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: December 2021

Learning and Information Transmission within Multinational Corporations[*]

Cheng CHEN Chang SUN Hongyong ZHANG[†]

Abstract

Firms face substantial uncertainty when doing business in new markets. We propose that multinational firms use "cross-market learning" to resolve such uncertainties. We develop a model of firm-level expectations formation with noisy signals from multiple markets and derive predictions on market entries and expectations formation over the firm's life cycle. Using a novel dataset of Japanese multinational corporations that includes sales expectations of each affiliate, we provide supportive evidence for the model's predictions. We find that firms rely on their performance in nearby markets to predict their profitability in a new market and make entry decisions. Such "cross-market learning" is less important after the firm has accumulated experience in the new market, but becomes more important if the uncertainty of the focal market is high or the firm has received more signals from the nearby markets.

Keywords: learning, expectation formation, multinational production

JEL classification: D83, F2

[*] 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, Ken Kikkawa, 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 in new markets, and this is particularly true for multinational corporations (MNCs) that produce and sell in multiple countries. Before a firm enters a particular market, it may have limited information about consumers' tastes and the costs of production. Given the large sunk entry costs, MNCs' entry decisions into new markets can be costly when information is imperfect. In this paper, we ask how and to what extent MNCs can resolve such uncertainties. In particular, we propose "cross-market" learning as a key mechanism, in which firms learn about their profitability in a new market based on their performance in nearby markets, and use direct measures of sales expectations to show the existence of such a learning mechanism.

To guide our empirical analysis, we first build a model in which a firm learns about its demand conditions in multiple markets based on similar ideas in the social learning literature (Foster and Rosenzweig, 1995; Jovanovic and Nyarko, 1996).[1] The model serves three purposes. First, it provides testable predictions of cross-market learning. Second, it informs us of ways to estimate key model parameters, which can be further used in the empirical analysis. Third, it makes a distinction between the implications of cross-market learning on market entries and those on expectations formation. Specifically, we show that though good performance in nearby markets increases both entry probabilities and sales expectations, the effects of other model parameters on this relationship (cross-derivatives in our model) are ambiguous for entry probabilities but unambiguous for expectations formation.[2] Therefore, it is crucial to use direct measures of firm-level expectations to test model predictions on the cross-derivatives.

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

[1] We assume that the only uncertainty comes from the demand side in our model. One can allow uncertainty on the supply side and reinterpret our model as a firm learning about its efficiency in different markets. Given that we do not observe prices and quantities separately in our data, we do not attempt to distinguish between learning about demand and supply.

[2] 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.

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.[3]

Several key testable predictions emerge from the model. First, the firm uses the average signal from the nearby market to forecast its expected sales in a new market and ignores signals from the remote market. Thanks to the positive correlation in the time-invariant demand, the firm can use information from the nearby market to reduce their uncertainty about 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.[4] Therefore, we focus on testing the comparative statics regarding firms' expectations formation, as the model yields unambiguous predictions along this dimension.

After the firm enters the new market, it continues to update its expectation of future sales given the signals observed, which now also include the signals from the new market. Again, due 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 (Jovanovic, 1982; Jovanovic and Nyarko, 1997) extended to cross-market learning. We show that the firm's expectation in the new market relies more on the average signal from the nearby market and less on the signal from the new market when (1) the firm is less experienced in the new market, and/or the firm is more experienced in the nearby market, and, (2) the signals from the new market are noisier (with higher variance of the transitory shocks). The intuition is that signals from the nearby market are more precise relative to those from the new market under these conditions.

We take advantage of a 22 year long panel dataset of Japanese MNCs to test the

[3]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 OA.1.6, 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.

[4]Online Appendix OA.1.5 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.

theoretical predictions. The dataset is at the affiliate-year level and it includes a measure of each affiliate's sales expectations for the next year. In Section 4, we provide more descriptions about the data and this unique variable, and show that the sales expectations are reliable and contain useful information that is used in actual firm decisions.

We first test our model's predictions on new market entries. 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.[5] By contrast, the average past signal of affiliates outside the region (referred to as "remote siblings") has a weak and statistically insignificant effect. Our baseline estimate suggests that a one-standard-deviation increase in the average nearby siblings' signal leads to an increase in entry probability by 0.28‰, approximately 25% of the average entry rate.

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 baseline sample to include regions which the firm has not entered yet and estimate the impact of different deciles of average nearby siblings' signal on the probability of market entry. Relative to firms without any nearby 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 the above finding of the heterogeneous effects demonstrates an important distinction between our learning mechanism and other mechanisms that lead to sequential entries in similar markets. For instance, the "extended gravity" literature (Morales et al., 2019) finds that an exporter's prior entry in nearby markets lowers the sunk entry costs into new markets and thus increases its entry probability into a new market.[6] Their mechanism may well exist for many MNCs in our data, as the presence

[5]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.

[6]Strictly speaking, Morales et al. (2019) study entries into new destinations by exporters, which cannot be directly compared to MNC entries. Using U.S. multinational firm data, Garetto et al. (2019) show that the entry probability conditional on having an existing affiliate in a market within the same continent is slightly larger than the unconditional probability. However, they do not consider the heterogeneity due to the performance of the existing affiliates.

of nearby siblings starts to show a positive and significant impact on subsequent entries into new markets when the siblings' signal is above the fourth decile. However, this is clearly 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 on subsequent entries may well depend on the historical performance of the existing affiliates.

Next, we explore our measure of affiliates' sales expectations after market entry and provide empirical support for additional theoretical predictions over the affiliate's life cycle. We find that strong average signals of nearby siblings raise the expectation for the next year's sales, while the average signals 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' signals on sales expectations hides rich underlying heterogeneity. Following the model's predictions, we further examine how market and affiliate characteristics affect the strength of learning. 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. In addition, model-consistent measures of market-level uncertainty, or the noisiness of the signals, can hinder the firm's learning in the new market and make it rely more on signals from nearby markets. Such heterogeneous effects are our key evidence for learning, as one may worry that our earlier findings of nearby siblings' signals on entry and sales expectations can be driven by correlated shocks within the firm 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. Moreover, taking advantage of our direct measures of sales forecasts, we perform a simple calibration and show that the coefficients in the expectations formation formula implied by the model are in the same ballpark of those estimated from the data for affiliates of different ages.

2 Literature Review

Our study contributes to five strands of the literature. First, our study contributes to the literature on learning in the international context. Existing studies have documented the role of learning in exporter dynamics, as well as the inter-market linkages through information acquisition or sunk cost reduction. For instance, Timoshenko (2015a,b) study how incorporating self-discovery by exporting firms into an otherwise standard heterogeneous firms models helps explain persistence in exporting and behavior of new exporters. Berman et al. (2017) and Arkolakis et al. (2018b) show that learning about demand is an important driver of firm dynamics.[7] Relatedly, 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. We contribute to this literature by directly detecting firm learning, thanks to the availability of firm-level expectations data. Our empirical approach can be extended further in future research, as firm-level expectations data are becoming increasingly available.[8]

A growing literature has focused on MNC dynamics, with or without the learning mechanism (Egger et al., 2014; Conconi et al., 2016; Gumpert et al., 2016; Garetto et al., 2019; 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. Conconi et al. (2016) and Gumpert et al. (2016) study the joint dynamics of exporters and MNCs with and without the learning mechanism. We differ from their studies by studying how cross-market learning and thus information transmission within the same MNC shape patterns of FDI entries and MNC dynamics. Finally, our current paper has a different focus from Chen et al. (2020) which study life-cycle dynamics of MNCs without the mechanism of cross-market learning within the same MNC.[9]

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

[7]Other importance contributions include Akhmetova and Mitaritonna (2013) Aeberhardt et al. (2014) and Cebreros (2016).

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

[9]As a result, affiliates in different countries that belong to the same MNC parent firm operate independently in Chen et al. (2020).

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.

Fourth, our paper connects to a large literature on learning and technology adoption (Foster and Rosenzweig, 1995; Jovanovic and Nyarko, 1996; Conley and Udry, 2010) as well as its applications in international trade (Fernandes and Tang, 2014; Kamal and Sundaram, 2016; Hamilton, 2018). Our model shares the same key ingredients as models in this literature, i.e., Bayesian updating and correlated signals, and we extend these models naturally to cross-market learning by the same firm.[10] Our main contribution to this literature is to directly measure firm's expectations in each market and use the expectations to test additional predictions from the model, such as life-cycle learning and the impact of uncertainty on learning.

Finally, our paper is also related to the work on inflation expectations and agents' decisions (Malmendier and Nagel, 2011, 2016; Coibion et al., 2018, 2020). Work on this topic finds that experience affects agents' expectations formation and thus actions. We contribute to this literature by showing how one key dimension of experience that is age affects firm-level expectations formation.

3 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.

[10] Our model has a key conceptual difference from the social learning literature: information "spillover" happens within the firm, so the firm can internalize such an "externality". Since we do not solve the full dynamics of the MNCs since their births, this distinction is only conceptual and does not affect our empirical tests.

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 innovation 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.

3.1 Setup

We study a single firm's problem. Suppose there are three foreign markets: markets 1 and 2 are in the same region, and market 3 is in another region. Without loss of generality, we focus on the firm's expectation in market 1, and refer to markets 2 and 3 as the "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

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

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 the 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}.

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

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

The above equation implies 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 where it has entered. To enter market 1, the firm has to pay a one-time entry cost F, where the cumulative distribution function of F is $G(\cdot)$.

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 the three market is

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

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.1, 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 OA.1.6, 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 involved.

3.2 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

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

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

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

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 OA.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 $\bar{a}_2$ on the entry probability depends on distributional assumptions of the idiosyncratic entry cost $G(\cdot)$. This is true, even if we assume that $G(\cdot)$ is log normal.[11] 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 life cycle is to derive and test theoretical predictions regarding expectations formation directly.

3.3 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. *[life-cycle learning] The weight it places on the average signal of itself (its nearby sibling) increases (decreases) with self age, and decreases (increases) with the total number of signals received from market 2.*

2. *[uncertainty impedes self-learning] 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.*

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.

[11] We prove that the cross derivative of the entry probability with respect to t_2 and $\bar{a}_2$ is ambiguous. See Online Appendix OA.1.5 for details.

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 OA.1.6, 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 thus show that all the results hold under the weaker assumption.

A convenient and probably unrealistic assumption of the our model is that temporary demand shocks are uncorrelated between the focal affiliate and its siblings in the same region. One interesting modification of our baseline model is to allow the temporary shocks to be positively correlated across destination economies within the same region. In Online Appendix OA.1.4, we provide such an extension and examine the impact of the correlation in temporary shocks on the learning parameters (the coefficient of the self signal and that of the sibling's signal). We are able to sign the impact under certain parameter restrictions. We also empirically examine the correlation in ε_{jt} in Appendix A.1. We find that the level of correlation is much lower than that among θ_j. Numerical simulations in Online Appendix OA.1.4 show that, if the correlation is as large as what we observe in the data, they have a negligible effect on the learning parameters β_1 and β_2.

4 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.

4.1 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 4.3.

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" and "transportation equipment" are the largest in services and manufacturing, respectively. 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 5, we examine the dynamic patterns of entry and the impact of siblings' signals formally.

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

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).

same region as the focal market and (2) the two markets belong to the same industry.[12] 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 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. The local-to-total-sales ratio may change over the affiliate's life cycle (Garetto et al., 2019). We calculate the 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%.[13] We define entry when a firm sets up its first horizontal affiliate in the destination market. Our main empirical results are robust to increasing this threshold to 95%. These results are presented in Columns 1 and 2 of Appendix Tables A.3 and A.4.

[12] 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 5.1, signals from different industries do not significantly affect entry probabilities.

[13] Garetto et al. (2019) find that the local sales shares decline as U.S. affiliates become older. Because of such fluctuations, we select our sample in slightly different ways for the entry regressions and the post-entry expectations formation regressions. In the entry regressions, we define a "horizontal affiliate" as one whose life-time average local-to-total sales ratio is above 85%, because we want to capture affiliates that are established mainly for local sales purposes. For the expectations formation regressions, we select the sample based on the current local sales share, because the learning mechanism that we propose may be very important for young affiliates that mainly serve the local market, meaning that we remove them from our sample when their local sales share drops below 85%. In Table A.4, we show that the expectations formation results are robust if we select the sample based on average local sales shares over the affiliates' life cycle.

4.2 Construction of Siblings' Signals

Consistent with our theory, we focus on firm learning about their idiosyncratic demand/supply conditions in particular markets. We construct our measures of signals using affiliates' local sales, which exclude potential vertical sales to parent firms or downstream affiliates of the same business group in other countries. We further 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 3 suggests that firms infer their market-specific demand using all past signals. Therefore, we construct the cumulative average of the past residual 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 destinations in the same region other than 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.[14] 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, requiring

[14] One may consider an alternative model in which firms may learn more when they sell to more customers. In this scenario, we should calculate sales-weighted signals instead of an simple average. In Online Appendix OA.2.6, we present results using sales-weighted signals. These signals are highly correlated with simple averages, and we observe similar results from the entry and expectations formation regressions. We think that there is not enough variation for us to distinguish our model and an alternative model where firms learn more when they sell more.

the observation to have at least one nearby and one remote siblings. Therefore, our main analysis can be at best seen as a test of the theory conditional on the firm having affiliates in both the nearby and remote markets. After we introduce our main results in Section 5.1, we consider firms that have not established any affiliates in a region or only have established affiliates in one region, i.e., Figure 1 and Table 7.

4.3 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.

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 Online Appendix OA.2.11, only 1.59% of the observations use their sales in year t as a forecast for sales in $t+1$. Our main regression results are basically unchanged after dropping these observations (see Online Appendix Table OA.17).[15] Second, we show in Table 2 that the sales forecasts have statistically significant and economically strong impacts on realized sales and employment in the future, even when we control for past 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 are unlikely to have incentives to misreport to avoid taxes or to manage stock market expectations.

5 Empirical Evidence

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

5.1 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

[15]In Online Appendix OA.2.5, we report a small but non-negligible fraction of affiliates (1.08%) report expected sales that coincide with the realized sales next period. These affiliates may be subject to another concern: managers and workers start shirking after they "hit the target". We perform robustness checks by dropping these observations and the results are similar.

Table 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
	(0.029)	(0.032)	(0.035)	(0.020)	(0.013)	(0.014)
$\log R_{it}$		0.121^a	0.121^a			
		(0.019)	(0.036)			
$\log R_{i,t-1}$			0.038^a			
			(0.011)			
$\log L_{it}$					0.514^a	0.505^a
					(0.023)	(0.030)
$\log L_{i,t-1}$						0.019
						(0.021)
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	0.44	0.46	0.45	0.16	0.37	0.36
R-squared	0.97	0.97	0.98	0.97	0.98	0.98

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 8 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.

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 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 any operations in those markets yet and thus has not received 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 3 shows the number of observations and next years' entries by year in the sample used in our baseline regressions on average. There are around 41,700 firm-

[16]As discussed in Section 4.2, for a particular focal market, siblings are affiliates of the same firm operating in the same industry but different countries.

market combinations in each year and 47 of them will see a new entry in the next year. The average entry rate is 1.1‰.[17]

Table 3: Number of observations and entries by year

Year	(1) # of obs.	(2) # of next year's entries	(3) entry rate (‰)
1995	21919	68	3.10
1996	27100	50	1.85
1997	27736	46	1.66
1998	30772	42	1.36
1999	36426	39	1.07
2000	36101	50	1.39
2001	33531	48	1.43
2002	38926	42	1.08
2003	40303	58	1.44
2004	41905	49	1.17
2005	44611	44	0.99
2006	44952	53	1.18
2007	44289	41	0.93
2008	46096	36	0.78
2009	47753	48	1.01
2010	47073	72	1.53
2011	48704	78	1.60
2012	51978	47	0.90
2013	54473	32	0.59
2014	54338	23	0.42
2015	56541	11	0.19
Total	875527	977	1.12

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 now introduce our econometric specification for this subsection. In particular, we run the following linear probability regression:

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

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 defined in equation (6). 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

[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 OA.2.12, we show our main empirical results are robust if we exclude 1995 from our sample.

the less extreme assumption that cross-region correlation in time-invariant demand is positive but smaller than that within region, we expect b_2 to be positive but smaller than b_1.[18]

Note that we add residual domestic performance $\tilde{r}_{ft}$ in our specification to control for productivity shocks to the entire multinational firm. The assumption here is that productivity shocks to the parent firm can be transmitted to all affiliates at the same rate. This is stronger than assuming the transmission rates are destination-specific, a typical assumption in the literature (Ramondo and Rodríguez-Clare, 2013; Tintelnot, 2017; Arkolakis et al., 2018a). In Online Appendix Section OA.2.3, we show that our main results are robust to controlling for $\tilde{r}_{ft}$ interacted with destination-industry fixed effects. We also consider parent firms' heterogeneous exposure to aggregate shocks, such as the banking shocks in Japan in the 1990s, by controlling for the interactions between parent firm characteristics (capital-labor ratio and firm sales) and year dummies. Finally, as discussed below, our empirical results are robust to controlling for firm-year fixed effects instead of firm fixed effects, which is a more flexible approach to control firm-level shocks regardless of their sources.

Before we show the regression results, Table 4 presents the summary statistics of the key regressors and related variables in the same sample as in Table 3. 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 being perfectly correlated. The correlation coefficients between any two of these variables are between 0.38 and 0.45.

Table 5 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

[18]Since the regressors are "generated", they contain estimation errors, which may cause biases in the standard errors. In Online Appendix Section OA.2.4, we perform bootstrap estimation for two core regression tables in the paper and find such biases tend to be small. Due to computational constraints, we present simple standard errors clustered at the firm level in all of our regressions in the paper.

Table 4: 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).

not the firm fixed effects. Both nearby siblings' signal and firms' domestic sales raise the probability of FDI entry in the next period. A one standard deviation increase in nearby siblings' signal raises the entry probability by $1.59 \times 0.174‰ = 0.28‰$, which is around 25% of the average entry probability (1.1‰). 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 5: Impact of siblings' experience on entry in the next period

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)
Average nearby signal	0.174^a	0.180^a	0.172^a
	(0.032)	(0.038)	(0.040)
Average remote signal	0.041	0.042	0.018
	(0.040)	(0.054)	(0.057)
Firm domestic sales	0.066^c	-0.142	
	(0.035)	(0.108)	
Destination-Ind-Year FE	✓	✓	✓
Firm FE		✓	
Firm-Year FE			✓
N	875527	875527	902527
# of Firms	1922	1922	1931
# of Firm-Markets	113998	113998	115183
# of Entries	977	977	1003
R-squared	0.064	0.067	0.088

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.

In addition to our linear probability model, we show that the previous results are robust if we 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 (see Conconi et al. (2016)):

$$h_{fsk}(t|\mathbf{X}) = h_j(t) \exp\left(b_1 \overline{r^{\text{nearby}}_{fskt}} + b_2 \overline{r^{\text{remote}}_{fskt}} + b_3 \tilde{r}_{ft}\right), \tag{8}$$

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). 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 6 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 $e^{1.59\times0.167} - 1 = 30\%$. 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 6: 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
	(0.027)	(0.031)	(0.042)	(0.050)
Average remote signal	0.049	0.026	-0.038	-0.046
	(0.031)	(0.036)	(0.059)	(0.061)
Firm domestic sales	0.062^c	0.049		
	(0.032)	(0.034)		
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 5. The number of observations differs because we do not count singletons due to the fixed effects in the linear regressions.

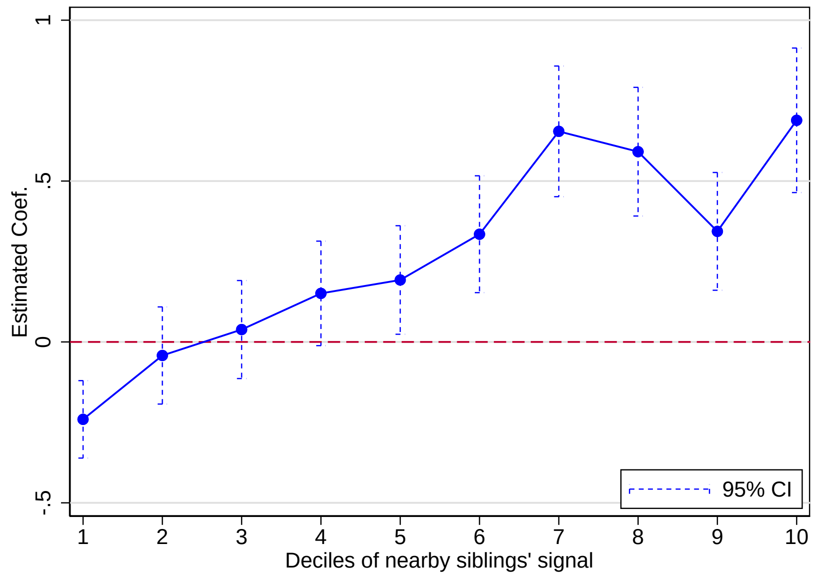
It is important to note that our evidence does not imply that the existence of a nearby sibling necessarily increases the likelihood of entry into other countries in the same region. Such a positive impact is realized only when the nearby siblings' signal is good enough. To demonstrate this point, we expand our sample to include regions which the firm has not entered yet. We then estimate the impact of 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 FDI context, 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 on 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 7, 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. We again find that only signals from siblings in the same region and industry can predict

[19] The details of the sample and regression results are presented in Online Appendix OA.2.2. 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: The coefficients of nearby siblings' signal decile dummies on entry probability, using 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 Online Appendix Table OA.3.

market entry. Therefore, learning effect is the strongest for this type of signals.[20]

Before we close this section and look at evidence based on affiliates' expectations post entry, we briefly discuss the potential mechanisms through which such learning effects operate. We first recognize that such a behavior is optimal from the MNC's perspective (Bayesian updating of beliefs). A MNC with frictionless information flows between subsidiaries and the parent company should use all the signals available and predict their future sales. Given that the signals are more correlated within a region, nearby siblings' signals are more important in predicting entry and expectations formation. We also conjecture that the pursuit of "regional strategies" by MNCs and the prevalence of regional headquarters (RHQs) may have further facilitated such information flows within a region ((Rugman and Verbeke, 2004)). For example, according to interviews with multiple MNCs and their RHQs, Nell et al. (2011) argue that RHQs play an

[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 7 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.

Table 7: Impact of siblings' experience on entry in the next period, horse race between signals from the same and different industries

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)	(4)	(5)	(6)
Avg nearby signal (same ind)	0.166^a	0.157^a	0.167^a	0.168^a	0.154^a	0.114^b
	(0.039)	(0.040)	(0.037)	(0.041)	(0.045)	(0.051)
Avg nearby signal (diff ind)	0.006	-0.019	0.009	-0.006	-0.020	-0.055
	(0.049)	(0.072)	(0.053)	(0.074)	(0.063)	(0.084)
Avg remote signal			0.004	0.136		
			(0.091)	(0.170)		
Avg remote signal (same ind)					0.045	0.021
					(0.062)	(0.070)
Avg remote signal (diff ind)					-0.014	-0.011
					(0.099)	(0.131)
Firm domestic sales	-0.005	0.025	-0.006	0.029	0.048	0.087
	(0.057)	(0.139)	(0.065)	(0.145)	(0.089)	(0.208)
Destination-Year FE	✓	✓	✓	✓	✓	✓
Industry-Year FE	✓	✓	✓	✓	✓	✓
Firm FE		✓		✓		✓
R-squared	0.01	0.02	0.01	0.02	0.02	0.02
N	458137	458136	447263	447261	319763	319763

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.

important role in facilitating the communication between regional affiliates and the parent firm. However, since we do not have proxies for information flows within the MNCs in our data, we cannot investigate this mechanism more thoroughly.

5.2 Life-cycle Learning 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:

$$\log E_t(R_{i,t+1}) = b_1\overline{r_{it}} + b_2\overline{r^{\text{nearby}}_{fskt}} + b_3\overline{r^{\text{remote}}_{fskt}} + b_4\tilde{r}_{ft} + \delta_{skt} + \delta_f + \epsilon_{i,t+1}, \tag{9}$$

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 (9) is almost the same as that of equation (7), except for the addition of the first regressor, $\overline{r_{it}}$. This variable is 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.

Column 1 of Table 8 presents the results from the baseline regression. The affiliate's own signal is a key determinant of future sales expectation, with a precisely estimated

Table 8: 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
	(0.011)	(0.024)	(0.026)	(0.009)
Average nearby signal	0.024^b	0.098^a	0.028	0.022^c
	(0.011)	(0.036)	(0.026)	(0.013)
Average remote signal	0.014	0.008	0.006	0.020
	(0.017)	(0.057)	(0.046)	(0.019)
Firm domestic sales	0.052^a	0.088^c	0.107^a	0.054^b
	(0.019)	(0.050)	(0.031)	(0.022)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
R-squared	0.88	0.88	0.89	0.90
N	32881	2182	3778	24160

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.

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 Part 1 of Proposition 2. In Columns 2 to 4 of Table 8, 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.

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 9. 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

Table 9: 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
	(0.009)	(0.019)	(0.010)	(0.020)
$\times$ log(self age)	0.091^a		0.092^a	
	(0.006)		(0.006)	
$\times \max\{\text{self age}, 10\}$		0.033^a		0.034^a
		(0.002)		(0.002)
Average nearby signal	0.023^b	0.181^a	0.034^a	0.200^a
	(0.012)	(0.026)	(0.012)	(0.029)
$\times$ log(self age)	-0.050^a		-0.052^a	
	(0.008)		(0.009)	
$\times \max\{\text{self age}, 10\}$		-0.019^a		-0.020^a
		(0.003)		(0.003)
Average remote signal	0.018	0.017	0.017	0.016
	(0.017)	(0.017)	(0.025)	(0.025)
Firm domestic sales	0.053^a	0.056^a		
	(0.019)	(0.018)		
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓	✓		
Firm-Year FE			✓	✓
Age FE	✓	✓	✓	✓
R-squared	0.89	0.89	0.91	0.91
N	32872	32872	31724	31724

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). The logarithm of affiliate age is also standardized to facilitate the interpretation of the coefficients. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

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 logarithm of 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.

Before we end this subsection, we examine how siblings' experience affects the strength of learning and test the other prediction in Part 1 of Proposition 2. 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.

Table 10 reports the results of regressions where we add interaction terms between the siblings' experience and self/siblings' average signals. 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 9. The effect of siblings' experience on learning is in general smaller than that of the affiliate's own experience.

We now show that the magnitude of the estimated age effect on the strength of learning is consistent with our theoretical model using a simple calibration, taking advantage of our direct measure of sales forecasts. In Online Appendix OA.1.3, we derive closed-form expressions for the coefficients of the average self and nearby siblings' signal in the expectation updating formula:

$$\beta_1 = \frac{(1-\rho_{12}^2)\lambda_2 + 1/t_2}{(1+1/\lambda_1 t_1)(\lambda_2 + 1/t_2) - \rho_{12}^2\lambda_2} \tag{10}$$

$$\beta_2 = \frac{\sigma_{\theta 1}}{\sigma_{\theta 2}} \frac{\rho_{12}/t_1}{(\lambda_1 + 1/t_1)(1+1/\lambda_2 t_2) - \rho_{12}^2\lambda_1}. \tag{11}$$

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 an earlier paper (Chen et al., 2020), we provide estimates for these parameters which imply a signal-to-noise ratio of 1.86.[21] The average age of nearby siblings is 15 according to Table 4 in the paper. We estimated ρ_{12} to be 0.38, using the model-consistent approach discussed in Appendix A.1 (the value in the second row of Table A.1). We then plug $\lambda_1 = \lambda_2 = 1.86$, $t_2 = 15$ and $\rho_{12} = 0.38$

[21] 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.

Table 10: 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
	(0.010)	(0.010)	(0.010)	(0.010)
× Self experience	0.092^a	0.093^a	0.091^a	0.093^a
	(0.006)	(0.006)	(0.006)	(0.006)
× Nearby siblings' experience	-0.011^c	-0.005	-0.004	-0.000
	(0.006)	(0.007)	(0.007)	(0.008)
Average nearby signal	0.039^a	0.060^a	0.033^b	0.051^a
	(0.015)	(0.016)	(0.014)	(0.014)
× Self experience	-0.051^a	-0.053^a	-0.052^a	-0.053^a
	(0.009)	(0.009)	(0.009)	(0.009)
× Nearby siblings' experience	0.021^b	0.032^a	0.016^b	0.025^a
	(0.009)	(0.011)	(0.008)	(0.010)
Nearby siblings' experience	0.023	0.000	0.023	0.015
	(0.018)	(0.022)	(0.017)	(0.020)
Average remote signal	0.019	0.020	0.019	0.019
	(0.018)	(0.025)	(0.018)	(0.026)
Firm domestic sales	0.052^a		0.053^a	
	(0.019)		(0.019)	
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓		✓	
Firm-Year FE		✓		✓
Age FE	✓	✓	✓	✓
R-squared	0.89	0.91	0.89	0.91
N	32872	31724	32862	31714

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 logarithm of self age, while the nearby siblings' experience is measured by the logarithm 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.

into equations (10) and (11). The implied coefficients under different values of t_1 are presented in Table 11, quantitatively similar to those estimated in Table 8 of the paper. For example, the coefficient of average nearby siblings' signal is estimated to be 0.098 for age one to age three affiliates in the data, while the model implies this coefficient to be 0.141, 0.088 and 0.063 for age one, two and three affiliates, respectively.

Table 11: 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	0.614	0.761	0.827	0.864	0.888	0.905	0.918	0.927	0.935	0.941
Coef. of Nearby Signal	0.141	0.088	0.063	0.050	0.041	0.035	0.030	0.027	0.024	0.022

Notes: The coefficients are calculated according to equations (10) and (11), respectively. We choose the following parameter values in addition to t_1: $\sigma_{\theta 1} = \sigma_{\theta 2}$, $\rho_{12} = 0.38$, $\lambda_1 = \lambda_2 = 1.86$, $t_2 = 15$.

5.3 Market Uncertainty and Cross-market Learning

In this subsection, we explore how the relationship between the affiliate's expectation and its nearby siblings' signal varies with the level of uncertainty in the affiliate's market, according to Part 2 of Proposition 2. 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 $2\sigma_{\varepsilon 1}^2$. Second, sufficiently old affiliates have almost discovered θ, meaning that the only source of their forecast errors is the temporary shock ε_t. Table 8 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 $\sigma_{\varepsilon 1}^2$. We also experiment with residual log sales growth and residual forecast errors from which we have removed destination-industry-year fixed effects (capturing aggregate shocks to all affiliates in the same market). The results are very similar.

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

$$\begin{aligned} \log E_t(R_{i,t+1}) \;\; = \;\; & b_1\overline{r_{it}} + b_2\overline{r_{fskt}^{\text{nearby}}} + b_3\overline{r_{fskt}^{\text{remote}}} + b_4\tilde{r}_{ft} \\ & b_5\overline{r_{it}} \times \hat{\sigma}_{\varepsilon 1,k} + b_6\overline{r_{fskt}^{\text{nearby}}} \times \hat{\sigma}_{\varepsilon 1,k} + \delta_{skt} + \delta_f + \epsilon_{i,t+1}. \end{aligned} \tag{12}$$

Our new estimation equation is equation (9) 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 destinations that have at least 20 observations of sales growth or forecast errors.[22] Proposition 2 predicts that b_5 and b_6 is are negatively and positively significant respectively.

Table 12 reports the regression results. In Columns 1 and 2, we approximate $\hat{\sigma}_{\varepsilon 1,k}$ 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

[22]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 12. Columns 3 and 4 construct this measure using the standard deviation of the forecast errors for affiliates above seven years, as discussed earlier. The signs of the two interaction terms are the same, but the magnitude of the coefficients falls. Finally, Columns 5–8 show that the results are robust when we use the standard deviation of residual sales growth or forecast errors, which exclude the systemic influence of destination-industry level trends.

Table 12: 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
	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)
$\times$ proxy of $\sigma_{\varepsilon 1}$	-0.056^a	-0.054^a	-0.028^a	-0.028^a	-0.056^a	-0.054^a	-0.028^a	-0.028^a
	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)
Average nearby signal	0.028^b	0.039^a	0.026^b	0.036^a	0.028^b	0.039^a	0.026^b	0.036^a
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
$\times$ proxy of $\sigma_{\varepsilon 1}$	0.035^a	0.037^a	0.022^a	0.019^a	0.035^a	0.037^a	0.022^a	0.019^a
	(0.009)	(0.009)	(0.007)	(0.007)	(0.009)	(0.009)	(0.007)	(0.007)
Average remote signal	0.018	0.019	0.016	0.017	0.018	0.019	0.016	0.017
	(0.017)	(0.024)	(0.017)	(0.025)	(0.017)	(0.024)	(0.017)	(0.025)
Firm domestic sales	0.054^a		0.054^a		0.054^a		0.054^a	
	(0.018)		(0.019)		(0.018)		(0.019)	
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Firm FE	✓		✓		✓		✓	
Firm-Year FE		✓		✓		✓		✓
Age FE	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.88	0.90	0.88	0.90	0.88	0.90	0.88	0.90
N	32872	31724	32855	31707	32872	31724	32853	31704

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 by the standard deviation of (residual) sales growth rates or that of (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.

5.4 Additional Evidence and Robustness

We discuss two robustness checks of our main empirical results in this subsection and refer the reader to the appendix for detailed regression tables. We also briefly discuss additional empirical results related to "learning from exporting" (Conconi et al., 2016;

Chen et al., 2020) and the impact of signals on affiliate exits. We present the additional results in the Online Appendix.

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 A.2, 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. This suggests that conditional on market uncertainty, the income levels of the focal markets do not affect the strength of learning.

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 a 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 robustness checks by restricting our sample to new entrants that have a small regional or global import shares. Specifically, 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 Appendix Table A.3 report the entry regressions with the restricted sample. 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 expectations formation regressions are also robust to using these two definitions of horizontal MP, which are reported in Columns 3–6 of Appendix Table A.4. Finally, as discussed in footnote 13, we select our sample for expectations formation regressions based on current local sales ratio, which may fluctuate over affiliates' life cycles. In Columns 7-8 of Table A.4, we select the sample based on the life-cycle average local sales ratio as for our entry regressions. We lose about 5% of the sample but the results are similar to those in Table A.2.

The literature has emphasized exporting as a mechanism through which MNCs can learn about the demand in potential markets (Conconi et al., 2016; Chen et al., 2020). The idea is that firms enter a foreign market by exporting first and "upgrading" to MP when the expected profitability is sufficiently high. They provide evidence that the number of years with export experience is positively associated with FDI entry and uncertainty reduction. We do not attempt to provide a full-fledged model that features both learning from exporting and learning from siblings. We do, however, provide alternative and complementary evidence to the literature in Online Appendix OA.2.1, in the spirit of the entry regressions presented in Section 5.1. We construct export "signals" in similar ways as siblings' signals, with the caveat that the measurement of arms-length exports by the parent firms in our data are far from being ideal. With this caveat in mind, we find that both the nearby siblings' signals and export experience (i.e., signals) increase the probability that a firm enters the new market in the same region, supporting both the "learning from exporting" and the "learning from siblings" mechanisms.

Finally, our paper focuses on expectations formation and a simple entry problem but abstracts from the full dynamics of multi-market entries and exits of MNCs, mainly due to the theoretical complexity of the problem (Tintelnot, 2017; Arkolakis and Eckert, 2017). In the Online Appendix OA.2.9, we provide discussions regarding affiliates' endogenous exits in a single market. We show that, under our simplifying assumptions, the affiliate's exit rate decreases in self and nearby siblings' signals. When taking this prediction to the data, we find that both the self and nearby siblings' signals have a negative impact on the probability of exits. However, the coefficient of the nearby siblings' signals is insignificant, which lends a weak support to the predictions of our extended.[23]

[23] A practical question here is how we measure "exits" of affiliates in the data. In the survey, some affiliates do respond in the year when they exit, and report that their status as "operation suspended"

6 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 key 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 to 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. We leave these promising approaches and interesting questions to future research.

or "dissolution or withdrawal" or "decline in control share" (below 10%). However, we are concerned that this strict definition of "exit" understates the overall exit rates because other affiliates may just stop responding when they exit. We therefore use two more general definitions of exits by including affiliates that stopped responding for at least two consecutive years (and plus those that report zero sales for at least two consecutive years). The exit rates are 7.29% and 7.31% under the two definitions, respectively.

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.

Arkolakis, Costas and Fabian Eckert, "Combinatorial Discrete Choice," Working Paper 2017.

__ , 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 Expectation Formation over the Firms' Life Cycle," CESifo Working Paper 8468 2020.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar, "How Do Firms Form Their Expectations? New Survey Evidence," *American Economic Review*, September 2018, *108* (9), 2671–2713.

— , — , **and Tiziano Ropele**, "Inflation Expectations and Firm Decisions: New Causal Evidence," *The Quarterly Journal of Economics*, February 2020, *135* (1), 165–219.

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.

Conley, Timothy G. and Christopher R. Udry, "Learning about a New Technology: Pineapple in Ghana," *American economic review*, 2010, *100* (1), 35–69.

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.

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.

Foster, Andrew D. and Mark R. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of political Economy*, 1995, *103* (6), 1176–1209.

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.

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**, "Learning by Doing and the Choice of Technology," *Econometrica*, November 1996, *64* (6), 1299.

— **and** — , "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.

Malmendier, Ulrike and Stefan Nagel, "Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?," *The Quarterly Journal of Economics*, 2011, *126* (1), 373–416.

— **and** — , "Learning from Inflation Experiences," *The Quarterly Journal of Economics*, February 2016, *131* (1), 53–87.

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

Nell, Phillip C., Björn Ambos, and Bodo B. Schlegelmilch, "The Benefits of Hierarchy? – Exploring the Effects of Regional Headquarters in Multinational Corporations," in Christian Geisler Asmussen, Torben Pedersen, Timothy M. Devinney, and Laszlo Tihanyi, eds., *Dynamics of Globalization: Location-Specific Advantages or Liabilities of Foreignness?*, Vol. 24 of *Advances in International Management*, Emerald Group Publishing Limited, January 2011, pp. 85–106.

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.

Rugman, Alan M and Alain Verbeke, "A Perspective on Regional and Global Strategies of Multinational Enterprises," *Journal of International Business Studies*, January 2004, *35* (1), 3–18.

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.

A Appendix

A.1 Within- and Cross-region Correlations in θ and ε

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 firm-market combination, 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.1 shows the within-region and cross-region correlations, pooling all within-region pairs and cross-region pairs, 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 global firm-level shocks that are not firm-market-specific. To address this issue, we compute two alternative proxies for θ. First, we remove the destination-industry-year fixed effects from log sales, so that the residual $\hat{e}_1$(sales) is arguably idiosyncratic demand. We then calculate the average of $\hat{e}_1$(sales) of affiliates that are at least seven years old within each firm-market. 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 global firm-level 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.1 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.

In our main model, we assume that the idiosyncratic transitory shocks ε_{jt} are i.i.d. over time and across affiliates. We now calculate the contemporary correlation of ε_{jt} between siblings. According to our model, the first difference in log sales is exactly $\Delta\varepsilon_{jt} \equiv \varepsilon_{jt} - \varepsilon_{j,t-1}$. We have used this result to estimate the variance σ_ε in Section 5.3.

Table A.1: Correlation of demand within and between regions for affiliates above age seven

Demand Measure	Corr. within Region	Corr. between Regions
Panel A: Corr(θ_i, θ_j)		
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]
Panel B: Corr($\varepsilon_{it}, \varepsilon_{jt}$)		
Δlog(sales)	0.096	0.062
	[166223]	[362710]
$\Delta\hat{e}_1$(sales)	0.043	0.025
	[156719]	[319896]
$\Delta\hat{e}_2$(sales)	0.099	0.068
	[160221]	[347933]

Notes: In Panel A, each observation is a firm-country-pair combination (two different countries). For each firm-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 of each affiliate. When the demand measure is $\hat{e}_1$(sales), we regress log local sales on the destination-industry-year fixed effects and use the residual to measure an affiliate'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. In Panel B, each observation is a affiliate-sibling-year combination (two different siblings). We calculate the correlation between the siblings of their growth in raw sales, growth in residual sales $\hat{e}_1$(sales) (controlling for destination-industry-year fixed effects) and growth in residual sales $\hat{e}_2$(sales) (further controlling parent sales). All the correlation coefficients are significant at 1%.

We now compute the correlation of this first difference between two siblings, i and j

$$Corr(\Delta\varepsilon_{it}, \Delta\varepsilon_{jt}) = \frac{Cov(\Delta\varepsilon_{it}, \Delta\varepsilon_{jt})}{\sqrt{Var(\Delta\varepsilon_{it})Var(\Delta\varepsilon_{jt})}} = \frac{Cov(\varepsilon_{it}, \varepsilon_{jt})}{\sqrt{Var(\varepsilon_{it})Var(\varepsilon_{jt})}} = Corr(\varepsilon_{it}, \varepsilon_{jt}),$$

where we have applied that $\{\varepsilon_{it}\}_i$ are independently and identically distributed over time (but not across affiliates). Similar to our calculation of θ_j, we consider three sales measures: log local sales, residual log local sales controlling for destination-industry-year fixed effects, $\hat{e}_1$sales, and residual log local sales further controlling for parent sales $\hat{e}_2$sales. We then take first difference and create all possible sibling pairs for each year and compute the correlation of the first differences. The number of observations is much larger than that used for calculating $Corr(\theta_i, \theta_j)$ since we do not aggregate over time and across affiliates within a destination. In general, we observe much lower within-region correlation than that of θ_j, ranging from 0.043 to 0.099.

A.2 Additional Robustness Checks

Table A.2: 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
	(0.010)	(0.010)	(0.010)	(0.010)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.026^a	-0.029^a		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.013^b	-0.010^c
			(0.006)	(0.006)
$\times$ log(self age)	0.086^a	0.086^a	0.091^a	0.088^a
	(0.007)	(0.007)	(0.006)	(0.007)
$\times$ Nearby siblings' experience	0.004	0.003	0.001	0.004
	(0.008)	(0.008)	(0.007)	(0.008)
$\times$ Destination income level		-0.004		0.014
		(0.011)		(0.009)
Average nearby signal	0.051^a	0.050^a	0.050^a	0.051^a
	(0.014)	(0.015)	(0.014)	(0.015)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.020^b	0.024^b		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.009	0.008
			(0.006)	(0.006)
$\times$ log(self age)	-0.047^a	-0.048^a	-0.052^a	-0.049^a
	(0.010)	(0.010)	(0.010)	(0.010)
$\times$ Nearby siblings' experience	0.023^b	0.024^b	0.024^b	0.024^b
	(0.010)	(0.010)	(0.010)	(0.010)
$\times$ Destination income level		0.006		-0.008
		(0.015)		(0.012)
Nearby siblings' experience	0.013	0.014	0.014	0.014
	(0.020)	(0.020)	(0.020)	(0.020)
Average remote signal	0.020	0.021	0.020	0.021
	(0.026)	(0.026)	(0.026)	(0.026)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	31714	31599	31697	31582
R-squared	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.

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

Def. of Horizontal Entry	Avg Local Sales Share ≥ 0.95		Avg Regional Import Share < 0.15		Avg Import Share < 0.15	
Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)	(4)	(5)	(6)
Average nearby signal	0.144^{a}	0.139^{a}	0.101^{a}	0.115^{a}	0.098^{a}	0.109^{a}
	(0.032)	(0.036)	(0.031)	(0.032)	(0.030)	(0.031)
Average remote signal	0.033	0.003	0.021	0.036	0.019	0.033
	(0.045)	(0.048)	(0.041)	(0.042)	(0.041)	(0.041)
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓
Firm FE	✓		✓		✓	
Firm-Year FE		✓		✓		✓
R-squared	0.06	0.08	0.07	0.09	0.07	0.09
N	902532	902527	902532	902527	902532	902527

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. The "average" is calculated withn an affiliate over time. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

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

Def. of Horizontal Affiliates	Local Sales Share ≥ 0.95		Avg Regional Import Share < 0.15		Avg Import Share < 0.15		Avg Local Sales Share ≥ 0.85	
Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average self signal	0.873^a	0.872^a	0.871^a	0.871^a	0.874^a	0.875^a	0.849^a	0.848^a
	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)	(0.013)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.026^a	-0.030^a	-0.031^a	-0.033^a	-0.032^a	-0.041^a	-0.020^b	-0.023^b
	(0.008)	(0.010)	(0.010)	(0.013)	(0.010)	(0.013)	(0.009)	(0.012)
$\times$ Self experience	0.088^a	0.089^a	0.092^a	0.092^a	0.096^a	0.097^a	0.084^a	0.085^a
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
$\times$ Nearby siblings' experience	0.008	0.007	0.011	0.010	0.013	0.011	-0.002	-0.003
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$\times$ Destination income level		-0.004		-0.003		-0.011		-0.004
		(0.012)		(0.013)		(0.014)		(0.014)
Average nearby signal	0.046^a	0.046^a	0.055^a	0.054^a	0.061^a	0.060^a	0.054^a	0.055^a
	(0.014)	(0.015)	(0.017)	(0.019)	(0.017)	(0.018)	(0.016)	(0.017)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.018^b	0.020^b	0.007	0.011	0.006	0.013	0.026^a	0.023^b
	(0.008)	(0.009)	(0.011)	(0.012)	(0.013)	(0.013)	(0.009)	(0.011)
$\times$ Self experience	-0.053^a	-0.054^a	-0.056^a	-0.056^a	-0.058^a	-0.059^a	-0.041^a	-0.041^a
	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
$\times$ Nearby siblings' experience	0.020^b	0.021^b	0.027^b	0.027^b	0.027^b	0.027^b	0.033^a	0.032^a
	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
$\times$ Destination income level		0.004		0.006		0.009		-0.005
		(0.015)		(0.019)		(0.019)		(0.016)
Nearby siblings' experience	0.009	0.010	0.004	0.005	-0.008	-0.007	0.042^c	0.043^c
	(0.021)	(0.021)	(0.022)	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)
Average remote signal	0.004	0.005	0.030	0.031	0.017	0.018	-0.004	-0.004
	(0.025)	(0.025)	(0.026)	(0.026)	(0.027)	(0.027)	(0.021)	(0.021)
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
N	26216	26102	24065	24021	23285	23247	30094	29998

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. Columns 3 to 6 select the sample based on the average of the corresponding shares over the affiliates' life cycles. Significance levels: a: 0.01, b: 0.05, c: 0.10.

Online Appendix for "Learning and Information Transmission within Multinational Corporations": Not for Publication

Cheng Chen Chang Sun Hongyong Zhang*

Sep 2021

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

OA.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 $\rho_{13} = \rho_{23} = 0$.

OA.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.

OA.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

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

OA.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:

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

OA.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

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

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

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

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

OA.1.3 Proof of Proposition 2

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

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

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 $E_{t-1}\left(A_{1t}\left[\varsigma w_{1t}/(\varsigma - 1)\right]^{1-\varsigma}\right)$. Since the posterior of a_{1t} is normal with mean $\bar{\mu}$ and variance $\bar{\Sigma} + \sigma_{\varepsilon 1}^2$ as discussed in Section OA.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

$$\beta_1 = \frac{(1-\rho_{12}^2)\lambda_2 + 1/t_2}{(1+1/\lambda_1 t_1)(\lambda_2 + 1/t_2) - \rho_{12}^2\lambda_2} \tag{13}$$

$$\beta_2 = \frac{\sigma_{\theta 1}}{\sigma_{\theta 2}}\frac{\rho_{12}/t_1}{(\lambda_1 + 1/t_1)(1+1/\lambda_2 t_2) - \rho_{12}^2\lambda_1} \tag{14}$$

$$\beta_3 = 0,$$

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

$$\bar{\mu} = \bar{\theta}_1 + \beta_1(\bar{a}_1 - \bar{\theta}_1) + \beta_2(\bar{a}_2 - \bar{\theta}_2), \tag{15}$$

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

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

■

OA.1.4 Correlated Temporary Shocks

A convenient and probably unrealistic assumption of the our model is that temporary demand shocks are uncorrelated between the focal affiliate and its siblings in the same region. One interesting modification of our baseline model is to allow the temporary shocks to be positively correlated across destination economies within the same region (i.e., the assumption we impose on time-invariant demand draws). In the remaining part of the model section, we consider a more realistic case in which temporary demand shocks are positively correlated within the region but not across regions. In general, it is hard to analyze comparative statics of the learning parameters with respect to the correlation of temporary demand shocks within the region, which forces us to prove Proposition 1 under parameter assumptions.

Proposition OA 1 *Assume that temporary demand shocks in markets 1 and 2 are positively correlated with a positive correlation coefficient of* $\rho_{12}^e (> 0)$. *Furthermore, we assume that*

ρ_{12} is not too large and λ_1 is not too small. Therefore, we have

1. *The weight the focal affiliate put on its nearby sibling's average signal decreases with the correlation coefficient of ρ^e_{12}.*

2. *The weight the focal affiliate put on its own average signal increases with the correlation coefficient of ρ^e_{12}.*

Proof. Since we assume there is no correlation of both time-invariant demand draws and temporary demand shocks across regions, the conditional mean of θ_1 given $\bar{a}_1, \bar{a}_2, \bar{a}_3$ can be expressed as

$$\bar{\mu} = \bar{\theta}_1 + \begin{bmatrix} \sigma^2_{\theta 1} & \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} \end{bmatrix} \begin{bmatrix} \sigma^2_{\theta 1} + \sigma^2_{\varepsilon 1}/t_1 & \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} + \frac{\rho^e_{12}\sigma_{\varepsilon 1}\sigma_{\varepsilon 2}}{\max\{t_1,t_2\}} \\ \rho_{12}\sigma_{\theta 1}\sigma_{\theta 2} + \frac{\rho^e_{12}\sigma_{\varepsilon 1}\sigma_{\varepsilon 2}}{\max\{t_1,t_2\}} & \sigma^2_{\theta 2} + \sigma^2_{\varepsilon 2}/t_2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{a}_1 - \bar{\theta}_1 \\ \bar{a}_2 - \bar{\theta}_2, \end{bmatrix}$$

which lead to the results that

$$\beta_1 = \frac{(1-\rho^2_{12})\lambda_2 + \frac{1}{t_2} - \rho_{12}\rho^e_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}}{(1+\frac{1}{\lambda_1 t_1})(\lambda_2 + \frac{1}{t_2}) - \rho^2_{12}\lambda_2 - 2\rho_{12}\rho^e_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0} - \left(\rho^e_{12}\right)^2\frac{1}{\lambda_1 t_0^2}} \tag{16}$$

$$\beta_2 = \frac{\sigma_{\theta 1}}{\sigma_{\theta 2}}\left(\frac{\frac{\rho_{12}\lambda_2}{\lambda_1 t_1} - \rho^e_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}}{(1+\frac{1}{\lambda_1 t_1})(\lambda_2 + \frac{1}{t_2}) - \rho^2_{12}\lambda_2 - 2\rho_{12}\rho^e_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0} - \left(\rho^e_{12}\right)^2\frac{1}{\lambda_1 t_0^2}}\right), \tag{17}$$

where

$$t_0 \equiv \max\{t_1, t_2\}.$$

Note that the common denominator in equations (16) and (17) is positive for sure. The first thing ot notice is that β_1 and β_2 can be negative now. This is more likely to happen when ρ^e_{12} is large. There two forces here. First, the focal affiliate wants to incorporate $\bar{a}_2$ into its forecast of θ_1 in a positive way, as θ_1 and θ_2 are positively correlated. At the same time, the focal affiliate also wants to tease out the series of temporary shocks in market 1 when forming the expectation for θ_1. When the i.i.d. temporary shocks in the two markets are highly and positively correlated, the focal affiliate can do so by taking the different between $\bar{a}_1$ and $\bar{a}_2$ which implies that the weight on $\bar{a}_2$ is negative.

For the weight put on self signal, we have

$$
\begin{aligned}
& Sign\left(\frac{\partial \beta_1}{\partial \rho_{12}^e}\right) \\
= \; & Sign\Bigg[-\rho_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\left((1+\frac{1}{\lambda_1 t_1})(\lambda_2+\frac{1}{t_2})-\rho_{12}^2\lambda_2-2\rho_{12}\rho_{12}^e\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}-\frac{\left(\rho_{12}^e\right)^2}{\lambda_1 t_0^2}\right) \\
& +\left(2\rho_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}+\frac{2\rho_{12}^e}{\lambda_1 t_0^2}\right)\left((1-\rho_{12}^2)\lambda_2+\frac{1}{t_2}-\rho_{12}\rho_{12}^e\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\right)\Bigg] \\
> \; & 0,
\end{aligned}
$$

when $\rho_{12} = 0.38$ and $\lambda_1 = \lambda_2 = 1.86$ (i.e., the calibrated values). In general, β_1 increases with ρ_{12}^e as long as ρ_{12} is not too large and λ_1 is not too small. The following condition is a sufficient condition:

$$(1-\rho_{12}^2)\lambda_2+\frac{1}{t_2} \geq \left(\lambda_2+\frac{1}{t_2}\right)\frac{1}{\lambda_1 t_1}.$$

However, when ρ_{12} is extremely large and λ_1 is extremely small, β_1 decreases with ρ_{12}^e.

For the weight put on nearby sibling's signal, we have

$$
\begin{aligned}
& Sign\left(\frac{\partial \beta_2}{\partial \rho_{12}^e}\right) \\
= \; & Sign\Bigg[-\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\left((1+\frac{1}{\lambda_1 t_1})(\lambda_2+\frac{1}{t_2})-\rho_{12}^2\lambda_2-2\rho_{12}\rho_{12}^e\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}-\frac{\left(\rho_{12}^e\right)^2}{\lambda_1 t_0^2}\right) \\
& +\left(2\rho_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}+\frac{2\rho_{12}^e}{\lambda_1 t_0^2}\right)\left(\frac{\rho_{12}\lambda_2}{\lambda_1 t_1}-\rho_{12}^e\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\right)\Bigg].
\end{aligned}
$$

Note that

$$
\begin{aligned}
& \Bigg[-\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\left((1+\frac{1}{\lambda_1 t_1})(\lambda_2+\frac{1}{t_2})-\rho_{12}^2\lambda_2-2\rho_{12}\rho_{12}^e\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}-\frac{\left(\rho_{12}^e\right)^2}{\lambda_1 t_0^2}\right) \\
& +\left(2\rho_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}+\frac{2\rho_{12}^e}{\lambda_1 t_0^2}\right)\left(\frac{\rho_{12}\lambda_2}{\lambda_1 t_1}-\rho_{12}^e\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\right)\Bigg] \\
= \; & \Bigg[-\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\left((1+\frac{1}{\lambda_1 t_1})(\lambda_2+\frac{1}{t_2})-\rho_{12}^2\lambda_2\right)+\left(2\rho_{12}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}+\frac{2\rho_{12}^e}{\lambda_1 t_0^2}\right)\frac{\rho_{12}\lambda_2}{\lambda_1 t_1}-\frac{\left(\rho_{12}^e\right)^2}{\lambda_1 t_0^2}\frac{\sqrt{\lambda_2/\lambda_1}}{t_0}\Bigg] \\
< \; & 0,
\end{aligned}
$$

when $\rho_{12} = 0.38$ and $\lambda_1 = \lambda_2 = 1.86$ (i.e., the calibrated values). In general, β_1 decreases with ρ_{12}^e as long as ρ_{12} is not too large and λ_1 is not too small. Otherwise, β_1 would increase with ρ^e_{12}. ■

Although we cannot find the exact range of parameter values in which the sign of comparative statics is unambiguously positive or negative, we can gain insights by considering a special case in which temporary shocks are perfectly correlated within the region and the focal affiliate and its nearby sibling are at the same age. In such a case, the difference between two average signals in the same region is simply $\bar{a}_1 - \bar{a}_2 = \theta_1 - \theta_2$ (where $\beta_1 = 1$ and $\beta_2 = -1$ in the formula of Bayesian updating), as the temporary shocks that have hit the two affiliates are perfectly canceled out. In addition, if we assume that there is no uncertainty concerning the nearby sibling's time-invariant demand draw (i.e., $\sigma^2_{\theta 2} = 0$), the focal affiliate can infer its time-invariant demand draw *perfectly* by taking the difference between the two average signals. In other words, the information value provided by the nearby sibling's signal is to tease out common temporary shocks, which leads to a negative coefficient of β_2 in the formula of Bayesian updating (if the time-invariant demand draws are uncorrelated). This insight has been pointed out in studies of tournament games and games of relative performance evaluation.[1] As the time-invariant demand draws are still correlated in our extended model, what we can show is that when the correlation of temporary demand shocks increases, the motive of doing "relative performance evaluation" (between the focal affiliate and the nearby sibling) becomes stronger.

Turning to the empirical side, we have to make it clear that the temporary shocks we are considering are firm-specific shocks. Thus, we can use residual sales to tease out aggregate persistent or temporary shocks that can be either correlated or uncorrelated across markets. According to the model, sufficiently old firms have almost learned the value of θ (the time-invariant demand shock) and the change in their residual sales over time (i.e., sales growth) is only caused by the temporary shocks, ε_{it}. Therefore, we calculate the growth rate of residual sales and correlate them across affiliates in different countries within the same multinational parent firm.[2] As a result, we obtain several measures for the correlation of firm-specific temporary shocks both within and across regions in Panel B, Table A.1 of the paper. There are several points that are worth mentioning. First, the correlations of temporary demand shocks are indeed positive and smaller when we focus on between-region correlations. Importantly, when we focus on the within-region correlation of temporary demand shocks, we find that it is much smaller than the within-region correlation of time-invariant demand draws. We also calculated proxies of ρ^e_{12} between countries within each region in Table OA.1. These values range from 0.03 to 0.2.

[1] Specifically, the value of doing a tournament game or relative performance evaluation is that common random shocks (i.e., lucks) that affect all agents' performance can be teased out by comparing performance between different agents.

[2] The methodology is documented in Section A.1 of the paper.

Table OA.1: Correlation of temporary idiosyncratic demand within each region

Demand Measure	Asia	North America	Latin America	Europe	Others
Δlog(sales)	0.079	0.110	0.073	0.178	0.193
	[127527]	[5437]	[2410]	[30058]	[791]
$\Delta\hat{e}_1$(sales)	0.037	0.062	0.066	0.062	0.132
	[123952]	[5180]	[1282]	[25839]	[466]
$\Delta\hat{e}_2$(sales)	0.086	0.097	0.104	0.149	0.203
	[122845]	[5386]	[2247]	[29129]	[614]

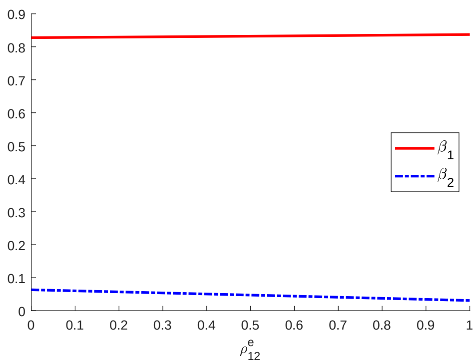
Notes: Each observation is an affiliate-sibling-year combination (two different siblings in two different countries within a particular region). The proxies for the temporary idiosyncratic demand shock are explained in the notes of Table A.1 of the paper. All the correlation coefficients are significant at 1%.

Based on those empirical estimates, we calculated the two weights, β_1 and β_2 over the focal affiliate's life cycles by imposing that $\rho^e_{12} = 0.03$ or $\rho^e_{12} = 0.2$ in Panel (a) of Figure OA.2. The curves for $\rho^e_{12} = 0.2$ and for $\rho^e_{12} = 0.03$ are almost indistinguishable from each other. This is even true when we consider more extreme values of the correlation, i.e., $\rho^e_{12} = 0.5$ and $\rho^e_{12} = 0$ in Panel (b). In these two panels, we have assumed the signal-to-noise ratios $\lambda_1 = \lambda_2 = 1.86$ as Chen et al. (2020). In Panels (c) and (d), we multiply λ_1 and λ_2 by three, respectively. Increasing λ_1 does affect the speed of learning, but it barely affects the difference between the case of low ρ^e_{12} and high ρ^e_{12}.

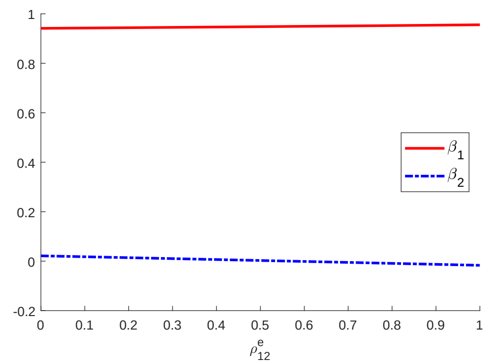
In total, we conclude that having a reasonable level of the correlation in temporary shocks only causes a quantitatively small bias in our estimated coefficients of self and sibling signals on expectation formation.

In total, we conclude that having a reasonable level of the correlation in temporary shocks only causes a quantitatively small bias in our estimated coefficients of self and sibling signals on expectation formation.

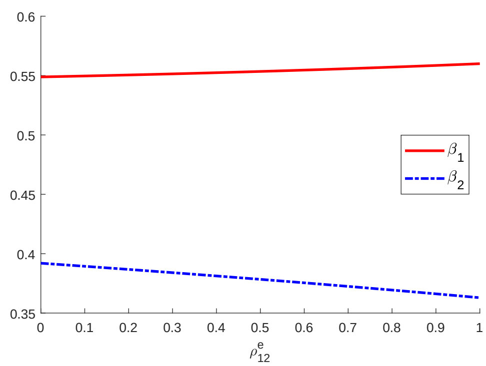
Figure OA.1: Correlation of temporary shocks and Learning Parameters



(a) $t_1 = 10, \rho_{12} = 0.38$



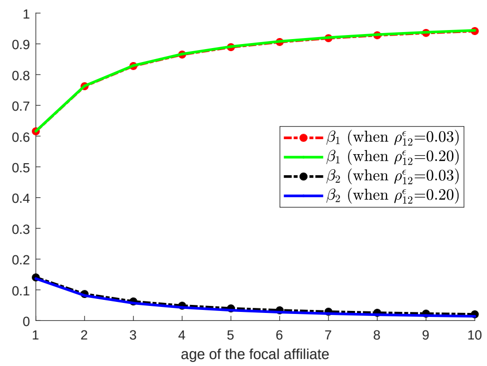
(b) $t_1 = 3, \rho_{12} = 0.38$



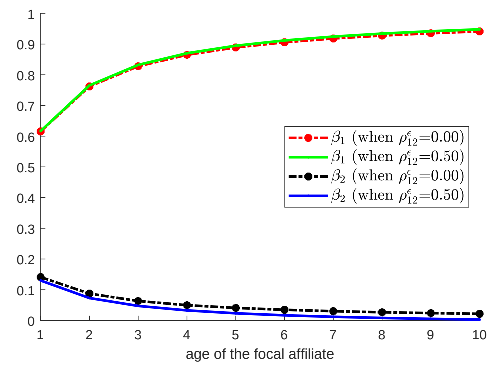
(c) $t_1 = 10, \rho_{12} = 0.9$

Notes: Other paramters $\lambda_1 = \lambda_2 = 1.86, t_2 = 15$.

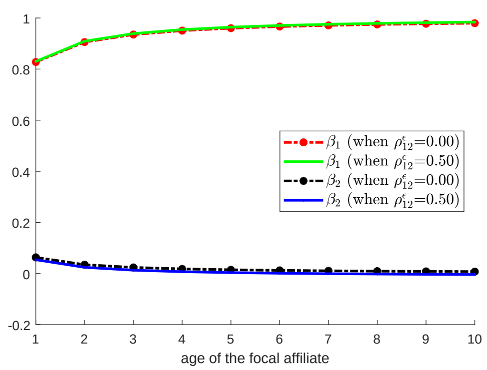
Figure OA.2: Correlation of temporary shocks and Learning over Life-cycle: correlation coefficients consistent with the data



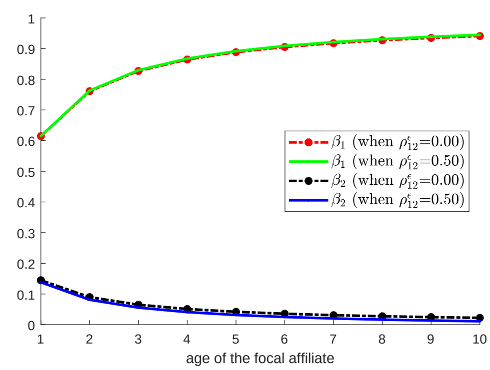
(a) Empirically relevant ρ^e_{12}, $\lambda_1 = \lambda_2 = 1.86$



(b) Extreme ρ^e_{12}, $\lambda_1 = \lambda_2 = 1.86$



(c) Extreme ρ^e_{12}, $\lambda_2 = 1.86, \lambda_1 = 3\lambda_2$



(d) Extreme ρ^e_{12}, $\lambda_1 = 1.86, \lambda_2 = 3\lambda_1$

Notes: Other parameters $\sigma_{\theta 1} = \sigma_{\theta 2} = 1.8, t_2 = 15, \rho_{12} = 0.38$.

OA.1.5 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

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

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

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

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

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

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

$$A \equiv \frac{\sigma_{\theta 1}\rho_{12}}{\sigma_{\theta 2}} \frac{\lambda_2 t_2}{1+\lambda_2 t_2}\left((\bar{a}_2-\bar{\theta}_2)-\frac{\sigma_{\theta 1}\sigma_{\theta 2}\rho_{12}}{2}\right). \tag{18}$$

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 (18), 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.

OA.1.6 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. *[life-cycle learning] The weight it places on its own average signal (the average signal of siblings in market 2) increases (decreases) with its age, and decreases (increases) with the total number of signals from market 2.*

2. *[uncertainty impedes self-learning] 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).*

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

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

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.[3] Therefore, we must have

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

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

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

■

OA.1.7 Effects of Signals on Endogenous Exit

In this section, we extend our model and analyze how the signals affect the exit rate of the focal affiliate. We assume that in each period, an affiliate has to pay a fixed operating cost, f_x, to stay in the market.

[3] 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.

Proposition OA 2 *Both the affiliate's own average past signal and its nearby sibling's average past signal negatively affect the exit probability of the focal affiliate.*

Proof. The value function of the incumbent as

$$V(t, \bar{\mu}_{t-1}) = \max_{p_t} E_{t-1} p_t^{-\varsigma} A_{1t} e^{a_{1t}} (p_t - w_{1t}) + \max\{E_t \beta V(t+1, \bar{\mu}_t) - f_x, 0\},$$

where β is the discount factor of the firm. Note that the state variable $\bar{\mu}_{t-1}$ (posterior mean of θ) depends on the average past signal and thus the age of the focal affiliate, t. In addition, it also depends on the age of the nearby sibling which we omit here for simplicity. Importantly, the firm decides whether to stay in the market (and pay the fixed per-period operation cost) at the beginning of each period (before observing the signal of the current period). Therefore, the final value function at the end of period t is simply

$$\max\{E_t \beta V(t+1, \bar{\mu}_t) - f_x, 0\}.$$

Now we prove that $E_t V(t+1, \bar{\mu}_t)$ increases with both $\bar{a}_1(t_1)$ and $\bar{a}_2(t_2)$ where t_1 and t_2 are the focal affiliate's age and the nearby sibling's age at period t. Note that the only uncertain variable in the value function is a_{1t} and

$$E_t(e^{a_{1,t+1}}) = e^{\bar{\mu}_t + \left(\bar{\Sigma} + \sigma_{\varepsilon 1}^2\right)/2},$$

where $\bar{\mu}_t$ is defined in equation (15). As $\bar{\mu}_t$ increases in $\bar{a}_1(t_1)$ and $\bar{a}_2(t_2)$ strictly, the expected per-period profit also increases in the two average signals strictly. Moreover, the choice set of p_{t+1} is the same, irrespective of the values of the two state variables in the value function.[4] Therefore, Theorem 4.7 of Stokey (1989) implies that the value function $E_t \beta V(t+1, \bar{\mu}_t)$ increases with $\bar{a}_1(t_1)$ and $\bar{a}_2(t_2)$. Accordingly, when $\bar{a}_1$ or $\bar{a}_2$ increases, the exit probability goes down. ■

[4]Note that the choice set of p_{t+1} is non-empty, compact-valued, and continuous with respect to $\bar{\mu}_t$. Also note that the expected profit function is bounded and continuous.

OA.2 Additional Empirical Results

OA.2.1 Learning from Exporting Experience

Several papers in the literature have emphasized the importance of MNE pre-entry exports to the market. Firms that are uncertain about the demand in a particular market can "test the market" by exporting, because the entry cost of exporting is likely to be lower than that of multinational production (MP). If the firms learn that their demand is high enough, they will establish a horizontal affiliate in that market. For example, Conconi et al. (2016) build a two-period model and show that under certain parameter values, firms enter a foreign market by exporting first and "upgrading" to MP when the expected profitability is sufficiently high. They provide evidence that the number of years of export experience is positively associated with FDI entry. Chen et al. (2020) build a multi-period dynamic model of export and MP and focus on predictions concerning forecasting errors. Using the same dataset as this paper, they show that affiliates whose parent firm has export experience before entry start with smaller forecast errors, consistent with the learning mechanism.

In this section, we provide alternative and complementary evidence to the literature, in the spirit of the entry regressions in Section 5.1 of the paper. We construct export "signals" in similar ways as siblings' signals, which is more informative about the level of demand in similar markets than indicators or the number of years of export experience. However, there are two caveats about the measurement of exports in the Japanese data. First, unlike Conconi et al. (2016), we only observe the parent firm's export to one of the seven regions (North America, Asia, Middle East, Europe, Latin America, Oceania and Africa), not its exports to a particular country. Second, the total exports to a particular region include exports to all countries, including those where the firm has entered as MNEs. Therefore, some of the exports may be intra-firm exports of intermediate inputs.

We cannot directly address the first caveat, but we argue that regional exports are informative about the overall level of demand at the region level. Given our assumption that nearby signals are correlated with the demand in the focal market, whether the exports are for consumers in the focal or nearby markets matters less. In this sense, the signals extracted from the regional arms-length exports are comparable to the nearby siblings' signals. For the second caveat, we try our best to exclude intra-firm exports. In our data, the existing affiliates report the total and intra-firm imports from Japan after 2009. We infer their intra-firm imports before 2009 by first calculating the average share of intra-firm imports in total imports from Japan across all affiliates of the same firm in the relevant regions after 2009, and multiply the total imports of an affiliate in a particular year before 2009 by that share. We exclude the intra-firm imports from the parent firm's total exports to the region, which represent the arms-length exports to the region. We also calculate a more conservative measure of arms-length exports to the regions by excluding all existing affiliates' import from Japan. This measure is actually quite close to the previous one since among Japanese affiliates, 90% of their imports from Japan are intra-firm.

With all the measurement caveats in mind, we first regress the log of parent exports by region and year fixed effects and obtain the residual, and use the cumulative average of these residual exports as a measure of the "average export signal". Table OA.2 replicates the regressions in Table 5 of the paper, controlling for average export signals. Both the nearby siblings' and export signals tend to increase the chance that a firm enters the new market in the same region. The effects of the export signals are significantly positive, and especially so when we control for firm-year fixed effects. The coefficients of the nearby siblings' signal are slightly smaller compared to those in Table 5, and the export signals are as quantitatively important as the nearby siblings' signals, though the coefficients are less precisely estimated. The results are robust regardless of whether we exclude intra-firm exports from the export measures.

In summary, we provide evidence that both the mechanism of learning from exporting and the mechanism of learning from nearby siblings exist in the Japanese data. The two mechanisms have similar quantitative importance regarding MP entry decisions.

Table OA.2: Impact of siblings' and export signals on entry in the next period

	All Parent Exports to Region		Exclude Siblings Imports from Japan		Exclude Intra-firm Imports from Japan	
Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)	(4)	(5)	(6)
Average nearby signal	0.136^a	0.153^a	0.150^a	0.153^a	0.152^a	0.155^a
	(0.036)	(0.045)	(0.036)	(0.046)	(0.035)	(0.045)
Average export signal	0.083^b	0.139^c	0.053	0.161^c	0.045	0.156^c
	(0.038)	(0.084)	(0.036)	(0.087)	(0.036)	(0.085)
Average remote signal	0.025	0.059	0.022	0.054	0.022	0.045
	(0.047)	(0.061)	(0.046)	(0.062)	(0.046)	(0.060)
Firm domestic sales	0.028		0.055		0.060	
	(0.046)		(0.045)		(0.045)	
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓
Firm-Year FE		✓		✓		✓
N	706487	718229	694723	706186	699979	711590
# of Firms	1551	1553	1541	1544	1547	1549
# of Firm-Markets	91846	92270	91042	91480	91466	91904
# of Entries	819	829	806	816	812	822
R-squared	0.062	0.086	0.062	0.087	0.062	0.087

Notes: Average export signal is the average of residual log exports, which in turn is obtained from a regression with year and region fixed effects. Different columns use different export measures. Columns 1 and 2 use the total export of parent firms to the region where the potential market belongs. Columns 3 and 4 exclude the imports of all existing affiliates in the region from Japan. Columns 5 and 6 exclude instead the intra-firm imports of these affiliates. The intra-firm imports are precise for years post 2009, but we impute the intra-firm imports before 2009 assuming that the share of intra-firm imports from Japan among all imports from Japan is the same as the average share of all sibling affiliates in the corresponding regions post 2009. Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

OA.2.2 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.[5]

If nearby siblings exist, we calculate their signal and group them into ten equally sized bins (deciles one to ten). We assign the observations with no nearby siblings 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 OA.3 shows, receiving signals in a higher decile tends to increase the entry probability, consistent with our findings in Table 5 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.

[5]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 OA.3: The impact of nearby siblings' signal on next period entry, using markets without nearby siblings as the base category

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)
Average nearby signal Q1	-0.241^a	-0.148^b
	(0.061)	(0.062)
Average nearby signal Q2	-0.042	0.063
	(0.077)	(0.081)
Average nearby signal Q3	0.039	0.123
	(0.078)	(0.078)
Average nearby signal Q4	0.151^c	0.272^a
	(0.083)	(0.086)
Average nearby signal Q5	0.193^b	0.295^a
	(0.086)	(0.088)
Average nearby signal Q6	0.335^a	0.432^a
	(0.093)	(0.093)
Average nearby signal Q7	0.654^a	0.758^a
	(0.104)	(0.107)
Average nearby signal Q8	0.591^a	0.713^a
	(0.102)	(0.100)
Average nearby signal Q9	0.344^a	0.503^a
	(0.093)	(0.097)
Average nearby signal Q10	0.689^a	0.851^a
	(0.115)	(0.121)
Firm domestic sales	-0.001	
	(0.009)	
Destination-Ind-Year FE	✓	✓
Firm-Ind FE	✓	
Firm-Ind-Year FE		✓
R-squared	0.02	0.02
N	13669307	13669307

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 5 of the paper because we include markets in regions where firms have no presence yet. These observations are used as the base category.

OA.2.3 Robustness to Heterogeneous Transmission and Exposure

In this section, we show that our results are robust to additional controls for heterogeneous transmission of parent shocks and parent firm heterogeneous exposures to aggregate shocks. In the paper, we sometimes control for parent- or MNE-level shocks using residual parent domestic sales. It is only an ideal control when the productivity shocks to the parent firms are transmitted to all affiliates at a constant rate. This is a stronger assumption than what the literature has assumed, i.e., a constant destination-specific transmission rate. (Ramondo and Rodríguez-Clare, 2013; Tintelnot, 2017; Arkolakis et al., 2018). In Column 1 of Table OA.4 and Columns 1 and 2 of Table OA.5, we control for interactions between residual parent domestic sales and destination-industry fixed effects, allowing the transmission to be destination-industry specific. Our main results are robust to this control.

We are also concerned that there may be parent- or firm-level shocks not captured by parent domestic sales, such as heterogeneous exposures to aggregate monetary and financial shocks. We postulate that such heterogeneous exposure is correlated with firm size and sufficiency of capital. Therefore, we control for parent firm size and their capital-labor ratios interacted with year fixed effects as a robustness check in Columns 2-4 of Table OA.4 and Columns 3-6 of Table OA.5. The results are very similar to those without these controls.

Table OA.4: Impact of siblings' experience on entry in the next period, controlling for heterogeneous transmission and HQ heterogeneous exposure to domestic shocks

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)	(4)	(5)
Average nearby signal	0.174^a	0.185^a	0.180^a	0.185^a	0.179^a
	(0.038)	(0.038)	(0.038)	(0.038)	(0.039)
Average remote signal	0.033	0.045	0.044	0.047	0.038
	(0.055)	(0.054)	(0.053)	(0.054)	(0.056)
Firm domestic sales		-0.159	-0.012	-0.021	
		(0.111)	(0.159)	(0.160)	
Destination-Ind-Year FE	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓
log Firm K/L × Year FE		✓		✓	✓
log Firm Sales × Year FE			✓	✓	✓
Destination FE × Domestic Sales	✓				✓
N	875527	863009	875527	863009	863009
R-squared	0.071	0.068	0.067	0.068	0.072

Notes: Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

Table OA.5: Full set of interaction terms in the expectation formation regressions, controlling for heterogeneous transmission and HQ heterogeneous exposure to domestic shocks

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)	(5)	(6)
Average self signal	0.867^a	0.866^a	0.869^a	0.868^a	0.869^a	0.869^a
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.031^a	-0.036^a	-0.030^a	-0.031^a	-0.030^a	-0.031^a
	(0.008)	(0.010)	(0.007)	(0.009)	(0.007)	(0.009)
× log(self age)	0.083^a	0.084^a	0.083^a	0.083^a	0.084^a	0.084^a
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
× Nearby siblings' experience	0.002	0.001	0.001	0.000	0.001	0.001
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
× Destination income level		-0.006		-0.002		-0.001
		(0.012)		(0.011)		(0.011)
Average nearby signal	0.029^b	0.029^c	0.032^b	0.031^b	0.034^b	0.033^b
	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.012	0.016	0.017^b	0.021^b	0.018^b	0.021^b
	(0.009)	(0.010)	(0.008)	(0.009)	(0.008)	(0.009)
× log(self age)	-0.046^a	-0.047^a	-0.046^a	-0.047^a	-0.045^a	-0.046^a
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
× Nearby siblings' experience	0.013	0.013^c	0.015^c	0.016^b	0.016^c	0.017^b
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
× Destination income level		0.005		0.005		0.004
		(0.014)		(0.014)		(0.014)
Nearby siblings' experience	0.016	0.016	0.020	0.021	0.018	0.018
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Average remote signal	0.021	0.021	0.019	0.019	0.018	0.018
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Firm domestic sales			0.058^a	0.059^a	0.021	0.021
			(0.020)	(0.020)	(0.026)	(0.026)
Destination-Ind-Year FE	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Age FE	✓	✓	✓	✓	✓	✓
Destination-Ind FE × Domestic Sales	✓	✓				
log Firm K/L × Year FE			✓	✓		
log Firm Sales × Year FE					✓	✓
N	32862	32749	32838	32725	32862	32749
R-squared	0.889	0.889	0.886	0.886	0.886	0.886

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.

OA.2.4 Bootstrap Estimation

Since our key regressors are cumulative residuals from regressing firm sales on a set of fixed effects, the inference in our main regressions suffer from the "generated regressor" problem. To assess the bias in standard errors, we perform bootstrap estimations of the two core tables (Tables 5 and A.2).

In particular, our bootstrap exercises are as follows. First, we randomly draw firms from the original affiliate-year level data with resampling. We draw blocks of firms instead of affiliates or affiliate-years because we worry about within-firm correlations in the error term – all our original standard errors are clustered at the firm level. We then estimate the regressors (as cumulative average of residual sales) using the bootstrapped samples and run the same regressions as in Tables 5 and A.2. Since the regressors are reestimated for each sample, this approach takes into account the potential estimation errors when generating the regressors. We perform 1000 bootstraps and present the results in Tables OA.6 and OA.7. In each column, we show the average point estimate, the standard deviation of the point estimate (in parentheses) and the 95% confidence interval (in brackets). In general, we find the bias in standard errors using simple OLS regressions is small. Due to computational constraints, we only use the bootstrapped regressions as a robustness check here and keep our original OLS regressions as the main evidence.

Table OA.6: Impact of siblings' experience on entry in the next period

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)
Average nearby signal	0.184 (0.033)	0.179 (0.041)	0.183 (0.044)
	[0.119, 0.253]	[0.098, 0.257]	[0.102, 0.272]
Average remote signal	0.041 (0.040)	0.040 (0.057)	0.024 (0.062)
	[-0.038, 0.122]	[-0.070, 0.150]	[-0.100, 0.142]
Firm domestic sales	0.065 (0.034)	-0.111 (0.098)	
	[-0.002, 0.134]	[-0.307, 0.084]	
Destination-Ind-Year FE	✓	✓	✓
Firm FE		✓	
Firm-Year FE			✓

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.

Table OA.7: Full set of interaction terms in the expectation formation regressions

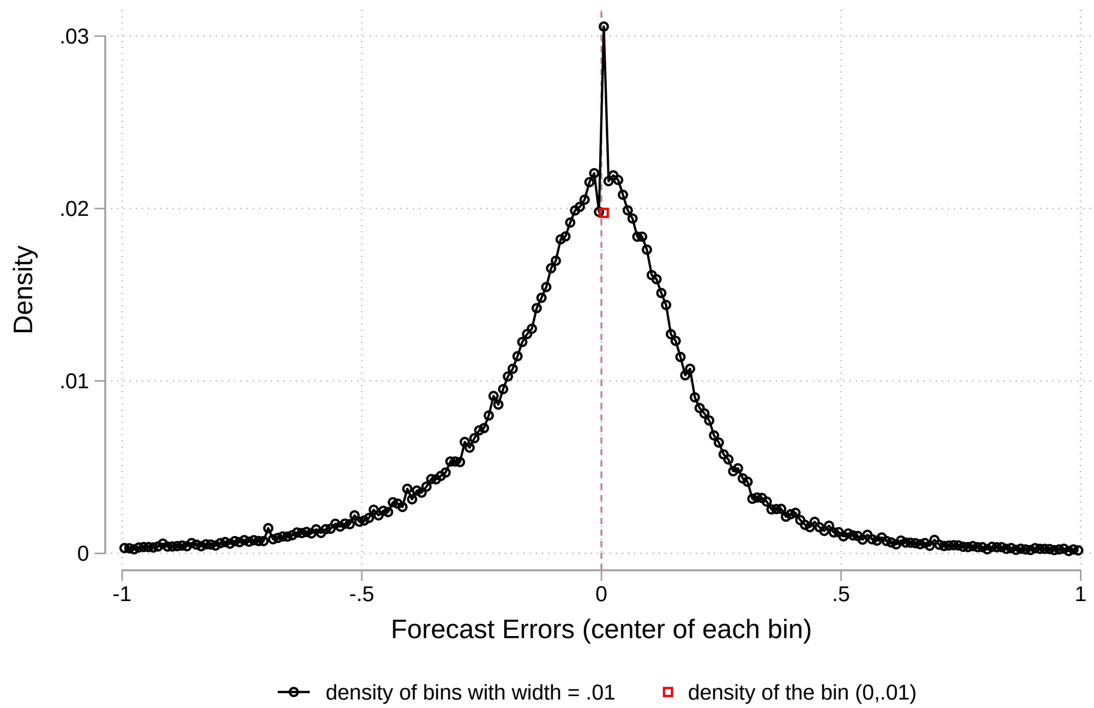
Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.872 (0.011)	0.872 (0.011)	0.871 (0.011)	0.870 (0.011)
	[0.849, 0.891]	[0.849, 0.891]	[0.847, 0.890]	[0.847, 0.889]
$\times \sigma_{\varepsilon 1}$	-0.027 (0.009)	-0.030 (0.012)	-0.015 (0.009)	-0.012 (0.009)
	[-0.046, -0.010]	[-0.055, -0.007]	[-0.033, 0.003]	[-0.030, 0.006]
$\times$ log(self age)	0.085 (0.007)	0.086 (0.007)	0.091 (0.007)	0.087 (0.007)
	[0.073, 0.098]	[0.072, 0.099]	[0.078, 0.104]	[0.074, 0.101]
$\times$ Nearby siblings' experience	0.006 (0.008)	0.005 (0.008)	0.003 (0.008)	0.006 (0.008)
	[-0.009, 0.020]	[-0.010, 0.021]	[-0.012, 0.017]	[-0.009, 0.021]
$\times$ Destination income level		-0.002 (0.013)		0.014 (0.010)
		[-0.026, 0.022]		[-0.004, 0.032]
Average nearby signal	0.052 (0.017)	0.052 (0.019)	0.051 (0.017)	0.053 (0.018)
	[0.018, 0.083]	[0.015, 0.087]	[0.018, 0.082]	[0.017, 0.087]
$\times \sigma_{\varepsilon 1}$	0.019 (0.010)	0.021 (0.013)	0.010 (0.010)	0.007 (0.011)
	[-0.001, 0.040]	[-0.004, 0.046]	[-0.009, 0.031]	[-0.014, 0.029]
$\times$ log(self age)	-0.049 (0.009)	-0.049 (0.009)	-0.053 (0.009)	-0.050 (0.009)
	[-0.066, -0.031]	[-0.066, -0.031]	[-0.071, -0.036]	[-0.068, -0.032]
$\times$ Nearby siblings' experience	0.023 (0.011)	0.023 (0.011)	0.024 (0.011)	0.023 (0.011)
	[-0.000, 0.044]	[0.001, 0.043]	[0.001, 0.045]	[0.001, 0.044]
$\times$ Destination income level		0.002 (0.017)		-0.010 (0.013)
		[-0.032, 0.036]		[-0.035, 0.017]
Nearby siblings' experience	0.008 (0.023)	0.009 (0.023)	0.010 (0.023)	0.009 (0.023)
	[-0.037, 0.053]	[-0.036, 0.054]	[-0.036, 0.054]	[-0.037, 0.053]
Average remote signal	0.019 (0.030)	0.020 (0.030)	0.019 (0.030)	0.020 (0.030)
	[-0.038, 0.077]	[-0.037, 0.079]	[-0.037, 0.078]	[-0.036, 0.079]
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓

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.

OA.2.5 Excluding Observations with Zero Forecast Errors

In this section, we examine the distribution of affiliates' forecast errors, with a special focus on the density of forecast errors around a small neighborhood of zero. We find that a small but non-negligible fraction of firms have exactly zero forecast errors. We discuss different interpretations of this finding and show that our results are robust to excluding this set of firms from our sample.

Figure OA.3: Density of forecast errors, $\log\left(\frac{R_{i,t+1}}{E_t(R_{i,t+1})}\right)$



Notes: Each circle represents the density of forecasting errors in a symmetric neighbourhood around the center of the bin. Each bin has equal width 0.01, with the left boundary closed and the right boundary open (e.g., [-0.02, -0.01), [-0.01, 0), [0, 0.01), etc). The red square denotes the fraction of observations with forecasting error in the range (0,0.01). We drop observations with forecasting errors below -1 and above 1, which accounts for 1.6% of the sample.

In Figure OA.3, we plot the share of firms in our expectation formation regressions in different bins of log forecast errors. Each bin has a width of 0.01, with the left boundary being inclusive. It is clear that the share of observations report forecast errors in the range [0,0.01) is much larger than the other bins in the neighborhood. A closer look at these observations reveals that this phenomenon is entirely driven by observations "bunching" at zero forecast errors, i.e., they perfectly predict their sales next period. In particular, the fractions of observations reporting forecast errors of zero and in the four neighborhoods of zero are displayed in Table OA.8.

Table OA.8: Fractions of observations reporting forecast errors in neighborhoods of zero

Range	[-0.02, -0.01)	[-0.01, 0)	0	(0,0.01)	[0.01, 0.02)
Share of Obs.	2.20%	1.98%	1.08%	1.98%	2.16%

We think that there are two possible interpretations for this bunching behavior. First, affiliates may just want to "hit their targets" and put less effort once they have satisfied the goals.[6] Second, we think that firms may have used their previous forecasts as anchors and simply report the same value as their current sales in the survey if their actual sales are quite close to the forecasts. Both are reasonable interpretations, but the evidence slightly favors the second one. If affiliates are trying to "hit the targets", it is a bit puzzling why they do not want to "beat the targets" by a small margin – we do not see extra mass in the ranges slightly above zero. In addition, we see that the density of forecast errors tends to decline as the bins are more distant from zero. However, this is not true when comparing the four bins around zero. We see a small increase when moving from [-0.01,0) to [-0.02,-0.01) and from (0,0.01) to [0.01,0.02). This suggests that affiliates may round their sales so that they have zero forecast errors when their sales are very close to the previous forecasts, e.g., when their true forecast errors are in the ranges of [-0.01,0) and (0,0.01).

Regardless of the cause of such bunching behavior, we are concerned that it may bias our estimates. We therefore perform robustness checks by excluding observations with zero forecast errors. Table OA.9 replicates Table A.2 in the paper after dropping these observations. The coefficients are almost unchanged.

OA.2.6 Sales weighted signals

In this section, we consider alternative measures of sibling signals that are cumulative averages of residual log sales weighted by the level of sales. In the paper, our preferred measure is a cumulative average where all past signals have equal weights. This measure is consistent with our simple model, but does not allow the possibility that firms learn more from signals with more sales activities. It is possible to construct models in which learning is positively correlated with the level of sales. For example, when firms reach more customers, they may draw a singal from each customer that they serve.

[6] We thank a referee for suggesting this possibility.

Table OA.9: Full set of interaction terms in the expectation formation regressions, excluding observations with zero forecast errors

Dep. Var: log $E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.867^a	0.866^a	0.866^a	0.864^a
	(0.010)	(0.010)	(0.010)	(0.010)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.026^a	-0.028^a		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.012^c	-0.010
			(0.006)	(0.006)
$\times$ log(self age)	0.086^a	0.086^a	0.091^a	0.088^a
	(0.007)	(0.007)	(0.006)	(0.007)
$\times$ Nearby siblings' experience	0.004	0.003	0.001	0.004
	(0.008)	(0.008)	(0.008)	(0.008)
$\times$ Destination income level		-0.002		0.014
		(0.011)		(0.009)
Average nearby signal	0.052^a	0.050^a	0.051^a	0.052^a
	(0.014)	(0.015)	(0.014)	(0.015)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.020^b	0.024^a		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.010^c	0.008
			(0.006)	(0.006)
$\times$ log(self age)	-0.046^a	-0.047^a	-0.050^a	-0.048^a
	(0.010)	(0.010)	(0.010)	(0.010)
$\times$ Nearby siblings' experience	0.024^b	0.025^a	0.025^a	0.025^a
	(0.010)	(0.010)	(0.010)	(0.010)
$\times$ Destination income level		0.007		-0.008
		(0.015)		(0.012)
Nearby siblings' experience	0.013	0.014	0.014	0.014
	(0.020)	(0.020)	(0.020)	(0.020)
Average remote signal	0.020	0.021	0.020	0.021
	(0.025)	(0.025)	(0.025)	(0.025)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	31522	31407	31505	31390
R-squared	0.906	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.

Table OA.10: Impact of siblings' experience on entry in the next period, sales-weighted signals

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)
Average nearby signal	0.210^a	0.202^a	0.225^a
	(0.032)	(0.037)	(0.041)
Average remote signal	0.122^a	0.095^c	0.118^b
	(0.039)	(0.050)	(0.057)
Firm domestic sales	-0.002	-0.152	
	(0.035)	(0.107)	
Destination-Ind-Year FE	✓	✓	✓
Firm FE		✓	
Firm-Year FE			✓
N	875523	875523	902523
# of Firms	1922	1922	1931
# of Firm-Markets	113996	113996	115181
# of Entries	977	977	1003
R-squared	0.064	0.067	0.088

Notes: Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

Table OA.11: Full set of interaction terms in the expectation formation regressions, sales-weighted signals

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.892^a	0.892^a	0.892^a	0.891^a
	(0.011)	(0.011)	(0.011)	(0.011)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.011	-0.009		
	(0.008)	(0.011)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.008	-0.007
			(0.007)	(0.007)
$\times$ log(self age)	0.078^a	0.077^a	0.080^a	0.078^a
	(0.007)	(0.007)	(0.007)	(0.007)
$\times$ Nearby siblings' experience	0.005	0.006	0.004	0.006
	(0.011)	(0.010)	(0.010)	(0.011)
$\times$ Destination income level		0.004		0.008
		(0.010)		(0.008)
Average nearby signal	0.014	0.014	0.014	0.014
	(0.012)	(0.012)	(0.012)	(0.012)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.001	0.004		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.003	0.004
			(0.005)	(0.005)
$\times$ log(self age)	-0.029^a	-0.029^a	-0.029^a	-0.029^a
	(0.008)	(0.008)	(0.008)	(0.008)
$\times$ Nearby siblings' experience	0.009	0.010	0.009	0.010
	(0.007)	(0.007)	(0.007)	(0.007)
$\times$ Destination income level		0.005		0.003
		(0.010)		(0.009)
Nearby siblings' experience	-0.019	-0.020	-0.019	-0.020
	(0.020)	(0.020)	(0.020)	(0.020)
Average remote signal	-0.002	-0.002	-0.002	-0.001
	(0.015)	(0.015)	(0.015)	(0.015)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	31714	31599	31697	31582
R-squared	0.913	0.913	0.913	0.913

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.

OA.2.7 Controlling (sibling) distance to Japan

In this section, we consider robustness checks by controlling for the focal affiliates' or the siblings' distance to Japan. In particular, Columns 1 and 2 in Table OA.12 replicate the expectation formation regressions adding interaction terms between signals and the log distance between the focal host country and Japan. This addresses the concerns that the proximity to the parent firms may affect the ability of the affiliates to adjust their expectations based on signals from itself and the nearby siblings. We see that the interaction terms with distance are insignificant, while the other interaction terms are not affected much compared to Table A.2 in the paper.[7]

Columns 3 and 4 augment the regressions with the average distance of nearby siblings to Japan and the interaction between this variable with signals. The impact of the average distance of siblings can be identified because for focal affiliates in the same destination, their siblings may be in different countries within the region, thus having different average distance to Japan. We again see insignificant effects of the average distance and the two interaction terms, while the other interaction terms are similar to our baseline results. In sum, our results are robust to controlling for the focal affiliate's and the siblings distance to Japan.

[7]One may also worry that the distance to Japan will affect the level of expected sales directly. However, this term is co-linear with the destination-industry-year fixed effects and cannot be identified.

Table OA.12: Full set of interaction terms in the expectation formation regressions, excluding observations with zero forecast errors

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.867^a	0.867^a	0.868^a	0.867^a
	(0.009)	(0.009)	(0.009)	(0.009)
$\times \sigma_{\varepsilon 1}$ (SD of sales growth)	-0.025^a	-0.027^a	-0.028^a	-0.030^a
	(0.007)	(0.008)	(0.007)	(0.008)
$\times$ log(self age)	0.082^a	0.083^a	0.083^a	0.083^a
	(0.006)	(0.006)	(0.006)	(0.006)
$\times$ Nearby siblings' experience	0.001	0.001	0.001	0.001
	(0.007)	(0.008)	(0.007)	(0.008)
$\times$ Destination dist. to Japan	0.007	0.008		
	(0.008)	(0.008)		
$\times$ Regional siblings dist. to Japan			0.002	0.002
			(0.008)	(0.009)
$\times$ Destination income level		-0.002		-0.002
		(0.011)		(0.012)
Average nearby signal	0.030^b	0.029^b	0.030^b	0.030^b
	(0.014)	(0.015)	(0.014)	(0.014)
$\times \sigma_{\varepsilon 1}$ (SD of sales growth)	0.015^b	0.019^b	0.017^b	0.020^b
	(0.007)	(0.008)	(0.007)	(0.009)
$\times$ log(self age)	-0.048^a	-0.048^a	-0.048^a	-0.048^a
	(0.009)	(0.009)	(0.009)	(0.009)
$\times$ Nearby siblings' experience	0.014^c	0.015^c	0.014^c	0.015^c
	(0.008)	(0.008)	(0.008)	(0.008)
$\times$ Destination dist. to Japan	-0.003	-0.004		
	(0.009)	(0.009)		
$\times$ Regional siblings dist. to Japan			0.002	0.001
			(0.010)	(0.010)
$\times$ Destination income level		0.006		0.005
		(0.014)		(0.015)
Nearby siblings' experience	0.018	0.018	0.018	0.018
	(0.016)	(0.016)	(0.016)	(0.016)
Average remote signal	0.021	0.021	0.021	0.021
	(0.018)	(0.018)	(0.018)	(0.018)
Regional siblings dist. to Japan			0.031	0.031
			(0.060)	(0.060)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	33321	33297	33420	33305
R-squared	0.885	0.885	0.885	0.885

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.

OA.2.8 Placebo Test: Predicting Rival Forecasts

As a placebo test, Table OA.13 replicates Table A.2 in the paper by replace the dependent variable with the average log expected sales of the focal affiliate's rivals. Rivals are defined to be those affiliates of other parent firms in the same destination and industry.[8]

We find the average self signal has a small negative impact on rivals' expectations. This is intuitive: better (relative) historical performance of the focal affiliates means stronger competition with the "rivals". Therefore, the rivals lower their expectations, the extent of which depends on multiple factors, such as how well the rivals observe these signals and the elasticity of substitution between products produced by different affiliates. We do not have a strong belief about the signs of the interaction terms and most of them are actually insignificant. Finally, we see from these regressions that the focal affiliate's nearby siblings' signals have a small and insignificant effect on the rivals' expectations. This suggests that firms may take into account the performance of rivals in the same destination and industry when forming their expectations, but do not take into account the rivals' performance in other markets.

[8] We thank a reviewer for proposing this placebo test.

Table OA.13: The impact of signals on rivals' expectations

Dep. Var: Average log $E_t(R_{i,t+1})$ of Rivals	(1)	(2)	(3)	(4)
Average self signal	-0.021^a	-0.020^a	-0.020^a	-0.020^a
	(0.001)	(0.001)	(0.001)	(0.001)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.004^b	0.003		
	(0.002)	(0.002)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.004^c	-0.005^b
			(0.002)	(0.002)
$\times$ log(self age)	0.001	0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
$\times$ Nearby siblings' experience	-0.001	-0.001	0.000	-0.000
	(0.002)	(0.001)	(0.001)	(0.001)
$\times$ Destination income level		-0.001		-0.004^a
		(0.002)		(0.001)
Average nearby signal	0.002	0.001	0.002	0.001
	(0.002)	(0.003)	(0.003)	(0.003)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.003^b	-0.001		
	(0.002)	(0.002)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.001	0.001
			(0.002)	(0.002)
$\times$ log(self age)	0.000	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
$\times$ Nearby siblings' experience	0.000	0.001	-0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
$\times$ Destination income level		0.004^c		0.005^a
		(0.002)		(0.002)
Nearby siblings' experience	-0.009^b	-0.008^b	-0.010^a	-0.008^b
	(0.004)	(0.003)	(0.004)	(0.003)
Average remote signal	0.005	0.005	0.004	0.003
	(0.004)	(0.004)	(0.004)	(0.004)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	31088	30984	31071	30967
R-squared	0.986	0.986	0.987	0.987

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.

OA.2.9 Impact of Signals on Affiliate Exits

In this section, we examine the impact of self and siblings' signals on exits. A practical question here is how we measure "exits" of affiliates. In the survey, some affiliates do respond in the year that they exit, and report that their status as "operation suspended" or "dissolution or withdrawal" or "decline in control share" (below 10%). However, we are concerned that this strict definition of "exit" will understate the overall exit rates because other affiliates may just stop responding when they exit. We therefore use two more general definitions of exits by including affiliates that stopped responding for at least two consecutive years (and plus those that report zero sales for at least two consecutive years).

We regress an indicator variable of whether the affiliate exits in the next year on self and sibling signals up to the current period in Online Appendix Table OA.14. We find that a better self signal significantly reduces the probability of exit next period. The coefficient in front of the nearby sibling's signal, though negative, is not precisely estimated and insignificantly different from zero. Therefore, we only find suggestive evidence for the model's predictions. This may be due to the difficulty of measuring affiliate exits precisely.

Table OA.14: Signals on affiliate exits

Dep. Var: Exit × 100	Basic Definition		Extended Definition	
	(1)	(2)	(3)	(4)
Average self signal	-1.117[a]	-1.307[a]	-1.129[a]	-1.319[a]
	(0.150)	(0.152)	(0.154)	(0.157)
Average nearby signal	-0.132	-0.146	-0.146	-0.157
	(0.196)	(0.196)	(0.199)	(0.199)
Average remote signal	0.212	0.183	0.217	0.187
	(0.349)	(0.349)	(0.351)	(0.351)
Firm domestic sales	0.666	0.655	0.632	0.619
	(0.452)	(0.453)	(0.454)	(0.455)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Age FE		✓		✓
N	40736	40721	40530	40515
# of Exits	2972	2968	2966	2962
R-squared	0.297	0.298	0.297	0.298

Notes: Dependent variable is an indicator of whether the affiliate exit in the next year (scaled by 100). Standard errors are clustered at the firm level. Significance levels: a: 0.01, b: 0.05, c: 0.10. Columns 1 and 2 define exit as affiliates that report "operation dissolved" or "dissolution or withdrawal" or "decline in control share" (below 10%), plus affiliates that stopped responding to the survey for at least two consecutive years. Column 3 and 4 further include cases where affiliates report zero sales for at least two consecutive years.

OA.2.10 Horse race between parent and affiliate/sibling experience

In this section, we run a horse race between the affiliate/sibling experience, measured by the number of signals received by the focal affiliate and its nearby siblings, and two measures of parent experience in multinational production: the time since the first affiliate was founded, and the current number of affiliates worldwide. Similar to the regressions in Table A.2, we interact the self and sibling signals with the two measures of parent global experience in Table OA.15. We find that the parent experience interaction terms are insignificant, but the other interaction terms have similar coefficients and standard errors as before. The positive interaction term between nearby siblings' signal and the total number of signals survives these horse race regressions. We therefore conclude that it is the number of the "relevant signals" rather than the parent firm's global experience that matters for the learning speed.

Table OA.15: Horse race between parent and affiliate/sibling experience

Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.867^a	0.866^a	0.867^a	0.867^a
	(0.009)	(0.009)	(0.009)	(0.009)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.030^a	-0.032^a	-0.030^a	-0.033^a
	(0.007)	(0.009)	(0.007)	(0.009)
$\times$ log(self age)	0.083^a	0.083^a	0.082^a	0.082^a
	(0.006)	(0.006)	(0.006)	(0.006)
$\times$ Nearby siblings' experience	-0.002	-0.002	-0.003	-0.004
	(0.008)	(0.009)	(0.008)	(0.008)
$\times$ log(total # of affiliates)	0.006	0.006		
	(0.008)	(0.008)		
$\times$ log(Parent MP Age)			0.010	0.010
			(0.009)	(0.009)
$\times$ Destination income level		-0.003		-0.004
		(0.011)		(0.011)
Average nearby signal	0.032^b	0.031^b	0.032^b	0.031^b
	(0.014)	(0.015)	(0.014)	(0.015)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.017^b	0.020^b	0.017^b	0.021^b
	(0.008)	(0.009)	(0.008)	(0.009)
$\times$ log(self age)	-0.047^a	-0.048^a	-0.047^a	-0.047^a
	(0.009)	(0.009)	(0.009)	(0.009)
$\times$ Nearby siblings' experience	0.015	0.016^c	0.015^c	0.016^c
	(0.009)	(0.009)	(0.009)	(0.009)
$\times$ log(total # of affiliates)	0.002	0.002		
	(0.010)	(0.011)		
$\times$ log(Parent MP Age)			0.000	-0.000
			(0.009)	(0.010)
$\times$ Destination income level		0.005		0.006
		(0.014)		(0.014)
Nearby siblings' experience	0.022	0.023	0.024	0.025
	(0.017)	(0.017)	(0.017)	(0.017)
Average remote signal	0.019	0.020	0.019	0.019
	(0.018)	(0.018)	(0.018)	(0.018)
Firm domestic sales	0.053^a	0.054^a	0.053^a	0.053^a
	(0.019)	(0.019)	(0.019)	(0.019)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
N	32862	32749	32862	32749
R-squared	0.886	0.886	0.886	0.886

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.

OA.2.11 The Frequency of "Naive" Forecasts

In Table OA.16, we show that it is very rare for firms to use a naive rule to make their sales forecasts. We calculate the expected growth rates 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 also checked whether firms simply use their previous growth rates as expected growth rates – we only see this in 0.03% of the observations in our sample .

Table OA.16: 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	0.1250	0.05
0.1000	0.08	0.1429	0.05
0.0526	0.07	0.3333	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 8 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.

Though it is difficult to tell whether forecasts that are the same as previous sales contain useful information or not, we conduct robustness checks in Table OA.17. It replicates all regressions in Table A.2 in the paper. Our main empirical results remain largely unchanged.

Table OA.17: 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
	(0.010)	(0.010)	(0.010)	(0.010)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.025^a	-0.028^a		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.013^b	-0.011^c
			(0.006)	(0.006)
× log(self age)	0.085^a	0.086^a	0.090^a	0.088^a
	(0.007)	(0.007)	(0.006)	(0.007)
× Nearby siblings' experience	0.005	0.004	0.003	0.005
	(0.008)	(0.008)	(0.007)	(0.008)
× Destination income level		-0.005		0.012
		(0.010)		(0.009)
Average nearby signal	0.054^a	0.053^a	0.053^a	0.055^a
	(0.014)	(0.015)	(0.014)	(0.015)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.020^a	0.025^a		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.011^c	0.009
			(0.006)	(0.006)
× log(self age)	-0.045^a	-0.046^a	-0.049^a	-0.047^a
	(0.010)	(0.010)	(0.010)	(0.010)
× Nearby siblings' experience	0.022^b	0.022^b	0.023^b	0.022^b
	(0.010)	(0.009)	(0.010)	(0.009)
× Destination income level		0.007		-0.008
		(0.015)		(0.012)
Nearby siblings' experience	0.010	0.011	0.012	0.011
	(0.020)	(0.020)	(0.020)	(0.020)
Average remote signal	0.025	0.026	0.025	0.026
	(0.025)	(0.025)	(0.025)	(0.025)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
R-squared	0.90	0.90	0.90	0.90
N	31101	30988	31084	30971

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.

OA.2.12 Excluding the Year of 1995

In our data, the entry rates in 1995 are higher than the other years. In Table OA.18 and OA.19, we replicate the entry and expectation formation regressions in Table 5 and A.2 in the paper, respectively after excluding the year 1995 from our sample. The main empirical results are robust.

Table OA.18: Impact of siblings' experience on entry in the next period, excluding 1995

Dep. Var: $\mathbb{1}(Enter_{spk,t+1}) \times 1000$	(1)	(2)	(3)
Average nearby signal	0.154^a	0.162^a	0.156^a
	(0.031)	(0.037)	(0.040)
Average remote signal	0.058	0.053	0.034
	(0.040)	(0.053)	(0.055)
Firm domestic sales	0.051	-0.110	
	(0.033)	(0.101)	
Destination-Ind-Year FE	✓	✓	✓
Firm FE		✓	
Firm-Year FE			✓
R-squared	0.06	0.06	0.09
N	853608	853608	879313

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.19: 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
	(0.010)	(0.010)	(0.010)	(0.010)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	-0.025^a	-0.029^a		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			-0.012^c	-0.010
			(0.006)	(0.006)
$\times$ log(self age)	0.086^a	0.087^a	0.092^a	0.089^a
	(0.007)	(0.007)	(0.006)	(0.007)
$\times$ Nearby siblings' experience	0.003	0.003	0.001	0.003
	(0.008)	(0.008)	(0.008)	(0.008)
$\times$ Destination income level		-0.004		0.013
		(0.011)		(0.009)
Average nearby signal	0.051^a	0.050^a	0.051^a	0.052^a
	(0.014)	(0.015)	(0.014)	(0.015)
$\times\sigma_{\varepsilon 1}$ (SD of sales growth)	0.019^b	0.023^b		
	(0.008)	(0.009)		
$\times\sigma_{\varepsilon 1}$ (SD of fore. err.)			0.009	0.007
			(0.006)	(0.006)
$\times$ log(self age)	-0.048^a	-0.048^a	-0.052^a	-0.050^a
	(0.010)	(0.010)	(0.010)	(0.010)
$\times$ Nearby siblings' experience	0.024^b	0.024^b	0.025^b	0.025^b
	(0.010)	(0.010)	(0.010)	(0.010)
$\times$ Destination income level		0.007		-0.007
		(0.015)		(0.012)
Nearby siblings' experience	0.013	0.014	0.014	0.014
	(0.020)	(0.020)	(0.020)	(0.020)
Average remote signal	0.019	0.020	0.019	0.021
	(0.026)	(0.026)	(0.026)	(0.026)
Destination-Ind-Year FE	✓	✓	✓	✓
Firm-Year FE	✓	✓	✓	✓
Age FE	✓	✓	✓	✓
R-squared	0.91	0.91	0.91	0.91
N	31586	31471	31569	31454

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

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.

Chen, Cheng, Tatsuro Senga, Chang Sun, and Hongyong Zhang, "Uncertainty, Imperfect Information and Expectation Formation over the Firms' Life Cycle," CESifo Working Paper 8468 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.

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

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

Stokey, Nancy L., *Recursive Methods in Economic Dynamics*, Harvard University Press, 1989.

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