Online Appendix for "Learning and Information Transmission within Multinational Corporations": Not for Publication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

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

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

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

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

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

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
$\Delta\log(\text{sales})$	0.079	0.110	0.073	0.178	0.193
	[127527]	[5437]	[2410]	[30058]	[791]
$\Delta\hat{e}_1(\text{sales})$	0.037	0.062	0.066	0.062	0.132
	[123952]	[5180]	[1282]	[25839]	[466]
$\Delta\hat{e}_2(\text{sales})$	0.086	0.097	0.104	0.149	0.203
	[122845]	[5386]	[2247]	[29129]	[614]

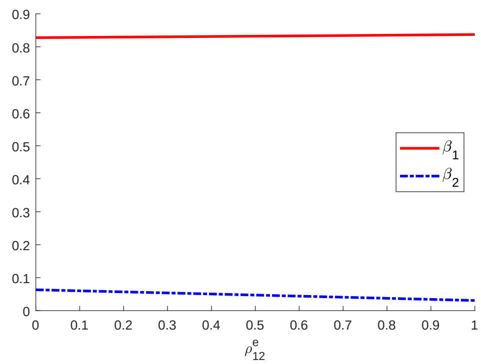
Notes: Each observation is an affiliate-sibling-year combination (two different siblings in two different countries within a particular region). The proxies for the temporary idiosyncratic demand shock are explained in the notes of Table A.1 of the paper. All the correlation coefficients are significant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

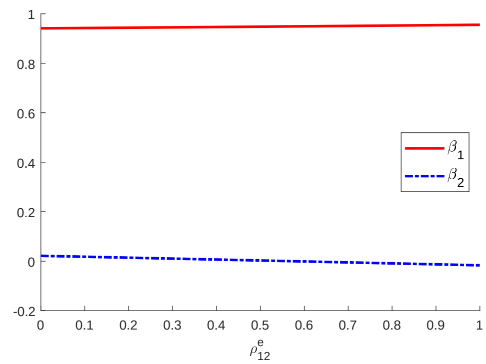
In total, we conclude that having a reasonable level of the correlation in temporary shocks only causes a quantitatively small bias in our estimated coefficients of self and sibling signals on expectation formation.

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.

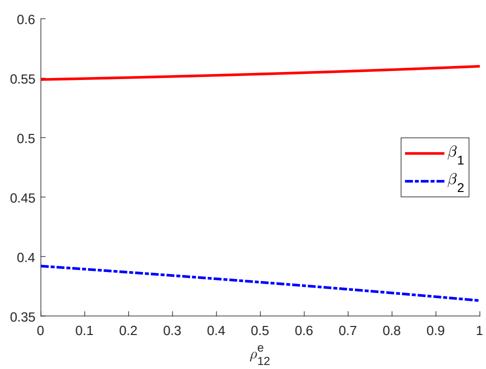
Figure OA.1: Correlation of temporary shocks and Learning Parameters



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



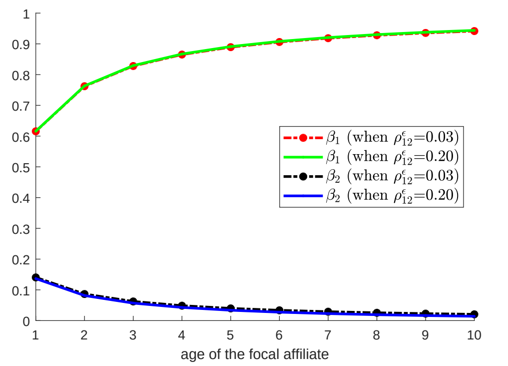
(b) $t_1 = 3, \rho_{12} = 0.38$



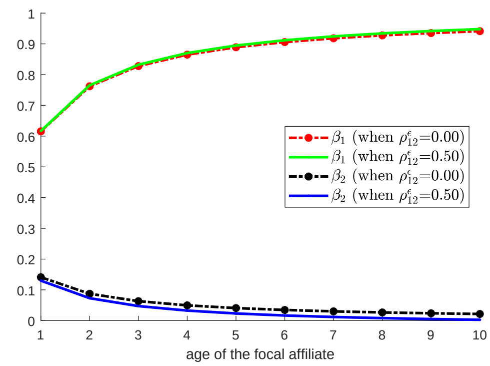
(c) $t_1 = 10, \rho_{12} = 0.9$

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

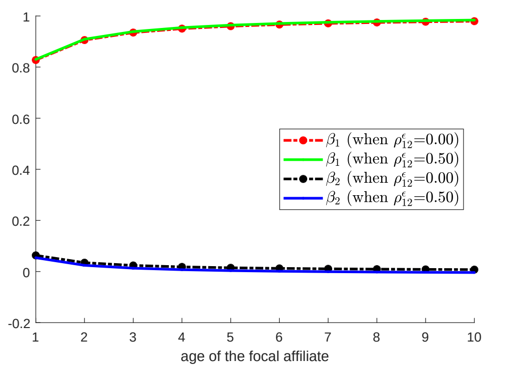
Figure OA.2: Correlation of temporary shocks and Learning over Life-cycle: correlation coefficients consistent with the data



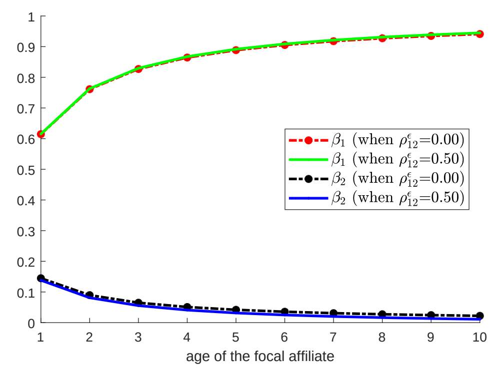
(a) Empirically relevant ρ^e_{12}, $\lambda_1 = \lambda_2 = 1.86$



(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

$$\begin{aligned}
\beta_1 &= \frac{\sigma_{\theta1}^2\sigma_{\theta2}^2\sigma_{\theta3}^2}{\Delta}\begin{bmatrix} 2\rho_{12}\rho_{13}\rho_{23} + (1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2 \\ -\rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{13}^2(1+\frac{1}{\lambda_2 t_2}) \end{bmatrix}, \\
\beta_2 &= \frac{\sigma_{\theta1}\sigma_{\theta2}\sigma_{\theta3}^2\sigma_{\varepsilon1}^2}{\Delta}\left[\frac{\rho_{12}}{t_1}(1+\frac{1}{\lambda_3 t_3}) - \rho_{13}\rho_{23}\frac{1}{t_1}\right], \\
\beta_3 &= \frac{\sigma_{\theta1}\sigma_{\theta3}\sigma_{\theta2}^2\sigma_{\varepsilon1}^2}{\Delta}\left[\frac{\rho_{13}}{t_1}(1+\frac{1}{\lambda_2 t_2}) - \rho_{12}\rho_{23}\frac{1}{t_1}\right],
\end{aligned}$$

where Δ equals

$$\sigma_{\theta1}^2\sigma_{\theta2}^2\sigma_{\theta3}^2\left[2\rho_{12}\rho_{13}\rho_{23} + (1+\frac{1}{\lambda_1 t_1})\left[(1+\frac{1}{\lambda_2 t_2})(1+\frac{1}{\lambda_3 t_3}) - \rho_{23}^2\right] - \rho_{12}^2(1+\frac{1}{\lambda_3 t_3}) - \rho_{13}^2(1+\frac{1}{\lambda_2 t_2})\right].$$

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 $\sigma_{\varepsilon1}$, 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) [0.119, 0.253]	0.179 (0.041) [0.098, 0.257]	0.183 (0.044) [0.102, 0.272]
Average remote signal	0.041 (0.040) [-0.038, 0.122]	0.040 (0.057) [-0.070, 0.150]	0.024 (0.062) [-0.100, 0.142]
Firm domestic sales	0.065 (0.034) [-0.002, 0.134]	-0.111 (0.098) [-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

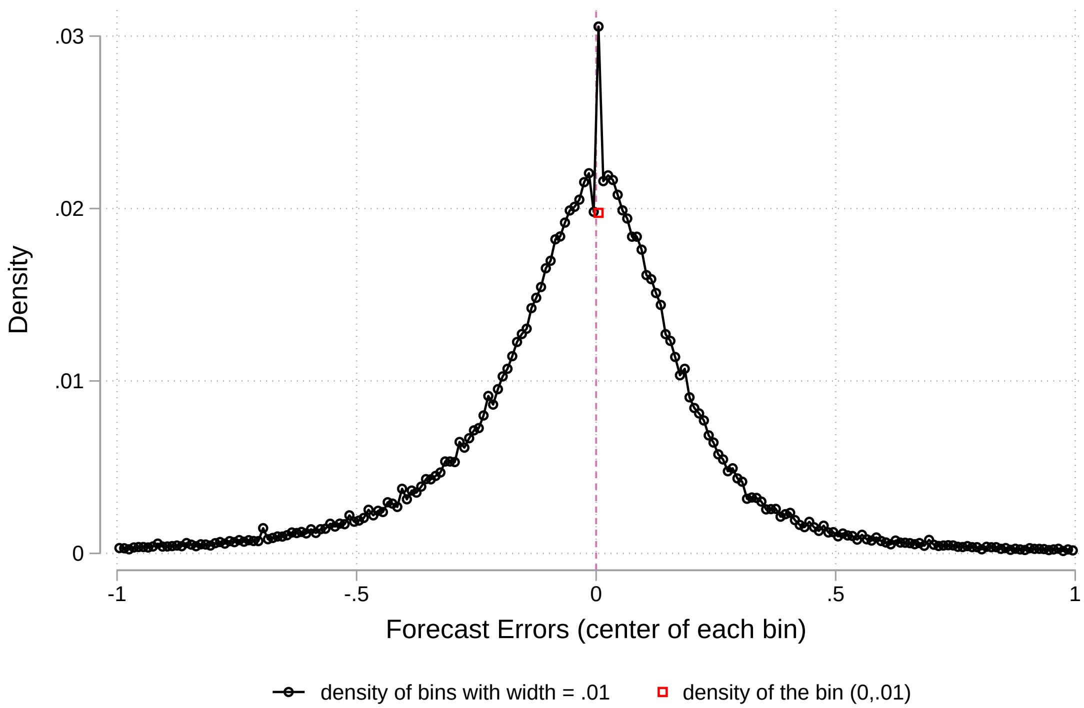
Dep. Var: $\log E_t(R_{i,t+1})$	(1)	(2)	(3)	(4)
Average self signal	0.872 (0.011)	0.872 (0.011)	0.871 (0.011)	0.870 (0.011)
	[0.849, 0.891]	[0.849, 0.891]	[0.847, 0.890]	[0.847, 0.889]
$\times \sigma_{\varepsilon 1}$	-0.027 (0.009)	-0.030 (0.012)	-0.015 (0.009)	-0.012 (0.009)
	[-0.046, -0.010]	[-0.055, -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)
$\times$ log(self age)	0.085^a	0.086^a	0.090^a	0.088^a
	(0.007)	(0.007)	(0.006)	(0.007)
$\times$ Nearby siblings' experience	0.005	0.004	0.003	0.005
	(0.008)	(0.008)	(0.007)	(0.008)
$\times$ 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)
$\times$ log(self age)	-0.045^a	-0.046^a	-0.049^a	-0.047^a
	(0.010)	(0.010)	(0.010)	(0.010)
$\times$ Nearby siblings' experience	0.022^b	0.022^b	0.023^b	0.022^b
	(0.010)	(0.009)	(0.010)	(0.009)
$\times$ 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.