Uncertainty, Imperfect Information and Learning in the International Market*

Cheng Chen Tatsuro Senga Chang Sun Hongyong Zhang†

July, 2018

Abstract

Using a unique data set of Japanese multinational firms' sales forecasts, we provide new evidence on imperfect information and learning at the firm level in the international market. We document three new facts concerning forecasts and forecast errors (FEs). First, firms make more precise forecasts as they become more experienced in the destination market, either through multinational production (MP) or exporting prior to MP entry. Second, past FEs are positively correlated with current forecasts as well as future FEs, which suggests the existence of imperfect information. Third, both (positive) correlations decline with firms' experience in the destination market. We then build and quantify a dynamic industry equilibrium model of trade and MP to match these facts and other salient features of firm dynamics. Counterfactual analysis shows that imperfect information enlarges productivity gains from liberalizing trade, and more so when multiple production modes are available.

Keywords. imperfect information and learning, uncertainty and firm expectations, trade and multinational production

JEL Classification. D83; D84; E23; F23; L2

---

*This research was conducted as a part of a research project funded by the Research Institute of Economy, Trade and Industry (RIETI). We would like to thank Costas Arkolakis, Scott Baier, Nick Bloom, Vasco Carvalho, Paola Conconi, Steven Davis, William Dougan, Taiji Furusawa, Gino Gancia, Stefania Garetto, Gene Grossman, Kyle Handley, Oleg Itskhoki, Kevin Lim, Yulei Luo, Eduardo Morales, Andreas Moxnes, Ezra Oberfield, Larry Qiu, Steve Redding, Kim Ruhl, Edouard Schaal, Nicholas Sly, Wing Suen, Heiwai Tang, Zhigang Tao, Olga Timoshenko, Felix Tintelnot, Mirko Wiederholt, Kei-Mu Yi, Lei Zhang and seminar participants at many institutions for helpful comments. Financial support from HKGRF, RIETI and Princeton University is greatly appreciated.

†Chen: University of Hong Kong, ccfour@hku.hk. Senga: Queen Mary University of London and RIETI, t.senga@qmul.ac.uk. Sun: University of Hong Kong, sunc@hku.hk. Zhang: RIETI, zhang-hong-yong@rieti.go.jp.

1 Introduction

A growing literature has highlighted the role of uncertainty in driving micro and macro performance (see, for example, Bloom (2009), Bloom et al. (2018), and Baker et al. (2016)). In fact, firms face uncertainty when making almost all decisions such as investment, hiring, and market entry.[1] A key part of such decisions is to form expectations about future outcomes; however, since we seldom observe firm's information directly, how firm expectations are formed remains unknown. This makes any attempt at isolating the source of uncertainty faced by firms difficult.

To study how firms form expectations and how firms resolve uncertainty over the life cycle, we use survey data with direct measures of firm expectations—a similar approach of Coibion et al. (2015) and Coibion and Gorodnichenko (2015) who use expectation data for households and professional forecasters. Using a unique data set of Japanese multinational firms' sales forecasts, we present direct evidence on firm-level uncertainty, imperfect information and learning in the international market. First, firms make more precise forecasts as they become more experienced in the destination market, either through multinational production (MP) or exporting prior to MP entry. Second, past forecast errors (FEs) are positively correlated with current forecasts as well as future FEs. Third, both (positive) correlations decline with firms' experience in the destination market. To account for these facts, we extend an industry equilibrium model of Jovanovic (1982), Timoshenko (2015), and Arkolakis et al. (2017) to allow the firm to make dynamic choices on its mode of service between exporting and multinational production. We use the model to show different implications of two types of uncertainty for resource allocation and aggregate productivity and how imperfect information interacts with productivity gains from trade liberalization.

Our dataset is unique in that we have a parent-affiliate matched panel on Japanese MNEs which contains forecasts for future sales at the affiliate level—a direct measure of firm's expectation. Specifically, affiliates of Japanese MNEs report their forecasted sales for the next year in an annual survey conducted by the Japanese government. Thanks to this feature, we are able to construct a measure of "forecast error" (FE) of sales, which is defined as the percentage deviation (or log point deviation) of the realized sales from the forecasted sales. We then treat the absolute value of FE as a measure of firm-level subjective uncertainty and document a set of novel facts concerning forecast and FEs. First, firm-level components explain most of the variation in FEs, while aggregate components explain a small fraction of it. Second, affiliate-level

[1] Uncertainty and imperfect information not only matter for individual-level decision making such as firm's investment (Guiso and Parigi (1999)), hiring (Bertola and Caballero (1994)), and market entry (Dixit (1989)), but also play an important role in triggering economic fluctuations (Bloom (2009)) and determining trade and foreign direct investment (FDI) flows.

(equivalently firm-level) subjective uncertainty declines with affiliate's age and its parent firm's previous export experience to the region where the affiliate is set up afterward.[2] The second finding suggests that firm-level subjective uncertainty decreases when its (or its parent firm's) operating experience in the market increases. However, it is silent on whether this is caused by learning which is related to imperfect information or by shocks whose variance is age-dependent and correlated with firm's experience.

In order to better understand the above findings, we borrow techniques from the studies of forecasting and expectation data (Mishkin (1983), Ito (1990), Andrade and Le Bihan (2013)) to detect the existence of imperfect information and learning in our data. As agents know information perfectly in full information rational expectation (FIRE) models, ex post FEs should not be correlated with any realized variable in the past. In particular, FEs in different periods should not be correlated. Moreover, the validity of these predictions is robust to different functional forms and distributional assumptions of the model.[3] When we look at the data, we find that the serial correlation of FEs is positively significant. In addition, we find that both past sales and past forecasts (for current sales) have predictive power for current sales, which cannot be rationalized using FIRE models.[4] In total, we provide evidence for the existence of imperfect information in our data.

Interestingly, the above evidence for imperfect information also exhibits age-dependent (or experience-dependent) patterns. Specifically, we find that the (positive) serial serial correlation of FEs and the (positive) covariance between past FEs and current forecasts all go down with firm age, which is true even after we have controlled for firm size (both parent and affiliate) and a battery of fixed effects. In addition, experienced affiliates (i.e., affiliates whose parent firms had previous export experience to the region where the affiliates are set up afterwards) start with *lower* levels of (positive) correlation and covariance compared to those inexperienced affiliates (i.e., affiliate whose parent firms had no previous export experience). In addition, we show past FEs are positively correlated with current forecasts, and this positive correlation decreases when the firm becomes older and when the firm's parent in Japan has the previous export experience. These evidence seems to suggest that at least a part of the reason why firm-level uncertainty declines with age or previous export experience is *learning*. I.e., firms learn their demand and supply conditions better and solve the problem of information imperfection partly, when they accumulate more experience via becoming older or having previous selling experience in the

[2]In order avoid confusion, we use the word "firm" to denote the affiliate (abroad) and the word "parent" (or parent firm) to denote the parent firm of the affiliate (in Japan).

[3]E.g., whether the distribution of the shocks is log normal and whether the variance of the shocks is time-varying.

[4]In FIRE models, only forecasted sales last period matter for current sales, as all shocks that unexpectedly affect current sales are random and not correlated with any variable in the past.

destination market before entry. Economically speaking, exporting and producing in the market generate *information value* in addition to profits.

Next, we build up a model featuring imperfect information and learning to rationalize the above empirical findings. We follow Jovanovic (1982) to set up the model and assume that firms face a downward sloping demand curve and differ in their fundamental firm-specific demand draws (i.e., shocks) which are time-invariant. Each period, the firm receives a transitory demand shock, which is independently and identically distributed (i.i.d.). These two shocks together determine the overall demand of the firm. Crucially, the firm cannot differentiate the demand draw from the transitory shock, and learns about the fundamental demand over time. The firm also knows the prior distribution of the draw before entry. After entry, the firm updates the belief about the demand draw using information on past sales in the Bayesian fashion. Thanks to the accumulation of market experience, the firm's posterior belief about its fundamental demand becomes more precise, when it operates in the market longer and accumulate experience via prior exporting. This explains why sales forecasts become more precise, when the affiliates become older *and* when their parent firms have previous export experience to the region where the affiliates are set up. Next, a bigger (in value) transitory shock increases both the current demand shock and the error of the forecast made last period. It is also going to increase the current forecast (for next period's sales), as the firm includes current demand shock into its posterior belief formed at the end of this period. In addition, this effect decreases with firm's age (and previous export experience), as the impact of new information on posterior beliefs goes down with age. Thus, Jovanovic model can explain why the covariance of past FEs and current forecast is positive, but decreases (to zero) over time. However, Jovanovic model implies zero serial correlation of FEs, as Bayesian updating with unbiased prior yields the best linear unbiased estimator (BLUE) for the fundamental demand θ based on past information. Therefore, we extend Jovanovic model in order to match the finding of positively correlated FEs.

We modify the Jovanovic model *at the minimum level* in order to match all the stylized empirical facts documented above. Specifically, we incorporate sticky information as in Mankiw and Reis (2002) into the model and assume that all entering firms do not know how to use past information to update their beliefs initially (i.e., the uninformed firms). Every period after entry, a randomly selected fraction uninformed firms become informed and figure out how to update their beliefs using past sales. When they become informed, they begin to utilize past sales to update their beliefs and will never become uninformed again. For the uninformed firms, they still use the prior belief when forecasting future sales. Under this setup, FEs are positively correlated over time, as uninformed firms always use the (same) prior belief to forecast future sales and create positively correlated FEs over time. Moreover, the positive correlation fades away with firm's age (and export experience), as fewer and fewer firms are uninformed over their life cycles. In short, the extended Jovanovic model with sticky information rationalizes all the three stylized empirical patterns.

We incorporate the extended Jovanovic model at the firm level into a dynamic industry equilibrium model in which firms endogenously choose to serve the foreign market via exporting or multinational production (MP) similar to Arkolakis et al. (2017). Different from Arkolakis et al. (2017), we allow the firm to make *dynamic* choices on its mode of service (i.e., exporting v.s. MP). Although MP helps firms save on the (variable) iceberg trade cost, it requires higher entry costs as in Helpman et al. (2004). Different from Helpman et al. (2004), we assume that there is a *dynamic interaction* between exporting and MP. Specifically, MP becomes attractive to firms, only when they are *certain* that their fundamental demand draws are good enough. Thus, firms do not want to start MP immediately after entry, if they are uncertain about its fundamental demand. Instead, the firm can export to the destination market before setting up an affiliate there, as exporting helps the firm solve the imperfect information problem and entails lower sunk entry costs (Conconi et al. (2016)). We then calibrate the model by utilizing the unique moments on FEs and other standard moments commonly used in the trade/MP literature. The calibrated model is able to capture the decline in absolute value of FEs over affiliates' life cycle, as well as the smaller absolute value of FEs for affiliates with previous export experience. It is also able to capture other salient features of the data, such as growth in exporters' sales, age-dependent volatility of affiliates' sales growth, which we do not directly target in the calibration. Finally, the calibrated model also generates predictions on serial correlation of FEs and the predictability of forecasts for future sales, which are consistent with the empirical findings.

We implement counterfactual exercises to show how imperfect information and different types of uncertainty affect productivity gains from trade and resource allocation. First, we focus on how the variance of time-invariant demand draws and that of the transitory shocks, which generate different sources of uncertainty, affect aggregate productivity and dynamic selection into trade and MP. We find that an increase in the variance of transitory shocks reduces aggregate productivity and welfare, while an increase in the variance of the time-invariant draws increases them. Second, we show that (productivity) gains from reducing trade costs are larger in a world with imperfect information (than in a world without), and the difference in the gains becomes larger when there are multiple production modes (i.e., trade plus MP) instead of a single production mode (i.e., trade only).

The key channel we emphasize is the *dynamic selection* into MP and into staying in the market. In the model, imperfect information causes inefficiency, and this is especially true at the extensive margin. Specifically, it is not always the case that the most efficient firms become MNEs and the least efficient firms exit in a world with imperfect information. Any shocks and policies that affect the dynamic selection impact aggregate productivity. First, increasing uncertainty due to more volatile transitory shocks reduce the signal-to-noise ratio and thus negatively affects firm learning. As a result, the selection into MP and into the market reflects firm's true efficiency less and lucks more (i.e., good transitory shocks), which leads to lower aggregate productivity and welfare. To the contrary, increasing uncertainty due to higher dispersion of time-invariant draws improves learning and therefore leads to higher aggregate productivity and welfare. Second, when the trade costs go down which make it easier for firms to stay in the market and learn about themselves, the allocation into different production modes gets improved. Therefore, in addition to the conventional sources for gains from trade, reducing trade costs generates information value in a world with imperfect information which amplifies gains from trade. Moreover, this information value becomes larger when there are multiple production modes than one production mode, which explains why the difference in gains from trade (between a world with imperfect information and a world without) is larger in a trade regime with both trade and MP than in a world with trade only.

The remainder of the paper is organized as follows. In Section 2, we review related literature. In Section 3, we document five new facts regarding firms' FEs in the international market. In Section 4, we build up an industry equilibrium model of trade and MP to rationalize the three new empirical findings. In Section 5, we calibrate the model and implement counterfactual analysis concerning the variance of the two types of demand shocks and how information imperfect affects productivity gains from trade. We conclude in Section 6.

2 Literature Review

In macroeconomics, researchers have long been interested in the information structure of agents and its implications for economic outcomes (Mackowiak and Wiederholt (2009); Andrade and Le Bihan (2013); Coibion and Gorodnichenko (2012); Coibion et al. (2015); Coibion and Gorodnichenko (2015)). However, none of these studies focus on the extensive margin of firm-level activities and on how firm heterogeneity affects the firm-level expectation formations. Our paper fill the gap in this literature by studying how firm-level uncertainty affects market entry, resource allocation and welfare.

A related literature studies the impact of uncertainty on firm-level and aggregate outcomes (Abel (1983); Bernanke (1983); Dixit and Pindyck (1994); Bloom et al. (2007); Bloom (2009); Bachmann et al. (2013); Bachmann and Bayer (2014); Baker et al. (2016); Fajgelbaum et al. (2017); Schaal (2017); Bloom et al. (2018)). Recent research in international trade also incorporates uncertainty and examines how it impacts exports (Handley (2014); Novy and Taylor (2014); Handley and Limão (2015); Handley and Limao (2017)) and MP (Ramondo et al. (2013); Fillat and Garetto (2015)). Conceptually, this literature treats uncertainty as a technology parameter that firms cannot influence, as most existing papers focus on aggregate uncertainty. We provide evidence that uncertainty faced by the firm is endogenous to firm activities. We also use different data moments to differentiate information imperfection from volatility/risk. We find that the two types of firm-level uncertainty have different impact on resource allocation. We also illustrate that different sources of uncertainty have qualitatively different implications for dynamic selection and productivity.

Next, our work is related to the literature on firm learning and technology choices in industry equilibrium (Jovanovic (1982); Irwin and Klenow (1994) Jovanovic and Nyarko (1994); Klenow (1998); Arkolakis et al. (2017)). Essentially, exporting and MP can be viewed as two technology choices which are positively correlated in term of efficiency level. We complement this literature by providing direct evidence for learning and studying how imperfect information and learning affect productivity gains from improving the efficiency of one production technology (i.e., exporting).

Finally, imperfect information and learning are more likely to exist in the international market than in the domestic market. This probably explains why international economists have already begun to explore implications of learning models for the exporter dynamics (Albornoz et al. (2012); Akhmetova and Mitaritonna (2013); Aeberhardt et al. (2014); Timoshenko (2015); Cebreros (2016); Conconi et al. (2016); Ruhl and Willis (2016)). Despite of the extensive studies in the literature, there is a lack of direct evidence for the existence of imperfect information and learning in the international market. Our study fills this gap by providing direct evidence for these phenomena.[5]

3 New Facts: Uncertainty Dynamics and Imperfect Information in the International Market

In this section, we first present facts regarding multinational firms' subjective uncertainty over their life cycles, which suggests the existence of imperfect information and gradual learning. Specifically, we introduce our data and show descriptive statistics of our measures for affiliate-level (i.e., firm-level) uncertainty. We then show how this measure changes with affiliate age

[5] Gumpert et al. (2016) studies the joint dynamics of exporting and MP under an AR(1) productivity process. We complement their work by focusing on learning as a mechanism of reducing uncertainty and by highlighting the information value generated by exporting for market entry.

and how it is correlated with parent firms' previous export experience. Next, we present key evidence that substantiates the existence of imperfect information and gradual learning in the international market.

3.1 Data

We combine two Japanese firm-level datasets prepared by the Ministry of Economy, Trade and Industry (METI): the Basic Survey of Japanese Business Structure and Activities ("parent firm survey" hereafter) and the Basic Survey on Overseas Business Activities ("FDI survey" hereafter). The parent firm survey provides information about business activities of Japanese firms and covers firms from a large set of industries that employ more than 50 workers and have more than 30 million Japanese yen in total assets.[6] Firms also report their exports to seven regions: North America, Latin America, Asia, Europe, Middle East, Oceania and Africa. Combined with the FDI survey, we are able to measure previous export experience in a region before an affiliate is established.

The FDI survey contains information about overseas subsidiaries of Japanese MNEs. This survey covers two types of overseas subsidiaries of Japanese MNEs: (1) direct subsidiaries with the share of equity owned by Japanese enterprises' being 10% or higher as of the end of the year, (2) level-two subsidiaries with the ratio of investment by Japanese subsidiaries of 50% or higher as of the end of the year. Tracing the identification codes over time, we are able to construct a panel of affiliates and parent firms from 1995 to 2014. The matched dataset contains on average 2300 parent firms and 14000 affiliates each year.[7] Similar to other surveys of multinational firms, this dataset contains information on affiliates' location, industry affiliation, sales, employment, investment etc. Finally, following the literature, we exclude multinational affiliates in tax haven economies from our sample due to the potential concern of profit-shifting using FDI.

More important for our study, the FDI survey asks not only about the realized sales in the previous fiscal year, but also about the *projected sales* for the next fiscal year. We use this variable as firms' expectations of future sales.

[6]The industries included are mining, manufacturing, wholesale and retail trade, and eating and drinking places.

[7]Affiliates with relatively small parent firms are lost in this process. We have approximated 3200 parent firms and 17,000 affiliates (per year) in the FDI survey, while 2300 parent firms and 14000 affiliates (per year) in the merged data. We use all the data in the FDI survey whenever possible (e.g., when examining the dynamics of forecast errors over affiliates' life cycle). We use the merged sample when estimating the effect of previous export experience on firms' initial subjective uncertainty.

3.2 A First Look at Forecast Errors

We define the deviation of the realized sales from the projected sales as the forecast error (FE) of the firm. In most of our empirical and quantitative analysis, we define FE to be the log point deviation of the realized sales from the projected sales as in equation (1):

$$FE_t^{\log} \equiv \log\left(R_{t+1}/E_t^S\left(R_{t+1}\right)\right), \tag{1}$$

where R_{t+1} is the realized sales in period $t+1$ and $E_t^S\left(R_{t+1}\right)$ denotes a firm's prediction in period t. FE in period t can also be stated as $FE_{t,t+1}^{\log}$, and we use $FE_{t,t+1}^{\log}$ and $FE_t^{\log}$ interchangeably in what follows. A positive (negative) forecast error means that the firm is under-predicting (over-predicting) its sales. The second measure is the percentage deviation of the projected sales from the realized sales

$$FE_t^{pct} = \frac{R_{t+1}}{E^S\left(R_{t+1}\right)} - 1.$$

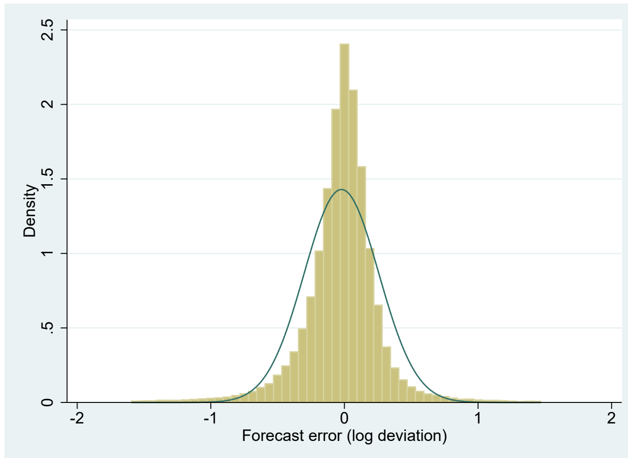
We use this measure for some of our robustness checks. Since we focus on firm-level uncertainty about their idiosyncratic business conditions, we want to exclude systemic FEs that are caused by unexpected aggregate shocks (e.g., recessions). We therefore construct a "residual forecast error" measure for robustness checks. We project our first measure $FE_t^{\log}$ onto country-year and industry-year fixed effects and obtain the residuals, $\hat{\epsilon}_{FE}$. The fixed effects only account for about 11% of the variation, which suggests that firm-level uncertainty plays a dominant role in generating firms' forecast errors. The facts we presented below are all robust to the two alternative measures of FE. As FEs calculated using above methods contain extreme values, we trim top and bottom one percent observations of FEs.

In Figure 1, we plot the distribution of our first measure of FEs, $FE^{\log}$, across all affiliates in all years. The FEs are centered around zero, and the distribution appears to be symmetric. The shape of the density is similar to a normal distribution, though the center and the tails seem to have more mass than the fitted normal distribution (solid line in the graph). This motivates us to assume firm-level shocks to be log-normal in our quantitative model.[8]

In Table 1, we report summary statistics regarding FEs. In the first two rows are about FEs and residual FEs. The mean of the residual FEs is zero by construction, while the mean of FE^{log} is close to zero. In the third row, we report the summary statistics of the absolute value of FEs, which we view as measures of firms' uncertainty. On average, firms under- or over-estimate their sales by 20 log points. In the fourth row, we show the statistics for the absolute value of the residual FEs. Since the country-year and industry-year fixed effects do not account for a

[8]By this assumption, the first measure of FEs has a log-normal distribution in our model. The normality assumption greatly simplifies our numerical implementation (see section 4).

Figure 1: Distribution of forecast errors



Note: Histogram of FE^{log} with fitted normal density (solid line).

large fraction of the variation, the mean (and median) and standard deviation of the absolute value of residual FEs are similar to those of $\left|FE^{\text{log}}\right|$. Patterns of manufacturing firms' FEs are similar to the overall patterns, as shown by the last two rows of the table.

Table 1: Summary Statistics of Forecast Errors

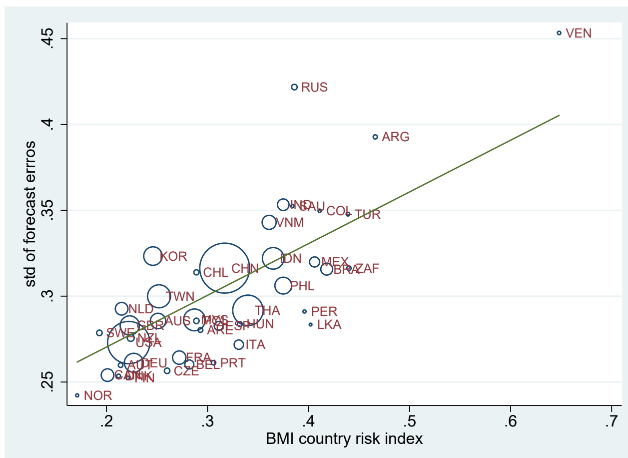
	Obs.	mean	std. dev.	median
FE^{log}	132,056	-0.024	0.300	-0.005
FE^{pct}	132,589	0.017	0.333	-0.006
$\hat{\epsilon}_{FE}$	131,760	-0.000	0.282	0.011
$\|FE^{log}\|$	132,056	0.200	0.224	0.130
$\|FE^{pct}\|$	132,589	0.204	0.264	0.130
$\|\hat{\epsilon}_{FE}\|$	131,760	0.184	0.213	0.116
FE^{log} - Manufacturing	91,580	-0.022	0.279	-0.003
$\|FE^{log}\|$ - Manufacturing	91,580	0.186	0.209	0.123

FE^{log} is the log deviation of the realized sales from the projected sales, while FE^{pct} is the percentage deviation of the realized sales from the projected sales. The last variable, $|\hat{\epsilon}_{FE^{log}}|$, is the absolute value of the residual forecast error, which we obtain by regressing FE^{log} on a set of industry-year and country-year fixed effects. Top and bottom one percent observations of forecast errors are trimmed.

3.3 Validation of Firm-level Forecasts

In this subsection, we show that the projected sales reported by Japanese MNEs' affiliates' and FEs constructed by us are reliable. First, the FDI survey is mandated by METI and not imposed by the parent firms of these affiliates. As a result, these sales forecasts are reported to the government (for policy and academic research) and should not be used by other firms (including the parent firm) required by the Japanese law. Therefore, the concern of strategic communication between the parent firm and the affiliate is unlikely to true in our dataset.[9] Second, Figure 2 shows that our constructed FEs are positively correlated with aggregate-level uncertainty such as the Country Risk Index (from the BMI research database), which makes intuitive sense. In addition, we regress our measures of FEs on the Country Risk Index and the standard deviation of real GDP growth rates in Table 2. The regression results show that aggregate-level uncertainty (at the destination economy) is highly positively correlated with affiliate-level uncertainty, and this is true even after we have controlled for parent fixed effects. In summary, sales forecasts reported by Japanese MNEs are reliable, and FEs constructed by us make intuitive sense.

Figure 2: Aggregate uncertainty and firm-level uncertainty



Note: Each circle represents a set of affiliates from a destination economy. The area of the circle reflects the number of observations in the destination economy.

3.4 Fact 1: Precision of Forecasts Increases over Affiliates' Life Cycles

In this subsection, we discuss how affiliates' subjective uncertainty about their sales evolves over their life cycles. We measure uncertainty using the absolute value of FEs. The first row in Table 3 shows the simple average of affiliates' $\left|FE^{\log}\right|$. As affiliates grow from age two to age ten, their FEs decline from 34% to 16%, which means they are better at predicting future sales as they become older. Similar patterns emerge when we use the absolute value of the residual FEs.

We further confirm these patterns formally by estimating an OLS regression of affiliate i's

[9] This point can also verified by the fact that the average of FEs is close to zero (i.e., no systematic over- or under-forecasting).

Table 2: Affiliates' uncertainty and country risk index

	(1) $\lvert FE^{log}\rvert$	(2) $\lvert FE^{pct}\rvert$	(3) $\lvert\hat{\epsilon}_{FE}\rvert$	(4) $\lvert FE^{log}\rvert$	(5) $\lvert FE^{pct}\rvert$	(6) $\lvert\hat{\epsilon}_{FE}\rvert$
Country risk index	0.275***	0.261***	0.264***			
	(0.042)	(0.041)	(0.049)			
$\sigma(\Delta log(GDP))$				1.061**	1.081***	0.988**
				(0.405)	(0.377)	(0.431)
N	130601	131105	130342	130522	131026	130276
R^2	0.149	0.151	0.140	0.146	0.150	0.137
Industry-year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Parent Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Mean of X	0.291			0.027		
Std. Dev. of X	0.062			0.010		

Standard errors are two-way clustered at country and parent firm level, * 0.10 ** 0.05 *** 0.01. Each column head lists the dependent variable of the regressions. $|FE^{log}|$ is the absolute log deviation of the realized sales from the projected sales; $|FE^{pct}|$ is the absolute percentage deviation of the realized sales from the projected sales; $|\hat{\epsilon}_{FE}|$ is the absolute value of the residual forecast error, which we obtain by regressing FE^{log} on a set of industry-year and country-year fixed effects. Country risk index (BMI research database) is an index from zero to one that measures the overall risk of the economy, such as an economic crisis or a sudden change in the political environment , with one being the most risky environment. $\sigma(\Delta log(GDP))$ is the standard deviation of real GDP growth rate of the host country since 1990, calculated from Penn World Table 9.0.

Table 3: Average (s.e.) of absolute forecast errors by age $FE_{t,t+1}$

	1	2	3	4	5	6	7	8	9	10
$\lvert FE^{log}\rvert$	0.364	0.298	0.258	0.231	0.215	0.215	0.205	0.199	0.194	0.177
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
$\lvert\hat{\epsilon}_{FE}\rvert$	0.349	0.280	0.244	0.217	0.205	0.202	0.191	0.185	0.178	0.161
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)

FE^{log} is the log deviation of the realized sales from the projected sales, while $\hat{\epsilon}_{FE}$ is the residual forecast error, which we obtain by regressing FE^{log} on a set of industry-year and country-year fixed effects.

FE in year t

$$|FE^{\log}|_{it} = \delta_t + \beta X_{it} + \delta_{ct} + \delta_s + \varepsilon_{it},$$

where δ_t is a vector of age dummies, δ_{ct} represents the country-year fixed effects and δ_s represents the (affiliate) industry fixed effects. We also control for affiliate or parent sales X_{it} in some regressions. We use age one as the base category, therefore the age fixed effects represent the difference in the absolute value of FEs between age t and age one. To further control for heterogeneity in uncertainty across affiliates, we also run a regression with affiliate fixed effects δ_i instead of the industry fixed effects δ_s.

We report the regression results in Table 4. Column 1 shows the baseline specification with industry and country-year fixed effects. It is clear that as affiliates become older, absolute value of their FEs declines. On average, affiliates that are at least ten years old have absolute FEs that are 16 log points lower. Most of the decline happens before age five. In column 2, we control for affiliates' sales and their parent firms' sales in Japan to address the concern that larger firms may have smaller uncertainty. Indeed, larger affiliates tend to have lower uncertainty. This may be because larger affiliates tend to diversify their products or these affiliates have better planning and thus more precise forecasts. Controlling for firm size does not alter the uncertainty-age profile.[10]

The uncertainty-age profile is also present, when we restrict our sample to entering and surviving affiliates. Endogenous exit may affect our estimates of the age effects for two reasons. First, affiliates with higher uncertainty may exit early as they are more likely to be hit by bad shocks. They may also delay their exit, since they have already paid the sunk cost (of FDI) and there is an option value of remaining in the market (Bloom (2009)). Second, FEs are censored as we do not observe realized sales for affiliates that have exited before the end of the year. To partially address these concerns, we focus on a subsample of affiliates that have survived and continuously appeared in the data from age one to age seven.

3.5 Fact 2: Learning about Market Demand through Exporting

In this subsection, we show that for affiliates that enter the destination country for the first time, they face lower subjective uncertainty if their parent firms have previous export experience to the region. Economically speaking, exporting generates information value in a world with imperfect information, which is similar to the information value generated by operating in the market

[10]Interestingly, affiliates with larger parent firms (measured by domestic sales) tend to have larger forecast errors. We conjecture that this is because larger parent firms may choose to enter riskier markets. This is confirmed by our regression in column 3, where we controlled for the subsidiaries' fixed effects and the parent firm size effect disappears.

Table 4: Age effects on the absolute forecast errors

Dep.Var: ($\|FE^{log}_{t,t+1}\|$)	(1)	(2)	(3)	(4)	(5)
Sample:		All Affiliates		Survived 7 years	Manufacturing
Age=2	-0.069	-0.065	-0.061	-0.069	-0.057
	(0.007)	(0.007)	(0.008)	(0.011)	(0.009)
Age=3	-0.107	-0.093	-0.080	-0.087	-0.077
	(0.007)	(0.008)	(0.008)	(0.011)	(0.009)
Age=4	-0.132	-0.116	-0.096	-0.098	-0.093
	(0.007)	(0.008)	(0.008)	(0.011)	(0.010)
Age=5	-0.146	-0.125	-0.098	-0.114	-0.092
	(0.007)	(0.007)	(0.008)	(0.011)	(0.010)
Age=6	-0.145	-0.124	-0.093	-0.115	-0.090
	(0.007)	(0.007)	(0.009)	(0.012)	(0.010)
Age=7	-0.156	-0.132	-0.098	-0.127	-0.092
	(0.007)	(0.007)	(0.009)	(0.011)	(0.010)
Age=8	-0.160	-0.134	-0.097	-0.123	-0.090
	(0.007)	(0.008)	(0.009)	(0.012)	(0.010)
Age=9	-0.164	-0.138	-0.098	-0.120	-0.088
	(0.007)	(0.007)	(0.009)	(0.013)	(0.010)
Age=10	-0.176	-0.139	-0.092	-0.121	-0.082
	(0.007)	(0.007)	(0.009)	(0.012)	(0.010)
log(Parent Domestic Sales)		0.008	0.002	0.011	0.002
		(0.001)	(0.002)	(0.001)	(0.002)
log(Affiliate Sales)		-0.025	-0.058	-0.033	-0.060
		(0.001)	(0.003)	(0.002)	(0.003)
N	131454	117419	111998	17157	83083
R^2	0.097	0.128	0.382	0.148	0.377
Affiliate Fixed Effect	No	No	Yes	No	Yes
Industry Fixed Effect	Yes	Yes	No	Yes	No
Country-year Fixed Effect	Yes	Yes	Yes	Yes	Yes

Standard errors are clustered at parent firm level. All coefficients are significant at 1% level, except for the log of parent firm's domestic sales in column 3. The dependent variable is the absolute value of forecast errors (log deviation), $|FE^{log}|$, in all regressions. Age is the age of the affiliate when making the forecasts. Regressions in columns 1, 2 and 3 include all affiliates, while the regression in column 4 only includes affiliate that continuously appeared in the sample from age 1 to age 7.

(i.e., age). The reduction in subjective uncertainty is economically significant compared to the average subjective uncertainty faced by entering affiliates.

We restrict our sample to first-time entrants into countries or regions that we identify using the founding year of the affiliates. We focus on affiliates in either the manufacturing sector or the wholesale and retail sector whose parent firms are in manufacturing.[11] As export data at the firm-destination country level are not available, we obtain information on parent firms' previous export experience at the regional level using the parent firm survey data.

We define previous export experience in a way similar to Conconi et al. (2016) and Deseatnicov and Kucheryavyy (2017). Due to the lumpiness of international trade, we define export entry if the firm does not export to the region for two consecutive years and starts exporting afterwards (the variable of *Exp Expe.* used in Table 6). Similarly, we define export exit if the firm stops exporting to the region for two consecutive years. For firms that have begun to export but have not exited yet, their previous export experience is positive and defined as the number of years since export entry. We assign zero year of export experience to firms that have exited export.

Comparing to existing studies of first-time entrants of Japanese MNEs (Deseatnicov and Kucheryavyy (2017)), our sample has fewer observations (see Table 5), as we only include first-time entrants that report sales at age two and projected sales at age one. However, we obtain very similar patterns regarding exporting and affiliate entry to existing studies.[12]

In Table 6, we provide evidence that previous export experience reduces the initial subjective uncertainty faced by the foreign affiliates that enter a country for the first time. We calculate the affiliates' absolute FEs at age one (log deviation of the realized sales at age two from the projected sales at age one) and regress this measure on various measures of previous export experience, controlling for industry fixed effects and country-year fixed effects. In columns 1 and 2, we use dummy variables that equal one if and only if the parent firm of the affiliate exported to the same region in the year (or in one of the two years) prior to MP entry. In column 3, we use the more sophisticated definition of export experience, and the dummy variable equals one if and only if export experience is positive. These regressions show that having previous export experience reduces absolute forecast errors by 13 log points. In column 4, we use a continuous

[11] Following Conconi et al. (2016), we include distribution-oriented FDI such as wholesale and retail in our analysis since affiliates in these industries may sell the same products as what the parent firms had previously exported. As a result, previous export experience helps reduce demand uncertainty for these distribution-oriented affiliates as well.

[12] In particular, the majority (73%) of the affiliates' parents in our sample have previous export experience to the region before their subsequent market entries (i.e., FDI). This share is higher than that of Norwegian MNE affiliates (39%) and French MNE affiliates (42%), as reported in Gumpert et al. (2016), but lower than that of Belgium MNE affiliates (86%), as reported in Conconi et al. (2016).

Table 5: Years of exporting experience before affiliate entry

	Frequency	Percent
0	187	28.4
1	48	7.3
2	44	6.7
3	46	7.0
4	35	5.3
5	43	6.5
6	35	5.3
7	28	4.3
8	30	4.6
9	19	2.9
10	38	5.8
11	32	4.9
12	18	2.7
13	22	3.3
14	15	2.3
15	18	2.7
Total	658	100.0

Only first-time entrant affiliates (into a country) that report their sales at age = 2, project sales at age = 1 and have nonmissing exporting experience are included in the sample.

measure of export experience instead of indicator variables. On average, one additional year of export experience reduces FE by 1.3 log points.

In Appendix 8.1, we provide a battery of robustness checks for Table 6. In one specification, we include parent firm size and affiliate size into the regression in order to control for firm-level heterogeneity between experienced and inexperienced affiliates. In another regression, we focus on horizontal FDI only (i.e., foreign affiliates that sell at least 1/3 of their sales in the hosting economy). In the final robustness check, we exclude intra-firm exports from total exports when measuring previous export experience of parent firms. All robustness checks yield qualitatively the same result as in the baseline regression. Taken together, we show that previous export experience is associated with lower initial uncertainty for first-time entering affiliates in the destination markets. This suggests that the existence of information value provided by exporting activities.

3.6 Fact 3: Correlation of Past Forecast Errors and Current Forecasts

In this subsection, we present evidence that is consistent with one key prediction of the learning model we are going to present in the theory section. Specifically, we show past FEs are positively correlated with current forecasts, and this positive correlation decreases when the firm becomes older and when the firm's parent in Japan has the previous export experience. The bottom two rows of Table 7 substantiate that the correlation coefficient is 0.08 and significantly different form zero at any conventional significance level. This holds for all firms, for manufacturing firms, and

Table 6: Forecast error and previous exporting

Dep.Var: $\lvert FE_{1,2}\rvert$	(1)	(2)	(3)	(4)
$Exp_{-1} > 0$	-0.159**			
	(0.065)			
$Exp_{-1} > 0$ or $Exp_{-2} > 0$		-0.151**		
		(0.064)		
Exp Expe. > 0			-0.132*	
			(0.070)	
Exp Expe.				-0.013**
				(0.006)
Industry FE	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes
N	553	561	658	658
R^2	0.486	0.499	0.472	0.472

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is affiliates' initial forecast error, which is calculated as the absolute log deviation of the realized sales at age = 2 from the projected sales (predicted by an affiliate at age = 1). We only include affiliates that are first-time entrants into a particular host country. Exporting experience (Exp Expe.) is defined at the continent level for each parent firm. Each column head indicates the different measure of exporting experience used in the regression.

for the first-time entrants into the the manufacturing sector. Then, we look at how firm age affects this positive covariance in the first four rows. It is clear that there is a declining trend for this positive correlation over the firm's life cycle, as firm age in Columns 1 and 3 is negatively significant. Moreover, the age dummies (relative to age one) in Columns 2 and 4 increase in absolute value over the firm's life cycle. Finally, we regress the correlation on the previous export experience for the first-time entrants, estimates in Columns 5 and 6 show that those firms with (positive) previous export experience *start* with lower levels of positive correlation between past FEs and current forecasts. In total, experience seems to reduce the positive correlation between past FEs and current forecasts, which suggests the existence of learning over the firm's life cycle.

3.7 Fact 4: Correlated Forecast Errors over Time

In this subsection, we present evidence on imperfect information and gradual learning in the international market. In the study of forecasting models (Mishkin (1983), Coibion and Gorodnichenko (2012), Andrade and Le Bihan (2013)), whether FEs made in different periods are positively correlated is used to detect the existence of information rigidity.[13] Intuitively, perfect information models imply that FEs (made by the forecasts in the current period for variables in next period) are all caused by *unexpected and contemporaneous* shocks that happen next

[13]The correlation of FEs over time refers to the serial correlation between $FE^{\log}_{t-1,t}$ and $FE^{\log}_{t,t+1}$, where $FE^{\log}_{t-1,t}$ refers to the error in period $t+1$ made by the forecast in period t. It is called the correlation between "rolling horizon" FEs as in Andrade and Le Bihan (2013).

Table 7: Effect of Firm Experience on $Cov(\log(R_t/E_{t-1}R_t), \log(E_tR_{t+1}))$

Dep.Var: $Cov(\log(R_t/E_{t-1}R_t), \log(E_tR_{t+1}))$	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	All		Manufacturing		Manu and First-time entrants	
Age	-0.002**		-0.002**			
	(0.001)		(0.001)			
Age=2		-0.030*		-0.037*		
		(0.017)		(0.020)		
Age=3		-0.040**		-0.067***		
		(0.017)		(0.021)		
Age=4		-0.044***		-0.060***		
		(0.017)		(0.021)		
Age=5		-0.041**		-0.046**		
		(0.017)		(0.021)		
Age=6		-0.040**		-0.051**		
		(0.017)		(0.020)		
Age=7		-0.045***		-0.060***		
		(0.017)		(0.020)		
Age=8		-0.042***		-0.061***		
		(0.016)		(0.020)		
Age=9		-0.045***		-0.055***		
		(0.016)		(0.020)		
Age=10		-0.047***		-0.062***		
		(0.015)		(0.019)		
$Exp_{-1} > 0$					-0.037	
					(0.079)	
$Exp_{-1} > 0$ or $Exp_{-2} > 0$						-0.074
						(0.087)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	117032	117032	81586	81586	354	358
R^2	0.014	0.014	0.014	0.014	0.356	0.360
$Corr(\log(R_t/E_{t-1}R_t), \log(E_tR_{t+1}))$	0.084		0.087		0.131	0.114
P-value	0.000		0.000		0.014	0.032

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is the product of residual of current forecast errors $\log(R_t/E_{t-1}R_t)$ and residual of sales forecasts for the next period $\log(E_tR_{t+1})$. The residuals are obtained by regressing the corresponding dependent variables on the same set of variables and fixed effects as shown in each column. Columns 1-2 include all firms. Columns 3-4 only include the manufacturing (or wholesale or retail) affiliates whose parent firms are in manufacturing. Column 5-6 further restricts to the sample of first-time entrants into particular host countries.

period. Thus, perfect information models imply that FEs made in two different periods are serially uncorrelated. Therefore, if the data exhibit serial correlation between FEs made in two different periods, imperfect information is present.[14] Moreover, if the positive serial correlation of FEs declines as time goes by, it suggests the existence of learning which solves the imperfect information problem over the firm's life cycle. We are going to show that both patterns exist in our data, which motivate an industry equilibrium featuring imperfect information and learning in the theory section.

Table 8: Serial correlation of forecast errors made in two consecutive years

	1	2	3	4	5
corr. ($FE^{\log}_{t-1,t}$, $FE^{\log}_{t,t+1}$)	0.124***	0.121***	0.145***	0.153***	0.146***
Manufacturing firms only?	No	Yes	Yes	Yes	Yes
Type of firms included	all firms	all manufacturing	entrants	survivors	entrants and survivors
N	178140	108135	11013	19968	9799

Notations: $FE^{log}_{t-1,t}$ is the log deviation of the realized sales in year t from the projected sales in year $t-1$, while $FE^{log}_{t,t+1}$ is the log deviation of the realized sales in year $t+1$ from the projected sales in year t. Top and bottom one percent observations of forecast errors are trimmed. Manufacturing firms including firms in wholesalers as well. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Manufacturing survivors refer to manufacturing affiliates that have survived for at least five years. Manufacturing entrants refers to manufacturing affiliates that entered the destination markets during our sample period.

We proceed our analysis in two steps. First, we present the (raw) correlation coefficients between FEs made in two consecutive periods in Table 8. The first two columns show that the correlation coefficient between FEs made in two consecutive years are positively and significantly correlated when we look at all the affiliates and affiliates in the manufacturing sector.[15] Next, when we focus on different subsamples of our data in Columns 3-5, the correlation coefficient is still positively significant and quantitatively similar t before. In short, the positive autocorrelation of FEs substantiates the existence of imperfect information in the data.

Interestingly, the (positive) serial correlation of FEs decreases with firm age. When we divide affiliates into four age groups, the correlation coefficient decreases as we move from a younger age cohort to an old age cohort, as shown by Table 9. As the serial correlation of FEs is indicative of information rigidity, Table 9 hints that information rigidity gets resolved (at least partially) over firm's life cycle.

Next, we run panel regressions to confirm the above two findings. Specifically, the regression

[14] The validity of this test is robust to different functional form and distributional assumptions we make in the model (e.g., whether the variance of the contemporaneous shock is age dependent and whether the fundamental shock is log normal).

[15] As what we have used in the last subsection, the manufacturing sector in the following empirical subsections includes the wholesale and retail sector as well.

Table 9: Serial correlation of forecast errors for different age groups

	age: 2-5	age: 6-8	age: 9-12	age$\geq$13
corr. $(FE^{\log}_{t-1,t}, FE^{\log}_{t,t+1})$	0.157***	0.123***	0.103***	0.109***
Manufacturing firms only?	Yes	Yes	Yes	Yes
N	13985	14278	18995	54021

Notations: $FE^{log}_{t-1,t}$ is the log deviation of the realized sales in year t from the projected sales in year $t-1$, while $FE^{log}_{t,t+1}$ is the log deviation of the realized sales in year $t+1$ from the projected sales in year t. Top and bottom one percent observations of forecast errors are trimmed. Manufacturing affiliates including those in retail and wholesale sectors as well. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

we run is

$$FE^{\log}_{i,t}(sales) = \beta FE^{\log}_{i,t-1}(sales) + \delta_{st} + \delta_{ct} + \varepsilon_{it}, \tag{2}$$

where δ_{ct} and δ_{st} represent the country-year and (affiliate) industry-year fixed effects. Index i refers to the affiliate. In the main specification, we include industry-year and country-year fixed effects into the regression. Regression results in Table 10 show that FEs made next year are positively and significantly correlated to FEs made in the current year, even after we have controlled for a battery of fixed effects. This is true for both the whole sample (i.e., manufacturing affiliates) and subsamples (e.g., the sample of survivors or entrants only). In addition, the conditional correlation coefficient is around 0.12, which is similar to the raw correlation coefficient reported in Table 8. Moreover, if we control for parent firm fixed effects in the above regression, the empirical result of the serial correlation of FEs is unchanged qualitatively, as shown by Table 24 in Appendix 8.2. In total, the data show the pattern of positively correlated FEs over time.

Table 10: Regression for the serial correlation of forecast errors

	(1)	(2)	(3)
		$FE^{\log}_{t,t+1}(sales)$	
$FE^{\log}_{t-1,t}(sales)$	0.106***	0.131***	0.120***
	(0.00689)	(0.0138)	(0.0187)
Type of firms included	manufacturing firms	manufacturing survivors	manufacturing entrants
Industry-year Fixed Effect	Yes	Yes	Yes
Country-year Fixed Effect	Yes	Yes	Yes
N	67790	13160	6787
adj. R^2	0.148	0.169	0.219

Notations: $FE^{log}_{t-1,t}$ is the log deviation of the realized sales in year t from the projected sales in year $t-1$, while $FE^{log}_{t,t+1}$ is the log deviation of the realized sales in year $t+1$ from the projected sales in year t. Top and bottom one percent observations of forecast errors are trimmed. Standard errors are in parentheses and clustered at the affiliate level. Manufacturing firms including firms in wholesalers as well. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Manufacturing survivors refer to manufacturing affiliates that have survived for at least five years. Manufacturing entrants refers to manufacturing affiliates that entered the destination markets during our sample period.

Now, we use regressions to show that the positive serial correlation of FEs gets attenuated, when the affiliate becomes olds and when its parent firm has previous export experience before the affiliate is set up. The regression equation we run is

$$\mathbb{1}\Big(Sign(FE^{\log}_{i,t}) = Sign(FE^{\log}_{i,t-1})\Big) = age_{i,t} + \delta_{ct} + affiliate_i + \varepsilon_{it}, \tag{3}$$

where δ_{ct} and $affiliate_i$ represent the country-year and the affiliate fixed effects. The indicator function, $\mathbb{1}\Big(Sign(FE^{\log}_{i,t}) = Sign(FE^{\log}_{i,t-1})\Big)$, equals one if FEs made in two consecutive years have the same sign (i.e., positive, negative or zero) and -1 otherwise. We use the a vector of age dummies (one to nine) or log age for the variable of $age_{i,t}$. In some of the regressions, we also control for affiliate size and parent size in order to tease out the (potential) size effect on the level of imperfect information. Results presented in Tables 11 and 25 (in the appendix) suggest that older firms tend to make less systematic FEs. In other words, older firms are less likely to over-forecast or under-forecast their next year's sales consecutively. This shows that experience helps firms learn about their demand and supply conditions in a world with imperfect information.[16]

Exactly following the same logic, we show that previous export experience helps first-time entering affiliates make less systematic FEs. Specifically, we run

$$\mathbb{1}\Big(Sign(FE^{\log}_{i,2}) = Sign(FE^{\log}_{i,1})\Big) = exp\ experience_i + \delta_{ct} + \delta_s + \varepsilon_{it}, \tag{4}$$

where δ_{ct} and δ_c represent the country-year and the (affiliate) industry fixed effects. As the dependent variable can only be defined for affiliates which are at least two years old, we utilize the correlation information of first-time entrants at age two in the above cross-sectional regression. Table 12 reveals that positive export experience substantially reduces the possibility that the first-time entrants make systematic FEs in the first two years after entry.[17] In total, our data of FEs show the existence of imperfect information at the firm level and its level decreases with firm's age or previous export experience.

[16] Although some of the estimates of the age dummies (and of the log of affiliate's age) become insignificant after we control for the affiliate size and the parent size, the quantitative magnitudes are stable across different specifications.

[17] Interestingly, affiliate size negatively affects the correlation of FEs (although insignificantly). This probably implies that larger firms face smaller information rigidity. Surprisingly, parent size positively affects the level of information rigidity, as they probably invest in projects with higher level of information rigidity in the destination economy.

Table 11: Age effects on the correlation of forecast errors (all firms)

Dep.Var: $\mathbb{1}(Sign(FE_{i,t}^{\log}) = Sign(FE_{i,t-1}^{\log}))$	(1)	(2)	(3)	(4)
Age=3	-0.0541**	-0.0302		
	(0.0312)	(0.0322)		
Age=4	-0.0247	-0.0177		
	(0.0324)	(0.0349)		
Age=5	-0.0593**	-0.0553		
	(0.0326)	(0.0348)		
Age=6	-0.0530	-0.0505		
	(0.0338)	(0.0357)		
Age=7	-0.0674***	-0.0426		
	(0.0343)	(0.0363)		
Age=8	-0.0982***	-0.0800***		
	(0.0345)	(0.0371)		
Age=9	-0.0847***	-0.0661**		
	(0.0349)	(0.0372)		
log(Affiliate Age)			-0.0578**	-0.0459
			(0.0302)	(0.0322)
log(Affiliate Sales)		-0.00525		-0.00597
		(0.00881)		(0.00884)
log(Parent Domestic Sales)		0.0127		0.0126
		(0.0117)		(0.0118)
N	92313	82861	92313	82861
R^2	0.193	0.201	0.193	0.201
Subsidiary FE	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable equals 1 if forecast errors made in two consecutive years have the same sign and -1 otherwise. Forecast error is calculated as the log deviation of the realized sales from the projected sales. Note that as the dependent variable can only be defined for affiliates which are at least two years old, the age dummies can be estimated from 3 (all relative to age 2 firms).

Table 12: Export experience and the correlation of forecast errors

Dep.Var: $\mathbb{1}(Sign(FE_{i,t}^{\log}) = Sign(FE_{i,t-1}^{\log}))$	(1)	(2)	(3)	(4)
$Exp_{-1} > 0$	-0.397**		-0.432**	
	(0.200)		(0.204)	
$Exp_{-1} > 0$ or $Exp_{-2} > 0$		-0.525**		-0.592***
		(0.203)		(0.209)
log(Parent Domestic Sales)			0.0738*	0.0710*
			(0.0376)	(0.0374)
log(Affiliate Sales)			-0.0282	-0.0272
			(0.0414)	(0.0407)
N	359	359	346	352
R^2	0.340	0.348	0.352	0.360
Industry FE	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable equals 1 if forecast errors made in two consecutive years have the same sign and -1 otherwise. Forecast error is calculated as the log deviation of the realized sales from the projected sales.

3.8 Fact 5: Both Current Sales and Forecasts Predict Future Sales

In this subsection, we present further evidence on the existence of imperfect information in the international market. Specifically, we show that the projected sales have statistically significant and economically strong impact on realized sales in the future. Specifically, we regress the realized log sales in year t on the logarithm of projected sales in year $t-1$ (for sales in year t), lagged log sales in year $t-1$ and $t-2$ and a set of fixed effects. Table 13 shows that sales forecasts not only positively and significantly affect the realized sales, but also have stronger predictive power than past sales.[18] This result verifies the reliability and usefulness of the data of sales forecasts reported by Japanese firms.[19] More importantly, it points out that a model with imperfect information is needed to rationalize this finding, as past sales should have no predictive power for future sales (conditioning on sales forecasts) in FIRE models.

Table 13: Both Current Sales and Forecasts Predict Future Sales

Dep.Var: $\log(R_t)$	(1)	(2)	(3)	(4)	(5)
Sample:	All	All	All	Manufacturing	Manu. & Survivors
$\log(E_{t-1}(R_t))$	0.968***	0.716***	0.660***	0.725***	0.777***
	(0.002)	(0.011)	(0.013)	(0.012)	(0.018)
$\log(R_{t-1})$		0.254***	0.251***	0.246***	0.186***
		(0.010)	(0.016)	(0.011)	(0.016)
$\log(R_{t-2})$			0.072***		
			(0.008)		
Industry-year FE	Yes	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes	Yes
N	134110	132636	111447	91716	13198
R^2	0.939	0.947	0.955	0.950	0.938

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is affiliates' log sales in period t. Regressors are affiliates' log forecasts about R_t at time $t-1$ and lagged log sales. Columns 1-3 include all firms. Column 4 only includes the manufacturing (or wholesale or retail) affiliates whose parent firms are in manufacturing. Column 5 further restricts to affiliates that have survived at least 7 years (from age one to age seven) in our sample.

4 An Industry Equilibrium Model of Firm Learning and Market Entry

In this section, we propose a dynamic industry equilibrium model of trade and MP to match all the above facts and to study how imperfect information and learning impact aggregate

[18]We find similar empirical results when regressing future investment or employment on sales forecasts and past sales. Results are available upon request.

[19]If the affiliates randomly or non-rationally reported their projected sales, the projected sales should have no or weak predictive power for future sales

productivity and gains from liberalizing trade. The model features firm learning about their demand (a la Arkolakis et al. (2017)) and information rigidity (similar to Mankiw and Reis (2002)). After describing the setup and equilibrium of the full model, we show that both mechanisms are needed to match all the facts under simplifying assumptions.

4.1 Setup: Demand and Supply

In our model, there are two countries, Japan and the foreign country. Each Japanese firm produces a differentiated variety and has to decide whether to serve the foreign market at all, and if so, whether through exporting or MP. We do not explicitly model Japanese domestic firms and ignore domestic demand for two reasons. First, it simplifies the model and helps us highlight between the trade-off between trade and MP. Second, since we do not have a representative sample of Japanese domestic firms, we avoid using moments on domestic firms to discipline related parameters.[20] In this sense, our model is partial equilibrium in its nature.

In the foreign country, the representative consumer has the following nested-CES preferences where the first nest is among composite goods produced by firms from different countries, indexed by i,

$$U_t = \left(\sum_i \chi_i^{\frac{1}{\delta}} Q_{it}^{\frac{\delta-1}{\delta}} \right)^{\frac{\delta}{\delta-1}},$$

and the second nest is among varieties $\omega \in \Sigma_{it}$ produced by firms from each country i,

$$Q_{it} = \left(\int_{\omega \in \Sigma_{it}} e^{\frac{a_t(\omega)}{\sigma}} q_t(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}. \tag{5}$$

In the first nest, the parameter χ_i is the demand shifter for country i goods, and the parameter δ is the Armington elasticity between goods produced by firms from different countries. In the second nest, the parameter σ is the elasticity between different varieties, and $a_t(\omega)$ is the demand shifter for variety ω. We assume that firms differ in their demand shifter, $a_t(\omega)$. After denoting foreign consumers' total expenditure as $\tilde{Y}_t$, we can express the demand for a particular Japanese variety, ω, as:

$$q_t(\omega) = \frac{\tilde{Y}_t}{\tilde{P}_t^{1-\delta}} \chi_{jp} P_{jp,t}^{\sigma-\delta} e^{a_t(\omega)} p_t(\omega)^{-\sigma}, \tag{6}$$

where $\tilde{P}_t$ is the aggregate price index for all goods, and $P_{jp,t}$ is the ideal price index for Japanese

[20] The Basic Survey of Japanese Business Structure and Activities does contain firms that do not export or conduct MP. However, since the threshold for the survey is quite high (50 workers and 30 mil yen of total assets), it misses a large number of small domestic firms. We think it is more representative for exporters and multinational parent firms.

goods. When the Armington elasticity δ equals 1, the first nest is Cobb-Douglas, and the expenditure on Japanese goods no longer depend on $P_{jp,t}$. When $\sigma = \delta$, the elasticities in the two CES nests are the same, which is the case in prominent trade models such as Eaton and Kortum (2002) and Melitz (2003). In our calibration, we set δ to be a value between 1 and σ.

In our model, we assume that the Japanese varieties make up a smaller fraction of foreign consumers' consumption and treat $\tilde{Y}_t$ and $\tilde{P}_t$ as exogenous. As a result, we can combine the exogenous terms in expression (6), $\tilde{Y}_t \tilde{P}_t^{\delta-1} \chi_{jp}$ into one variable, Y_t, and call it the aggregate demand shifter. In addition, since we only focus on Japanese firms, we suppress the subscript jp in the following analysis. The CES preference over different varieties of Japanese goods implies the ideal price index

$$P_t \equiv \left(\int_{\omega \in \Sigma_t} e^{a_t(\omega)} p_t(\omega)^{1-\sigma} d\omega \right)^{1/(1-\sigma)}. \tag{7}$$

We use the term, $A_t \equiv Y_t P_t^{\sigma-\delta}$ to denote the aggregate demand condition faced by all firms in period t and rewrite the firm-level demand function as

$$q_t(\omega) = A_t e^{a_t(\omega)} p_t(\omega)^{-\sigma}. \tag{8}$$

We specify firm-specific demand as follows. For each firm, the demand uncertainty comes from the demand shifter $a_t(\omega)$. We assume that $a_t(\omega)$ is the sum of a time-invariant fundamental demand draw $\theta(\omega)$ and a transitory shock $\varepsilon_t(\omega)$ as in Arkolakis et al. (2017):

$$a_t(\omega) = \theta(\omega) + \varepsilon_t(\omega), \ \varepsilon_t(\omega) \overset{i.i.d.}{\sim} N\left(0, \sigma_\varepsilon^2\right). \tag{9}$$

Firms understand that $\theta(\omega)$ is drawn from a normal distribution $N\left(\bar{\theta}, \sigma_\theta^2\right)$, and the independent and identically distributed (i.i.d.) transitory shock, $\varepsilon_t(\omega)$, is drawn from another normal distribution $N\left(0, \sigma_\varepsilon^2\right)$. Since θ is time-invariant, we call it the "fundamental demand (draw)" in what follows. We assume that the firm does not know the value of its fundamental demand. Each period, the firm only observes the realize demand, $a_t(\omega)$, and cannot distinguish between its two components. Instead, the firm forms an posterior belief about the distribution of θ according to the Bayes' rule.

In addition, we assume that there are two types of firms in the economy in the same spirit as in sticky information models a la Mankiw and Reis (2002): the informed firms and the uninformed firms. Initially, all entering firms are uninformed in the sense that they use the prior distribution of θ (i.e., $N(\bar{\theta}, \sigma_\theta^2)$) to form the expectation. At the end of each period, $1-\alpha$ fraction of the remaining uninformed firms become informed. From the time when they become informed, they begin to update the beliefs by utilizing realized demand shifters and will never become uninformed again. The remaining uninformed firms still use the prior distribution of θ to form posterior beliefs.

Firms produce only using labor. Different from the demand side, we assume that firms are homogeneous in labor productivity. In order to produce q units of output, the firm has to employ the same amount of workers. We deviate from ex-ante heterogeneity in productivity as modeled in Arkolakis et al. (2017), as our data (see Table 26 in the appendix) show that experienced MNEs are larger and more productive than inexperienced ones. Heterogeneity in ex-ante labor productivity would imply the opposite (see Appendix 8.3 for details). Japanese firms can either employ domestic workers to produce at a constant wage w, or employ foreign workers at w^* and conduct MP. Therefore, the implicit assumption is that Japanese exporters and multinationals employ a small fraction of the labor force both domestically and abroad, so their activities cannot affect the wages. Both wages are normalized to one when we calibrate our model.

The industry structure features monopolistically competition. There is an exogenous mass of potential entrants J (from Japan) that decide whether or not to enter the foreign market every period. Each entrant draws a fundamental demand shifter θ from a normal distribution, $N\left(\bar{\theta}, \sigma_\theta^2\right)$, and a MP sunk entry cost f_m^e from a log-normal distribution, $logN(\mu_{f_m^e}, \sigma_{f_m^e}^2)$. The entrant knows f_m^e but does not know θ. If the firm chooses to enter the market, it also has to decide how to serve the foreign market. A potential entrant can either serve the foreign market via exporting, which involves a sunk entry cost of f_x^e, or serve the foreign market by setting up an affiliate with the cost f_m^e. Both sunk costs are paid in units of domestic labor. If neither mode is profitable, the potential entrant simply exits and obtains zero payoff.

Similar to entrants, incumbents do not know the exact value of θ. However, the informed incumbents have better information for the fundamental demand than entrants, as they can utilize information on past sales to update the beliefs for the fundamental demand. The uninformed incumbents have the same information as the entrants, as they cannot utilize information on past sales to update their beliefs. In each period, the incumbents first receive an exogenous death shock with probability η. For surviving firms, they have to decide whether to change their mode of service. They can keep their service mode unchanged, or switch to another mode of service (i.e., from exporting to MP or from MP to exporting). In addition, they can also choose to permanently exit the market. Firms have to pay a one-time entry cost, f_x^e, in order to start exporting. However, we assume that incumbent MNEs can switch to exporting without paying this sunk cost. Firms also have to pay a fixed cost each period in order to remain exporting (with a fixed cost of f_x) or doing MP (with a fixed cost of f_m). Therefore, firms with low demand draws will choose to exit.

For firms that serve the foreign market, they decide how much to produce in period t before the overall demand shock a_t is realized. After the demand shifter in period t is realized, they choose the price p_t in order to sell all the products they have produced, as we assume there is no storage technology and firms cannot accumulate inventories. For informed incumbent firms, they update their beliefs about the fundamental demand after observing the demand shifter in period t. Additionally, a randomly selected $1-\alpha$ fraction of uninformed firms become informed at the end of each period.

In the steady-state equilibrium, all entrants and incumbents make optimal decisions on the choice of production modes, quantities and prices. The distribution of firms over all state variables is invariant over time. In the appendix, we discuss the value functions and the definition of the steady-state equilibrium in detail.

In the calibration exercise, we normalize the sunk entry cost into exporting, f_x^e, to zero, as we do not use the information on domestic production of Japanese firms to calibrate this parameter. In addition, since we do not observe that multinational affiliates exit more when they are young, we set the fixed per-period operation cost of MP, f_m, to zero. In other words, all MNE exits are caused by the exogenous death shock.

4.2 Belief Updating

In this subsection, we discuss how the firm forms its ex post belief for the fundamental demand draw. For an informed firm, we assume it has observed its realized value in the past t periods: $a_1, a_2, \ldots, a_t$. Therefore, this firm switched from the "uninformed" status to the "informed status" at $t = 1$. Since both the prior and the shocks are normally distributed, we apply the Bayes' rule and the posterior belief about θ is normal with mean μ_t and variance σ_t^2 where

$$\mu_t = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + t\sigma_\theta^2}\bar{\theta} + \frac{t\sigma_\theta^2}{\sigma_\varepsilon^2 + t\sigma_\theta^2}\bar{a}_t, \tag{10}$$

and

$$\sigma_t^2 = \frac{\sigma_\varepsilon^2 \sigma_\theta^2}{\sigma_\varepsilon^2 + t\sigma_\theta^2}. \tag{11}$$

The history of signals $(a_1, a_2, \ldots, a_t)$ is summarized by age t and the average

$$\bar{a}_t \equiv \frac{1}{t}\sum_{i=1}^{t} a_i \text{ for } t \geq 1; \quad \bar{a}_0 \equiv \bar{\theta}.$$

Therefore, the firm believes that the overall demand shock in period $t+1$, $a_{t+1} = \theta + \varepsilon_{t+1}$, has a normal distribution with mean μ_t and variance $\sigma_t^2 + \sigma_\varepsilon^2$. For an uninformed firm, its belief for the mean and variance of θ is the same as the prior belief.

4.3 Static Optimization of Per-Period Profit

We study the firm's static optimization problem in the steady state in this subsection. As all aggregate variables such as wages, the ideal price index and expenditure on Japanese goods are constant in the steady state, we omit the subscript t whenever possible. In each period, conditional on the mode of service, a firm's decision about how much to produce is a static problem. Firms hire labor and produce q_t quantity of output to maximize expected per-period profit given its belief about the demand shock a_t. The realized per-period profit for an affiliate ($o = m$) or an exporter ($o = x$) is

$$\pi_{o,t} = p_t(a_t)q_t - MC_o \times q_t - wf_o,$$

where the marginal cost of production depend on the mode of service

$$MC_o = \begin{cases} \tau w & \text{if } o = x; \\ w^* & \text{if } o = m, \end{cases}$$

where w and w^* denote the domestic and foreign wages, respectively.

Firms set the price after observing the realized demand a_t to sell all the output so the price and quantity must have the relationship as in equation (6). Firms choose quantity q_t to maximize the expected per-period profit $E_{a_t|\bar{a}_{t-1},t}\left(\pi_{o,t}\right)$. (For an uninformed firm, $t = 1$) Solving the first-order conditions, we obtain

$$q_{o,t} = \left(\frac{\sigma - 1}{\sigma}\right)^{\sigma} \left(\frac{b\left(\bar{a}_{t-1}, t-1\right)}{MC_o}\right)^{\sigma} \frac{Y}{P^{\delta-\sigma}}, \tag{12}$$

where

$$b\left(\bar{a}_{t-1}, t-1\right) = E_{a_t|\bar{a}_{t-1},t-1}\left(e^{a_t/\sigma}\right) = \exp\left\{\frac{\mu_{t-1}}{\sigma} + \frac{1}{2}\left(\frac{\sigma_{t-1}^2 + \sigma_{\varepsilon}^2}{\sigma^2}\right)\right\}. \tag{13}$$

The corresponding price is

$$p_{o,t}\left(a_t\right) = \frac{\sigma}{\sigma - 1} e^{a_t/\sigma} \frac{MC_o}{b\left(\bar{a}_{t-1}, t-1\right)}. \tag{14}$$

The resulting per-period profit function is

$$E\pi_{o,t} = \frac{(\sigma - 1)^{\sigma-1}}{\sigma^{\sigma}} \frac{b\left(\bar{a}_{t-1}, t-1\right)^{\sigma}}{{MC_o}^{\sigma-1}} \frac{Y}{P^{\delta-\sigma}} - wf_o. \tag{15}$$

In Appendix 7.1, we characterize firm's dynamic choice of market entry and production mode. As a result, we state value functions for various types of firms (e.g., exporters, multinational affiliates). In Appendix 7.2, we state the definition of the recursive competitive equilibrium of our model and specify the law of motions for the state variables. We skip these content here to save space.

4.4 Intuition on Matching Facts about Forecasts and FEs

In this subsection, we show how our model is able to match facts 1-5 presented in Section 3. We illustrate the intuition using a special case in which there is no endogenous switching of production modes. The detailed proof can be found in Appendix 7.5. In Appendix 7.3 and 7.4, we show that perfect information models and the original Jovanovic model can not be used to rationalize all the five empirical facts documented in Section 3.

Proposition 1 *When there is no endogenous switching of production modes, the forecasts and forecast errors of exporters and multinational firms' sales have the following properties*

1. *Variance of forecast errors declines with years of experience.*
2. *Forecast errors made in the last period are positively correlated with current forecasts. Moreover, the covariance between these two terms declines with years of experience.*
3. *Forecast errors made in two consecutive periods by the same firm are positively correlated when $\alpha > 0$ but are uncorrelated when $\alpha = 0$. When $\alpha > 0$, the positive correlation between these two terms declines with years of experience.*
4. *Both forecasted sales in the past and past sales have predictive power for current sales.*

Proof. See Appendix 7.5. ■

Here we discuss the intuition behind the proof. Given the CES demand and given that we define the forecast error as $FE_{t,t+1} = log(R_{t+1}/E_t R_{t+1})$, FE of sales is a linear function of FE of θ. Loosely speaking, we can instead consider the properties of the forecasts of θ. When there is no selection, the variance of the posterior belief coincides with the variance of the forecast errors of θ since θ is distributed as the prior.

Both learning and the reduction in information rigidity over time contribute to the first property. Due to learning, informed firms accumulate more experience and have clearer information about their fundamental demand, when they operate in the market for a longer period of time. Second, as more firms become informed over time and informed firms make more accurate forecasts (compared to the uninformed firms), the variance of FEs goes down with firm's market experience. These two forces together rationalize Fact 1. When there is endogenous mode switching, since exporting also helps firms to accumulate experience, it is a natural result the learning mechanism also helps to match Fact 2.

The second property comes from the fact that informed firms incorporate the current realizations into forecasts about future sales. As firms accumulate more experience, the forecast errors contain less information about θ thus less correlated with firms' forecasts about θ. This is consistent with Fact 3.

Third, the third property rationalize Fact 4 in Section 3. The positive correlation is caused by uninformed firms (being uninformed in two consecutive periods), as their forecasts do not change across two consecutive periods. In the proof, we show that firms that are informed in both periods and firms that switched from the uninformed status to the informed status, their forecast errors are serially uncorrelated. Therefore, information rigidity is necessary to match Fact 4, while Jovanovic-type learning alone cannot do so. As the share of uninformed firms decline with experience, the positive autocorrelation also declines.

The final prediction depends on the existence of both the informed firms and the uninformed firms. For the informed firms, as forecasted sales are BLUE for next period's sales, only forecasted sales matter for sales in the next period. For the uninformed firms, as their forecasts are constant over time, only past sales matter for sales in the next period. The coexistence of the informed firms and the uninformed firms rationalize Fact 5. Taken together, we need both Bayesian learning and sticky information in order to rationalize all the five empirical facts we have documented in the empirical section.

5 Quantitative Analysis

In this section, we describe the procedures for calibrating the dynamic model proposed above. The calibrated model is able to capture the decline in absolute value of FEs over affiliates' life cycle and for affiliates with previous export experience (Facts 1-2). It is also able to match the correlation of FEs and the predictive power of forecasts and past sales for current sales qualitatively (Facts 4-5). In addition, the calibrated model is able to capture other salient features of the data, such as growth in exporters' sales and the decline in exit rates over their life cycles, which we do not directly target in the calibration. After calibrating the model, we consider two counterfactual experiments: an increase in the variance of the two types of shocks and reductions in the iceberg trade costs.

5.1 Calibration

We first normalize a set of model primitives which are not identified since our model is not a general equilibrium model. Specifically, aggregate demand shifter Y, the wage rate in Japan, w, and the wage rate in the foreign countries, w^* are normalized to one. The mean of the logarithm of the fundamental demand draw, μ_θ, is normalized to zero. We also normalize the entry costs into exporting, f_x^e, to zero, as we abstract from Japanese domestic firms in the model.[21]

Next, we calibrate a set of parameters without solving our model, as listed in Table 14. We set the elasticity of substitution between varieties, σ, to 4, a common value in the literature (see Bernard et al. (2003); Arkolakis et al. (2013)). The Armington elasticity among goods from different countries, δ, is set to 2, an intermediate value between the Cobb-Douglas case ($\delta = 1$) and the elasticity between different varieties σ. We set the discount factor, β, to 0.96, which implies a real interest rate of four percent.

The exogenous death rate η and the FDI per-period fixed costs f_m are crucial for the exit rates of multinational affiliates. Because there is strong selection in the model, affiliates' exit rate would decline over their life cycles if the FDI per-period fixed costs are positive. However, we did not find a significant decline for affiliates' exit rate over their life cycle, even for affiliates without export experience.[22] Therefore, we postulate that $f_m = 0$ and set η to 0.03 so that the model can match the average exit rate of affiliates (3%). Appendix 8.4 provides evidence.

Table 14: Parameters calibrated without solving the model

Parameters	Description	Value	Source
σ	Elasticity of substitution between Japanese goods	4	Bernard et al. (2003)
δ	Armington elasticity between goods from different countries	2	
β	Discount factor	0.96	4% real interest rate
η	Exogenous death rate	0.03	Average exit rates of multinational affiliates
f_m	FDI per-period fixed costs	0	Flat profile of affiliates' exit rate over their life cycles

Three parameters that govern information rigidity and learning can also be backed out without calibrating the model. Since we have shut down the endogenous exit for multinational affiliates, there is no selection on the fundamental demand draw, θ, among multinational affiliates after entry. Also, as entrants have the same prior belief for their fundamental demand shocks, they choose to become MNEs based on the realized entry cost of FDI, f_m^e and do not en-

[21]Specifically, moments that can be used to pin down f_x^e, such as the share of exporters relative to domestic firms, are not available. In this case, we should interpret the entry costs into FDI, f_m^e, as the entry costs into FDI *relative to exporting.*

[22]See Appendix 8.4 for details.

dogenously exit after entry. Thus, the formulas derived in Appendix 7.4 and 7.5 can be directly applied to calculated variance and auto-covariance of FEs for inexperienced multinationals. In particular, we target the variance of FEs at age one and age ten of inexperienced MNEs, since they are most informative about σ_θ and σ_ε, respectively. We also target the covariance of FEs at age one and two, since the positive autocovariance is caused by information rigidity in our model, which is governed by the parameter α.[23] The calibrated value of α is 0.21, which implies that 79% of uninformed firms become informed each period and less than one percent of the firms are still uninformed after three years.

Table 15: Parameters related to forecast errors and moments

Parameters	Value	Description	Moments	Data	Model
σ_θ	2.05	Std of time-invariant shock	Var. of FE at age 1	0.48	0.48
σ_ϵ	0.90	Std of transitory shock	Var. of FE at age 10	0.24	0.24
α	0.21	prob of awaking	Cov of $FE_{1,2}$ and $FE_{2,3}$	0.034	0.034

The remaining four parameters are jointly calibrated by solving the equilibrium and matching four moments. The parameters are: the per-period fixed cost of exporting, f_x, the mean and standard deviation of log FDI entry cost, $\mu_{f_m^e}$ and $\sigma_{f_m^e}$, and the iceberg trade costs τ. The four targeted moments are the average exit rate of exporters, the fraction of exporters among active firms, the fraction of experienced affiliates at age one and the share of exports in total sales (i.e., total exports plus total sales of MNEs' affiliates).

In Table 16, we list the moments in an order such that, loosely speaking, each moment is the most informative about the parameter in the same row. A higher export per-period fixed cost raises the exporter exit rate, while a higher average FDI entry cost increases the fraction of exporters among all firms selling in the foreign market. In the model, only firms with small enough FDI entry costs become inexperienced MNEs in the first period. Thus, $\sigma_{f_m^e}$ affects the share of inexperienced MNEs positively. Finally, the iceberg trade costs have a large impact on the intensive margin of export (i.e., the sales share of exporters among all active firms).

5.1.1 Untargeted Moments

We now turn to untargeted moments of the calibrated model. We first examine the dynamics of FEs for affiliates with and without export experience. We then consider the dynamics of

[23]In practice, due to partial-year effects, age-one firms in the data may have less information than age-one firms in the model. We therefore assume that age-one firms in the data correspond to a mixture of age-zero and age-one firms in the model (with equal weights), and age-two firms in the data correspond to a mixture of age-one and age-two firms in the model (with equal weights). We then adjust the formulas derived in Appendix 7.4 and 7.5 accordingly.

Table 16: Parameters calibrated by solving the model and matching moments

Parameters	Value	Description	Moments	Data	Model
f_x	0.0053	export fixed cost	average exit rate of exporters	0.10	0.11
$\mu_{f_m^e}$	1.59	mean of log FDI entry cost	fraction of exporters among active firms	0.70	0.69
$\sigma_{f_m^e}$	2.45	Std of log FDI entry cost	fraction of experienced MNEs at age 1	0.73	0.75
τ	1.46	iceberg trade cost	Exporter sales share	0.21	0.21

exporters and affiliates in terms of sales growth and endogenous exit over their life cycles. Finally, we discuss other untargeted moments related to facts 4-5 and growth volatility of sales.

Dynamics of Forecast Errors We examine the changes in $\left|FE^{\log}\right|$ over affiliates' life cycles in Figure 3. We first estimate the age effects on $\left|FE^{\log}\right|$ for affiliates that enter a host country for the first time. We interact the age fixed effects with a dummy variable indicating whether the parent firm has previous export experience in the same region. We plot the estimated fixed effects for experienced and inexperienced MNEs in the left panel of Figure 3, using the age-one inexperienced MNEs as the base group. In the right panel, we plot the average $\left|FE^{\log}\right|$ by affiliate age predicted by the calibrated model, again normalizing the average $\left|FE^{\log}\right|$ to zero for inexperienced MNEs at age one.[24]

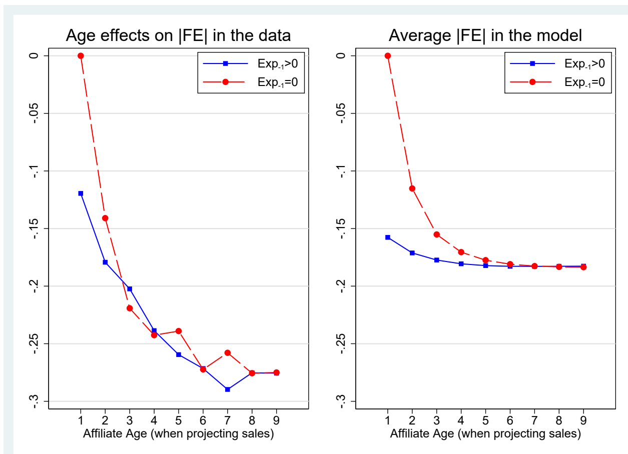
Consistent with the data, the model predicts that average $|FE^{\log}|$ declines over affiliates' life cycles and that the initial $\left|FE^{\log}\right|$ is lower for affiliates whose parent firms have export experience. However, the model predicts a much smaller decline in $\left|FE^{\log}\right|$ for experienced MNEs and a larger difference between experienced MNEs and inexperienced MNEs at age one.[25] The main reason for this discrepancy is that firms in the model learn very fast - there is almost no uncertainty about θ after four periods. This implies experienced firms starts with very precise forecast about θ and the main source of FEs is the transitory transitory shock, ε. A possible improvement we can make is to introduce learning about both demand and supply and assume that exporting only helps the firm learn foreign demand (and not supply conditions in the destination economy).

Dynamics of Sales In Figure 4, we compare the age profile of exports in the model to that in the data. The red dashed line represents the average log of sales by exporter age predicted by the model, while the blue solid line represents the corresponding moments in the data. The two lines are very well aligned. Note that learning and information rigidity per se do not generate growth in sales, since firms may receive either better or worse signals than their time-invariant

[24]Note that as we target at the variance of FEs at age one and ten in the calibration, absolute value of FEs at age one and ten are untargeted moments.

[25]For example, in the model, average $\left|FE^{\log}\right|$ of experienced MNEs drops by 0.02 over their life cycles, while the corresponding empirical moment is 0.16.

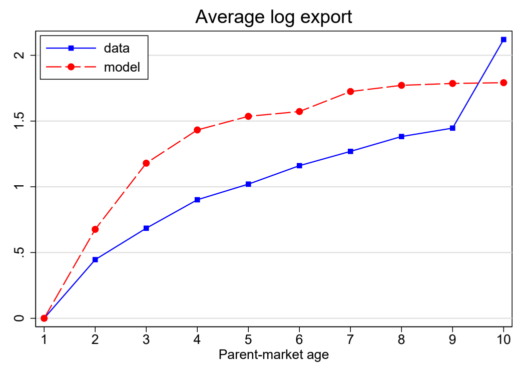
Figure 3: Forecast error - age profile: data v.s. model



Note: The left panel shows the estimated age effects on average $|FE^{log}|$ for affiliates in the data, while the right panel shows the average $|FE^{log}|$ by affiliate age in the model. To calculate the average $|FE^{log}|$ at age t in the model, we adjust the partial-year effects by averaging the forecast errors of affiliates at age $t-1$ and age t, since most affiliates enter into the destination market in the middle of each fiscal year. The blue solid line shows the estimated age effects for affiliates whose parent firms have previous export experience, i.e., $Exp_{-1} > 0$, while the red dashed line shows the estimated age effects for affiliates without previous export experience, i.e., $Exp_{-1} = 0$. The age effect of the affiliates without previous export experience is normalized to zero.

demand. The growth of average firm size is a result of learning *together with selection* on θ in the model. In contrast, the model does not capture the growth of average log sales of the multinational affiliates, as we shut down endogenous exits of affiliates to match the flat profile of exit rate (with firm age). Other mechanisms, such as the accumulation of customer capital, is needed to match the dynamics of sales for multinational affiliates (e.g., Fitzgerald et al. (2016)).

Figure 4: Export-age profile: data v.s. model



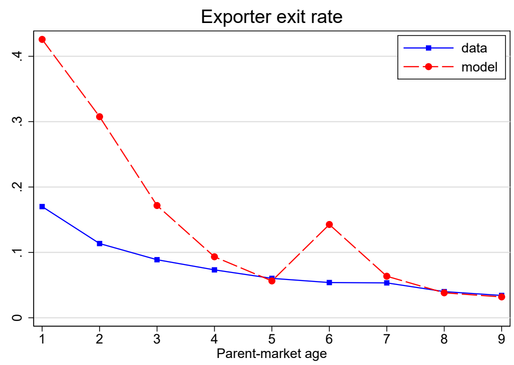
Note: The blue solid line represents average log(exports) by exporter age in the data, while the red dashed line represents the average log of exports by exporter age in the model.

Dynamics of Exits Figure 5 compares the exporter exit rates by their ages in the model to those in the data. Consistent with the data, the model predicts a series of declining exit rates with age for exporters. Overall, the model predicts higher exporter exit rates (by each age group), even though the model is able to match the average exporter exit rate in the data. This occurs as the distribution of exporters is more skewed towards old firms in the model compared to that in the data.

Standard Deviation of Sales Growth In Figure 6, we compare the age profile of the standard deviation of sales growth in the model to that in the data. As discussed in Appendix 7.4.2, Jovanovic-type learning models can *endogenously* generate age-declining volatility of sales growth *without* assuming age-declining variance of time-variant shocks. In Jovanovic-type learning models, sales growth volatility is largely determined by the volatility of firm's belief (for the fundamental demand) over time, as the firm uses the belief to determine output. As firm's belief fluctuates more when they are younger than older, sales growth volatility is also higher when firms are younger than older. Moreover, the variance of FEs is highly related to the variance of firm's belief and thus the variance of sales (and output) growth. As we target at the variance of FEs at age one and age ten in the calibration, our calibrated model matches the age profile of standard deviations of sales growth (of multinational affiliates) reasonably well.

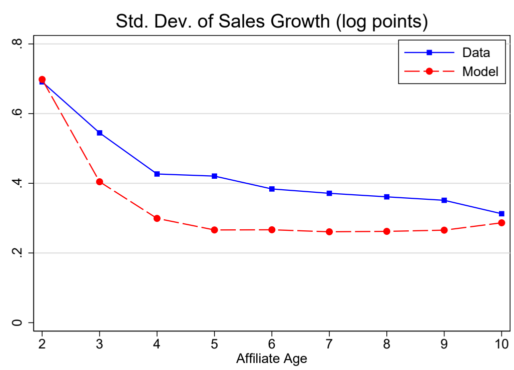
Other Untargeted Moments Using the calibrated parameters, we also simulate data to

Figure 5: Exit rate - age profile: data v.s. model



Note: The blue solid line represents the estimated age effects on the probability of exiting the export market in the data, while the red dashed line represents the average exit rate of exporters by age in the model.

Figure 6: Standard Deviation of sales growth - age profile: data v.s. model



Note: The blue solid line represents the average standard deviation of sales growth of multinational affiliates by age in the data, while the red dashed line represents the average standard deviation of sales growth of multinational affiliates by age in the model.

Table 17: Autocorrelation of Forecast Errors

Dep.Var: $FE^{\log}(t, t+1)$	Data		Model	
	(1)	(2)	(3)	(4)
$FE^{\log}(t-1, t)$	0.093***	-0.096***	0.028***	-0.084***
	(0.009)	(0.009)	(0.008)	(0.007)
Year FE	No	No	Yes	Yes
Country-year FE	Yes	Yes	No	No
Industry-year FE	Yes	Yes	No	No
Affiliate FE	No	Yes	No	Yes
N	44161	42501	23294	23038
R^2	0.166	0.366	0.002	0.138

Standard errors are clustered at affiliate level, * 0.10 ** 0.05 *** 0.01. In the first column, only first-time entrants into particular countries are included.

show that our calibrated model can match the positive serial correlation of FEs and the predictive power of forecasts for realized sales *qualitatively* well (i.e., facts 4 and 5). These results are shown by Tables 17 and 18. In summary, our calibrated model is able to capture the patterns of the exporter dynamics, the FE dynamics and age-declining sales growth volatility observed in the data, which are not targeted in the calibration.

Table 18: Both forecast and past sales predict future sales

Dep.Var: $\log(R_t)$	Data (1)	Model (2)
$\log(E_{t-1}(R_t))$	0.693***	0.965***
	(0.013)	(0.008)
$\log(R_{t-1})$	0.278***	0.036***
	(0.013)	(0.008)
N	60034	23294
R^2	0.951	0.984
Level of Fixed Effects	Country-Year, Industry-Year	Year

Standard errors are clustered at affiliate level, * 0.10 ** 0.05 *** 0.01. In the first column, only first-time entrants into particular countries are included.

5.2 Ex ante Uncertainty and Ex post Volatility

In this section, we use the calibrated model to examine the effects of uncertainty on trade/MP patterns as well as welfare. We distinguish between two types of uncertainty: (ex-ante) uncertainty about the fundamental demand before entry, captured by σ_θ, and (ex-post) uncertainty in the temporary demand shock σ_ε. We are interested in the effects of these parameters because we find that our firm-level idiosyncratic uncertainty measures are correlated with economic policy uncertainty (EPU, see appendix 8.5) and evidence suggests that these two parameters are not perfectly correlated across countries/regions (see appendix 8.6). We provide more empirical exploration on the forecast errors and their relation to EPU in the appendix.[26]

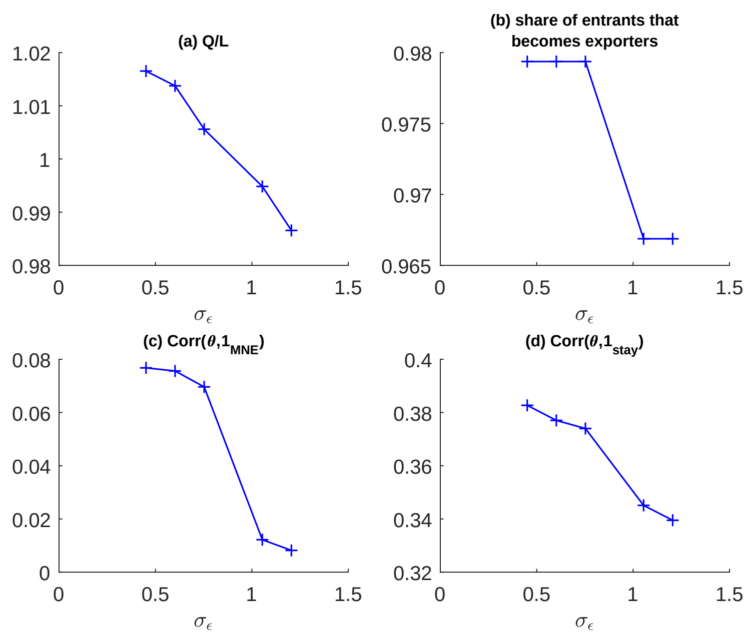
We first increase the ex-post uncertainty σ_ε from 0.5 to 1.2. Since a higher σ_ε increases the mean of the demand shifter $e^{(\theta+\varepsilon)/\sigma}$, we subtract $\Delta\sigma_\varepsilon^2/2\sigma^2$ from $\bar{\theta}$ (note that $\bar{\theta} = 0$ in the baseline calibration) to ensure that it is a mean-preserving spread.

Figure 7 shows four equilibrium outcomes for different values of σ_ε. A bigger σ_ε implies a smaller signal-to-noise ratio, $\sigma_\theta^2/\sigma_\varepsilon^2$, which leads to less effective learning and thus weaker incentives to use exporting as a way to learn about the fundamental demand. Therefore, more

[26] It is unclear through which mechanism EPU is correlated with firm-level uncertainty. One explanation is that governments that conduct uncertain aggregate economic policies may also engage in policies that generates time-varying firm-level distortions in the product or factor markets. In contrast, a more "rule-based" government will engage less in such policies at the micro and aggregate level.

entrants choose to do MP directly without learning via exporting (panel b). With perfect information, only firms with the highest θ become multinationals given a certain level of MP entry cost. In contrast, firms make "mistakes" under imperfect information. When σ_ε is larger, learning becomes less effective so firms make more "mistakes" at the margins of MP entry and export exit, which causes a lower correlation between θ and conducting MP (panel c) and a lower correlation between θ and staying in the market (panel d). This eventually causes lower aggregate labor productivity (panel a).[27]

Figure 7: Counterfactuals with respect to σ_ε



We now increase ex-ante uncertainty σ_θ from 1.6 to 3.4. We again subtract $\Delta\sigma_\theta^2/2\sigma^2$ from $\bar{\theta}$ to ensure that these are mean-preserving spreads for the demand shifter $e^{(\theta+\varepsilon)/\sigma}$.[28] The effect of an increase in ex ante uncertainty on the dynamic selection is exactly the opposite to the effect of an increase in ex post volatility, as it implies a higher signal-to-noise ratio and therefore more effective learning. In short, the dynamic selection is positively affected by such

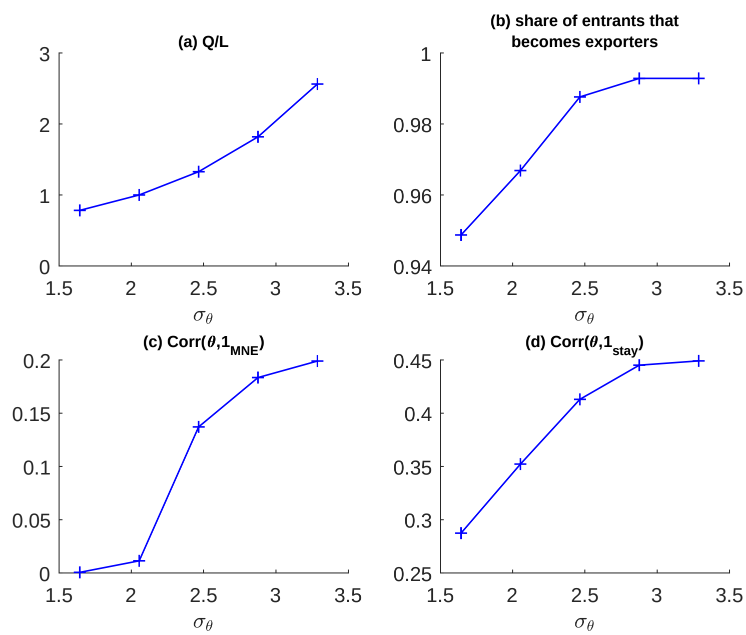
[27] Since our model is partial equilibrium in nature, it is not ideal for welfare analysis. We think aggregate labor productivity is a better measure of welfare than the ideal price index, as it takes into account the resources drawn into the "foreign market sector" by payment of fixed and entry costs.

[28] A caveat here is that the variance of the time-invariant demand draws reflects the heterogeneity of producers in an economy. It is driven by production technologies and consumer preferences in the economy and not directly related to any macro or industry policies we can think of. However, we still implement such a counterfactual analysis, as we want to contrast the impact of ex ante uncertainty (i.e., the variance of θ) on the dynamic selection with the effect of ex post volatility (i.e., the variance of ε).

an increase, as both the correlation between θ and being a multinational firm (panel c) and the correlation between θ and staying in the market (panel d) increase when σ_θ becomes bigger. Eventually, these effects translate into higher aggregate productivity when the variance of ex ante uncertainty increases, which is the opposite to the effect of an increase in ex post volatility.

One point worth mentioning is that since varieties are substitutes to each other, consumers prefer a larger heterogeneity in the demand shifter a_t. Therefore, both higher ex-ante and ex-post uncertainty can generate a mechanical welfare improvement due to the substitution effect. The result that higher σ_ε lower aggregate labor productivity suggests the information effect dominates the substitution effect. On the other hand, the effect of higher σ_θ on labor productivity overstates the information effect.

Figure 8: Counterfactuals with respect to σ_θ



In summary, firm-level uncertainty can be caused by both ex-ante and ex-post uncertainty. However, the effects of these two types of uncertainty on the dynamic selection and aggregate productivity are completely different. Therefore, in order to understand the effect of uncertainty on aggregate and dynamic economic outcomes, it is important to distinguish between different sources of uncertainty.

5.3 Gains from Trade Liberalization and the Role of Imperfect Information

In this subsection, we investigate how the existence of imperfect information affects gains in aggregate productivity from reducing trade costs in a world with multiple production modes (i.e., exporting and MP). The key mechanism is the increasing information value of exporting, when the trade costs go down which makes it easier to stay in the market. Crucially, this increasing information value associated with trade liberalization only exists in a world with imperfect information, and is magnified when there are two production modes instead of one.

To quantify such complementarity, we calculate "gains from trade" under alternative assumptions of the model as in Ramondo and Rodríguez-Clare (2013). In particular, we lower the iceberg trade cost τ from 1.46 (the calibrated value) to 1 (no variable trade costs). We then calculate the change in labor productivity ($GT \equiv (Q'/L')/(Q/L)$). Moreover, we look at two alternative worlds: one with perfect information and the other one with imperfect information.[29] We also investigate productivity gain from trade in two alternative trade regimes: one with both trade and MP and the other one with trade only (as we push up MP entry costs to infinity in this case). The gain in aggregate productivity under the trade-only regime is denoted by GT^* (a hypothetic world), and the gain in aggregate productivity under the other regime is denoted by GT (i.e., the calibrated world).

Table 19 shows two interesting results. First, the gain in aggregate productivity is always larger in a world with imperfect information than in a world without, and this does not depend on the type of trade regime. When the trade costs go down which makes it easier for firms to stay in the market and learn about themselves, the allocation into different production modes gets improved. Therefore, in addition to the conventional channels for gains from trade (e.g., the productivity effect, the variety effect etc.), reducing trade costs generates increasing information value in a world with imperfect information which amplifies gains from trade. Second, the difference in the aggregate productivity gain from trade (between a world with imperfect information and a world without) is larger in a world with both trade and MP than in a world with trade only. The additional information value generated by trade liberalization (in a world with imperfect information) is amplified when multiple production modes are available, as there are more choices to be made at the extensive margins. In addition, there is little change in the productivity gain from reducing trade costs, when we shut down the sticky information part *only* by setting α to zero (i.e., Jovanovic-type learning is still present). This is shown by the middle row of Table 19 ($\alpha = 0$). Therefore, the existence of gradual learning accounts for most of the variation in the gain from trade, when we move from an imperfect information world to a

[29] In a world with perfect information, we assume that firms know their fundamental demand perfectly and before the entry. As a result, the model becomes a static model, as the choices at the extensive margins are static.

world with perfect information.[30] In total, the existence of imperfect information complements gains from trade, and this complementarity becomes bigger when both exporting and MP are feasible.

Table 19: Complementarity between Trade and MP

Welfare Measure	Labor Productivity Q/L		
	GT	GT^*	GT/GT^*
Imperfect Info (full model)	1.17	1.47	0.79
Imperfect Info ($\alpha = 0$)	1.16	1.47	0.79
Perfect Info	1.08	1.46	0.74

5.4 Efficiency-Improving Subsidies

In this subsection, we study the social planner's problem and the difference in resource allocation between the social planner (who faces the same imperfect information structure as the decentralized economy) and the decentralized economy. We show that the difference in resource allocation only exists at the extensive margin. Specifically, the social planner would make the same output choices as what firms choose in the decentralized economy. However, the entry/exit choices may be different between the social planner's solution and the decentralized economy. This further rationalizes why productivity gain from reducing trade costs should be larger in an imperfect information world with multiple production modes than in an imperfect information world with a single production mode.

Proposition 2 *The social planner would choose the same levels of output as what firms choose in the decentralized economy, when both of them face the same imperfect information structure.*

Proof. See Appendix 7.6. ■

6 Conclusion

In this paper, we use a unique data set of Japanese MNEs, which contains information on sales forecasts at the affiliate level, to detect information imperfection and learning in the international market. We document several new and important stylized facts concerning affiliates' forecasts and FEs. We view these facts as direct evidence for the existence of age-dependent

[30]Note that the calibrated α is small in the model (0.21) and less than one percent of firms are still informed after three period. This probably explains why the sticky information part does not matter for the variation in gain from trade quantitatively.

firm-level uncertainty, imperfect information and learning. We then build up a dynamic industry equilibrium model of trade and MP, which features information rigidity and learning, in order to explain the documented facts.

To quantify the role of learning and uncertainty in exporters and multinational firms' dynamics, we extend the standard firm learning model (Jovanovic (1982); Arkolakis et al. (2017)) into an international setting. Specifically, we add sticky information a la Mankiw and Reis (2002) and the endogenous choice between exporting and MP into the canonical firm learning model and calibrate the extended model to match moments related to FEs and aggregate trade/MP variables. The calibrated model is able to replicate the new facts about affiliates' FEs, as well as other salient features of the data.

We then conduct counterfactual analysis regarding firm-level uncertainty. We find that changing the variance of the time-invariant demand draws and that of the temporary demand shocks have different implications for the entry margins of trade and MP and aggregate productivity. In particular, a higher variance of time-invariant draws implies a higher signal-to-noise ratio and leads to more effective learning (and stronger incentives to use exporting to obtain information). To the contrary, a higher variance of the transitory shocks implies a lower signal-to-noise ratio and leads to less effective learning. As a result, dynamic selection is negatively affected by ex post volatility, which leads to lower aggregate productivity. This exercise shows that when we analyze how uncertainty affects aggregate outcomes, it is crucial to distinguish between the two sources of uncertainty discussed above. We also show that productivity gains from reducing trade costs are larger in a world with imperfect information (than in a world without). Moreover, the difference in the gains (between a world with imperfect information and a world without) becomes larger, when there are multiple production modes instead of a single production mode. This shows the information value generated by trade liberalization, whose aggregate implications have not been studied very much in the existing literature.

7 Theory Appendix

7.1 Dynamic Choices of Production Modes and Value Functions

In each period, an entrant or incumbent firm can choose among three different service modes: exiting, exporting (denoted as x) or FDI (denoted as m). To become an exporter or MNE, a firm must pay a sunk cost. A firm's state variables include its mode of service in previous period, o, the draw of sunk entry cost into FDI, f_m^e, its current market experience, t, the history of demand shocks, $\bar{a}_{t-1}$ (from the period when the firm becomes informed), and whether or not the firm is informed, $in \in \{0, 1\}$. Since firms make optimal decisions based on their belief about θ rather than the true value of θ, these variables are sufficient to characterize the value functions and policy functions of the firm.

An incumbent exporter can choose to stay exporting, become an MNE or exit. If it wants to be an MNE, it has to pay the sunk cost f_m^e in units of domestic labor.[31] We derive the value function for the informed incumbent firm first. For an informed exporting firm, the value function prior to choosing the mode of service (and after the death shock) in period t is given by

$$V\left(x, f_m^e, t, \bar{a}_{t-1}, in = 1\right) = \max_{o' \in \{x, m, exit\}} E_t \left\{ \begin{array}{c} \pi_{x,t} + \beta(1-\eta) V\left(x, f_m^e, t+1, \bar{a}_t, in = 1\right), \\ \pi_{m,t} - w f_m^e + \beta(1-\eta) V\left(m, f_m^e, t+1, \bar{a}_t, in = 1\right), \\ V_{exit} \end{array} \right\}, \tag{16}$$

where $t \geq 2$ and the value of exiting, V_{exit}, is normalized to zero. We denote the optimal choice of the service mode that maximizes equation (16) as $o'\left(x, f_m^e, t, \bar{a}_{t-1}, in = 1\right) \in \{x, m, exit\}$. All expectations in equation (16) are calculated using the firm's subjective belief about the distribution of the demand shock a_t in the current period. A note here is that $\bar{a}_{t-1}$ is calculated from the time when the firm becomes informed. In other words, it equals $\frac{1}{t-t_0+1} \Sigma_{i=t_0}^{t} a_i$ where $t_0 (\leq t)$ is the period when the firm switches from being uninformed to being informed. In particular, it does not necessarily equal $\frac{1}{t} \Sigma_{i=1}^{t} a_i$. Since a multinational affiliate does not need to pay the sunk cost when switching to exporting, the value of being an informed incumbent MNE prior to the endogenous choice of service mode (and after the death shock) in period t is

[31] As we abstract from the analysis of the labor market, whether the MNE uses domestic labor to pay the entry cost is irrelevant.

given by

$$V\left(m, f_m^e, t, \bar{a}_{t-1}, in=1\right) = \max_{o' \in \{x,m,exit\}} E_t \left\{ \begin{array}{c} \pi_{x,t} + \beta(1-\eta)V\left(x, f_m^e, t+1, \bar{a}_t, in=1\right), \\ \pi_{m,t} + \beta(1-\eta)V\left(m, f_m^e, t+1, \bar{a}_t, in=1\right), \\ V_{exit} \end{array} \right\}. \tag{17}$$

We denote the optimal choice of the service mode for such a firm as $o'\left(m, f_m^e, t, \bar{a}_{t-1}, in=1\right) \in \{x, m, exit\}$.

Now, we derive the value functions for the uninformed incumbent firms. For an uninformed exporting firm, the value function prior to choosing the mode of service (and after the death shock) in period t is given by

$$V\left(x, f_m^e, t, \bar{a}_0, in=0\right) = \max_{o' \in \{x,m,exit\}} E_t \left\{ \begin{array}{c} \pi_{x,t} + \beta(1-\eta)[\alpha V\left(x, f_m^e, t+1, \bar{a}_0, in=0\right) \\ +(1-\alpha)V\left(x, f_m^e, t+1, a_t, in=1\right)], \\ \pi_{m,t} - w f_m^e + \beta(1-\eta)[\alpha V\left(m, f_m^e, t+1, \bar{a}_0, in=0\right) \\ +(1-\alpha)V\left(m, f_m^e, t+1, a_t, in=1\right)], V_{exit} \end{array} \right\}, \tag{18}$$

and we denote the optimal choice of the service mode for such a firm as $o'\left(x, f_m^e, t, \bar{a}_0, in=0\right) \in \{x, m, exit\}$. Note that if the firm is uninformed at the beginning of period t, it will become informed at the end of period t with probability $1-\alpha$. In addition, the history of demand shocks changes from $\bar{a}_0$ (i.e., the prior belief) to a_t when the firm becomes informed at the end of period t, as it utilizes past information on sales only from the period when it becomes informed. All expectations in equation (18) are calculated using firms' subjective belief about the distribution of the demand shock a_t in the current period. Since a multinational affiliate does not need to pay the sunk cost in the case of switching to exporting, the value of being an uninformed incumbent MNE prior to the endogenous choice of service mode (and after the death shock) in period t is

$$V\left(m, f_m^e, t, \bar{a}_0, in=0\right) = \max_{o' \in \{x,m,exit\}} E_t \left\{ \begin{array}{c} \pi_{x,t} + \beta(1-\eta)[\alpha V\left(x, f_m^e, t+1, \bar{a}_0, in=0\right) \\ +(1-\alpha)V\left(x, f_m^e, t+1, a_t, in=1\right)], \\ \pi_{m,t} + \beta(1-\eta)[\alpha V\left(m, f_m^e, t+1, \bar{a}_0, in=0\right) \\ +(1-\alpha)V\left(m, f_m^e, t+1, a_t, in=1\right)], V_{exit} \end{array} \right\}. \tag{19}$$

We denote the optimal choice of the service mode in period t for an uninformed incumbent MNE as $o'\left(m, f_m^e, t, \bar{a}_0, in=0\right) \in \{x, m, exit\}$.

For the entrant, it is always uninformed and simply chooses the service mode that yields the highest value:

$$o'\left(ent,f_m^e,1,\bar{a}_0,in=0\right)=\arg\max_{o'\in\{x,m,exit\}}E_1\left\{\begin{array}{c}\pi_{x,t}-wf_x^e+\beta(1-\eta)[\alpha V\left(x,f_m^e,2,\bar{a}_0,in=0\right)\\+(1-\alpha)V\left(x,f_m^e,2,a_1,in=1\right)]\\\pi_{m,t}-wf_m^e+\beta(1-\eta)[\alpha V\left(m,f_m^e,2,\bar{a}_0,in=0\right)\\+(1-\alpha)V\left(m,f_m^e,2,a_1,in=1\right)],V_{exit}\end{array}\right\}.\tag{20}$$

7.2 Steady-state Recursive Competitive Equilibrium

We discuss the stationary equilibrium of the model in this subsection as follows.

1. value functions $V\left(o,f_m^e,t,a_{t-1},in\right)$, $o\in\{x,m\}$ that satisfy equations (16), (17), (18) and (19);

2. policy functions of the mode choice $o'\left(o,f_m^e,t,\bar{a}_{t-1},in\right)$ ($o\in\{ent\}$ if $t=1$ while $o\in\{x,m\}$ if $t\geq 2$) that maximize value functions defined in equations (16), (17), (18) and (19);

3. policy functions of optimal output $q_o,o\in\{m,x\}$ that satisfy equation (12).

4. prices given the demand shock in the current period $p_o\left(a_t\right),o\in\{m,x\}$ that satisfy equation (14);

5. a measure function of firms $\lambda\left(o,f_m^e,t,\bar{a}_t,\theta,in\right)$, $o\in\{x,m,ent\}$ that is consistent with the aggregate law of motion. This measure function of firms is defined at the beginning of each period and after the exogenous exit takes place. In particular, an exogenous mass J of entrants draw θ from a log-normal distribution $G_\theta\left(\cdot\right)$ each period, respectively. Therefore, the measure of entrants with state variables θ is

$$\lambda\left(ent,f_m^e,1,\bar{a}_0,\theta,in=0\right)=Jg_\theta\left(\theta\right)g_{f_m^e}\left(f_m^e\right),$$

where $g_\theta\left(\cdot\right)$ and $g_{f_m^e}\left(\cdot\right)$ are the density functions of the log normal distribution for θ and f_m^e respectively. The cumulative demand shock $\bar{a}_0$ for entrants is defined to be the same as the prior mean of θ. All entrants are uninformed in the first period. The measure function of the exporters (and that of MNEs) is the fixed point of the law of motion of this measure function. Given any Borel set of $\bar{a}_t$, Δ_1, measures of uninformed and informed firms with service mode $o'\in\{x,m\}$ at the beginning of period $t+1$ satisfy

$$\begin{aligned}&\lambda\left(o',f_m^e,t+1,\Delta_1,\theta,in=0\right)\\=\;&\alpha\sum_{o\in\{x,m,ent\}}\int\begin{array}{c}\mathbf{1}\left(\bar{a}_t\in\Delta_1,o'\left(o,f_m^e,t,\bar{a}_{t-1},in=0\right)=o'\right)\\\times\left(1-\eta\right)\Pr\left(\bar{a}_t|\bar{a}_{t-1},\theta\right)\lambda\left(o,f_m^e,t,d\bar{a}_{t-1},\theta,in=0\right)\end{array}\end{aligned}$$

and

$$\begin{aligned}
&\lambda\left(o', f_m^e, t+1, \Delta_1, \theta, in=1\right) \\
= \; &\sum_{o\in\{x,m,ent\}} \int \begin{array}{c} \mathbf{1}\left(\bar{a}_t \in \Delta_1, o'\left(o, f_m^e, t, \bar{a}_{t-1}, in=1\right)=o'\right) \\ \times(1-\eta)\Pr\left(\bar{a}_t|\bar{a}_{t-1},\theta\right)\lambda\left(o, f_m^e, t, d\bar{a}_{t-1}, \theta, in=1\right) \end{array} \\
&+(1-\alpha)\sum_{o\in\{x,m,ent\}} \int \begin{array}{c} \mathbf{1}\left(\bar{a}_t \in \Delta_1, o'\left(o, f_m^e, t, \bar{a}_{t-1}, in=0\right)=o'\right) \\ \times(1-\eta)\Pr\left(\bar{a}_t|\bar{a}_{t-1},\theta\right)\lambda\left(o, f_m^e, t, d\bar{a}_{t-1}, \theta, in=0\right) \end{array},
\end{aligned}$$

where $\mathbf{1}(...)$ is an indicator function. Note that at the beginning of the first period (i.e., $t=1$), all firms' service modes are defined to be *ent* (i.e., entrants). Therefore, there are no exporters or MNEs at age one in the stationary distribution, i.e., $\lambda\left(o, f_m^e, 1, \bar{a}_0, \theta, in\right)=0$ for $o\in\{x,m\}$. In contrast, at the beginning of each later period ($t\geq 2$), all incumbents' modes of service are either exporting or FDI (i.e., x or m). Thus, there are no entrants at $t\geq 2$, i.e., $\lambda\left(o, f_m^e, t, \bar{a}_{t-1}, \theta, in\right)=0$ for $o=ent$ and $t\geq 2$.

6. the price index P is constant over time and must be consistent with consumer optimization (7):

$$P^{1-\sigma}=\sum_{t\geq 1}\sum_{\substack{in\in\{0,1\}\\ o\in\{x,m,ent\}}}\int_{f_m^e,\bar{a}_{t-1},\theta,a_t} \begin{array}{c} e^{a_t}p_o\left(a_t, q_o\left(b\left(\bar{a}_{t-1}, t-1\right), in\right)\right)^{1-\sigma} \\ \times\lambda\left(o, df_m^e, t, d\bar{a}_{t-1}, d\theta, in\right)d\Pr\left(a_t|\theta\right) \end{array}.$$

Given each guess of P, we can solve for the value functions and policy functions, as well as the measure function for firms, λ. We iterate on the value of P until it converges in the simulation nd calibration.

7.3 Full Information Rational Expectation Models

In this subsection, we derive FE in FIRE models in detail. Base on the optimal output choice derived in the main text we have the log of *realized* sales as

$$\log(R_t(\theta))=\log(A_t)+\frac{\sigma\theta+\varepsilon_t+\frac{(\sigma-1)\sigma_\varepsilon^2}{2\sigma}}{\sigma}+(\sigma-1)\Big[\log\Big(\frac{\sigma-1}{\sigma w}\Big)\Big].$$

Since the firm knows θ in FIRE models, its best estimate for a_t is θ as ε_t is a transitory i.i.d. shock. Therefore, the conditional distribution of a_t in any period is $N\left(\theta,\sigma_\varepsilon^2\right)$, and the log of forecasted sales is

$$\log\Big(E_{t-1}^S(R_t)\Big)=\log(A_t)+\frac{\sigma\theta+\frac{\sigma_\varepsilon^2}{2}}{\sigma}+(\sigma-1)\Big[\log\Big(\frac{\sigma-1}{\sigma w}\Big)\Big],$$

which leads to

$$FE^{\log}_{t-1,t}(sales) = \frac{\varepsilon_t}{\sigma} - \frac{\sigma^2_\varepsilon}{2\sigma^2}. \tag{21}$$

Note that when the firm forecasts its sales one period in advance, it still faces an *unpredictable* random shock (i,e., ε_t) which causes FEs even in the full information model. Note that as the contemporaneous transitory shock in demand follows the log normal distribution, higher variance of the transitory shock pushes up the forecasted sales, price and output.

Next, we consider the case in which the fundamental demand shock, θ, follows an AR(1) process.[32] Specifically, we assume that θ is time-varying and follows:

$$\theta_{t+1} = \rho_1 \theta_t + \zeta_{t+1}, \tag{22}$$

where ζ_t is a draw from a normal distribution $N\left(0, \sigma^2_\zeta\right)$ and serially uncorrelated. The resulting FE and serial correlation of FEs are

$$FE^{\log}_{t-1,t}(sales) = \frac{\varepsilon_t}{\sigma} - \frac{\sigma^2_\varepsilon}{2\sigma^2} + \frac{\zeta_t}{\sigma} - \frac{\sigma^2_\zeta}{2\sigma^2} \tag{23}$$

and

$$cov(FE^{\log}_{t-1,t}, FE^{\log}_{t,t+1}) = cov\left(\frac{\varepsilon_t + \zeta_t}{\sigma}, \frac{\varepsilon_{t+1} + \zeta_{t+1}}{\sigma}\right) = 0,$$

as contemporaneous innovations in θ are serially uncorrelated by definition.

Finally, we discuss how endogenous exit would change full information rational expectation (FIRE) model's prediction on the serial correlation of FEs. Different from the case without selection, we now have endogenous selection in the sense that firms with the fundamental demand draws below a certain cutoff, $\bar{\theta}_t$, exit the market in period t. Therefore, survivors in both periods t and $t+1$ must have

$$\rho_1 \theta_{t-1} + \zeta_t \geq \bar{\theta}_t \quad \rho_1(\rho_1 \theta_{t-1} + \zeta_t) + \zeta_{t+1} \geq \bar{\theta}_{t+1}, \tag{24}$$

as firms choose whether or not to exit based on the realization of θ_t in period t and forecast is made in the end of period $t-1$. Conditional on θ_{t-1} and survival in both periods, there is a negative correlation between ζ_t and ζ_{t-1} implied by equation (24). The intuition is that a better contemporaneous fundamental demand shock last period makes survival in this period easier. This leads to a negative correlation between the contemporaneous fundamental shocks in two consecutive periods, conditioning on survival.

Finally, we discuss serial correlation between past FEs and current forecasts in FIRE models.

[32] The analysis for AR(m) process is similar.

First, we discuss the case in which θ is time-invariant. In this case, we have

$$cov(FE^{\log}_{t-1,t}, Forecast^{\log}_{t,t+1}) = cov\Big(\frac{\varepsilon_t}{\sigma}, \theta\Big) = 0.$$

Next, we discuss the case in which θ follows an AR(1) process specified in equation (22). As the Japanese affiliate forecasts its sales in period t at the beginning of period t (in the data), it knows the realization of θ_t when making the forecast for sales in period t. Therefore, FE at period t is still

$$FE^{\log}_{t-1,t}(sales) = \frac{\varepsilon_t}{\sigma} - \frac{\sigma^2_\varepsilon}{2\sigma^2}.$$

As a result, we have

$$cov(FE^{\log}_{t-1,t}, Forecast^{\log}_{t,t+1}) = cov\Big(\frac{\varepsilon_t}{\sigma}, \theta_{t+1}\Big) = 0.$$

In total, FIRE models cannot be used to rationalize FEs which are correlated with future FEs and future forecasts.

7.4 Jovanovic Model and Bayesian Learning

In this subsection, we discuss the empirical predictions of the original Jovanovic model (without information rigidity) in terms of our five empirical findings. Note that the original Jovanovic model is a special case of our full model with $\alpha = 0$. The next proposition shows that the FEs in two consecutive periods are uncorrelated:

Proposition 3 *Forecast errors made in two consecutive periods by the same firm are uncorrelated.*

Proof. We derive realized sales and forecasted sales first. The log of realized sales and the log of forecasted sales are

$$\log(R_t(\theta)) = \log(A_t) + \frac{\theta + \varepsilon_t}{\sigma} + (\sigma - 1)\log b\left(\bar{a}_{t-1}, t-1\right) + (\sigma - 1)\Big[\log\Big(\frac{\sigma - 1}{\sigma w}\Big)\Big], \tag{25}$$

and

$$\log\Big(E^S_{t-1}(R_t)\Big) = \log(A_t) + \sigma \log b\left(\bar{a}_{t-1}, t-1\right) + (\sigma - 1)\Big[\log\Big(\frac{\sigma - 1}{\sigma w}\Big)\Big]$$

respectively. The resulting FE becomes

$$FE^{\log}_{t-1,t}(sales) = \frac{\varepsilon_t + (\theta - \mu_{t-1})}{\sigma} - \frac{\sigma^2_{t-1} + \sigma^2_\varepsilon}{2\sigma^2}. \tag{26}$$

Jovanovic model predicts that FEs made in two consecutive periods are uncorrelated. To see this, we rewrite FE as

$$
\begin{aligned}
FE^{\log}_{t-1,t}(sales) &= \frac{(1-\zeta(t-1,\lambda))(\theta-\bar{\theta})+\varepsilon_t-\zeta(t-1,\lambda)\frac{\Sigma_{i=1}^{t-1}\varepsilon_i}{t-1}}{\sigma} \\
&\quad -\frac{\sigma^2_{t-1}+\sigma^2_{\varepsilon}}{2\sigma^2},
\end{aligned}
$$

where the signal-to-noise ratio is defined as

$$
\lambda \equiv \frac{\sigma^2_{\theta}}{\sigma^2_{\varepsilon}}; \quad \zeta(t-1,\lambda) \equiv \frac{(t-1)\lambda}{1+(t-1)\lambda}.
$$

The implied covariance of FEs in two consecutive periods is

$$
cov(FE^{\log}_{t-1,t}, FE^{\log}_{t,t+1}) = \frac{1}{\sigma^2}\left[\frac{\lambda\sigma^2_{\varepsilon}}{(1+\lambda t)(1+\lambda(t-1))} - \frac{\lambda t\sigma^2_{\varepsilon}}{t(1+\lambda t)} + \frac{\lambda t\lambda(t-1)\sigma^2_{\varepsilon}}{t(1+\lambda t)(1+\lambda(t-1))}\right] = 0.
$$

■

The above result is a common property of Bayesian updating with non-biased prior (i.e., the agent has the same prior as the data generating process). In our model, every firm's prior mean for θ is $\bar{\theta}$. As Bayesian updating with unbiased prior yields the best linear unbiased estimator (BLUE) for the fundamental demand θ based on past information, FEs in period t (i.e., $FE^{\log}_{t,t+1}$) are uncorrelated to any variable that has been realized up to period t, which include FEs and (log) sales in periods $t-1$. Therefore, Jovanovic model cannot generate positively correlated FEs over time. Moreover, forecast in period t should perfectly predicts sales in period $t+1$, as it is BLUE for sales in period $t+1$. In Appendix 7.4.1, we show that this result holds in the steady state, even if we allow θ to follow an AR(1) process.

Although the Jovanovic model cannot generate positive correlation of FEs over time, it can explain the fact that the variance of FEs goes down with firm age and market experience. Based on equation (26), we can derive the variance of FEs at age t as

$$
\begin{aligned}
var(FE^{\log}_{t-1,t}(sales)) &= \frac{var(\varepsilon_t+\theta-\mu_{t-1})}{\sigma^2} \\
&= \frac{var(\varepsilon_t)+var(\theta)+var(\mu_{t-1})-2cov(\theta,\mu_{t-1})}{\sigma^2},
\end{aligned}
$$

as ε_t is independent of θ and μ_{t-1}. The next proposition shows that this variance decreases with firm age and previous experience in the market:

Proposition 4 *Variance of the forecast errors declines with years of experience t.*

Proof. We can rewrite the expression of the variance of FE derived in the main text as:

$$var(FE^{\log}_{t-1,t}(sales)) = \frac{(1+\frac{\zeta(t-1,\lambda)^2}{t-1})\sigma_\varepsilon^2}{\sigma^2} + \frac{(1-\zeta(t-1,\lambda))^2\sigma_\theta^2}{\sigma^2}, \tag{27}$$

where

$$\zeta(t-1,\lambda) \equiv \frac{(t-1)\lambda}{1+(t-1)\lambda}.$$

Differentiating equation (27) with respect to t leads to

$$-\frac{\partial\zeta(t-1,\lambda)}{\partial t}\sigma_\varepsilon^2\Big[2(1-\zeta(t-1,\lambda))\lambda - \frac{2\zeta(t-1,\lambda)}{t-1}\Big] - \sigma_\varepsilon^2\frac{\zeta(t-1,\lambda)^2}{(t-1)^2},$$

which equals

$$-\sigma_\varepsilon^2\frac{\zeta(t-1,\lambda)^2}{(t-1)^2} < 0.$$

Therefore, variance of FEs declines with t. ■

Thanks to the accumulation of experience, the variance of FEs becomes smaller, when the affiliate becomes older and when the affiliate's parent firm has the previous export experience.[33] Importantly, the Jovanovic model can also generate a positive correlation between past FEs and current forecasts, which declines with firm age and market experience. This is consistent with the empirical finding in Section 3.6.

Proposition 5 *Past forecast errors are positively correlated with current forecasts. Moreover, this positive correlation declines with years of experience t.*

Proof. Simple calculation shows that covariance of past FEs and current forecast is

$$\begin{aligned} & cov(FE^{\log}_{t-1,t}, forecast^{\log}_{t,t+1}) \\ = \ & \frac{1}{\sigma}\frac{\lambda\sigma_\varepsilon^2}{1+\lambda(t-1)} > 0, \end{aligned}$$

where we have used the following property of belief updating in Jovanovic model:

$$\mu_t = \mu_{t-1} + \frac{\lambda(a_t-\mu_{t-1})}{1+\lambda t} \quad FE^{\log}_{t-1,t} = \frac{a_t-\mu_{t-1}}{\sigma} - \frac{\sigma_{t-1}^2+\sigma_\varepsilon^2}{2\sigma^2}.$$

Note that the positive correlation (or covariance) goes down when the firm becomes older or has accumulated experience (via previous export experience) before subsequent FDI entry. When the

[33] This result generalizes to the absolute value of FEs immediately, as average FEs is zero for each age cohort.

firm is infinitely old, this positive correlation disappears as the firm learns about its fundamental demand perfectly. ■

In summary, Jovanovic model can explain the first three empirical facts documented above, but not the last two. This motivates us to extend Jovanovic model *at the minimum level* to rationalize all the five empirical findings.[34]

7.4.1 Time-varying Fundamental Demand Shock

In this subsection of the appendix, we discuss the case when the fundamental demand θ is time-varying. In particular, we impose that θ_t follows an AR(1) structure:

$$\theta_{i,t} = \rho\theta_{i,t-1} + \zeta_{i,t}$$

and

$$a_{i,t} = \theta_{i,t} + \varepsilon_{i,t}.$$

We assume the firm can only observe $a_0, ..., a_{t-1}$ up to the beginning of period t. This is an variant of Muth (1960)'s model. The optimal forecasting rule follows:

$$forecast_{i,t} = (\rho - K_{t-1})forecast_{i,t-1} + K_{t-1}a_{i,t-1},$$

where K_{t-1} is is the Kalman gain. Note that $forecast_{i,t}$ is the belief formed at the beginning of period t without observing $a_{i,t}$. To be consistent with the notation in earlier sections, we denote forecast with μ. We also drop the firm subscript i for simplicity.

Forecast error (FE) for the hidden state variable θ_{t+1} is

$$\begin{aligned}
e_{t+1} &= \theta_{t+1} - \mu_{t+1} \\
&= \theta_{t+1} - (\rho - K_t)\mu_t - K_t a_t \\
&= \rho\theta_t + \zeta_{t+1} - (\rho - K_t)\mu_t - K_t(\theta_t + \varepsilon_t) \\
&= (\rho - K_t)e_t + \zeta_t - K_t\varepsilon_t.
\end{aligned}$$

Now, we calculate variance of both sides and denote $\Sigma_t \equiv Var(e_t)$ to obtain

$$\Sigma_{t+1} = (\rho - K_t)^2\Sigma_t + \sigma_\zeta^2 + K_t^2\sigma_\varepsilon^2.$$

[34] Jovanovic model has implications for age-dependent volatility of growth rate of output. Specifically, the model predicts that the volatility of growth rate of output decreases with firm age, which is a salient feature of our data. For details, see Appendix 7.4.2.

Given Σ_t, we can use first order condition to derive the optimal Kalman gain as

$$K_t = \frac{\rho \Sigma_t}{\Sigma_t + \sigma_\varepsilon^2}.$$

We discuss the correlation of FEs in the steady state now (i.e., $t \to \infty$). The two equations that pin down the steady-state Kalman gain and variance of FEs are

$$\begin{aligned} K &= \rho\Sigma/(\Sigma + \sigma_\varepsilon^2) \\ \Sigma &= (\rho - K)^2\Sigma + \sigma_\zeta^2 + K^2\sigma_\varepsilon^2. \end{aligned}$$

We can solve these equations analytically:

$$K = \frac{\sqrt{(1+\lambda-\rho^2)^2 + 4\rho^2\lambda} - (1+\lambda-\rho^2)}{2\rho},$$

where $\lambda = \sigma_\zeta^2/\sigma_\varepsilon^2$ is the noise-to-signal ratio.

Now, we prove the key result of this subsection: $cov(e_{t+1}, a_s) = 0$ for any $s \leq t$ in the steady state. Since it is the steady state, we write $K_t = K$. Iterating backwards, one can express μ_{t+1} (forecast of θ_{t+1} with information prior to $t+1$) as

$$\begin{aligned} \mu_{t+1} &= (\rho - K)\mu_t + Ka_t \\ &= K\sum_{j=0}^{\infty}(\rho - K)^j a_{t-j}. \end{aligned}$$

Thus, we have

$$\begin{aligned} e_{t+1} &= \theta_{t+1} - \mu_{t+1} \\ &= \rho\theta_t + \zeta_{t+1} - K\sum_{j=0}^{\infty}(\rho - K)^j a_{t-j}. \end{aligned}$$

Covariance between a_s and e_{t+1} is (for $s \leq t$)

$$Cov(a_s, e_{t+1}) = \rho Cov(\theta_t, a_s) - K\sum_{j=0}^{\infty}(\rho - K)^j Cov(a_{t-j}, a_s).$$

Note that a_s and θ_t can be rewritten as

$$\begin{aligned}\theta_t &= \sum_{j=0}^{\infty}\rho^j\zeta_{t-j}\\ a_s &= \theta_s+\varepsilon_s=\sum_{j=0}^{\infty}\rho^j\zeta_{s-j}+\varepsilon_s.\end{aligned}$$

Therefore, covariance between θ_t and a_s is

$$Cov(\theta_t,a_s)=\rho^{t-s}\sigma_\theta^2,$$

where $\sigma_\theta^2=\sigma_\zeta^2/(1-\rho^2)$ is the steady-state variance of θ.

For the covariance between a_{t-j} and a_s, there are three cases:

$$\begin{aligned}Cov(a_{t-j},a_s) &= Cov(\sum_{m=0}^{\infty}\rho^m\zeta_{s-m}+\varepsilon_s,\sum_{m=0}^{\infty}\rho^m\zeta_{t-j-m}+\varepsilon_{t-j})\\ &= \begin{cases}\rho^{t-j-s}\sigma_\theta^2 & \text{if } t-j>s\\ \sigma_\theta^2+\sigma_\varepsilon^2 & \text{if } t-j=s.\\ \rho^{s-(t-j)}\sigma_\theta^2 & \text{if } t-j<s\end{cases}\end{aligned}$$

Adding up each part, we have

$$\begin{aligned}Cov(a_s,e_{t+1}) &= \rho^{t-s+1}\sigma_\theta^2-K\sum_{j=0}^{t-s}(\rho-K)^j\rho^{t-j-s}\sigma_\theta^2\\ &\quad -K(\rho-K)^{t-s}\sigma_\varepsilon^2-K\sum_{j=t-s+1}^{\infty}(\rho-K)^j\rho^{s-(t-j)}\sigma_\theta^2\\ &= \rho^{t-s+1}\sigma_\theta^2-\rho^{t-s+1}\left(1-\left(\frac{\rho-K}{\rho}\right)^{t-s+1}\right)\sigma_\theta^2\\ &\quad -K(\rho-K)^{t-s}\sigma_\varepsilon^2-\frac{\rho K(\rho-K)^{t-s+1}}{1-\rho(\rho-K)}\sigma_\theta^2\\ &= \frac{(\rho-K)^{t-s+1}}{1-\rho(\rho-K)}\sigma_\zeta^2-K(\rho-K)^{t-s}\sigma_\varepsilon^2\\ &= (\rho-K)^{t-s}\sigma_\varepsilon^2\left(\frac{\lambda(\rho-K)}{1-\rho(\rho-K)}-K\right)\\ &= 0.\end{aligned}$$

Therefore, FE at period $t+1$ is uncorrelated to any variable that has been relied up to period t. As FE for log sales in period $t+1$ is proportional to

$$e_{t+1} + \varepsilon_{i,t} = \frac{\theta_{t+1} + \varepsilon_{i,t+1} - \mu_{t+1}}{\sigma},$$

the correlation between log sales FEs and any variable that has been relied up to period t is also zero, as the i.i.d. transitory shock $\varepsilon_{i,t+1}$ is also independent of any variable that has been relied up to period t.

7.4.2 Volatility of Output Growth

In this subsection of the appendix, we show that the Jovanovic model predicts that the volatility of growth rate of output decreases with firm age, which is a salient feature of our data. The learning model provides a rationale for this, even in the case in which the volatility of productivity or demand shocks does not vary with age.

Proposition 6 *Growth volatility of output decreases with firm age. Growth volatility of sales decreases with firm age at least when the firm is young.*

Proof. Based on equation (25), we know that log firm output at age t is proportional to (after omitting constants and aggregate variables which are constant in the steady state)

$$\sigma \log(b(\bar{a}, t-1)), \tag{28}$$

and it is proportional to

$$\sigma \log(b(\bar{a}, t)), \tag{29}$$

at age $t+1$. Therefore, the growth rate of output or quantity (or the log difference) is proportional to[35]

$$\Big(\frac{\lambda(\theta - \bar{\theta}) + (1+\lambda(t-1))\lambda\varepsilon_t - \lambda^2\Sigma_{i=1}^{t-1}\varepsilon_i}{(1+\lambda t)(1+\lambda(t-1))}\Big).$$

The variance of the growth rate at age t ($t \geq 2$) is

$$var(growth\ rate)_t = \Big[\frac{\lambda^2\sigma_\theta^2 + \lambda^4\sigma_\varepsilon^2(t-1)}{(1+\lambda t)^2(1+\lambda(t-1))^2} + \Big(\frac{\lambda}{(1+\lambda t)}\Big)^2\sigma_\varepsilon^2\Big],$$

[35]Here, we omit the term related to $\frac{1}{2}\left(\frac{\sigma_{t-1}^2+\sigma_\varepsilon^2}{\sigma}\right)$ and $\frac{1}{2}\left(\frac{\sigma_t^2+\sigma_\varepsilon^2}{\sigma}\right)$, as they do not vary across firms.

which can be reduced to

$$var(growth\ rate)_t = \Big(\frac{\lambda^2\sigma_\varepsilon^2}{(1+\lambda t)(1+\lambda(t-1))}\Big).$$

Thus, the growth volatility of output decreases with firm age.

Next, we explore how the growth volatility of sales varies with firm age. Based on equation (25), we know that log firm sales at age t are proportional to (after omitting constants and aggregate variables which are constant in the steady state)

$$(\sigma-1)\log(b(\bar{a},t-1)) + \frac{a_t}{\sigma}, \tag{30}$$

and it is proportional to

$$(\sigma-1)\log(b(\bar{a},t)) + \frac{a_{t+1}}{\sigma} \tag{31}$$

at age $t+1$. Therefore, the growth rate of sales (or the log difference) is proportional to

$$\Big(\frac{\sigma-1}{\sigma}\Big)\Big(\frac{\lambda(\theta-\bar{\theta})+(1+\lambda(t-1))\lambda\varepsilon_t-\lambda^2\Sigma_{i=1}^{t-1}\varepsilon_i}{(1+\lambda t)(1+\lambda(t-1))}\Big)+\frac{\varepsilon_{t+1}-\varepsilon_t}{\sigma}.$$

Note that there is an additional term of $\frac{\varepsilon_{t+1}-\varepsilon_t}{\sigma}$ due to the shock to the price. The variance of the growth rate at age t ($t \geq 2$) is

$$\begin{aligned} var(growth\ rate)_t \quad &= \quad \Big(\frac{\sigma-1}{\sigma}\Big)^2\Big[\frac{\lambda^2\sigma_\theta^2+\lambda^4\sigma_\varepsilon^2(t-1)}{(1+\lambda t)^2(1+\lambda(t-1))^2}+\Big(\frac{\lambda}{(1+\lambda t)}-\frac{1}{\sigma-1}\Big)^2\sigma_\varepsilon^2\Big] \\ &\quad +\frac{\sigma_\varepsilon^2}{\sigma^2}, \end{aligned}$$

which can be further reduced to

$$\Big(\frac{\sigma-1}{\sigma}\Big)^2\Big[\frac{\lambda^3\sigma_\varepsilon^2}{(1+\lambda t)^2(1+\lambda(t-1))}+\Big(\frac{\lambda}{(1+\lambda t)}-\frac{1}{\sigma-1}\Big)^2\sigma_\varepsilon^2\Big]+\frac{\sigma_\varepsilon^2}{\sigma^2}.$$

In general, we do not have a result concerning how the growth volatility of sales evolves with firm age, as the term of $\Big(\frac{\lambda}{(1+\lambda t)}-\frac{1}{\sigma-1}\Big)^2$ can either increase or decrease with firm age. However, when λ is not too small and firm is young. the growth volatility of sales also goes down with firm age. However, the term of $\Big(\frac{\lambda}{(1+\lambda t)}-\frac{1}{\sigma-1}\Big)^2$ increases with firm age when the firm is old enough. ■

Several things are worth mentioning. First, the declining profile of the volatility of output (and sales) growth ceases to exist when $\lambda = 0$ (i.e., the fundamental uncertainty disappears).

This verifies that if there is no need to learn which is true when the fundamental uncertainty goes to zero, there is no age-dependent volatility of output growth. Therefor, it is exactly the learning mechanism that triggers the declining volatility of firm growth. Second, as the cross derivative of $\frac{\lambda^2\sigma_\varepsilon^2}{(1+\lambda t)(1+\lambda(t-1))}$ with respect to λ and t is negative, the reduction in the level of growth volatility is larger when the signal-to-noise ratio is higher. Economically speaking, a bigger λ implies faster learning and a quicker convergence to the volatility of sales growth at the steady state level.

Now, the question is why the growth volatility declines with firm age. A quick glance at equations (28) and (29) shows that two components contribute to the growth volatility of output. The first term is related to the change in the average experience (i.e., $\log(b(\bar{a},t)) - \log(b(\bar{a},t-1))$ which goes down with firm age. Intuitively, firm's belief and average experience fluctuate substantially when it is young, as there is not too much experience there. However, when the firm is old enough, its average experience stops fluctuating significantly. The second term is the contemporaneous productivity shock which is age independent. In short, the declining volatility of firm belief with age triggers the declining volatility of firm's output growth. And, the learning mechanism generates the declining volatility of output growth endogenously.

The final question is why the growth volatility of *sales* does not always go down with firm age. The key is that the shock to the price in the period t (i.e., ε_t) is correlated with the belief formed at the end of period t, which is used to forecast sales in period $t+1$. And, the volatility of this term (i.e., ε_t) does not always decrease with firm age. This makes deriving an unambiguous result on how age affects the growth volatility of sales impossible.

7.5 Proof for Proposition 1

For the first part of the proposition, the proof is as follows: **Proof.** The variance of FEs at age t now becomes

$$\begin{aligned}
& var(FE_{t-1}^{\log})_{sticky} \\
= \; & \frac{\sigma_\theta^2 + \sigma_\varepsilon^2 + (1-\alpha^{t-1})\zeta(t-1,\lambda)^2\Big(\sigma_\theta^2 + \frac{\sigma_\varepsilon^2}{t-1}\Big) - 2(1-\alpha^{t-1})\zeta(t-1,\lambda)\sigma_\theta^2}{\sigma^2} \\
= \; & \frac{[1-(2\zeta(t-1,\lambda)-\zeta(t-1,\lambda)^2)(1-\alpha^{t-1})]\sigma_\theta^2 + \Big(1 + \frac{(1-\alpha^{t-1})\zeta(t-1,\lambda)^2}{t-1}\Big)\sigma_\varepsilon^2}{\sigma^2} \\
= \; & \frac{\sigma_\theta^2+\sigma_\varepsilon^2}{\sigma^2} - (1-\alpha^{t-1})\Big(\frac{\sigma_\theta^2+\sigma_\varepsilon^2}{\sigma^2} - var(FE_{t-1}^{\log})\Big),
\end{aligned}$$

where $var(FE^{\log}_{t-1})$ is the variance of FEs at age t in the original Jovanovic model. Note that $var(FE^{\log}_{t-1})$ (i.e., variance of FEs in Jovanovic model) is always smaller than the variance of FEs made by firms using the prior belief (i.e., $\frac{\sigma_\theta^2+\sigma_\varepsilon^2}{\sigma^2}$). Moreover, $var(FE^{\log}_{t-1})$ and $(1-\alpha^{t-1})$ decreases and increases with firm age respectively. Therefore, we have that

$$\frac{\sigma_\theta^2+\sigma_\varepsilon^2}{\sigma^2} - (1-\alpha^{t-1})\Big(\frac{\sigma_\theta^2+\sigma_\varepsilon^2}{\sigma^2} - var(FE^{\log}_{t-1}))\Big),$$

decreases with firm age t. In total, $var(FE^{\log}_{t-1})_{sticky}$ decreases with firm age and previous export experience. ■

For the second part of the proposition, the proof is as follows: **Proof.** We shows that past FEs are also positively correlated with current forecasts in our full model. We can decompose the proof into three parts. First, there are α^t fraction of firms that are still uninformed by the end of period t which use $\bar{\theta}$ (and plus some constant terms) to predict future sales in both period $t-1$ and period t. For those firms, past FEs are not correlated with current forecasts which are a constant. Second, there are $1-\alpha^{t-1}$ fraction of firms that are informed by the end of period $t-1$. They are behave in the same way as in the Jovanovic model. FEs are positively correlated with current forecasts. The third type of firms is the firm that switches from being uninformed to being informed with the fraction of $\alpha^{t-1}(1-\alpha)$ among all firms. For those firms, we have

$$FE^{\log}_{t-1,t} \sim \frac{\theta+\varepsilon_t-\bar{\theta}}{\sigma}; \qquad Forecast^{log}_{t,t+1} \sim \frac{\bar{\theta}+\lambda(\theta+\varepsilon_t)}{1+\lambda},$$

where $\sim$ means "proportional to" and λ is the signal-to-noise ratio. Thus, we have

$$cov(FE^{\log}_{t-1,t}, Forecast^{log}_{t,t+1}) = \frac{\lambda(\sigma_\varepsilon^2+\sigma_\theta^2)}{\sigma(1+\lambda)} > 0.$$

It is also straightforward to observe that the (positive) covariance goes down when the switching firm becomes older after the switching.

We know that we have three groups of firms: uninformed firms, informed firms and switching firms. The overall covariance follows:

$$cov(X,Y) = E(cov(X,Y|Z)) + cov(E(X|Z), E(Y|Z)),$$

where $X \equiv FE^{\log}_{t-1,t}$, $Y \equiv Forecast^{log}_{t,t+1}$, and Z denotes one of the three groups of firms. The selection into the three types of firms is random in the model, which implies zero correlation between average FEs last period and average forecasts this period among each group of firms.[36] As a result, the second term above is zero. Therefore, the result that $cov(X, Y|Z) > 0$ for two groups of firms (and zero for the remaining group) implies that the overall covariance is positive. ■

For the third part of the proposition, the proof is as follows: **Proof.** We prove the positive serial correlation of FEs first. The correlation of FEs in two consecutive periods (i.e., period t and period $t+1$) can be decomposed into three parts. First, there are α^t fraction of firms that are still uninformed by the end of period t. Second, there are $1-\alpha^{t-1}$ fraction of firms that are informed by the end of period $t-1$. The third type of firms is the firm that switches from being uninformed to being informed with the fraction of $\alpha^{t-1}(1-\alpha)$ among all firms. As in the original Jovanovic model, the serial correlation (and covariance) of FEs is zero for $1-\alpha^{t-1}$ fraction of firms that are informed by the end of period $t-1$. For firms that are still uninformed by the end of period t, the correlations is

$$cov(FE_{t-1}^{\log}, FE_t^{\log}) = cov\Big(\frac{\varepsilon_t + \theta - \bar{\theta}}{\sigma}, \frac{\varepsilon_{t+1} + \theta - \bar{\theta}}{\sigma}\Big) = \frac{\sigma_\theta^2}{\sigma^2},$$

as they use the prior mean, $\bar{\theta}$, to forecast the fundamental demand in both periods. For the switchers, the correlation is

$$\begin{aligned} & cov(FE_{t-1}^{\log}, FE_t^{\log}) \\ = \;& cov\Big(\frac{\varepsilon_t + \theta - \bar{\theta}}{\sigma}, \frac{\varepsilon_{t+1} + \theta - \mu_t}{\sigma}\Big) \\ = \;& \frac{1}{\sigma^2} cov\Big(\varepsilon_t + \theta - \bar{\theta}, \theta - \mu_t\Big) \\ = \;& \frac{1}{\sigma^2} cov\Big(\varepsilon_t + \theta - \bar{\theta}, \frac{\theta - \bar{\theta}}{1+t\lambda} - \frac{\lambda \sum_{t=1}^t \varepsilon_t}{1+t\lambda}\Big) \\ = \;& \frac{1}{\sigma^2}\Big(\frac{\sigma_\theta^2}{1+t\lambda} - \frac{\lambda\sigma_\varepsilon^2}{1+t\lambda}\Big) = 0. \end{aligned}$$

Therefore, the average covariance of $FE_{t-1,t}^{\log}$ and $FE_{t,t+1}^{\log}$ is

$$\frac{\alpha^t \sigma_\theta^2}{\sigma^2}.$$

[36] In fact, $E(X|Z)$ should be zero among each group of firms.

The correlation coefficient is

$$corr.(FE_{t-1}^{\log}, FE_t^{\log}) = \frac{cov(FE_{t-1}^{\log}, FE_t^{\log})}{\sqrt{var(FE_{t-1}^{\log})}\sqrt{var(FE_t^{\log})}} = \frac{\frac{\alpha^t \sigma_\theta^2}{\sigma^2}}{\frac{\sigma_\varepsilon^2+\sigma_\theta^2}{\sigma^2}},$$

as

$$var(FE_{t-1}^{\log}) = var(FE_t^{\log}) = \frac{\sigma_\varepsilon^2 + \sigma_\theta^2}{\sigma^2}$$

for the uninformed firms. Therefore, both the covariance and the correlation coefficient are strictly positive and decrease when the firm becomes older or has previous export experience.

We know that we have three groups of firms: uninformed firms, informed firms and switching firms. The overall correlation coefficient becomes:

$$cov(X, Y) = E(cov(X, Y|Z)) + cov(E(X|Z), E(Y|Z)),$$

where $X \equiv FE_{t-1,t}^{\log}$, $Y \equiv FE_t^{\log}$, and Z denotes one of the three groups of firms. The selection into the three types of firms is random in the model, which implies zero correlation between average FEs last period and average FEs this period among each group of firms.[37] As a result, the second term above is zero. Therefore, the result that $cov(X, Y|Z) > 0$ for one group of firms (and zero for the remaining two groups of firms) implies that the overall correlation is positive.

Another type of regression we can run is to take the first-order difference of FEs (or include the affiliate fixed effects), when we run the regression specified in equation (2). In these two cases, our model (with both Jovanovic-type learning and sticky information) predicts a negative coefficient, which is consistent with the regression results. Detailed proofs and regression results are available upon request. ■

For the last part of the proposition, the proof is as follows: **Proof.** we use the property of conditional expectation to prove that when we run the following regression using the sample of informed firms:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \tag{32}$$

where $y = sales_t^{\log}$, $x_1 = forecast_{t-1,t}^{\log}$ and $x_2 = sales_{t-1}^{\log}$. the estimated coefficients are $\beta_1 = 1$ and $\beta_2 = 0$.

According to Hayashi (n.d.) (Proposition 2.8) we know that $\beta\mathbf{x}$ is the "best linear predictor" of y as it minimizes the mean square error (among all linear predictors). For the informed firms, we know that they use information on past sales to form BLUE for the fundamental demand

[37] In fact, both $E(X|Z)$ and $E(Y|Z)$ should be zero among each group of firms.

draw using Bayesian updating. This implies that

$$E(\theta|a_{t-1},\dots,a_1)=\mu_{t-1}$$

Thus, we have

$$\begin{aligned}E(sales_t^{\log}|a_{t-1},\dots,a_1) &= E(\frac{\sigma-1}{\sigma}\mu_{t-1}+\theta+\varepsilon_t|a_{t-1},\dots,a_1)+con.\\ &= \frac{\sigma-1}{\sigma}\mu_{t-1}+cons=forecast_{t-1,t}^{\log}+con,\end{aligned}$$

where *con* is a constant term which does not vary across firms (e.g., domestic and foreign wage rates in the steady state etc.). Therefore, $forecast_{t-1,t}^{\log}$ is the conditional expectation. According to Hayashi (n.d.) (Proposition 2.7), conditional expectation is the "best predictor" (i.e., the one that minimizes the mean squared error) among all estimators. Since it takes linear form, it must also be the "best linear predictor" and coincide with the linear regression coefficients $\beta=E(\mathbf{x}\mathbf{x}')^{-1}E(\mathbf{x}y)$. Therefore, $\beta_1=1$ and $\beta_2=0$ are the estimated coefficients.

For the uninformed firms, they use (fixed) prior to forecast future sales. In equation (32), we have $x_1=forecast_{t-1,t}^{\log}=\bar{\theta}+con$ and $x_2=sales_{t-1}^{\log}$. The first order condition of the estimation equation is

$$E[y-(\beta_0+\beta_1x_1+\beta_2x_2)]=0,$$

$$E[y-(\beta_0+\beta_1x_1+\beta_2x_2)]x_1=0,$$

and

$$E[(y-(\beta_0+\beta_1x_1+\beta_2x_2))x_2]=0.$$

If $\beta_0=0$, $\beta_1=\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\theta^2}$ and $\beta_2=\frac{\sigma_\theta^2}{\sigma_\varepsilon^2+\sigma_\theta^2}$, we have

$$E[y-(\beta_0+\beta_1x_1+\beta_2x_2)]=\frac{1}{\sigma}\Big[(\sigma-1)\bar{\theta}+E\theta-\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\theta^2}\sigma\bar{\theta}-\frac{\sigma_\theta^2}{\sigma_\varepsilon^2+\sigma_\theta^2}[(\sigma-1)\bar{\theta}+E\theta]\Big]=0.$$

As x_1 is a constant, we also have

$$E[(y-(\beta_0+\beta_1x_1+\beta_2x_2))x_1]=x_1E[y-(\beta_0+\beta_1x_1+\beta_2x_2)]=0.$$

Furthermore, we have

$$Cov[y-(\beta_0+\beta_1x_1+\beta_2x_2),x_2]=\frac{1}{\sigma^2}Cov\Big(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\theta^2}\theta+\varepsilon_t-\frac{\sigma_\theta^2}{\sigma_\varepsilon^2+\sigma_\theta^2}\varepsilon_{t-1},\theta+\varepsilon_{t-1}\Big)=0.$$

As

$$E[y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2)] = 0,$$

we must have

$$E[(y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2))x_2] = 0.$$

Therefore, $\beta_0 = 0$, $\beta_1 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\theta^2}$ and $\beta_2 = \frac{\sigma_\theta^2}{\sigma_\varepsilon^2+\sigma_\theta^2}$ are indeed the estimated coefficients of the OLS regression for the uninformed firms. ■

7.6 Proof for Proposition 2

Proof. We want to show that the loss of welfare (compared to the social planner's solution or the constrained optimum) is due to the difference in allocation at the extensive margin only. We shut down the existence of sleeping firms in this section. for simplicity. As a result, the number of state variables become four:

$$(mode, f_e^I, \bar{a}_t, t).$$

Since our model is not a general equilibrium model, a proper measure for welfare is labor productivity. Specifically, the social planner wants to maximize

$$\max_{I_x(f_e^I,\bar{a},t), I_I(f_e^I,\bar{a},t), q_x(\bar{a},t), q_I(\bar{a},t)} \ln(C^*) - \ln(L), \tag{33}$$

where

$$\begin{aligned} C^* &= \Bigg(\Sigma_{t=1}^{\infty} \int_{f_e^I} \int_{\bar{a}} I_x(f_e^I, \bar{a}, t) q_x(\bar{a}, t)^{\frac{\sigma-1}{\sigma}} b(\bar{a}, t) m(x, f_e^I, \bar{a}, t) \\ &\quad + \Sigma_{t=1}^{\infty} \int_{f_e^I} \int_{\bar{a}} I_I(f_e^I, \bar{a}, t) q_I(\bar{a}, t)^{\frac{\sigma-1}{\sigma}} b(\bar{a}, t) m(I, f_e^I, \bar{a}, t) \Bigg)^{\frac{\sigma}{\sigma-1}}, \end{aligned} \tag{34}$$

where $m(., f_e^I, \bar{a}, t)$ is the steady-state measure of firms in that cell and

$$\begin{aligned} L &= \Sigma_{t=1}^{\infty} \int_{f_e^I} \int_{\bar{a}} \Big(\tau q_x(\bar{a}, t) + f \Big) I_x(f_e^I, \bar{a}, t) m(x, f_e^I, \bar{a}, t) \\ &\quad + \Sigma_{t=1}^{\infty} \int_{f_e^I} \int_{\bar{a}} q_I(\bar{a}, t) I_I(f_e^I, \bar{a}, t) m(I, f_e^I, \bar{a}, t) \\ &\quad + \Sigma_{t=1}^{\infty} \int_{f_e^I} \int_{\bar{a}} f_e^I [I_I(f_e^I, \bar{a}, t) * (1 - I_I(f_e^I, \bar{a}, t-1))] m(I, f_e^I, \bar{a}, t), \end{aligned} \tag{35}$$

where $I_I(f_e^I, \bar{a}, 0) = 0$. Different from Arkolakis et al. (2017) is that we have elastic supply of labor L now. If the social planner wants to make more (young) exporting firms survive, it has to hire more labor to pay for the fixed cost. As a result, both real consumption and labor inputs increase.

We derive F.O.C.s for output choices:

$$\Big(C^{*\frac{1-\sigma}{\sigma}} q_x(\bar{a},t)^{-\frac{1}{\sigma}} b(\bar{a},t) - \frac{\tau}{L}\Big) I_x(f_e^I, \bar{a}, t) m(x, f_e^I, \bar{a}, t) = 0,$$

where $b(\bar{a}, 0)$ is the prior belief (of entrants). The above F.O.C. leads to the optimal output of exports as

$$q_x(\bar{a},t) = \Big(\frac{b(\bar{a},t)L}{\tau C^{*\frac{\sigma-1}{\sigma}}}\Big)^{\sigma}. \tag{36}$$

Similarly, the optimal MP output is

$$q_I(\bar{a},t) = \Big(\frac{b(\bar{a},t)L}{C^{*\frac{\sigma-1}{\sigma}}}\Big)^{\sigma}. \tag{37}$$

We now solve for the endogenous composite consumption C^*. Substituting equations (36) and (37) into the composite consumption yields

$$\Big(\frac{C^*}{L}\Big)^{\sigma} = \Bigg(\Sigma_{t=1}^{\infty} \int_{f_e^I} \int_{\bar{a}} \tau^{1-\sigma} b(\bar{a},t)^{\sigma} I_x(f_e^I, \bar{a}, t) m(x, f_e^I, \bar{a}, t) + \Sigma_{t=1}^{\infty} \int_{f_e^I} \int_{\bar{a}} b(\bar{a},t)^{\sigma} I_I(f_e^I, \bar{a}, t) m(I, f_e^I, \bar{a}, t)\Bigg)^{\frac{\sigma}{\sigma-1}},$$

where the right hand side is simply proportional to the ideal price index (I assume that the ideal price index is the same using the subjective or objective belief for θ which is probably true). Thus, we have

$$\Big(\frac{C^*}{L}\Big)^{\sigma} = \Big(\frac{\sigma-1}{\sigma} P\Big)^{-\sigma}.$$

Therefore, the output choices can be simplified to

$$q_x(\bar{a},t) = \frac{1}{\tau^{\sigma}} \Big(\frac{b(\bar{a},t) C^* P\big(\frac{\sigma}{\sigma-1}\big)}{C^{*\frac{\sigma-1}{\sigma}}}\Big)^{\sigma} = \big(\frac{\sigma}{\sigma-1}\big)^{\sigma} \frac{b(\bar{a},t)^{\sigma}}{\tau^{\sigma}} \frac{Y}{P^{1-\sigma}} \tag{38}$$

and

$$q_I(\bar{a},t) = \Big(\frac{b(\bar{a},t) C^* P\big(\frac{\sigma}{\sigma-1}\big)}{C^{*\frac{\sigma-1}{\sigma}}}\Big)^{\sigma} = \big(\frac{\sigma}{\sigma-1}\big)^{\sigma} b(\bar{a},t)^{\sigma} \frac{Y}{P^{1-\sigma}}, \tag{39}$$

where $Y \equiv C^* P$ is the total expenditure on Japanese goods. The social would choose exactly the same levels of output as what the decentralized economy achieves. In other words, there is no distortion at the intensive margin in the learning model with varying labor supply. ■

8 Empirical Appendix

8.1 Robustness Checks for Learning from Exporting

In this subsection, we provide several robustness checks for the baseline regression described in Subsection 3.5. Specifically, Table 20 presents the same pattern as in Table 6, when we restrict our sample to first-time entrants into regions instead of countries. The effect of export experience is larger but at the same time more noisy due to the reduced size of our sample. To be conservative, we prefer to use estimates from the sample of first-time entrants into countries in our quantitative exercises.

Table 20: Forecast error and previous exporting (first entrants into continents)

Dep.Var: $\lvert FE_{1,2}\rvert$	(1)	(2)	(3)	(4)
$Exp_{-1} > 0$	-0.303*			
	(0.168)			
$Exp_{-1} > 0$ or $Exp_{-2} > 0$		-0.175		
		(0.173)		
$Exp \quad Expe. > 0$			-0.226	
			(0.169)	
$Exp \quad Expe.$				-0.034*
				(0.020)
Industry FE	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes
N	153	152	185	185
R^2	0.601	0.589	0.592	0.607

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is affiliates' initial forecast error, which is calculated as the absolute log deviation of the realized sales at age = 2 from the projected sales (predicted by an affiliate at age = 1). We only include affiliates that are first-time entrants into a particular continent. Exporting experience (Exp Expe.) is defined at the continent level for each parent firm.

The relationship between previous export experience and FEs at age one is robust to controlling for firm size. As we discussed in the previous section, bigger firms may have smaller subjective uncertainty. In addition, Table 26 in the appendix shows that experience MNEs tend to set up bigger and more productive (in terms of sales per worker) affiliates upon FDI entry relative to inexperienced MNEs. Therefore, firm size and productivity are also correlated with previous export experience of first-time entrants. In order to address these concerns, we control for parent firm's employment (or sales) and employment (or sales) of its affiliate in Table 21. Previous export experience still has a significantly negative impact on the initial subjective uncertainty, and the magnitude of the effect does not vary much. Consistent with the evidence in Table 4, parent firm size is not strongly correlated with affiliate subjective uncertainty while affiliate size is negatively associated with its subjective uncertainty.

Our final robustness checks are related to the type of FDI and exports measured in our data.

Table 21: Forecast error and previous exporting - control firm size

	(1)	(2)	(3)	(4)	(5)	(6)
$Exp_{-1} > 0$	-0.151**	-0.115*				
	(0.063)	(0.062)				
$Exp_{-1} > 0$ or $Exp_{-2} > 0$			-0.147**	-0.121*		
			(0.063)	(0.064)		
$Exp \quad Expe. > 0$					-0.113*	-0.077
					(0.065)	(0.063)
log(Parent Employment)	0.017		0.021		0.009	
	(0.023)		(0.022)		(0.021)	
log(Affiliate Employment)	-0.031		-0.020		-0.045**	
	(0.020)		(0.018)		(0.018)	
log(Parent Domestic Sales)		0.018		0.021		0.018
		(0.017)		(0.016)		(0.016)
log(Affiliate Sales)		-0.054***		-0.052***		-0.058***
		(0.014)		(0.013)		(0.014)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	549	534	557	543	654	625
R^2	0.493	0.535	0.503	0.541	0.485	0.532

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is the absolute log deviation of the realized sales at age = 2 from the projected sales (predicted by an affiliate at age = 1). We only include affiliates that are first-time entrants into a particular continent. Exporting experience (Exp Expe.) is defined at the continent level for each parent firm.

Learning about uncertain foreign demand through exporting is more relevant for horizontal than vertical FDI. In columns 1-3 of Table 22, we try to exclude possible affiliates of firms doing vertical FDI by restricting our sample to affiliates that never export more than 1/3 of their sales back to Japan. This does not affect the estimated effect of previous export experience. In columns 4-6, we refine our measure of parent firms' export experience. Specifically, we redefine export experience to be zero, if all of the parent firm's exports to a certain region are intra-firm trade. The estimated effects are less significant than other specifications, but the magnitude remains stable.

Finally, in Table 23, we reproduce Table 6 using the residual forecast error as the dependent variable. The results are very similar. We can also replicate the results of the other specifications mentioned above using this alternative measure of FE. These results are available upon request.

8.2 Robustness Checks for Correlation of Forecast Errors

We run the regression in equation (3) and control for parent firm fixed effects. The empirical result on the serial correlation of FEs is unchanged, as shown by the following table:

We run the regression in equation (3) by utilizing observations in the manufacturing (and wholesale and retail) sector. The empirical result of the declining (positive) serial correlation of FEs with age is unchanged, as shown by the following table:

Table 22: Forecast error and previous exporting - exclude vertical FDI and affiliated export

	Exclude vertical FDI			Exclude affiliated export		
Dep.Var: $\lvert FE_{1,2}\rvert$	(1)	(2)	(3)	(4)	(5)	(6)
$Exp_{-1} > 0$	-0.166**			-0.099		
	(0.073)			(0.067)		
$Exp_{-1} > 0$ or $Exp_{-2} > 0$		-0.155**			-0.141**	
		(0.072)			(0.067)	
$Exp \quad Expe. > 0$			-0.159**			-0.114
			(0.078)			(0.071)
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	456	464	551	441	446	551
R^2	0.542	0.549	0.529	0.545	0.554	0.524

a Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is the absolute log deviation of the realized sales at age = 2 from the projected sales (predicted by an affiliate at age = 1). We only include affiliates that are first-time entrants into a particular continent. Exporting experience (Exp Expe.) is defined at the continent level for each parent firm.

b In columns 1-3, we exclude affiliates whose sales share back to Japan is larger than one third in at least one year. In columns 4-6, in addition to excluding vertical FDI, we further refine our measure of exporting experience by excluding intra-firm exports from parent firm to affiliates in a particular continent.

Table 23: Forecast error and previous exporting

Dep.Var: $\lvert\hat{\epsilon}_{FE,(1,2)}\rvert$	(1)	(2)	(3)	(4)
$Exp_{-1} > 0$	-0.139**			
	(0.066)			
$Exp_{-1} > 0$ or $Exp_{-2} > 0$		-0.142**		
		(0.064)		
$Exp \quad Expe. > 0$			-0.121*	
			(0.070)	
$Exp \quad Expe.$				-0.013**
				(0.006)
Industry FE	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes
N	552	560	657	657
R^2	0.462	0.475	0.446	0.447

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is affiliates' initial residual forecast error. We only include affiliates that are first-time entrants into a particular host country. Exporting experience (Exp Expe.) is defined at the continent level for each parent firm. Each column head indicates the different measure of exporting experience used in the regression.

Table 24: Regression for the serial correlation of sales forecast errors (including parent firm fixed effects)

	(1) $FE^{pct}_{t,t+1}$	(2) $FE^{\log}_{t,t+1}$	(3) $\hat{\epsilon}_{FE^{log}_{t,t+1}}$	(4) $FE^{pct}_{t,t+1}$	(5) $FE^{\log}_{t,t+1}$	(6) $\hat{\epsilon}_{FE^{log}_{t,t+1}}$
$FE^{pct}_{t-1,t}$	0.0656***			0.0703***		
	(0.00600)			(0.00757)		
$FE^{\log}_{t-1,t}$		0.0642***			0.0631***	
		(0.00526)			(0.00665)	
$\hat{\epsilon}_{FE^{log}_{t-1,t}}$			0.0641***			0.0629***
			(0.00526)			(0.00665)
Type of firms	all	all	all	manufacturing	manufacturing	manufacturing
Parent firm FE Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Industry-year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Country-year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
N	112766	109775	109765	74353	72792	72789
R^2	0.170	0.191	0.088	0.186	0.209	0.095

FE^{log} is the log deviation of the realized sales from the projected sales, while FE^{pct} is the percentage deviation of the realized sales from the projected sales. The last variable, $\hat{\epsilon}_{FE^{log}}$, is the residual forecast error, which we obtain by regressing FE^{log} on a set of industry-year and country-year fixed effects. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Top and bottom one percent observations of forecast errors are trimmed. Standard errors are in parentheses and clustered at the affiliate level. Manufacturing affiliates including those in retail and wholesale sectors as well.

8.3 Size Differences between Experienced and Inexperienced Affiliates

In Table 26, we compare the size/productivity between experienced and inexperienced affiliates. We find that experienced affiliates are larger in terms of sales upon entry. They also have higher labor productivities. This empirical fact motivates us to assume firms have ex-ante heterogeneity in MP entry costs, instead of heterogeneity in productivity. If there is heterogeneity in (labor) productivity, more productive firms are more likely to enter MP without export experience, which is inconsistent with the fact. When firms have heterogenous MP entry costs which are not correlate with their fundamental demand draws, firms with lower entry costs (but average fundamental demand draws) enter MP without export experience. Experienced multinational affiliates are larger upon entry as only exporters with better (than average) fundamental demand draws switch from exporting to MP.

8.4 Export Experience, Age and Affiliates' Exit Rates

In this section, we show that previous export experience and the affiliate's age do not affect the exit rates of the affiliates. In Columns 1 and 2 of Table 27, we regress the exit dummy on previous exporting experience for affiliates (first-time entrants) at age one. Export experience does not have a significant impact on the exit rate, and this pattern is robust to alternative

Table 25: Age effects on the correlation of forecast errors (manufacturing)

Dep.Var: $\mathbb{1}(Sign(FE^{\log}_{i,t}) = Sign(FE^{\log}_{i,t-1}))$	(1)	(2)	(3)	(4)
Age=3	-0.0596	-0.0556		
	(0.0366)	(0.0374)		
Age=4	-0.0400	-0.0424		
	(0.0373)	(0.0392)		
Age=5	-0.0672**	-0.0741**		
	(0.0377)	(0.0398)		
Age=6	-0.0841***	-0.0897***		
	(0.0385)	(0.0404)		
Age=7	-0.0654	-0.0559		
	(0.0401)	(0.0413)		
Age=8	-0.114***	-0.105***		
	(0.0396)	(0.0417)		
Age=9	-0.0928***	-0.0837***		
	(0.0396)	(0.0416)		
log(Affiliate Age)			-0.0607**	-0.0461
			(0.0323)	(0.0342)
log(Affiliate Sales)		-0.0104		-0.0112
		(0.00959)		(0.00962)
log(Parent Domestic Sales)		0.0205**		0.0206**
		(0.0113)		(0.0113)
N	77493	72412	77493	72412
R^2	0.193	0.199	0.193	0.198
Subsidiary FE	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable equals 1 if forecast errors made in two consecutive years have the same sign and -1 otherwise. Forecast error is calculated as the log deviation of the realized sales from the projected sales. Note that as the dependent variable can only be defined for affiliates which are at least two years old, the age dummies can be estimated from 3 (all relative to age 2 firms).

Table 26: Exporting Experience and Firm Size/Productivity

Dependent Var:	log(Affiliate Sales)		log(Sales/Emp)	
	(1)	(2)	(3)	(4)
$Exp_{-1} > 0$	0.297		0.413*	
	(0.284)		(0.223)	
$Exp_{-1} > 0$ or $Exp_{-2} > 0$		0.600**		0.592**
		(0.279)		(0.242)
Industry FE	Yes	Yes	Yes	Yes
Country-year FE	Yes	Yes	Yes	Yes
N	811	808	778	778
R^2	0.572	0.577	0.648	0.652

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is affiliates' sales or labor productivity at age 1. We only include affiliates that are first-time entrants into a particular host country. Exporting experience is defined at the continent level for each parent firm.

definitions of export experience.

In Column 3 and 4, we look at first-time entrants at all ages. In addition to export experience, we add interaction terms between export experience and affiliate age. Neither the export experience dummy nor the interaction terms are significant. This suggests that the exit rates do not decline with age.

Table 27: Export Experience, Age and Affiliate Exits

Dep.Var: Exit Dummy	Age = 1		All Ages	
	(1)	(2)	(3)	(4)
$Exp_{-1} > 0$	0.006		-0.007	
	(0.005)		(0.005)	
$Exp_{-1} > 0 \,\vert\, Exp_{-2} > 0$		0.006		-0.005
		(0.005)		(0.006)
Age $\times (Exp_{-1} = 0)$			0.000	
			(0.001)	
Age $\times (Exp_{-1} > 0)$			-0.000	
			(0.000)	
Age $\times (Exp_{-1} = 0 \,\&\, Exp_{-2} = 0)$				0.000
				(0.001)
Age $\times (Exp_{-1} > 0 \,\vert\, Exp_{-2} > 0)$				-0.000
				(0.000)
N	1285	1288	19249	18821
R^2	0.226	0.228	0.047	0.047
Industry FE	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes

Standard errors are clustered at parent firm level, * 0.10 ** 0.05 *** 0.01. Dependent variable is a dummy variable, which equals one when the affiliate exits next year. We only include affiliates that are first-time entrants into a particular host country. Exporting experience is defined at the continent level for each parent firm. The first two columns include only affiliates at age 1. The last two columns include affiliates of all ages.

The above empirical results stand in contrast to the predictions of models of firm learning and endogenous exits. With endogenous exits, experienced firms and older firms are less likely to exit after entry into MP. To reconcile these patterns, we set the per-period fixed MP cost to zero. Therefore, affiliates only exit the market when they receive the exogenous death shock. This also implies that export experience will not affect the exit rate of affiliates at age one.

8.5 Aggregate Uncertainty and Firm-level Uncertainty

In this subsection, we present evidence to show that aggregate uncertainty and firm-level subjective uncertainty are positively correlated in our data. First, similar to our analysis in Section 3.3, we regress absolute value of FEs on BMI risk index (at the country level) and a set of control variables using observations of Japanese affiliates who are at least eight years old.[38] As Table

[38]The variance of FEs made by old enough affiliates mainly reflects the volatility of transitory (demand and supply) shocks.

28 shows, country-level risks such as the probability of an economic crisis and the stability of government policies positively affect firm-level volatility. This result indicates that macro stabilization policies and rule-based policies (instead of discretionary policies) at the aggregate level probably reduce the volatility of firm-level demand and supply conditions.

Table 28: Firm-level Uncertainty and Country-level Risks

	(1) $\lvert FE^{log}\rvert$	(2) $\lvert FE^{pct}\rvert$	(3) $\lvert\hat{\epsilon}_{FE^{log}}\rvert$
Country risk index	0.0702**	0.0547**	0.0846**
	(0.0302)	(0.0272)	(0.0357)
log($sales$)	-0.0209***	-0.0197***	-0.0162***
	(0.00113)	(0.00105)	(0.00102)
N	65280	65224	65379
R^2	0.198	0.175	0.202
Firm Age	≥ 8	≥ 8	≥ 8
Industry-year Fixed Effect	Yes	Yes	Yes
Parent Fixed Effect	Yes	Yes	Yes
Age Fixed Effect	Yes	Yes	Yes

Standard errors are clustered at the country level, * 0.10 ** 0.05 *** 0.01.

Next, we obtain data on economic policy uncertainty (EPU) for 19 economies for years between 1996 and 2007 and calculate the average EPU for these 19 economies over the 12 years. Then, we correlate the variance of FEs made by old enough affiliates (i.e., affiliates that are at least ten years old) in each of those 19 economies to the EPU index. Table 29 shows that there is a strong positive correlation between the EPU index and the variance of FEs constructed by us. This shows that country-level policy uncertainty seems to be positively associated with firm-level volatility. In total, we conclude that volatility of firm-level demand and supply conditions seems to be correlated to economic policies at the aggregate or industry level.

Table 29: Correlation between EPU and firm-level volatility

	FE^{pct}	FE^{log}	$\hat{\epsilon}_{FE^{log}}$
Economic Policy Uncertainty Index	0.2910	0.1740	0.1873
Type of Firms	all	all	all
obs.	19	19	19

8.6 Calibrating Regional Specific σ_θ and σ_ε

One motivation for examining the impact of σ_θ and σ_ε is that these parameters vary across regions and they are not perfectly correlated. Here we calibrate them for four regions separately

(Asia excluding China, China, North America and Europe) following the same strategy in Section 5.1. Table 30 shows considerable variations in both σ_θ and σ_ε.

Table 30: Moments and parameters for different regions

Region/Country	Asia (non-China)	China (P.R.C.)	North America	Europe
Moments				
Var of $FE_{1,2}$	0.48	0.62	0.45	0.42
Var of FE_{10+}	0.24	0.28	0.23	0.26
Parameters				
σ_θ	2.09	2.78	1.91	1.60
σ_ϵ	0.91	1.08	0.87	0.98

Note: As before, we only use non-experienced affiliates when calculating moments related to the variance and auto-covariance of FEs for the above four regions.

References

Abel, Andrew B., "Optimal investment under uncertainty," *The American Economic Review*, 1983, *73* (1), 228–233.

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.

Andrade, Philippe and Hervé Le Bihan, "Inattentive professional forecasters," *Journal of Monetary Economics*, 2013, *60* (8), 967–982.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple, "Innovation and Production in the Global Economy," Working Paper 18972, National Bureau of Economic Research April 2013.

_ , Theodore Papageorgiou, and Olga Timoshenko, "Firm Learning and Growth," Working Paper 2017.

Bachmann, Rüdiger and Christian Bayer, "Investment Dispersion and the Business Cycle," *American Economic Review*, April 2014, *104* (4), 1392–1416.

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

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, "Measuring economic policy uncertainty," *The Quarterly Journal of Economics*, 2016, *131* (4), 1593–1636.

Bernanke, Ben S., "Irreversibility, uncertainty, and cyclical investment," *The Quarterly Journal of Economics*, 1983, *98* (1), 85–106.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Bertola, Guiseppe and Ricardo J. Caballero, "Irreversibility and Aggregate Investment," *The Review of Economic Studies*, 1994, *61* (2), 223–246.

Bloom, Nicholas, "The impact of uncertainty shocks," *Econometrica*, 2009, *77* (3), 623–685.

_ , Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry, "Really uncertain business cycles," *Econometrica*, 2018, *86* (3), 1031–1065.

Bloom, Nick, Stephen Bond, and John Van Reenen, "Uncertainty and investment dynamics," *The Review of Economic Studies*, 2007, *74* (2), 391–415.

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

Coibion, Olivier and Yuriy Gorodnichenko, "What can survey forecasts tell us about information rigidities?," *Journal of Political Economy*, 2012, *120* (1), 116–159.

_ and _ , "Information rigidity and the expectations formation process: A simple framework and new facts," *The American Economic Review*, 2015, *105* (8), 2644–2678.

_ , _ , and Saten Kumar, "How do firms form their expectations? new survey evidence," Technical Report, National Bureau of Economic Research 2015.

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.

Deseatnicov, Ivan and Konstantin Kucheryavyy, "Exports and FDI Entry Decision: Evidence from Japanese foreign-affiliated firms," Working Paper 17-E-036, RIETI March 2017.

Dixit, Avinash, "Entry and exit decisions under uncertainty," *Journal of political Economy*, 1989, *97* (3), 620–638.

Dixit, Avinash K. and Robert S. Pindyck, *Investment under uncertainty*, Princeton university press, 1994.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Fajgelbaum, Pablo D, Edouard Schaal, and Mathieu Taschereau-Dumouchel, "Uncertainty traps," *The Quarterly Journal of Economics*, 2017, *132* (4), 1641–1692.

Fillat, José L. and Stefania Garetto, "Risk, Returns, and Multinational Production," *The Quarterly Journal of Economics*, September 2015.

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi, "How Exporters Grow," Working Paper 2016.

Guiso, Luigi and Giuseppe Parigi, "Investment and Demand Uncertainty," *The Quarterly Journal of Economics*, February 1999, *114* (1), 185–227.

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

Handley, Kyle, "Exporting under trade policy uncertainty: Theory and evidence," *Journal of international Economics*, 2014, *94* (1), 50–66.

_ **and Nuno Limão**, "Trade and Investment under Policy Uncertainty: Theory and Firm Evidence," *American Economic Journal: Economic Policy*, November 2015, *7* (4), 189–222.

_ **and Nuno Limao**, "Policy Uncertainty, Trade and Welfare: Theory and Evidence for China and the U.S," Working Paper 2017.

Hayashi, Fumio, "Econometrics."

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.

Irwin, Douglas A and Peter J Klenow, "Learning-by-doing spillovers in the semiconductor industry," *Journal of political Economy*, 1994, *102* (6), 1200–1227.

Ito, Takatoshi, "Foreign Exchange Rate Expectations: Micro Survey Data," *The American Economic Review*, 1990, pp. 434–449.

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.," Technical Report, National Bureau of Economic Research 1994.

Klenow, Peter J, "Learning curves and the cyclical behavior of manufacturing industries," *Review of Economic Dynamics*, 1998, *1* (2), 531–550.

Mackowiak, Bartosz and Mirko Wiederholt, "Optimal sticky prices under rational inattention," *American Economic Review*, 2009, *99* (3), 769–803.

Mankiw, Gregory N. and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *The Quarterly Journal of Economics*, 2002, *117* (4), 1295–1328.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, October 2003, *71* (6), 1695–1725.

Mishkin, Frederic S, "The Econometric Methodology," in "A Rational Expectations Approach to Macroeconomics: Testing Policy Ineffectiveness and Efficient-Markets Models," University of Chicago Press, 1983, pp. 9–43.

Muth, John F, "Optimal properties of exponentially weighted forecasts," *Journal of the American Statistical Association*, 1960, *55* (290), 299–306.

Novy, Dennis and Alan M. Taylor, "Trade and Uncertainty," Working Paper 19941, National Bureau of Economic Research February 2014.

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, "The Proximity-Concentration Tradeoff under Uncertainty," *The Review of Economic Studies*, July 2013.

Ruhl, Kim J. and Jonathan L. Willis, "New Exporter Dynamics," Working Paper 2016.

Schaal, Edouard, "Uncertainty and unemployment," *Econometrica*, 2017, *85* (6), 1675–1721.

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