RIETI Discussion Paper Series 19-E-053

Learning and Information Transmission within Multinational Corporations (Revised)

CHEN, Cheng
Clemson University

SUN, Chang
University of Hong Kong

ZHANG, Hongyong
RIETI

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 19-E-053
First draft: July 2019
Revised: September 2020

Learning and Information Transmission within Multinational Corporations[*]

Cheng CHEN   Chang SUN   Hongyong ZHANG[†]

Abstract

We propose that multinational firms learn about their profitability in a particular market by observing their performance in nearby markets. We first develop a model of firm expectations formation with noisy signals from multiple markets and derive predictions on expectations formation and market entries. Using a dataset of Japanese multinational corporations that includes sales expectations of each affiliate, we provide evidence supporting the model's predictions. We find that a positive signal about demand inferred from nearby markets raises the probability of entry into a new market, or raises the firm's sales expectation in an existing (focal) market. The latter effect is stronger when (1) the firm is less experienced in the focal market (2) the signals from the focal market are noisier and (3) the firm is more experienced in markets where signals are extracted.

Keywords: multinational production, learning, expectation formation, information transmission
JEL classification: F1; F2; D83

[*] This research is conducted as a part of project "Studies on the Impact of Uncertainty and Structural Change in Overseas Markets on Japanese Firms" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study utilizes the data of the questionnaire information based on "the Basic Survey on Overseas Business Activities" which is conducted by the Ministry of Economy, Trade and Industry (METI). We would like to thank Scott Baier, Taiji Furusawa, Jason Garred, Jung Hur, Toshiyuki Matsuura, Eduardo Morales, Masayuki Morikawa, Hiromichi Ozeki, Heiwai Tang, Eiichi Tomiura, Sumiko Urabe, Daniel Xu, Makoto Yano and Maurizio Zanardi for helpful comments. Financial support from HKGRF (project codes: 17500618, 17507916, 27502318), JSPS KAKENHI (grant numbers: 17H02531, 17H02554) and RIETI is greatly appreciated.

[†] Chen: Clemson University, chencheng1983613@gmail.com. Sun: the University of Hong Kong, sunc@hku.hk. Zhang: RIETI, zhang-hongyong@rieti.go.jp.

1 Introduction

Firms face substantial uncertainty when doing businesses, and this is particularly true for multinational corporations (MNCs) that produce and sell in multiple locations. Before a MNC enters a particular market, it may have limited information about consumers' tastes and the costs of production there. Given the large sunk entry costs, MNCs' entry decisions into new markets can be costly when information is imperfect. Naturally, a MNC may learn about a new market from its experience in nearby markets in order to alleviate the information problem. In this paper, we propose "cross-market" learning as a mechanism to resolve MNCs' uncertainty about new markets and implement a systematical empirical study of such behavior using a dataset of Japanese MNCs. In this dataset, we observe not only MNCs' entries into each market but also their affiliates' sales expectations post entry. Thanks to this feature, we are able to show that MNCs learn about the destination markets not only from their' local experience, but also from their experience in nearby markets. In short, we find that multinational production (MP) generates *information value* and information generated by MP is *transmitted* within the firm boundary. As how the agent forms expectations is a central component of learning models, our study provides new and direct evidence on learning and information transmission across geographic locations within the firm boundary.[1]

In order to guide our empirical analysis, we first build a model in which a firm learns about its demand conditions in multiple markets for three purposes.[2] First, it informs us of ways to estimate key parameters of the model, which we use to validate model assumptions and further test model predictions. Second, it generates testable predictions regarding market entry and expectations formation. Importantly, the model shows that theoretical predictions regarding how the learning parameters affect a firm's *probability of entry* into a new market are ambiguous, while the model has unambiguous predictions on how the firm forms its sales expectations.[3] Therefore, it is crucial to use

[1]This is true for a large class of models of learning in the international context. See Akhmetova and Mitaritonna (2013); Aeberhardt et al. (2014); Egger et al. (2014); Timoshenko (2015a,b); Conconi et al. (2016); Cebreros (2016); Berman et al. (2017).

[2]We focus on learning about demand in our model. However, one can also recast our model and interpret it as one in which firms learn about location-specific supply conditions.

[3]One example is that our learning model has an ambiguous prediction on how the precision of signals affects the probability of market entry (into the destination market), although it unambiguously predicts that the weight of these signals used in the expectations formation formula increases with the their precision.

direct measures of firm-level expectations to test predictions of the learning model.

In the model, the firm's demand shifter in a particular market and period is the sum of a time-invariant component and a transitory shock. The firm does not know the exact value of the time-invariant component, but has to infer it based on its prior and observed signals (demand shifters) in the past. Without loss of generality, we assume that the firm operates in two other markets besides the focal market. One is close to the focal market, and its time-invariant component is positively correlated with that in the focal market. The positive correlation can be caused by similar consumer preferences over the characteristics of the products or services that the firm provides. The third market is remote from the first two markets, so its time-invariant demand component is assumed to be uncorrelated with those in the first two.[4]

Several key testable predictions emerge from the model. First, the firm uses the average signal from the nearby market to forecast its expected profit in a new market and ignores signals from the remote market. Thanks to the positive correlation in the time-invariant demand, information from the nearby market is transmitted within the MNC and used to inform its decision to enter the new market. Moreover, although better signals from nearby markets increase the entry probability (into the new market), how the precision of such signals affects this positive effect depends on distributional assumptions and accordingly are ambiguous.[5] Therefore, we focus on testing predictions of comparative statics exercises regarding firms' expectations formation, as the model yields unambiguous predictions along this dimension.

After the firm enters the new market, it continues to update the expectation of future sales given the signals observed, which include the signals from the new market now. Thanks to the positive correlation in the time-invariant demand, the model predicts that the sales expectation depends on the signals from both this new market and the nearby market. Importantly, the model also yields other testable predictions related to the key mechanism of the life-cycle learning model (e.g., Jovanovic (1982), Jovanovic and Nyarko (1997)) extended to incorporate information transmission within the firm. We show that the firm's expectation in the new market relies more on the average signal

[4] These assumptions are motivated by our empirical finding that only past sales in the same industry and region can predict entry and sales expectation in a particular market. In Online Appendix 1.5, we show the model predictions are robust even if we allow the demand in the third market to have a positive but weaker correlation with those in the first two markets.

[5] Online Appendix 1.4 discusses this point in details. We show that this is the case even if we assume a log normal distribution for the entry cost. Despite this theoretical ambiguity, previous works rely heavily on the entry margin to establish the existence of firm learning.

from the nearby market and less on the signal from the new market when (1) the firm is less experienced in the new market, (2) the signals from the new market are noisier (with higher variance of the transitory shocks), and (3) the firm is more experienced in the nearby market. The intuition is that signals from the nearby market are more precise relative to those from the new market under these conditions.

Using a 22-year panel dataset of Japanese MNCs that contains affiliate-level sales expectations, we provide empirical evidence for the theoretical predictions derived in our model. First, we show that the strong average past sales (average "signal") of affiliates in markets within the same region (referred to as "nearby siblings") raises the probability of entry into a new market.[6] By contrast, the average past signal of affiliates outside the region (referred to as "remote siblings") has a weak and statistically insignificant impact on entry. Our baseline estimate suggests that a one-standard-deviation increase in the average nearby siblings' signal leads to an increase in entry probability by 0.028%, which is about 25% of the average entry rate.

Similar to the "extended gravity" documented by Morales et al. (2019), the cross-market learning highlighted here indicates interdependence across markets when firms make market entry decisions. However, there is an important distinction between cross-market learning and "extended gravity", which we show empirically. In Morales et al. (2019), "extended gravity" implies that a firm's presence in a nearby market *unambiguously* increases its entry probability into a new market.[7] In contrast, we find that the sign of such an impact depends crucially on the performance of the existing affiliates in the nearby markets. In particular, firms with prior presence in nearby markets actually have *lower* probabilities of entering a new market compared to those without, if their existing affiliates' signals are sufficiently bad. This fact distinguishes the learning mechanism from mechanisms that lead to the "extended gravity". It also distinguishes the transmission of information about market profitability from transferring other intangibles such as production or management know-how inside the firm: having more information does not necessarily imply more entries and higher sales. It has a positive impact on these variables, only when the signal is good enough.[8]

[6]To ensure the information spillover within MNCs concerns an individual firm's demand or supply conditions, we use average past sales net of aggregate components, taking out the destination-industry-year fixed effects.

[7]In their structural model, Morales et al. (2019) assume a firm's presence in a nearby market may reduce the entry cost into new markets.

[8]Note that, in our framework as well as many others featuring imperfect information, having more (precise) information is always beneficial to the firm, as the firm's expected profit increases with the precision of the information, taking entry costs into account.

Second, we explore our measure of affiliates' sales expectations after market entry and provide empirical support for additional theoretical predictions. We start by showing that the sales expectations in our dataset are reliable and contain relevant information that is used in actual firm-level decisions. We then proceed to regressions and find that the strong average signal of nearby siblings raises the expectation for the next year's sales, while the average signal of remote siblings has no significant impact. The elasticity of sales expectations with respect to the strength of nearby siblings' signal is 0.024.

The average effect of nearby siblings' signal on sales expectations hides rich underlying heterogeneity. Following the model's predictions, we further examine how market and affiliate characteristics affect the strength of the learning effect. We find the elasticity of expected sales with respect to the nearby siblings' signal is larger if the affiliate in the focal market is younger and/or the siblings in the nearby markets are older. We also construct two model-consistent measures of signal noisiness and show that the learning effect from the nearby markets is stronger if the signals are noisier in the focal market. These findings are consistent with the model predictions. One may worry that the average learning effect regarding entry and expectations formation is driven by correlated shocks within the firm and across markets, despite that we control for market-year and firm (or firm-year) fixed effects in all our regressions. However, we find it difficult to rationalize the heterogeneous learning effects using an explanation based on correlated shocks. Therefore, we argue that the learning mechanism is present, even if the temporary demand/supply shocks are indeed correlated across markets.

Our study contributes to three strands of the literature. First, our paper is related to the literature on the flow of intangibles within the firm boundary. Using the commodity flow data of the U.S., Atalay et al. (2014) find that vertical ownership is *not* primarily used to facilitate transfers of goods. Instead, they argue that the flow of intangibles is a crucial factor for us to understand intra-firm relationships. Echoing their finding, Ramondo et al. (2016) document a similar pattern for U.S. MNCs. Several papers have investigated various channels through which intangibles are transferred within the firm boundary (Keller and Yeaple, 2013; Fan, 2017; Bilir and Morales, 2018). Using the same data of U.S. MNCs, Bilir and Morales (2018) find that headquarters' innovations increase affiliate performance, although affiliates' innovations do not affect performance at other firm sites. We complement this literature by substantiating the existence of information sharing within the firm boundary and across geographic locations.

Second, our study contributes to the literature on MNCs' location choices.[9] We document that multinational affiliates in different markets are linked via the information channel. In a recent study, Garetto et al. (2019) provide evidence that, for U.S. MNCs, their presence in a country only has a slightly positive and sometimes insignificant effect on subsequent entries into similar countries. As we discussed above, we find that the MNC's prior presence in a market has a positive impact on subsequent entries in similar markets only when it receives a sufficiently good signal from the existing market. Therefore, the lack of the "extended gravity" in the entry patterns of U.S. MNCs does not necessarily imply the MNC makes entry decisions in different markets independently. There may be heterogeneous effects of current presence on subsequent entries, depending on the MNC's performance in existing markets.

Finally, our study is related to a growing literature on learning and exporter/MNC dynamics. Existing studies have documented the role of learning and self-discovery in exporter dynamics, as well as the inter-market linkages through information acquisition or sunk cost reduction.[10] Studies of MNC dynamics are relatively scant, with the exceptions of recent works such as Garetto et al. (2019), Gumpert et al. (2016), Bilir and Morales (2018), and Chen et al. (2020). Egger et al. (2014) show that the dynamic entry patterns of German MNCs are consistent with a two-period model featuring cross-market learning. We complement the existing work by showing that cross-market learning not only exists prior to entry, but also after market entry. Moreover, the coefficients that we estimate map directly to the expectations formation formula under the assumption of Bayesian updating. As a result, we can quantify how the firm absorbs new information in forming its belief, which is crucial if one wants to quantify how the arrival of new information affects market entries and firm growth.[11] Our empirical approach of detecting firm learning can be extended further in future research, as firm-level expectations data (concerning firm-specific variables) are becoming increasingly available.[12]

[9]For example, see Egger et al. (2014); Tintelnot (2017); Wang (2017); Arkolakis et al. (2018a); Alviarez (2019); and Head and Mayer (2019).

[10]See Akhmetova and Mitaritonna (2013); Aeberhardt et al. (2014); Timoshenko (2015a,b); Cebreros (2016); Berman et al. (2017) for studies on exporter self-discovery. Albornoz et al. (2012) examine cross-market learning among exporters, and Morales et al. (2019) use a novel moment inequality approach to quantify reductions in entry costs into a new market if the firm has already exported to similar markets. Fernandes and Tang (2014) document between-firm information spillovers in the export market, which we do not study in this paper.

[11]In Appendix A.5, we perform a simple calibration and show that the coefficients in the expectations formation formula implied by the model are in line with those estimated from the data.

[12]Papers that use firm-level expectations include Gennaioli et al. (2016), Bloom et al. (2017), and

The remainder of the paper is organized as follows. In the next section, we present a learning model, from which we derive testable predictions concerning market entries and expectations formation. In Section 3, we describe our data and construct key variables used in our empirical analysis. We test empirical predictions of the model in Section 4 and conclude in Section 5. We present additional empirical analysis in the paper appendix, and relegate proofs, extensions of the model and various robustness checks to the online appendix.

2 Model

In this section, we develop a simple model of firm learning that features both self-discovery in a particular market (Jovanovic, 1982; Arkolakis et al., 2018b) and learning about the focal market from other markets (Albornoz et al., 2012). As the firm's information on market-level demand conditions is imperfect, the firm has to form an expectation of these conditions in the destination market both before and after market entry. Before entering the foreign market, the firm learns its demand conditions in the destination market *imperfectly* from the performance of its affiliates in nearby markets. After observing the performance of nearby siblings, the firm decides whether to enter the destination market and is more likely to enter when its nearby affiliates have better past sales performance.

The key feature of our model rests on the expectations formation after market entry. If the firm enters the foreign market, its affiliate in that market updates its expectation of demand conditions over the life-cycle. Different from previous studies (e.g., Timoshenko (2015b), Berman et al. (2017)), we allow the affiliate to learn its demand conditions both from its own performance (i.e., average past sales) and from the performance of its nearby siblings.

2.1 Setup

We study a partial equilibrium model with a single firm. Suppose there are three foreign markets: markets 1 and 2 are in the same region, and market 3 is in another region. We focus on the firm's expectation in market 1, and refer to markets 2 and 3 as the

Altig et al. (2019) for American firms, Bachmann et al. (2013), Bachmann and Elstner (2015), and Enders et al. (2019) for German firms, Boneva et al. (2018) for firms in the U.K., and Ma et al. (2019) for Italian firms.

"nearby" and "remote" markets, respectively. We first study the case in which the firm is considering entering market 1 and then the problem of expectations formation after it has entered market 1.

We assume that consumers in all foreign markets have CES preferences. The firm's demand function in market j is

$$q_{jt} = A_{jt} e^{a_{jt}} p_{jt}^{-\sigma}, \tag{1}$$

where t denotes time and σ is the elasticity of substitution. The variable A_{jt} is the aggregate demand shifter and a_{jt} is firm-specific demand in market j. For each market j, the firm faces demand uncertainty, which comes from demand shifter a_{jt}. We assume that a_{jt} is the sum of a time-invariant market-specific demand draw θ_j and a transitory shock ε_{jt}:

$$a_{jt} = \theta_j + \varepsilon_{jt},\ \varepsilon_{jt} \overset{i.i.d.}{\sim} N\left(0, \sigma_{\varepsilon j}^2\right). \tag{2}$$

The firm understands that θ_j is drawn from a normal distribution $N\left(\bar{\theta}_j, \sigma_{\theta j}^2\right)$, and the independent and identically distributed (i.i.d.) transitory shock, ε_{jt}, is drawn from another normal distribution $N\left(0, \sigma_{\varepsilon j}^2\right)$. On the supply side, we assume that to produce q units of output in market j, all firms have to employ one unit of labor at the wage rate w_{jt}.[13]

The timing of the model is stated as follows. After a firm enters market j, its affiliate in that market makes its output choice after observing the demand shifter, a_{jt}, in period t. As a result, realized sales are

$$R_{jt} = A_{jt} e^{a_{jt}} \left(\frac{\sigma w_{jt}}{\sigma - 1}\right)^{1-\sigma}. \tag{3}$$

The above equation shows that the logarithm of realized sales is the sum of a_{jt} and a term that only consists of aggregate variables. Therefore, we construct a measure of a_{jt} in our empirical analysis by taking out the market-year fixed effects in log sales.

Before the firm enters market 1, it forms an expectation of θ_1 based on the realized sales in the other markets. To enter, the firm has to pay a one-time entry cost F, where the cumulative distribution function of F is $G(\cdot)$. We use π_{1t} to denote the discounted

[13] We only allow firm heterogeneity on the demand side but not on the supply side for simplicity. One can allow firm heterogeneity on the supply side and reinterpret our model as MNCs learning about their productivity in different markets. Given that we do not separately observe prices and quantities, we do not attempt to distinguish between learning about demand and learning about supply.

future profit flows in market 1 and $G(\pi_{1t})$ is thus the probability of entering market 1, which increases in π_{1t}. After the firm enters market 1, the affiliate there forms an expectation of its sales in the next period using its own past sales as well as the those of its siblings.

The fundamental assumption of the model is that the firm does not know the value of θ_j and therefore has to form a belief about its distribution to make its entry decision. After entry, the firm updates its belief about θ_j over time. Naturally, the sources of information the firm uses to form its expectations in market j are the key predictions of the model. These are determined by the extent to which demand shocks θ_j are correlated across markets.

We introduce the interdependence of demand shocks across markets as follows. The variance-covariance matrix of the firm's demand draws in this three-country world is denoted as

$$\mathbf{V}\left(\begin{bmatrix}\theta_1\\ \theta_2\\ \theta_3\end{bmatrix}\right)=\begin{bmatrix}\sigma_{\theta 1}^2 & \sigma_{12}^2 & \sigma_{13}^2\\ \sigma_{12}^2 & \sigma_{\theta 2}^2 & \sigma_{23}^2\\ \sigma_{13}^2 & \sigma_{23}^2 & \sigma_{\theta 3}^2\end{bmatrix}.$$

We further define $\rho_{ij} \equiv \sigma_{ij}^2/\sigma_{\theta i}\sigma_{\theta j}$ as the correlation between θ_i and θ_j. We make the following assumption on these correlation coefficients:

Assumption 1 $\rho_{12} > \rho_{13} = \rho_{23} = 0$.

In Appendix A.2, we provide a model-consistent method of estimating within- and cross-region correlations in θ (i.e., ρ_{12} and ρ_{13}). Within-region correlation is always higher than cross-region correlation, but the latter is also positive. We assume that the cross-region correlation is zero in our model for simplicity. As shown in Online Appendix 1.5, our model predictions continue to hold even if we allow the cross-region correlation to be positive but smaller than the within-region correlation although the mathematical derivations are more complicated.

2.2 Determinants of Market Entry

According to the assumption of a random market entry cost, the probability of entering market 1 in period t is $G(\pi_{1t})$, where π_{1t} is the discounted expected profit from this market in all future periods and $G(\cdot)$ is the cumulative distribution function of the entry cost. To understand how siblings' signals affect the entry probability, we need to know how they affect π_{1t}. In particular, π_{1t} can be written as

$$\pi_{1t} = E_{t-1} \sum_{\tau=t}^{\infty} A_{1\tau} \left(\frac{\sigma w_{1t}}{\sigma - 1} \right)^{1-\sigma} \eta^{\tau-t} e^{a_{1\tau}}, \tag{4}$$

where the expectation is taken given the information up to period $t-1$ and η denotes the discount factor. Further assuming that the firm-specific demand draws are independent of the aggregate variables and taking into account the fact that $a_{1t} = \theta_1 + \varepsilon_{1t}$, where ε_{1t} is i.i.d. normal, we have

$$\pi_{1t} = e^{\sigma_{\varepsilon 1}^2/2} E_{t-1} \left(e^{\theta_1} \right) \times E_{t-1} \sum_{\tau=t}^{\infty} A_{1\tau} \left(\frac{\sigma w_{1t}}{\sigma - 1} \right)^{1-\sigma} \eta^{\tau-t}.$$

Therefore, it is sufficient to examine how $E_{t-1} \left(e^{\theta_1} \right)$ responds to siblings' signals. Assuming that the sibling has received t_2 signals from market 2, we can prove the following proposition:

Proposition 1 *Under Assumption 1, the firm only uses signals from market 2 to forecast its "would-be" demand in market 1 and ignores signals from market 3. The firm's expected profit and entry probability in market 1 increase with the average past signals in market 2,* $\bar{a}_2 \equiv \sum_{\tau=t-t_2}^{t-1} a_{2\tau}/t_2$.

Proof. See Online Appendix 1.2. ■

The intuition behind this result is that a firm's demand conditions across markets within the same region are correlated. Therefore, nearby siblings' past sales contain information value, when the firm forecasts its demand in the market that it may enter in the future. Naturally, when the forecast is above a certain threshold, the MNC chooses to enter market 1.

In the next subsection, we will examine how various parameters such as t_2 affect the expectations formation post entry. However, how t_2 affects the positive effect of an increase in $\bar{a}_2$ on the entry probability depends on distributional assumptions and accordingly are ambiguous. This is true, even if we assume a log normal distribution for the entry cost.[14] On the contrary, we will show that our learning model has *unambiguous* predictions regarding how the firm forms sales expectations over its life cycle (post entry). Therefore, we argue that the best way to provide evidence on learning over the

[14]We prove that the cross derivative of the entry probability with respect to t_2 and $\bar{a}_2$ is ambiguous. See Online Appendix 1.4 for details.

life cycle is to derive and test theoretical predictions regarding expectations formation directly.

2.3 Expectations Formation after Market Entry

After the firm enters market 1, it continues to update its belief for θ_1. Now the firm can use signals from both markets 1 and 2 to update its posterior. The following proposition characterizes the firm's (or equivalently, the affiliate's) forecasting rule for its sales in market 1.

Proposition 2 *Under Assumption 1, an affiliate in market 1 uses its own average past signal and the average past signal of its siblings in market 2 to form its expectation of future sales, with positive weights put on both average signals. All else equal, the weights it places on the average signals of itself and its nearby siblings have the following properties:*

1. *The weight it places on the average signal of itself (its nearby sibling) increases (decreases) with self age.*
2. *The weight it places on the average signal of itself (its nearby sibling) decreases (increases) with the standard deviation of the transitory shocks in its market.*
3. *The weight it places on the average signal of itself (its nearby sibling) decreases (increases) with the total number of signals in market 2.*

Why do diverging age profiles for the two weights show up in the expectations formation formula? When the number of signals from market i ($i \in 1, 2$) increases (while fixing the number of signals from the other market), the precision of signals increases both in absolute terms and in relative terms (compared with the signals from the other market). As a result, the affiliate's expectation of sales in market i relies more on signals from market i. On the contrary, the precision of signals from the other market stays unchanged in absolute terms and decreases in relative terms (compared with the signals from market i) when the number of signals from market i increases. This results in the affiliate placing a lower weight on the signals from the other market in the expectations formation process. Similarly, when the affiliate's own signal becomes less precise, its forecast depends more on nearby siblings' signals and less on its own signals, all other things being equal.

It is worth discussing how the results would change if we allow the signals from market 3 to be informative as well. In Online Appendix 1.5, we derive model predictions under a weaker assumption $\rho_{12} > \rho_{13} = \rho_{23} \geq 0$. In this more general setting, we find that the average past signal from market 3 is also used to predict the would-be profit before the firm enters market 1 and to predict future sales thereafter. However, when ρ_{12} is sufficiently larger than ρ_{13} and ρ_{23}, the firm places higher weights on the signals from market 2 than those from market 3 when forming its expectations. Finally, we also derive the effects of the other model parameters on learning as in Proposition 2. We can thus show that all the results hold under the weaker assumption.

3 Data

In this section, we describe our data and discuss how we construct the key variables in our main empirical specifications. Given our emphasis on the direct measure of affiliate-level sales expectations, we also devote a subsection to discuss the credibility of this measure.

3.1 Basic Description and the Definition of Markets

We draw our data from the Basic Survey on Overseas Business Activities (Kaigai Jigyo Katsudo Kihon Chosa) conducted by the Ministry of Economy, Trade and Industry (METI) of the Japanese government ("the survey" hereafter). This survey is mandatory and conducted annually via self-declaration survey forms (one for the parent firm and another for each foreign affiliate) sent to the parent firm at the end of each fiscal year. The survey form for parent firms includes variables on the firm's domestic sales, employment, industry classifications, and so on, while the survey for foreign affiliates collects information on their sales, employment, location, and industry.

Based on the annual survey, we construct a panel dataset of parent–affiliate pairs from 1995 to 2016 that includes both manufacturing and non-manufacturing firms. Each parent–affiliate pair is traced throughout the period using time-consistent identification codes. Compared with other standard multinational datasets such as the U.S. BEA survey, our data is novel in that it contains information on affiliate-level expectations. Specifically, the affiliates of Japanese MNCs are asked to report their forecasted sales for the next year. This enables us to provide evidence of learning that directly uses affiliate-level expectations. Since this measure is rare in firm-level datasets, we examine its credibility in Section 3.3 and Appendix A.1.

In our data, affiliates are classified into 29 industries, including 16 manufacturing and 9 services sectors. In terms of the total number of affiliates abroad, wholesale and retail is the largest industry in services, and "transportation equipment" is the largest in manufacturing. Regarding geographic distribution, Table 1 shows the number of firms with presence in the most popular destinations in 2016, after dropping affiliates in tax haven countries listed in Gravelle (2009). China and the United States are the largest markets for Japanese multinationals. Interestingly, for firms that operate in two destinations the top combination is China-Thailand, which may be seen as suggestive evidence that geographic closeness between host countries is important for understanding multinational location choices. In Section 4, we examine the dynamic patterns of entry and the impact of siblings' signals formally.

Table 1: Most Popular Destinations in 2016

Destination	# Firms	Destinations	# Firms	Destinations	# Firms
CHN	2784	CHN-THA	230	CHN-THA-USA	73
USA	679	CHN-USA	179	CHN-IDN-THA	38
THA	526	CHN-VNM	115	CHN-KOR-TWN	24
VNM	258	CHN-TWN	86	CHN-THA-VNM	24
TWN	183	CHN-KOR	83	CHN-DEU-USA	21

Notes: The table shows the most popular destinations or destination combinations for firms operate in one, two and three destinations. Destination abbreviations: CHN (China), USA (the United States), THA (Thailand), VNM (Vietnam), TWN (Taiwan), KOR (South Korea), IDN (Indonesia), DEU (Germany).

We define markets at the destination-industry level. For a (potential) market of a Japanese firm, we define "nearby" and "remote" markets by first grouping all destinations into seven geographic regions: North America, Latin America, Asia (excluding the Middle East), the Middle East, Europe, Oceania, and Africa. A nearby market is a destination-industry pair that satisfies two conditions: (1) the destination is in the same region as the focal market and (2) the two markets belong to the same industry.[15] Similarly, a "remote" market is in the same industry but located in a different region. "Nearby" and "remote" siblings are existing affiliates of the same firm in nearby and remote markets, respectively. Consistent with our model setup, we require the firm to

[15] We focus on within-industry learning for two reasons. First, firms in our sample do not typically set up foreign affiliates in multiple industries (average number of industries is 1.6). Second, as we show in Section 4.1, signals from different industries do not significantly affect entry probabilities.

have at least one nearby and one remote siblings.

We focus on horizontal FDI by defining an entry only when a firm first sets up an affiliate in a market and this affiliate has high local sales shares. We calculate the local-to-total-sales ratio for each affiliate-year and use the average ratio of each affiliate over time to determine the nature of FDI. In our baseline regressions, we define affiliates to be "horizontal" only when this ratio is above 85%. We define entry when a firm sets up its first horizontal affiliate in the destination market. Our main empirical results are robust to more strict definitions of horizontal FDI, as discussed in Online Appendix 2.3.

3.2 Construction of Siblings' Signals

Consistent with our theory, we focus on firm learning about their idiosyncratic demand/supply conditions in particular markets. We therefore tease out the aggregate components in affiliates' performance by regressing the affiliate's log local sales on the destination–industry-year fixed effects. Suppose we denote the log local sales of affiliate i in year t as r_{it}; we then run the following regression:

$$r_{it} = \hat{\delta}_{skt} + \tilde{r}_{it}, \tag{5}$$

where $\hat{\delta}_{skt}$ denotes the estimated destination-industry-year fixed effects. Destinations and industries are denoted by k and s, respectively. We use the residual from this regression (denoted as $\tilde{r}_{it}$) as a measure of the affiliate's exceptional performance relative to its peers in the same market. Similarly, we project firms' domestic sales on the domestic-industry-year fixed effects and use the residual sales as a control for productivity shocks common across all affiliates of the same firm.

We are now ready to define the two key regressors in our empirical analyses. The model in Section 2 suggests that firms infer their market-specific demand using all past signals. Therefore, we construct the cumulative average of the past sales of existing affiliates as follows:

$$\overline{r_{fskt}^{\text{nearby}}} \equiv \frac{1}{N(\tau \leq t, i \in I_{fsk})} \sum_{\tau \leq t, i \in I_{fsk}} \tilde{r}_{i\tau}, \qquad \overline{r_{fskt}^{\text{remote}}} \equiv \frac{1}{N(\tau \leq t, i \in I_{fsk}^c)} \sum_{\tau \leq t, i \in I_{fsk}^c} \tilde{r}_{i\tau}, \tag{6}$$

where I_{fsk} denotes the set of firm f's affiliates in industry s and in other countries that are in the same region as destination k. The set I_{fsk}^c includes the affiliates of the same firm in industry s but in other regions. The $N(\cdot)$ function denotes the number of signals observed until time t. In the following analysis, we refer to $\overline{r_{fskt}^{\text{nearby}}}$ and $\overline{r_{fskt}^{\text{remote}}}$ as nearby and remote siblings' signals. We use both variables as regressors in our main specifications, therefore requiring the observation to have at least one nearby and one remote siblings.

3.3 Validation of Affiliate-level Forecasts

One unique feature of our dataset is that each affiliate reports its expected sales for the next year, when it fills out the survey of the current year. As such information is rarely available in firm-level datasets, we discuss why this measure is reliable and contains useful information that matters for actual firm decisions. Additional statistics and results are presented in Appendix A.1.

First, in our sample, it is very rare for firms to use a naive rule to make their sales forecasts. For example, as is shown in Appendix A.1, only 1.59% of the observations have a forecast for year $t + 1$ that is exactly the same as sales in year t. Our main regression results are basically unchanged after dropping these observations (See Online Appendix 2.2). Second, we show that the sales forecasts have statistically significant and economically strong impacts on realized sales and employment in the future, even when we control for previous sales and employment. Finally, the MNC survey is mandated by METI under the Statistics Law, so the information in the survey is confidential and cannot be applied for purposes beyond the scope of the survey, such as tax collection. Firms therefore do not have incentives to misreport to avoid taxes or to manage stock market expectations.

4 Empirical Evidence

We now examine the empirical predictions of the model (Propositions 1 and 2).

4.1 Market Entries

In this subsection, we study how the past sales of existing affiliates affect the probability of Japanese firms' entering new markets in order to provide empirical evidence for Proposition 1. We first transform our affiliate-year-level dataset into a firm-market-year-level dataset, where a "market" refers to a destination-industry pair. In principle, each firm can enter a potential market in any year. We keep the market-year combinations in which the firm has not yet established any affiliates in that market (not restricted to horizontal affiliates) and study the probability of setting up a horizontal affiliate there in the next year. Since we include nearby and remote siblings' signals as regressors, our sample also requires the focal market to have at least one nearby and one remote siblings. For instance, suppose firm A has set up affiliates in industry s and regions r_1, r_2 and r_3. We consider firm A's entries into any of the remaining destination markets in these three regions (in industry s).[16] We do not consider its entries into other regions or industries since it does not have operations in those markets and thus does not receive signals. For new markets in r_1, signals from existing affiliates in r_1 are "nearby signals" while signals from r_2 and r_3 are "remote signals".

Table 2 shows the number of observations and next years' entries by year in the sample use in our baseline regressions. There are around 41,700 observations (firm-market combinations) in each year and 47 of them will see a new entry in the next year. The average entry rate is 0.11%.[17]

We now introduce our econometric specification. In particular, we run the following linear probability regression:[18]

$$\Pr(Enter_{fsk,t+1} = 1) = b_1 \overline{r_{fskt}^{\text{nearby}}} + b_2 \overline{r_{fskt}^{\text{remote}}} + b_3 \tilde{r}_{ft} + \delta_{skt} + \delta_f + \epsilon_{fk,t+1}, \tag{7}$$

where the dependent variable is a binary variable indicating whether firm f enters destination k and industry s in year $t+1$. The independent variables are nearby and remote siblings' signals up to year t. We also control for the firms' domestic performance, $\tilde{r}_{ft}$, which is the residual of log domestic sales after teasing out the domestic industry-year fixed effects. We control for various fixed effects in our regressions, such as market-year fixed effects (δ_{skt}) and firm fixed effects (δ_f). According to Proposition 1, we expect b_1 to be positive while b_2 to be zero. Under the less extreme assumption that cross-region correlation in time-invariant demand is positive but smaller than that within region,

[16] As discussed in Section 3.2, for a particular focal market, siblings are affiliates of the same firm operating in the same industry but different countries.

[17] The entry rate in 1995 is higher than those in the other years. Note that we define entry using the founding year of each affiliate reported in the survey instead of using their first appearance in the data, so the higher entry rate in 1995 is not an artifact. In Online Appendix 2.4, we show our main empirical results are robust if we exclude 1995 from our sample.

[18] We use the linear probability regression as our main specification since it allows us to control for a rich set of fixed effects. The results are similar, both qualitatively and quantitatively, to alternative non-linear specifications (e.g., the Cox regression model used by Conconi et al. (2016)). We report these results in Appendix A.3.

Table 2: Number of observations and entries by year

Year	(1) # of obs.	(2) # of next year's entries	(3) entry rate (%)
1995	21919	68	0.31
1996	27100	50	0.18
1997	27736	46	0.17
1998	30772	42	0.14
1999	36426	39	0.11
2000	36101	50	0.14
2001	33531	48	0.14
2002	38926	42	0.11
2003	40303	58	0.14
2004	41905	49	0.12
2005	44611	44	0.10
2006	44952	53	0.12
2007	44289	41	0.09
2008	46096	36	0.08
2009	47753	48	0.10
2010	47073	72	0.15
2011	48704	78	0.16
2012	51978	47	0.09
2013	54473	32	0.06
2014	54338	23	0.04
2015	56541	11	0.02
Total	875527	977	0.11

Notes: Column 1 shows the number of observations by year in our baseline regression. Column 2 shows the number of the next year's entries among the observations in Column 1. Column 3 calculates the entry rates (Column 2/Column 1).

we expect b_2 to be positive but smaller than b_1.

Before we show the regression results, Table 3 presents the summary statistics of the key regressors and related variables in the same sample as in Table 2. The median observation has one nearby sibling and two remote siblings, and the average number of siblings (1.7 and 3.7) is larger than the median, suggesting that their distributions are right-skewed. Although many firms entered new destinations during our sample period, they established operations in developed regions (e.g., North America and Europe) long time ago. This is reflected by the average age of nearby and remote siblings, with medians of 13.8 and 15.5, respectively. Finally, there is substantial variability in the siblings' signals. For example, the 75th percentile of nearby siblings' signal is 197 log points higher than the 25th percentile, which translates into a 618% difference in past sales. The three regressors (nearby siblings' signal, remote siblings' signal, and residual parent sales) are also far from perfectly correlated. The correlation coefficients between any two of these variables are between 0.38 and 0.45.

Table 4 reports the estimation results of equation (7). In Column 1, we estimate the equation controlling for the destination-year and industry-year fixed effects but not the firm fixed effects. Both nearby siblings' signal and firms' domestic sales raise

Table 3: Summary statistics of siblings and parents

	Obs.	mean	std. dev.	25 pct.	median	75 pct.
Number of nearby siblings	875,527	1.677	2.247	1	1	2
Average age of nearby siblings	875,478	15.38	9.410	8.750	13.83	20.20
Average nearby signal	875,527	-0.258	1.585	-1.182	-0.157	0.789
Number of remote siblings	875,527	3.695	5.364	1	2	4
Average age of remote siblings	875,510	16.46	7.905	11	15.50	20.91
Average remote signal	875,527	0.0321	1.432	-0.784	0.0824	0.960
Residual sales of parents	875,527	-0.226	1.794	-1.388	-0.122	1.046

Notes: Nearby siblings are affiliates of the same firm in the same region and industry but a different destination. Remote siblings are affiliates of the same firm in the same industry but other regions. We calculate the signals as the cumulative average residual sales following the definition in equation (6).

the probability of FDI entry in the next period. A one standard deviation increase in nearby siblings' signal raises the entry probability by $1.59 \times 0.0174\% = 0.028\%$, which is around 25% of the average entry probability (0.11%). By contrast, remote siblings' signal does not have a significant impact on the probability of FDI entry. In Column 2, we further control for firm fixed effects to tease out time-invariant firm characteristics. Column 3 shows that the results are robust when we drop firms' domestic sales but control for firm-year fixed effects.

Table 4: Impact of siblings' experience on entry in the next period

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 100$	(1)	(2)	(3)
Average nearby signal	0.0174^{a}	0.0180^{a}	0.0172^{a}
	(0.00318)	(0.00377)	(0.00403)
Average remote signal	0.00414	0.00418	0.00179
	(0.00405)	(0.00536)	(0.00566)
Firm domestic sales	0.00660^{c}	-0.0142	
	(0.00350)	(0.0108)	
Destination-Ind-Year FE	✓	✓	✓
Firm FE		✓	
Firm-Year FE			✓
N	875527	875527	902527
R^2	0.064	0.067	0.088
# of Firms	1922	1922	1931
# of Firm-Markets	113998	113998	115183
# of Entries	977	977	1003

Notes: The dependent variable indicates whether the firm enters a particular destination in the next year. We calculate the signals as the cumulative average residual sales following the definition in equation (6). Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

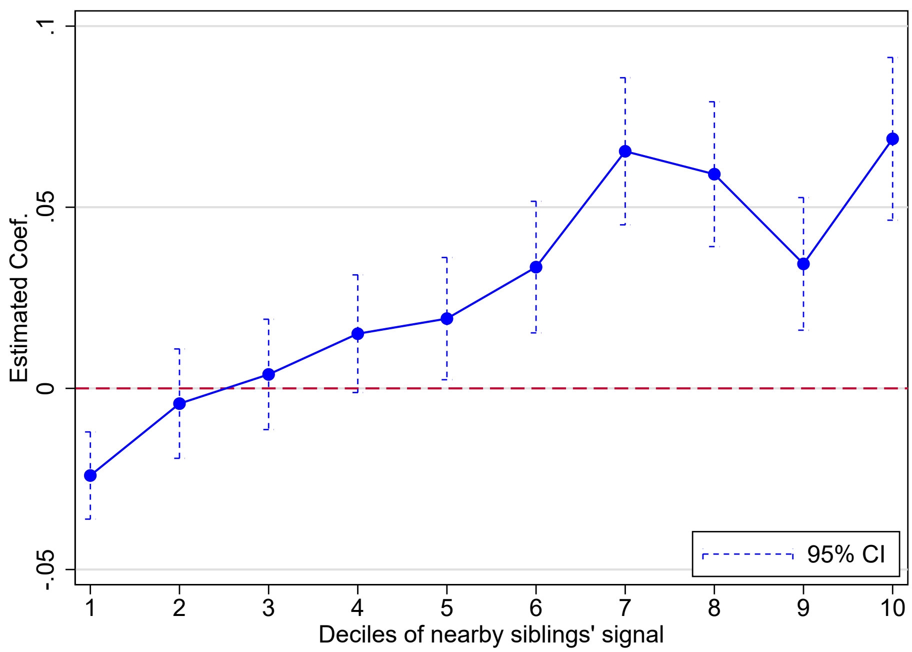
It is important to note that our evidence does not imply that the existence of a nearby sibling necessarily increases the likelihood of entry into other countries in the same region. Such a positive impact is realized only when the nearby siblings' signal is good enough. To demonstrate this point, we expand our sample to include regions which the firm has not entered yet. We then estimate the impact on different deciles of average nearby siblings' signal on the probability of market entry, using the observations without any nearby siblings as the base category, and controlling firm domestic performance, and destination-industry-year and firm-industry fixed effects.[19] We then plot the coefficients of decile dummies in Figure 1. Consistent with our earlier evidence, stronger siblings' signal raises the entry probability. However, compared to firms without any siblings, having a sibling only significantly raises the probability of entry when the siblings' signal is above the fourth decile. When the siblings' signal is in the lowest decile, the entry probability is actually significantly lower than that of a firm without any presence in the region.

We think that this result demonstrates an important distinction between our learning mechanism and other mechanisms that lead to sequential entries in similar markets. For example, Morales et al. (2019) construct and estimate an empirical model where an exporter's prior entry in nearby markets lowers the sunk entry costs into new markets, which can explain the "extended gravity" patterns in market entry. Their mechanism may well exist in our multinational firm data, as the presence of nearby siblings starts to show a positive impact on subsequent entries into new markets when the siblings' signal is as low as the third decile. However, this is not the case for the lowest two deciles. In a recent study, Garetto et al. (2019) provide evidence that the presence of a U.S. MNC in a country only has a slightly positive and sometimes insignificant effect on the probability of its entry into another similar country. We conjecture that the effects of prior presence of subsequent entries may well depend on the historical performance of the existing affiliates.

We have so far defined markets at destination-industry levels. In Table 5, we perform horse race regressions and show that signals from other industries cannot predict market entry, even if those signals come from the same region. In Columns 1-2, we regress the entry dummy on the average signals of siblings in the same region and industry and of siblings in the same region but different industries. We see that only the signal of siblings in the same region and industry has predictive power for the next period's entry. In Columns 3-4, we add remote sibling signals, and further separate remote sibling signals into those in the same industry and those in different industries. The results suggest

[19] The details of the sample and regression results are presented in Appendix A.4. We also report the results controlling for firm-industry-year fixed effects instead of firm-industry fixed effects. The results are similar.

Figure 1: Impact of Nearby Sibling Signal Deciles on Entry Probability (%)



Notes: Firm-country-industry cells without any nearby siblings are the base category (horizontal line at $y = 0$). The details of the sample and the regression results are presented in Table A.5.

that only signals from siblings in the same region and industry can predict market entry. Therefore, learning effect is the strongest for this type of signals.[20]

Table 5: Impact of siblings' experience on entry in the next period, horse race between signals from the same and different industries

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
Avg nearby signal (same ind)	0.0166^a	0.0157^a	0.0167^a	0.0168^a	0.0154^a	0.0114^b
	(0.00393)	(0.00397)	(0.00370)	(0.00414)	(0.00450)	(0.00507)
Avg nearby signal (diff ind)	0.000588	-0.00188	0.000871	-0.000621	-0.00204	-0.00554
	(0.00495)	(0.00717)	(0.00528)	(0.00736)	(0.00630)	(0.00835)
Avg remote signal			0.000420	0.0136		
			(0.00908)	(0.0170)		
Avg remote signal (same ind)					0.00447	0.00213
					(0.00615)	(0.00698)
Avg remote signal (diff ind)					-0.00142	-0.00107
					(0.00991)	(0.0131)
Firm domestic sales	-0.000523	0.00250	-0.000637	0.00290	0.00481	0.00866
	(0.00567)	(0.0139)	(0.00651)	(0.0145)	(0.00890)	(0.0208)
Destination-Year FE	✓	✓	✓	✓	✓	✓
Industry-Year FE	✓	✓	✓	✓	✓	✓
Firm FE		✓		✓		✓
N	458137	458136	447263	447261	319763	319763
R^2	0.014	0.018	0.014	0.017	0.019	0.022
# of Firms	1039	1038	973	971	489	489
# of Firm-Markets	58663	58662	57375	57373	42952	42952
# of Entries	571	571	551	551	406	406

Notes: The dependent variable indicates whether the firm enters a particular destination in the next year. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

4.2 Expectations Formation after Market Entries

In this subsection, we use our measure of affiliates' sales forecasts to study how past signals affect the formation of expectations and to test Proposition 2. The baseline regression specification is as follows:

$$\log E_t(R_{i,t+1}) = b_1\overline{r_{it}} + b_2\overline{r^{\text{nearby}}_{fskt}} + b_3\overline{r^{\text{remote}}_{fskt}} + b_4\tilde{r}_{ft} + \delta_{skt} + \delta_f + \epsilon_{i,t+1}, \tag{8}$$

where we examine how the affiliate's own signal and its siblings' signals affect its expected sales in the next year. The right hand of equation (8) is almost the same as that of equation (7), except for the addition of the first regressor, $\overline{r_{it}}$. This variable is

[20]A caveat is that, as we add more signals into the horse race regressions, the number of observations shrinks. For example, Columns 1-2 in Table 5 requires that, for the focal market, the firm has at least one sibling in the same region-industry and one sibling in the same region but different industry. Columns 3-4 require an additional sibling in the other regions, whether in the same industry or not, while the last two columns further require one sibling in the same industry but different region and one sibling in a different industry and different region.

a measure of the affiliate's own signal, which is defined as the cumulative average of its residual log local sales $\tilde{r}_{i\tau}, \tau \leq t$. Proposition 2 predicts that both b_1 and b_2 are positive.

Table 6: Impact of siblings' signal on expected sales in the next year, baseline and by age group

Dep. Var: $\log E_t(R_{i,t+1})$ Sample:	(1) all ages	(2) $1 \leq$ age ≤ 3	(3) $4 \leq$ age ≤ 6	(4) age ≥ 7
Average self signal	0.823^a	0.550^a	0.805^a	0.935^a
	(0.0111)	(0.0242)	(0.0263)	(0.00912)
Average nearby signal	0.0244^b	0.0982^a	0.0282	0.0218^c
	(0.0112)	(0.0358)	(0.0259)	(0.0126)
Average remote signal	0.0140	0.00800	0.00641	0.0198
	(0.0171)	(0.0573)	(0.0459)	(0.0187)
Firm domestic sales	0.0524^a	0.0878^c	0.107^a	0.0537^b
	(0.0191)	(0.0498)	(0.0308)	(0.0216)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
N	32881	2182	3778	24160
R^2	0.878	0.882	0.894	0.900
# of Firms	989	386	511	858
# of Affiliates	7152	1385	2033	5265

Notes: Dependent variable is the logarithm of expected sales in the next year. We calculate the signals as the cumulative average residual sales following the definition in equation (6). Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10. The number of observations in Columns 2–4 does not add up to that in Column 1 because we have excluded the singletons (observations whose variation is completely absorbed by the fixed effects) when calculating these numbers, and the set of singletons depends on the subsample.

Column 1 of Table 6 presents the results from the baseline regression. The affiliate's own signal is a key determinant of future sales expectation, with a precisely estimated coefficient of 0.823. Nearby siblings' signals also positively affect expectations. If the average past sales of all nearby siblings increase by one log point, the affiliate's expected sales increase by 0.024 log points. By contrast, remote siblings' signals have a positive but insignificant impact, which is consistent with the evidence we presented for market entries in the previous subsection. We next explore the heterogeneous effects of the nearby siblings' signals and test the additional predictions in Proposition 2.

4.2.1 Siblings' Signal Matters More when the Affiliate is Younger

We now show that the impact of nearby siblings' signals on sales expectations is stronger when the affiliate considered is younger. In Columns 2 to 4 of Table 6, we divide the sample into affiliates of different ages. We find that the impact of nearby siblings' signal is higher for younger affiliates, whereas the impact of self-experience is higher for older affiliates. When affiliates are no older than three years, the coefficient of nearby siblings' signal is four times the average effect in Column 1, while the coefficient of the affiliate's own signal is one third smaller. When affiliates are older, the coefficients of the average nearby siblings' signal are much smaller and becomes insignificant or marginally significant. In Appendix A.5, we perform a simple calibration of the model and show that the magnitude of these coefficients is in line with those implied by the model.

Table 7: Impact of siblings' signal (interacted with affiliate age) on expected sales in the next year

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.868^a	0.596^a	0.868^a	0.587^a
	(0.00935)	(0.0194)	(0.00982)	(0.0203)
$\times$ log(self age)	0.0907^a		0.0917^a	
	(0.00575)		(0.00602)	
$\times \max\{\text{self age}, 10\}$		0.0331^a		0.0344^a
		(0.00212)		(0.00221)
Average nearby signal	0.0235^b	0.181^a	0.0342^a	0.200^a
	(0.0115)	(0.0262)	(0.0118)	(0.0285)
$\times$ log(self age)	-0.0503^a		-0.0522^a	
	(0.00850)		(0.00923)	
$\times \max\{\text{self age}, 10\}$		-0.0193^a		-0.0204^a
		(0.00282)		(0.00316)
Average remote signal	0.0179	0.0173	0.0168	0.0159
	(0.0175)	(0.0173)	(0.0250)	(0.0254)
Firm domestic sales	0.0534^a	0.0561^a		
	(0.0188)	(0.0183)		
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓	✓		
Firm-Year FE			✓	✓
Age FE	✓	✓	✓	✓
N	32872	32872	31724	31724
R^2	0.885	0.885	0.905	0.905

Notes: Dependent variable is the logarithm of expected sales in the next year. We calculate the signals as the cumulative average residual sales following the definition in equation (6). Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

To confirm the increasing (declining) impact of the affiliate's own (nearby siblings') signal on the expectations formation, we interact these two signals with affiliate age in Table 7. Since some affiliates in our data are old, we create two age measures to capture the non-linear effects of age: the logarithm of affiliate age and affiliates' age capped at 10. We further control for the direct impact of age on expected sales using the affiliate age fixed effects. The logarithm of affiliate age is also standardized to facilitate the interpretation of the coefficients. Taking the estimates in Column 1 as an example, we find that a one standard deviation increase in the log affiliate age raises the impact of the affiliate's own signals by 0.091 and reduces the impact of nearby siblings' signals by

0.050. In Columns 3 and 4, we replace firms' domestic sales and the firm fixed effects with the firm-year fixed effects and the patterns are similar.

4.2.2 Siblings' Signal Matters More in Markets with Noisier Signals

In this subsection, we explore how the relationship between the affiliate's expectation and its nearby siblings' signal varies with proxies of the signal noisiness in the affiliate's market. We first construct two measures of $\sigma_{\varepsilon 1}$ that are consistent with our model. First, log sales in our model are proportional to $\theta + \varepsilon_t$. Hence, subtracting log sales in period $t-1$ from that in period t can remove the time-invariant component θ. The variance of the log sales growth rates in the focal market is thus proportional to $2\sigma^2_{\varepsilon 1}$. Second, sufficiently old affiliates have almost discovered θ, meaning that the only source of their forecast errors is the temporary shock ε_t. Table 6 suggests that learning from siblings is very weak after seven years in the market. We therefore use the standard deviation of forecast errors of affiliates with at least seven years of experience as a proxy for $\sigma^2_{\varepsilon 1}$.

We perform the following regression to examine the impact of $\sigma_{\varepsilon 1}$:

$$\begin{aligned} \log E_t(R_{i,t+1}) \quad = \quad & b_1\overline{r_{it}} + b_2\overline{r^{\text{nearby}}_{fskt}} + b_3\overline{r^{\text{remote}}_{fskt}} + b_4\tilde{r}_{ft} \\ & b_5\overline{r_{it}} \times \hat{\sigma}_{\varepsilon 1,k} + b_6\overline{r^{\text{nearby}}_{fskt}} \times \hat{\sigma}_{\varepsilon 1,k} + \delta_{skt} + \delta_f + \epsilon_{i,t+1}. \end{aligned} \tag{9}$$

Our new estimation equation is equation (8) with the addition of two new terms: the interaction terms between signal noisiness in destination k and the signals of the affiliate and of its nearby siblings. The destination-level signal noisiness measure, $\hat{\sigma}_{\varepsilon 1,k}$, is defined as the standard deviation of the log sales growth of all the Japanese affiliates in destination k, or the standard deviation of the sales forecast errors of affiliates at least seven years old. To ensure these measures are precise, we only include countries that have at least 20 observations of sales growth or forecast errors.[21] Proposition 2 predicts that b_5 is negative while b_6 is positive.

Table 8 reports the regression results. In Columns 1 and 2, we approximate $\hat{\sigma}_{\varepsilon,k}$ using the standard deviation of the sales growth rates in destination k, which are further standardized to facilitate the interpretation of the coefficients. Column 2 replaces the firms' domestic sales control and firm fixed effects in Column 1 with firm-year fixed

[21]Ideally, one would want to calculate a proxy $\sigma_{\varepsilon 1}$ at the destination-industry level because it is our definition of a "market". However, this causes more measurement errors in $\sigma_{\varepsilon 1}$ since we have fewer observations in each cell. We decide to aggregate the sales growth rates at the destination level instead.

effects. The results in these two columns show that b_5 is negative, while b_6 is positive, which confirms the model's prediction. As shown in Column 1, a one standard deviation increase in $\hat{\sigma}_{\varepsilon 1,k}$ lowers the coefficient of the affiliate's own signals by 0.056 and raises the coefficient of nearby siblings' signals by 0.035.

We experiment with alternative measures of $\hat{\sigma}_{\varepsilon 1,k}$ in the other columns of Table 8. Columns 3 and 4 construct this measure using the standard deviation of the forecast errors for affiliates above seven years, as discussed above. The signs of the two interaction terms are the same, but the magnitude of the coefficients falls. Finally, since sales growth rates and forecast errors may be affected by aggregate shocks (e.g., the affiliates in a destination tend to overpredict their sales before a recession hits), we also use measures of residual sales growth and residual forecast errors by removing the destination-industry-year fixed effects before calculating the standard deviation. The results shown in Columns 5–8 are almost the same as those in Columns 1–4.

Table 8: Effect of market noisiness on learning

Proxy constructed using	Sales Growth		Fore. Err.		Res. Sales Growth		Res. Fore. Err.	
Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average self signal	0.844^a	0.847^a	0.838^a	0.840^a	0.844^a	0.847^a	0.838^a	0.840^a
	(0.0115)	(0.0122)	(0.0115)	(0.0121)	(0.0115)	(0.0122)	(0.0115)	(0.0121)
× proxy of $\sigma_{\varepsilon 1}$	-0.0560^a	-0.0537^a	-0.0286^a	-0.0284^a	-0.0560^a	-0.0537^a	-0.0286^a	-0.0284^a
	(0.00802)	(0.00822)	(0.00719)	(0.00741)	(0.00802)	(0.00822)	(0.00718)	(0.00741)
Average nearby signal	0.0277^b	0.0391^a	0.0262^b	0.0360^a	0.0277^b	0.0391^a	0.0262^b	0.0359^a
	(0.0113)	(0.0114)	(0.0111)	(0.0112)	(0.0113)	(0.0114)	(0.0111)	(0.0112)
× proxy of $\sigma_{\varepsilon 1}$	0.0350^a	0.0371^a	0.0220^a	0.0194^a	0.0350^a	0.0371^a	0.0220^a	0.0194^a
	(0.00879)	(0.00867)	(0.00676)	(0.00670)	(0.00879)	(0.00867)	(0.00675)	(0.00670)
Average remote signal	0.0181	0.0196	0.0159	0.0173	0.0181	0.0196	0.0160	0.0173
	(0.0171)	(0.0240)	(0.0174)	(0.0249)	(0.0171)	(0.0240)	(0.0174)	(0.0249)
Firm domestic sales	0.0544^a		0.0537^a		0.0544^a		0.0536^a	
	(0.0185)		(0.0191)		(0.0185)		(0.0191)	
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓		✓		✓		✓	
Firm-Year FE		✓		✓		✓		✓
Age FE	✓	✓	✓	✓	✓	✓	✓	✓
N	32858	31708	32841	31691	32858	31708	32839	31688
R^2	0.882	0.902	0.881	0.901	0.882	0.902	0.881	0.901

Notes: Dependent variable is the logarithm of expected sales in the next year. We calculate the signals as the cumulative average residual sales following the definition in equation (6). Market noisiness $\sigma_{\varepsilon 1}$ is proxied for using the standard deviation of (residual) sales growth rates or (residual) forecast errors, which are indicated in the column heads. The proxies are standardized. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

4.2.3 Siblings' Signal Matters More when Siblings are More Experienced

In this subsection, we examine how siblings' experience affects the strength of learning and test the last prediction of Proposition 2.

To perform the statistical test, we first need to construct measures of sibling experience. Since siblings' signal is calculated by aggregating all siblings' past sales in nearby markets, the correct notion of siblings' experience is the number of signals observed by the firm. However, since some siblings entered before 1995, the earliest year of our data, we cannot observe their performance before 1995 and cannot include them in the siblings' signal measure. We thus construct two variables to measure siblings' age. First, consistent with our notion of average past signals (residual log local sales), we calculate the number of signals used in this calculation, i.e., $N(\tau \leq t, i \in I_{fsk})$ in equation (6). Second, we calculate the sum of nearby siblings' ages, which assumes that the firm uses all past signals of the nearby siblings to forecast sales in the focal market. To capture the non-linear effect, we use the logarithms of both variables in our regressions; they are also standardized to facilitate interpretation.

We empirically test the model prediction using a variation of equation (8). We include interaction terms between nearby siblings' experience and the signals of the affiliate itself and of its nearby siblings. At the same time, we control for the first-order terms of nearby siblings' experience and the signals of the affiliate and of its nearby siblings. We also control for the interaction terms between the age of the affiliate (i.e., self experience) and the signals because the affiliate's age is positively correlated with its nearby sibling's age, whereas it has the opposite effect on learning, as shown in Section 4.2.1.

Table 9 reports the results. Although the interaction term between the nearby siblings' experience and the affiliate's own signal is significantly negative in only one specification (Column 1), the interaction term between nearby siblings' experience and their own signal is significantly positive in all specifications, suggesting that siblings' signals matters more if they are older. Depending on the specification, a one standard deviation increase in the nearby siblings' experience raises the coefficient of the nearby siblings' signal by around 50%. The estimated effects are similar regardless of whether siblings' experience is measured by the number of observed signals or total age. Finally, the coefficients of the interaction terms of the affiliate's age and the signals are similar to those in Table 7. The effect of siblings' experience on learning is in general smaller than that of the affiliate's own experience.

Table 9: Interaction of siblings' signal with siblings' experience

Sibling Experience Measure:	# Signals		Total Age	
Dep. Var: log $E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.869^a	0.870^a	0.868^a	0.869^a
	(0.00961)	(0.0100)	(0.00961)	(0.00999)
× Self experience	0.0917^a	0.0935^a	0.0915^a	0.0933^a
	(0.00582)	(0.00611)	(0.00582)	(0.00608)
× Nearby siblings' experience	-0.0110^c	-0.00474	-0.00379	-0.000190
	(0.00649)	(0.00682)	(0.00737)	(0.00767)
Average nearby signal	0.0394^a	0.0601^a	0.0331^b	0.0507^a
	(0.0146)	(0.0161)	(0.0135)	(0.0142)
× Self experience	-0.0514^a	-0.0530^a	-0.0518^a	-0.0534^a
	(0.00864)	(0.00937)	(0.00869)	(0.00944)
× Nearby siblings' experience	0.0209^b	0.0322^a	0.0160^b	0.0251^a
	(0.00853)	(0.0109)	(0.00815)	(0.00972)
Nearby siblings' experience	0.0229	0.000337	0.0228	0.0150
	(0.0177)	(0.0219)	(0.0167)	(0.0203)
Average remote signal	0.0188	0.0204	0.0185	0.0187
	(0.0178)	(0.0252)	(0.0180)	(0.0260)
Firm domestic sales	0.0523^a		0.0527^a	
	(0.0187)		(0.0187)	
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓		✓	
Firm-Year FE		✓		✓
Age FE	✓	✓	✓	✓
N	32872	31724	32862	31714
R^2	0.886	0.905	0.885	0.905

Notes: Dependent variable is the logarithm of expected sales in the next year. We calculate the signals as the cumulative average residual sales following the definition in equation (6). Self experience is the log of self age, while the nearby siblings' experience is measured by the log of total number of signals or total age of the nearby siblings, indicated by the column head. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

4.3 Robustness Checks

We discuss two robustness checks of our main empirical results in this subsection and refer the reader to the online appendix for detailed regression tables.

First, in our expectations formation regressions, we considered several factors that may affect the weights that affiliates place on the signals of itself and of its nearby siblings. These factors may be correlated with each other and/or correlated with other confounding variables. In Table OA.1, we rerun the regressions including the full set of factors considered above and obtain similar results as before. This suggests that affiliate age, market noisiness, and siblings' experience all have separate effects on learning as predicted by the model. We also show that our results are robust to adding the interaction of signals and focal market income levels.

The second challenge to our empirical analysis is the presence of regional value chains. We know from earlier work that Japanese firms may have established regional value chains, especially in Asia (Hayakawa and Matsuura, 2011). For example, if the Thai affiliate of a Japanese firm produces electronic components that are both sold in Thailand and exported to its Chinese affiliate for final assembly, supply shocks to the Thai affiliate can cause positive correlations in the local sales in Thailand and the expected sales of the Chinese affiliate. To address this concern, we perform our baseline entry and expectations formation regressions restricting our sample to new entrants that have a small regional or global import shares. As reported in Online Appendix 2.3, this aggressive strategy reduces our sample size by about one quarter, and reduces the impact of the nearby siblings' signal on entry by one third but does not change its impact on expectation formation. Therefore, our main findings are not simply driven by regional or global value chains.

5 Conclusion

In this study, we use a novel dataset of Japanese MNCs to provide evidence that MNCs learn about profitability in the destination market by observing the performance of their affiliates in similar markets. Specifically, the strong past sales of siblings in nearby markets raises the probability of the firm entering a particular market. In addition, after market entry, the strong sales performance of siblings in nearby markets also raises the expectation of future sales held by the affiliate in the focal market. Importantly, such an impact declines over the affiliate's life-cycle, while self-discovery becomes more important as the affiliate ages. We also show that the effect of learning from nearby siblings is stronger if the destination market's signals are noisier and when siblings are more experienced. We view these findings as evidence of cross-market learning and information transmission within MNCs. The simple model we provide here rationalizes all the empirical findings and is thus a good starting point for studying MNC dynamics and interdependence across markets.

There are at least three fruitful avenues for future research. First, constructing a structural model would be useful to estimate the structural parameters of the model (e.g., correlations of the time-invariant demand across markets, variances of the time-invariant demand and the transitory shock) and conduct counterfactual analysis. Second, incorporating information transmission within MNCs into a quantitative MP framework (e.g., Helpman et al. (2004) and Ramondo and Rodríguez-Clare (2013)) would help quantify the role of learning within MNCs in determining their entry and production patterns. Finally, the current study does not consider information spillovers across MNCs, which may also influence their activities abroad and have strong policy implications.[22] We leave these promising approaches and interesting questions to future research.

[22]See, for example, Fernandes and Tang (2014), Kamal and Sundaram (2016), Hamilton (2018) for evidence in the context of exporting.

References

Aeberhardt, Romain, Ines Buono, and Harald Fadinger, "Learning, incomplete contracts and export dynamics: Theory and evidence from French firms," *European Economic Review*, May 2014, *68*, 219–249.

Akhmetova, Zhanar and Cristina Mitaritonna, "A Model of Firm Experimentation under Demand Uncertainty: an Application to Multi-Destination Exporters," Working Paper, CEPII research center April 2013.

Albornoz, Facundo, Héctor F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas, "Sequential exporting," *Journal of International Economics*, 2012, *88* (1), 17–31.

Altig, David, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Brent H Meyer, and Nicholas Parker, "Surveying business uncertainty," Technical Report, National Bureau of Economic Research 2019.

Alviarez, Vanessa, "Multinational production and comparative advantage," *Journal of International Economics*, July 2019, *119*, 1–54.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple, "Innovation and Production in the Global Economy," *American Economic Review*, August 2018, *108* (8), 2128–2173.

__ , Theodore Papageorgiou, and Olga A. Timoshenko, "Firm learning and growth," *Review of Economic Dynamics*, 2018, *27*, 146–168.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson, "Vertical Integration and Input Flows," *American Economic Review*, April 2014, *104* (4), 1120–1148.

Bachmann, Rüdiger and Steffen Elstner, "Firm Optimism and Pessimism," *European Economic Review*, October 2015, *79*, 297–325.

__ , __ , and Eric R Sims, "Uncertainty and Economic Activity: Evidence from Business Survey Data," *American Economic Journal: Macroeconomics*, 2013, *5* (2), 217–249.

Berman, Nicolas, Vincent Rebeyrol, and Vincent Vicard, "Demand learning and firm dynamics: evidence from exporters," Working Paper 2017.

Bilir, L. Kamran and Eduardo Morales, "Innovation in the Global Firm," Working Paper 2018.

Bloom, Nicholas, Steven J. Davis, Lucia Foster, Brian Lucking, Scott Ohlmacher, and Itay Saporta-Eksten, "Business-Level Expectations and Uncertainty," SSRN Scholarly Paper 3085377, Social Science Research Network, Rochester, NY December 2017.

Boneva, Lena, James Cloyne, Martin Weale, and Tomasz Wieladek, "Firms' Price , Cost and Activity Expectations: Evidence from Micro Data," Working Paper 2018.

Cebreros, Alfonso, "The Rewards of Self-discovery: Learning and Firm Exporter Dynamics," Working Paper, Banco de México 2016.

Chen, Cheng, Tatsuro Senga, Chang Sun, and Hongyong Zhang, "Uncertainty, Imperfect Information and Learning in the International Market," Working Paper 2020.

Conconi, Paola, André Sapir, and Maurizio Zanardi, "The internationalization process of firms: From exports to FDI," *Journal of International Economics*, March 2016, *99*, 16–30.

Egger, Peter, Matthias Fahn, Valeria Merlo, and Georg Wamser, "On the genesis of multinational foreign affiliate networks," *European Economic Review*, January 2014, *65*, 136–163.

Enders, Zeno, Franziska Hünnekes, and Gernot J Müller, "Firm expectations and economic activity," CESifo Working Paper 2019. tex.publisher: CESifo Working Paper.

Fan, Jingting, "Talent, Geography, and Offshore R&D," Working Paper 2017.

Fernandes, Ana P. and Heiwai Tang, "Learning to export from neighbors," *Journal of International Economics*, September 2014, *94* (1), 67–84.

Garetto, Stefania, Lindsay Oldenski, and Natalia Ramondo, "Multinational Expansion in Time and Space," Working Paper 25804, National Bureau of Economic Research May 2019.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer, "Expectations and investment," *NBER Macroeconomics Annual*, 2016, *30* (1), 379–431.

Gravelle, Jane G., "Tax Havens: International Tax Avoidance and Evasion," *National Tax Journal*, 2009, *62* (4), 727–753.

Gumpert, Anna, Andreas Moxnes, Natalia Ramondo, and Felix Tintelnot, "Multinational Firms and Export Dynamics," Working Paper 2016.

Hamilton, Ben, "Learning, Externalities, and Export Dynamics," Working Paper 2018.

Hayakawa, Kazunobu and Toshiyuki Matsuura, "Complex vertical FDI and firm heterogeneity: Evidence from East Asia," *Journal of the Japanese and International Economies*, September 2011, *25* (3), 273–289.

Head, Keith and Thierry Mayer, "Brands in motion: How frictions shape multinational production," Working Paper 2019.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple, "Export versus FDI with Heterogeneous Firms," *The American Economic Review*, March 2004, *94* (1), 300–316.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 649–670.

— **and Yaw Nyarko**, "Stepping-stone mobility," *Carnegie-Rochester Conference Series on Public Policy*, June 1997, *46*, 289–325.

Kamal, Fariha and Asha Sundaram, "Buyer-seller relationships in international trade: Do your neighbors matter?," *Journal of International Economics*, September 2016, *102*, 128–140.

Keller, Wolfgang and Stephen Ross Yeaple, "The Gravity of Knowledge," *American Economic Review*, 2013, *103* (4), 1414–44.

Ma, Yueran, Tiziano Ropele, David Sraer, and David Thesmar, "A quantitative analysis of distortions in managerial forecasts," Working Paper 2019.

Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Extended Gravity," *The Review of Economic Studies*, November 2019, *86* (6), 2668–2712. Publisher: Oxford Academic.

Ramondo, Natalia and Andrés Rodríguez-Clare, "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, April 2013, *121* (2), 273–322.

— **, Veronica Rappoport, and Kim J. Ruhl**, "Intrafirm trade and vertical fragmentation in U.S. multinational corporations," *Journal of International Economics*, January 2016, *98*, 51–59.

Timoshenko, Olga A., "Learning versus sunk costs explanations of export persistence," *European Economic Review*, October 2015, *79*, 113–128.

— , "Product switching in a model of learning," *Journal of International Economics*, 2015, *95* (2), 233–249.

Tintelnot, Felix, "Global Production with Export Platforms," *The Quarterly Journal of Economics*, February 2017, *132* (1), 157–209.

Wang, Zi, "Headquarters Gravity," Working Paper 2017.

A Appendix

A.1 Validation of Affiliate-level Forecasts

In this section, we show that the variable of expected sales in the MNC survey is reliable and contains useful information that matters for actual firm decisions.

First, we show that it is very rare for firms to use a naive rule to make their sales forecasts. In Table A.1, we present the expected growth rates, which is calculated as the ratio of the affiliate's forecast for year $t+1$ to its realized sales in year t minus one. If an affiliate simply uses its realized sales in year t to predict their sales next year, the expected growth rate will be zero. As one can see from the table, only 1.59% of the observations in our sample have a zero expected growth rate. The frequency of the other top cases is all below 0.1%. For the affiliates reporting zero expected growth rates, it is difficult to tell whether they are making a naive forecast or making a serious forecast with the expectation that their sales growth will be very close to zero. We conduct robustness checks in Online Appendix 2.2 by dropping all observations with zero expected growth rates. Our main empirical results remain largely unchanged.

Table A.1: The Most Frequent Values of Expected Growth Rates

Top 1-5		Top 6-10	
$E_t(R_{t+1})/R_t - 1$	Freq. (%)	$E_t(R_{t+1})/R_t - 1$	Freq. (%)
0.0000	1.59	0.0417	0.06
0.1111	0.09	0.2000	0.06
0.2500	0.09	0.3333	0.05
0.1000	0.08	0.1250	0.05
0.0526	0.07	0.1429	0.05

Notes: This table shows the most frequent values of expected growth rates among all the affiliate-year observations that are in our baseline regressions using the variable of sales expectations (Column 1 of Table 6 in the paper). Total number of observations is 29,958. It is smaller than that in our baseline regression because some affiliates do not report their current sales. Our data contains more observations than those in our baseline regressions since our regressions only include affiliates with at least one nearby and one remote siblings. However, if we compute the expected growth rates over all the observations in the dataset, the results are similar. They are available upon request.

Second, we show that the sales forecasts have statistically significant and economically strong impacts on realized sales and employment in the future. Specifically, we regress the realized sales in year $t+1$ on the sales forecast made in year t and a set of fixed effects, and the results are reported in Table A.2. The first three columns show that the sales forecast in year t positively and significantly predicts the realized sales in year $t+1$. Importantly, the effect of the sales forecast is not eliminated after we include the realized sales in year t as a control in Column 2. Its coefficients are actually much larger than the realized sales. Further including the realized sales in year $t-1$ does not change this pattern (Column 3). Columns 4-6 show that the sales forecasts also have strong predicative power for future employment, even if we control for the current and past employment. These findings easily reject the hypothesis that firms fill out this survey question with random guesses. By contrast, they indicate that firms take these forecasts seriously and that the forecasts contain more information about the affiliates' future than typical observables such as past sales and employment.

Table A.2: Sales Forecasts Predict Affiliates' Future Outcomes

Dep. Var.	log total sales $\log(R_{i,t+1})$			log employment $\log(L_{i,t+1})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\log E_t(R_{i,t+1})$	0.619^a	0.526^a	0.517^a	0.291^a	0.130^a	0.135^a
	(0.0294)	(0.0315)	(0.0346)	(0.0199)	(0.0135)	(0.0144)
$\log R_{it}$		0.121^a	0.121^a			
		(0.0189)	(0.0364)			
$\log R_{i,t-1}$			0.0380^a			
			(0.0113)			
$\log L_{it}$					0.514^a	0.505^a
					(0.0233)	(0.0302)
$\log L_{i,t-1}$						0.0194
						(0.0208)
Affiliate FE	✓	✓	✓	✓	✓	✓
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓
N	26040	26040	22726	25935	25863	22962
# of Firms (cluster)	782	782	706	784	784	711
Within R-squared	0.443	0.456	0.446	0.156	0.372	0.364
R-squared	0.974	0.975	0.977	0.969	0.978	0.979

Notes: The dependent variable is affiliate i's log total sales or total employment in year $t+1$. We use R to denote sales and L to denote employment. $E_t(R_{i,t+1})$ refers to the affiliate's expectation in year t for its sales in year $t+1$. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10. We restrict our sample to those with at least one nearby and one remote siblings as in Column 1 of Table 6 in the paper. We have fewer observations here because we require a longer panel (at least two years for each affiliate). We also run the same regressions using all the observations in our dataset, and the results are similar. They are available upon request.

Finally, the MNC survey is mandated by METI under the Statistics Law, so the information in the survey cannot be applied for purposes beyond the scope of the survey, such as tax collection. Firms do not have incentives to misreport due to tax purposes. Moreover, unlike earnings forecasts announced by public firms, the sales forecasts reported to METI are confidential, so firms do not have the incentive to misreport strategically to manage the expectations of the stock market. In total, the empirical patterns described above assure us that the sales forecasts contained in the

MNC survey are reliable and suitable for our empirical analysis.

A.2 Within- and Cross-region Correlations in θ

In this section, we compare the within-region and cross-region correlations of time-invariant demand θ. To measure such correlations, we first try to extract model-consistent measures of θ from the data. According to the model, sufficiently old firms have almost learned the value of θ and the variability in their sales is only caused by ε. Therefore, if we average over a large number of realized (log) sales of old firms, we can obtain a proxy for θ. We perform this exercise for each parent-firm-market, only taking observations when the affiliate is at least seven years old. We then obtain a parent-firm-market-level dataset. We pair each market in which a parent firm has entered with all the other markets it has presence. For each pair of markets 1 and 2, we can calculate the correlation in θ_1 and θ_2 across all firms with presence in both markets. The correlation can be calculated for two markets within the same region or in different regions. Row 1 of Table A.3 shows the within-region and cross-region correlations, pooling all within-region pairs and cross-region pairs together, respectively. The within-region correlation is around 0.41, higher than the cross-region correlation.

One concern about this calculation is that the proxy for θ is contaminated by other factors such as aggregate shocks and firm-level global shocks that are not firm-market-specific. To address this issue, we compute two alternative proxies for θ. First, we remove the country-year and industry-year fixed effects from log sales, so that the residual $\hat{e}_1$(sales) is arguably idiosyncratic demand. We then calculate the average within a parent-firm-market for the affiliate that are at least seven years old. Second, we use a different residual $\hat{e}_2$(sales) obtained by regressing log sales on log parent firm domestic sales as well as the above fixed effects. This further removes the firm-level global shocks that are not firm-market-specific. We use this $\hat{e}_2$(sales) to construct a third proxy for θ. Rows 2 and 3 of Table A.3 show the correlations of θ constructed in these ways within and across regions, respectively. These correlations are smaller than that in row 1, whereas the within-region correlation is always larger than the cross-region correlation, and the differences are around 0.1.

Table A.3: Correlation of demand within and between regions for affiliates at all ages

Demand Measure	Corr. within Region	Corr. between Regions
log(sales)	0.414	0.315
	[13270]	[28639]
$\hat{e}_1$(sales)	0.379	0.298
	[12574]	[25704]
$\hat{e}_2$(sales)	0.328	0.230
	[13033]	[28062]

Notes: Each observation is an HQ-country-country pair (two different countries). For each HQ-country cell, we take the average of sales for all affiliates at least seven years old. When the demand measure is log(sales), we simply use the logarithm of local sales for each affiliate. When the demand measure is $\hat{e}_1$(sales), we regress log local sales on the country-year and industry-year fixed effects and use the residual to measure a firm's idiosyncratic demand. When the demand measure is $\hat{e}_2$(sales), we further control for parent sales in Japan beyond the fixed effects to obtain residual sales. All the correlation coefficients are significant at 1%.

A.3 Cox Regression Results for Market Entry

In the paper, we use a linear probability model to study how previous signals affect subsequent entries in nearby markets. Here we follow Conconi et al. (2016) to model the hazard ratio of firm f that enters destination k and industry s between time t and $t+1$ using the Cox regression model:

$$h_{fsk}(t|\mathbf{X}) = h_j(t)\exp\left(b_1 \overline{r_{fskt}^{\text{nearby}}} + b_2 \overline{r_{fskt}^{\text{remote}}} + b_3 \tilde{r}_{ft}\right), \tag{10}$$

where $h_j(t)$ is the hazard ratio for strata j and the terms in the exponential function are defined in the same way as in equation (7) in the paper. The key assumption of this model is that the regressors shift the hazard function $h_j(t)$ proportionally. The hazard functions within each stratum are allowed to differ and do not need to be estimated. We specify strata at different levels to check the robustness of the results.

Table A.4 shows the results from the Cox regression models, which are qualitatively similar to those from the linear probability model. When we set the strata at the market or market-year level, both the nearby siblings' signal and the firms' domestic sales have a positive impact on the hazard of FDI entry. According to the estimates in Column 1, a one standard deviation increase in the average nearby siblings' signal raises the hazard ratio by $e^{1.59\times 0.167} - 1 = 30\%$. Since the subject of the survival analysis is at the firm-market-year level, we cannot specify the strata at a level finer than the firm-market level. In Columns 3 and 4, we set the strata at the firm and firm-year levels, respectively and obtain slightly larger effects of the average nearby siblings' signal.

Table A.4: Impact of siblings' experience on entry in the next period (survival analysis)

	(1)	(2)	(3)	(4)
Average nearby signal	0.167^a	0.178^a	0.241^a	0.208^a
	(0.0274)	(0.0306)	(0.0425)	(0.0498)
Average remote signal	0.0494	0.0261	-0.0378	-0.0460
	(0.0309)	(0.0357)	(0.0588)	(0.0610)
Firm domestic sales	0.0624^c	0.0492		
	(0.0320)	(0.0341)		
N	881049	881049	907868	907868
# of Firms	1923	1923	1932	1932
# of Firm-Markets	114469	114469	115642	115642
# of Entries	1030	1030	1063	1063
Log likelihood	-3950.1	-2885.8	-4127.5	-3847.6
Strata	Destination-Ind	Destination-Ind-Year	Firm	Firm-Year

Notes: Results of the Cox regression models. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10. Note that the sample is exactly the same as those in Table 4 in the paper. The number of observations differs because we do not count singletons due to the fixed effects in the linear regressions.

A.4 The Impact of Nearby Sibling Signal Deciles on Entry Probability

In this section, we compare the entry probabilities among three types of firms for a given region r: (1) multinationals that have presence in the region and have received good signals, (2) multinationals that have presence in the region but have received bad signals, and (3) multinationals that have no existing affiliates in the region. Note that our baseline entry regression focuses on firms that already have presence in the region and excludes multinationals in group (3). To highlight the difference between firms with and without presence in the region, we expand our sample to include markets in regions where firms have no presence yet. We also focus on the impact of nearby siblings' presence/signals and do not require the firm to have established an affiliate in a remote market. This increases our sample size substantially.[23]

If nearby siblings exist, we calculate their signal and group them into ten equally sized bins (deciles one to ten). We assign the decile to be "zero" if no nearby sibling exists, and use this group as the base category. Therefore, when we run a linear probability model of entry on decile dummies, the coefficient indicates the difference in the entry probability between each decile and the observations with no nearby siblings. Besides the decile dummies, we also include destination-industry-year and firm-industry (or firm-industry-year) fixed effects. As Table A.5 shows, receiving signals in a higher decile tends to increase the entry probability, consistent with our findings in Table 4 in the paper. However, we find that the presence of nearby siblings significantly lowers the probability of entry, if the signal is sufficiently bad (in the lowest decile). We see this as a key distinction between the learning mechanism and other mechanisms that lead to sequential entries into similar markets.

A.5 Coefficients Estimated from the Data and Implied by the Model

In this section, we perform a simple calibration of our model and show that the weights that firms put on self and nearby siblings' signals predicted by the model are similar to

[23]For each firm, we only include industries in which they eventually enter in at least one destination. This is to make sure that the firm does have the technological capability of operating in these industries. We implicitly added the same restriction in our baseline regressions, since we require the firm to have at least one sibling in the same region and industry (i.e., the nearby sibling). However, we do not restrict the firm to have operations in a remote market in the current regression, as we are not doing a horse race between nearby siblings' and remote siblings' signals.

Table A.5: The impact of nearby siblings' signal on next period entry, using markets without nearby siblings as the base category

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 100$	(1)	(2)
Average nearby signal Q1	-0.0241^{a}	-0.0148^{b}
	(0.00614)	(0.00623)
Average nearby signal Q2	-0.00421	0.00626
	(0.00770)	(0.00810)
Average nearby signal Q3	0.00386	0.0123
	(0.00777)	(0.00779)
Average nearby signal Q4	0.0151^{c}	0.0272^{a}
	(0.00828)	(0.00861)
Average nearby signal Q5	0.0193^{b}	0.0295^{a}
	(0.00861)	(0.00883)
Average nearby signal Q6	0.0335^{a}	0.0432^{a}
	(0.00926)	(0.00929)
Average nearby signal Q7	0.0654^{a}	0.0758^{a}
	(0.0104)	(0.0107)
Average nearby signal Q8	0.0591^{a}	0.0713^{a}
	(0.0102)	(0.00999)
Average nearby signal Q9	0.0344^{a}	0.0503^{a}
	(0.00934)	(0.00968)
Average nearby signal Q10	0.0689^{a}	0.0851^{a}
	(0.0115)	(0.0121)
Firm domestic sales	-0.0000705	
	(0.000898)	
Destination-Ind-Year FE	✓	✓
Firm-Ind FE	✓	
Firm-Ind-Year FE		✓
N	13669307	13669307
R^2	0.016	0.025
# of Firms	8363	8363
# of Firm-Markets	1646318	1646318
# of Entries	2724	2724

Notes: Dependent variable is an indicator variable indicating whether the headquarters enters a particular destination next year. Standard errors are clustered at headquarters (HQ) level. Significance levels: a: 0.01, b: 0.05, c: 0.10. The number of observations is much larger than that in Table 4 of the paper because we include markets in regions where firms have no presence yet. These observations are used as the base category.

those estimated from the data.

In Online Appendix 1.3, we derive closed-form expressions for the coefficients of the average self and nearby siblings' signal in the expectation updating formula:

$$\beta_1 = \frac{(1-\rho_{12}^2)\lambda_2 + 1/t_2}{(1+1/\lambda_1 t_1)(\lambda_2+1/t_2) - \rho_{12}^2\lambda_2} \tag{11}$$

$$\beta_2 = \frac{\sigma_{\theta 1}}{\sigma_{\theta 2}} \frac{\rho_{12}/t_1}{(\lambda_1+1/t_1)(1+1/\lambda_2 t_2) - \rho_{12}^2\lambda_1}. \tag{12}$$

To gauge the values of β_1 and β_2, we first impose symmetry within a region so that markets 1 and 2 have the same σ_θ and σ_ε. In Chen et al. (2020), we provide estimates for these parameters which imply a signal-to-noise ratio of 1.86.[24] The average age of nearby siblings is 15 according to Table 3 in the paper. We estimated ρ_{12} to be 0.41, using the model-consistent approach discussed in Appendix A.2 (see the first row of Table A.3). We then plug $\lambda_1 = \lambda_2 = 1.86$, $t_2 = 15$ and $\rho_{12} = 0.41$ into equations (11) and (12). The implied coefficients under different values of t_1 are presented in Table A.6, quantitatively similar to those estimated in Table 6 of the paper. For example, the coefficient of average nearby siblings' signal is estimated to be 0.098 in the data, while the model implies this coefficient to be 0.155, 0.097 and 0.070 for age one, two and three affiliates, respectively.

Table A.6: Model-implied coefficients of average self and nearby siblings' signals

Self Age t_1	1	2	3	4	5	6	7	8	9	10
Coef. of Self Signal	0.608	0.756	0.823	0.861	0.886	0.903	0.916	0.925	0.933	0.939
Coef. of Nearby Signal	0.155	0.097	0.070	0.055	0.045	0.038	0.033	0.030	0.026	0.024

Notes: The coefficients are calculated according to equations (11) and (12), respectively. We choose the following parameter values in addition to t_1: $\sigma_{\theta 1} = \sigma_{\theta 2}$, $\rho_{12} = 0.41$, $\lambda_1 = \lambda_2 = 1.86$, $t_2 = 15$.

[24] The estimation relies on the result that the forecast errors of old firms are dominated by ε, while uncertainty about θ and ε drives the forecast errors of young firms together.

Online Appendix for "Learning and Information Transmission within Multinational Corporations"

Cheng Chen Chang Sun Hongyong Zhang*

July, 2020

*Chen: Clemson University, cchen9@clemson.edu. Sun: University of Hong Kong, sunc@hku.hk. Zhang: RIETI, zhang-hong-yong@rieti.go.jp.

1 Additional Theoretical Results

In this theory appendix, we first discuss the forecasting problem in the general case in which $\rho_{12} > \rho_{13} = \rho_{23} > 0$ and then prove Propositions 1 and 2 as a special case in which $\rho_{13} = \rho_{23} = 0$.

1.1 Expectation Formation in the General Case

Before we consider the expectation formation before and after entering market 1, we show that the average past signals in each market are sufficient statistics for the posterior distribution of θ_1. To see this, without loss of generality, suppose the firm has entered all three markets and observed signals $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$, where the bold letters represent the entire vector of the signals from a particular market. Using Bayes' rule and denoting the density functions with $f(\cdot)$, we have

$$
\begin{aligned}
f(\theta_1|\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3) &= \frac{f(\theta_1,\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3)}{f(\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3)} \propto f(\theta_1,\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3) \\
&= \int_{\theta_2,\theta_3} f(\theta_1,\theta_2,\theta_3,\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3) d\theta_2 d\theta_3 \\
&= \int_{\theta_2,\theta_3} f(\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3|\theta_1,\theta_2,\theta_3) f(\theta_1,\theta_2,\theta_3) d\theta_2 d\theta_3 \\
&= \int_{\theta_2,\theta_3} f(\theta_1,\theta_2,\theta_3) \prod_{i=1}^{3} f(\boldsymbol{a}_i|\theta_i) d\theta_2 d\theta_3 \qquad (1) \\
&= \int_{\theta_2,\theta_3} f(\theta_1,\theta_2,\theta_3) \prod_{i=1}^{3} \frac{f(\theta_i|\boldsymbol{a}_i) f(\boldsymbol{a}_i)}{f(\theta_i)} d\theta_2 d\theta_3 \qquad (2) \\
&\propto \int_{\theta_2,\theta_3} f(\theta_1,\theta_2,\theta_3) \prod_{i=1}^{3} \frac{f(\theta_i|\bar{a}_i) f(\bar{a}_i)}{f(\theta_i)} d\theta_2 d\theta_3 \qquad (3) \\
&= f(\theta_1,\bar{a}_1,\bar{a}_2,\bar{a}_3) \propto f(\theta_1|\bar{a}_1,\bar{a}_2,\bar{a}_3). \qquad (4)
\end{aligned}
$$

We have used the fact that conditional on θ_i, each element in $\boldsymbol{a}_i$ is independent to obtain step (1), applied Bayes' rule to obtain step (2), used the well-known result that $\bar{a}_i$ is a sufficient statistic if one wants to predict θ_i with $\boldsymbol{a}_i$ alone (e.g., Jovanovic (1982)) when deriving step (3), and finally obtained equation (4) by rolling back the derivations above (with $\bar{a}_i$ instead of $\boldsymbol{a}_i$). Therefore, we have simplified the problem: we just need to use the joint distribution of $\theta_1, \bar{a}_1, \bar{a}_2, \bar{a}_3$ to derive the posterior distribution of θ_1.

1.1.1 Before Entering Market 1

Before the firm enters market 1, it uses $\bar{a}_2$ and $\bar{a}_3$ to predict θ_1 given the joint normal distribution:

$$\begin{bmatrix} \theta_1 \\ \bar{a}_2 \\ \bar{a}_3 \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{\theta}_1 \\ \bar{\theta}_2 \\ \bar{\theta}_3 \end{bmatrix}, \begin{bmatrix} \sigma_{\theta 1}^2 & \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} & \rho_{13}\sigma_{\theta 1}\sigma_{\theta 3} \\ \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} & \sigma_{\theta 2}^2 + \sigma_{\varepsilon 2}^2/t_2 & \rho_{23}\sigma_{\theta 2}\sigma_{\theta 3} \\ \rho_{13}\sigma_{\theta 1}\sigma_{\theta 3} & \rho_{23}\sigma_{\theta 2}\sigma_{\theta 3} & \sigma_{\theta 3}^2 + \sigma_{\varepsilon 3}^2/t_3 \end{bmatrix} \right).$$

We denote the number of signals received in market j up to the current period as t_j, and the signal-to-noise ratio in market j as $\lambda_j \equiv \sigma_{\theta j}^2/\sigma_{\varepsilon j}^2$.

Using the formula of the conditional distribution under joint normal distributions, $\theta_1|\bar{a}_2, \bar{a}_3$ is distributed as normal with mean $\bar{\mu}$ and variance $\bar{\Sigma}$. One can obtain the conditional mean of θ_1

$$\bar{\mu} = \bar{\theta}_1 + \beta_2(\bar{a}_2 - \bar{\theta}_2) + \beta_3(\bar{a}_3 - \bar{\theta}_3),$$

where

$$\beta_2 = \frac{\sigma_{\theta 1}\sigma_{\theta 2}}{\sigma_{\varepsilon 2}^2} \frac{\rho_{12}(\lambda_3 + 1/t_3) - \rho_{13}\rho_{23}\lambda_3}{(\lambda_2 + 1/t_2)(\lambda_3 + 1/t_3) - \rho_{23}^2\lambda_2\lambda_3} \tag{5}$$

$$\beta_3 = \frac{\sigma_{\theta 1}\sigma_{\theta 3}}{\sigma_{\varepsilon 3}^2} \frac{\rho_{13}(\lambda_2 + 1/t_2) - \rho_{12}\rho_{23}\lambda_2}{(\lambda_2 + 1/t_2)(\lambda_3 + 1/t_3) - \rho_{23}^2\lambda_2\lambda_3}. \tag{6}$$

The conditional variance is

$$\bar{\Sigma} = \sigma_{\theta 1}^2 - \beta_2\sigma_{12}^2 - \beta_3\sigma_{13}^2 = \sigma_{\theta 1}^2 - \sigma_{\theta 1}^2 \frac{\rho_{12}^2\lambda_2(\lambda_3 + 1/t_3) - 2\rho_{12}\rho_{13}\rho_{23}\lambda_2\lambda_3 + \rho_{13}^2\lambda_3(\lambda_2 + 1/t_2)}{(\lambda_2 + 1/t_2)(\lambda_3 + 1/t_3) - \rho_{23}^2\lambda_2\lambda_3}.$$

1.1.2 After Entering Market 1

After the firm enters market 1, it uses all three average past signals $\bar{a}_1, \bar{a}_2, \bar{a}_3$ to form the posterior of θ_1. The joint distribution of $\theta_1, \bar{a}_1, \bar{a}_2, \bar{a}_3$ is

$$\begin{bmatrix} \theta_1 \\ \bar{a}_1 \\ \bar{a}_2 \\ \bar{a}_3 \end{bmatrix} \sim N \left(\begin{bmatrix} \bar{\theta}_1 \\ \bar{\theta}_1 \\ \bar{\theta}_2 \\ \bar{\theta}_3 \end{bmatrix}, \begin{bmatrix} \sigma_{\theta 1}^2 & \sigma_{\theta 1}^2 & \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} & \rho_{13}\sigma_{\theta 1}\sigma_{\theta 3} \\ \sigma_{\theta 1}^2 & \sigma_{\theta 1}^2 + \sigma_{\varepsilon 1}^2/t_1 & \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} & \rho_{13}\sigma_{\theta 1}\sigma_{\theta 3} \\ \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} & \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} & \sigma_{\theta 2}^2 + \sigma_{\varepsilon 2}^2/t_2 & \rho_{23}\sigma_{\theta 2}\sigma_{\theta 3} \\ \rho_{13}\sigma_{\theta 1}\sigma_{\theta 3} & \rho_{13}\sigma_{\theta 1}\sigma_{\theta 3} & \rho_{23}\sigma_{\theta 2}\sigma_{\theta 3} & \sigma_{\theta 3}^2 + \sigma_{\varepsilon 3}^2/t_3 \end{bmatrix} \right).$$

According to the formula of the conditional distribution of joint normal distributions, the conditional mean of θ_1 given $\bar{a}_1, \bar{a}_2, \bar{a}_3$ is

$$\bar{\mu} = \bar{\theta}_1 + \begin{bmatrix} \sigma_{\theta 1}^2 & \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} & \rho_{13}\sigma_{\theta 1}\sigma_{\theta 3} \end{bmatrix} A^{-1} \begin{bmatrix} \bar{a}_1 - \bar{\theta}_1 \\ \bar{a}_2 - \bar{\theta}_2 \\ \bar{a}_3 - \bar{\theta}_3, \end{bmatrix}$$

where A denotes the submatrix of the variance-covariance matrix after removing Row 1 and Column 1.

Therefore, the conditional mean of θ_1 is linear in $\bar{a}_i - \bar{\theta}_i$:

$$\bar{\mu} = \bar{\theta}_1 + \beta_1(\bar{a}_1 - \bar{\theta}_1) + \beta_2(\bar{a}_2 - \bar{\theta}_2) + \beta_3(\bar{a}_3 - \bar{\theta}_3),$$

where

$$
\beta_1 = \frac{\sigma_{\theta 1}^2 \sigma_{\varepsilon 2}^2 \sigma_{\varepsilon 3}^2 \begin{bmatrix} (\lambda_2 + 1/t_2)(\lambda_3 + 1/t_3) + 2\rho_{12}\rho_{13}\rho_{23}\lambda_2\lambda_3 \\ -\rho_{23}^2\lambda_2\lambda_3 - \rho_{12}^2\lambda_2(\lambda_3 + 1/t_3) - \rho_{13}^2\lambda_3(\lambda_2 + 1/t_2) \end{bmatrix}}{\Delta}, \tag{7}
$$

$$
\beta_2 = \frac{\sigma_{\theta 1}\sigma_{\theta 2}\sigma_{\varepsilon 1}^2 \sigma_{\varepsilon 3}^2 \left[\frac{\rho_{12}}{t_1}(\lambda_3 + 1/t_3) - \rho_{13}\rho_{23}\frac{\lambda_3}{t_1} \right]}{\Delta}, \tag{8}
$$

$$
\beta_3 = \frac{\sigma_{\theta 1}\sigma_{\theta 3}\sigma_{\varepsilon 1}^2 \sigma_{\varepsilon 2}^2 \left[\frac{\rho_{13}}{t_1}(\lambda_2 + 1/t_2) - \rho_{12}\rho_{23}\frac{\lambda_2}{t_1} \right]}{\Delta}, \tag{9}
$$

and Δ is the determinant of matrix A, which is positive. ($(\bar{a}_1, \bar{a}_2, \bar{a}_3)$ has a non-degenerate multivariate normal distribution, meaning that the covariance matrix must be positive-definite with a positive determinant.) The conditional variance of θ_1, $\bar{\Sigma}$, can be expressed as follows:

$$
\bar{\Sigma} = (1 - \beta_1)\sigma_{\theta 1}^2 - \beta_2\sigma_{12}^2 - \beta_3\sigma_{13}^2. \tag{10}
$$

1.2 Proof of Proposition 1

Proof. Under Assumption 1, we can simplify equations (5) and (6) as

$$
\beta_2 = \frac{\sigma_{\theta 1}\sigma_{\theta 2}}{\sigma_{\varepsilon 2}^2}\frac{\rho_{12}}{\lambda_2 + 1/t_2}, \quad \beta_3 = 0.
$$

Therefore, the firm only uses signals from market 2 to form its expectation of market 1.

Next, we study how the average signal from market 2 affects the entry probability. We can rewrite the conditional mean and variance of θ_1 as

$$
\bar{\mu} = \bar{\theta}_1 + \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}}\left(1 - \frac{1}{1 + \lambda_2 t_2}\right)(\bar{a}_2 - \bar{\theta}_2) \tag{11}
$$

and

$$\bar{\Sigma} = \sigma_{\theta 1}^2 - \sigma_{\theta 1}^2 \rho_{12}^2 \frac{\lambda_2 t_2}{1 + \lambda_2 t_2}. \tag{12}$$

The firm's probability of entering market 1 is $G(\pi_{1t})$ and

$$\frac{\partial G(\pi_{1t})}{\partial \bar{a}_2} = g(\pi_{1t}) B_t e^{\bar{\mu} + \frac{\bar{\Sigma}}{2}} \frac{\sigma_{\theta 1} \rho_{12}}{\sigma_{\theta 2}} \frac{\lambda_2 t_2}{1 + \lambda_2 t_2} > 0,$$

where

$$B_t \equiv e^{\sigma_{\varepsilon 1}^2/2} E_{t-1} \sum_{\tau=t}^{\infty} A_{1\tau} \left(\frac{\sigma w_{1t}}{\sigma - 1} \right)^{1-\sigma} \eta^{\tau - t}.$$

We can conclude that the entry probability increases with the average signal from market 2, $\bar{a}_2$. ■

1.3 Proof of Proposition 2

Proof. Recall that the firm's sales in market 1 can be expressed as

$$R_{1t} = A_{1t} e^{a_{1t}} \left(\frac{\sigma w_{1t}}{\sigma - 1} \right)^{1-\sigma}.$$

Here, we maintain the assumption that the aggregate variables A_{1t}, w_{1t} are independent of the demand draw θ_1. Therefore, we can write the expected sales as

$$E_{t-1}(R_t) = E_{t-1}(e^{a_{1t}}) e^{b_{t-1}},$$

where b_{t-1} is the log of $E_{t-1}\left(A_{1t}\left(\sigma w_{1t}/(\sigma - 1)\right)^{1-\sigma}\right)$. Since the posterior of a_{1t} is normal with mean $\bar{\mu}$ and variance $\bar{\Sigma} + \sigma_{\varepsilon 1}^2$ as discussed in Section 1.1.2, we have

$$\log E_{t-1}(R_t) = \bar{\mu} + \left(\bar{\Sigma} + \sigma_{\varepsilon 1}^2\right)/2.$$

In this expression, only the term $\bar{\mu}$ is affected by the signals. Therefore, to understand how the signals affect the log of expected revenue, it is sufficient to examine how they affect $\bar{\mu}$.

Under Assumption 1, we can simplify equations (7) to (8) as

$$\begin{aligned}
\beta_1 &= \frac{(1-\rho_{12}^2)\lambda_2 + 1/t_2}{(1+1/\lambda_1 t_1)(\lambda_2 + 1/t_2) - \rho_{12}^2\lambda_2} \qquad (13)\\
\beta_2 &= \frac{\sigma_{\theta 1}}{\sigma_{\theta 2}} \frac{\rho_{12}/t_1}{(\lambda_1 + 1/t_1)(1+1/\lambda_2 t_2) - \rho_{12}^2\lambda_1} \qquad (14)\\
\beta_3 &= 0,
\end{aligned}$$

and the firm forms its expectation of θ_1 using the following rule:

$$\bar{\mu} = \bar{\theta}_1 + \beta_1(\bar{a}_1 - \bar{\theta}_1) + \beta_2(\bar{a}_2 - \bar{\theta}_2),$$

Both β_1 and β_2 are positive.

We are now ready to characterize how the effects of signals on expected revenue are affected by the other model parameters. It is straightforward to show that

$$\frac{\partial \beta_1}{\partial t_1} > 0, \quad \frac{\partial \beta_1}{\partial t_2} < 0, \quad \frac{\partial \beta_2}{\partial t_1} < 0, \quad \frac{\partial \beta_2}{\partial t_2} > 0.$$

The noisiness of signals from market 1, $\sigma_{\varepsilon 1}$, only enters β_1 and β_2 via $\lambda_1 \equiv \sigma_{\theta 1}^2/\sigma_{\varepsilon 1}^2$. Since β_1 increases with λ_1 and β_2 decreases with λ_1 (holding all the other parameters fixed), we must have

$$\frac{\partial \beta_1}{\partial \sigma_{\varepsilon 1}} < 0, \quad \frac{\partial \beta_2}{\partial \sigma_{\varepsilon 1}} > 0.$$

This completes the proof of all three properties discussed in Proposition 2. ■

1.4 Effects of t_2 on the Entry Probability

In this section, we examine how t_2 affects the entry probability and how it affects the partial derivative of $G(\pi_{1t})$ with respect to $\bar{a}_2$.

First, calculation shows

$$\frac{\partial G(\pi_{1t})}{\partial t_2} = g(\pi_{1t}) B_t e^{\bar{\mu}+\frac{\bar{\Sigma}}{2}} \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{\lambda_2 t_2}{1+\lambda_2 t_2} \left((\bar{a}_2 - \bar{\theta}_2) - \frac{\sigma_{\theta 1}\sigma_{\theta 2}\rho_{12}}{2} \right).$$

Therefore, $\frac{\partial G(\pi_{1t})}{\partial t_2} > 0$ if and only if $\bar{a}_2 > \bar{\theta}_2 + \frac{\sigma_{\theta 1}\sigma_{\theta 2}\rho_{12}}{2}$ (i.e., $\bar{a}_2$ is sufficiently large).

Next, we discuss signs of $\frac{\partial^2 \ln(\pi_{1t})}{\partial \bar{a}_2 \partial t_2}$ and $\frac{\partial^2 \pi_{1t}}{\partial \bar{a}_2 \partial t_2}$. Simple calculation shows

$$\frac{\partial^2 \ln(\pi_{1t})}{\partial \bar{a}_2 \partial t_2} = \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{1}{1+\lambda_2 t_2^2} > 0,$$

and

$$\frac{\partial^2 \pi_{1t}}{\partial \bar{a}_2 \partial t_2} = \frac{\partial \pi_{1t}}{\partial \bar{a}_2} \left[1 + \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{\lambda_2 t_2}{1+\lambda_2 t_2} \left((\bar{a}_2 - \bar{\theta}_2) - \frac{\sigma_{\theta 1}\sigma_{\theta 2}\rho_{12}}{2} \right) \right],$$

which is positive if and only if

$$1 + \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{\lambda_2 t_2}{1+\lambda_2 t_2} \left((\bar{a}_2 - \bar{\theta}_2) - \frac{\sigma_{\theta 1}\sigma_{\theta 2}\rho_{12}}{2} \right) > 0.$$

I.e., when $\bar{a}_2$ is not too small, $\frac{\partial^2 \pi_{1t}}{\partial \bar{a}_2 \partial t_2} > 0$.

Third, the relationship between entry probability, $G(\pi_{1t})$, and the nearby sibling's signal, $\bar{a}_2$, is mediated by various parameters such as t_2. One may conjecture that the sign of $\frac{\partial^2 G(\pi_{1t})}{\partial \bar{a}_2 \partial t_2}$ is unambiguous (at least under simple parameter restrictions). However, we are going to show the sign of this cross derivative is actually *ambiguous*.

Consider the cross derivative of $G(\pi_{1t})$ with respect to $\bar{a}_2$ and t_2, which can be written as

$$\frac{\partial^2 G(\pi_{1t})}{\partial \bar{a}_2 \partial t_2} = \frac{\partial}{\partial t_2} \left(g(\pi_{1t}) \frac{\partial \pi_{1t}}{\partial \bar{a}_2} \right) = g'(\pi_{1t}) \frac{\partial \pi_{1t}}{\partial t_2} \frac{\partial \pi_{1t}}{\partial \bar{a}_2} + g(\pi_{1t}) \frac{\partial^2 \pi_{1t}}{\partial \bar{a}_2 \partial t_2},$$

where $\pi_{1t} = B_t \exp(\bar{\mu} + \bar{\Sigma}/2)$. The above expression can be rewritten as

$$\frac{\partial^2 G(\pi_{1t})}{\partial \bar{a}_2 \partial t_2} = \frac{\partial \pi_{1t}}{\partial \bar{a}_2} \left[g'(\pi_{1t}) \pi_{1t} A + g(\pi_{1t})(1+A) \right],$$

where

$$A \equiv \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{\lambda_2 t_2}{1+\lambda_2 t_2} \left((\bar{a}_2 - \bar{\theta}_2) - \frac{\sigma_{\theta 1}\sigma_{\theta 2}\rho_{12}}{2} \right). \tag{15}$$

Therefore, $\frac{\partial^2 G(\pi_{1t})}{\partial \bar{a}_2 \partial t_2}$ has an ambiguous sign, as the value of $g(\pi_{1t})$ and the sign of $g^{'}(\pi_{1t})$ all depend on the value of π_{1t} and the functional assumption of $g(\cdot)$. Without knowing the distributional assumption of the entry cost, we cannot determine the sign of the above expression.

Finally, we discuss whether the sign of $\frac{\partial^2 G(\pi_{1t})}{\partial \bar{a}_2 \partial t_2}$ has a systemic pattern, if the entry cost is assumed to follow a log normal normal $N(\mu_e, \sigma_e^2)$. In such a case, we have

$$\begin{aligned}
\frac{\partial^2 G(\pi_{1t})}{\partial \bar{a}_2 \partial t_2} &= \frac{\partial^2 \Phi(\ln(\pi_{1t}))}{\partial \bar{a}_2 \partial t_2} \\
&= \frac{\partial}{\partial t_2} \left(\phi(\ln(\pi_{1t})) \frac{\partial \ln(\pi_{1t})}{\partial \bar{a}_2} \right) \\
&= \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{1}{1+\lambda_2 t_2^2} \left[\phi'(\ln(\pi_{1t})) A + \phi(\ln(\pi_{1t})) \right] \\
&= \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{1}{1+\lambda_2 t_2^2} \phi(\ln(\pi_{1t})) \left(1 - A \frac{\pi_{1t} - \mu_e}{\sqrt{\sigma_e^2}} \right),
\end{aligned}$$

where A is defined in equation (15), Φ and ϕ denote the CDF and PDF of the normal distribution with mean μ_e and variance σ_e^2. The last step comes from the definition of PDF of the log normal distribution. We know $\phi(\ln(\pi_{1t}))$ is positive and both A and π_{1t} strictly increase with $\bar{a}_2$. In particular, both A and π_{1t} approach infinity when $\bar{a}_2$ goes to infinity, which leads to $\frac{\partial^2 \Phi(\ln(\pi_{1t}))}{\partial \bar{a}_2 \partial t_2} < 0$. However, we do not know the sign of $1 - A\frac{\pi_{1t}-\mu_e}{\sqrt{\sigma_e^2}}$ (and thus $\frac{\partial^2 \Phi(\ln(\pi_{1t}))}{\partial \bar{a}_2 \partial t_2}$) in general. In total, our learning model has an ambiguous prediction on how the number of signals affects the positive impact of a better average signal on the entry probability.

1.5 Model Predictions with Positive Cross-region Correlations

In this subsection, we discuss how our model predictions change when we allow ρ_{13} and ρ_{23} to be positive. In particular, we make the following assumption instead of Assumption 1.

Assumption 1' $\rho_{12} > \rho_{23} = \rho_{13} > 0$.

Under this alternative assumption, we have two propositions analogous to Propositions 1 and 2.

Proposition 1' *Assume Assumption 1' holds. Before the firm enters market 1, it uses signals from both markets 2 and 3 to forecast its "would-be" demand in market 1. The firm's expected profit and entry probability in market 1 increases with the average past signals* $\bar{a}_2 \equiv \sum_{\tau=t-t_2}^{t-1} a_{2\tau}/t_2$. *and* $\bar{a}_3 \equiv \sum_{\tau=t-t_3}^{t-1} a_{3\tau}/t_3$.

Proof. Since $0 < \rho_{23} = \rho_{13} < \rho_{12}$, one can simplify equations (5) and (6) and show

$$\beta_2 > 0, \beta_3 > 0.$$

Because the average past signals only affect the expected profit and entry probability via the conditional mean of θ_1 ($\bar{\mu}$), both margins increase with $\bar{a}_2$ and $\bar{a}_3$. ■

Proposition 2' *Under Assumption 1', an affiliate in market 1 uses its own average past signal, that of its siblings in market 2, and that of its siblings in market 3 to form its expectation of future sales, with positive weights on all average signals. All else equal, the weights it places on its own average signal and those of the sibling in market 2 have the following properties:*

1. *The weight it places on its own average signal (the average signal of siblings in market 2) increases (decreases) with its age;*

2. *the weight it places on its own average signal (the average signal of siblings in market 2) decreases (increases) with the standard deviation of the time-varying idiosyncratic shocks in its market (market noisiness);*

3. *The weight it places on its own average signal (the average signal of siblings in market 2) decreases (increases) with the total number of signals in market 2.*

Proof. Similar to the proof of Proposition 2, we simplify equations (7) to (9) under the new assumption. Specifically, we rewrite the expressions for β_1 and β_2 as

$$\begin{aligned}
\beta_1 &= \frac{\sigma_{\theta 1}^2\sigma_{\theta 2}^2\sigma_{\theta 3}^2}{\Delta}\left[\begin{array}{c} 2\rho_{12}\rho_{13}\rho_{23} + (1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2 \\ -\rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{13}^2(1+\frac{1}{\lambda_2 t_2}) \end{array}\right], \\
\beta_2 &= \frac{\sigma_{\theta 1}\sigma_{\theta 2}\sigma_{\theta 3}^2\sigma_{\varepsilon 1}^2}{\Delta}\left[\frac{\rho_{12}}{t_1}(1+\frac{1}{\lambda_3 t_3}) - \rho_{13}\rho_{23}\frac{1}{t_1}\right], \\
\beta_3 &= \frac{\sigma_{\theta 1}\sigma_{\theta 3}\sigma_{\theta 2}^2\sigma_{\varepsilon 1}^2}{\Delta}\left[\frac{\rho_{13}}{t_1}(1+\frac{1}{\lambda_2 t_2}) - \rho_{12}\rho_{23}\frac{1}{t_1}\right],
\end{aligned}$$

where Δ equals

$$\sigma_{\theta 1}^2\sigma_{\theta 2}^2\sigma_{\theta 3}^2\left[2\rho_{12}\rho_{13}\rho_{23} + (1+\frac{1}{\lambda_1 t_1})\left[(1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2\right] - \rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{13}^2(1+\frac{1}{\lambda_2 t_2})\right].$$

It is straightforward to show that

$$\beta_1, \beta_2, \beta_3 > 0.$$

Regarding the effect of the signals moderated by t_1, t_2 and $\sigma_{\varepsilon 1}$, we take the partial derivative of β_1 and β_2 with respect to these parameters. Three points are worth mentioning. First, the numerator of β_1 does not depend on t_1 and $\sigma_{\varepsilon 1}$ and the numerator of β_2 does not depend on t_2. Second, Δ increases with $\sigma_{\varepsilon 1}$ and decreases with t_1 and t_2. Therefore, we must have

$$\frac{\partial \beta_1}{\partial \sigma_{\varepsilon 1}} < 0, \quad \frac{\partial \beta_1}{\partial t_1} > 0, \quad \frac{\partial \beta_2}{\partial t_2} > 0.$$

Third, the numerator of β_2 increases proportionately with $\sigma_{\varepsilon 1}$ and decreases proportionately with t_1. However, the determinant of matrix A, Δ, increases less proportionately with $\sigma_{\varepsilon 1}$ and decreases less proportionately with t_1.[1] Therefore, we must have

$$\frac{\partial \beta_2}{\partial \sigma_{\varepsilon 1}} > 0, \quad \frac{\partial \beta_2}{\partial t_1} < 0.$$

[1] This is true, as $2\rho_{12}\rho_{13}\rho_{23} + \left[(1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2\right] - \rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{13}^2(1+\frac{1}{\lambda_2 t_2})$ is strictly positive.

Finally, we analyze how β_1 varies with t_2. We rewrite β_1 as

$$\beta_1 = \begin{bmatrix} (1+\frac{1}{\lambda_2 t_2})[(1+\frac{1}{\lambda_3 t_3})(1+\frac{1}{\lambda_1 t_1}) - \rho_{13}^2] \\ +2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2(1+\frac{1}{\lambda_1 t_1}) \end{bmatrix}^{-1} \begin{bmatrix} (1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3} - \rho_{13}^2) \\ +2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2 \end{bmatrix}.$$

We prove that $\frac{1}{\beta_1}$ decreases with $1+\frac{1}{\lambda_2 t_2}$ in what follows:

$$\frac{1}{\beta_1} = 1 + \begin{bmatrix} (1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3} - \rho_{13}^2) \\ +2\rho_{12}\rho_{13}\rho_{23} - \rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2 \end{bmatrix}^{-1} \left[\frac{1}{\lambda_1 t_1}(1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3}) - \frac{\rho_{23}^2}{\lambda_1 t_1}\right] > 1.$$

The calculation shows that

$$Sign\left[\frac{\partial \log\left(\frac{1}{\beta_1} - 1\right)}{\partial \log\left(1+\frac{1}{\lambda_2 t_2}\right)}\right] = Sign\left[-\left[(1+\frac{1}{\lambda_3 t_3})\rho_{12} - \rho_{13}\rho_{23}\right]^2\right] < 0.$$

Since $1+\frac{1}{\lambda_2 t_2}$ decreases with t_2, we have

$$\frac{\partial \beta_1}{\partial t_2} < 0.$$

■

2 Additional Empirical Results

2.1 Full Set of Interaction Terms in the Expectation Formation Regressions

Table OA.1: Full set of interaction terms in the expectation formation regressions

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.869^a	0.869^a	0.868^a	0.867^a
	(0.00999)	(0.00986)	(0.00995)	(0.00987)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.0261^a	-0.0293^a		
	(0.00765)	(0.00922)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.0128^b	-0.0104^c
			(0.00629)	(0.00622)
$\times$ log(self age)	0.0856^a	0.0860^a	0.0913^a	0.0881^a
	(0.00661)	(0.00676)	(0.00643)	(0.00691)
$\times$ Nearby siblings' experience	0.00383	0.00309	0.00107	0.00361
	(0.00762)	(0.00778)	(0.00748)	(0.00765)
$\times$ Destination income level		-0.00362		0.0137
		(0.0112)		(0.00899)
Average nearby signal	0.0507^a	0.0497^a	0.0503^a	0.0515^a
	(0.0143)	(0.0154)	(0.0142)	(0.0152)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.0197^b	0.0238^b		
	(0.00786)	(0.00929)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.00944	0.00761
			(0.00606)	(0.00630)
$\times$ log(self age)	-0.0471^a	-0.0479^a	-0.0516^a	-0.0495^a
	(0.00972)	(0.00977)	(0.00960)	(0.00986)
$\times$ Nearby siblings' experience	0.0232^b	0.0239^b	0.0244^b	0.0242^b
	(0.00976)	(0.00955)	(0.00965)	(0.00953)
$\times$ Destination income level		0.00638		-0.00790
		(0.0148)		(0.0121)
Nearby siblings' experience	0.0128	0.0136	0.0145	0.0136
	(0.0200)	(0.0202)	(0.0201)	(0.0202)
Average remote signal	0.0197	0.0208	0.0198	0.0211
	(0.0255)	(0.0256)	(0.0258)	(0.0257)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	31714	31599	31697	31582
R^2	0.905	0.905	0.905	0.905

Notes: Dependent variable is the logarithm of expected sales in the next year. Nearby siblings' experience is the total number of nearby siblings' signals. Host country income level is measured as the log of real GDP per capita in 2005. All moderator variables are standardized. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

2.2 Excluding Observations with Zero Expected Growth Rates

Table OA.2 replicates the regressions in Table OA.1 after excluding affiliates whose expected growth rates are zero. The results are robust.

Table OA.2: Full set of interaction terms in the expectation formation regressions, excluding observations with zero expected growth rates

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.859^a	0.858^a	0.857^a	0.856^a
	(0.00962)	(0.00952)	(0.00951)	(0.00951)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.0245^a	-0.0285^a		
	(0.00755)	(0.00865)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.0127^b	-0.0107^c
			(0.00577)	(0.00559)
$\times$ log(self age)	0.0853^a	0.0858^a	0.0905^a	0.0877^a
	(0.00652)	(0.00667)	(0.00629)	(0.00679)
$\times$ Nearby siblings' experience	0.00524	0.00433	0.00275	0.00492
	(0.00763)	(0.00776)	(0.00748)	(0.00763)
$\times$ Destination income level		-0.00461		0.0121
		(0.0105)		(0.00869)
Average nearby signal	0.0537^a	0.0526^a	0.0535^a	0.0547^a
	(0.0143)	(0.0155)	(0.0141)	(0.0151)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.0204^a	0.0247^a		
	(0.00764)	(0.00915)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.0107^c	0.00898
			(0.00589)	(0.00615)
$\times$ log(self age)	-0.0447^a	-0.0455^a	-0.0491^a	-0.0470^a
	(0.00969)	(0.00976)	(0.00957)	(0.00983)
$\times$ Nearby siblings' experience	0.0215^b	0.0221^b	0.0226^b	0.0224^b
	(0.00966)	(0.00943)	(0.00953)	(0.00938)
$\times$ Destination income level		0.00660		-0.00795
		(0.0149)		(0.0121)
Nearby siblings' experience	0.0104	0.0113	0.0118	0.0110
	(0.0201)	(0.0202)	(0.0201)	(0.0202)
Average remote signal	0.0247	0.0257	0.0252	0.0261
	(0.0251)	(0.0252)	(0.0254)	(0.0252)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	31101	30988	31084	30971
R^2	0.905	0.904	0.904	0.904

Notes: Dependent variable is the logarithm of expected sales in the next year. Nearby siblings' experience is the total number of nearby siblings' signals. Host country income level is measured as the log of real GDP per capita in 2005. All moderator variables are standardized. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

2.3 Stricter Definitions of Horizontal MP

Our theory applies to horizontal rather than vertical multinational production (MP). In our main specifications, we carefully constructed our sample by focusing on affiliates with at least 85% of their sales in the local market. In Columns 1 and 2 of Tables OA.3 and OA.4, we set the threshold as 95% and check the robustness of our results. This reduces the number of entries in the entry regressions and the number of observations in the expectation formation regressions. For the expectation formation regressions, we only present the specifications with the full set of moderator variables. The results are similar to those obtained before.

Table OA.3: Robustness of the entry regressions: Stricter definitions of horizontal MP

Def. of Horizontal Entry	Local Sales Share ≥ 0.95		Regional Import Share < 0.15		Import Share < 0.15	
Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
Average nearby signal	0.0144^a	0.0139^a	0.0101^a	0.0115^a	0.00984^a	0.0109^a
	(0.00321)	(0.00364)	(0.00309)	(0.00322)	(0.00303)	(0.00309)
Average remote signal	0.00327	0.000324	0.00207	0.00363	0.00187	0.00331
	(0.00448)	(0.00481)	(0.00413)	(0.00424)	(0.00415)	(0.00413)
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓
Firm FE	✓		✓		✓	
Firm-Year FE		✓		✓		✓
N	902532	902527	902532	902527	902532	902527
R^2	0.065	0.085	0.073	0.095	0.072	0.094
# of Firms	1931	1931	1931	1931	1931	1931
# of Firm-Markets	115184	115183	115184	115183	115184	115183
# of Entries	830	830	746	746	720	720

Notes: Dependent variable is an indicator variable indicating whether the firm enters a particular destination in the next year. Siblings' signals are the average of past residual sales. The local sales share is the ratio of local sales to total sales. The regional import share is the ratio of imports from other countries in the same region to total sales. The import share is the ratio of imports from the rest of the world (excluding Japan) to total sales. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

To show that our results are not simply driven by regional value chains, we calculate the regional and total import shares for each affiliate. The regional import share is defined as affiliates' imported inputs from countries in the same region (excluding Japan) divided by total sales, while the total import share is the ratio of imports from all countries excluding Japan to total sales. We restrict our sample to affiliates whose import shares are less than 15%, since we see these affiliates as not well integrated into the regional value chains or the global value chains.

Columns 3–6 of Table OA.3 report the entry regressions with the restricted sample.

Table OA.4: Robustness of the expectation formation regressions: Stricter definitions of horizontal MP

Def. of Horizontal Affiliates	Local Sales Share ≥ 0.95		Regional Import Share < 0.15		Import Share < 0.15	
Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)	(5)	(6)
Average self signal	0.873^a	0.872^a	0.871^a	0.871^a	0.874^a	0.875^a
	(0.0107)	(0.0105)	(0.0114)	(0.0112)	(0.0121)	(0.0120)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.0264^a	-0.0298^a	-0.0311^a	-0.0333^a	-0.0320^a	-0.0410^a
	(0.00775)	(0.00951)	(0.00983)	(0.0126)	(0.00996)	(0.0128)
$\times$ Self experience	0.0883^a	0.0886^a	0.0917^a	0.0921^a	0.0959^a	0.0969^a
	(0.00724)	(0.00739)	(0.00781)	(0.00793)	(0.00807)	(0.00820)
$\times$ Nearby siblings' experience	0.00795	0.00709	0.0108	0.0103	0.0130	0.0113
	(0.00828)	(0.00843)	(0.00865)	(0.00888)	(0.00909)	(0.00927)
$\times$ Destination income level		-0.00379		-0.00280		-0.0112
		(0.0121)		(0.0130)		(0.0137)
Average nearby signal	0.0465^a	0.0459^a	0.0549^a	0.0535^a	0.0615^a	0.0602^a
	(0.0141)	(0.0152)	(0.0170)	(0.0186)	(0.0167)	(0.0179)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.0177^b	0.0204^b	0.00697	0.0114	0.00633	0.0131
	(0.00758)	(0.00949)	(0.0109)	(0.0124)	(0.0128)	(0.0132)
$\times$ Self experience	-0.0529^a	-0.0537^a	-0.0557^a	-0.0562^a	-0.0578^a	-0.0589^a
	(0.0105)	(0.0106)	(0.0118)	(0.0120)	(0.0128)	(0.0129)
$\times$ Nearby siblings' experience	0.0204^b	0.0209^b	0.0267^b	0.0270^b	0.0271^b	0.0272^b
	(0.0100)	(0.00988)	(0.0110)	(0.0106)	(0.0108)	(0.0108)
$\times$ Destination income level		0.00444		0.00637		0.00949
		(0.0151)		(0.0191)		(0.0188)
Nearby siblings' experience	0.00906	0.00966	0.00406	0.00466	-0.00771	-0.00670
	(0.0211)	(0.0213)	(0.0224)	(0.0228)	(0.0230)	(0.0232)
Average remote signal	0.00446	0.00538	0.0301	0.0313	0.0169	0.0180
	(0.0247)	(0.0248)	(0.0263)	(0.0265)	(0.0272)	(0.0272)
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
N	26216	26102	24065	24021	23285	23247
R^2	0.909	0.909	0.909	0.909	0.912	0.912

Notes: Dependent variable is the logarithm of expected sales in the next year. Standard errors are clustered at the firm level. Self-experience is the log of affiliate age. Nearby siblings' experience is the total number of nearby siblings' signals. Host country income level is the log of real GDP per capita in 2005. All moderator variables are standardized. The local sales share is the ratio of local sales to total sales. The regional import share is the ratio of imports from other countries in the same region to total sales. The import share is the ratio of imports from the rest of the world (excluding Japan) to total sales. Significance levels: a: 0.01, b: 0.05, c: 0.10.

Requiring the regional import share to be lower than 15% reduces the number of entries by around one-quarter. Compared with the earlier results, the coefficient of the average nearby siblings' signal falls, suggesting that part of the earlier results are driven by integration into regional value chains. Nevertheless, restricting the sample does not eliminate these effects. Columns 5 and 6 require entering affiliates to have an import share below 15%. The import share is higher than the regional import share by definition, and thus we drop more entries. However, since most of the imported inputs are from the same region, this criterion only drops slightly more entries compared with Columns 3 and 4. We obtain similar results as in those two columns. The results from the expectation formation regressions are also robust to using these two definitions of horizontal MP, which are reported in Table OA.4.

2.4 Excluding the Year of 1995

In our data, the entry rates in 1995 are higher than the other years. In Table OA.5 and OA.6, we replicate the entry and expectation formation regressions in Table 4 in the paper and Table OA.1, respectively after excluding the year 1995 from our sample. The main empirical results are robust.

Table OA.5: Impact of siblings' experience on entry in the next period, excluding 1995

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 100$	(1)	(2)	(3)
Average nearby signal	0.0154^a	0.0162^a	0.0156^a
	(0.00312)	(0.00374)	(0.00404)
Average remote signal	0.00578	0.00528	0.00337
	(0.00401)	(0.00530)	(0.00551)
Firm domestic sales	0.00510	-0.0110	
	(0.00325)	(0.0101)	
Destination-Ind-Year FE	✓	✓	✓
Firm FE		✓	
Firm-Year FE			✓
N	853608	853608	879313
R^2	0.062	0.065	0.086
# of Firms	1914	1914	1922
# of Firm-Markets	112869	112869	114043
# of Entries	909	909	931

Notes: The dependent variable indicates whether the firm enters a particular destination in the next year. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

Table OA.6: Full set of interaction terms in the expectation formation regressions, excluding the year 1995 from our sample

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.869^a	0.868^a	0.868^a	0.867^a
	(0.00997)	(0.00984)	(0.00992)	(0.00986)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.0249^a	-0.0287^a		
	(0.00773)	(0.00933)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.0121^c	-0.00995
			(0.00633)	(0.00626)
$\times$ log(self age)	0.0862^a	0.0866^a	0.0917^a	0.0887^a
	(0.00663)	(0.00677)	(0.00645)	(0.00691)
$\times$ Nearby siblings' experience	0.00338	0.00253	0.000650	0.00300
	(0.00764)	(0.00780)	(0.00751)	(0.00769)
$\times$ Destination income level		-0.00437		0.0127
		(0.0111)		(0.00895)
Average nearby signal	0.0513^a	0.0502^a	0.0509^a	0.0520^a
	(0.0144)	(0.0155)	(0.0142)	(0.0153)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.0188^b	0.0232^b		
	(0.00795)	(0.00936)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.00891	0.00728
			(0.00608)	(0.00631)
$\times$ log(self age)	-0.0475^a	-0.0484^a	-0.0519^a	-0.0500^a
	(0.00974)	(0.00979)	(0.00963)	(0.00988)
$\times$ Nearby siblings' experience	0.0236^b	0.0243^b	0.0247^b	0.0246^b
	(0.00980)	(0.00958)	(0.00969)	(0.00956)
$\times$ Destination income level		0.00685		-0.00713
		(0.0148)		(0.0122)
Nearby siblings' experience	0.0127	0.0135	0.0143	0.0135
	(0.0201)	(0.0202)	(0.0202)	(0.0203)
Average remote signal	0.0192	0.0203	0.0192	0.0205
	(0.0256)	(0.0257)	(0.0259)	(0.0258)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	31586	31471	31569	31454
R^2	0.905	0.905	0.905	0.905

Notes: Dependent variable is the logarithm of expected sales in the next year. Nearby siblings' experience is the log of total number of nearby siblings' signals. Host country income level is measured as 2005 real GDP per capita. All moderator variables are standardized. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

References

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 649–670.