \lambda \kappa n_j A_j \theta_j e^{-\lambda z_j} = 0 \\
\frac{\partial \mathcal{L}}{\partial n_h} &= w_0 (1 + c_0 z_h) - \kappa = 0 \\
\frac{\partial \mathcal{L}}{\partial z_h} &= n_h w_0 c_0 - \sum_{j=0}^{1} \phi_j = 0 \\
\frac{\partial \mathcal{L}}{\partial \xi_j} &= q_j - n_j A_j (1 - e^{-\lambda \bar{z}}) = 0 \\
\frac{\partial \mathcal{L}}{\partial \phi_j} &= \bar{z} - z_h - z_j = 0 \\
\frac{\partial \mathcal{L}}{\partial \kappa} &= \sum_{j=0}^{1} n_j A_j \theta_j e^{-\lambda z_j} - n_h = 0
\end{aligned}$$

A.1.1.3 Comparative statics

Domestic production. By the implicity function theorem,

$$\frac{dz_h}{dx_0} = -\frac{\frac{d(1.13)}{dx_0}}{\frac{d(1.13)}{dz_h}}$$

The sign of $\frac{dz_h}{dx_0}$ is given by $-\frac{d(1.13)}{dx_0}$ because $\frac{d(1.13)}{dz_h} = \lambda\theta_0 e^{-\lambda(\bar{z}-z_h)}(2c_0 + \lambda(1 + c_0 z_h)) > 0$.

$$\frac{d(1.13)}{d\bar{z}} = -\lambda\theta_0 e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1 + c_0 z_h)) < 0 \qquad \Rightarrow \frac{dz_h}{d\bar{z}} > 0$$
$$\frac{d(1.13)}{d\theta_0} = e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1 + c_0 z_h)) > 0 \qquad \Rightarrow \frac{dz_h}{d\theta_0} < 0$$
$$\frac{d(1.13)}{dc_0} = \theta_0 e^{-\lambda(\bar{z}-z_h)}(1 + \lambda z_h) - 1 < 0 \qquad \Rightarrow \frac{dz_h}{dc_0} > 0$$
$$\frac{d(1.13)}{d\lambda} = \theta_0 e^{-\lambda(\bar{z}-z_h)}((1 + c_0 z_h) - (\bar{z} - z_h)(c_0 + \lambda(1 + c_0 z_h))) \qquad \Rightarrow \frac{dz_h}{d\lambda} \quad \text{ambiguous}$$
$$\frac{d(1.13)}{dq_0} = \frac{d(1.13)}{dA_0} = \frac{d(1.13)}{dw_0} = 0 \qquad \Rightarrow \frac{dz_h}{dq_0} = \frac{dz_h}{dA_0} = \frac{dz_h}{dw_0} = 0$$

$\frac{dz_0}{d\theta_0} > 0$ and $\frac{dz_0}{dc_0} < 0$ by $z_0 = \bar{z} - z_h$. w_0, A_0 and q_0 do not affect z_0.

Note that $\frac{dz_h}{d\bar{z}} < 1$ by $\lambda\theta_0 e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1 + c_0 z_h)) < \lambda\theta_0 e^{-\lambda(\bar{z}-z_h)}(2c_0 + \lambda(1 + c_0 z_h))$. This implies that $\frac{dz_0}{d\bar{z}} = 1 - \frac{dz_h}{d\bar{z}} > 0$.

Domestic and foreign production. As the production knowledge and the number of managers are a function of managerial knowledge, the comparative statics for the managerial knowledge are derived first.

Managerial knowledge (4). By the implicit function theorem,

$$\frac{dz_h}{dx_j} = -\frac{\frac{d(1.12)}{dx_j}}{\frac{d(1.12)}{dz_h}}$$

$\frac{d(1.12)}{dz_h} > 0$ by the second order condition for a minimum. This implies

$$\frac{dz_h}{d\bar{z}} = -\left(\frac{\partial(1.12)}{\partial z_h}\right)^{-1}\left(-\lambda(1 + c_0 z_h)\frac{q_{\bar{j}}}{A_{\bar{j}}}\theta_{\bar{j}} e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1 + c_0 z_h))\right) > 0$$

where the subscript $\bar{j}$ refers to the country with binding knowledge constraint $z_{\bar{j}} = \bar{z} - z_h$.

Concerning the other parameters, it is necessary to distinguish two possible cases.

- *The knowledge constraint $z_j = \bar{z} - z_h$ is binding in the home country, slack in the foreign country.*

$$
\begin{aligned}
\frac{\partial z_h}{\partial A_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} \lambda(1+c_0 z_h)\frac{q_0 w_0 c_0}{A_0^2} \quad < 0 \\
\frac{\partial z_h}{\partial A_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} (-c_0)\frac{q_1 w_1 c_1}{A_1^2} \quad > 0 \\
\frac{\partial z_h}{\partial \theta_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} \lambda(1+c_0 z_h) q_0 e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) \quad < 0 \\
\frac{\partial z_h}{\partial \theta_1} &= 0 \\
\frac{\partial z_h}{\partial w_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} \lambda(1+c_0 z_h) \\
&\quad \left(q_0\theta_0 e^{-\lambda(\bar{z}-z_h)}(c_0 + \lambda(1+c_0 z_h)) - \frac{q_0 c_0}{A_0}\right) \quad > 0 \\
\frac{\partial z_h}{\partial w_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} c_0\frac{q_1 c_1}{A_1} \quad < 0 \\
\frac{\partial z_h}{\partial c_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} \left(\frac{q_1 w_1 c_1}{A_1} + \lambda(1+c_0 z_h)\left(q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0(1+\lambda z_h)\right.\right. \\
&\quad \left.\left. - \frac{q_0 w_0}{A_0}\right) + \lambda z_h\left(q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) - \frac{q_0 w_0 c_0}{A_0}\right)\right) \quad > 0 \\
\frac{\partial z_h}{\partial c_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} c_0\frac{q_1 w_1}{A_1} \quad < 0 \\
\frac{\partial z_h}{\partial q_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} \lambda(1+c_0 z_h) \\
&\quad \left(\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) - \frac{w_0 c_0}{A_0}\right) \quad > 0 \\
\frac{\partial z_h}{\partial q_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_0 = \bar{z} - z_h\right)^{-1} c_0\frac{w_1 c_1}{A_1} \quad < 0
\end{aligned}
$$

$\frac{\partial z_h}{\partial c_0} > 0$ follows from $\frac{q_1 w_1 c_1}{A_1} + \lambda(1+c_0 z_h)\left(q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0(1+\lambda z_h) - \frac{q_0 w_0}{A_0}\right) = -\lambda(1+c_0 z_h)\frac{\lambda}{c_0}$ and $q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0(c_0 + \lambda(1+c_0 z_h)) - \frac{q_0 w_0 c_0}{A_0} < 0$ by (1.12).

- *The knowledge constraint $z_j = \bar{z} - z_h$ is binding in the foreign country, slack in the home country.*

$$
\frac{\partial z_h}{\partial A_0} = -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \frac{-c_0^2 q_0 w_0}{A_0^2} \quad > 0
$$

$$
\begin{aligned}
\frac{\partial z_h}{\partial A_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \frac{q_1 w_1 c_1}{A_1^2} \quad < 0 \\
\frac{\partial z_h}{\partial \theta_0} &= 0 \\
\frac{\partial z_h}{\partial \theta_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \lambda(1+c_0 z_h) q_1 e^{-\lambda(\bar{z}-z_h)} w_0 (c_0 + \lambda(1+c_0 z_h)) < 0 \\
\frac{\partial z_h}{\partial w_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \\
&\quad \left(c_0^2 \frac{q_0}{A_0} + \lambda(1+c_0 z_h) q_1 \theta_1 e^{-\lambda(\bar{z}-z_h)} (c_0 + \lambda(1+c_0 z_h))\right) < 0 \\
\frac{\partial z_h}{\partial w_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \frac{-q_1 c_1}{A_1} \quad > 0 \\
\frac{\partial z_h}{\partial c_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \left(2 c_0 \frac{q_0 w_0}{A_0} + \lambda(1+c_0 z_h) q_1 \theta_1 e^{-\lambda(\bar{z}-z_h)} w_0 (1 + \lambda z_h)\right. \\
&\quad \left. + \lambda z_h \left(q_1 \theta_1 e^{-\lambda(\bar{z}-z_h)} w_0 (c_0 + \lambda(1+c_0 z_h)) - \frac{q_1 w_1 c_1}{A_1}\right)\right) \quad < 0 \\
\frac{\partial z_h}{\partial c_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \frac{-q_1 w_1}{A_1} \lambda(1+c_0 z_h) \quad > 0 \\
\frac{\partial z_h}{\partial q_0} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} c_0^2 \frac{w_0}{A_0} \quad < 0 \\
\frac{\partial z_h}{\partial q_1} &= -\left(\frac{\partial(1.12)}{\partial z_h}|z_1 = \bar{z} - z_h\right)^{-1} \\
&\quad \lambda(1+c_0 z_h) \left(\theta_1 e^{-\lambda(\bar{z}-z_h)} w_0 (c_0 + \lambda(1+c_0 z_h)) - \frac{w_1 c_1}{A_1}\right) > 0
\end{aligned}
$$

In principle, the knowledge constraint may be binding for both countries if $\theta_{\bar{j}} A_{\bar{j}} w_j c_j = \theta_j A_j w_{\bar{j}} c_{\bar{j}}$. We do not derive the comparative statics for this case, though, as changes in the variables will lead to the occurrence of either of the other two cases.

Number of production workers (1).

$$
\begin{aligned}
\frac{\partial n_j}{\partial q_j} &= \frac{1}{A_j(1 - e^{-\lambda \bar{z}})} \quad > 0 \\
\frac{\partial n_j}{\partial A_j} &= -\frac{q_j}{A_j^2 (1 - e^{-\lambda \bar{z}})} \quad < 0 \\
\frac{\partial n_j}{\partial \bar{z}} &= -\frac{q_j \lambda e^{-\lambda \bar{z}}}{A_j (1 - e^{-\lambda \bar{z}})^2} \quad < 0
\end{aligned}
$$

Production knowledge (2).

- *The knowledge constraint $z_j = \bar{z} - z_h$ is binding in the home country, slack in the foreign country.*

$$\begin{aligned}
& \frac{\partial z_0}{\partial x_j} = -\frac{\partial z_h}{\partial x_j} \\
\Rightarrow \quad & \frac{\partial z_0}{\partial A_0} > 0,\ \frac{\partial z_0}{\partial \theta_0} > 0,\ \frac{\partial z_0}{\partial w_0} < 0,\ \frac{\partial z_0}{\partial c_0} < 0,\ \frac{\partial z_0}{\partial q_0} < 0 \\
& \frac{\partial z_1}{\partial A_1} = -\frac{\frac{w_1 c_1}{\lambda w_0(1+c_0 z_h)A_1^2\theta_1} + \frac{c_0 w_1 c_1 \partial z_h/\partial A_1}{\lambda w_0(1+c_0 z_h)^2 A_1\theta_1}}{-\lambda e^{-\lambda z_1}} \quad > 0 \\
& \frac{\partial z_1}{\partial \theta_1} = \frac{1}{\lambda\theta_1} \quad > 0 \\
& \frac{\partial z_1}{\partial w_1} = -\frac{-\frac{c_1}{\lambda w_0(1+c_0 z_h)A_1\theta_1} + \frac{c_0 w_1 c_1 \partial z_h/\partial w_1}{\lambda w_0(1+c_0 z_h)^2 A_1\theta_1}}{-\lambda e^{-\lambda z_1}} \quad < 0 \\
& \frac{\partial z_1}{\partial c_1} = -\frac{-\frac{w_1}{\lambda w_0(1+c_0 z_h)A_1\theta_1} + \frac{c_0 w_1 c_1 \partial z_h/\partial c_1}{\lambda w_0(1+c_0 z_h)^2 A_1\theta_1}}{-\lambda e^{-\lambda z_1}} \quad < 0 \\
& \frac{\partial z_1}{\partial q_1} = -\frac{\frac{c_0 w_1 c_1 \partial z_h/\partial q_1}{\lambda w_0(1+c_0 z_h)^2 A_1\theta_1}}{-\lambda e^{-\lambda z_1}} \quad < 0
\end{aligned}$$

- *The knowledge constraint $z_j = \bar{z} - z_h$ is binding in the foreign country, slack in the home country.*

$$\begin{aligned}
& \frac{\partial z_1}{\partial x_j} = -\frac{\partial z_h}{\partial x_j} \\
\Rightarrow \quad & \frac{\partial z_1}{\partial A_1} > 0,\ \frac{\partial z_1}{\partial \theta_1} > 0,\ \frac{\partial z_1}{\partial w_1} < 0,\ \frac{\partial z_1}{\partial c_1} < 0,\ \frac{\partial z_1}{\partial q_1} < 0 \\
& \frac{\partial z_0}{\partial A_0} = -\frac{\frac{c_0}{\lambda(1+c_0 z_h)A_0^2\theta_0} + \frac{c_0^2 \partial z_h/\partial A_0}{\lambda(1+c_0 z_h)^2 A_0\theta_0}}{-\lambda e^{-\lambda z_0}} \quad > 0 \\
& \frac{\partial z_0}{\partial \theta_0} = \frac{1}{\lambda\theta_0} \quad > 0 \\
& \frac{\partial z_0}{\partial w_0} = -\frac{\frac{c_0^2 \partial z_h/\partial w_0}{\lambda(1+c_0 z_h)^2 A_0\theta_0}}{-\lambda e^{-\lambda z_0}} \quad < 0 \\
& \frac{\partial z_0}{\partial c_0} = -\frac{-\frac{1}{\lambda(1+c_0 z_h)A_0\theta_0} + \frac{c_0 \partial z_h/\partial c_0 + z_h}{\lambda(1+c_0 z_h)^2 A_0\theta_0}}{-\lambda e^{-\lambda z_0}} \quad < 0 \quad \text{if } z_h < \lambda(1+c_0 z_h) \\
& \frac{\partial z_0}{\partial q_0} = -\frac{\frac{c_0^2 \partial z_h/\partial q_0}{\lambda(1+c_0 z_h)^2 A_0\theta_0}}{-\lambda e^{-\lambda z_0}} \quad < 0
\end{aligned}$$

Number of managers (3).

$$\frac{\partial n_h}{\partial A_j} = -\frac{\lambda e^{-\lambda z_j} q_j \theta_j \frac{\partial z_j}{\partial A_j}}{1-e^{-\lambda \bar{z}}} \quad < 0$$
$$\frac{\partial n_h}{\partial \theta_j} = \frac{q_j e^{-\lambda z_j} - \lambda e^{-\lambda z_j} q_j \theta_j \frac{\partial z_j}{\partial \theta_j}}{1-e^{-\lambda \bar{z}}}$$
$$\frac{\partial n_h}{\partial w_j} = -\frac{\lambda e^{-\lambda z_j} q_j \theta_j \frac{\partial z_j}{\partial w_j}}{1-e^{-\lambda \bar{z}}} \quad > 0$$
$$\frac{\partial n_h}{\partial c_j} = -\frac{\lambda e^{-\lambda z_j} q_j \theta_j \frac{\partial z_j}{\partial c_j}}{1-e^{-\lambda \bar{z}}} \quad > 0$$
$$\frac{\partial n_h}{\partial q_j} = \frac{\theta_j e^{-\lambda z_j} - \lambda e^{-\lambda z_j} q_j \theta_j \frac{\partial z_j}{\partial q_j}}{1-e^{-\lambda \bar{z}}} \quad > 0$$
$$\frac{\partial n_h}{\partial \bar{z}} = -\frac{\lambda e^{-\lambda \bar{z}}}{(1-e^{-\lambda \bar{z}})^2} \sum_j q_j \theta_j e^{-\lambda z_j} - \frac{1}{1-e^{-\lambda \bar{z}}} \sum_j \lambda e^{-\lambda z_j} q_j \theta_j \frac{\partial z_j}{\partial \bar{z}} \quad < 0$$

A.1.2 Profit maximization

A.1.2.1 The effect of the parameters on marginal costs

- $\phi_j = 0$, $\xi_j = \frac{1}{A_j(1-e^{-\lambda \bar{z}})} \left(w_j(1+c_j z_j) + \frac{1}{\lambda} w_j c_j \right)$

Effect of production quantity in country j on marginal costs in country j :

$$\frac{\partial \xi_j}{\partial q_j} = \frac{1}{A_j(1-e^{-\lambda \bar{z}})} w_j c_j \frac{\partial z_j}{\partial q_j} \quad < 0$$

Effect of production quantity in country $\bar{j}$ on marginal costs in country j :

$$\frac{\partial \xi_j}{\partial q_{\bar{j}}} = \frac{1}{A_j(1-e^{-\lambda \bar{z}})} w_j c_j \frac{\partial z_j}{\partial q_{\bar{j}}} \quad > 0$$

Effect of other characteristics of country j on marginal costs in country j :

$$\frac{\partial \xi_j}{\partial \theta_j} = \frac{1}{A_j(1-e^{-\lambda \bar{z}})} w_j c_j \frac{1}{\lambda \theta_j} \quad > 0$$
$$\frac{\partial \xi_j}{\partial w_j} = \frac{1}{A_j(1-e^{-\lambda \bar{z}})} \left(1 + c_j z_j + \frac{1}{\lambda} \frac{c_0}{1+c_0 z_h} w_j c_j \frac{\partial z_h}{\partial w_j} \right) \quad > 0$$
$$\frac{\partial \xi_j}{\partial c_j} = \frac{w_j}{A_j(1-e^{-\lambda \bar{z}})} \left(z_j + c_j \frac{1}{\lambda} \frac{c_0}{1+c_0 z_h} \frac{\partial z_h}{\partial c_j} \right) \quad > 0$$
$$\frac{\partial \xi_j}{\partial A_j} = \frac{1}{A_j(1-e^{-\lambda \bar{z}})} \left(-\frac{1}{A_j} w_j \left(1 + c_j z_j + \frac{1}{\lambda} c_j \right) + w_j c_j \frac{\partial z_j}{\partial A_j} \right) \quad < 0$$

- $\phi_{\bar{j}} \neq 0$, $\xi_{\bar{j}} = \frac{1}{A_{\bar{j}}(1-e^{-\lambda\bar{z}})}\left(w_{\bar{j}}(1+c_{\bar{j}}(\bar{z}-z_h)) + w_0(1+c_0 z_h)A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}\right)$

Effect of production quantity in country $\bar{j}$ on marginal costs in country $\bar{j}$:

$$\frac{\partial\xi_{\bar{j}}}{\partial q_{\bar{j}}} = \frac{1}{A_{\bar{j}}(1-e^{-\lambda\bar{z}})}(-w_{\bar{j}}c_{\bar{j}} + A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)))\frac{\partial z_h}{\partial q_{\bar{j}}} \quad < 0$$

Effect of production quantity in country j on marginal costs in country $\bar{j}$:

$$\frac{\partial\xi_{\bar{j}}}{\partial q_j} = \frac{1}{A_{\bar{j}}(1-e^{-\lambda\bar{z}})}(-w_{\bar{j}}c_{\bar{j}} + A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)))\frac{\partial z_h}{\partial q_j} \quad > 0$$

Effect of other characteristics of country $\bar{j}$ on marginal costs in country $\bar{j}$:

$$\begin{aligned}
\frac{\partial\xi_{\bar{j}}}{\partial \theta_{\bar{j}}} &= \frac{1}{A_{\bar{j}}(1-e^{-\lambda\bar{z}})}\Bigg((-w_{\bar{j}}c_{\bar{j}} + A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)))\frac{\partial z_h}{\partial q_{\theta j}} \\
&+ A_{\bar{j}}w_0(1+c_0 z_h)e^{-\lambda(\bar{z}-z_h)}\Bigg) \quad > 0 \\
\frac{\partial\xi_{\bar{j}}}{\partial w_{\bar{j}}} &= \frac{1}{A_{\bar{j}}(1-e^{-\lambda\bar{z}})}\Bigg(1+c_{\bar{j}}(\bar{z}-z_h) + A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}(1+c_0 z_h) \\
&+ \frac{\partial z_h}{\partial w_{\bar{j}}}(-w_{\bar{j}}c_{\bar{j}} + A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)))\Bigg) \quad > 0 \\
\frac{\partial\xi_{\bar{j}}}{\partial c_{\bar{j}}} &= \frac{1}{A_{\bar{j}}(1-e^{-\lambda\bar{z}})}\Bigg(w_{\bar{j}}(\bar{z}-z_h) + w_0 z_h A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)} \\
&+ \frac{\partial z_h}{\partial c_{\bar{j}}}(-w_{\bar{j}}c_{\bar{j}} + A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)))\Bigg) \quad > 0 \\
\frac{\partial\xi_{\bar{j}}}{\partial A_{\bar{j}}} &= \frac{1}{A_{\bar{j}}(1-e^{-\lambda\bar{z}})}\Bigg(-\frac{1}{A_{\bar{j}}}w_{\bar{j}}(1+c_{\bar{j}}(\bar{z}-z_h) \\
&+ \frac{\partial z_h}{\partial A_{\bar{j}}}(-w_{\bar{j}}c_{\bar{j}} + A_{\bar{j}}\theta_{\bar{j}}e^{-\lambda(\bar{z}-z_h)}w_0(c_0+\lambda(1+c_0 z_h)))\Bigg) \quad < 0
\end{aligned}$$

The effects of w_j, c_j and A_j are mitigated by organizational adjustments. $\frac{\partial\xi_j}{\partial w_j} > 0$, $\frac{\partial\xi_j}{\partial c_j} > 0$ and $\frac{\partial\xi_j}{\partial A_j} < 0$ because by the envelope property, the minimum value function with variable z_h is at least as concave as the minimum value function with fixed z_h.

A.1.2.2 The effect of the parameters on output and sales

Differentiating Equation (1.19) with respect to q_j yields:

$$\frac{d\pi^I(\cdot)}{dq_j} = \frac{\sigma-1}{\sigma}q_j(\bar{z}_i)^{-\frac{1}{\sigma}}Q_j^{\frac{1}{\sigma}} - \xi_j = 0 \qquad \text{(F.O.C.)}$$

$$\frac{d^2\pi^I(\cdot)}{dq_j^2} = -\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_j(\bar{z}_i)^{-\frac{1}{\sigma}-1}Q_j^{\frac{1}{\sigma}} - \frac{d\xi_j}{dq_j} < 0 \qquad \text{(S.O.C.)}$$

We differentiate Equation (1.21) to determine how output varies with parameters x_j:

$$\frac{dq_j}{dx_j} = -\frac{Q_j\sigma\left(\frac{\sigma}{\sigma-1}\xi_j\right)^{-\sigma-1}\frac{\sigma}{\sigma-1}\frac{d\xi_j}{dx_j}}{1+Q_j\sigma\left(\frac{\sigma}{\sigma-1}\xi_j\right)^{-\sigma-1}\frac{\sigma}{\sigma-1}\frac{d\xi_j}{dq_j}}$$

To determine the sign of this term, note that its denominator is obtained by multiplying the second order condition by $-\left(\frac{1}{\sigma}\frac{\sigma-1}{\sigma}q_j(\bar{z}_i)^{-\frac{1}{\sigma}-1}Q_j^{\frac{1}{\sigma}}\right)^{-1}$. The denominator is thus the product of two negative terms, so it is positive. The sign of $\frac{dq_j}{dx_j}$ given by $-\frac{d\xi_j}{dx_j}$: output is decreasing in the communication costs, in wages, and in the knowledge acquisition costs, and increasing in the labor productivity.

The sales also vary inversely with marginal costs:

$$\begin{aligned}\frac{dp(q_j)q_j}{dx_j} &= (1-\sigma)Q_j\left(\frac{\sigma}{\sigma-1}\xi_j\right)^{-\sigma}\frac{\sigma}{\sigma-1}\left(\frac{d\xi_j}{dx_j}+\frac{d\xi_j}{dq_j}\frac{dq_j}{dx_j}\right)\\ &= (1-\sigma)Q_j\left(\frac{\sigma}{\sigma-1}\xi_j\right)^{-\sigma}\frac{\sigma}{\sigma-1}\frac{d\xi_j}{dx_j}\frac{1}{1+Q_j\sigma\left(\frac{\sigma}{\sigma-1}\xi_j\right)^{-\sigma-1}\frac{\sigma}{\sigma-1}\frac{d\xi_j}{dq_j}}\end{aligned}$$

The sign of $\frac{dp(q_j)q_j}{dx_j}$ is $-\frac{d\xi_j}{dx_j}$ by $\sigma > 1$.

The arguments of ξ_j, q_j are omitted for simplicity of exposition.

A.1.3 General equilibrium

A.1.3.1 Closed economy

- Comparative statics: the marginal costs of production ξ

$$\begin{aligned}\frac{\partial\xi}{\partial\bar{z}} &= -\frac{\lambda e^{-\lambda\bar{z}}}{A(1-e^{-\lambda\bar{z}})^2}(w(1+c(\bar{z}-z_h))+A\theta we^{-\lambda(\bar{z}-z_h)}(1+cz_h))\\ &+\frac{1}{A(1-e^{-\lambda\bar{z}})}(wc-\lambda A\theta e^{-\lambda(\bar{z}-z_h)}w(1+cz_h))\\ &= -\frac{1}{A(1-e^{-\lambda\bar{z}})^2}(\lambda e^{-\lambda\bar{z}}w(1+c(\bar{z}-z_h))+e^{-\lambda\bar{z}}wc-A\theta wce^{-\lambda(\bar{z}-z_h)})\end{aligned}$$

$$
\begin{aligned}
&< 0 \quad \text{by Assumption 1} \\
\frac{\partial \xi}{\partial q} &= 0 \quad \text{by } \frac{dz_h}{dq} = 0 \\
\frac{\partial \xi}{\partial c} &= \frac{1}{A(1-e^{-\lambda \bar{z}})}(w(\bar{z}-z_h) + \theta A w e^{-\lambda(\bar{z}-z_h)} z_h) \quad > 0 \\
\frac{\partial \xi}{\partial w} &= \frac{1}{A(1-e^{-\lambda \bar{z}})}(1 + c(\bar{z}-z_h) + \theta A e^{-\lambda(\bar{z}-z_h)}(1 + c z_h)) \quad > 0 \\
\frac{\partial \xi}{\partial \theta} &= \frac{1}{(1-e^{-\lambda \bar{z}})} e^{-\lambda(\bar{z}-z_h)} w(1 + c z_h) \quad > 0 \\
\frac{\partial \xi}{\partial A} &= -\frac{1}{A^2(1-e^{-\lambda \bar{z}})} w(1 + c(\bar{z}-z_h)) \quad < 0
\end{aligned}
$$

- Comparative statics: cut-off productivity level

$$
\begin{aligned}
\frac{d\bar{z}^*}{dx_j} &= -\frac{\frac{d\xi_j}{dx_j}}{\frac{d\xi_j}{d\bar{z}^*}} \quad \text{by the implicit function theorem} \\
\Rightarrow \quad \operatorname{sgn}\left(\frac{d\bar{z}^*}{dx_j}\right) &= \operatorname{sgn}\left(\frac{d\xi_j}{dx_j}\right) \quad \text{as } \frac{d\xi_j}{d\bar{z}^*} < 0
\end{aligned}
$$

A.1.3.2 Open Economy

- Comparative statics: the profits from foreign direct investment

$$
\begin{aligned}
\pi_0^I(\bar{z}_i) &= \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}\Big(Q_0 \xi_0(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1)^{1-\sigma} \\
&\quad + Q_1 \xi_1(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1)^{1-\sigma}\Big) \\
\frac{\partial \pi_0^I}{\partial \bar{z}_i} &= -\left(\frac{\sigma}{\sigma-1}\right)^{-\sigma}\Bigg(Q_0 \xi_0(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1)^{-\sigma}\frac{\partial \xi_0}{\partial \bar{z}} \\
&\quad + Q_1 \xi_1(\bar{z}_i, q_0^I(\bar{z}_i), q_1^I(\bar{z}_i), w_0, w_1)^{-\sigma}\frac{\partial \xi_1}{\partial \bar{z}}\Bigg) \\
&= -\frac{\partial C}{\partial \bar{z}_i} \quad \geq 0 \quad \text{by Assumption 1}
\end{aligned}
$$

Generally,

$$
\frac{\partial \pi_0^I}{\partial x_j} = -\frac{\partial C}{\partial x_j}
$$

Correspondingly,

$$
\begin{aligned}
\frac{\partial \pi_0^I}{\partial w_1} &= -\left(\frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}(1+c_1 z_1+\frac{1}{\lambda}c_1)\right) \quad < 0 && \text{for } z_1 > \bar{z}_i - z_h \\
\frac{\partial \pi_0^I}{\partial w_1} &= -\left(\frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}(1+c_1(\bar{z}_i - z_h))\right) \quad < 0 && \text{for } z_1 = \bar{z}_i - z_h \\
\frac{\partial \pi_0^I}{\partial c_1} &= -\left(\frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}(w_1 z_1+\frac{1}{\lambda}w_1)\right) \quad < 0 && \text{for } z_1 > \bar{z}_i - z_h \\
\frac{\partial \pi_0^I}{\partial c_1} &= -\left(\frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}(w_1(\bar{z}_i - z_h))\right) \quad < 0 && \text{for } z_1 = \bar{z}_i - z_h \\
\frac{\partial \pi_0^I}{\partial \theta_1} &= -\left(\frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}w_1 c_1 \frac{1}{\lambda \theta_1}\right) \quad < 0 && \text{for } z_1 > \bar{z}_i - z_h \\
\frac{\partial \pi_0^I}{\partial \theta_1} &= -\left(\frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}w_0(1+c_0 z_h)A_1 e^{-\lambda(\bar{z}_i - z_h)}\right) \quad < 0 && \text{for } z_1 = \bar{z}_i - z_h \\
\frac{\partial \pi_0^I}{\partial A_1} &= -\left(-\frac{q_1}{A_1^2(1-e^{-\lambda \bar{z}})}w_1(1+c_1 z_1+\frac{1}{\lambda}c_1)\right) \quad > 0 && \text{for } z_1 > \bar{z}_i - z_h \\
\frac{\partial \pi_0^I}{\partial A_1} &= -\left(-\frac{q_1}{A_1^2(1-e^{-\lambda \bar{z}})}w_1(1+c_1(\bar{z}_i - z_h))\right) \quad > 0 && \text{for } z_1 = \bar{z}_i - z_h
\end{aligned}
$$

The effect of λ:

$$
\begin{aligned}
\frac{\partial^2 \pi_0^I}{\partial \theta_1 \partial \lambda} &= -\left(-\frac{q_1 \bar{z}_i e^{-\lambda \bar{z}_i}}{A_1(1-e^{-\lambda \bar{z}_i})^2}w_1 c_1 \frac{1}{\lambda \theta_1} - \frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}w_1 c_1 \frac{1}{\lambda^2 \theta_1}\right) \\
&< 0 \quad \text{for } z_1 > \bar{z}_i - z_h \\
\frac{\partial^2 \pi_0^I}{\partial \theta_1 \partial \lambda} &= -\left(-\frac{q_1 \bar{z}_i e^{-\lambda \bar{z}_i}}{A_1(1-e^{-\lambda \bar{z}_i})^2}w_0(1+c_0 z_h)A_1 e^{-\lambda(\bar{z}_i - z_h)}\right. \\
&\quad \left. -(\bar{z}_i - z_h)e^{-\lambda(\bar{z}_i - z_h)}\frac{q_1}{A_1(1-e^{-\lambda \bar{z}_i})}w_0(1+c_0 z_h)A_1\right) \\
&< 0 \quad \text{for } z_1 = \bar{z}_i - z_h
\end{aligned}
$$

- Comparative statics: a multinational firm's cut-off knowledge level $\bar{z}^I$

$$
\begin{aligned}
\frac{d\bar{z}^I}{dx_j} &= -\frac{\frac{d\pi^X}{dx_j} - \frac{d\pi^I}{dx_j}}{\frac{d\pi^X}{d\bar{z}^I} - \frac{d\pi^I}{d\bar{z}^I}} \quad \text{by the implicit function theorem} \\
\Rightarrow \quad \operatorname{sgn}\left(\frac{d\bar{z}^I}{dx_j}\right) &= \operatorname{sgn}\left(\frac{d\pi^X}{dx_j} - \frac{d\pi^I}{dx_j}\right) \quad \text{as } \frac{d\pi^X}{d\bar{z}^I} - \frac{d\pi^I}{d\bar{z}^I} < 0 \\
\Rightarrow \frac{d\bar{z}^I}{d\tau_1} &< 0,\ \frac{d\bar{z}^I}{dw_1} > 0,\ \frac{d\bar{z}^I}{dc_1} > 0,\ \frac{d\bar{z}^I}{d\theta_1} > 0,\ \frac{d\bar{z}^I}{dA_1} < 0
\end{aligned}
$$

A.1.4 Robustness

The cost minimization problem is given by

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

$$\begin{aligned} \text{s.t.} \quad n_j A_j (1 - e^{-\lambda \bar{z}}) &\geq q_j \qquad \forall j \\ z_j &\geq \bar{z} - z_h^j \quad \forall j \\ n_h &\geq \sum_{j=0}^{1} n_j A_j \theta_j e^{-\lambda z_j} \\ n_h j \geq 0,\; z_h^j &\geq 0, z_h j \leq \bar{z} \\ n_j \geq 0,\; z_j &\geq 0, z_j \leq \bar{z} \qquad \forall j \end{aligned}$$

- Both knowledge constraints are binding: $z_j = \bar{z} - z_h^j$.
- The number of production workers is determined as before. The number of managers responsible for production in country j is given by $n_h^j = \frac{q_j \theta_j e^{-\lambda(\bar{z} - z_h)}}{1 - e^{-\lambda \bar{z}}}$.
- Managerial knowledge levels are given by $\frac{w_j c_j}{A_j} = \theta_j e^{-\lambda(\bar{z} - z_h)} w_0 (c_0 + \lambda(1 + c_0 z_h^j))$.
- The comparative statics follow from similar considerations as in the domestic case, see Section A.1.3.1.

A.2 Additional regression results

Table A.1: Regression results for the within-firm differences in productivity across countries, all affiliates

	1	2	3	4	5	6	7	8	9
Log GDP	0.143***	0.059***	0.077***	0.097***	0.073***	0.039***	0.051***	0.101***	0.055***
	(0.006)	(0.008)	(0.011)	(0.011)	(0.008)	(0.009)	(0.009)	(0.012)	(0.008)
Log flight time	−0.160***	−0.047***	−0.008	−0.035**	−0.038***	−0.058***	−0.084***	−0.063***	−0.102***
	(0.008)	(0.011)	(0.013)	(0.014)	(0.011)	(0.013)	(0.014)	(0.018)	(0.025)
Linguistic distance	−0.631***	−0.066*	-0.055^{+}	−0.068*	-0.055^{+}	−0.134***	−0.049	−0.052	−0.062*
	(0.029)	(0.029)	(0.033)	(0.031)	(0.029)	(0.036)	(0.033)	(0.035)	(0.029)
Time difference to Germany	0.001	0.020***	0.036***	0.041***	0.019***	0.024***	0.018***	0.044***	0.019***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)
Public expenditure per pupil, % GDP per capita		0.017***	0.018***	0.012***	0.018***	0.019***	0.019***	0.016***	0.016***
		(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.004)	(0.002)
GDP per employee		0.208***	0.160***	0.109***	0.211***	0.232***	0.189***	0.142***	0.215***
		(0.008)	(0.012)	(0.015)	(0.008)	(0.009)	(0.009)	(0.017)	(0.008)
Log compensation per employee			-0.012^{+}						
			(0.006)						
Trust, survey 1996				0.255***				0.211***	
				(0.032)				(0.039)	
Statutory tax rate					−0.007***			−0.008***	
					(0.001)			(0.002)	
Log costs to enforce contracts						0.021			
						(0.024)			
Log time to enforce contracts						−0.002		0.172***	
						(0.021)		(0.019)	
# of procedures to enforce contracts						0.012***			
						(0.002)			
Log costs to register property							−0.002		
							(0.007)		
Log time to register property							−0.082***	−0.077***	
							(0.008)	(0.008)	
# of procedures to register property							0.011**		
							(0.004)		
Log distance									0.041*
									(0.018)
Constant	5.004***	3.915***	3.791***	3.246***	3.858***	3.746***	4.484***	2.627***	3.951***
	(0.077)	(0.119)	(0.155)	(0.156)	(0.118)	(0.190)	(0.155)	(0.260)	(0.119)
R^2	0.092	0.191	0.133	0.136	0.192	0.199	0.203	0.152	0.192
Observations	159,870	125,449	97,370	105,509	125,214	87,781	77,235	63,240	125,449
# parents	8,987	8,325	7,368	7,794	8,316	6,738	6,367	5,921	8,325
# country combinations	9,362	9,185	8,775	8,800	9,184	6,851	6,239	5,933	9,185

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Dependent variable:* foreign productivity defined as $\ln \frac{\text{foreign sales}}{\text{foreign employees}}$. *Covariate definitions:* see Table 1.1. Sample based on all foreign affiliates which meet uniform reporting requirements (at least 10% participation and balance sheet total of at least 5 million euro or at least 50% participation if balance sheet total between 3 and 5 million euro).

Table A.2: Regression results for the within-firm differences in productivity across countries, including wages

	Alternative determinants, including wages					
	1	2	3	4	5	6
Log GDP	0.021	0.081***	0.057***	0.110***	0.059**	0.088***
	(0.015)	(0.011)	(0.014)	(0.014)	(0.022)	(0.011)
Log flight time	0.059**	−0.009	−0.038*	−0.066***	−0.035	0.076*
	(0.019)	(0.013)	(0.016)	(0.017)	(0.023)	(0.035)
Linguistic distance	0.021	−0.080*	−0.275***	−0.094*	−0.066	−0.076*
	(0.036)	(0.035)	(0.045)	(0.043)	(0.053)	(0.034)
Time difference to Germany	0.057***	0.036***	0.033***	0.031***	0.045***	0.037***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.006)	(0.003)
Public expenditure per pupil,	0.011***	0.019***	0.012**	0.023***	0.013**	0.021***
% of GDP p.c.	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)
GDP per employee	0.211***	0.137***	0.202***	0.106***	0.183***	0.135***
	(0.019)	(0.013)	(0.015)	(0.015)	(0.029)	(0.013)
Log compensation per employee	−0.033***	−0.009	0.019*	−0.005	0.005	−0.013*
	(0.008)	(0.006)	(0.008)	(0.006)	(0.011)	(0.006)
Trust, survey 1996	0.005				0.092+	
	(0.034)				(0.050)	
Statutory tax rate		0.004*			−0.003	
		(0.002)			(0.003)	
Log costs to enforce contract			0.172***			
			(0.034)			
Log time to enforce contracts			0.060+		0.197***	
			(0.033)		(0.027)	
# procedures to enforce contracts			0.013***			
			(0.002)			
Log costs to register property				0.028**		
				(0.010)		
Log time to register property				−0.073***	−0.075***	
				(0.011)	(0.009)	
# procedures to register property				−0.007		
				(0.006)		
Log distance						−0.056**
						(0.022)
Constant	4.353***	3.666***	2.698***	3.945***	2.916***	3.682***
	(0.198)	(0.159)	(0.288)	(0.234)	(0.483)	(0.162)
R^2	0.132	0.136	0.144	0.143	0.148	0.137
Observations	87,371	91,600	62,856	55,385	52,530	91,600
# parents	6,775	6,867	5,469	5,182	5,109	6,867
# country combinations	8,498	8,628	6,399	5,821	5,714	8,628

Standard errors in parentheses. $^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Dependent variable:* foreign productivity defined as $\ln \frac{\text{foreign sales}}{\text{foreign employees}}$. *Covariate definitions:* see Table 1.1. Sample based on majority owned foreign affiliates.

Table A.3: Regression results for the cut-off productivity levels, additional regressors

	1	2	3	4	5	6	7
Domestic productivity							
Log GDP	−0.178**	−0.177***	−0.194***	−0.171***	−0.170**	−0.172***	−0.173***
	(0.049)	(0.030)	(0.032)	(0.032)	(0.048)	(0.031)	(0.030)
Log flight time	0.325***	0.203***	0.190***	0.233***	0.277***	0.207***	0.214***
	(0.058)	(0.058)	(0.054)	(0.055)	(0.064)	(0.052)	(0.052)
Linguistic distance	0.472**	0.406***	0.277*	0.452***	0.510***	0.339**	0.347**
	(0.126)	(0.098)	(0.107)	(0.096)	(0.105)	(0.106)	(0.115)
Time difference to Germany	0.045**	0.010	0.012	0.007	0.043**	0.013	0.012
	(0.015)	(0.009)	(0.008)	(0.009)	(0.013)	(0.009)	(0.009)

(Continued on next page)

Table A.3: Regression results for the cut-off productivity levels, additional regressors

	1	2	3	4	5	6	7
GDP per employee	0.067	−0.007	0.043	−0.011	0.085	0.072	0.074
	(0.055)	(0.033)	(0.043)	(0.039)	(0.068)	(0.052)	(0.054)
Public expenditure per pupil,	0.022	0.008	0.007	0.014	0.032*	0.013	0.012
% GDP p.c.	(0.012)	(0.011)	(0.010)	(0.011)	(0.013)	(0.010)	(0.010)
Bilateral trust	0.063				−0.098		
	(0.093)				(0.199)		
Statutory tax rate		0.008			−0.000		
		(0.005)			(0.004)		
Cost of enforcing contracts			0.001				
			(0.001)				
Time to enforce contracts			−0.000		−0.000		
			(0.000)		(0.000)		
# of proc. to enforce contracts			0.026*				
			(0.011)				
Cost of registering property				0.001			
				(0.010)			
Time to register property				−0.000	−0.001		
				(0.001)	(0.001)		
# of proc. to register property				−0.003			
				(0.016)			
Rule of Law						−0.175*	
						(0.068)	
Government Effectiveness							−0.182*
							(0.078)
R^2	0.387	0.327	0.338	0.342	0.391	0.329	0.329
Observations	5,280	8,232	8,200	7,231	4,627	8,280	8,280
Global productivity							
Log GDP	−0.153**	−0.187***	−0.173***	−0.165***	−0.140***	−0.160***	−0.158***
	(0.041)	(0.031)	(0.034)	(0.037)	(0.033)	(0.035)	(0.033)
Log flight time	0.219***	0.214***	0.211***	0.223***	0.151*	0.232***	0.235***
	(0.057)	(0.054)	(0.048)	(0.050)	(0.057)	(0.051)	(0.053)
Linguistic distance	0.597***	0.423**	0.262	0.435***	0.601***	0.374**	0.385**
	(0.113)	(0.130)	(0.132)	(0.116)	(0.092)	(0.128)	(0.134)
Time difference to Germany	0.059***	0.005	0.010	0.004	0.056***	0.006	0.005
	(0.015)	(0.008)	(0.010)	(0.009)	(0.012)	(0.010)	(0.010)
GDP per employee	0.170**	0.082*	0.127**	0.090	0.171**	0.127	0.120
	(0.051)	(0.039)	(0.045)	(0.051)	(0.052)	(0.069)	(0.066)
Public expenditure per pupil,	0.017	0.013	0.010	0.014	0.028*	0.016	0.015
% GDP p.c.	(0.015)	(0.010)	(0.008)	(0.010)	(0.012)	(0.009)	(0.009)
Bilateral trust	0.152				−0.043		
	(0.107)				(0.120)		
Statutory tax rate		0.014*			0.003		
		(0.007)			(0.005)		
Cost of enforcing contracts			−0.000				
			(0.001)				
Time to enforce contracts			−0.000		−0.000		
			(0.000)		(0.000)		
# of proc. to enforce contracts			0.024*				
			(0.009)				
Cost of registering property				0.005			
				(0.011)			
Time to register property				−0.000	−0.001**		
				(0.001)	(0.000)		
# of proc. to register property				0.017			
				(0.017)			
Rule of Law						−0.101	
						(0.079)	
Government Effectiveness							−0.087
							(0.083)
R^2	0.465	0.366	0.366	0.364	0.475	0.358	0.357
Observations	6,237	10,228	10,184	8,958	5,442	10,309	10,309

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All regressions include a constant that is omitted for space considerations.

Table A.4: Regression results for the cut-off productivity levels, alternative regressors

Domestic productivity	1	2	3	4	5	6
Log GDP	−0.159***	−0.196***		−0.185***	−0.182***	−0.146***
	(0.029)	(0.039)		(0.028)	(0.029)	(0.023)
Log flight time	0.212***	0.185**	0.209***	0.169***	0.172***	0.331***
	(0.055)	(0.053)	(0.056)	(0.035)	(0.034)	(0.052)
Linguistic distance			0.399***	0.460***	0.493***	0.373***
			(0.100)	(0.098)	(0.096)	(0.090)
Time difference to Germany	0.008	−0.017	0.009	0.018*	0.017*	0.010
	(0.008)	(0.012)	(0.009)	(0.009)	(0.008)	(0.012)
GDP per employee	−0.002	0.002	−0.002	0.082	0.079	
	(0.036)	(0.038)	(0.036)	(0.042)	(0.042)	
Public expenditure per pupil, % GDP p.c.	0.007	0.002	0.006			0.042***
	(0.011)	(0.013)	(0.011)			(0.009)
Linguistic distance, off. language(s) to German	0.401***					
	(0.098)					
Share of pop. speaking English		0.001				
		(0.001)				
Log market potential			−0.173***			
			(0.031)			
PISA Maths score				−0.002*		
				(0.001)		
PISA Science score					−0.002*	
					(0.001)	
Log productivity p. emp., OECD						0.017
						(0.026)
Constant	7.761***	7.335***	6.795***	7.133***	7.131***	4.731***
	(0.531)	(0.687)	(0.461)	(0.530)	(0.538)	(0.425)
R^2	0.325	0.340	0.326	0.355	0.355	0.398
Observations	8,293	6,209	8,280	3,604	3,604	5,332

Global productivity	1	2	3	4	5	6
Log GDP	−0.152***	−0.154***		−0.161***	−0.159***	−0.058*
	(0.032)	(0.037)		(0.032)	(0.033)	(0.027)
Log flight time	0.235***	0.154**	0.232***	0.162***	0.165***	0.304***
	(0.052)	(0.055)	(0.052)	(0.044)	(0.043)	(0.049)
Linguistic distance			0.409**	0.521***	0.559***	0.328*
			(0.128)	(0.120)	(0.113)	(0.134)
Time difference to Germany	0.003	0.008	0.004	0.018	0.017	0.034*
	(0.009)	(0.013)	(0.009)	(0.011)	(0.010)	(0.013)
GDP per employee	0.084	0.038	0.082	0.179***	0.176***	
	(0.045)	(0.051)	(0.045)	(0.048)	(0.048)	
Public expenditure per pupil, % GDP p.c.	0.013	0.014	0.013			0.058***
	(0.010)	(0.012)	(0.010)			(0.011)
Linguistic distance, off. language(s) to German	0.410**					
	(0.121)					
Share of pop. speaking English		0.004*				
		(0.002)				
Log market potential			−0.165***			
			(0.036)			
PISA Maths score				−0.003*		
				(0.001)		
PISA Science score					−0.003*	
					(0.001)	
Log productivity p. emp., OECD						−0.023
						(0.028)
Constant	4.305***	4.890***	3.251***	8.158***	8.168***	2.156***
	(0.550)	(0.603)	(0.436)	(0.633)	(0.661)	(0.454)
R^2	0.355	0.396	0.356	0.386	0.385	0.447
Observations	10,337	7,493	10,309	4,468	4,468	6,256

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.3 The Heckman selection model

One possibility for addressing the self-selection of firms across destinations is to estimate a Heckman selection model.

The Heckman selection model consists of two equations, a selection equation and the regression equation. The model does not allow using parent–year fixed effects. Instead, we use parent sector dummies and include domestic productivity to take differences across firms into account. This implies that the Heckman selection model is different from the model estimated above. Whereas our regression above strictly compares differences in the performance of one and the same firm across countries without taking the self-selection of the set of countries into account, the Heckman selection model explicitly models the investment decision of firms. At the same time, the Heckman selection model does not recognize which observations belong to the same firm, so the results do not cleanly distinguish between within-firm differences across countries and differences within countries across firms.

The model is specified as follows:

$$
\begin{aligned}
y^*_{1ijt} &= \eta_0 + \eta_1 \tilde{P}_{it} + \eta_2 \tilde{Q}_{jt} + \eta_3 \tilde{\theta}_{jt} + \eta_4 \tilde{c}_{jt} + \eta_5 \tilde{w}_{jt} + \eta_6 \tilde{A}_{jt} + \eta_7 \tilde{f}^I_{jt} + \alpha_s + \alpha_t + u_{1ijt} \\
y_{1ijt} &= \mathbf{1}(y^*_{1ijt} \geq 0) \\
y^*_{2ijt} &= \theta_0 + \theta_1 \tilde{P}_{it} + \theta_2 \tilde{Q}_{jt} + \theta_3 \tilde{\theta}_{jt} + \theta_4 \tilde{c}_{jt} + \theta_5 \tilde{w}_{jt} + \theta_6 \tilde{A}_{jt} + \alpha_s + \alpha_t + u_{2ijt} \\
y_{2ijt} &= y_{1ijt} y^*_{2ijt}
\end{aligned}
$$

where $\tilde{P}_{it}$ denotes domestic productivity, and $\tilde{f}^I_{jt}$ denotes the fixed costs of investment that are used as exclusion restriction.

The first two equations describe the self-selection of firms across countries. y^*_{1ijt} denotes the latent difference of the productivity level of firm i in country j in period t and the firm-specific cut-off productivity level $\bar{y}^*_{1ijt}$. Only if the latent difference is positive, i.e., if firm productivity is above the threshold $\bar{y}^*_{1ijt}$, does the firm decide to invest in a country. In this case, we observe firm productivity y_{2ijt} as described by the second set of equations.

The Heckman selection model rests on strict assumptions on the unobserved error terms u_{1ijt}, u_{2ijt}: they are assumed to be jointly normally distributed, so each is assumed to have a normal density. It is possible to test whether this assumption is approximately fulfilled in our sample by running an OLS regression of the investment decision and the observed foreign productivities on the covariates and analyzing the resulting residuals. Figures A.1 and A.2 present kernel density estimates of the distribution of the residuals.

Figure A.1: Kernel density, residual foreign productivity

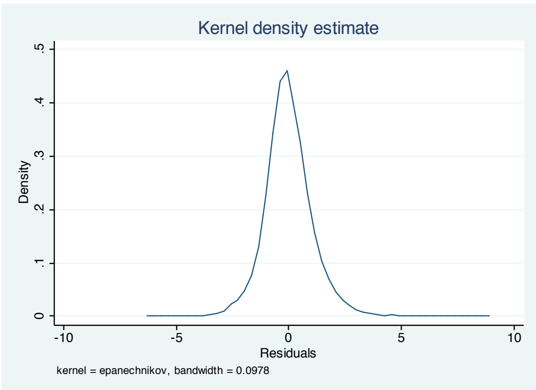
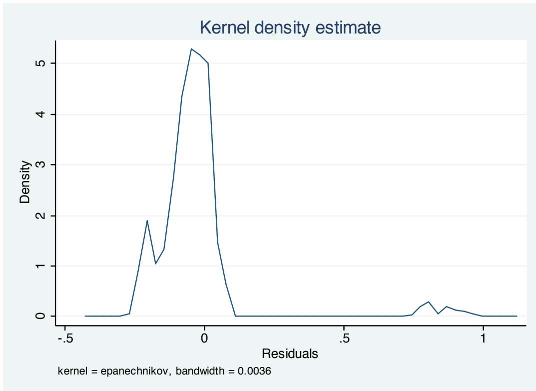


Figure A.2: Kernel density, residual investment equation



Figures depict kernel density estimates of the residuals of an OLS regression of the foreign productivities and the investment decision on the covariates, respectively.

It is evident that the assumption of normally distributed errors is defendable for the distribution of foreign productivities. However, it is difficult to argue that the assumption is fulfilled for the extensive margin: the distribution is not symmetric, it is uneven and has mass far in the right tail. The assumptions of the Heckman selection model are thus not fulfilled.

Correspondingly, it is not surprising that the regression results presented in Table A.5 are not stable. In the selection equation, the costs of starting a business, GDP and the communication costs are consistent with expectations. Paradoxically, a shorter time to start a business and higher labor productivity are predicted to depress the probability of investment. The coefficient pattern in the regression equation is incoherent: the effects of GDP and communication costs are consistent with the theoretical predictions in columns 3 to 5, but the signs flip in columns 1, 2 and 6. In addition, Mill's ratio is significant at the 0.1% level in columns 1-3 and 6, but insignificant in column 5. Overall, the Heckman selection model does not yield meaningful results.

Table A.5: Heckman selection model

	1	2	3	4	5	6
Foreign productivity						
Domestic productivity	0.312***	0.317***	0.342***	0.323***	0.319***	0.306***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)
Log GDP	−0.028**	−0.006	0.261***	0.182***	0.163***	−0.107***
	(0.011)	(0.012)	(0.019)	(0.019)	(0.028)	(0.029)
Linguistic distance, official language(s) to German		−0.526***	−0.819***	−0.692***	−0.436***	0.126**
		(0.037)	(0.033)	(0.033)	(0.041)	(0.044)
Log flighttime			−0.221***	−0.210***	−0.060*	0.171***
			(0.016)	(0.014)	(0.024)	(0.025)
Time difference to Germany				0.018***	0.049***	0.010***
				(0.002)	(0.002)	(0.003)
Total public expenditure on education, share of GDP					0.131***	0.043***
					(0.008)	(0.008)
GDP per employee						0.300***
						(0.006)
Selection equation: investment decision						
Domestic productivity	0.037***	0.037***	0.038***	0.039***	0.042***	0.041***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Log GDP	0.350***	0.349***	0.359***	0.368***	0.346***	0.357***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Linguistic distance, official language(s) to German		−0.760***	−0.382***	−0.390***	−0.381***	−0.410***
		(0.015)	(0.016)	(0.016)	(0.017)	(0.018)
Log flighttime			−0.276***	−0.267***	−0.323***	−0.332***
			(0.003)	(0.004)	(0.005)	(0.005)
Time difference to Germany				0.004***	−0.001	0.000
				(0.001)	(0.001)	(0.001)
Total public expenditure on education, share of GDP					−0.070***	−0.065***
					(0.003)	(0.003)
GDP per employee						−0.025***
						(0.004)
Cost of starting a business	−0.005***	−0.004***	−0.002***	−0.002***	−0.004***	−0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Time to start a business	−0.001***	−0.000**	0.002***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	−6.372***	−5.738***	−4.739***	−4.890***	−3.902***	−3.890***
	(0.042)	(0.044)	(0.047)	(0.054)	(0.064)	(0.065)
Mill's ratio	−0.574***	−0.513***	0.356***	0.151**	−0.152	−0.545***
Number of observations	1,112,575	1,107,922	979,577	933,035	629,022	543,749

Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Dependent variable: foreign productivity - ln (foreign sales/foreign employees), investment decision - 1 if in country. *Covariate definitions:* see Table 1.1.

A.4 Discussion: A monitoring based model

A multinational firm 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: $q_j = n_j a_j A_j$. A_j denotes the maximum amount of output that can be produced with one unit of labor input and full effort $a_j = 1$, and corresponds to the labor productivity in 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. We assume that the managers exert full effort $a_h = 1$ in supervision, consistent with the literature. Production workers know that they are monitored at any point in time with probability P_j.

The workers receive the firm specific 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

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

This implies that $w_j = \frac{1}{P_j} g(a_j^*)$. 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 multinational firm is given by

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

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 cross border 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 productivity revealed in Section 3.5.

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 multinational firms 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. This implication is at odds with the empirical evidence.

In contrast, in the knowledge based model of the organization of multinational firms, firms with higher overall knowledge level $\bar{z}$ become multinationals. They are predicted to pay higher domestic wages. As the following equations demonstrate, the knowledge level of production workers in the home country z_0 is increasing in the overall knowledge level of the firm, independently from whether the constraint $z_j = \bar{z} - z_h$ is binding in the home country ($\phi_0 \neq 0$) or in the foreign country ($\phi_1 \neq 0$).

$$
\begin{aligned}
& \frac{dz_h}{d\bar{z}} = -\frac{-\lambda(1+c_0 z_h)q_1\theta_1 e^{-\lambda(\bar{z}-z_h)}\lambda w_0(c_0+\lambda(1+c_0 z_h))}{d\ (1.12)/dz_h} \quad > 0 \quad \text{for } \phi_1 \neq 0 \\
\Rightarrow \quad & \frac{dz_0}{d\bar{z}} = \frac{c_0}{\lambda(1+c_0 z_h)}\frac{dz_h}{d\bar{z}} \quad > 0 \\
& \frac{dz_h}{d\bar{z}} = -\frac{-\lambda(1+c_0 z_h)q_0\theta_0 e^{-\lambda(\bar{z}-z_h)}\lambda w_0(c_0+\lambda(1+c_0 z_h))}{d\ (1.12)/dz_h} \quad > 0 \quad \text{for } \phi_0 \neq 0 \\
\Rightarrow \quad & \frac{dz_0}{d\bar{z}} = 1 - \frac{dz_h}{d\bar{z}} \quad > 0 \quad \text{by } \frac{dz_h}{d\bar{z}} < 1; \\
& \frac{dz_h}{d\bar{z}} < 1 \quad \text{by} \\
& \frac{dz_h}{d\bar{z}} = \frac{q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0(c_0+\lambda(1+c_0 z_h))\lambda(1+c_0 z_h)}{\lambda q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0((c_0+\lambda(1+c_0 z_h))^2+\lambda c_0(1+c_0 z_h)) - \lambda\frac{q_0 w_0 c_0^2}{A_0}} \\
& \quad = \frac{1}{\lambda}\frac{1}{1+\frac{q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0 c_0(c_0+2\lambda(1+c_0 z_h)) - \frac{q_0 c_0^2 w_0}{A_0}}{q_0\theta_0 e^{-\lambda(\bar{z}-z_h)} w_0(c_0+\lambda(1+c_0 z_h))\lambda(1+c_0 z_h)}} \\
& \quad < 1 \quad \text{if } q_0\theta_0 e^{-\lambda(\bar{z}-z_h)}\lambda(1+c_0 z_h) > c_0 q_1\theta_1 e^{-\lambda z_1}, \\
& \qquad\qquad \text{which is the case by } \frac{d(1.12)}{dz_h} > 0.
\end{aligned}
$$

Appendix B

Appendix to Chapter 2

B.1 Theory

B.1.1 Optimal organizational structure

Optimization problem at production locations. Given the number of layers, the optimal number of employees and the optimal knowledge level per layer of the single establishments are derived by differentiating the Lagrangian equation. The first order conditions are:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_j}{\partial z^0_{L,j}} &= \frac{d_j}{\eta\theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} w_j (1 + c_j z^0_{L,j})\lambda + \frac{d_j}{\eta\theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} w_j c_j \\
&\quad - \xi_j \frac{d_j}{\eta\theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} \lambda A_j (1 - e^{-\lambda \bar{Z}_H}) - \phi_j + \chi_j = 0 \qquad \text{(B.1)} \\
\frac{\partial \mathcal{L}_j}{\partial z^\ell_{L,j}} &= \frac{d_j}{\eta\theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} w_j (1 + c_j z^0_{L,j})\lambda \\
&\quad + \frac{d_j}{\theta_j} e^{\lambda \sum_{k=\ell}^{L-1} z^k_{L,j}} w_j c_j + \sum_{k=1}^{\ell} \frac{d_j}{\theta_j} e^{\lambda \sum_{m=k}^{L-1} z^m_{L,j}} \lambda w_j (1 + c_j z^k_{L,j}) \\
&\quad - \xi_j A_j (1 - e^{-\lambda \bar{Z}_H}) \frac{d_j}{\eta\theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} \lambda - \phi_j + \chi_j = 0 \qquad \text{(B.2)} \\
\frac{\partial \mathcal{L}_j}{\partial d_j} &= \frac{1}{\eta\theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} w_j (1 + c_j z^0_{L,j}) + \sum_{\ell=1}^{L-1} \frac{1}{\theta_j} e^{\lambda \sum_{k=\ell}^{L-1} z^k_{L,j}} w_j (1 + c_j z^\ell_{L,j}) + p_j \\
&\quad - \xi_j A_j (1 - e^{-\lambda \bar{Z}_H}) \frac{1}{\eta\theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z^\ell_{L,j}} = 0 \qquad \text{(B.3)}
\end{aligned}
$$

$$\frac{\partial \mathcal{L}_j}{\partial \xi_j} = q_j - A_j(1-e^{-\lambda \bar{Z}_H})\frac{d_j}{\eta\theta_j}e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} = 0 \tag{B.4}$$

$$\frac{\partial \mathcal{L}_j}{\partial \phi_j} = \underline{Z}_H - \sum_{\ell=0}^{L-1} z_{L,j}^{\ell} = 0 \tag{B.5}$$

$$\frac{\partial \mathcal{L}_j}{\partial \chi_j} = \sum_{\ell=0}^{L-1} z_{L,j}^{\ell} - \bar{Z}_H = 0 \tag{B.6}$$

We use the following notation: $Z_{L,j}^{L-1}$ denotes total knowledge at the establishment, i.e., $Z_{L,j}^{L-1} = \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}$. The constraint $Z_{L,j}^{L-1} = \underline{Z}_H$ is binding for $j = \bar{j}$ and slack for $j = \hat{j}$.

$Z_{L,j}^{L-1} = \bar{Z}_H$ cannot be optimal, because otherwise, headquarter capacities would be used without benefits in terms of output, so $\chi_j = 0 \quad \forall j$. Equation (B.4) determines d_j. $\xi_{\bar{j}}$ follows from Equation (B.3), $\xi_{\hat{j}}$ follows from Equation (B.1) with $\phi_{\hat{j}} = 0$. $z_{L,j}^0$ follows from Equation (B.5) for $j = \bar{j}$ and from Equation (B.3) for $j = \hat{j}$ after inserting $\xi_{\hat{j}}$. $z_{L,j}^1$ results from the difference of Equations (B.2) and (B.1). $z_{L,j}^{\ell}$, $\ell > 1$, $L-1 > 1$ results from the difference of Equation (B.2) for ℓ and for $\ell-1$. $\phi_{\bar{j}}$ results from Equation (B.1) after substituting for $\xi_{\bar{j}}$.

Optimization problem of the firm. The optimal organizational structure of the firm is determined by differentiating the overall cost function with respect to $\underline{Z}_H$, $\bar{Z}_H$ and $\{p_j\}_{j=0}^{1}$.

With κ denoting the multiplier for $\sum_j d_j = 1$, the first order conditions are given by:

$$\frac{\partial \mathcal{L}_{MNE}}{\partial \underline{Z}_H} = -\phi_H + \sum_j \phi_j - \sum_j p_j \frac{\partial d_j}{\partial \underline{Z}_H} + \kappa \sum_j \frac{\partial d_j}{\partial \underline{Z}_H} = 0 \tag{B.7}$$

$$\frac{\partial \mathcal{L}_{MNE}}{\partial \bar{Z}_H} = \phi_H - \sum_j \chi_j - \sum_j \xi_j \lambda e^{-\lambda \bar{Z}_H} A_j \frac{d_j}{\eta\theta_j} e^{\lambda Z_{L,j}^{L-1}} - \sum_j p_j \frac{\partial d_j}{\partial \bar{Z}_H} + \kappa \sum_j \frac{\partial d_j}{\partial \bar{Z}_H} = 0 \tag{B.8}$$

$$\frac{\partial \mathcal{L}_{MNE}}{\partial p_j} = \kappa \frac{\partial d_j}{\partial p_j} - p_j \frac{\partial d_j}{\partial p_j} = 0 \tag{B.9}$$

$$\frac{\partial \mathcal{L}_{MNE}}{\partial \kappa} = \sum_j d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) - 1 = 0 \tag{B.10}$$

$p_j = \kappa$ for $Z_{L,j}^{L-1} > \underline{Z}_H$ follows from Equation (B.9). Note that $\frac{\partial d_j}{\partial p_j} = 0$ for $Z_{L,j}^{L-1} = \underline{Z}_H$, so $p_{\bar{j}}$ is undefined. Equation (B.7) defines κ. $\bar{Z}_H$ follows from substituting for κ in Equation (B.8). $\underline{Z}_H$ is defined by the constraint (B.10).

B.1.2 Comparative statics

Ad Proposition 1. We keep the following notation: the constraint $Z_{L,j}^{L-1} \equiv \sum_{\ell=0}^{L-1} z_{L,j}^{\ell} = \underline{Z}_H$ is binding for $j = \bar{j}$ and slack for $j = \hat{j}$. Define:

$$\gamma_1 = \frac{1}{\eta\theta_{\hat{j}}} w_{\hat{j}} c_{\hat{j}} e^{\lambda Z_{L,\hat{j}}^{L-1}} + \frac{1}{\eta\theta_{\bar{j}}} w_{\bar{j}} c_{\bar{j}} \frac{q_{\hat{j}} A_{\bar{j}}}{q_{\bar{j}} A_{\hat{j}}} e^{\lambda \underline{Z}_H} e^{-2\lambda \bar{Z}_H} \qquad > 0, \tag{B.11}$$

$$\gamma_3 = \frac{1}{\eta\theta_{\bar{j}}} e^{\lambda \underline{Z}_H} (1 - e^{-\lambda \bar{Z}_H}) \left(w_{\bar{j}} c_{\bar{j}} (1 + e^{-\lambda \bar{Z}_H}) - w_0 c_0 \frac{A_{\bar{j}}}{q_{\bar{j}}} \left(1 + \sum_{\ell=1}^{L-1} \frac{1}{\eta} e^{\lambda \sum_{k=0}^{\ell-1} z_{L,\bar{j}}^{k}} \right) \right), \tag{B.12}$$

$$\gamma_4 = -e^{-\lambda \underline{Z}_H} \left(1 + \sum_{\ell=1}^{L-1} \frac{1}{\eta} e^{\lambda \sum_{k=0}^{\ell-1} z_{L,\bar{j}}^{k}} \right) - \frac{1}{\eta\theta_{\bar{j}}} \frac{w_{\bar{j}} c_{\bar{j}}}{w_0 c_0} e^{-2\lambda \bar{Z}_H} \qquad < 0, \tag{B.13}$$

$$\gamma_6 = \frac{q_{\hat{j}} \theta_{\hat{j}} A_{\bar{j}}}{q_{\bar{j}} \theta_{\bar{j}} A_{\hat{j}}} e^{-\lambda Z_{L,\hat{j}}^{L-1}} \left(1 + \sum_{\ell=1}^{L-1} \frac{1}{\eta} e^{\lambda \sum_{k=0}^{\ell-1} z_{L,\hat{j}}^{k}} \right) + \frac{1}{\eta\theta_{\bar{j}}} \frac{q_{\hat{j}} A_{\bar{j}}}{q_{\bar{j}} A_{\hat{j}}} \frac{w_{\hat{j}} c_{\hat{j}}}{w_0 c_0} e^{-2\lambda \bar{Z}_H} \qquad > 0. \tag{B.14}$$

In the following, we assume that $\gamma_3 > 0$. This is the case whenever the marginal costs of increasing $\underline{Z}_H$, $w_{\bar{j}} c_{\bar{j}} \frac{\eta q_{\bar{j}} e^{-\lambda \sum_{\ell=0}^{L-2} z_{L,j}^{\ell}}}{A_{\bar{j}} (1 - e^{-\lambda \bar{Z}_H})}$, is larger than the marginal benefit of increasing $\underline{Z}_H$, $w_0 c_0$. If this assumption does not hold, $\kappa \leq 0$, so the assumption does not seem to be particularly strong.

We differentiate the optimum Equations (2.20), (2.21) and (2.15)—after substituting for p_j—with respect to the different model parameters x_j. That is, for each model parameter x_j, we differentiate each of the equations and take both direct effects of the parameters on the equations and indirect effects via $\frac{d\bar{Z}_H}{dx_j}$, $\frac{dz_{L,\hat{j}}^0}{dx_j}$ and $\frac{dz_{L,\bar{j}}^0}{dx_j}$ into account. Substituting for $\frac{d\bar{Z}_H}{dx_j}$ yields a system of two linear equation in two unknowns, $\frac{dz_{L,\hat{j}}^0}{dx_j}$ and $\frac{dz_{L,\bar{j}}^0}{dx_j}$:

$$\gamma_1 \frac{dz_{L,\hat{j}}^0}{dx_j} = \gamma_2 + \gamma_3 \frac{dz_{L,\bar{j}}^0}{dx_j} \tag{B.15}$$

$$\gamma_4 \frac{dz_{L,\bar{j}}^0}{dx_j} = \gamma_5 + \gamma_6 \frac{dz_{L,\hat{j}}^0}{dx_j} \tag{B.16}$$

This implies:

$$\frac{dz_{L,\hat{j}}^0}{dx_j} = \left(\gamma_1 - \frac{\gamma_3 \gamma_6}{\gamma_4} \right)^{-1} \left(\gamma_2 + \frac{\gamma_3 \gamma_5}{\gamma_4} \right) \tag{B.17}$$

$$\frac{dz_{L,\bar{j}}^0}{dx_j} = \left(\gamma_4 - \frac{\gamma_6 \gamma_3}{\gamma_1} \right)^{-1} \left(\gamma_5 + \frac{\gamma_6 \gamma_2}{\gamma_1} \right) \tag{B.18}$$

As

$$\gamma_1 - \frac{\gamma_3\gamma_6}{\gamma_4} > 0 \quad \text{and} \quad \gamma_4 - \frac{\gamma_6\gamma_3}{\gamma_1} < 0,$$

the signs of the derivatives are determined by γ_2 and γ_5, which vary with x_j.

In the following, we list the expressions for γ_2 and γ_5 and derive the resulting implications for $\frac{dz^0_{L,\hat{j}}}{dx_j}$ and $\frac{dz^0_{L,\bar{j}}}{dx_j}$. In doing so, we distinguish between locations $j = \bar{j}$ with $Z^{L-1}_{L,\bar{j}} = \underline{Z}_H$ and $j = \hat{j}$ with $Z^{L-1}_{L,\hat{j}} > \underline{Z}_H$. As the knowledge acquisition costs c_0 and the wages w_0 additionally affect the costs that arise at the headquarters, we additionally distinguish between $j = 0$ and $j \neq 0$ for the two model parameters.

B.1.2.1 Effect of communication costs

B.1.2.1.1 $\theta_{\hat{j}}$

$$\gamma_2 = \frac{1}{\lambda\theta^2_{\hat{j}}} c_{\hat{j}} w_{\hat{j}} e^{\lambda z^{L-1}_{L,\hat{j}}} > 0 \tag{B.19}$$

$$\gamma_5 = -\frac{1}{\lambda\theta_{\bar{j}}} e^{-\lambda Z^{L-1}_{L,\hat{j}}} \frac{q_{\hat{j}} A_{\bar{j}}}{q_{\bar{j}} A_{\hat{j}}} < 0 \tag{B.20}$$

$\Rightarrow \frac{dz^0_{L,\hat{j}}}{d\theta_{\hat{j}}} > 0$, $\frac{dz^0_{L,\bar{j}}}{d\theta_{\hat{j}}}$ ambiguous.

B.1.2.1.2 $\theta_{\bar{j}}$

$$\gamma_2 = -\frac{1}{\lambda\theta^2_{\bar{j}}} \left(w_{\bar{j}} c_{\bar{j}} e^{\lambda z^{L-1}_{L,\bar{j}}} - w_0 c_0 \frac{A_{\bar{j}}(1 - e^{-\lambda \bar{Z}_H})}{\eta q_{\bar{j}} e^{-\lambda \underline{Z}_H}} \right) < 0 \tag{B.21}$$

$$\gamma_5 = -\frac{1}{\lambda\theta^2_{\bar{j}}} \frac{A_{\bar{j}}}{q_{\bar{j}}} \left(\frac{1}{\eta}(1 - e^{-\lambda \bar{Z}_H}) - \frac{q_{\hat{j}}\theta_{\hat{j}}}{A_{\hat{j}}} e^{-\lambda Z^{L-1}_{L,\hat{j}}} \right) < 0 \tag{B.22}$$

$\Rightarrow \frac{dz^0_{L,\hat{j}}}{d\theta_{\bar{j}}}$ ambiguous, $\frac{dz^0_{L,\bar{j}}}{d\theta_{\bar{j}}} > 0$.

B.1.2.2 Effect of wages

B.1.2.2.1 $w_{\hat{j}}$, $w_{\hat{j}} = w_0$

$$\begin{aligned}\gamma_2 =& -\frac{1}{\lambda\theta_{\hat{j}}} c_0 e^{\lambda z_{L,\hat{j}}^{L-1}} - c_o \frac{A_{\bar{j}}(1-e^{-\lambda\bar{Z}_H})}{\lambda\eta q_{\bar{j}}\theta_{\bar{j}} e^{-\lambda\underline{Z}_H}} \\ &+ \frac{1}{\lambda\theta_{\bar{j}}} e^{\lambda\underline{Z}_H} e^{-2\lambda\bar{Z}_H} \frac{w_{\bar{j}}}{w_0}\left(1 + c_{\bar{j}} z_{L,\bar{j}}^0 + \frac{1}{\lambda} c_{\bar{j}}\right) < 0 \qquad \text{(B.23)}\\ \gamma_5 =& -\frac{\lambda\eta q_{\hat{j}}\theta_{\hat{j}}}{A_{\hat{j}}(1-e^{-\lambda\bar{Z}_H})} e^{-\lambda Z_{L,\hat{j}}^{L-1}}\left(1 + \sum_{\ell=1}^{L-1}\frac{1}{\eta} e^{\lambda\sum_{k=0}^{\ell-1} z_{L,\hat{j}}^k}\right) \\ &- \frac{e^{-2\lambda\bar{Z}_H}}{(1-e^{-\lambda\bar{Z}_H})^2}\left(\frac{\eta q_{\bar{j}}\theta_{\bar{j}}}{A_{\bar{j}}} e^{-\lambda\underline{Z}_H} + \frac{\eta q_{\hat{j}}\theta_{\hat{j}}}{A_{\hat{j}}} e^{-\lambda Z_{L,\hat{j}}^{L-1}}\right)\lambda\frac{q_{\hat{j}}}{A_{\hat{j}}} < 0 \qquad \text{(B.24)}\end{aligned}$$

$\Rightarrow \frac{dz_{L,\hat{j}}^0}{dw_{\hat{j}}}$ ambiguous, $\frac{dz_{L,\bar{j}}^0}{dw_{\hat{j}}} > 0$.

B.1.2.2.2 $w_{\hat{j}}$, $w_{\hat{j}} \neq w_0$

$$\gamma_2 = -\left(\frac{1}{\lambda\theta_{\hat{j}}} c_{\hat{j}} e^{\lambda z_{L,\hat{j}}^{L-1}} + \frac{q_{\hat{j}}A_{\bar{j}}}{q_{\bar{j}}A_{\hat{j}}}\frac{1}{\eta\theta_{\bar{j}}} e^{\lambda\underline{Z}_H} e^{-2\lambda\bar{Z}_H}\left(1 + c_{\hat{j}} z_{L,\hat{j}}^0 + \frac{1}{\lambda}c_{\hat{j}}\right)\right) < 0 \qquad \text{(B.25)}$$

$$\gamma_5 = \frac{1}{\eta\theta_{\bar{j}}}\frac{A_{\bar{j}}q_{\hat{j}}}{A_{\hat{j}}q_{\bar{j}}}\frac{1}{w_0c_0} e^{-2\lambda\bar{Z}_H}\left(1 + c_{\hat{j}} z_{L,\hat{j}}^0 + \frac{1}{\lambda}c_{\hat{j}}\right) > 0 \qquad \text{(B.26)}$$

$\Rightarrow \frac{dz_{L,\hat{j}}^0}{dw_{\hat{j}}} < 0$, $\frac{dz_{L,\bar{j}}^0}{dw_{\hat{j}}}$ ambiguous.

B.1.2.2.3 $w_{\bar{j}}$, $w_{\bar{j}} = w_0$

$$\gamma_2 = \frac{1}{\theta_{\hat{j}}}\frac{w_{\hat{j}}c_{\hat{j}}}{w_0c_0} e^{\lambda z_{L,\hat{j}}^{L-1}} + \frac{A_{\bar{j}}}{q_{\bar{j}}}\frac{\lambda}{\eta\theta_{\bar{j}}} e^{-2\lambda\bar{Z}_H} e^{\lambda\underline{Z}_H}\frac{q_{\hat{j}}w_{\hat{j}}}{A_{\hat{j}}w_0c_0}\left(1 + c_{\hat{j}} z_{L,\hat{j}}^0 + \frac{1}{\lambda}c_{\hat{j}}\right) > 0 \qquad \text{(B.27)}$$

$$\gamma_5 = -\frac{1}{\eta\theta_{\bar{j}}}\frac{A_{\bar{j}}}{q_{\bar{j}}} e^{-2\lambda\bar{Z}_H}\frac{q_{\hat{j}}w_{\hat{j}}(1 + c_{\hat{j}} z_{L,\hat{j}}^0 + \frac{1}{\lambda}c_{\hat{j}})}{A_{\hat{j}}w_0^2c_0} < 0 \qquad \text{(B.28)}$$

$\Rightarrow \frac{dz_{L,\hat{j}}^0}{w_{\bar{j}}} > 0$, $\frac{dz_{L,\bar{j}}^0}{w_{\bar{j}}}$ ambiguous.

B.1.2.2.4 $w_{\bar{j}}$, $w_{\bar{j}} \neq w_0$

$$\gamma_2 = \frac{1}{\lambda\theta_{\bar{j}}} c_{\bar{j}} e^{\lambda z_{L,\bar{j}}^{L-1}} - \frac{1}{\eta\theta_{\bar{j}}} e^{\lambda \underline{Z}_H} e^{-2\lambda\bar{Z}_H} \left(1 + c_{\bar{j}} z_{L,\bar{j}}^0 + \frac{1}{\lambda} c_{\bar{j}}\right) > 0 \tag{B.29}$$

$$\gamma_5 = \frac{1}{\eta\theta_{\bar{j}}} e^{-2\lambda\bar{Z}_H} \frac{1}{w_0 c_0} \left(1 + c_{\bar{j}} z_{L,\bar{j}}^0 + \frac{1}{\lambda} c_{\bar{j}}\right) > 0 \tag{B.30}$$

$\Rightarrow \frac{z_{L,\hat{j}}^0}{w_{\bar{j}}}$ ambiguous, $\frac{dz_{L,\bar{j}}^0}{dw_{\bar{j}}} < 0$.

B.1.2.3 Knowledge acquisition costs

B.1.2.3.1 $c_{\hat{j}}$, $c_{\hat{j}} = c_0$

$$\begin{aligned}
\gamma_2 = & -\frac{1}{\lambda\theta_{\hat{j}}} w_0 e^{\lambda z_{L,\hat{j}}^{L-1}} - \frac{w_0 A_{\bar{j}}(1 - e^{-\lambda\bar{Z}_H})}{\lambda\eta q_{\bar{j}}\theta_{\bar{j}} e^{-\lambda\underline{Z}_H}} \\
& + \frac{w_0 c_0 A_{\bar{j}}}{\lambda\eta q_{\bar{j}}\theta_{\bar{j}}} e^{-2\lambda\bar{Z}_H} e^{\lambda\underline{Z}_H} \left(\frac{q_{\hat{j}}\lambda}{A_{\hat{j}} c_0^2} + \frac{q_{\bar{j}} w_{\bar{j}}}{A_{\bar{j}} w_0 c_0^2} \lambda (1 + c_{\bar{j}} z_{L,\bar{j}}^0 + \frac{1}{\lambda} c_{\bar{j}})\right) < 0 \qquad \text{(B.31)} \\
\gamma_5 = & -\frac{\lambda e^{-2\lambda\bar{Z}_H}}{(1 - e^{-\lambda\bar{Z}_H})^2} \eta \left(\frac{q_{\bar{j}}\theta_{\bar{j}} e^{-\lambda\underline{Z}_H}}{A_{\bar{j}}} + \frac{q_{\hat{j}}\theta_{\hat{j}}}{A_{\hat{j}}} e^{-\lambda Z_{L,\hat{j}}^{L-1}}\right) \frac{q_{\bar{j}} w_{\bar{j}}}{A_{\bar{j}} w_0 c_0^2} \left(1 + c_{\bar{j}} z_{L,\bar{j}}^0 + \frac{1}{\lambda} c_{\bar{j}}\right) \\
& < 0 \qquad \text{(B.32)}
\end{aligned}$$

$\Rightarrow \frac{dz_{L,\hat{j}}^0}{dc_{\hat{j}}}$ ambiguous, $\frac{dz_{L,\bar{j}}^0}{dc_{\hat{j}}} > 0$.

B.1.2.3.2 $c_{\hat{j}}$, $c_{\hat{j}} \neq c_0$

$$\gamma_2 = -\frac{1}{\lambda\theta_{\hat{j}}} w_{\hat{j}} e^{\lambda z_{L,\hat{j}}^{L-1}} - \frac{1}{\eta\theta_{\bar{j}}} \frac{A_{\bar{j}} q_{\hat{j}}}{A_{\hat{j}} q_{\bar{j}}} e^{\lambda\underline{Z}_H} e^{-2\lambda\bar{Z}_H} w_{\hat{j}} \left(z_{L,\hat{j}}^0 + \frac{1}{\lambda}\right) < 0 \tag{B.33}$$

$$\gamma_5 = \frac{1}{\eta\theta_{\bar{j}}} \frac{A_{\bar{j}}}{q_{\bar{j}}} e^{-2\lambda\bar{Z}_H} \frac{q_{\hat{j}} w_{\hat{j}}}{A_{\hat{j}} w_0 c_0} \left(z_{L,\hat{j}}^0 + \frac{1}{\lambda}\right) > 0 \tag{B.34}$$

$\Rightarrow \frac{dz_{L,\hat{j}}^0}{dc_{\hat{j}}} < 0$, $\frac{dz_{L,\bar{j}}^0}{dc_{\hat{j}}}$ ambiguous.

B.1.2.3.3 $c_{\bar{j}}$, $c_{\bar{j}} = c_0$

$$\gamma_2 = \frac{1}{\lambda\theta_{\bar{j}}} w_0 e^{\lambda z_{L,\bar{j}}^{L-1}} - w_0 \frac{A_{\bar{j}}(1 - e^{-\lambda\bar{Z}_H})}{\lambda\eta q_{\bar{j}}\theta_{\bar{j}} e^{-\lambda\underline{Z}_H}}$$

$$+\frac{w_0 c_0 A_{\bar{j}}}{\eta q_{\bar{j}}\theta_{\bar{j}}} e^{\lambda \underline{Z}_H} e^{-2\lambda \bar{Z}_H}\left(\frac{q_{\bar{j}} w_{\bar{j}}}{A_{\bar{j}} w_0 c_0^2}+\frac{q_{\hat{j}} w_{\hat{j}}(1+c_{\hat{j}} z^0_{L,\hat{j}}+\frac{1}{\lambda}c_{\hat{j}})}{A_{\hat{j}} w_0 c_0^2}\right)>0 \tag{B.35}$$

$$\gamma_5=-\frac{q_{\bar{j}} w_{\bar{j}}}{A_{\bar{j}} w_0 c_0^2}-\frac{q_{\hat{j}} w_{\hat{j}}(1+c_{\hat{j}} z^0_{L,\hat{j}}+\frac{1}{\lambda}c_{\hat{j}})}{A_{\hat{j}} w_0 c_0^2}<0 \tag{B.36}$$

$\Rightarrow \frac{dz^0_{L,\hat{j}}}{dc_{\bar{j}}}>0$, $\frac{dz^0_{L,\bar{j}}}{dc_{\bar{j}}}$ ambiguous.

B.1.2.3.4 $c_{\bar{j}}$, $c_{\bar{j}} \neq c_0$

$$\gamma_2=\frac{1}{\lambda\theta_{\bar{j}}} w_{\bar{j}} e^{\lambda z^{L-1}_{L,\bar{j}}}-\frac{1}{\eta\theta_{\bar{j}}} w_{\bar{j}} e^{-2\lambda\bar{Z}_H} e^{\lambda\underline{Z}_H}\left(z^0_{L,\bar{j}}+\frac{1}{\lambda}\right)>0 \tag{B.37}$$

$$\gamma_5=\frac{e^{-2\lambda\bar{Z}_H}}{\eta\theta_{\bar{j}}}\frac{w_{\bar{j}}}{w_0 c_0}(z^0_{L,\bar{j}}+\frac{1}{\lambda})>0 \tag{B.38}$$

$\Rightarrow \frac{dz^0_{L,\hat{j}}}{dc_{\bar{j}}}$ ambiguous, $\frac{dz^0_{L,\bar{j}}}{dc_{\bar{j}}}<0$.

B.1.2.4 Effect of labor productivity

B.1.2.4.1 $A_{\hat{j}}$

$$\gamma_2=\frac{A_{\bar{j}}}{\eta q_{\bar{j}}\theta_{\bar{j}}}\frac{q_{\hat{j}} w_{\hat{j}}(1+c_{\hat{j}} z^0_{L,\hat{j}}+\frac{1}{\lambda}c_{\hat{j}})}{A^2_{\hat{j}}} e^{\lambda\underline{Z}_H} e^{-2\lambda\bar{Z}_H}>0 \tag{B.39}$$

$$\gamma_5=\frac{A_{\bar{j}}}{q_{\bar{j}}\theta_{\bar{j}}}\frac{1}{\eta\lambda A_{\hat{j}}}\left(\frac{\eta q_{\hat{j}}\theta_{\hat{j}}}{A_{\hat{j}}} e^{-\lambda Z^{L-1}_{L,\hat{j}}}-e^{-2\lambda\bar{Z}_H}\frac{\lambda q_{\hat{j}} w_{\hat{j}}}{A_{\hat{j}} w_0 c_0}(1+c_{\hat{j}} z^0_{L,\hat{j}}+\frac{1}{\lambda}c_{\hat{j}})\right)>0 \tag{B.40}$$

$\Rightarrow \frac{dz^0_{L,\hat{j}}}{dA_{\hat{j}}}$ ambiguous, $\frac{dz^0_{L,\bar{j}}}{A_{\hat{j}}}<0$.

B.1.2.4.2 $A_{\bar{j}}$

$$\gamma_2=\frac{w_0 c_0}{\lambda\eta\theta_{\bar{j}} q_{\bar{j}} e^{-\lambda\underline{Z}_H}}\left(-1+e^{-\lambda\bar{Z}_H}\left(1+e^{-\lambda\bar{Z}_H}\frac{\lambda q_{\bar{j}} w_{\bar{j}}}{A_{\bar{j}} w_0 c_0}(1+c_{\bar{j}} z^0_{L,\bar{j}}+\frac{1}{\lambda}c_{\bar{j}})\right)\right)<0 \tag{B.41}$$

$$\text{by}\quad -1+e^{-\lambda\bar{Z}_H}\left(1+e^{-\lambda\bar{Z}_H}\frac{\lambda q_{\bar{j}} w_{\bar{j}}}{A_{\bar{j}} w_0 c_0}(1+c_{\bar{j}} z^0_{L,\bar{j}}+\frac{1}{\lambda}c_{\bar{j}})\right)$$

$$<-1+e^{-2\lambda\bar{Z}_H}\left(e^{\lambda\bar{Z}_H}+e^{\lambda\bar{Z}_H}-1\right)=-(1-e^{-\lambda\bar{Z}_H})^2$$

$$\gamma_5 = \frac{1}{\lambda A_{\bar{j}}} e^{-\lambda \underline{Z}_H} - \frac{1}{\eta \theta_{\bar{j}}} e^{-2\lambda \bar{Z}_H} \frac{w_{\bar{j}}(1 + c_{\bar{j}} z^0_{L,\bar{j}} + \frac{1}{\lambda} c_{\bar{j}})}{A_{\bar{j}} w_0 c_0} > 0 \tag{B.42}$$

$\Rightarrow \frac{dz^0_{L,\hat{j}}}{dA_{\bar{j}}} < 0$, $\frac{dz^0_{L,\bar{j}}}{A_{\bar{j}}}$ ambiguous.

Ad Proposition 2. To determine the effect of the model parameters on the difference in knowledge levels, we derive the derivative of the difference in knowledge levels with respect to $z^0_{L,j}$ and use the insights from Proposition 1.

The difference in the knowledge levels of two adjacent layers is given by

$$\Delta = z^\ell_{L,j} - z^{\ell-1}_{L,j}$$
$$\frac{d\Delta}{dz^{\ell-1}_{L,j}} = e^{\lambda z^{\ell-1}_{L,j}} - 1 \quad > 0$$

By Equation (2.16), the knowledge at higher layers is a function of $z^0_{L,j}$. The difference between the knowledge at higher layer and production knowledge increases in $z^0_{L,j}$ by

$$\Delta = z^\ell_{L,j} - z^0_{L,j}$$
$$\frac{d\Delta}{dz^0_{L,j}} = \frac{1}{\eta} e^{\lambda \sum_{k=0}^{\ell-1} z^k_{L,j}} - 1 \quad > 0$$

The difference in knowledge levels is thus increasing in θ_j and $w_j \neq w_0$ by the arguments in Proposition 1.

The effect of knowledge acquisition costs is ambiguous as

$$\frac{d\Delta}{dc_j} = \left(\frac{1}{\eta} e^{\lambda \sum_{k=0}^{\ell-1} z^k_{L,j}} - 1 \right) \frac{dz^0_{L,j}}{dc_j} + \frac{1}{c_j^2}$$

and $\frac{dz^0_{L,j}}{dc_j} < 0$ for $c_j \neq c_0$.

B.1.3 Endogenous size of the headquarters

If the number of headquarter layers is endogenous, it is necessary to separately analyze the cost minimization problem at the headquarters given $\bar{Z}_H, \underline{Z}_H$. The cost minimization problem is given by:

$$C^H(\underline{Z}_H, \bar{Z}_H, \sum_j \delta_j) = \min_{H \geq 0} C_H(\underline{Z}_H, \bar{Z}_H, \sum_j \delta_j) \tag{B.43}$$

$$C_H(\underline{Z}_H, \bar{Z}_H, \sum_j \delta_j) = \min_{\{n_H^h, z_H^h\}_{h=0}^H \geq 0} \left[\sum_{h=0}^{H} n_H^h w_0 (1 + c_0 z_H^h) \right] \tag{B.44}$$

$$\text{s.t.} \quad n_H^H = 1 \tag{B.45}$$

$$n_H^0 = \sum_j \delta_j \tag{B.46}$$

$$n_H^h = \sum_j \left(\frac{\delta_j}{\theta_j} e^{-\lambda(\underline{Z}_H - Z_{L,j}^{L-1})} e^{-\lambda \sum_{k=0}^{h-1} z_H^k} \right) \tag{B.47}$$

$$\sum_{h=0}^{H} z_H^h = \bar{Z}_H - \underline{Z}_H \tag{B.48}$$

where $H+1$ denotes the number of hierarchical layers at the headquarters, $h = 0, ..., H$ refers to the single layers and δ_j counts the problems that are sent from the establishments at the different locations $j = 0, 1$. The optimal organizational structure at the headquarters is derived for any arbitrary combination of $\{\delta_j\}_{j=0}^1$.

As before, the number of layers $H+1$ is determined via a discrete decision: the optimal number of layers yields the lowest costs (Equation B.43). The firm chooses the optimal number of employees n_H^h and the optimal knowledge levels z_H^h per headquarter layer h given the optimal number of layers in order to minimize the costs that occur at the headquarters (Equation B.44). In doing so, the number of employees at the highest layer is restricted to one (Equation B.45). At the lowest headquarter layer $h = 0$, a sufficient number of employees has to be hired such that all problems $\frac{1}{\eta\theta_j}\delta_j$ sent from the different production locations can be listened to (Equation B.46). The number of employees at the intermediate layers is endogenously determined. In doing so, it is necessary to take knowledge overlaps between establishments and headquarters into account as they decrease the probability that a solution is found in a lower headquarter layer (Equation B.47). If there is no knowledge overlap, the probability that a solution is found in the lowest layer is equal to $e^{-\lambda z_H^0}$. If there is a knowledge overlap, the probability is lower: for sure, the solution is not found in the interval from $Z_{L,\hat{j}}^{L-1} - \underline{Z}_H$. The probability that a solution is found is given by $e^{-\lambda(\underline{Z}_H + z_H^0 - Z_{L,\hat{j}}^{L-1})}$. That is why δ_j takes account of the origin of problems. Finally, knowledge at all headquarter layers is restricted to sum up to headquarter knowledge $\bar{Z}_H - \underline{Z}_H$ (equation B.48). The cost minimization problem is thus more involved than before, but as the delineation of headquarters is unobserved in the data, it does not yield new testable insights.

To endogenize $[\underline{Z}_H, \bar{Z}_H]$, the firm solves

$$\begin{aligned}
\min_{\underline{Z}_H, \bar{Z}_H, \{\delta_j, p_j\}_{j=0}^J} \quad & C_H(\underline{Z}_H, \bar{Z}_H, \sum_j \delta_j) \\
& + \sum_{j=0}^{1} \Big(C_j(d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j), \underline{Z}_H, \bar{Z}_H, p_j) - p_j d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) \Big) && \text{(B.49)} \\
\text{s.t.} \quad & \bar{Z}_H \geq \underline{Z}_H && \text{(B.50)} \\
& \delta_j = d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) \quad \forall j && \text{(B.51)}
\end{aligned}$$

The number of problems sent from the establishment at j, d_j, has to be equal to the number of problems anticipated in the optimization problem of the headquarters, δ_j.

The first order conditions are analogous to the first order conditions in Appendix B.1.1. This yields:

$$z_H^0 = \frac{1}{\lambda \theta_{\bar{j}}} \frac{w_{\bar{j}} c_{\bar{j}}}{w_0 c_0} e^{\lambda z_{L,\bar{j}}^{L-1}} - \frac{1}{\lambda} \frac{\sum_{j=0}^{1} \frac{q_j \theta_j}{A_j} e^{-\lambda Z_{L,j}^{L-1}}}{\frac{q_{\bar{j}} \theta_{\bar{j}}}{A_{\bar{j}}} e^{-\lambda \underline{Z}_H}} - \frac{1}{c_0} \tag{B.52}$$

$$\begin{aligned}
e^{\lambda \bar{Z}_H} + e^{-\lambda \bar{Z}_H} \frac{\sum_{j=0}^{1} 1(Z_{L,j}^{L-1} > \underline{Z}_H) q_j / A_j}{\sum_{j=0}^{1} q_j / A_j} &= 1 + \sum_{j=0}^{1} \frac{\lambda q_j w_j}{A_j w_0 c_0} \left(1 + c_j z_{L,j}^0 + \frac{c_j}{\lambda} \right) \\
&- \frac{\sum_{j=0}^{1} 1(Z_{L,j}^{L-1} > \underline{Z}_H) q_j / A_j}{\sum_j q_j / A_j} \left(\sum_{j=0}^{1} \frac{\eta q_j \theta_j e^{-\lambda Z_{L,j}^{L-1}}}{A_j} - 1 \right)
\end{aligned} \tag{B.53}$$

with $Z_{L,\bar{j}}^{L-1} = \underline{Z}_H$ for $j = \bar{j}$.

With z_H^0 defined in this manner, $\underline{Z}_H$ is given by

$$e^{\lambda \underline{Z}_H} = e^{-\lambda \sum_{h=0}^{H-1} z_H^h} \sum_{j=0}^{1} \frac{\eta q_j}{A_j (1 - e^{-\lambda \bar{Z}_H})}. \tag{B.54}$$

This equation follows from (B.47) and (B.45).

B.1.4 Knowledge gaps

The analysis of the optimal firm organizational structure proceeds in two steps. First, it is necessary to study the optimization problem of an establishment with cumulative local knowlede of less than $\underline{Z}_H$, i.e. with a knowledge gap to headquarter knowledge. Second, we study the optimization problem at the firm level.

Optimization problem at production locations. The optimization problem is modified to take into account that only $A_j(1 - e^{-\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H})$ units of output are produced per unit of labor input in the establishment at location j in case of knowledge gaps. The corresponding Lagrangian equation is given by:

$$\begin{aligned} \mathcal{L}_j = {} & \frac{d_j}{\theta_j \eta} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} w_j(1 + c_j z_{L,j}^0) + \sum_{\ell=1}^{L-1} \frac{d_j}{\theta_j} e^{\lambda \sum_{k=\ell}^{L-1} z_{L,j}^{k}} w_j(1 + c_j z_{L,j}^{\ell}) + p_j d_j \\ & + \underline{\phi}_j \left[\sum_{\ell=0}^{L-1} z_{L,j}^{\ell} - \underline{Z}_H \right] \\ & + \underline{\xi}_j \left[q_j - A_j(1 - e^{-\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H}) \frac{d_j}{\eta \theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} \right] \end{aligned} \tag{B.55}$$

We underline multipliers to highlight that they are different from the multipliers used above. The distinction is relevant for the optimization problem at the firm level, that is analyzed below.

The following first order conditions are affected by this modification:

$$\begin{aligned} \frac{\partial \mathcal{L}_j}{\partial z_{L,j}^0} &= \frac{d_j}{\eta \theta_j} \lambda e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} w_j(1 + c_j z_{L,j}^0) + \frac{d_j}{\eta \theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} w_j c_j \\ &- \underline{\xi}_j \frac{d_j}{\eta \theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} \lambda A_j (1 + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H}) + \underline{\phi}_j = 0 \end{aligned} \tag{B.56}$$

$$\begin{aligned} \frac{\partial \mathcal{L}_j}{\partial z_{L,j}^{\ell}} &= \frac{d_j}{\eta \theta_j} \lambda e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} w_j(1 + c_j z_{L,j}^0) + \frac{d_j}{\theta_j} e^{\lambda \sum_{k=\ell}^{L-1} z_{L,j}^{k}} w_j c_j \\ &+ \sum_{k=1}^{\ell} \frac{d_j}{\theta_j} e^{\lambda \sum_{m=k}^{L-1} z_{L,j}^{m}} \lambda w_j (1 + c_j z_{L,j}^{k}) \\ &- \underline{\xi}_j \frac{d_j}{\eta \theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} \lambda A_j (1 + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H}) + \underline{\phi}_j = 0 \end{aligned} \tag{B.57}$$

$$\frac{\partial \mathcal{L}_j}{\partial d_j} = \frac{1}{\eta \theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} w_j(1 + c_j z_{L,j}^0) + \sum_{\ell=1}^{L-1} \frac{1}{\theta_j} e^{\lambda \sum_{k=\ell}^{L-1} z_{L,j}^{k}} w_j(1 + c_j z_{L,j}^{\ell}) + p_j$$

$$
\begin{aligned}
& - \quad \underline{\xi}_j A_j(1 - e^{-\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H}) \frac{1}{\eta \theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} = 0 && \text{(B.58)} \\
\frac{\partial \mathcal{L}_j}{\partial \underline{\xi}_j} & = \quad q_j - A_j(1 - e^{-\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H}) \frac{d_j}{\eta \theta_j} e^{\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} = 0 && \text{(B.59)}
\end{aligned}
$$

Correspondingly, the optimal values adjust. The amount of headquarter services employed in local production at an establishment with a knowledge gap is given by

$$
d_j = \frac{\eta \theta_j q_j e^{-\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}}}{A_j(1 - e^{-\lambda \sum_{\ell=0}^{L-1} z_{L,j}^{\ell}} + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H})} \tag{B.60}
$$

Assuming that the knowledge constraint $\sum_{\ell=0}^{L-1} z_{L,j}^{\ell} = \underline{Z}_H$ is not binding, as required for a knowledge gap, the optimal knowledge level is given by:

$$
p_j = \frac{1}{\theta_j \lambda} e^{\lambda z_{L,j}^{L-1}} w_j c_j + \frac{w_j(1 + c_j z_{L,j}^0 + \frac{1}{\lambda} c_j)}{\eta \theta_j A_j(1 + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H})} \tag{B.61}
$$

Finally, the marginal costs of production are determined by

$$
\underline{\xi}_j = \frac{w_j(1 + c_j z_{L,j}^0 + \frac{1}{\lambda} c_j)}{A_j(1 + e^{-\lambda \underline{Z}_H} - e^{-\lambda \bar{Z}_H})} \tag{B.62}
$$

The span of control and the relationship of knowledge levels within the establishment are given by the same expressions in the main text (cf. Equations 2.14 and 2.16).

Optimization problem of the firm. As above, the firm solves the following problem to endogenize $[\underline{Z}_H, \bar{Z}_H]$:

$$
\begin{aligned}
\min_{\underline{Z}_H, \bar{Z}_H, \{p_j\}_{j=0}^{J}} \quad & w_0(1 + c_0(\bar{Z}_H - \underline{Z}_H)) \\
& + \sum_{j=0}^{1} \Big(C_j(d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j), \underline{Z}_H, \bar{Z}_H, p_j) - p_j d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) \Big) && \text{(B.63)} \\
\text{s.t.} \quad & \bar{Z}_H \geq \underline{Z}_H && \text{(B.64)} \\
& 1 = \sum_{j=0}^{1} d_j(q_j, \underline{Z}_H, \bar{Z}_H, p_j) && \text{(B.65)}
\end{aligned}
$$

Due to the different cost structure at the establishments, the optimum values of $\underline{Z}_H$ and $\bar{Z}_H$ adjust. The optimal upper bound of headquarter knowledge $\bar{Z}_H$ is given by equating the marginal costs of an additional unit of knowledge with the marginal benefit, i.e. additional output.

$$w_0 c_0 = \lambda e^{-\lambda \bar{Z}_H} \sum_{j=0}^{1} \xi_j \frac{d_j}{\eta \theta_j} e^{-\lambda \sum_{\ell=0}^{L-1} z^{\ell}_{L,j}} \tag{B.66}$$

The optimal lower bound of headquarter knowledge $\underline{Z}_H$ is determined by equating the marginal benefit of increasing $\underline{Z}_H$, i.e. acquiring *less* headquarter knowledge, with the marginal cost in terms of less output that is produced at the establishment with the knowledge gap. Note that output at establishments with knowledge overlaps is unaffected.

$$-w_0 c_0 = -\lambda e^{-\lambda \underline{Z}_H} \sum_{j=0}^{1} \mathbf{1} \left(\sum_{\ell=0}^{L-1} z^{\ell}_{L,j} < \underline{Z}_H \right) \underline{\xi}_j \frac{d_j}{\eta \theta_j} e^{-\lambda \sum_{\ell=0}^{L-1} z^{\ell}_{L,j}} \tag{B.67}$$

where the indicator function $\mathbf{1}(\cdot)$ determines whether there is a knowledge gap. We keep the negative signs to highlight that $w_0 c_0$ quantify the marginal benefit in this equation, not the marginal costs as in the equation for $\bar{Z}_H$.

It is not possible to simplify these expressions in order to obtain more tractable terms. We therefore impose the assumption that no knowledge gaps are possible.

B.2 Empirics

B.2.1 Imputation

Table B.1: Fraction of censored observations, by layer

Layer	0	1	2	3
# observations	5,582,625	776,496	793,869	191,725
Share censored	0.044	0.231	0.473	0.612

12.5% of observations in the sample are censored because the daily wage is above the limit that is subject to social security contributions. The fraction of censored observations varies across layers and tends to be higher in higher layers, as Table B.1 shows.

We follow the procedures suggested by Dustmann, Ludsteck, and Schönberg (2009), Card, Heining, and Kline (2013) and Gartner (2005) and impute the censored observations using a series of Tobit regressions for mutually exclusive sub-groups of the data.

It is typically recommended to use multiple imputation approaces to take the additional uncertainty introduced through the imputation process into account (e.g. Schafer, 1999). The large number of observations in our data set does not allow to use multiple imputation, however.

We define five mutually exclusive education groups (missing; primary, lower secondary or intermediate schooling certificate but no vocational training; primary, lower secondary or intermediate schooling certificate and vocational training; upper secondary school certificate with or without vocational training; university training) and six mutually exclusive age groups (below 20, 20-29, 30-39, 40-49, 50-59, above 59) (as Card, Heining, and Kline, 2013). For each of the 30 age $\times$ education groups, we fit a tobit model for log observed remuneration. We use a gender dummy, a dummy for not being German, age, establishment size as measured by full time employees, district dummies, sector dummies and layer dummies as well as log distance, the similarity of dialects and the similarity of religion as covariates. We thus include all covariates that we use in the wage regression as covariates in the imputation procedure, as recommended in the literature (Schafer, 1999). As we use firm fixed effects in our regression and run regressions separately by layer, we additionally include the share of university graduates per establishment, the average coworkers' wage per establishment and the share of censored observations in the establishment as well as the average coworkers' wage in the same layer and firm and the share of censored observations in the same layer and firm to account for dependence within establishments and layers of the same firm. We include a dummy that is one for establishments with a single worker (similar to Card, Heining, and Kline, 2013, who use establishment fixed effects).

We impute remuneration based on the estimated model parameters and a random draw from the associated left censored distribution. That is, if log remuneration $y \sim N(X'\beta, \sigma)$

and $u \sim U(0,1)$

$$y_{imputed} = X'\beta + \sigma\Phi^{-1}[k + u(1-k)] \quad \text{with } k = \Phi\left[\frac{c - X'\beta}{\sigma}\right].$$

B.2.2 Classification of occupations into layers

Table B.2: List of dropped occupations

Occupation code	Description
11	Farmers (*Landwirte*)
12	Winemakers (*Weinbauern*)
42	Milkers (*Melker*)
555	Disabled people (*Behinderte*)
666	Participants of programs for professional reintegration after longer periods of illness (*Rehabilitanden*)
888	Personal care assistants (*Pflegepersonen*)
892	Members of the clergy (*Angehörige geistlicher Berufe*
924	Graduates from home economics school (*Haushaltsschulabschluss*)
971	Family members working in the firm, outside of agriculture (*Mithelfende Familienangehörige außerhalb d. Landwirtschaft*)
981	Apprentices, profession not yet defined (*Auszubildende mit noch nicht feststehendem Ausbildungsberuf*)
982	Trainees, profession not yet defined (*Praktikanten, Volontäre mit noch nicht feststehendem Beruf*)
983	Unemployed, no further details (*Arbeitskräfte (arbeitsuchend) m. nicht bestimmtem Beruf*)
991	Workers, no further details (*Arbeitskräfte ohne nähere Tätigkeitsangabe*)
995	Employees in early retirement programs (*Vorruhestands- oder Altersgeldbezieher*)
997	Recipients of compensation allowances (*Ausgleichsgeldbezieher*)
.z	No information (*k. Angabe*)

To match occupations that are not matched in our baseline procedure detailed in section 2.3.2, we use the following criteria. We assign occupations to the core area whenever the mode is missing because two layers are equally frequent ("core area, mode missing") or mode and core area are contradictory ("core area"). In cases where the core area is missing, we assign occupations to the most frequent layer ("only layer assigned, core area missing"). The remaining observations are matched based on plausibility considerations.

Table B.3: List of occupations that are manually assigned to layers

Layer	Occupation	Description	Criterion
0	21	Animal breeders (*Tierzüchter*)	Only layer assigned, core area missing
0	22	Fishermen (*Fischer*)	Plausibility
0	41	Farm workers (*Landarbeitskräfte*)	Only layer assigned, core area missing
0	61	Foresters (*Forstverwalter, Förster, Jäger*)	Plausibility
0	141	Chemical workers (*Chemiebetriebswerker*)	Core area
0	163	Bookbinders (*Buchbinderberufe*)	Core area, mode missing
0	173	Pressmen (*Buchdrucker (Hochdruck)*)	Core area, mode missing
0	286	Watchmakers (*Uhrmacher*)	Core area, mode missing
0	302	Goldsmiths (*Edelmetallschmiede*)	Core area, mode missing
0	305	Instrument makers (*Musikinstrumentenbauer*)	Core area, mode missing
0	412	Food preservation workers (*Fertiggerichte-, Obst-, Gemüsekonservierer, -zubereiter*)	Only layer assigned, core area missing
0	466	Civil engineering (*Sonstige Tiefbauer*)	Core area, mode missing
0	531	Unskilled worker, no further details (*Hilfsarbeiter ohne nähere Tätigkeitsangabe*)	Plausibility
0	541	Machine operators in energy industry (*Energiemaschinisten*)	Core area, mode missing
0	685	Pharmaceutical assistants (*Apothekenhelferinnen*)	Core area, mode missing
0	691	Banking professionals (*Bankfachleute*)	Core area, mode missing
0	692	Building society professionals (*Bausparkassenfachleute*)	Core area
0	773	Cashiers (*Kassierer*)	Core area
0	781	Clerks (*Bürofachkräfte*)	Core area
0	791	Security guards (*Werkschutzleute, Detektive*)	Core area, mode missing
0	793	Doormen (*Pförtner, Hauswarte*)	Core area, mode missing
0	801	Soldiers, border guards, police officers (*Soldaten, Grenzschutz-, Polizeibedienstete*)	Core area, mode missing
0	802	Professional firefighters (*Berufsfeuerwehrleute*)	Core area, mode missing
0	804	Chimney sweepers (*Schornsteinfeger*)	Core area, mode missing
0	805	Health workers (*Gesundheitssichernde Berufe*)	Core area
0	814	Bailiffs (*Rechtsvollstrecker*)	Core area
0	823	Librarians, employees in archives and museums (*Bibliothekare, Archivare, Museumsfachleute*)	Core area
0	834	Decorators (*Dekorationen-, Schildermaler*)	Core area, mode missing
0	838	Artists, professional athletes (*Artisten, Berufssportler, künstlerische Hilfsberufe*)	Plausibility

Table B.3: List of occupations that are manually assigned to layers

Layer	Occupation	Description	Criterion
0	902	Personal grooming service providers (*Sonstige Körperpfleger*)	Core area
0	911	Barkeepers, hoteliers (*Gastwirte, Hoteliers, Gaststättenkaufleute*)	Plausibility
1	491	Interior decorators (*Raumausstatter*)	Core area, mode missing
1	687	Traveling salesmen (*Handelsvertreter, Reisende*)	Core area, mode missing
1	855	Dietitians, pharmacy technicians (*Diätassistenten, Pharmazeutisch-technische Assistenten*)	Core area
2	32	Agricultural engineers, agricultural advisers (*Agraringenieure, Landwirtschaftsberater*)	Core area
2	52	Landscape architects (*Gartenarchitekten*)	Core area
2	861	Social workers (*Sozialarbeiter, Sozialpfleger*)	Core area
2	862	Children's home professionals (*Heimleiter, Sozialpädagogen*)	Core area
2	864	Kindergarten workers (*Kindergärtnerinnen, Kinderpflegerinnen*)	Core area
2	872	High school teachers (*Gymnasiallehrer*)	Core area, mode missing
2	922	Consumer counselors (*Verbraucherberater*)	Core area, mode missing
3	751	Enterpreneurs, managing directors, heads of business area (*Unternehmer, Geschäftsführer*)	Plausibility
3	762	Senior directive administrative officer	Plausibility

B.2.3 Plausibility checks on delineation of layers

This section compares summary statistics provided in Caliendo, Monte, and Rossi-Hansberg (2012) and summary statistics from our data to evaluate the quality of the classification of occupations into layers.

Table B.4: Distribution of remuneration by layer

Distribution of daily remuneration in levels, IAB data

Layer	N	Mean	SD	p5	p25	p50	p75	p95
0	5,585,080	92.584	36.805	39.615	65.900	88.356	113.881	169.521
1	776,933	130.474	41.066	58.304	99.136	134.277	170.479	179.605
2	794,252	149.989	37.284	72.758	126.420	167.981	179.605	179.605
3	191,902	153.374	39.090	67.929	134.860	178.918	179.605	179.917
Total	7,348,167	104.383	43.166	41.748	71.530	97.5234	134.839	170.605

Distribution of average hourly remuneration in levels in 2005 Euros, Caliendo, Monte, and Rossi-Hansberg (2012)

Layer	Mean	p5	p25	p50	p75	p95
Blue collars	20.83	9.67	12.88	15.24	18.34	26.94
Clerks	19.06	9.74	13.03	15.63	19.01	28.33
Supervisors	26.30	13.14	18.14	21.87	26.44	38.94
Senior staff	47.91	19.56	28.93	35.96	44.62	69.01
CEO, directors	75.60	22.33	38.86	54.62	75.07	132.17

Distribution of observed remuneration by layer following Table 1 in Caliendo, Monte, and Rossi-Hansberg (2012).

Blue collar workers and clerks are grouped to layer 0 in Caliendo, Monte, and Rossi-Hansberg (2012), supervisors are assigned to layer 1, senior staff to layer 2 and CEOs/ directors to layer 3. As in Caliendo, Monte, and Rossi-Hansberg (2012), all moments of the wage distribution in our data increase in the hierarchical level. The maximum is an exception. The remuneration is usually only reported up to the highest level that is subject to social security contributions; this value may be exceeded due to one-time or bonus payments (vom Berge, König, and Seth, 2013, p. 41).

The difference in the level of remuneration is startling at first glance. Supervisors in France earn 26.30 Euro per hour, so 197.25 Euro on a working day with 7.5 hours, whereas employees in layer 1 in Germany earn 130.47 Euro per day on average. Note however that the German data compute daily remuneration based on the number of *calendar* days, not working days. The yearly earnings of workers in layer 1 in Germany are thus about 47,621.55 Euros and the yearly remuneration of French supervisors with 35 working hours per week are around 47,866.00 Euros.

The number of layers per firm or establishment is similar, as Table B.5 shows.

Table B.5: Number of layers per establishment

# layers	1	2	3	4	All	Mean # of layers
# establishments	58,846	22,595	11,903	8,618	102,146	1.709
# firms, CMR (2012)					72,918	1.50

Distribution of number of layers of establishments of multi-establishment firms following Table 2 in Caliendo, Monte, and Rossi-Hansberg (2012). "CMR (2012)" is an abbreviation for Caliendo, Monte, and Rossi-Hansberg (2012).

B.2.4 Additional tables

Table B.6: Bivariate correlations, log individual imputed remuneration and communication costs

Observed remuneration					Residual remuneration				
A. All establishments									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	0.067***	0.123***	0.159***	0.106***	Log distance	0.071***	0.136***	0.161***	0.107***
Dialect	0.018***	0.062***	0.084***	0.016***	Dialect	0.033***	0.083***	0.099***	0.030***
B. Establishments of multi-establishment firms									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	-0.151***	-0.073***	-0.066***	-0.207***	Log distance	-0.040***	0.027***	0.012***	-0.131***
Dialect	-0.103***	-0.044***	-0.049***	-0.179***	Dialect	-0.045***	0.014***	0.009***	-0.127***
C. Establishments of multi-establishment firms without headquarter function									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	-0.134***	-0.092***	-0.059***	-0.203***	Log distance	-0.030***	0.009***	0.019***	-0.125***
Dialect	-0.095***	-0.055***	-0.050***	-0.196***	Dialect	-0.039***	0.005*	0.005*	-0.143***
D. Manufacturing establishments of multi-establishment firms without headquarter function									
Layer	0	1	2	3	Layer	0	1	2	3
Log distance	-0.142***	-0.099***	-0.028***	-0.160***	Log distance	0.042***	0.026***	0.098***	-0.030***
Dialect	-0.100***	-0.059***	-0.016***	-0.169***	Dialect	0.029***	0.021***	0.080***	-0.067***

The table displays bivariate correlations between log imputed remuneration at the employee level per layer and communication costs, measured using log distance and the dissimilarity of dialects. We calculate log distance as the log of observed distance plus 1 to be able to include single-establishment firms in Panel A. The left part of the table uses log imputed remuneration in levels. The right part of the table uses the residual from a linear regression of log imputed remuneration on district dummies. Panel A includes all establishments. Panel B includes only establishments of multi-establishment firms. Panel C drops establishments with headquarter function. The last panel only uses establishments in the manufacturing sector. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.7: Regression results for the span of control, manufacturing firms

	2 Layers		3 Layers		All layers	
	1	1	2	2	3	3
Log distance to HQ	−1.538	−1.807	0.381	0.210	−2.961	−2.951
	(1.217)	(1.090)	(0.295)	(0.249)	(2.090)	(2.303)
Hourly wages	−0.341	−0.225	−0.188	−0.102	−1.263	−0.820
	(0.400)	(0.340)	(0.207)	(0.164)	(1.236)	(1.087)
Hourly labor productivity	0.141	0.120	0.211^{+}	0.123	0.622	0.616
	(0.166)	(0.172)	(0.124)	(0.080)	(0.587)	(0.589)
Log average distance to university	-5.925^{+}	−2.966	2.501	3.298	13.819	9.248
	(3.481)	(2.660)	(2.658)	(2.262)	(8.923)	(11.537)
Constant	55.567^{*}	32.913^{*}	−27.481	−19.783	−64.330	−38.587
	(22.114)	(15.251)	(22.261)	(15.276)	(46.349)	(60.731)
R-squared	0.363	0.022	0.102	0.024	0.137	0.035
Firm fixed effects	N	Y	N	Y	N	Y
Sector dummies	Y	N	Y	N	Y	N
# firms	88	88	120	120	149	149
# observations	304	304	396	396	516	516
Dissimilarity of dialects	−0.122	-0.122^{+}	0.036	0.010	−0.210	−0.189
	(0.076)	(0.064)	(0.027)	(0.021)	(0.142)	(0.154)
Hourly wages	−0.271	−0.152	−0.192	−0.104	−1.212	−0.750
	(0.397)	(0.325)	(0.206)	(0.166)	(1.217)	(1.079)
Hourly labor productivity	0.135	0.122	0.211+	0.122	0.639	0.622
	(0.163)	(0.167)	(0.123)	(0.080)	(0.606)	(0.612)
Log average distance to university	-8.386^{*}	-4.993^{+}	3.132	3.568	10.075	6.215
	(4.021)	(2.651)	(2.988)	(2.578)	(7.249)	(9.849)
Constant	55.333^{*}	29.039^{+}	−27.781	−19.782	−67.057	−45.103
	(21.618)	(15.548)	(22.396)	(15.437)	(48.385)	(65.329)
R-squared	0.367	0.024	0.104	0.024	0.135	0.031
Firm fixed effects	N	Y	N	Y	N	Y
Sector dummies	Y	N	Y	N	Y	N
# firms	88	88	120	120	149	149
# observations	304	304	396	396	516	516

Dependent variable: span of control, defined as $s_j = n^{\ell-1}_{L,j}/n^{\ell}_{L,j}$. Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix C

Appendix to Chapter 3

C.1 List of countries classified as tax havens

The following countries are classified as tax havens according to Hines and Rice (1994, p. 178):

Andorra, Anguilla, Antigua and Barbuda, Bahamas, Bahrain, Barbados, Belize, Bermuda, British Virgin Islands, Cayman Islands, Channel Islands, Cook Island, Cyprus, Dominica, Gibraltar, Grenada, Hong Kong, Ireland, Isle of Man, Jordan, Lebanon, Liberia, Liechtenstein, Luxembourg, Macao, Maldives, Malta, Marshall Islands, Monaco, Montserrat, Netherlands Antilles, Panama, Saint Kitts and Nevis, Saint Lucia, Saint Martin, Saint Vincent and the Grenadines, Singapore, Switzerland, Turks and Caicos Islands, Vanuatu.

We are unable to distinguish investment in Monaco and Saint Martin from investment in France. Therefore, these tax havens are neglected in our analysis.

The OECD's list of tax havens contains the following countries (OECD, 2000): Andorra, Anguilla, Antigua and Barbuda, Aruba, Bahamas, Bahrain, Barbados, Belize, British Virgin Islands, Cook Islands, Dominica, Gibraltar, Grenada, Guernsey/ Sark/ Alderney, Isle of Man, Jersey, Liberia, Liechtenstein, Maldives, Marshall Islands, Monaco, Montserrat, Nauru, Netherlands Antilles, Niue, Panama, Samoa, Seychelles, St Lucia, St. Christopher and Nevis, St. Vincent and the Grenadines, Tonga, Turks & Caicos, US Virgin Islands, Vanuatu.

C.2 German taxation of foreign income

Foreign income, whether in the form of dividends from foreign subsidiaries or income earned in foreign branches, is virtually tax exempt in Germany.

Dividends distributed by national or foreign affiliates as well as capital gains are tax free. Only 5% of dividends and capital gains are taxed as non-deductible operating expenditures (§8b KStG, the German corporate income tax code). This is an important difference with the U.S. tax system, since a U.S. tax is due when the parent company receives dividends from foreign affiliates, and the parent company is entitled to claim an indirect tax credit for income taxes paid by foreign affiliates.

Income earned in foreign branches is not taxed if Germany has a double taxation treaty with the host country, as Germany generally stipulates tax exemption of foreign income in double taxation treaties (Höhn and Höring, 2010, p. 116). Due to the tight network of double taxation treaties, income earned in foreign branches is de facto exempt from taxation in Germany.

An exception to these general rules is the anti-avoidance provision of German tax law (part of the German "Außensteuergesetz (AStG)" commonly referred to as "CFC-legislation"). The anti-deferral rules apply if a German parent controls an affiliate or branch abroad which earns income from passive investment that accounts for more than 10% of total income and is taxed at a rate of less than 25%. In this case, the passive income of the branch or affiliate is apportioned to the parent and subject to German tax independent of repatriation (§§7-9 AStG).

Passive income is defined in a negative way as income which is not active, that is, income which is not generated through agriculture, production, trade, services, dividends, disposal of shares, and, subject to further requirements, banking, insurance, renting or leasing. Income from borrowing or lending is classified as active income if capital is raised in foreign capital markets only and from unrelated parties and lent to active foreign businesses or permanent establishments (§8 AStG). Until recent changes for the years from 2011 onwards (draft Jahressteuergesetz 2010, i.e. tax law for the year 2010), these rules did not apply if the nominal tax rate was higher than 25%, even if the effective tax burden was much lower, as for example in Malta or Panama.

Since the judgment of the European Court of Justice in the Case of Cadbury Schweppes, the provisions explicitly provide for the opportunity to demonstrate substantive activities if the affiliate is located in the EU or EEA, which include Ireland, Luxembourg and Liechtenstein on the list of tax havens. The rules do not apply if the affiliate can be demonstrated to participate in the host country markets, to employ qualified personnel and generate its own income (Bundesministerium der Finanzen, 2007).

A number of recent papers analyze the effect of the anti-avoidance regulation and yield a nuanced picture on the effect of these provisions on profit shifting by multinational enterprises. On the one hand, Ruf and Weichenrieder (2012) report that the CFC rules significantly reduced passive investment in low-tax jurisdictions. They take a multinational's location decisions as given and define passive income as the total financial assets of an affiliate minus equity holdings in and lending to affiliated enterprises to avoid double counting. Using the same set up and a regression discontinuity approach, Egger and Wamser (2010) find that the CFC rules are also associated with less investment in fixed assets around the threshold from non-applicability to applicability. On the other hand, Overesch and Wamser (2010) provide evidence that the German CFC rules do not affect internal lending of foreign affiliates in low-tax locations to other foreign subsidiaries. They find that internal debt shares react positively to tax rate differentials between different locations and that CFC rules do not influence this relationship. Whether and how these provisions affect profit shifting through other strategies such as transfer pricing has not yet been explored.

Overall, this research suggests that the CFC provisions do not foreclose tax planning by MNEs per se, but they render it cumbersome. MNEs can still strategically relocate activities to low tax countries and tax havens, but they have to generate income from active local investment and may not benefit from simply setting up a "letterbox company". The significance of this requirement is that using a tax haven may entail considerable fixed costs, as MNEs must generate active income to benefit fully from their tax haven investments.

C.3 Mathematical Appendix

Proof of Proposition 1

Recall the constrained maximization problem described in equations (3.2) and (3.3). We assume first that this constraint is fulfilled and subsequently reconsider what happens if this is not the case. The first order condition for ψ_i is thus

$$1 - (1 - \tau_i) - (1 - \tau_i)\frac{a\psi_i}{\rho_i} = 0 \tag{C.1}$$

which implies

$$\psi_i^{*th} = \frac{\tau_i \rho_i}{a(1 - \tau_i)} \tag{C.2}$$

Inserting ψ_i^{*th} into condition (3.3) produces a condition for a and τ_i that must be fulfilled for (3.3) to hold.

$$\rho_i \quad - \quad \frac{\tau_i \rho_i}{a(1 - \tau_i)} - \frac{{\tau_i}^2 \rho_i}{2a(1 - \tau_i)^2} \geq 0 \tag{C.3}$$

$$\Leftrightarrow \qquad \tau_i \leq 1 - \sqrt{\frac{1}{2a + 1}} \tag{C.4}$$

Consider now values of τ_i and a such that the constraint (3.3) is not fulfilled for ψ_i^{*th} as determined by the first order condition. In this case, ψ_i is chosen such that condition (3.3) is satisfied with equality, which yields

$$\bar{\psi}_i = \frac{\rho_i}{a}\left(\sqrt{2a + 1} - 1\right) \tag{C.5}$$

There are no more positive profits reported by the affiliate in the non-haven country and hence $\bar{\psi}_i$ equals the multinational's profit from investing in country i, reallocated to the tax haven, after incurring fixed cost c_i.

Q.E.D.

Proof of Proposition 2

Recall the condition (3.6) that determines $\tilde{n}$, such that the multinational chooses $d_i = 1$ for all countries $i = 1, ..., \tilde{n}$ if he holds a tax haven affiliate. Using ψ_i^{*th} as determined by the first order condition for ψ_i, this condition simplifies to

$$(1-\tau_{\tilde{n}})\rho_{\tilde{n}} + \frac{\tau_{\tilde{n}}^2 \rho_{\tilde{n}}}{2a(1-\tau_{\tilde{n}})} - c_{\tilde{n}} \geq 0 > (1-\tau_{\tilde{n}+1})\rho_{\tilde{n}+1} + \frac{\tau_{\tilde{n}+1}^2 \rho_{\tilde{n}+1}}{2a(1-\tau_{\tilde{n}+1})} - c_{\tilde{n}+1} \; . \tag{C.6}$$

Consider now the multinational's situation if it has no tax haven affiliate. In this case, profit-shifting is directed to the country charging the lowest tax rate among those in which the multinational holds an affiliate, denoted by $\underline{\tau}$.Then the profit maximization problem is the following

$$\max_{d_i,\psi_i} \sum_{i=1}^{n} d_i \left[(1-\underline{\tau})\psi_i + (1-\tau_i)(\rho_i - \psi_i - \frac{a}{2}\frac{\psi_i^2}{\rho_i}) - c_i \right] \tag{C.7}$$

with $d_i \in \{0, 1\}$, subject to the same constraint (3.3) as above. The first order condition yields

$$\psi_i^{*nth} = \frac{(\tau_i - \underline{\tau})\rho_i}{a(1-\tau_i)} \; . \tag{C.8}$$

Note that for the parameter condition on τ_i and a assumed above, this optimal ψ_i^{*nth} also satisfies constraint (3.3). The first order condition for d_i yields that the multinational chooses $d_i = 1$ for all countries $i = 1, ..., \hat{n}$ and $d_i = 0$ otherwise, where $\hat{n}$ is determined by the condition that

$$\begin{aligned}(1-\underline{\tau})\psi_{\hat{n}} + (1-\tau_{\hat{n}})(\rho_{\hat{n}} - \psi_{\hat{n}} - \frac{a}{2}\frac{\psi_{\hat{n}}^2}{\rho_{\hat{n}}}) - c_{\hat{n}} \geq 0 \\ > (1-\underline{\tau})\psi_{\hat{n}+1} + (1-\tau_{\hat{n}+1})(\rho_{\hat{n}+1} - \psi_{\hat{n}+1} - \frac{a}{2}\frac{\psi_{\hat{n}+1}^2}{\rho_{\hat{n}+1}}) - c_{\hat{n}+1} \; .\end{aligned} \tag{C.9}$$

Using ψ_i^{*nth} as determined by the first order condition for ψ_i, this condition simplifies to

$$(1-\tau_{\hat{n}})\rho_{\hat{n}} + \frac{(\tau_{\hat{n}} - \underline{\tau})^2 \rho_{\hat{n}}}{2a(1-\tau_{\hat{n}})} - c_{\hat{n}} \geq 0 > (1-\tau_{\hat{n}+1})\rho_{\hat{n}+1} + \frac{(\tau_{\hat{n}+1} - \underline{\tau})^2 \rho_{\hat{n}+1}}{2a(1-\tau_{\hat{n}+1})} - c_{\hat{n}+1} \; . \tag{C.10}$$

A comparison with $(C.6)$ shows that $\tilde{n} \geq \hat{n}$, since the profits realized from each country are potentially larger if it is possible to reduce taxes by reallocating income to a tax haven.

For the multinational, investing in a tax haven is worth the set up cost c_0 if and only if

$$\begin{aligned} &\sum_{i=1}^{\tilde{n}} \left[\psi_i + (1-\tau_i)(\rho_i - \psi_i - \frac{a}{2}\frac{{\psi_i}^2}{\rho_i}) - c_i \right] - c_0 \\ \geq &\sum_{i=1}^{\hat{n}} \left[(1-\underline{\tau})\psi_i + (1-\tau_i)(\rho_i - \psi_i - \frac{a}{2}\frac{{\psi_i}^2}{\rho_i}) - c_i \right] \end{aligned} \tag{C.11}$$

Inserting the optimal ψ_i^{*th} and ψ_i^{*nth} and simplifying yields the following condition:

$$Inc_{th} = \sum_{i=1}^{\hat{n}} \frac{\rho_i \underline{\tau}(2\tau_i - \underline{\tau})}{2a(1-\tau_i)} + \sum_{i=\hat{n}+1}^{\tilde{n}} \left[(1-\tau_i)\rho_i + \frac{{\tau_i}^2 \rho_i}{2a(1-\tau_i)} - c_i \right] - c_0 \geq 0, \tag{C.12}$$

where Inc_{th} denotes the net benefit ("*Inc*entive") from investing in a tax haven. If this net benefit is positive the multinational chooses to invest in a tax haven.

Q.E.D.

C.4 Additional tables

Table C.1: Summary statistics, full sample

	Definition	Full sample			Manufacturing firms			Service firms		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Have haven	*Indicator variable; 1 if parent firm holds at least one affiliate in at least one tax haven in a given year, 0 otherwise.*	19165	.179	.383	11603	.170	.376	6733	.199	.399
# parent employees (in 1,000)	*Number of employees, parent firm.*	19165	1.188	8.176	11603	1.286	8.408	6733	1.060	8.245
Ln (# parent employees)	*Natural logarithm of number of parent employees.*	18039	5.437	1.833	11448	5.898	1.425	5828	4.523	2.169
# non-haven employees (in 1,000)	*Sum of number of employees in affiliates which are not located in tax havens, reduced according to share of participating interests.*	18158	.437	2.211	11153	.505	2.619	6185	.326	1.328
Ln (# non-haven employees)	*Natural logarithm of the number of foreign non-haven employees.*	17266	4.472	1.742	10824	4.672	1.665	5696	4.124	1.807
Average foreign non-haven tax rate	*Average of statutory tax rates faced by a firm's foreign non-haven affiliates weighted by number of affiliate employees adjusted by degree of participation.*	17189	.295	.058	10799	.299	.056	5670	.293	.058
Average foreign non-h. tax rate at 2001 loc.	*Average of statutory tax rates faced by a firm's affiliates at non-haven locations of 2001 weighted by GDP.*	13859	.315	.058	8830	.320	.055	4479	.308	.060

Manufacturing firms: firms classified NACE 1500-3700, *service firms:* firms classified NACE 5000-9300, with the before mentioned sample restrictions. *Full sample* additionally includes parent firms in the sectors agriculture, mining, electricity and water supply, and construction.

Table C.2: Regression results, full sample

	OLS	IV	FE	FE IV
Average foreign non-haven	0.509***	0.963***	0.046	0.126
tax rate	(0.083)	(0.120)	(0.101)	(0.144)
Parent size	-0.022**	-0.024*	-0.007	-0.013
	(0.011)	(0.012)	(0.014)	(0.015)
Parent size, squared	0.004***	0.005***	0.002	0.003*
	(0.001)	(0.001)	(0.002)	(0.002)
Non-haven size	-0.065***	-0.070***	-0.040***	-0.045***
	(0.012)	(0.016)	(0.012)	(0.014)
Non-haven size, squared	0.012***	0.013***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	-0.005	-0.118**	0.087*	0.088
	(0.038)	(0.053)	(0.048)	(0.059)
# of observations	16410	12755	16410	12755
# of parents	3680	2569	3680	2569
R-squared overall	0.10	0.14	0.13	0.14
R-squared within	–	–	0.02	0.02
Instrument	N	Y	N	Y
F-statistic	–	5687.41***	–	660.74***
Standard errors	Cluster	Cluster	Cluster	Bootstrap

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

Table C.3: Regression results, instrumentation based on first period's locations

	Manufacturing firms				Service firms			
	OLS	2SLS	FE	FE IV	OLS	2SLS	FE	FE IV
Ave. foreign	0.566***	0.902***	0.175*	0.323*	0.508***	0.945***	-0.085	0.168
n.h. tax rate	(0.096)	(0.125)	(0.094)	(0.190)	(0.162)	(0.176)	(0.124)	(0.173)
Parent size	-0.061***	-0.049**	-0.013	-0.016	0.024	0.018	-0.002	-0.010
	(0.021)	(0.021)	(0.003)	(0.026)	(0.016)	(0.016)	(0.013)	(0.012)
Parent size,	0.008***	0.007***	0.003	0.004	-0.000	0.000	0.001	0.001
squared	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Foreign non-	-0.051***	-0.049***	-0.052***	-0.044***	-0.064***	-0.061**	-0.037*	-0.045**
haven size	(0.015)	(0.015)	(0.014)	(0.014)	(0.023)	(0.024)	(0.020)	(0.022)
For. n.h. size,	0.011***	0.011***	0.009***	0.008***	0.011***	0.012***	0.009***	0.009***
squared	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Constant	0.022	-0.120*	0.070	0.013	-0.103	-0.185**	0.141***	0.104
	(0.062)	(0.067)	(0.069)	(0.083)	(0.075)	(0.076)	(0.053)	(0.064)
# of obs.	10662	10170	10662	10170	5052	4833	5052	4833
# of parents	2321	2246	2321	2246	1270	1226	1270	1226
R-squared overall	0.16	0.17	0.16	0.17	0.10	0.11	0.08	0.09
R-squared within	–	–	0.02	0.02	–	–	0.02	0.02
Instrument	N	Y	N	Y	N	Y	N	Y
F-Statistics	–	5969.57***	–	510.62***	–	3221.87***	–	787.27***
Standard errors	Cluster	Cluster	Cluster	Bootstrap	Cluster	Cluster	Cluster	Bootstrap

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

Table C.4: Regression results, firms without haven activity in 2001

	Manufacturing firms				Service firms			
	OLS	2SLS	FE	FE IV	OLS	2SLS	FE	FE IV
Ave. foreign	0.200***	0.241***	0.016	0.196**	0.144	0.050	-0.135	-0.109
n.h. tax rate	(0.065)	(0.070)	(0.080)	(0.099)	(0.107)	(0.103)	(0.098)	(0.156)
Parent size	-0.012	-0.000	0.007	0.012	0.014	0.000	-0.009	-0.010
	(0.011)	(0.009)	(0.032)	(0.040)	(0.010)	(0.011)	(0.010)	(0.012)
Parent size,	0.002	0.000	0.001	0.000	-0.000	0.001	0.000	0.001
squared	(0.001)	(0.001)	(0.004)	(0.005)	(0.001)	(0.002)	(0.001)	(0.002)
Foreign non-	-0.018*	0.005	-0.021**	-0.014	-0.017	-0.001	-0.016	-0.020
haven size	(0.010)	(0.007)	(0.009)	(0.009)	(0.014)	(0.011)	(0.015)	(0.021)
For. n.h. size,	0.004***	0.001	0.004***	0.003**	0.004*	0.001	0.004*	0.005
squared	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
Constant	0.017	-0.062*	-0.032	-0.096	-0.024	-0.006	0.141***	0.083
	(0.034)	(0.032)	(0.076)	(0.094)	(0.042)	(0.039)	(0.053)	(0.057)
# of obs.	8594	6549	8594	8594	4123	2893	4123	2893
# of parents	1943	1333	1943	1943	1075	653	1075	653
R-squared overall	0.18	0.02	0.03	0.01	0.05	0.03	0.02	0.01
R-squared within	–	–	0.02	0.02	–	–	0.02	0.03
Instrument	N	Y	N	Y	N	Y	N	Y
F-Statistics	–	3681.92***	–	568.14***	–	2754.72***	–	511.67***
Standard errors	Cluster	Cluster	Cluster	Bootstrap	Cluster	Cluster	Cluster	Bootstrap

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

Table C.5: Limited dependent variable models, coefficients

	Manufacturing firms				Service firms			
	Probit	IV probit	Logit	FE logit	Probit	IV probit	Logit	FE logit
Ave. foreign	3.117***	5.578***	5.940***	10.883**	2.700***	5.874***	5.176***	-2.524
n.h. tax rate	(0.612)	(0.984)	(1.173)	(5.538)	(1.005)	(1.272)	(1.858)	(8.751)
Parent size	-0.194*	-0.164	-0.314	0.325	0.239**	0.173*	0.504**	0.257
	(0.108)	(0.127)	(0.221)	(1.356)	(0.096)	(0.095)	(0.210)	(1.579)
Parent size,	0.029***	0.026**	0.049**	0.059	-0.012	-0.007	-0.027	-0.010
squared	(0.010)	(0.011)	(0.019)	(0.144)	(0.009)	(0.010)	(0.020)	(0.151)
Non-haven size	-0.078	-0.086	-0.003	-2.134**	-0.208**	-0.196	-0.353	-1.215
	(0.079)	(0.096)	(0.168)	(1.002)	(0.104)	(0.123)	(0.216)	(1.044)
Non-haven size,	0.032***	0.034***	0.045***	0.341***	0.039***	0.039***	0.068***	0.242*
squared	(0.008)	(0.010)	(0.017)	(0.104)	(0.012)	(0.013)	(0.024)	(0.125)
Constant	-2.528***	-3.387***	-5.017***	—	-2.825***	-3.677***	-5.292***	—
	(0.409)	(0.553)	(0.846)	—	(0.448)	(0.555)	(0.888)	—
Observations	10662	8533	10662	920	5052	3751	5052	447
Standard errors	Cluster	Cluster	Cluster	Bootstrap	Cluster	Cluster	Cluster	Bootstrap

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

Table C.6: Limited dependent variable models, average marginal effects

	Manufacturing firms				Service firms			
	Probit	IV probit	Logit	FE logit	Probit	IV probit	Logit	FE logit
Ave. foreign	0.578***	1.103***	0.593***	0.118	0.530***	1.211***	0.553***	-0.395
n.h. tax rate	(0.114)	(0.201)	(0.117)	(0.380)	(0.200)	(0.276)	(0.200)	(1.356)
Parent size	-0.036*	-0.032	-0.031	0.004	0.047**	0.036*	0.054**	0.040
	(0.019)	(0.025)	(0.022)	(0.009)	(0.019)	(0.020)	(0.022)	(0.248)
Parent size,	0.005***	0.005**	0.005**	0.001	-0.002	-0.001	-0.003	-0.002
squared	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.024)
Non-haven size	-0.014	-0.017	-0.000	-0.023	-0.041**	-0.040	-0.038	-0.190
	(0.015)	(0.019)	(0.017)	(0.089)	(0.020)	(0.025)	(0.023)	(0.167)
Non-haven size,	0.006***	0.007***	0.005***	0.004	0.008***	0.008***	0.007***	0.038*
squared	(0.002)	(0.002)	(0.002)	(0.014)	(0.002)	(0.003)	(0.003)	(0.019)
Observations	10662	8533	10662	920	5052	3751	5052	447

Standard errors in parentheses. Standard errors are estimated using the delta method. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

Table C.7: Regression results, including sector holding companies

	Manufacturing firms				Service firms			
	OLS	2SLS	FE	FE IV	OLS	2SLS	FE	FE IV
Ave. foreign	0.684***	1.196***	0.093	0.286	0.457***	1.265***	-0.119	-0.038
n.h. tax rate	(0.094)	(0.150)	(0.083)	(0.175)	(0.171)	(0.216)	(0.119)	(0.205)
Parent size	-0.029***	-0.030**	-0.008	-0.008	0.018	0.016	0.005	-0.015
	(0.010)	(0.013)	(0.017)	(0.021)	(0.014)	(0.016)	(0.015)	(0.014)
Parent size,	0.004***	0.004***	0.002	0.002	-0.000	-0.000	-0.000	0.002
squared	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Foreign non-	-0.052***	-0.050***	-0.049***	-0.045***	-0.053***	-0.053**	-0.004	-0.024
haven size	(0.012)	(0.015)	(0.013)	(0.017)	(0.018)	(0.023)	(0.020)	(0.017)
For. n.h. size,	0.012***	0.012***	0.009***	0.009***	0.011***	0.012***	0.004*	0.006***
squared	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Constant	-0.067*	-0.215***	0.117**	0.049	-0.070	-0.243***	0.114*	0.190***
	(0.040)	(0.063)	(0.048)	(0.076)	(0.070)	(0.082)	(0.058)	(0.064)
# of obs.	13611	10900	13611	10900	6372	4768	6372	4768
# of parents	3035	2213	3035	2213	1598	1070	1598	1070
R-squared overall	0.13	0.13	0.13	0.13	0.10	0.11	0.06	0.10
R-squared within	–	–	0.02	0.02	–	–	0.01	0.02
Instrument	N	Y	N	Y	N	Y	N	Y
F-Statistics	–	5359.76***	–	493.40***	–	2089.10***	–	659.43***
Standard errors	Cluster	Cluster	Cluster	Bootstrap	Cluster	Cluster	Cluster	Bootstrap

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies. The significance level of the tax variable in the fixed effects IV specification for manufacturing firms is 10.1%.

Table C.8: Regression results, excluding Switzerland

	Manufacturing firms				Service firms			
	OLS	2SLS	FE	FE IV	OLS	2SLS	FE	FE IV
Ave. foreign	0.453***	0.780***	0.035	0.327**	0.298**	0.597***	-0.012	0.014
n.h. tax rate	(0.074)	(0.116)	(0.076)	(0.152)	(0.142)	(0.163)	(0.087)	(0.124)
Parent size	-0.057***	-0.064***	-0.010	-0.009	0.004	-0.002	-0.006	-0.004
	(0.016)	(0.020)	(0.024)	(0.028)	(0.009)	(0.011)	(0.007)	(0.010)
Parent size,	0.006***	0.007***	0.003	0.003	0.000	0.001	0.000	-0.000
squared	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)
Foreign non-	-0.075***	-0.078***	-0.043***	-0.030**	-0.069***	-0.074***	-0.036**	-0.055***
haven size	(0.013)	(0.016)	(0.014)	(0.015)	(0.020)	(0.026)	(0.018)	(0.020)
For. n.h. size,	0.013***	0.013***	0.007***	0.006***	0.011***	0.012***	0.008***	0.010***
squared	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Constant	0.090**	0.027	0.054	-0.050	-0.006	-0.046	0.080**	0.099**
	(0.046)	(0.065)	(0.062)	(0.082)	(0.061)	(0.071)	(0.033)	(0.049)
# of obs.	10662	8533	10662	8533	5052	3751	5052	3751
# of parents	2321	1696	2321	1696	1270	832	1270	832
R-squared overall	0.19	0.19	0.18	0.19	0.11	0.11	0.08	0.09
R-squared within	–	–	0.02	0.01	–	–	0.03	0.04
Instrument	N	Y	N	Y	N	Y	N	Y
F-Statistics	–	4064.39***	–	465.73***	–	1872.06***	–	717.80***
Standard errors	Cluster	Cluster	Cluster	Bootstrap	Cluster	Cluster	Cluster	Bootstrap

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

Table C.9: Regression results using tax rates from Djankov, Ganser, McLiesh, Ramalho, and Shleifer (2010)

	Manufacturing		Services	
	OLS	2SLS	OLS	2SLS
Average foreign non-haven	0.517***	0.537***	0.536**	0.917***
tax rate	(0.122)	(0.153)	(0.209)	(0.309)
Parent size	-0.078**	-0.080**	0.026	0.016
	(0.030)	(0.036)	(0.019)	(0.020)
Parent size, squared	0.009***	0.009***	-0.000	0.001
	(0.003)	(0.003)	(0.002)	(0.002)
Foreign non-haven size	-0.069***	-0.082***	-0.038	-0.039
	(0.020)	(0.027)	(0.032)	(0.040)
Foreign non-haven size,	0.013***	0.014***	0.008**	0.008*
squared	(0.002)	(0.003)	(0.004)	(0.004)
Constant	0.149*	0.195*	-0.086	-0.135
	(0.090)	(0.111)	(0.085)	(0.117)
# of observations	1482	1254	686	543
# of parents	1482	1254	686	543
R-squared	0.19	0.18	0.08	0.08
Instrument	N	Y	N	Y
F-statistics instrument	–	2106.93***	–	1063.09***
Standard errors	Cluster	Cluster	Cluster	Cluster

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

C.5 Endogeneity of non-haven size

To address the endogeneity of non-haven size, we focus our analysis on the years 2002 and 2008.[1] We restrict our regression sample to firms that changed their tax haven use between 2002 and 2008. Thus, our strategy mimics fixed effects logit, where the coefficients are identified given that the dependent variable has changed. We define a new dependent variable that takes the value zero if a firm holds a tax haven affiliate in 2002 and does not in 2008, and the value one if a firm does not have a tax haven affiliate in 2002 and does so in 2008:

$$\Delta_i = \begin{cases} 0 & \text{if } y_{i2002} = 1 \text{ and } y_{i2008} = 0 \\ 1 & \text{if } y_{i2002} = 0 \text{ and } y_{i2008} = 1 \end{cases} \tag{C.13}$$

About 60% of firms start using a tax haven, around 40% of firms close down their tax haven activities. As independent variables, we use changes in parent size, the size of non-haven activities and the average foreign non-haven tax rate. This way, we partial out the firm fixed effect.

To construct the instrument, we focus on affiliates that a multinational holds in both 2002 and 2008, because tax rate changes at these locations are exogenous to any locational changes that a firm has made after opening an affiliate in a tax haven or closing down its haven activities. To take the endogeneity of the foreign non-haven size into account, we use the insight that affiliate growth can be very well explained by GDP growth with a coefficient that is not significantly different from one (Desai, Foley, and Hines, 2006). We inflate the size of the foreign non-haven affiliates in 2002 by GDP growth between 2002 and 2008 and use the resulting hypothetical change in the foreign non-haven size as an instrument for the actual change in foreign non-haven size between 2008 and 2002. Only actual non-haven size includes affiliates which may have been opened or closed due to the decision to invest or stop investing in a tax haven. Note that we can only include size linearly in our regression because our strategy yields only one instrument.

[1] Note that considerable fraction of variation is lost in this manner, because firms exit before 2008 or enter after 2002, or because they revise their decision to invest in a tax haven more than once during our period of analysis.

Table C.10: Regression results, endogenous non-haven size

	Manufacturing firms		Service firms	
	1	2	3	4
Δ ave. foreign	0.052*	0.048	0.052	0.056
n.h. tax rate	(0.027)	(0.036)	(0.030)	(0.030)
Δ parent size	0.120***	0.122***	0.134	0.134
	(0.041)	(0.041)	(0.143)	(0.134)
Δ non-haven size	0.153**	0.181	0.128***	0.205**
	(0.069)	(0.128)	(0.044)	(0.101)
Constant	0.704***	0.683***	0.731***	0.686***
	(0.123)	(0.170)	(0.162)	(0.189)
# of parents	88	88	32	32
R-squared	0.20	0.20	0.22	0.18
Endogenous variable	Tax	Tax & Size	Tax	Tax & Size
F-statistics (tax)	27.35***	32.86***	27.24***	14.77***
F-statistics (size)	–	42.80***	–	2.56*
Partial R-squared (tax)	–	0.16	–	0.45
Partial R-squared (size)	–	0.12	–	0.17
Standard errors	Robust	Robust	Robust	Robust

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses.
The significance level of the coefficient of non-haven size is 15.6% in column 2. The coefficient of the average foreign non-haven tax rate has a P-value of 18.7% in column 2 and 24.7% in column 4.
Dependent variable: dummy variable; 1 if a parent firm did not hold an affiliate in a tax haven in 2002, but does so in 2008, zero if it did hold a tax haven affiliate in 2002 and does not so in 2008. Independent variables: differences in the average foreign non-haven tax rate, the number of parent employees and the number of non-haven employees between 2002 and 2008.

We instrument the difference of the average foreign non-haven tax rates with the difference of tax rates had a firm refrained from adjustments in its location choices. The idea of this alternative instrument is to capture changes in the firm's average tax rate that are exogenous to the firm and do not depend on changes in tax haven use.

We calculate the sum of the differences of the tax rates interacted with GDP at the locations where a firm is present in both 2002 and 2008 and weight the single differences with the difference in GDP:

$$T_{it} = \frac{\sum_{l=1}^{L}(\tau_{l,2008} \cdot GDP_{l,2008} - \tau_{l,2002} \cdot GDP_{l,2002})}{\sum_{l=1}^{L}(GDP_{l,2008} - GDP_{l,2002})}, \tag{C.14}$$

where $l = 1, ..., L$ are the locations of a firm in both 2002 and 2008.

Table C.10 presents the estimation results, which, for manufacturing firms, are largely consistent with estimated tax effects obtained earlier using different identification methods and much larger samples. In the specification that instruments only for the average foreign non-haven tax rate, the 0.052 coefficient indicates that firms whose average foreign non-haven tax rates increase are more likely than others to add a tax haven affiliate. This effect is significant at the 5.9% level in a two-tailed test, implying that it is possible to reject (using a one-tailed test) that high non-haven foreign tax rates discourage tax haven affiliate ownership, as reported by Desai, Foley, and Hines (2006) for U.S. firms. If both non-haven size and taxation are instrumented, the resulting coefficients are of similar magnitude but lower significance levels, reflecting in part the small sample size. Firms with growing parent companies are more likely than others to add tax haven affiliates, which is consistent with earlier findings.

It is difficult to draw ambitious inferences about the behavior of service firms from the regressions reported in columns 3-4, largely due to the problem of using instrumental variables methods with a small (32 firms) sample, though greater size of non-haven foreign operations is associated with adding tax haven operations, which is again consistent with patterns appearing in other regressions.

C.6 Alternative instrumentation strategy

This section presents regression results based on an alternative instrumentation strategy similar to the one used in Desai, Foley, and Hines (2006). We instrument the parent's foreign non-haven tax rate with the competitors' average foreign non-haven tax rate. The competitors are defined as the other firms in the same sector. Firms in the same sector react to similar incentives in choosing their investment destinations because similar location factors are beneficial for them. At the same time, the competitors' investment decision is exogenous to whether a certain firm in the sector invests in a tax haven.

Endogenous and correlated group effects as discussed by Manski (1993) are unlikely to contaminate the regression results, i.e. it is unlikely that a firm's tax haven investment decision is directly affected by the choices of other firms in the same sector, or that the tax haven investment by firms in the same sector is correlated because entry in a tax haven of some firms exerts competitive pressure on the remaining firms in the sector to follow suit. As highlighted by Manski (1993), the existence and detection of group effects presupposes that the reference group relevant to an individual is correctly specified and that the group mean behavior can be correctly perceived by the individual group members. It is unlikely that the firms in the same sector in our data constitute the reference group to a firm. The sector classification is assigned to the firm by the statistical department of the Bundesbank. It is coarse and the resulting firm groups are sizeable. Firms' reference groups are determined by their choice of products and markets. Further, for some endogenous effect to exist, a firm would have to be able to observe the other firms' mean tax haven investment and activities. There is no evident way how firms could obtain this information. For sake of reputation, firms are unlikely to announce that they have opened a tax haven affiliate and intend to reallocate taxable income there.

Nonetheless, we prefer the instrumentation strategy used in the main text of this chapter because it is difficult to rule out the possibility that firms in similar industries are influenced by correlated omitted variables that influence the location of all of their foreign investments, including tax havens and non-havens. As it happens, the estimation results based on competitors' average foreign non-haven tax rates are very similar to the esti-

Table C.11: Regression results, instrumentation as Desai, Foley, and Hines (2006)

	Manufacturing firms				Service firms			
	1	2	3	4	5	6	7	8
	OLS	2SLS	FE	FE IV	OLS	2SLS	FE	FE IV
Ave. foreign	0.007***	0.027***	0.002***	0.039*	0.007***	0.029	-0.000	-0.046
n.h. tax rate	(0.001)	(0.007)	(0.001)	(0.023)	(0.001)	(0.021)	(0.001)	(0.165)
Parent size	-0.064***	-0.060***	-0.013	-0.008	0.024	0.023	-0.002	0.007
	(0.021)	(0.021)	(0.025)	(0.030)	(0.016)	(0.017)	(0.013)	(0.053)
Parent size,	0.008***	0.007***	0.003	0.002	-0.000	-0.000	0.001	-0.001
squared	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.006)
Foreign non-	-0.051***	-0.042***	-0.052***	-0.036	-0.063***	-0.046	-0.037*	-0.051
haven size	(0.015)	(0.015)	(0.014)	(0.024)	(0.024)	(0.029)	(0.020)	(0.051)
Foreign n.h. size,	0.011***	0.011***	0.009***	0.006*	0.011***	0.010***	0.009***	0.011
squared	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.010)
Constant	-0.006	-0.597***	0.051	-0.995	-0.160**	-0.742	0.119**	1.357
	(0.060)	(0.212)	(0.068)	(0.684)	(0.069)	(0.604)	(0.055)	(4.443)
# of observations	10661	10661	10661	10661	5052	5047	5052	5047
# of parents	2320	2320	2320	2320	1270	1269	1270	1269
R-squared	0.17	0.09	0.01	–	0.11	–	0.02	–
F-Statistics	–	40.95***	–	6.53**	–	5.76**	–	3.36*
Standard errors	Cluster	Cluster	Cluster	Bootstrap	Cluster	Cluster	Cluster	Bootstrap

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Standard errors are clustered at the level of the parent firm or bootstrapped with 200 replications.
The coefficients of foreign non-haven size in column 4 are significant at 12.5% (linear term) and 5.3% (squared term). The F-statistic for the instrument is significant at 1.07%. In column 6, the coefficient of linear non-haven size has a P-value of 10.7%. *Regression sample:* column 1-4 manufacturing firms, i.e. firms classified NACE 1500-3700; column 5-8 service firms, i.e. firms classified NACE 500-9300, except NACE 65xx, 70xx, 7490, 75xx, 95xx. Dependent variable: dummy variable which denotes whether a parent owns at least one affiliate in at least one tax haven in a given year. Independent variables: see Table 3.2.

mation results in the main text, so there is reason to expect that the impact of potential correlated omitted variables is not particularly strong.

Table C.11 corresponds to Table 3.3 in the main text. The regression results are very similar; merely the significance levels are higher for the group of service firms in Table 3.3.

Table C.12 corresponds to Table 3.4. We do not report regressions with interaction terms, as both the R&D intensity and the interaction term usually turned out insignificant. Effects in Table C.12 are weaker than effects in Table 3.4, possibly reflecting the fact that the strength of the instrument as measured by the F-statistic is much higher if the instrumentation is based on firms' initial location decisions.

Table C.12: Regression results including sector R&D intensity, instrumentation as Desai, Foley, and Hines (2006)

	Manufacturing		Services	
	1	2	3	4
	OLS	IV	OLS	IV
Average foreign nh.	0.007***	0.033***	0.006***	-0.053
tax rate	(0.001)	(0.010)	(0.001)	(0.065)
Parent size	-0.060**	-0.057**	0.014	0.013
	(0.022)	(0.023)	(0.015)	(0.015)
Parent size, squared	0.008***	0.007***	0.002	0.003**
	(0.002)	(0.002)	(0.002)	(0.002)
Non-haven size	-0.051**	-0.038	-0.076***	-0.115**
	(0.021)	(0.025)	(0.022)	(0.048)
Non-haven size,	0.011***	0.010***	0.013***	0.016***
squared	(0.003)	(0.003)	(0.002)	(0.004)
Sector R&D intensity	-0.001	-0.007	0.009	0.027
	(0.006)	(0.007)	(0.006)	(0.021)
Constant	-0.026	-0.755***	-0.112**	1.544
	(0.077)	(0.291)	(0.050)	(1.805)
# of observations	9915	9915	3860	3860
# of sectoral groups	23	23	14	14
R-squared	0.17	0.04	0.16	–
F-Statistics	–	11.88***	–	4.78*
Standard errors	Cluster	Cluster	Cluster	Cluster

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Standard errors are clustered at the level of the parent firm.
Dependent variable: dummy variable which denotes whether a parent owns at least one affiliate in at least one tax haven in a given year. Independent variables: see Table 3.2; *sector R&D intensity:* calculated as expenditures on innovation over total sales by sector.

Finally, Table C.15 displays results estimated on imputed data and Tables C.13 and C.14 present summary statistics to gauge the quality of the imputation.

As indicated in the main text, almost a fifth of observations on parent firms in services drop from the regression sample because the number of parent or non-haven employees is zero, so size is not defined.

We impute the data in order to use all available observations and ensure that our results do not reflect sample selection based on size. We impute the sales variable because zero observed sales are a result of the reporting requirements. Sales are surveyed in million euros, so they are zero for any firm with a turnover of strictly less than 500,000 euros. This implies that a reported figure x disguises possible true sales values ranging from $1,000,000x - 500,000$ to $1,000,000x + 500,000$, except for 0 which disguises the interval $[0; 499,999]$.

We use a model for grouped data developed in the statistical literature by Heitjan and Rubin (1990), imputing parent and non-haven sales based on the assumption that sales are log-normally distributed. This assumption is reasonable given the distribution of observed sales in our data. The imputation procedure basically consists of two steps. First, we estimate the relationship of the natural log of parent sales and the natural log of non-haven sales and the other variables employed in our analysis using the EM-algorithm proposed by Hasselblad, Stead, and Galke (1980). Then, we follow Heitjan and Rubin (1990) and impute plausible values for parent sales and non-haven sales given the other variables as well as the coefficients derived in the first step and based on the assumption of a normally distributed error term. Conditional on the other covariates, we generate 120 sets of plausible data, as Heitjan and Rubin (1990) use a similar number of imputations in their study.

Tables C.13 and C.14 provide summary statistics on several sets of imputed sales compared to the observed parent and non-haven sales. The left parts of the tables show statistics for all parent-years, that is, including those units with missing observed sales. As expected, the mean of the imputed sales are lower and the standard deviation is higher. The right part of the table shows the statistics only for those units with non-missing observed parent and non-haven sales. The moments of the distributions of the imputed data sets are very close to the corresponding moments of the observed data. We are thus assured of not introducing some artificial correlation or bias into our analysis through our imputation procedure.

We re-run our analyses and re-calculate coefficients and standard errors based on the formulas developed by Rubin (1987) which take into account that our data are imputed.

Table C.15 provides the estimated coefficients for regressions using data on service firms. The use of imputed data does not appear to produce major changes in the coefficient and significance patterns. In particular, higher average non-haven foreign tax rates continue to be associated with greater likelihood of tax haven affiliate ownership in the uninstrumented regressions and with no effect on tax haven affiliate ownership in the instrumented regressions. There are small differences in the estimated effects of parent size on tax haven affiliate ownership, but these are largely unimportant, so this evidence suggests that the difference in the effect of taxes on tax haven investment between manufacturing and service firms is unlikely to be attributable to sample selection among service firms.

Instead of this rather complex procedure, one could opt for a more pragmatic solution such as plugging in “1” in place of the zeros. One could argue that “1” is close to zero relative to the other values observed, so measurement error should be negligible, but nonetheless, all observations could be used in the analysis. We prefer imputation because plugging in “1” would create an artificial censoring value. According to a recent literature started by Rigobon and Stoker (2007, 2009), this could introduce further bias in our analysis because previously missing observations then pile up at $ln(1)$. For the same reason, plugging in any other value below 500,000 instead of the zeros is not a viable option. Further, the imputation procedure takes into account the correlation between the sales variables and the other variables employed in our analysis, so it deals with the missing data in a statistically appropriate way.

Table C.13: Comparison of distributions of observed and imputed parent sales

	All parent-years					Only non-zero observed sales				
	N	Mean	SD	Skew.	Kurt.	N	Mean	SD	Skew.	Kurt.
Ln (parent sales), observed	16978	18.240	1.688	.111	3.877	16978	18.240	1.688	.111	3.877
Ln (parent sales), imputed	18048	17.896	2.138	-.679	4.239	16978	18.238	1.692	.090	3.922
Ln (parent sales), imputed	18048	17.896	2.140	-.684	4.252	16978	18.239	1.691	.095	3.914
Ln (parent sales), imputed	18048	17.897	2.137	-.676	4.230	16978	18.238	1.692	.091	3.920
Ln (parent sales), imputed	18048	17.899	2.132	-.669	4.225	16978	18.239	1.690	.100	3.908
Ln (parent sales), imputed	18048	17.896	2.139	-.682	4.244	16978	18.238	1.692	.091	3.924
Ln (parent sales), imputed	18048	17.899	2.133	-.668	4.216	16978	18.239	1.690	.098	3.908
Ln (parent sales), imputed	18048	17.897	2.138	-.678	4.240	16978	18.238	1.693	.089	3.924
Ln (parent sales), imputed	18048	17.898	2.135	-.674	4.227	16978	18.239	1.690	.010	3.906
Ln (parent sales), imputed	18048	17.895	2.143	-.690	4.265	16978	18.238	1.692	.090	3.924

Table C.14: Comparison of distributions of observed and imputed non-haven sales

	All parent-years					Only non-zero observed sales				
	N	Mean	SD	Skew.	Kurt.	N	Mean	SD	Skew.	Kurt.
Ln (non-haven sales), observed	17201	16.858	1.589	.458	3.670	17201	16.858	1.589	.458	3.670
Ln (non-haven sales), imputed	18048	16.655	1.809	-.062	3.767	17201	16.859	1.590	.443	3.714
Ln (non-haven sales), imputed	18048	16.657	1.803	-.042	3.736	17201	16.860	1.587	.456	3.688
Ln (non-haven sales), imputed	18048	16.657	1.803	-.040	3.720	17201	16.860	1.588	.456	3.680
Ln (non-haven sales), imputed	18048	16.658	1.801	-.035	3.708	17201	16.859	1.591	.443	3.702
Ln (non-haven sales), imputed	18048	16.657	1.804	-.045	3.726	17201	16.860	1.589	.446	3.708
Ln (non-haven sales), imputed	18048	16.657	1.801	-.034	3.706	17201	16.858	1.591	.439	3.717
Ln (non-haven sales), imputed	18048	16.658	1.802	-.040	3.726	17201	16.861	1.586	.463	3.675
Ln (non-haven sales), imputed	18048	16.657	1.802	-.035	3.699	17201	16.859	1.590	.447	3.696
Ln (non-haven sales), imputed	18048	16.658	1.801	-.036	3.717	17201	16.860	1.589	.446	3.707

Table C.15: Regression results based on imputed data, service firms

	OLS	IV	FE	FE IV
Average foreign non-haven	0.005***	0.034***	0.000	-0.029
tax rate	(0.001)	(0.001)	(0.015)	(0.208)
Ln (parent sales)	-0.065*	-0.064**	-0.024	-0.016
	(0.034)	(0.025)	(0.036)	(0.043)
Ln (parent sales), squared	0.002**	0.002***	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.002)
Ln (non-haven sales)	-0.176***	-0.090*	-0.172**	-0.217
	(0.055)	(0.055)	(0.072)	(0.033)
Ln (non-haven sales),	0.007***	0.004*	0.006***	0.008
squared	(0.002)	(0.002)	(0.002)	(0.011)
Constant	1.536***	0.0757	1.414	2.4513
	(0.144)	(0.427)	(0.866)	(7.684)
# of observations	6140	6140	6140	6140
# of imputations	120	120	120	120
Instrument	N	Y	N	Y
Mean F-statistics	—	12.03***	—	10.30***
Standard errors	Cluster	Cluster	Cluster	Bootstrap

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications contain year dummies.

List of Figures

List of Tables

Bibliography

Ambos, T. C., and B. Ambos (2009): "The impact of distance on knowledge transfer effectiveness in multinational corporations," *Journal of International Management*, 15(1), 1–14.

Angrist, J. D., and J.-S. Pischke (2009): *Mostly harmless econometrics*. Princeton: Princeton University Press.

Antràs, P. (2003): "Firms, contracts, and trade structure," *The Quarterly Journal of Economics*, 118(4), 1375–1418.

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 E. Helpman (2004): "Global sourcing," *Journal of Political Economy*, 112(3), 552–580.

Bahar, D. (2013): "Heavier than air? Knowledge transmission within the multinational firm," *mimeo*, Harvard University.

Bayard, K., and K. R. Troske (1999): "Examining the employer-size wage premium in the manufacturing, retail trade, and service industries using employer-employee matched data," *The American Economic Review*, 89(2), 99–103.

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 (2009): "The distinct effects of information technology and communication technology on firm organization," NBER Working Paper No. 14975.

Bloom, N., R. Sadun, and J. Van Reenen (2010): "Recent advances in the empirics of organizational economics," *Annual Review of Economics*, 2(1), 105–137.

——— (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.

Brainard, S. L. (1997): "An empirical assessment of the proximity-concentration trade-off between multinational sales and trade," *The American Economic Review*, 87(4), 520–544.

Buettner, T., M. Overesch, U. Schreiber, and G. Wamser (2009): "Taxation and capital structure choice—Evidence from a panel of German multinationals," *Economics Letters*, 105(3), 309–311.

Buettner, T., and G. Wamser (2013): "Internal debt and multinational's profit shifting—Empirical evidence from firm-level panel data," *National Tax Journal*, 66(1), 63–96.

Bundesministerium der Finanzen (2007): "Hinzurechnungsbesteuerung nach dem Außensteuergesetz (AStG)," BMF Schreiben vom 08.01.2007.

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2012): "The anatomy of French production hierarchies," NBER Working Paper No. 18259.

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

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," NBER Working Paper No. 18522.

Carr, D. L., J. R. Markusen, and K. E. Maskus (2001): "Estimating the knowledge-capital model of the multinational enterprise," *The American Economic Review*, 91(3), 693–708.

Chen, M. X., and M. O. Moore (2010): "Location decision of heterogeneous multinational firms," *Journal of International Economics*, 80(2), 188–199.

Clausing, K. (2001): "The impact of transfer pricing on intrafirm trade," in *International taxation and multinational activity*, ed. by J. R. Hines, pp. 173–94. Chicago: University of Chicago Press.

——— (2003): "Tax-motivated transfer pricing and U.S. intrafirm trade prices," *Journal of Public Economics*, 87(9-10), 2207–2223.

——— (2006): "International tax avoidance and U.S. international trade," *National Tax Journal*, 59(2), 269–87.

Cremer, J., L. Garicano, and A. Prat (2007): "Language and the theory of the firm," *The Quarterly Journal of Economics*, 122(1), 373–407.

Crystal, D. (2003): *English as a global language*. Cambridge: Cambridge University Press.

De Long, D. W., and L. Fahey (2000): "Diagnosing cultural barriers to knowledge management," *The Academy of Management Executive*, 14(4), 113–127.

Defever, F. (2012): "The spatial organization of multinational firms," *Canadian Journal of Economics*, 45(2), 672–697.

Desai, M. A., C. F. Foley, and J. R. Hines (2004): "A multinational perspective on capital structure choice and internal capital markets," *Journal of Finance*, 59(6), 2451–2488.

——— (2006): "The demand for tax haven operations," *Journal of Public Economics*, 90(3), 513–531.

Deutscher Bundestag (2010): "Entwurf eines Jahressteuergesetzes 2010 (JStG 2010)," Bundestagsdrucksache 17/2249, published on 06/21/2010.

Dharmapala, D., and J. R. Hines (2009): "Which countries become tax havens?," *Journal of Public Economics*, 93(9-10), 1058–1068.

Dischinger, M., and N. Riedel (2011): "Corporate taxes and the location of intangible assets within multinational firms," *Journal of Public Economics*, 95(7), 691–707.

Djankov, S., T. Ganser, C. McLiesh, R. Ramalho, and A. Shleifer (2010): "The effect of corporate taxes on investment and entrepreneurship," *American Economic Journal: Macroeconomics*, 2(3), 31–64.

Drucker, J. (2010): "Google 2.4% rate shows how $60 billion lost to tax loopholes," http://www.bloomberg.com/news/2010-10-21/google-2-4-rate-shows-how-60-billion-u-s-revenue-lost-to-tax-loopholes.html, published on 10/21/2010.

Durlauf, S., S. Navarro, and D. A. Rivers (2010): "Understanding aggregate crime regressions," *Journal of Econometrics*, 158(2), 306–317.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009): "Revisiting the German wage structure," *The Quarterly Journal of Economics*, 124(2), 843–881.

Efron, B., and R. Tibshirani (1998): *An introduction to the bootstrap*. London: Chapman & Hall.

Egger, P., W. Eggert, C. Keuschnigg, and H. Winner (2010): "Corporate taxation, debt financing and foreign-plant ownership," *European Economic Review*, 54(1), 96–107.

Egger, P., and G. Wamser (2010): "The impact of CFC legislation on multinational firms—Evidence from a two-dimensional regression discontinuity approach," CEPR Discussion Paper No. 8460.

Ekholm, K., R. Forslid, and J. R. Markusen (2007): "Export-platform foreign direct investment," *Journal of the European Economic Association*, 5(4), 776–795.

Eurostat (2013): "Statistics on European cities," http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Statistics_on_European_cities, accessed on 12/11/2013.

Falck, O., S. Heblich, A. Lameli, and J. Südekum (2012): "Dialects, cultural identity, and economic exchange," *Journal of Urban Economics*, 72(2), 225–239.

FAZ (2013): "Die größten Unternehmen 2013," FAZ online, published on 07/03/2013.

Fitzenberger, B., A. Osikominu, and R. Völter (2005): "Imputation rules to improve the education variable in the IAB employment subsample," ZEW - Centre for European Economic Research Discussion Paper No. 05-10.

Foss, N. J., and T. Pedersen (2002): "Transferring knowledge in MNCs: The role of sources of subsidiary knowledge and organizational context," *Journal of International Management*, 8(1), 49–67.

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," *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.

——— (2009): "Specialization, firms, and markets: The division of labor within and between law firms," *Journal of Law, Economics and Organization*, 25(2), 339–371.

Garicano, L., and E. Rossi-Hansberg (2006): "Organization and inequality in a knowledge economy," *The Quarterly Journal of Economics*, 121(4), 1383–1435.

Gartner, H. (2005): "The imputation of wages above the contribution limit with the German IAB employment sample," FDZ-Methodenreport 02/2005.

Ghemawat, P. (2007): "Managing differences: The central challenge of global strategy," *Harvard Business Review*, 85(5), 59–68.

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.

Grossman, G. M., E. Helpman, and A. Szeidl (2006): "Optimal integration strategies for the multinational firm," *Journal of International Economics*, 70(1), 216–238.

Grubert, H., and J. Slemrod (1998): "The effect of taxes on investment and income shifting to Puerto Rico," *The Review of Economics and Statistics*, 80(3), 365–373.

Gruhl, A., A. Schmucker, and S. Seth (2012): "The Establishment History Panel 1975-2010: Handbook version 2.2.1," FDZ-Datenreport 04/2012.

Guiso, L., P. Sapienza, and L. Zingales (2009): "Cultural biases in economic exchange?," *The Quarterly Journal of Economics*, 124(3), 1095–1131.

Hansen, M. T., and B. Løvås (2004): "How do multinational companies leverage technological competencies? Moving from single to interdependent explanations," *Strategic Management Journal*, 25(8-9), 801–822.

Harding, T., and B. S. Javorcik (2011): "Roll out the red carpet and they will come: Investment promotion and FDI inflows," *The Economic Journal*, 121(557), 1445–1476.

Harris, D., R. Morck, J. Slemrod, and B. Yeung (1993): "Income shifting in U.S. multinational corporations," in *Studies in international taxation*, ed. by A. Giovannini, R. Hubbard, and J. Slemrod, pp. 277–308. Chicago: University of Chicago Press.

Hasselblad, V., A. G. Stead, and W. Galke (1980): "Analysis of coarsely grouped data from the lognormal distribution," *Journal of the American Statistical Association*, 75(372), 771–778.

Heitjan, D. F., and D. B. Rubin (1990): "Inference from coarse data via multiple imputation with application to age heaping," *Journal of the American Statistical Association*, 85(410), 304–314.

Helpman, E. (1984): "A simple theory of international trade with multinational corporations," *Journal of Political Economy*, 92(3), 451–471.

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

Herzog, T. N., F. J. Scheuren, and W. E. Winkler (2007): *Data quality and record linkage techniques*. New York: Springer.

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.

Hines, J. R., and E. M. Rice (1994): "Fiscal paradise: Foreign tax havens and American business," *The Quarterly Journal of Economics*, 109(1), 149–182.

Höhn, N., and J. Höring (2010): *Das Steuerrecht international agierender Unternehmen. Grenzüberschreitende Steuerplanung*. Wiesbaden: Gabler.

Huizinga, H., and L. Laeven (2008): "International profit shifting within multinationals: A multi-country perspective," *Journal of Public Economics*, 92(5), 1164–1182.

IBFD (2002-2006): *European tax handbook*. Amsterdam: International Bureau of Fiscal Documentation (IBFD).

——— (2007-2009): *Global corporate tax handbook*. Amsterdam: International Bureau of Fiscal Documentation (IBFD).

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

Jacobebbinghaus, P., and S. Seth (2007): "The German integrated employment biographies sample IEBS," *Schmollers Jahrbuch*, 127(2), 335–342.

Jäger, S. (2013): "The return to higher education: Evidence from university openings in Germany," *mimeo*, Harvard University.

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

Javorcik, B. S. (2004): "Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages," *The American Economic Review*, 94(3), 605–627.

Joint Committee on Taxation (2010): "Present law and background related to possible income shifting and transfer pricing," U.S. House Committee on Ways and Means.

Kaldor, N. (1934): "The equilibrium of the firm," *The Economic Journal*, 44(173), 60–76.

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.

Karkinsky, T., and N. Riedel (2012): "Corporate taxation and the choice of patent location within multinational firms," *Journal of International Economics*, 88(1), 176–185.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2010): "The Worldwide Governance Indicators: Methodology and analytical issues," *Policy Research Working Paper 5430*, World Bank.

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

Keller, W., and S. R. Yeaple (2013): "The gravity of knowledge," *The American Economic Review*, 103(4), 1414–44.

Kremer, M. (1993): "The O-ring theory of economic development," *The Quarterly Journal of Economics*, 108(3), 551–575.

Lagerström, K., and M. Andersson (2003): "Creating and sharing knowledge within a transnational team? The development of a global business system," *Journal of World Business*, 38(2), 84–95.

Lipponer, A. (2009): "Microdatabase Direct investment—MiDi. A brief guide," Technical documentation, Deutsche Bundesbank.

Lucas, L., B. Jopson, and V. Houlder (2012): "Starbucks ground down," Financial Times, published on 12/07/2012.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60(3), 531–542.

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

Markusen, J. R., and K. E. Maskus (2001): "Multinational firms: reconciling theory and evidence," in *Topics in empirical international economics: A festschrift in honor of Robert E. Lipsey*, ed. by M. Blomström, and L. S. Goldberg, pp. 71–98. Chicago: University of Chicago Press.

——— (2002): "Discriminating among alternative theories of the multinational enterprise," *Review of International Economics*, 10(4), 694–707.

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

Muendler, M.-A., and S. O. Becker (2010): "Margins of multinational labor substitution," *The American Economic Review*, 100(5), 1999–2030.

Nunn, N. (2007): "Relationship-specificity, incomplete contracts, and the pattern of trade," *The Quarterly Journal of Economics*, 122(2), 569–600.

OECD (2000): "Towards global tax co-operation. Progress in identifying and eliminating harmful tax practices," Paris: OECD Publications.

——— (2013): "Action plan on base erosion and profit shifting," Paris: OECD Publications.

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.

Overesch, M., and G. Wamser (2010): "Bilateral internal debt financing, tax planning, and the effectiveness of anti-avoidance rules," *mimeo*, University of Mannheim.

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., V. Rappoport, and K. J. Ruhl (2013): "Horizontal vs. vertical FDI: Revisiting evidence from US multinationals," *mimeo*, University of California at San Diego.

Rigobon, R., and T. M. Stoker (2007): "Estimation with censored regressors: Basic issues," *International Economic Review*, 48(4), 1441–1467.

——— (2009): "Bias from censored regressors," *Journal of Business & Economic Statistics*, 27(3), 340–353.

Rubin, D. B. (1987): *Multiple imputation for nonresponse in surveys*. Hoboken: John Wiley & Sons.

Ruf, M., and A. J. Weichenrieder (2012): "The taxation of passive foreign investment. Lessons from German experience," *Canadian Journal of Economics*, 45(4), 1504–1528.

Schafer, J. L. (1999): "Multiple imputation: a primer," *Statistical Methods in Medical Research*, 8(1), 3–15.

Schmieder, J. F., T. Von Wachter, and S. Bender (2012): "The effects of extended unemployment insurance over the business cycle: Evidence from regression discontinuity estimates over 20 years," *The Quarterly Journal of Economics*, 127(2), 701–752.

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

Staiger, D., and J. H. Stock (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65(3), 557–586.

Stock, J. H., and M. Yogo (2002): "Testing for weak instruments in linear IV regression," NBER Technical Working Paper No. 284.

The Economist (2012): "Wake up and smell the coffee," The Economist Print Edition, published on 12/15/2012.

Tintelnot, F. (2012): "Global production with export platforms," *mimeo*, Pennsylvania State University.

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

Troske, K. R. (1999): "Evidence on the employer size-wage premium from worker-establishment matched data," *The Review of Economics and Statistics*, 81(1), 15–26.

UNCTAD (2004): *World investment report 2004. The shift towards services.* Geneva: United Nations Publications.

vom Berge, P., M. König, and S. Seth (2013): "Sample of integrated labour market biographies (SIAB) 1975-2010," FDZ-Datenreport 01/2013.

von Wachter, T., and S. Bender (2006): "In the right place at the wrong time: The role of firms and luck in young workers' careers," *The American Economic Review*, 96(5), 1679–1705.

Wamser, G. (2011): "Foreign (in)direct investment and corporate taxation," *Canadian Journal of Economics*, 44(4), 1497–1524.

Weichenrieder, A. (2009): "Profit shifting in the EU: Evidence from Germany," *International Tax and Public Finance*, 16(3), 281–297.

Weichenrieder, A. J., and J. Mintz (2010): *The indirect side of direct investment: Multinational company finance and taxation.* Cambridge: The MIT Press.

Wooldridge, J. D. (2002): *Econometric analysis of cross section and panel data.* Cambridge: The MIT Press.

Yeaple, S. R. (2003): "The role of skill endowments in the structure of US outward foreign direct investment," *The Review of Economics and Statistics*, 85(3), 726–734.

——— (2009): "Firm heterogeneity and the structure of U.S. multinational activity," *Journal of International Economics*, 78(2), 206–215.

——— (2013): "Scale, scope, and the international expansion strategies of multiproduct firms," *mimeo*, Pennsylvania State University.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Datum:

Unterschrift: