RIETI Discussion Paper Series 18-E-010

Uncertainty, Imperfect Information, and Expectation Formation over the Firm's Life Cycle (Revised)

CHEN, Cheng
Clemson University

SENGA, Tatsuro
RIETI

SUN, Chang
University of Hong Kong

ZHANG, Hongyong
RIETI

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

RIETI Discussion Paper Series 18-E-010
First Draft: March 2018
Revised: September 2020
Revised: July 2022

Uncertainty, Imperfect Information, and Expectation Formation over the Firm's Life Cycle*

Cheng CHEN Tatsuro SENGA Chang SUN Hongyong ZHANG†

Abstract

Using a long-panel data set of Japanese firms that contains firm-level sales forecasts, we provide evidence on firm-level uncertainty and imperfect information over their life cycles. We find that firms make non-negligible and positively auto-correlated forecast errors. However, they make more precise forecasts and less auto-correlated forecast errors when they become more experienced. We then build a model of heterogeneous firms with endogenous entry and exit where firms gradually learn about their demand by using a noisy signal. We present our novel approach to cleanly isolate the learning mechanism from other mechanisms by using expectations data over time. We combine the model with our data to perform a non-parametric decomposition of forecast errors and find that learning accounts for between 20% to 40% of the overall decline in forecast errors over the life cycle. Our model shows that, within the context of our cross-regional data, productivity gains from removing information frictions ranges from 3% to 12%. We find a prominent role of firm entry and exit in generating high productivity gains.

Keywords: firm expectations, forecast errors, uncertainty, learning, productivity
JEL classification: D83; D84; E22; E23; F23; L2

* This research is conducted as a part of project "Studies on Firm Management and Internationalization under the Growing Fluidity of the Japanese Economys" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study utilizes the data of the questionnaire information based on "the Basic Survey on Overseas Business Activities" which is conducted by the Ministry of Economy, Trade and Industry (METI). We thank Costas Arkolakis, Nick Bloom, Vasco Carvalho, Gene Grossman, Kyle Handley, Christian Hellwig, Aubhik Khan, Caliendo Lorenzo, Yulei Luo, Eduardo Morales, Masayuki Morikawa, Steve Redding, Jane Ryngaert, Edouard Schaal, Peter Schott, Wing Suen, Stephen Terry, Eiichi Tomiura, Mirko Wiederholt, Makoto Yano and participants at various seminars and conferences for their helpful comments. Financial support from HKGRF (project codes: 17500618, 17507916, 27502318), JSPS KAKENHI (grant numbers: 17H02531, 17H02554, 22K01451), Princeton University and RIETI is greatly appreciated.
† Chen: Clemson University, chencheng1983613@gmail.com. Senga: Queen Mary University of London, keio University, and RIETI, t.senga@qmul.ac.uk. Sun: the University of Hong Kong, sunc@hku.hk. Zhang: RIETI, zhang-hongyong@rieti.go.jp.

1 Introduction

A growing literature has highlighted the importance of uncertainty and imperfect information in driving firm dynamics and aggregate productivity.[1] In fact, firms face uncertainty and imperfect information when making almost all decisions in a dynamic environment, including investment, hiring, and market entry.[2] A key part of these decisions is to form expectations of future outcomes, such as sales and profits. However, as we seldom observe firms' expectations over their life cycles directly, how firms respond to and resolve uncertainty and imperfect information over time remains unknown. This makes it difficult to quantify the degree of informational imperfections and to evaluate how much they matter for aggregate outcomes, such as aggregate productivity.

In this paper, we make empirical progress by using panel data with quantitative measures of sales expectations at the firm level. Unlike other data sets featuring old and large firms, ours is designed to include young firms since their inception, allowing us to track how firms resolve uncertainty and form expectations over their life cycle.[3] Our main finding is that the precision of sales forecasts increases over a firm's life cycle. Specifically, we present evidence that the variance of forecast errors declines with firms' experience of operation. Moreover, although each firm's forecast errors are positively autocorrelated, the autocorrelation of forecast errors declines with firms' experience, a new fact that has not been unraveled by existing studies. This suggests that firms accumulate experience and make more informed decisions over time, the way of thinking in the theoretical literature of learning.[4]

To investigate its aggregate implications, we build a model of heterogeneous firms based on Jovanovic (1982) and augment the model with two key modifications: (1) firms gradually learn about their demand using payoff-irrelevant signals and (2) firms are subject to idiosyncratic productivity shocks whose volatility varies over the life cycle. As will be shown later, both learning and the age-dependent volatility channels contribute to the age-declining *variance* of forecast errors, while only learning can drive the age-declining *autocorrelation* of forecast errors. We thus can use the age profile of autocorrelation of forecast errors observed in data to pin down parameters that govern learning in the model cleanly. This way, we

[1] See, for example, Bloom (2009) and Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2018), for seminal works.

[2] It is commonly understood that uncertainty matters for individual-level decision-making, such as investment (Guiso and Parigi, 1999), hiring (Bertola and Caballero, 1994), market entry (Dixit, 1989), and trade Handley (2014).

[3] The Ifo business climate survey used in Bachmann, Elstner, and Sims (2013) and Enders, Hünnekes, and Müller (2019) covers a wide range of firms over a long time; however, the major components are qualitative firm-level expectations, with only a smaller set of supplementary quantitative questions available (Bachmann and Elstner, 2015).

[4] See, for example, Jovanovic (1982), among others.

can further distinguish the effect of learning from other confounding mechanisms on the observed age-declining variance of forecast errors. This decomposition is demonstrated within our data to uncover cross-region/country differences in the degree of imperfect information and the potential gains from eliminating it. Our quantitative exercise reveals that (1) the contribution of learning to the change in the variance of forecast errors over the firm's life cycle ranges between 20% to 40% and (2) eliminating imperfect information can lead to productivity gains among firms ranging from 3% to 12%, and we find that entries and exits of firms (i.e., the extensive margin) play a prominent role in generating large gains.

The data set we use is a parent firm-affiliated firm-matched 20-year panel data set on Japanese multinational firms—taken from annual business surveys conducted by the Japanese government. Our data set has three distinctive features: (1) it contains both quantitative expectations and realized outcomes for each firm, which allows us to calculate forecast errors at the firm level; (2) its panel structure and the inclusion of many young firms enable the analysis of within-firm variation of forecast errors over their life cycle; and (3) it is a confidential, mandatory survey enforced by the government, which leads to high response rates (70% on average) and high quality.

Exploring our data set, we show the following features of forecast errors made by an individual firm regarding its sales. First, firms's forecast errors are close to zero on average, and firm-year-level components explain most of the variations in forecast errors (compared with aggregate components such as country-year and industry-year fixed effects). Second, firms make more precise forecasts as they operate longer. Moreover, this result survives even when we control for firm size and measures of market/product diversification at the firm level, suggesting that accumulation and diversification of customer and product portfolios alone cannot explain this fact. Third, past forecast errors are positively correlated with current and future forecast errors, which is similar to Coibion and Gorodnichenko (2012)'s finding regarding households' and professional forecasters' forecast errors. Importantly, we also show that this positive autocorrelation declines over the life cycle of firms. Additionally, we find that firms in countries with better management practices and in countries with smaller time differences from Japan make less serially correlated forecast errors. These stylized facts suggest that firms become better informed as they accumulate more experience. In addition, low-quality management and inefficient within-firm communications could be drivers of information frictions.

To build a model that can reproduce the aforementioned facts about forecast errors as well as other salient empirical regularities regarding firm dynamics, we take a model of heterogeneous firms with life cycles based on Jovanovic (1982) and Arkolakis, Papageorgiou, and Timoshenko (2018). Firms face a downward sloping demand curve in a setting where the firm-specific demand shifter is heterogeneous across firms and unknown to them. The firm-specific demand shifter is constant over the life cycle of firms and they will gradually learn about it.

Two features in our model are important for us to be able to distinguish the effect of learning from other confounding factors on the observed age-declining variance of forecast errors. First, firms in our model face information constraints. We assume that firms learn about their demand from a noisy signal, which is purely informational and *does not* affect firms' per-period profits (i.e., payoff-irrelevant). We show that this information structure allows us to reproduce both the age-declining variance and the age-declining autocorrelation of forecast errors, as compared to other models like Jovanovic (1982), where the variance of forecast errors declines but forecast errors are serially *uncorrelated*.[5] Second, we introduce idiosyncratic shocks to firm-level productivity and assume that its volatility decreases as firms become older, following Atkeson and Kehoe (2005). Therefore, not only learning but also the age-declining volatility can contribute to the age-declining variance of forecast errors in the model, reflecting factors other than learning in much the same way as reality, where there is a host of other reasons that explain why the precision of firms' sales forecasts, such as accumulation and diversification of customer and product portfolios. Importantly, though, the learning mechanism generates the age-declining autocorrelation of forecast errors, while the age-declining volatility mechanism does not. With two moments in hand—variance and covariance of forecast errors—we can decompose the observed forecast errors into the learning effect and other effects. This approach using our model and data reveals how much learning contributes the declining variance of forecast errors. Our decomposition exercise shows that learning components are small in general, explaining about 10 to 20% of the variance of the forecast errors. However, we show that the contribution of learning to the *change* in the variance of forecast errors over the firm's life cycle is large, ranging between 20 to 40%. To the best of our knowledge, our analysis is the first to isolate the evolution of firms' beliefs over their life cycle directly from panel data on firm expectations and succeeds in reproducing the age-declining characteristics of firm-level forecast errors.

In the final part of the paper, we demonstrate our approach that incorporates both the learning and other channels by calibrating our model to infer the learning parameters and other key parameters governing firm dynamics. Our counterfactual experiment of eliminating imperfect information reveals not only a substantial gain in overall productivity, but also the role of firm entry and exit in driving it. In a world with extensive margin choices,

[5] The age-declining autocorrelation of forecast errors implies a deviation from full-information rational expectations, but it can reflect either deviations from full-information or departures from rational expectations. We account for the age-declining autocorrelation of forecast errors by a model of information constraints under rational expectations in the spirit of Coibion and Gorodnichenko (2012, 2015).

eliminating imperfect information leads to not only more informed static decisions (i.e., employment), but also more informed dynamic decisions on entry and exiting. When we allow endogenous entry and exit under imperfect information, providing better information improves average firm-specific demand in the economy substantially. This selection channel leads to larger productivity gains in our model compared with the model without endogenous entry and exit, despite that the number of active firms (varieties) drops after informational imperfection is eliminated. For instance, productivity gains are 3.49% when we remove imperfect information in the model without endogenous entry and exit. However, if we allow endogenous entry and exit productivity gains are 6.35%. We also confirm this result when we implement cross-regional analysis, where we calibrate our model to match data moments for eight regions/countries in the world. We show that the degree of imperfect information and the associated aggregate implications vary across regions/countries, broadly consistent with the view that low-quality management and inefficient within-firm communications can lead to more severe information frictions and therefore productivity losses.[6]

Related literature: While economists have long speculated on how agents form expectations, it is the lack of direct expectations data that has made the treatment of agents' expectations an assumption-based approach. A growing literature breaks with this tradition by collecting and analyzing direct expectations data as in Bloom, Davis, Foster, Lucking, Ohlmacher, and Saporta-Eksten (2020) and Altig, Barrero, Bloom, Davis, Meyer, and Parker (2020), among others.[7] Notably, the seminal works by Coibion and Gorodnichenko (2012) and Coibion and Gorodnichenko (2015) have demonstrated how to best model and calibrate a theoretical framework and thus highlighted the usefulness of such a direct-measure-oriented approach.[8] One feature of our paper is to study firms' expectations of micro objects such as their own sales, instead of macro objects like the GDP growth rate and the inflation rate. In this regard, our approach is in line with that of Enders et al. (2022) and Born, Enders, Müller, and Niemann (2022), who investigate production and price adjustment by firms and how they are causally influenced by firm-level expectations. Our focus in the paper is firms' entry and exit decisions, providing insights into young firms' post-entry dynamics as em-

[6]The caveat here is that our data only contain Japanese firms in various regions/countries, and Japanese firms in one region/country may not be representative for all firms in that region/country. Therefore, the calibrated parameters and the implied productivity gains across regions should be taken with caution.

[7]Other papers that study micro-level expectations include Gennaioli, Ma, and Shleifer (2016) for U.S. firms, Bachmann and Elstner (2015), Bachmann, Elstner, and Hristov (2017), Triebs and Tumlinson (2013), and Enders, Hünnekes, and Müller (2022) for German firms, Boneva, Cloyne, Weale, and Wieladek (2020) for firms in the United Kingdom, Tanaka, Bloom, David, and Koga (2019) for Japanese firms, and Coibion, Gorodnichenko, and Ropele (2020) for Italian firms.

[8]Recent papers that have studied how agents form expectations and respond to shocks include Coibion, Gorodnichenko, and Kumar (2018), Baker, McElroy, and Sheng (2020), and Enders et al. (2022).

phasized by, for example, Sedláček and Sterk (2017) and Foster, Haltiwanger, and Syverson (2016).

Our paper contributes to the recent literature on firm-level uncertainty by showing how it evolves over the life cycle of firms. While it is extensively documented that firm-level uncertainty varies over time and across firms, within-firm variation of uncertainty over the life cycle has received less attention. For example, Bloom et al. (2018) and Kehrig (2015) show the cross-sectional dispersion of firm-level output and productivity fluctuates coutercyclically, while Vavra (2014) and Berger and Vavra (2019) show that the dispersion of price changes fluctuates coutercyclically. Moreover, Bachmann and Bayer (2014) show that the cross-sectional dispersion of firm-level investment rates fluctuates procyclically. We focus on how firm-level uncertainty fluctuates within firms and share the spirit with Baley and Blanco (2019) and Baley, Figueiredo, and Ulbricht (2022) who study how uncertainty fluctuates or creates a cycle within firms, in the context of price adjustment and worker-firm match quality, respectively. Ilut, Valchev, and Vincent (2020) also study a life-cycle profile of uncertainty within firms in a model where new firms accumulate information signals over the life cycle. In the business cycle context, the role of information accumulation at the firm level has been studied by Ilut and Saijo (2021), who also use forecast data to validate the structural model.

Finally, we identify and quantify misallocation due to imperfect information along the life cycle of firms. Our focus on the life cycle of firms is reminiscent of Hsieh and Klenow (2014), and our quantitative exercise is related to David, Hopenhayn, and Venkateswaran (2016), who also substantiate the role of imperfect information in determining allocative efficiency. What distinguishes our paper from theirs is that we show that productivity losses through extensive margin dynamics—firms' entries and exits—are substantial. Regarding the importance of the extensive margin, our paper complements the results from Midrigan and Xu (2014) and Buera, Kaboski, and Shin (2011), among other papers on misallocation, although they focus on financial frictions.[9]

2 Empirical Facts

In this section, we construct our panel of Japanese firms operating in foreign markets to document the properties of the forecast errors and their relationship with firms' experience. First, the forecast errors made by firms become smaller as they become more experienced. Second, the forecast errors are positively autocorrelated, but the serial correlation declines as they become more experienced. In addition, firms in countries with better management

[9]Related literature includes Khan and Thomas (2013) and Buera and Moll (2015) who studied the role of financial frictions in generating capital misallocation and its aggregate implications.

and/or in countries with small time differences from Japan show smaller serial correlation of forecast errors. These facts indicate that firms become better informed as they accumulate more experience, and management and within-firm communication could be one driver of information frictions.

2.1 Data and the Reliability of Sales Forecasts

Our main data source is the Basic Survey on Overseas Business Activities ("foreign activities survey" hereafter) conducted by the Ministry of Economy, Trade and Industry (METI). The survey contains information on overseas affiliated firms of Japanese parent companies, including the affiliated firms' location, industry, sales, and employment. The survey covers two types of overseas businesses: (1) direct (first-tier) affiliated firms with more than 10% of the equity share capital owned by Japanese parent companies, and (2) second-tier affiliated firms with more than 50% of the equity share capital owned by Japanese parent companies. Dropping tax haven countries documented in Gravelle (2009), our baseline regression sample contains on average 1781 parent companies and 6922 affiliated firms in a typical year from 1995 to 2013. Our sample covers Japanese firms operating in 96 countries and 29 industries, including both manufacturing and services. In Online Appendix Section 1.1, we report descriptive statistics regarding sub-samples in different time periods and the distribution of firms across regions and industries in a typical year. The unit of analysis in our empirical investigation is the affiliated firm by year. We slightly change the terminology: We refer to the affiliated firms as "firms" and to all the affiliated firms belonging to the same parent company as a "business group".

The unique feature of the foreign activities survey is that each firm reports its sales forecast for the next year when it fills out the survey of the current year. Because such information is rarely available in firm-level data sets, we show that the sales forecasts are reliable and contain useful information that affect actual firm decisions.

First, we show that firms do not use naive rules to make their sales forecasts. In Table 1, we present the expected growth rates, calculated as the ratio of the firm's forecast for year $t+1$ to its realized sales in year t minus one. If firms simply use their realized sales in year t to predict their sales next year, the expected growth rate will be zero. In Table 1, only 3.35% of the observations in our sample have a zero expected growth rate. The shares of the other frequent cases are all extremely low. For the firms reporting zero expected growth rates, it is difficult to tell whether they are making a naive forecast or a serious forecast with the expectation that their sales growth will be close to zero. We therefore conduct robustness checks of our main regressions in Online Appendix Tables OA.8 and OA.15 by dropping all observations with zero expected growth rates. Our empirical results remain largely unchanged.

Table 1: The Most Frequent Values of Expected Growth Rates

Top 1-5		Top 6-10	
$E_t(R_{t+1})/R_t - 1$	Freq. (%)	$E_t(R_{t+1})/R_t - 1$	Freq. (%)
0.0000	3.35	0.0714	0.11
0.1111	0.22	0.3333	0.11
0.2500	0.20	0.0417	0.11
0.0526	0.17	0.0870	0.11
0.2000	0.14	1.0000	0.10

Notes: The table reports the most frequent values of expected growth rates among all firm-year observations. Zero means that the firm expect the next year's sales to be exactly the same as this year's.

Second, we show that the sales forecasts have statistically significant and economically strong impacts on future firm outcomes. Specifically, we regress the realized sales in year $t+1$ on the sales forecast made in year t and a set of fixed effects, and the results are reported in Table 2. The first three columns of Table 2 show that the sales forecasts in year t positively and significantly predict the realized sales in year $t+1$. Importantly, the effect of the sales forecast does not disappear when we include the realized sales in year t as a control variable in Column 2. The coefficient of sales forecast is much larger than that of realized sales in the previous year. Further including the realized sales in year $t-1$ does not change this pattern (Column 3). Columns 4–6 show that the sales forecasts also have strong predicative power for future employment, even if we control for current and past employment. These findings easily reject the hypothesis that firms fill out this survey question with random guesses. By contrast, firms make these forecasts seriously, and the forecasts contain more information on the firms' future conditions than realized outcomes in the past.

Finally, the foreign activities survey is mandated by METI under the Statistics Law; thus, the information in the survey cannot be applied to purposes beyond the scope of the survey, such as tax collection. Firms have no incentive to misreport because of tax purposes. Moreover, unlike earnings forecasts announced by public firms, the sales forecasts reported to METI are confidential; thus, firms have no incentive to misreport strategically and manage the expectations of the stock market investors. In total, the aforementioned empirical patterns assure us that the sales forecasts contained in the foreign activities survey are reliable and suitable for our empirical analysis.

Table 2: Sales Forecasts Predict Firms' Future Outcomes

Dep. Var.	log total sales $\log(R_{i,t+1})$			log employment $\log(L_{i,t+1})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\log E_t(R_{i,t+1})$	0.673^a	0.550^a	0.584^a	0.301^a	0.132^a	0.132^a
	(0.011)	(0.012)	(0.015)	(0.013)	(0.007)	(0.007)
$\log R_{it}$		0.138^a	0.080^a			
		(0.008)	(0.016)			
$\log R_{i,t-1}$			0.063^a			
			(0.007)			
$\log L_{it}$					0.511^a	0.505^a
					(0.011)	(0.014)
$\log L_{i,t-1}$						0.057^a
						(0.007)
Country-Year FE	Y	Y	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
N	128937	127277	106785	127485	126534	106819
# of B-groups (cluster)	4951	4931	3859	4938	4924	3837
Within R-squared	0.477	0.484	0.493	0.163	0.384	0.392
R-squared	0.961	0.964	0.967	0.958	0.970	0.972

Notes: The dependent variable is firm i's log total sales or total employment in year $t+1$. We use R to denote sales and L to denote employment. $E_t(R_{i,t+1})$ refers to the firm's expectation in year t for its sales in year $t+1$. Standard errors are clustered at the business group level. Significance levels: a: 0.01, b: 0.05, c: 0.10.

2.2 Forecast Errors

We now describe how firms' forecast errors evolve over their life cycles. Our main measure of forecast errors is the log point deviation of the realized sales from the sales forecast as

$$FE^{\log}_{t,t+1} \equiv \log\left(R_{t+1}/E_t\left(R_{t+1}\right)\right),$$

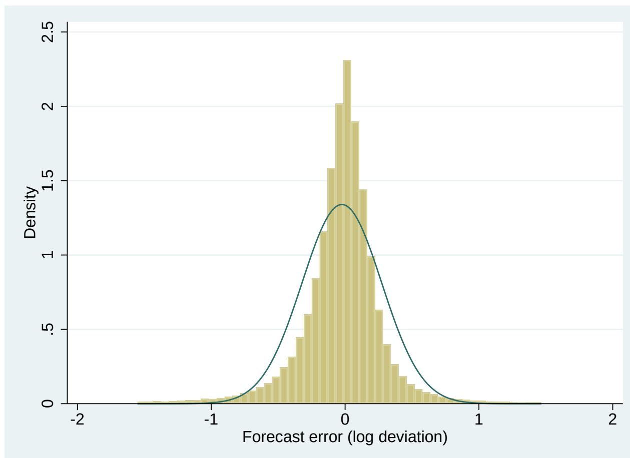
where R_{t+1} is the realized sales in period $t+1$ and $E_t\left(R_{t+1}\right)$ denotes a firm's time t forecast of its sales next period. A positive (negative) forecast error means that the firm under-predicts (over-predicts) its sales. In Online Appendix Tables OA.5, OA.6, OA.13, OA.14, we show that our key empirical results are robust to two alternative definitions of forecast errors: the percentage deviation, and the residual of raw forecast errors after removing aggregate components such as industry and country-year fixed effects.[10] We also trim the top and bottom one percent of observations of the forecast errors, to exclude outliers.

In Figure 1, we plot the distribution of our leading measure of forecast errors, $FE^{\log}_{t,t+1}$,

[10]The aggregate components explain approximately 11% of the variation in forecast errors. Recent work has substantiated that firms may have heterogeneous exposure to aggregate shocks, which implies that the "simple" residual forecast errors we construct may still be affected by the aggregate economic conditions. Therefore, we construct alternative residual forecast errors by explicitly considering firms' heterogeneous exposure to aggregate shocks. For these alternative residual forecast errors, aggregate components explain approximately 23% of the variation in forecast errors, but our main empirical findings are robust to these alternative measures. Detailed discussions are in Section 1.3.6 of the online appendix.

across all firms in all years. The forecast errors are centered around zero, and the distribution appears to be symmetric. The shape of the density is similar to a normal distribution, although the center and the tails have more mass than the fitted normal distribution (solid line in the graph). The average forecast error across all firm-year observations is –0.024, with a median of –0.005 and a standard deviation of 0.298. The absolute value of $FE^{\log}_{t,t+1}$ is 0.2, which implies that firms on average over forecast or under forecast their sales by 20%.

Figure 1: Distribution of the Forecast Errors



Notes: Histogram of $FE^{\log}_{t,t+1}$ with the fitted normal density (solid line).

Fact 1: Precision of Forecasts Increases as Firms Become More Experienced

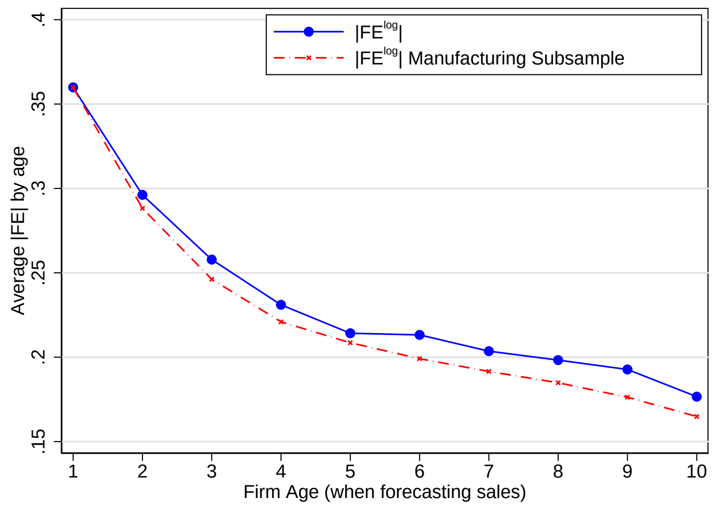
Figure 2 presents the average absolute value of forecast errors by age cohorts, where age is top-coded at ten. The precision of sales forecasts increases as the firm ages. Specifically, as firms age from one to ten years, the absolute forecast errors decline from 36% to 18% on average. Moreover, the decline occurs mainly in the first five years after entry. For concreteness, we also present these statistics for a subsample in the manufacturing sector. The patterns are similar.

We further confirm these patterns formally by an OLS regression of firm i's absolute forecast error in year t:

$$|FE^{\log}_{it,t+1}| = \delta_n + \beta X_{it} + \delta_{ct} + \delta_{st} + \varepsilon_{it}, \tag{1}$$

where δ_n is a vector of age dummies, δ_{ct} represents the country-year fixed effects, and δ_{st} represents the industry-year fixed effects. Time-varying controls such as firm size are denoted by X_{it}. We use age one as the base category; therefore, the age fixed effects represent the difference in the absolute forecast errors between age n and age one. To further control for heterogeneity across firms, we also run regressions with firm fixed effects δ_i.

Figure 2: $|FE^{\text{log}}|$ Declines with Firm Age



Note: Average absolute value of FE^{log} by age cohorts.

Column 1 in Table 3 shows the baseline specification with industry and country-year fixed effects. As firms become older, the absolute forecast errors decline. On average, firms that are at least ten years old have absolute forecast errors 17 log points lower. In Columns 2 and 3, we control for the size of the firms and their parent companies in Japan (measured by log employment). Although larger firms tend to have smaller absolute forecast errors, the age effects survive.[11]

One possible explanation of the age effects is that firms gradually improve their management quality or capability such that they can obtain a more diversified portfolio of destination markets and products. We show that the age effects are not totally driven by such diversification. To examine the importance of market diversification, we control for the concentration of firm sales across markets, measured by the Herfindahl–Hirschman Index (HHI) of the shares of firm sales in six markets (the finest classification available in the survey): the host country (local market), Japan, Asia, North America, Europe, and the rest of the world. Column 4 shows that a higher value of market-level concentration increases the absolute forecast errors, but controlling for it has a limited impact on the age effects. Unfortunately, our main data source does not provide a breakdown of firm sales by products. To examine

[11]Tanaka et al. (2019) report that older firms make more precise forecasts than younger firms do, on the basis of cross-section results. By contrast, our finding is based on within-firm variation with the firm fixed effects, thereby pointing to the life cycle pattern of forecast errors.

the importance of product diversification, we focus on the subset of firms in China which we managed to match with the Chinese customs data between 2000 and 2009. For the matched observations, we calculate a product sales HHI of using the customs' records on firm exports at the HS 6-digit product level. Column 5 shows that a higher value of product concentration also has a positive impact on absolute forecast errors, but it is not significant because of the small sample size. The age effects, however, are robust to controlling for the product concentration.[12]

To evaluate the robustness of our results, we restrict our sample to (1) surviving entrants and (2) firms in manufacturing. Column 6 reports the result for a subsample of firms that have survived and continuously appeared in the data from age one to seven, which shows that our results are not driven by endogenous exits and nonreporting. Column 7 focuses on the manufacturing subsample, and the results are similar. In the Online Appendix 1.3, we further show that the age effects (1) are robust to various alternative measures of forecast errors, including those that explicitly take firms' heterogeneous exposure to aggregate shocks into account, and (2) are not driven by the fact that firms enter in different months of a fiscal year and that "age-one" firms actually have fewer than 12 months of experience (the so-called "partial-year effects"), and (3) are not due to age-dependent biases in the level of forecast errors.[13] For the last robustness check, we design a two-step procedure to address age-dependent biases in forecast errors. In particular, we run a first-stage regression on the level of forecast errors, and project the squared residuals from the first stage on the same set of independent variables in the second stage. These regressions show that the conditional variance of forecast errors declines significantly with firm age. Interested readers can refer to Section 1.3.2 of the online appendix.

Fact 2: Forecast Errors are Positively Autocorrelated but Less So as Firms Become More Experienced

A growing literature has highlighted the serial correlation of forecast errors in various contexts. For example, Ryngaert (2017) and Coibion and Gorodnichenko (2012) demonstrated that professional forecasters' forecast errors of future inflation rates are autocorrelated, indicating their imperfect information on macroeconomic conditions. Instead of using expectations data on macroeconomic outcomes, we utilize data on the sales expectations of

[12]We provide details on the construction of these measures, the matching between China customs data and our main data source, and additional robustness checks regarding product and market diversification in the Online Appendix 1.3.4.

[13]Age-dependent biases in the level of forecast errors occur, for example, when young firms over-predict future sales as they are too optimistic. This can cause biases of the age effects on the absolute forecast errors.

Table 3: Age Effects on the Absolute Forecast Errors

Sample:	All Firms				China	Survivors	Manufacturing
Dep.Var: $\lvert FE^{\log}_{t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\mathbb{1}(\text{Age}_t = 2)$	-0.066^a	-0.059^a	-0.063^a	-0.049^a	-0.048	-0.068^a	-0.072^a
	(0.007)	(0.007)	(0.008)	(0.009)	(0.031)	(0.010)	(0.011)
$\mathbb{1}(\text{Age}_t = 3)$	-0.102^a	-0.089^a	-0.088^a	-0.067^a	-0.073^b	-0.093^a	-0.104^a
	(0.007)	(0.007)	(0.008)	(0.008)	(0.030)	(0.010)	(0.011)
$\mathbb{1}(\text{Age}_t = 4)$	-0.128^a	-0.113^a	-0.110^a	-0.085^a	-0.077^a	-0.108^a	-0.127^a
	(0.007)	(0.007)	(0.008)	(0.009)	(0.030)	(0.011)	(0.011)
$\mathbb{1}(\text{Age}_t = 5)$	-0.142^a	-0.125^a	-0.116^a	-0.094^a	-0.089^a	-0.121^a	-0.128^a
	(0.007)	(0.007)	(0.008)	(0.009)	(0.030)	(0.012)	(0.011)
$\mathbb{1}(\text{Age}_t = 6)$	-0.142^a	-0.124^a	-0.114^a	-0.092^a	-0.074^b	-0.120^a	-0.131^a
	(0.007)	(0.007)	(0.008)	(0.009)	(0.032)	(0.013)	(0.011)
$\mathbb{1}(\text{Age}_t = 7)$	-0.152^a	-0.131^a	-0.120^a	-0.100^a	-0.075^b	-0.134^a	-0.138^a
	(0.007)	(0.007)	(0.008)	(0.009)	(0.032)	(0.014)	(0.011)
$\mathbb{1}(\text{Age}_t = 8)$	-0.156^a	-0.133^a	-0.121^a	-0.100^a	-0.085^b	-0.125^a	-0.140^a
	(0.007)	(0.007)	(0.009)	(0.009)	(0.033)	(0.016)	(0.012)
$\mathbb{1}(\text{Age}_t = 9)$	-0.160^a	-0.135^a	-0.122^a	-0.104^a	-0.086^b	-0.126^a	-0.143^a
	(0.007)	(0.007)	(0.008)	(0.009)	(0.034)	(0.017)	(0.012)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.172^a	-0.137^a	-0.121^a	-0.103^a	-0.079^b	-0.129^a	-0.137^a
	(0.007)	(0.007)	(0.009)	(0.009)	(0.037)	(0.019)	(0.012)
$\log(\text{Emp})_t$		-0.021^a	-0.024^a	-0.023^a	-0.030^a	-0.035^a	-0.025^a
		(0.001)	(0.002)	(0.002)	(0.011)	(0.005)	(0.002)
$\log(\text{Parent Emp})_t$		0.001	0.001	-0.001	0.001	0.010	0.001
		(0.001)	(0.003)	(0.003)	(0.010)	(0.007)	(0.003)
HHI Market Sales at t				0.015^a			
				(0.005)			
HHI HS6 Product Exports at t					0.006		
					(0.014)		
Industry-year FE	Y	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y	Y
Firm FE			Y	Y	Y	Y	Y
N	131230	128429	123111	104598	8177	21982	76823
R^2	0.104	0.122	0.366	0.376	0.393	0.357	0.363

Notes: Standard errors are clustered at the business group level. Significance levels: c: 0.10, b: 0.05, a: 0.01. The dependent variable is the absolute value of forecast errors in all regressions. Age refers to the age of the firm when making the forecasts. Regressions in Columns 1–4 include all firms. Column 5 only includes firms in China that can be matched to the Chinese customs data. Survivors (Column 6) refer to firms that have continuously appeared in the sample from age one to seven. Column 7 focuses on firms in manufacturing.

individual firms and show that their forecast errors are positively autocorrelated over time. Importantly, we document that the serial correlation of forecast errors declines with the firm's age.

Table 4: Correlation of $FE^{\log}_{t,t+1}$ and $FE^{\log}_{t-1,t}$, Overall and by Age Group

Sample	All ages	Age 2-4	Age 5-7	Age $\geq$ 8
All industries	0.137	0.170	0.152	0.120
	[96452]	[10410]	[13801]	[72241]
Manufacturing	0.139	0.193	0.151	0.116
	[60123]	[5828]	[8591]	[45704]

Notes: $FE^{\log}_{t,t+1}$ is the log deviation of the realized sales in year $t+1$ from the sales forecast made in year t. Age is measured at the end of year t. Number of observations used for each correlation is shown in the brackets below. All correlation coefficients are significant at the 1% level.

Table 4 presents the serial correlation of forecast errors, for the entire sample and different age groups. Among all firm-year observations, we find that the correlation coefficient between $FE^{\log}_{t,t+1}$ and $FE^{\log}_{t-1,t}$ is 0.137. This result suggests that firms tend to make systematic errors in forecasting their sales. The remaining three columns present the serial correlation for different age groups. When firms become more experienced, the positive correlation is reduced, indicating that firms become more informed and make smaller systematic errors when forecasting. Such patterns are robust when focusing on the manufacturing subsample. We find similar patterns when using alternative definitions of forecast errors (see Online Appendix Table OA.12). Importantly, our results are robust when using percentage forecast errors, $\frac{R_{t+1}-E_t(R_{t+1})}{E_t(R_{t+1})}$, and are not an artifact of the log transformation.

We next confirm this pattern by running the AR(1) type of regressions at the firm level. This allows us to control for the time-varying firm characteristics and various sets of fixed effects to rule out confounding factors. In particular, we run the following regression:

$$FE^{\log}_{i,t+1,t+2} = \beta_1 FE^{\log}_{i,t,t+1} + \beta_2 FE^{\log}_{i,t,t+1} \times Age_{it} + \beta_3 X_{it} + \delta_{st} + \delta_{ct} + \delta_g + u_{it}, \qquad (2)$$

where Age_{it} denotes the firm's age at time t and X_{it} denotes the firm's other time-varying characteristics such as employment at time t. In all regressions, we control for the industry-year, country-year, and business group fixed effects, denoted by δ_{st}, δ_{ct} and δ_g, respectively. In some regressions, we replace the business group fixed effects by business group-firm age fixed effects.

Table 5 shows the regression results. To capture the nonlinear effect of the firm's age, we use either age top-coded at ten or the log of age. According to the estimates in Column 1, the AR(1) coefficient starts at 0.098 at age one and each additional year of experience reduces it by 0.006. When controlling for business group-firm age fixed effects instead of

Table 5: AR(1) Regressions and the Effect of Age

Dep.Var: $FE^{\log}_{t+1,t+2}$	All firms				Manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$FE^{\log}_{t,t+1}$	0.104^a	0.100^a	0.131^a	0.123^a	0.114^a	0.112^a	0.137^a	0.134^a
	(0.014)	(0.013)	(0.018)	(0.017)	(0.019)	(0.019)	(0.025)	(0.025)
$\times \max\{\text{Age}_t, 10\}$	-0.006^a		-0.008^a		-0.008^a		-0.009^a	
	(0.002)		(0.002)		(0.002)		(0.003)	
$\times \log(\text{Age}_t)$		-0.018^a		-0.023^a		-0.027^a		-0.031^a
		(0.006)		(0.007)		(0.009)		(0.011)
$\log(\text{Emp})_t$	0.003^a	0.003^a	0.002^b	0.002^b	0.002^c	0.002^c	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\log(\text{Parent Emp})_t$	-0.010^b	-0.010^b	-0.010^b	-0.010^b	-0.010^c	-0.010^c	-0.014^b	-0.014^b
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Business Group FE	Y	Y			Y	Y		
Busi.Group-Age FE			Y	Y			Y	Y
N	93478	93478	84839	84839	58630	58630	52510	52510
R^2	0.205	0.205	0.274	0.274	0.229	0.229	0.300	0.300

Notes: Standard errors are clustered at the business group level, c: 0.10, b: 0.05, a: 0.01.

business group fixed effects, the AR(1) coefficients as well as the impact of firm age are higher. Results are similar when we focus on firms in the manufacturing sample (Columns 5–8).

Fact 3: Potential Drivers of Information Frictions

Our data offer a wide coverage of countries where Japanese firms operate. This subsection explores how serial correlation of forecast errors are correlated with various characteristics of each country, using similar specifications as in Table 5, to shed light on potential drivers of underlying differences in informational imperfection across countries.

We focus on three country characteristics: (1) management; (2) time zone differences; and, (3) real GDP per capita. As suggested by Bloom, Kawakubo, Meng, Mizen, Riley, Senga, and Van Reenen (2021), better managed firms are able to make more accurate forecasts about their own sales growth. We therefore use country-level average management scores that are used in Bloom, Lemos, Sadun, Scur, and Van Reenen (2014) as a measure of the management quality in each country. Second, the literature has identified time zone differences as barriers to communication within (multinational) firms (Gumpert (2018) and Bahar (2020)), which possibly lead to severe information frictions. Finally, we examine real GDP per capita at the beginning of our sample (1995), which is a proxy for the overall development level of the countries. We interact the country characteristics with the (one-period) lagged forecast error to see how they affect the AR(1) coefficient. These results are by no means causal and the list of drivers we study here is not exhaustive. However, they still shed light on why information frictions at the firm level differ.

Table 6: AR(1) Coefficient and Country Characteristics

	Dep.Var: $FE^{\log}_{t+1,t+2}$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$FE^{\log}_{t,t+1}$	0.1264^a	0.1121^a	0.1077^a	0.0701^a	0.0643^a	0.0606^a	0.0837^a	0.0705^a	0.0670^a
	(0.0080)	(0.0068)	(0.0070)	(0.0094)	(0.0075)	(0.0078)	(0.0101)	(0.0083)	(0.0085)
× Management Score (WMS 2015)	-0.0131^c			-0.0087			-0.0229^a		
	(0.0070)			(0.0071)			(0.0081)		
× Time Diff from Japan		0.0098			0.0142^b			0.0116	
		(0.0066)			(0.0066)			(0.0073)	
× log GDP p.c. 1995			-0.0112^c			-0.0077			-0.0178^a
			(0.0058)			(0.0058)			(0.0066)
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Business Group FE				Y	Y	Y			
Busi.Group-Age FE							Y	Y	Y
N	62005	96100	96100	61200	95152	95152	53433	86271	86271
R^2	0.130	0.135	0.135	0.207	0.201	0.201	0.283	0.270	0.270

Notes: Standard errors are clustered at the business group level. Significance levels: c: 0.1, b: 0.05, a: 0.01. Management score is from the World Management Survey up to 2015 . Management score, time zone differences and log GDP per capita are all standardized to facilitate interpretation of the coefficients.

Table 6 reports the regression results. Country characteristics are all standardized to facilitate interpretation. In Columns 1–3, we control for industry-year and country-year fixed effects, while we further control for business group or business group-firm age fixed effects in the other columns. In general, we find that the management score and GDP per capita are negatively associated with the AR(1) coefficient of forecast errors, while time zone differences affect the coefficient positively. In the most demanding specifications (Columns 7–9), we see that a one standard deviation of management score and GDP per capita reduces the AR(1) coefficient by 0.023 and 0.018, respectively. A one standard deviation of time zone differences increases the coefficient by 0.012. If we view the AR(1) coefficient as a measure for information frictions, these results are consistent with our hypotheses that better management, more similar time zones, and higher development levels are negatively associated with the severity of firm-level information frictions.

In the context of our sample of multinational firms, the result regarding management capabilities can be thought of as suggesting that better managed firms have superior practices of monitoring, one important component of management practices defined in Bloom and Van Reenen (2007), that help them collect and digest crucial information to make better forecasts, leading to less correlated forecast errors. Time zone differences can also reflect information barriers with respect to intrafirm communication between parents and affiliates. More similar time zones with more overlapped working hours make communication easier and thus reduce information frictions within the firm. Finally, the economic development level of an economy affects the supply of good managers (see Hjort, Malmberg, and Schoellman (2021)). Therefore, a higher level of GDP per capita can lead to better managers that work in multinational firms (and other firms) in the host country. As a result, those firms' managers make less correlated forecast errors.

As we have shown that the AR(1) coefficients are also affected by firm age, we examine the robustness of the above results by introducing a horse race between country characteristics and firm age in Online Appendix Table OA.16. We find that age still significantly reduces the AR(1) coefficients, and the country characteristics have the expected effects as in Table 6. We also run a horse race between time zone differences and GDP per capita, and find that the former and the latter significantly increases and reduces the AR(1) coefficient, respectively.[14]

3 Model

We develop a dynamic industry equilibrium model with Jovanovic (1982)-type learning embedded as in Arkolakis et al. (2018). We enrich the model by having two distict features. Firstly, each firm updates ex post beliefs about its permanent demand observing a noisy signal. Secondly, the variance of shocks to idiosyncratic productivity decline exogenously over the life cycle. As will become clear below, this setup helps us match the aforementioned stylized facts and provides a framework to quantify informational imperfections using our data.

3.1 Setup

In our model, time is discrete with periods t = 1,2,..., and the representative consumer spends income Y_t on goods produced by monopolistically competitive firms. Consumer utility from consuming $q_t(\omega)$ units of different products ω can be expressed using the quantity of the following CES aggregate:

$$Q_t = \left(\int_{\omega \in \Omega_t} e^{\frac{\theta(\omega)}{\sigma}} q_t(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

where σ is the elasticity between different varieties, $\theta(\omega)$ is the demand shifter for variety ω, and Ω_t denotes the set of varieties available at time t. We can express the demand for a particular variety, ω, as:

$$q_t(\omega) = Y_t P_t^{\sigma-1} e^{\theta(\omega)} p_t(\omega)^{-\sigma}, \tag{4}$$

[14]The correlation between the management score and GDP per capita in our sample is 0.94. We therefore do not have enough variations to separately identify the impact of these two variables on the AR(1) coefficient. In contrast, the correlation between time zone differences and GDP per capita is 0.60.

where P_t is the price index of the industry:

$$P_t \equiv \left(\int_{\omega \in \Omega_t} e^{\theta(\omega)} p_t(\omega)^{1-\sigma} d\omega \right)^{1/(1-\sigma)}. \tag{5}$$

The firm-specific demand, $\theta(\omega)$, is unknown to the firm but it understands that $\theta(\omega)$ is drawn from a normal distribution $N\left(\bar{\theta}, \sigma_\theta^2\right)$. We assume that the firm cannot fully uncover its permanent demand draw $\theta(\omega)$ from sales observation, faced with constraints in collecting and processing information. Instead, the firm receives a *noisy* signal about permanent demand draw $\theta(\omega)$ and needs to learn about it over the life cycle:

$$s_t(\omega) = \theta(\omega) + \varepsilon_t(\omega), \tag{6}$$

where $\varepsilon_t(\omega)$ is an i.i.d. noise term and drawn from a normal distribution $N(0, \sigma_\varepsilon^2)$. The noise term can reflect errors in managing and sharing financial data inside the firm, and thus managers are unable precisely back out the implied demand draw $\theta(\omega)$ from available information like realized sales.[15]

As we will show below, our chosen information structure generates the aforementioned age-declining serially correlated forecast errors about sales (Fact 2).[16] The key is that $\varepsilon_t(\omega)$ is payoff irrelevant, being purely informational and orthogonal to firms' per-period profits. If $\varepsilon_t(\omega)$ is a real term and payoff relevant as in Jovanovic (1982) and Arkolakis et al. (2018), sales forecast errors are serially *uncorrelated*. In Online Appendix 2.3, we show that sales forecast errors are serially *uncorrelated* if $\varepsilon_t(\omega)$ is a real term and payoff relevant as in Jovanovic (1982) and Arkolakis et al. (2018) and there are no endogenous exits.[17] In addition, alternative information environments and shock processes, such as perfect information and

[15] It can also reflect the fact that management practices such as monitoring are far from being perfect inside the firm (e.g., Bloom and Van Reenen (2007), Bloom and Van Reenen (2010a)). As a result, managers are unable to measure employees' efforts perfectly and thus cannot use realized sales to *precisely* back out the implied demand draw $\theta(\omega)$.

[16] Not only a constraint in collecting and processing information but also a lack of knowledge about underlying model structures can lead to serially correlated forecast errors. To make our quantitative decomposition of forecast errors transparent, we incorporate only the former but not the latter. See Ryngaert (2017) for the quantitative importance of each channel for inflation forecasts.

[17] Suppose, for example, that the firm-specific demand shifter consists of a permanent component and a transitory one, both of which are payoff-relevant and not separately observable, as in Arkolakis et al. (2018). Then, firm's information set in period t includes all the realized demand shifter and forecasts made up to period t (forecasts are functions of realized demand shifters). As a result, forecast errors for period $t+1$ is independent of their lagged values, due to the nature of conditional expectations and the fact that the lagged forecast errors are functions of previous demand shifters up to period t. See Online Appendix 2.3 for further details. In our model, however, the realized demand shifter in our model θ (every period) is never in firm's information set in period t, as the firm cannot perfectly observe it. Inside the firm's information set in period t are the series of the noisy signals and the forecasts made in the past.

learning about an time-varying, AR(1) process of the firm demand θ_t, imply zero forecasting errors (see Online Appendix 2.1 and 2.4). On the other hand, endogenous exits are unlikely to drive the results. We show that they generate negative correlated forecasting errors under perfect information with AR(1) type of shocks (see Online Appendix 2.2).

Output is linear in labor with $q_t = \varphi_t l_t$ and firms hire workers at the wage rate of w. Firms' labor productivities follow an AR(1) process, where the variance of the shock is age-dependent. Replacing the time subscript with firm age n, we can write the productivity process as

$$\log \varphi_n = \mu_\varphi + \rho \log \varphi_{n-1} + \nu_n, \quad \nu_n \sim N(0, \sigma^2_{\nu_n}).$$

This setup allows for age-dependent volatility. When we parameterize the model, we follow Atkeson and Kehoe (2005) and assume that σ_{ν_n} declines according to a quadratic function up to an age cutoff. There are multiple reasons why the variance of forecast errors or sales growth rates declines over firms' life cycles, including learning, accumulation of customers, and diversification of product portfolios. The age-dependent volatility captures mechanisms other than learning in a "reduced-form" way. We incorporate this term into our model, as we want to isolate the contribution of the learning mechanism (to the decline of the variance of forecast errors) from contributions made by the other alternative mechanisms mentioned above. In short, we acknowledge the possible existence of other alternative mechanisms in our empirical setting and let the data tell us the contribution purely made by the learning mechanism. As we will discuss in Section 4, information on autocovariance *and* variance of the forecast errors helps us *separately* identify learning and age-dependent volatility, as age-dependent productivity shocks do not affect the autocovariance of forecast errors.

The timing of the events in a given period t is as follows. At the beginning of each period, the incumbents receive an exogenous exit shock with probability η randomly. Surviving incumbents choose between staying in the market by paying a fixed cost f in units of labor and exiting (permanently). Conditional on staying in the market, firms decide how many workers to hire, l_t, before the current labor productivity φ_t is realized. At the end of the period, the productivity φ_t is realized and the firm chooses the price p_t to sell all the products produced, as we assume there is no storage technology and firms cannot accumulate inventories. Finally, firms observe the new signals s_t and update beliefs.

In each period, there is a unit mass of potential entrants that decide whether to enter the market. Each entrant makes a permanent demand draw θ from a normal distribution, $N\left(\bar{\theta}, \sigma^2_\theta\right)$ and an initial labor productivity φ_0 from a log normal distribution, $\log N\left(0, \sigma^2_{\varphi_0}\right)$. Potential entrants have perfect information about φ_0 and know the distribution of θ that is consistent with the data generating process, $N\left(0, \sigma^2_\theta\right)$. In other words, they know the unbiased prior distribution of θ, but not the value of θ they draw. Potential entrants have to make a decision on whether to enter the market. Those with high labor productivity φ_0 will enter and produce in this market.

3.2 Belief Updating

In this subsection, we discuss how a firm forms the ex post belief for its permanent demand. At the beginning of a period, a firm that is $n+1$ ($n \geq 1$) years old has observed noisy signals of the permanent demand draw in the past n periods: $s_1, s_2, \ldots, s_n$. Because both the prior and the noisy signals are normally distributed, Bayes' rule implies that the posterior belief about θ is normally distributed with mean μ_n and variance σ_n^2:

$$\mu_n = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n\sigma_\theta^2}\bar{\theta} + \frac{n\sigma_\theta^2}{\sigma_\varepsilon^2 + n\sigma_\theta^2}\bar{s}_n, \quad \sigma_n^2 = \frac{\sigma_\varepsilon^2\sigma_\theta^2}{\sigma_\varepsilon^2 + n\sigma_\theta^2}, \tag{7}$$

where the history of signals $(s_1, s_{2,} \ldots, s_n)$ is summarized by age n and the average signal of the permanent demand draw:

$$\bar{s}_n \equiv \frac{1}{n}\sum_{i=1}^{n} s_i \;\text{ for } n \geq 1; \quad \bar{s}_0 \equiv \bar{\theta}.$$

For age-one firms (i.e., entrants), their belief for the mean and variance of θ is the same as the prior belief:

$$\mu_0 = \bar{\theta}, \;\; \sigma_0^2 = \sigma_\theta^2.$$

3.3 Static Optimization of Per-Period Profit

In this subsection, we study the firm's static optimization problem. As we focus on firms' behavior in the steady state (i.e., the stationary equilibrium) in what follows, we omit the subscript t whenever possible, and use age subscript n when necessary. In each period, the firm's output decision is a static choice. Given the belief about θ and φ_n, an age-n firm hires l_n workers to maximize its expected per-period profit, $E\left(\pi_n|\varphi_{n-1}, \bar{s}_{n-1}, n\right)$. The realized per-period profit is $\pi_n = p_n q_n - w l_n - w f$, where $q_n = \varphi_n l_n$ and firms set price p_n to clear the market according to equation (4). Maximizing $E\left(\pi_t|\varphi_{t-1}, \bar{s}_{n-1}, n\right)$, the optimal employment is

$$l_n = \left(\frac{\sigma - 1}{\sigma}\right)^{\sigma}\left(\frac{b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right)}{w}\right)^{\sigma} Y P^{\sigma-1}, \tag{8}$$

where

$$b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right) \equiv E\left(e^{\frac{\theta}{\sigma}} \varphi_n^{\frac{\sigma-1}{\sigma}} | \varphi_{n-1}, \bar{s}_{n-1}, n\right)$$
$$= \exp\left\{\frac{\mu_{n-1}}{\sigma} + \frac{\sigma_{n-1}^2}{2\sigma^2} + \frac{\sigma-1}{\sigma}\left((1-\rho)\mu_\varphi + \rho \log \varphi_{n-1}\right) + \frac{(\sigma-1)^2\sigma_{\nu_n}^2}{2\sigma^2}\right\}, \tag{9}$$

and n is the firm's age. The resulting price and expected per-period profit function are:

$$p_n = \left(YP^{\sigma-1}e^{\theta}\right)^{\frac{1}{\sigma}} q_n^{\frac{1}{\sigma}} = \frac{\sigma}{\sigma-1} e^{\frac{\theta}{\sigma}} \varphi_n^{-1/\sigma} \frac{w}{b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right)}; \tag{10}$$

$$E\pi_n = (\sigma-1)^{\sigma-1}\sigma^{-\sigma}YP^{\sigma-1}\frac{b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right)^{\sigma}}{w^{\sigma-1}} - wf. \tag{11}$$

3.4 Dynamic Optimization and Equilibrium Definition

In each period, the potential entrant chooses whether to enter the market and the incumbent firm chooses whether to stay in the market or exit. For an incumbent firm that is $n+1$ years old, its state variables include the labor productivity φ_n, the history of demand signals summarized by $\bar{s}_n$, and its age n in the last period. The incumbent firm's value function (after the random death shock is realized) satisfies:

$$V(\varphi_n, \bar{s}_n, n) = \max\{0, E_n\pi_{n+1} + \beta(1-\eta)E_nV(\varphi_{n+1}, \bar{s}_{n+1}, n+1)\}, \quad n \geq 1. \tag{12}$$

If the firm chooses to exit permanently, it receives a value of zero. For a potential entrant, its values function is:

$$V(\varphi_0, \bar{s}_0, 0) = \max\{0, E_0\pi_1 + \beta(1-\eta)E_0V(\varphi_1, \bar{s}_1, 1)\}, \tag{13}$$

as its productivity already evolves once when the entrant chooses to stay in the market and thus produce. In total, we have the value function that applies to both the potential entrant and the incumbent firm:

$$V(\varphi_n, \bar{s}_n, n) = \max\{0, E_n\pi_{n+1} + \beta(1-\eta)E_nV(\varphi_{n+1}, \bar{s}_{n+1}, n+1)\}, \quad n \geq 0. \tag{14}$$

We denote the corresponding policy function as $o(\varphi_n, \bar{s}_n, n)$, which can be either staying or exiting. The definition of equilibrium is contained in the Appendix.

4 Decomposing Forecast Errors

In this section, we show how our model matches Facts 1 and 2 presented in Section 2. As will be clear below, learning contributes to both (1) the age-declining variance of forecast errors, and (2) the age-declining covariance of forecast errors, while the age-dependent volatility only generates the former. This insight from our model allows us to decompose the variance of forecast errors into learning and age-dependent volatility components. We illustrate the intuitions by using a special case in which there is no endogenous entry and exit of firms.

Proposition 1 *When there is no endogenous entry and exit, the forecasts and forecast errors of firm sales have the following properties.*

1. *The variance of forecast errors declines with age.*

2. *Forecast errors made in two consecutive periods by the same firm are positively correlated. The positive covariance declines with age.*

3. *The difference between the variance of forecast errors (made at age n) and the autocovariance of forecast errors (made at age $n-1$ and n) has a one-to-one relationship with the (age-dependent) volatility of productivity shocks.*

Proof. See Appendix 7.1. ■

Both life cycle learning and age-dependent volatility contribute to the age-declining variance of forecast errors. Thanks to learning, firms accumulate more experience and thus have clearer information on their permanent demand when they become older, which makes the variance of forecast errors smaller. In addition, as we assume the variance of productivity shocks, σ_{ν_n}, declines with firm age, the variance of forecast errors also shrinks when the firm becomes older.

The above proposition also rationalizes the finding of the serially correlated forecast errors presented in Section 2.2, as firms adjust their posterior beliefs *gradually*. In other words, firms incorporate new signals partially into their posterior beliefs. As a result, the firm is more likely to under-predict (or over-predict) its next year's sales, if it has underpredicted (or overpredicted) its current year's sales. This leads to the positive autocorrelations of forecast errors.[18] Moreover, as a more experienced firm makes smaller forecast errors, the autocovariance of forecast errors declines with years of experience.

[18]However, this *does not* mean that firms make non-zero forecast errors on average, as positive and negative errors are averaged out across a large number of firms.

4.1 Nonparametric Decomposition

The above proposition illustrates how we can back out the learning parameters (σ_θ and σ_ε) and age-dependent volatility separately by using the panel data of forecast errors. To make the intuitions salient, we assume away endogenous exits. Under this assumption, the forecast errors of sales at age n is

$$FE_{n,n+1} \equiv \log \frac{R_{n+1}}{E_n R_{n+1}} = \underbrace{\frac{\theta}{\sigma} - \log E_n(e^{\frac{\theta}{\sigma}})}_{FE^{\theta}_{n,n+1}} + \underbrace{\frac{\sigma-1}{\sigma}\log \varphi_{n+1} - \log E_n(\varphi_{n+1}^{\frac{\sigma-1}{\sigma}})}_{FE^{\varphi}_{n,n+1}}, \tag{15}$$

where the first two terms, denoted by $FE^{\theta}_{n,n+1}$, represent the forecast errors that arise because of the firm's imperfect information about θ. The third and fourth terms, denoted by $FE^{\varphi}_{n,n+1}$, represent the forecast errors that come from the unpredictable innovation in the firm's AR(1) productivity process. As is shown in Appendix 7.1, the term $FE^{\varphi}_{n,n+1}$ is linear in the innovation term ν_{n+1}, which is uncorrelated with $FE^{\varphi}_{n-1,n}$ (linear in ν_n). By contrast, the term $FE^{\theta}_{n,n+1}$ is serially correlated since firms never observe θ and gradually update their belief about θ with noisy signals. The calculation shows that the covariance and variance of $FE_{n,n+1}$ are:

$$Cov(FE_{n-1,n}, FE_{n,n+1}) = \frac{\sigma_n^2}{\sigma^2}; \quad Var(FE_{n,n+1}) = \frac{\sigma_n^2}{\sigma^2} + \frac{(\sigma-1)^2\sigma_{\nu_n}^2}{\sigma^2}. \tag{16}$$

We can perform a non-parametric decomposition of $Var(FE_{n,n+1})$ into the learning component and the age-dependent volatility component, by using the two formulas together. Specifically, the covariance of forecast errors is only related to learning, as age-dependent volatility does not enter into the expressions. When we take the difference between the variance and the autocovariance of forecast errors, the only term that is left is the (age-dependent) variance of the firm's productivity shocks (multiplied by a constant):

$$Var(FE_{n,n+1}) - Cov(FE_{n-1,n}, FE_{n,n+1}) = \frac{(\sigma-1)^2\sigma_{\nu_n}^2}{\sigma^2}. \tag{17}$$

Note that our decomposition is "non-parametric" in the sense that we do not impose any structure on σ_{ν_n}.

Following this logic, we use Table 7 to implement the decomposition exercise. Columns (1) and (2) of the table are variance and covariance of forecast errors at age n in the data, while Column (5) is the difference between the two, capturing age-dependent volatility. In terms of levels, the learning component (covariance terms) are in general small, explaining about 10 to 20% of the variance of the forecast errors (Column 3, the ratio of Column 2 to Column 1). However, they have a larger contribution to the *change* in the variance of forecast errors over the firm's life cycle, ranging between 20 to 40% (Column 4). This is because the variance of shocks to labor productivity does not diminish to zero when firms are sufficiently old, which levels up the overall variance of forecast errors and makes the ratios in Column 3 small. Both learning and age-dependent volatility are important to account for the life cycle dynamics of firms' forecast errors.

Table 7: Non-parametric Decomposition of Learning and Age-Dependent Volatility Assuming Away Selection

Age n	(1) $Var(FE_{n,n+1})$	(2) $Cov(FE_{n-1,n}, FE_{n,n+1})$	(3) % Level	(4) $\%\Delta$ from $n=2$	(5) $\frac{(\sigma-1)^2\sigma^2_{\nu_{n+1}}}{\sigma^2}$
1	0.242				
2	0.174	0.034	19.8		0.139
3	0.135	0.019	14.5	38.3	0.115
4	0.110	0.020	18.5	22.0	0.089
5	0.098	0.013	12.9	28.8	0.086
6	0.097	0.014	14.5	26.5	0.083
7	0.088	0.014	16.0	23.7	0.074
8	0.087	0.008	9.1	30.5	0.079
9	0.081	0.009	10.9	27.6	0.072
10	0.069	0.008	11.9	25.0	0.061
11	0.069	0.008	11.3	25.4	0.061

Notes: Columns (1) and (2) report the variance and covariance of log forecast errors of firms at different ages in our data. Column (3) reports the ratio, $\frac{Cov(FE_{n-1,n}, FE_{n,n+1})}{Var(FE_{n,n+1})}$, in percentage terms. Column (4) reports the share contributed by the reduction in $Cov(FE_{n-1,n}, FE_{n,n+1})$ in the overall reduction in $Var(FE_{n,n+1})$. Mathematically, it equals $\frac{Cov(FE_{n-1,n}, FE_{n,n+1}) - Cov(FE_{1,2}, FE_{2,3})}{Var(FE_{n,n+1}) - Var(FE_{2,3})}$. Column (5) reports the difference between Columns (1) and (2). According to the equation (17), this term is driven by age-dependent volatility. Note that all these decomposition are under the assumption that there is no endogenous entry and exit.

5 Quantitative Analysis

In this section, we quantitatively assess the aggregate implications of imperfect information. We first describe the procedures used for calibrating our model and show the mapping from the model elements to the empirical facts presented in Section 2. We show that our calibrated model can capture the dynamics of firms' forecast errors, as well as other features of the data. Using similar procedures, we recalibrate some of the model parameters by different regions in the world, and evaluate the gains from a reduction in the information friction in each region.

5.1 Calibration

We first normalize a set of parameters not separately identified from others. Specifically, aggregate demand shifter, Y, and the wage rate, w, are normalized to one. The mean of the logarithm of the permanent demand, $\bar{\theta}$, is normalized to zero. Next, we set the elasticity of substitution between the varieties, σ, to four, a common value in the literature (see Bernard, Eaton, Jensen, and Kortum, 2003). We set the discount factor, β, to 0.96, which implies a real interest rate of 4% per annum. The exogenous death rate, η, is needed so that very large firms will still exit as observed in the data. We set it to 0.03, matching the exit rate of the largest 5% of firms above age ten. We impose an age threshold so that learning and age-dependent volatility are no longer important for these firms. Only extremely negative shocks to labor productivity and the exogenous death shock will induce exits (Table 8).

Table 8: Parameters Calibrated Without Solving the Model

Parameters	Description	Value	Source
σ	Elasticity of substitution between different varieties	4	Bernard et al. (2003)
β	Discount factor	0.96	4% real interest rate
η	Exogenous death rate	0.03	Exit rate of the largest 5% of firms above age ten

In our calibration, learning is parameterized by the two parameters, σ_θ and σ_ε. Guided by the decomposition exercise in section 4.1, two natural candidate moments are the covariance of FEs for the youngest firms and the oldest firms. Loosely speaking, conditional on other parameters, we calibrate σ_θ and σ_ε so that the model can match the autocovariance of FEs at ages of one and two, and the autocovariance of FEs above age ten.

Learning contributes the age-declining variance of FEs but only partially, as discussed in section 4.1. We let the age-dependent volatility reproduce the rest of the age-declining variance of FEs. We parameterize the age-dependent volatility using a quadratic function following Atkeson and Kehoe (2005)

$$\sigma_{\nu_n} = \begin{cases} \kappa_0 + \kappa_1 \left(\frac{10-n}{10}\right)^2 & \text{if } n < 10 \\ \kappa_0 & \text{if } n \geq 10. \end{cases}$$

Therefore, σ_{ν_n} starts from a value of $\kappa_0 + \kappa_1$ and drops to and stays at κ_0 after age ten. We calibrate the two parameters so that the model can match the variance of forecast errors above age ten and the variance of forecast errors at age one.[19]

[19]When mapping the model to the data, we use a mix of "age one" and "age two" firms to mimic "age one" firms in the data. In the model, "age one" firms have not received their first signal s_1 and are making their extensive margin and employment decisions based on their prior, while "age two" firms have received one signal. In the data, "age one" firms are defined as those established in any month of the current fiscal year. Therefore, firms that entered late in the fiscal year may have little information about θ and will make their prediction like an "age one" firm in the model, while other firms that entered early in the year may have received s_1 and behave like an "age two" firm in the model. We assume that any entrant has an equal chance to enter at the beginning or at the end of a fiscal year. When we match the variance and covariance

We are left with the choice of the remaining two other parameters: the per-period fixed cost, f, and the AR(1) coefficient of the labor productivity process, ρ. For the former, we target the average exit rate of incumbent firms. For the latter, we first compute the "adjusted labor productivity" as

$$\log \check{A}_n = \log R_n - \frac{\sigma-1}{\sigma}\log l_n = \frac{\theta}{\sigma} + \frac{\sigma-1}{\sigma}\log\varphi_n + \frac{1}{\sigma}\log(Y) + \frac{\sigma-1}{\sigma}\log(P),$$

where θ is firm-specific but time-invariant and Y and P are aggregate variables that do not vary across firms. The coefficient before $\log l_n$ is important – with this adjustment, the term related to expectation, $b(\varphi_{n-1}, \bar{s}_{n-1}, n-1)$, drops out from the labor productivity measure. We then use the following data moment

$$\frac{Var[\log(\check{A}_{n+1}/\check{A}_{n-1})]}{Var[\log(\check{A}_{n+1}/\check{A}_n)]} - 1, \quad n \geq 10 \tag{18}$$

to calibrate ρ.[20] Note that without selection, this formula provides an unbiased estimate for the persistence parameter in a stationary AR(1) process, even in small samples (Lo and MacKinlay, 1988). In our modified setting, taking the one- and two-period differences in $\check{A}_n$ removes the permanent demand shock θ. In addition, focusing on old firms ensures that σ_{ν_n} is constant and we can apply the same argument in Lo and MacKinlay (1988). Endogenous exits break the one-to-one mapping between this moment and ρ. However, we find that selection creates a very small bias and that this moment tightly pins down ρ.

Table 9: Parameters Calibrated by Solving the Model and Matching Moments

Parameters	Value	Description	Moments	Data	Model
f	0.0093	fixed cost	average exit rate of incumbents	0.093	0.093
σ_θ	0.96	std of θ	$Cov(FE_{t-1}, FE_t)$ at age one	0.034	0.034
σ_ε	1.36	std of ε	$Cov(FE_{t-1}, FE_t)$ above age ten	0.008	0.008
κ_0	0.33	$\sigma_{\nu_n} = \kappa_0 + \kappa_1(1-n/10)^2$	Var(FE) above age ten	0.069	0.069
κ_1	0.28	$\sigma_{\nu_n} = \kappa_0 + \kappa_1(1-n/10)^2$	Var(FE) above at age one	0.242	0.241
ρ	0.67	persistence in productivity	$\frac{Var[\log(\check{A}_{n+1}/\check{A}_{n-1})]}{Var[\log(\check{A}_{n+1}/\check{A}_n)]} - 1$	0.664	0.666

In Table 9, we list the parameters and moments in an order such that loosely, the moment provides the most information on the parameter in the same row. All moments are matched

of FEs of "age-one" firms in the data, the model counterpart is the same moments of a mix of "age one" and "age two" firms. The shares of "age one" and "age two" firms in this mix are close to 50% each, with the latter being slightly smaller due to the endogenous exit. We use the same strategy for other firm ages.

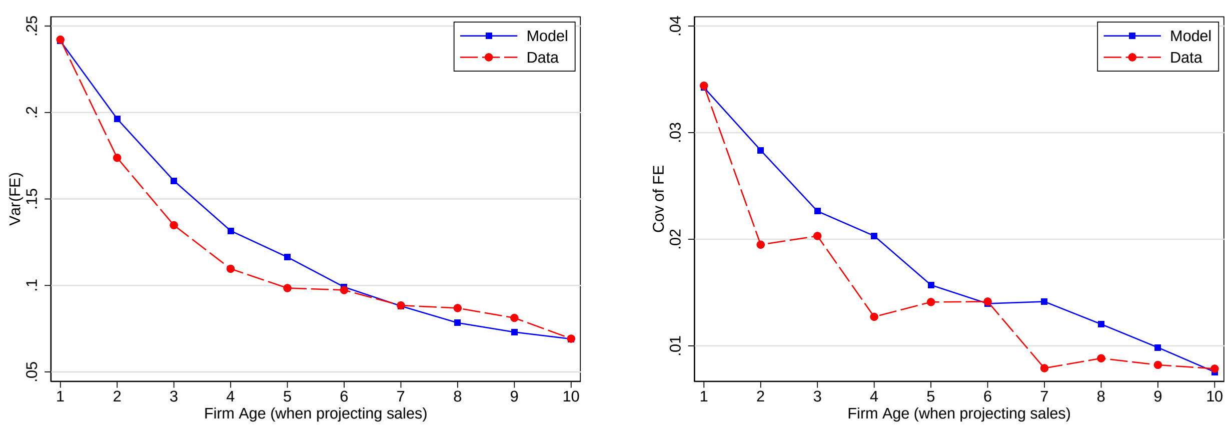
[20] Since P and Y do not vary across firms, they drop out from the variance. Moreover, θ also drops out from the difference in the logarithm of "adjusted labor productivity", as it is time-invariant.

precisely. The calibrated σ_θ and σ_ε are 0.96 and 1.36, respectively, implying a signal-to-noise ratio of 0.50. We find the value of ρ to be 0.67, very close to the data counterpart of equation (18).

5.2 Dynamics of Forecast Errors and Sales

In this section, we examine how our calibrated model performs regarding untargeted moments, focusing on moments of forecast errors and sales. In Figure 3, we plot the age profile of the variance and covariance of forecast errors. In the calibration, we match these moments for the youngest and oldest firms with two parameters related to learning and two parameters related to age-dependent volatility. The variance and covariance of forecast errors at other firm ages (between two and nine), though not directly targeted, track the data quite closely. Therefore, the parameterization does not cost us much in terms of matching the dynamics of forecast errors compared with the more flexible "non-parametric" decomposition in Table 7.

Figure 3: Moments of Forecast Errors, Model vs. Data



(a) $Var(FE_{n,n+1})$ (b) $Cov(FE_{n,n+1}, FE_{n+1,n+2})$

Next, we examine how our model performs regarding the age-declining AR(1) coefficients reported in Table 5. In particular, we simulate a 20-year panel of firms using the calibrated model parameters. We choose the number of simulated firms such that the simulated regressions have similar number of observations as our AR(1) regressions in Table 5. From this simulated data, we can calculate untargeted coefficients obtained from firm-level regressions. In Columns (1) and (2) of Table 10, we report coefficients of the AR(1) together with the impact of firm age (capped at age 10) in the data, with or without business group fixed effects. The model implied coefficients (Column 3) are consistent with the data: we see a positive autocorrelation overall, and a smaller coefficient when firms are older. Since we do not target these regression coefficients in our calibration, the overall AR(1) coefficient is a bit higher than that observed in the data, while the impact of age on the AR(1) coefficient is slightly smaller. However, the point estimate of the age effect in the simulated panel is within the 95% confidence interval of the estimates from the data reported in the first two columns.

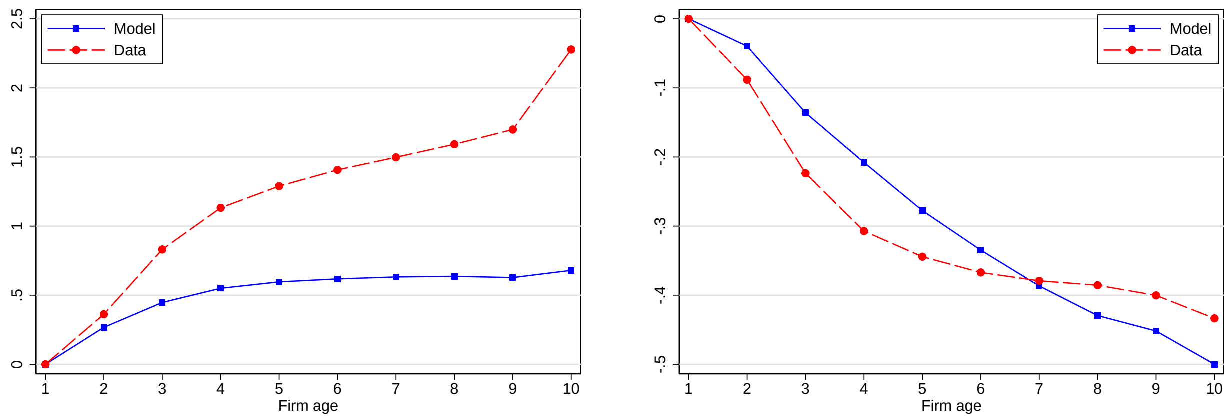
Table 10: AR(1) of Forecast Errors interacted with Age

	Dep Var: $FE^{\log}_{t+1,t+2}$		
	data		model
	(1)	(2)	(3)
$FE^{\log}_{t,t+1}$	0.141^a	0.102^a	0.168^a
	(0.013)	(0.014)	(0.006)
$FE^{\log}_{t,t+1} \times \text{Age}_t$	-0.005^a	-0.006^a	-0.003^a
	(0.002)	(0.002)	(0.001)
Year FE			Y
Industry-Year FE	Y	Y	
Country-year FE	Y	Y	
Business Group FE		Y	
Age FE	Y	Y	Y
N	96153	95196	94608
R^2	0.138	0.204	0.031

Notes: Standard errors are clustered at parent-firm level for columns (1) and (2) and at the firm level for column (3). Significance levels: c 0.10 b 0.05 a 0.01. Note that business group, country and industry fixed effects are not applicable to the simulated panel of firms. In the simulated regression, we control for year and age fixed effects to best mimic columns (1) and (2).

In Figure 4, we examine the model's performance in terms of moments related to firm sales, which are not directly targeted in our calibration. Panel (a) plot the average log sales of firms of different ages. There is growth in average firm sales over their life cycles both in the model and in the data. However, the growth rate tends to decline as firms become older. The main difference between the data and the model is that average firm size still grows after age ten in the data but not too much in the model. The decline in the rate of firm growth is a key feature of learning models, which has been used to estimate the learning parameters in Arkolakis et al. (2018). Our model implies slower growth and a quicker diminish of the growth (over the firm's life cycle) than the data. This is expected, as we do not target these moments in the calibration, and there are other mechanisms that explain firm growth (e.g., the accumulation of customer capital as in Foster et al. (2016)). Regarding second moments, our model successfully generates the decline in the standard deviation of the sales growth rates observed in the data, as reported in Panel (b) of the figure.

Figure 4: Moments of Sales, Model vs. Data



(a) Average Log(Sales)

(b) Change in SD(Δlog(Sales))

5.3 Aggregate Implications

In this subsection, we study the effect of information frictions on aggregate outcomes such as allocative efficiency and productivity. We first set the parameter values as described above and study comparative statistics of varying σ_ε, the parameter that governs the noisiness of the signals that firms receive. This highlights various channels that operate through intensive and extensive margins, affecting allocative efficiency and productivity in the economy, and we find that endogenous firm entry and exit (the extensive margin) play a quantitatively important role in generating large productivity gains from eliminating informational frictions from the economy. Finally, we demonstrate this in our cross-regional analysis where we apply our calibration approach to infer the learning parameters as well as other key parameters governing firm dynamics in each region. Our experiment confirms not only the degree of information frictions, but also the extent to which entries and exits of firms matter for the size of productivity gains from removing information frictions.

5.3.1 Comparative statistic: intensive and extensive margins

We first consider a change in the information environment by changing the value of σ_ε with other parameters held fixed at the values described above. Our baseline σ_ε is 1.36, and we vary it between 0.10 and 2.50, the highest value corresponding to the region with the highest σ_ε as we show in our by-region calibration in Section 5.3.3. We also consider a case where information about θ is perfect in that entrants know the true value of θ.

Figure 5 plots the impact of information frictions on aggregate outcomes. We compare our baseline model to a version of our model without endogenous entry and exit. In Figure

5, the blue curves with dots summarize the comparative statistics with respect to σ_ε in the baseline model. The red curves with square markers indicate the same comparative statistics with respect to σ_ε in the model where we set the per-period fixed costs f to zero. In both models, price index increases with σ_ε (top left panel), while labor productivity decreases with σ_ε (top right panel), with the slope being steeper in the baseline model.

Table 11: Aggregate Outcomes under Different σ_ε

Panel A: $f = 0.0093$ (benchmark)	(1)	(2)	(3)
Statistics	High Info. Friction $\sigma_\varepsilon = 2.50$	Baseline Info. Friction $\sigma_\varepsilon = 1.36$	Perfect Info.
Mass of Active Firms	11.224	10.359	9.046
Incumbents Average θ	0.591	0.764	1.046
Incumbents Average $\theta + (\sigma - 1)\log\varphi$	0.187	0.231	0.315
Q/L	3.482	3.623	3.853
Δ% Q/L	-3.88		6.36
Panel B: $f = 0$	(1)	(2)	(3)
Statistics	High Info. Friction $\sigma_\varepsilon = 2.50$	Baseline Info. Friction $\sigma_\varepsilon = 1.36$	Perfect Info.
Mass of Active Firms	32.333	32.333	32.333
Incumbents Average θ	0	0	0
Incumbents Average $\theta + (\sigma - 1)\log\varphi$	0	0	0
Q/L	4.528	4.624	4.794
Δ% Q/L	-2.08		3.66

Notes: This table reports equilbrium outcomes under a high level of information frictions ($\sigma_\varepsilon = 2.50$), baseline model ($\sigma_\varepsilon = 1.36$) and perfect information, with different values of fixed costs (baseline value, 0.0093, and alternative value, 0). As is explained in footnote 21, the term $\theta + (\sigma - 1)\log\varphi$ can be interpreted as "firm capability", which uniquely determines a firm's size in a perfect information static model.

These productivity losses from informational imperfection stem from the effects that operate through both intensive and extensive margins. For the intensive margin, it is shown by that the correlation between firm capability ($\log\phi \equiv (\sigma-1)\log\varphi+\theta$) and production scale ($\log b$) decreases with σ_ε (middle left panel).[21] This is because more severe informational imperfection tends to make firms with low demand θ produce too much, and vice versa for firms with high demand. Imprecise knowledge about demand θ makes output choice far from the optimal level at the intensive margin, which can also be seen in that the correlation between the true demand θ and the average of past noisy signals $\bar{s}$ decreases with σ_ε (middle right panel).

For the extensive margin, it can be seen that the average value of demand shifters across firms decreases with σ_ε (bottom left panel). More severe informational imperfection tends to make firms with low demand draws but high value for the noise term enter and stay

[21] Firm capability term $\log\phi$ is a combination of firm labor productivity φ and its permanent demand shifter θ. Scaling $\log\varphi$ by the coefficient $\sigma - 1$ ensures that this term solely determines firm-level output in a perfect information static model. In our dynamic imperfect information model, b is the only firm-level variable that determines expected profit (see equations (9) and (11)). In a perfect information static model, $\log b$ is linear in $\log\phi$ thus the correlation is one.

and to make firms with high demand draws but low value for the noise term exit, a similar mechanism studied in the context of selection into exporting in Sager and Timoshenko (2021). A less stringent selection under more severe informational imperfection can also be seen from the result that the mass of active firms increases with σ_ε (bottom right panel).[22] These effects from selection do not show up without endogenous entry and exit of firms, as depicted by the red curves with square markers in the bottom panels. In this alternative model, all potential entrants (other than those that exit exogenously) are active in production. Thus, the information friction affects the price index and labor productivity only through the intensive margin.

Table 11 shows the quantitative implications and highlight the role of selection. Labor productivity increases by 6.35 % in our baseline model with endogenous entry and exit, while it increases by 3.66 % in the alternative model where the extensive margin does not play a role. Our comparative statistics show not only a substantial gain in overall productivity with eliminating informational frictions, but also the role of firm entry and exit in driving it.

5.3.2 Heterogeneous effects across different age groups of firms

One feature of our model is the gradual resolution of uncertainty over the life cycle of firms. Entrants and young firms face more severe informational imperfection and learn their true values of demand shifter over time, while deciding in each period whether to stay or exit from the market. We proceed with the analysis to see how firms in different age groups are affected differently by the elimination of the information friction and how much each age group's productivity change contributes to the overall productivity gains in the economy.

First, we define the average productivity of age-n ($n \geq 1$) firms as

$$A_n \equiv \frac{Q_n}{L_n^{prod}} = \frac{\left(\int_{\omega\in\Omega_n} e^{\frac{\theta(\omega)}{\sigma}} q_n(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right)^{\frac{\sigma}{\sigma-1}}}{L_n^{prod}}, \tag{19}$$

where L_n^{prod} is the number of workers used in production of all age-n firms, excluding workers used to pay for the fixed cost. Ω_n is the set of active age-n firms and $q_n(\omega)$ is the output of the firm that produces variety ω. Note that entrants are age-one firms, while incumbents are older than one. According to the definition of average productivity of firms of all ages,

[22] Note that the mass of potential entrants is fixed in the model.

we can write

$$A \equiv \frac{\left(\sum_n \int_{\omega\in\Omega_n} e^{\frac{\theta(\omega)}{\sigma}} q_n(\omega)^{\frac{\sigma-1}{\sigma}} d\omega\right)^{\frac{\sigma}{\sigma-1}}}{L^{prod}}$$
$$= \left(\sum_{n=1}^{N} A_n^{\frac{\sigma-1}{\sigma}} \left(\frac{L_n^{prod}}{L^{prod}}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} = \left(\sum_{n=1}^{N} A_n^{\frac{\sigma-1}{\sigma}} \left(\frac{\bar{L}_n^{prod} M_n}{\bar{L}^{prod}\sum_{n=1}^{N} M_n}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{20}$$

where $L^{prod} = \sum_{n=1}^{N} L_i^{prod}$, and N is the maximum age that we consider in the simulation. In addition, $\bar{L}_n^{prod}$ is the average employment of production workers of age-n firms, and $\bar{L}^{prod}$ is the average employment of production workers of all firms. $M_n \equiv \int_{\omega\in\Omega_n} d\omega$ is the measure of age-n firms that are active. We then define the normalized productivity $\tilde{A}_n = A_n M_n^{\frac{1}{1-\sigma}}$. Note that the difference between $\tilde{A}_n$ and A_n is that the former does not take into account the variety effect, reflected by the number of active firms in our model.

Finally, the log (or percentage) change in average labor productivity can be decomposed as

$$\frac{dA}{A} = \sum_{n=1}^{N}\left[contri_n\left(d\log(\tilde{A}_n) + \frac{\sigma}{\sigma-1}\frac{dfrac_n}{frac_n} + d\log\left(\frac{\bar{L}_n^{prod}}{\bar{L}^{prod}}\right)\right)\right] + \frac{1}{\sigma-1}\frac{dM}{M}, \tag{21}$$

where the weight is defined as $contri_n \equiv frac_n \tilde{A}_n^{\frac{\sigma-1}{\sigma}} \left(\frac{\bar{L}_n^{prod}}{\bar{L}^{prod}}\right)^{\frac{\sigma-1}{\sigma}} / \sum_{n=1}^{N} frac_n \tilde{A}_n^{\frac{\sigma-1}{\sigma}} \left(\frac{\bar{L}_n^{prod}}{\bar{L}^{prod}}\right)^{\frac{\sigma-1}{\sigma}}$ and $frac_n$ is the fraction of active firms that are n years old among all active firms. The total mass of active firms is simply denoted by $M = \sum_{n=1}^{N} M_n$.

There are four terms related to the change in average productivity in equation (21). First, term $d\log(\tilde{A}_n)$ is the change in normalized productivity for each age group. Second, $\frac{\sigma}{\sigma-1}\frac{dfrac_n}{frac_n}$ reflects the change in population shares for different age groups. Third, $d\log\left(\frac{\bar{L}_n^{prod}}{\bar{L}^{prod}}\right)$ is the change of the average size of age-n firms (relative to the overall mean). The final term, $\frac{1}{\sigma-1}\frac{dM}{M}$, reflects the variety effect. Figure 6 plots these terms, when we move from our baseline level of imperfect information (with $\sigma_\varepsilon = 1.36$) to perfect information wherein all entrants know the true value of θ. In Figure 6, the blue curves with dots show the results for our baseline model where firms can endogenously enter and exit, while the red curves with square markers indicate the results for an alternative model without such endogenous entry and exit, by setting the per-period fixed costs f to zero.

Panel (a) illustrates that the group-specific productivity increases for all groups, but the gains are larger among young firms than among old firms. We also can see that these disproportionately larger gains among young firms are more pronounced in our baseline model with endogenous entry and exit (blue curves with dots). As shown in Panel (b), the population shares of young firms drop significantly, while old firms above age nine increase their population shares in our baseline model with endogenous entry and exit, while this mechanism is absent in the alternative model (red curves with square markers). Selection gets tougher when the information friction becomes less severe, and this leads to a "better" selected group of firms operating in the economy. We discussed this in the previous section but we now see this extensive margin effect operates more prominently among young firms, highlighting the importance of post-entry selection especially among young incumbent firms. Relatedly, the average relative size of firms measured by employment increases among young firms but decreases among old firms, as shown by Panel (c). In addition, the same disparity between young and old firms caused by the existence of the extensive margin can be seen by comparing the two curves in Panel (c). Finally, from Panel (d) we can see that the mass of firms declines for each age group after σ_ε declines in our baseline model with endogenous entry and exit.[23]

5.3.3 Cross-regional analysis

This subsection presents our cross-regional analysis. As suggested by Fact 3 in Section 2.2, firms may face different levels of informational imperfection in different regions due to communication barriers and differences in capabilities of local managers and/or in quality of management practices. But empirically pinpointing the drivers is beyond the scope of our paper and thus we are still left to speculate on what form of policies would be ideal for countries where firms suffer informational frictions. Instead, this subsection uses our model and calibration approach described above to explore the degree of informational imperfection and the potential productivity gains from eliminating information frictions across regions. As demonstrated in Subsection 5.3.1, the potential gains from eliminating information frictions in the model depends not only on the degree of informational imperfection, but also parameters such as the fixed operating costs. Thus, our quantitative investigation takes into account cross-regional heterogeneity in these parameters and sets their values to be consistent with the corresponding moments.

We exploit cross-regional differences in the parameters governing learning (σ_θ and σ_ε), age-dependent volatility of labor productivity (κ_0 and κ_1), and the fixed cost (f). We use data from eight major regions/countries of the world: Africa, Middle East, Latin America, Eastern Europe, ASEAN countries, China, Western Europe, and the United States. Similar to the baseline calibration, we target the covariance and variance of the youngest and oldest firms, together with the incumbent exit rates in each region. These eight regions do not

[23]This panel shows the result only for the baseline model with endogenous entry and exit because the mass of firms is fixed in the alternative model without endogenous entry and exit.

exhaust all foreign countries that Japanese multinational firms operate in, but they cover the majority of firm sales in our data. In addition, they also display significant differences in the income levels and business environments, and contain countries that are relatively homogeneous within each region. Online Appendix Table OA.17 provides the full list of countries in each region.

Panel A of Table 12 presents the calibrated parameters by region and the corresponding model moments. Each set of parameters enables us to precisely match the data moments, so we omit them from the table to save space and report them in Online Appendix OA.18. We find that Africa, Middle East, Latin America, and Eastern Europe have higher values of σ_ε than the other regions, which are driven by their higher covariance of forecast errors targeted in the calibration. Firms in Latin America and Eastern Europe have higher values of σ_θ than the other regions. There is also some variation in the calibrated age-dependent volatility. One thing to notice is that the high value of σ_θ calibrated for Latin America pushes its age-dependent volatility parameter κ_1 to zero, as learning combined with the oldest firms' volatility are sufficient to generate the observed variance of FEs of the youngest firms in this region. In contrast, while the variance of forecast errors is high in ASEAN, the covariance of forecast errors is low relative to other regions, leading to the age-dependent volatility picking up most of the dynamics of forecast errors with the implied size of the learning parameters being small.

Our calibration also reveals large differences in the per-period fixed cost across regions. For example, firms in the United States, Western Europe, Africa and Middle East face the highest fixed costs, which are associated with the high exit rates in these regions. While the values of the learning parameters in China, σ_θ and σ_ε, are similar to those in the United States and Western Europe, the exit rates of Japanese firms in China are much lower, which implies a lower fixed cost f. In contrast, firms in Latin America and Eastern Europe face much higher levels of uncertainty and information frictions due to higher values of σ_θ and σ_ε. To match similar exit rates as in the United States and Western Europe, the model suggests that the fixed costs of operating in these countries are also low.

Equipped with the calibrated economies by region, we then assess the productivity gain from eliminating informational imperfection. Moving from the calibrated economy to perfect information, $\sigma_\varepsilon = 0$, we report the increase in labor productivity in percentage terms in the last column of Panel A of Table 12. ASEAN, Western Europe and the United States have the lowest gains in labor productivity, 3.54%, 6.95% and 7.11%, respectively. The gains from removing informational imperfection are only slightly larger in Latin America and China. Firms in Africa, Middle East, and Eastern Europe have the highest gains from eliminating the information friction, 12.56%, 12.16%, and 9.45%, respectively.

Table 12: Calibration and Counterfactuals by Region

Region	Parameters						Model Moments					Info. Gains
	σ_θ	σ_ε	$\sigma^2_\theta/\sigma^2_\varepsilon$	σ_{ν_1}	$\sigma_{\nu_{10}}$	f	Cov_1	Cov_{10}	Var_1	Var_{10}	exit rate	%Δ Q/L
Panel A: change five parameters												
Africa	0.86	2.57	0.11	0.51	0.37	0.0152	0.040	0.020	0.186	0.100	0.105	12.56
Middle East	0.83	2.64	0.10	0.58	0.45	0.0142	0.038	0.019	0.226	0.134	0.102	12.16
Eastern Europe	1.41	1.80	0.62	0.58	0.32	0.0079	0.068	0.014	0.283	0.072	0.101	9.44
Latin America	1.62	1.66	0.95	0.39	0.39	0.0070	0.073	0.013	0.218	0.097	0.103	7.13
ASEAN	0.44	1.50	0.09	0.70	0.34	0.0075	0.008	0.006	0.264	0.072	0.078	3.54
China	1.12	1.48	0.57	0.64	0.31	0.0074	0.044	0.010	0.276	0.065	0.089	7.16
Western Europe	0.91	1.47	0.39	0.50	0.31	0.0131	0.034	0.009	0.179	0.065	0.106	6.95
United States	0.78	1.49	0.27	0.52	0.31	0.0147	0.028	0.009	0.180	0.063	0.110	7.11
Panel B: change four parameters, fix f												
Africa	0.86	2.57	0.11	0.51	0.37	0.0093	0.039	0.020	0.184	0.098	0.073	9.77
Middle East	0.83	2.64	0.10	0.58	0.45	0.0093	0.037	0.018	0.225	0.130	0.074	9.83
Eastern Europe	1.41	1.80	0.62	0.58	0.32	0.0093	0.068	0.015	0.281	0.072	0.112	9.86
Latin America	1.62	1.66	0.95	0.39	0.39	0.0093	0.071	0.014	0.219	0.098	0.125	7.84
ASEAN	0.44	1.50	0.09	0.70	0.34	0.0093	0.011	0.006	0.268	0.072	0.102	3.87
China	1.12	1.48	0.57	0.64	0.31	0.0093	0.042	0.010	0.280	0.065	0.104	7.70
Western Europe	0.91	1.47	0.39	0.50	0.31	0.0093	0.034	0.010	0.177	0.066	0.081	5.93
United States	0.78	1.49	0.27	0.52	0.31	0.0093	0.029	0.008	0.177	0.063	0.077	5.55

Notes: Panel (A) shows the results when we re-calibrate five parameters for each region $(\sigma_\theta, \sigma_\varepsilon, \kappa_1, \kappa_0, f)$. We present age-dependent volatility $\sigma_{\nu_1}, \sigma_{\nu_{10}}$ instead of κ_1, κ_0 to facilitate interpretation. We target five moments in this calibration, $Cov(FE_{n-1,n}, FE_{n,n+1})$ for $n = 1$ and $n \geq 10$, $Var(FE_{n,n+1})$ for $n = 1$ and $n \geq 10$ and incumbent exit rates, respectively. $\%\Delta Q/L$ is the percentage change in labor productivity when we change the model from the calibrated imperfect information case to perfect information. Panel (B) reports the results when we re-calibrate the learning and uncertainty related parameters but keep the fixed costs at the baseline value $f = 0.0093$. We target the first four moments but do not attempt to match the exit rates in the data. The model matches the data moments well (other than the untargeted exit rates in Panel B). To save space, we report the data moments in Online Appendix Table OA.18. A full list of countries in each region can be found in Online Appendix Table OA.17.

In general, regions with larger σ_ε and σ_θ tend to have larger gains from moving toward perfect information. It is clear that higher σ_ε leads to noisier signals and potentially more misallocation at both the intensive and extensive margins. A higher value of σ_θ, on the other hand, increases the benefit of eliminating the information friction, as there is much more to learn over the life cycle.[24] For instance, if we just focus on the learning parameters $\sigma_\theta, \sigma_\varepsilon$, our model would imply that the gains from eliminating informational imperfection should be higher in China and Latin America than Western Europe and the United States. However, when disciplined by firms' exit rates in different regions, we find that the fixed costs in Latin America and China are much smaller than those in Western Europe and the United States. This in turn reduces the gains from eliminating informational imperfection at the extensive margin, and thus dampens the difference between these regions. Indeed, as is reported in Panel B of Table 12, when we keep the fixed cost at the baseline level for all regions, the gains from eliminating informational imperfection in Latin America and

[24]Firms in Latin America and Eastern Europe have higher values of σ_θ than the other regions, and their signal-to-noise ratios are the highest among the eight regions. This is broadly consistent with a view that firms acquire information optimally by paying a cost, which makes σ_ε (or equivalently, the signal-to-noise ratio, $\sigma^2_\theta/\sigma^2_\varepsilon$) endogenous to the level of σ_θ (see Sims (2003); Luo (2008); Mackowiak and Wiederholt (2009)).

China become about 1.77% and 2.26% higher than those numbers in Western Europe and the United States, instead of being 0.01% and 0.32% higher. Another impact of keeping the fixed cost constant across the regions is that the gain from eliminating informational imperfection for Africa and the Middle East becomes smaller and closer to that for Eastern Europe, as the calibrated fixed cost is the highest in Africa among all regions.

These calibrated parameters and the associated productivity gains are based on the sample of Japanese firms, and therefore we need caution before generalizing our results. On the one hand, one could argue that local firms may face with less severe information frictions than foreign owned firms because of possible barriers such as time zone differences that make intrafirm communication with parents firms difficult. On the other hand, one could also argue that foreign owned firms are able to make more accurate forecasts due to their superior management practices, as extensively documented in the literature (e.g. Bloom and Van Reenen (2007) and Bloom and Van Reenen (2010b)). Our view is that the derived results from our quantitative exercises is broadly consistent with that firms in regions/countries where better managers/management practices are available face with less severe informational constraints, while factors that make intrafirm communication difficult such as large time differences lead to more severe informational constraints. The result that firms gains more in Africa, Middle East, and Eastern Europe than in China, ASEAN, Western Europe and the United States appears to be consistent with the former argument. In contrast, the relatively less severe imperfect information and the small associated productivity losses in ASEAN and China are consistent with the latter argument. We believe that our cross-region analysis is a step ahead of the literature on firm-level information imperfection and misallocation, which usually focuses on a small number of countries; however, additional empirical investigation would be useful to provide more concrete causal evidence.

5.3.4 Sensitivity analysis

In this subsection, we compare the preceding results obtained in an industry equilibrium model where total expenditure and wage rates are exogenous to those under general equilibrium. To do so, we add two more conditions: (1) total expenditure by the consumers that equals their labor income plus aggregate firm profits, and (2) total labor demand that equals total (inelastic) labor supply. Columns under "Fixed J" in Table 13 present the similar comparative statics as those in Table 11, now under general equilibrium. We find similar quantitative predictions: increasing σ_ε from the baseline value to 2.50 lowers the aggregate labor productivity by 4.0% and moving toward perfect information increases it by 5.9%.

The columns under "free entry" in Table 13 go a step further and allow the mass of potential entrants, J, to be determined by a zero net profit condition. Instead of assuming that potential entrants can draw their initial productivity φ_0 without any cost, we assume that they have to pay an entry cost f_e in order to make such draws. We set this entry cost to the expected net profit of entrants in our baseline equilibrium. This entry cost will ensure that $J = 1$ is consistent with a free-entry, industry equilibrium model. As is shown in the "free entry" columns in Table 13, the equilibrium mass of potential entrants is close to one, which makes the free-entry general equilibrium model comparable with our baseline model with a fixed J.[25] The loss of productivity from varying σ_ε from 1.36 (the baseline value) to 2.50 reduces productivity by 4.5%, while moving toward perfect information raises productivity by 7.5%. Our message from these two general equilibrium settings is that the impact of eliminating the information friction on aggregate labor productivity is not sensitive to our assumption that wages, aggregate expenditures, and the mass of potential entrants are exogenous.

Table 13: The Impact of σ_ε Under General Equilibrium

	Fixed J			Free entry		
	High Info. Friction $\sigma_\varepsilon = 2.50$	Baseline $\sigma_\varepsilon = 1.36$	Perfect Info.	High Info. Friction $\sigma_\varepsilon = 2.50$	Baseline $\sigma_\varepsilon = 1.36$	Perfect Info.
J	1.000	1.000	1.000	1.043	1.072	1.245
aggregate profits	0.174	0.180	0.203	0.053	0.059	0.057
Mass of Active	12.616	11.858	10.439	11.908	11.115	9.741
P	0.329	0.319	0.306	0.330	0.317	0.295
Emp	1.000	1.000	1.000	1.000	1.000	1.000
Mean θ	0.563	0.729	0.974	0.590	0.774	1.106
Mean φ	0.065	0.066	0.092	0.074	0.078	0.126
Mean ϕ	0.173	0.215	0.290	0.184	0.232	0.340
Labor Prod	3.258	3.395	3.594	3.187	3.335	3.584
Δ% Labor Prod	-4.03		5.88	-4.45		7.47

Notes: The first three columns show parameters and equilibrium outcomes in an alternative general equilibrium model where the potential mass of entrants J is fixed at one while total expenditure equals labor income plus aggregate firm profits (and total labor equala total inelastic labor supply which is normalized at one). The last three columns show results from a model where we further allow J to be endogenous and determined by a free-entry condition. We set $\sigma_\varepsilon = 2.5$ in the "High Info. Friction" case.

6 Conclusion

In this paper, we use firm-level panel data on sales forecasts to directly detect imperfect information and learning over the life cycle of firms. We provide novel evidence of imperfect information faced by firms and its gradual resolution: the variance of forecast errors declines with firms' experience and the covariance of forecast errors is positive and declines with firms' experience. We then develop a quantitative model of heterogeneous firms that learn about

[25]Note that the equilibrium J is not exactly one since we are now considering a general equilibrium model instead of an industry equilibrium model. We choose f_e so that J is exactly one in an industry equilibrium model, consistent with our baseline.

their demand over the life cycle and show how it can be used to decompose the variance of forecast errors observed in data into learning and any other components. The contribution of learning to the change in the variance of forecast errors over the firm's life cycle ranges between 20% to 40%. We also demonstrate how to use firm-level data on forecast errors with a clean mapping from data to the learning parameters in the model. We calibrate the model to quantify cross-region/country differences in the degree of informational imperfections and the potential gains from eliminating informational imperfections. We find the prominent role of entry and exit of firms in driving the potential gains from eliminating informational imperfections, ranging from 3.5% in ASEAN countries to 12.5% in Africa.

7 Appendix

7.1 Proof for Proposition 1

We derive several expressions concerning forecast errors first. The firm's revenue can be expressed as

$$\begin{aligned} R_n = p_n q_n &= \left(YP^{\sigma-1}e^{\theta}\right)^{1/\sigma} q_n^{1-1/\sigma} \\ &= \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} YP^{\sigma-1}\left(\frac{b(\varphi_{n-1},\bar{s}_{n-1},n-1)}{w}\right)^{\sigma-1} e^{\theta/\sigma}\varphi_n^{1-1/\sigma}. \end{aligned}$$

Therefore, (log) forecast error of sales are

$$\begin{aligned} FE^{log}_{n,n+1} \equiv \log R_{n+1} - \log E_n R_{n+1} &= \frac{\theta}{\sigma} + \frac{\sigma-1}{\sigma}\log\varphi_{n+1} - \log E_n(e^{\theta/\sigma}\varphi_{n+1}^{\frac{\sigma-1}{\sigma}}) \\ &= \underbrace{\frac{\theta}{\sigma} - \log E_n(e^{\theta/\sigma})}_{FE^{\theta}_{n,n+1}} + \underbrace{\frac{\sigma-1}{\sigma}\log\varphi_{n+1} - \log E_n(\varphi_{n+1}^{\frac{\sigma-1}{\sigma}})}_{FE^{\varphi}_{n,n+1}} \\ &= \frac{\theta-\mu_n}{\sigma} - \frac{\sigma_n^2}{2\sigma^2} + \frac{(\sigma-1)\nu_{n+1}}{\sigma} - \frac{(\sigma-1)^2\sigma^2_{\nu_{n+1}}}{2\sigma^2}. \end{aligned} \tag{22}$$

From equation (22), it is straightforward to show that, without selection on θ,

$$Cov(FE^{log}_{n-1,n}, FE^{log}_{n,n+1}) = \frac{\sigma_n^2}{\sigma^2} = \frac{\sigma_\theta^2\sigma_\varepsilon^2}{(\sigma_\varepsilon^2 + n\sigma_\theta^2)\sigma^2}, \quad Var(FE^{log}_{n,n+1}) = \frac{\sigma_n^2 + (\sigma-1)^2\sigma^2_{\nu_{n+1}}}{\sigma^2}.$$

For the first part of the proposition, we can rewrite equation (22) as

$$FE^{\log}_{n-1,n} = \frac{(1-\zeta(n-1,\lambda))(\theta-\bar{\theta}) - \zeta(n-1,\lambda)\frac{\sum_{i=1}^{n-1}\varepsilon_i}{n-1}}{\sigma} - \frac{\sigma^2_{n-1}}{2\sigma^2} + \frac{(\sigma-1)\nu_n}{\sigma} - \frac{(\sigma-1)^2\sigma^2_{\nu_n}}{2\sigma^2}, \tag{23}$$

where

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

Note that λ defined above is the signal-to-noise ratio. Based on equation (23), we calculate the variance of forecast error as

$$\begin{aligned} Var(FE^{\log}_{n-1,n}) &= \frac{\zeta(n-1,\lambda)^2\sigma_\varepsilon^2}{(n-1)\sigma^2} + \frac{(1-\zeta(n-1,\lambda))^2\lambda\sigma_\varepsilon^2}{\sigma^2} \\ &= \frac{\sigma_\varepsilon^2}{\sigma^2}\left(\frac{\lambda}{1+(n-1)\lambda}\right) + \frac{(\sigma-1)^2\sigma^2_{\nu_n}}{\sigma^2}. \end{aligned} \tag{24}$$

One can see that the variance of forecast errors declines with n, as both the first and the second terms decrease with n.

For the second part of the proposition, one can calculate that

$$Cov(FE^{log}_{n-1,n}, FE^{log}_{n,n+1}) = \frac{\sigma_n^2}{\sigma^2} = \frac{\sigma_\theta^2 \sigma_\varepsilon^2}{(\sigma_\varepsilon^2 + n\sigma_\theta^2)\sigma^2} > 0,$$

as long as we have random informational shocks, ε_i (i.e., $\sigma_\varepsilon^2 > 0$). This means as long as we have random ε_i, the forecast errors in two consecutive periods are positively correlated.

Finally, it is straightforward to observe and calculate that

$$Var(FE^{log}_{n,n+1}) - Cov(FE^{log}_{n-1,n}, FE_{n,n+1}) = \frac{(\sigma-1)^2 \sigma_{\nu_n}^2}{\sigma^2},$$

which means that the difference between the variance of forecast errors (made at age n) and the autocovariance of forecast errors (made at age $n-1$ and age n) has a one-to-one relationship with the (age-dependent) volatility of productivity shocks.

7.2 Definition of Equilibrium

Definition 1 *A steady-state equilibrium of the model is defined as follows:*

1. *policy functions of optimal employment $l(\varphi_{n-1}, \bar{s}_{n-1}, n-1)$ that maximizes the per-period profit function as in equation (8);*

2. *firms' prices in the current period $p(\theta, \varphi_n, b(\varphi_{n-1}, \bar{s}_{n-1}, n-1))$ that clear the market, i.e., equation (10);*

3. *value functions, $V(\varphi_{n-1}, \bar{s}_{n-1}, n-1)$, and policy functions $o(\varphi_{n-1}, \bar{s}_{n-1}, n-1)$, of whether to stay (= 1) or exit (= 0) , that are consistent with equation (14);*

4. *a measure function of firms $\lambda(\varphi_{n-1}, \bar{s}_{n-1}, n-1, \theta)$ that is consistent with the aggregate law of motion. This measure function of firms is defined at the beginning of each period (i.e., after the exogenous exit takes place but before the endogenous mode switching happens). In particular, in each period, an exogenous mass J of entrants draw θ and φ_0 from the corresponding distributions. Therefore, the measure of entrants with state variables (φ_0, θ) is*

   $$\lambda(d\varphi_0, \bar{s}_0, 0, d\theta) = (1-\eta) J g_\theta(\theta)\, d\theta \times g_{\varphi_0}(\varphi_0)\, d\varphi_0,$$

   where $g_\theta(\cdot)$ and $g_{\varphi_0}(\cdot)$ are the density functions of the distributions for θ and φ_0, *respectively. The measure function for incumbent firms should be a fixed point of the aggregate law of motion, i.e., given any Borel set of $\bar{s}_n$, Δ_s, and any Borel set of φ_n, Δ_φ, measures of firms with $n \geq 2$ satisfy*

$$\lambda\left(\Delta_\varphi, \Delta_s, n, \theta\right) = \int_{\varphi_{n-1}, \bar{s}_{n-1}, \theta} \begin{array}{c} \mathbf{1}\left(\bar{s}_n \in \Delta_s, \varphi_n \in \Delta_\varphi\right) \times o(\varphi_{n-1}, \bar{s}_{n-1}, n-1)\times \\ (1-\eta)\Pr\left(\bar{s}_n | \bar{s}_{n-1}, \theta\right)\Pr\left(\varphi_n|\varphi_{n-1}\right)\lambda\left(d\varphi_{n-1}, d\bar{s}_{n-1}, n-1, d\theta\right) \end{array}.$$

5. *the price index P is constant over time and must be consistent with consumer optimization (5):*

$$P^{1-\sigma} = \sum_{n\geq 1}\int_{\varphi_{n-1}, \bar{s}_{n-1}, \theta} \begin{array}{c} e^{\theta} \times p\left(\theta, \varphi_n, b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right)\right)^{1-\sigma} \times (1-\eta)\times \\ o\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right) \times \lambda\left(d\varphi_{n-1}, d\bar{s}_{n-1}, , n, d\theta\right) \end{array}.$$

7.3 Aggregate Labor Productivity

We define aggregate labor productivity as the aggregate output divided by total labor input, including labor used for production as well as paying fixed costs and entry costs. The aggregate output follows our definition of the CES composite of different varieties in equation (3) in the paper. In the steady state, we can express the CES composite integrating over the mass of firms with different state variables:

$$Q = \left(\sum_{n\geq 1}\int_{\varphi_{n-1}, \bar{s}_{n-1}, \theta} \begin{array}{c} e^{\theta/\sigma} q\left(\varphi_n, b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right)\right)^{\frac{\sigma-1}{\sigma}} \times (1-\eta)\times \\ o(\varphi_{n-1}, \bar{s}_{n-1}, n-1) \times \lambda\left(d\varphi_{n-1}, d\bar{s}_{n-1}, n-1, d\theta\right) \end{array}\right)^{\frac{\sigma}{\sigma-1}}.$$

Labor is used for paying variable as well as fixed costs. Denote the demand for labor from variable costs as L^{prod}, and the demand for labor from fixed costs as L^{fixed}, we have:

$$L^{prod} = \sum_{n\geq 1}\int_{\varphi_{n-1}, \bar{s}_{n-1}, \theta} \begin{array}{c} l\left(b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right)\right)^{\frac{\sigma-1}{\sigma}} \times (1-\eta)\times \\ o(\varphi_{n-1}, \bar{s}_{n-1}, n-1) \times \lambda\left(d\varphi_{n-1}, d\bar{s}_{n-1}, n-1, d\theta\right) \end{array};$$

$$L^{fixed} = f\sum_{n\geq 1}\int_{\varphi_{n-1}, \bar{s}_{n-1}, \theta} (1-\eta)o(\varphi_{n-1}, \bar{s}_{n-1}, n-1) \times \lambda\left(d\varphi_{n-1}, d\bar{s}_{n-1}, n-1, d\theta\right).$$

Finally, aggregate labor productivity is defined as

$$\frac{Q}{L} = \frac{Q}{L^{prod} + L^{fixed}}.$$

Note that there is no entry costs in our baseline model. We introduced entry costs f^e which firms have to pay to draw φ_0, θ in the "Free Entry" model in Section 5.3.4. In this case, L should also include labor used for entry, i.e., $L = L^{prod} + L^{fixed} + L^{entry}$, where $L^{entry} = Jf^e$.

References

Altig, D., J. M. Barrero, N. Bloom, S. J. Davis, B. Meyer, and N. Parker (2020): "Surveying business uncertainty," *Journal of Econometrics*.

Arkolakis, C., T. Papageorgiou, and O. A. Timoshenko (2018): "Firm learning and growth," *Review of Economic Dynamics*, 27, 146–168.

Atkeson, A. and P. j. Kehoe (2005): "Modeling and Measuring Organization Capital," *Journal of Political Economy*, 113, 1026–1053.

Bachmann, R. and C. Bayer (2014): "Investment Dispersion and the Business Cycle," *American Economic Review*, 104, 1392–1416.

Bachmann, R. and S. Elstner (2015): "Firm Optimism and Pessimism," *European Economic Review*, 79, 297–325.

Bachmann, R., S. Elstner, and A. Hristov (2017): "Surprise, Surprise – Measuring Firm-Level Investment Innovations," *Journal of Economic Dynamics and Control*, 83, 107–148.

Bachmann, R., S. Elstner, and E. R. Sims (2013): "Uncertainty and Economic Activity: Evidence from Business Survey Data," *American Economic Journal: Macroeconomics*, 5, 217–249.

Bahar, D. (2020): "The hardships of long distance relationships: time zone proximity and the location of MNC's knowledge-intensive activities," *Journal of International Economics*, 125, 103311.

Baker, S. R., T. S. McElroy, and X. S. Sheng (2020): "Expectation formation following large, unexpected shocks," *Review of Economics and Statistics*, 102, 287–303.

Baley, I. and A. Blanco (2019): "Firm Uncertainty Cycles and the Propagation of Nominal Shocks," *American Economic Journal: Macroeconomics*, 11, 276–337.

Baley, I., A. Figueiredo, and R. Ulbricht (2022): "Mismatch Cycles," *Journal of Political Economy*.

Berger, D. and J. Vavra (2019): "Shocks versus Responsiveness: What Drives Time-Varying Dispersion?" *Journal of Political Economy*, 127, 2104–2142.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93, 1268–1290.

Bertola, G. and R. J. Caballero (1994): "Irreversibility and Aggregate Investment," *The Review of Economic Studies*, 61, 223–246.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

Bloom, N., S. J. Davis, L. Foster, B. Lucking, S. Ohlmacher, and I. Saporta-Eksten (2020): "Business-Level Expectations and Uncertainty," Working Paper 28259, National Bureau of Economic Research.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018): "Really Uncertain Business Cycles," *Econometrica*, 86, 1031–1065.

Bloom, N., T. Kawakubo, C. Meng, P. Mizen, R. Riley, T. Senga, and J. Van Reenen (2021): "Do Well Managed Firms Make Better Forecasts?" Working Paper 29591, National Bureau of Economic Research.

Bloom, N., R. Lemos, R. Sadun, D. Scur, and J. Van Reenen (2014): "JEEA-FBBVA Lecture 2013: The new empirical economics of management," *Journal of the European Economic Association*, 12, 835–876.

Bloom, N. and J. Van Reenen (2007): "Measuring and explaining management practices across firms and countries," *The quarterly journal of Economics*, 122, 1351–1408.

——— (2010a): "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives*, 24, 203–224.

——— (2010b): "Why do management practices differ across firms and countries?" *Journal of economic perspectives*, 24, 203–24.

Boneva, L., J. Cloyne, M. Weale, and T. Wieladek (2020): "Firms' Price, Cost and Activity Expectations: Evidence from Micro Data," *The Economic Journal*, 130, 555–586.

Born, B., Z. Enders, G. J. Müller, and K. Niemann (2022): "Firm expectations about production and prices: Facts, determinants, and effects," *Handbook of Economic Expectations.*

Buera, F. J., J. P. Kaboski, and Y. Shin (2011): "Finance and Development: A Tale of Two Sectors," *American Economic Review*, 101, 1964–2002.

Buera, F. J. and B. Moll (2015): "Aggregate implications of a credit crunch: The importance of heterogeneity," *American Economic Journal: Macroeconomics*, 7, 1–42.

Coibion, O. and Y. Gorodnichenko (2012): "What Can Survey Forecasts Tell Us About Information Rigidities?" *Journal of Political Economy*, 120, 116–159.

——— (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105, 2644–78.

Coibion, O., Y. Gorodnichenko, and S. Kumar (2018): "How Do Firms Form Their Expectations? New Survey Evidence," *American Economic Review*, 108, 2671–2713.

Coibion, O., Y. Gorodnichenko, and T. Ropele (2020): "Inflation Expectations and Firm Decisions: New Causal Evidence," *The Quarterly Journal of Economics*, 135, 165–219.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016): "Information, Misallocation, and Aggregate Productivity," *The Quarterly Journal of Economics*, 131, 943–1005.

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

Enders, Z., F. Hünnekes, and G. J. Müller (2019): "Monetary policy announcements and expectations: Evidence from german firms," *Journal of Monetary Economics.*

——— (2022): "Firm expectations and economic activity," *Journal of the European Economic Association.*

Foster, L., J. Haltiwanger, and C. Syverson (2016): "The slow growth of new plants: learning about demand?" *Economica*, 83, 91–129.

Gennaioli, N., Y. Ma, and A. Shleifer (2016): "Expectations and investment," *NBER Macroeconomics Annual*, 30, 379–431.

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

Guiso, L. and G. Parigi (1999): "Investment and Demand Uncertainty," *The Quarterly Journal of Economics*, 114, 185–227.

Gumpert, A. (2018): "The Organization of Knowledge in Multinational Firms," *Journal of the European Economic Association*, 16, 1929–1976.

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

Hjort, J., H. Malmberg, and T. Schoellman (2021): "The missing middle managers: Labor costs, firm structure, and development," .

Hsieh, C.-T. and P. J. Klenow (2014): "The life cycle of plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084.

Ilut, C. and H. Saijo (2021): "Learning, confidence, and business cycles," *Journal of Monetary Economics*, 117, 354–376.

Ilut, C., R. Valchev, and N. Vincent (2020): "Paralyzed by Fear: Rigid and Discrete Pricing Under Demand Uncertainty," *Econometrica*, 88, 1899–1938.

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

Kehrig, M. (2015): "The Cyclical Nature of the Productivity Distribution," Working Paper.

Khan, A. and J. K. Thomas (2013): "Credit shocks and aggregate fluctuations in an economy with production heterogeneity," *Journal of Political Economy*, 121, 1055–1107.

Lo, A. W. and A. C. MacKinlay (1988): "Stock market prices do not follow random walks: Evidence from a simple specification test," *The review of financial studies*, 1, 41–66.

Luo, Y. (2008): "Consumption dynamics under information processing constraints," *Review of Economic Dynamics*, 11, 366–385.

Mackowiak, B. and M. Wiederholt (2009): "Optimal sticky prices under rational inattention," *American Economic Review*, 99, 769–803.

Midrigan, V. and D. Y. Xu (2014): "Finance and misallocation: Evidence from plant-level data," *American economic review*, 104, 422–58.

Ryngaert, J. (2017): "What Do (and Don't) Forecasters Know about U.S. Inflation?" Working Paper.

Sager, E. and O. A. Timoshenko (2021): "Demand Uncertainty, Selection, and Trade," Working Paper.

Sedláček, P. and V. Sterk (2017): "The growth potential of startups over the business cycle," *American Economic Review*, 107, 3182–3210.

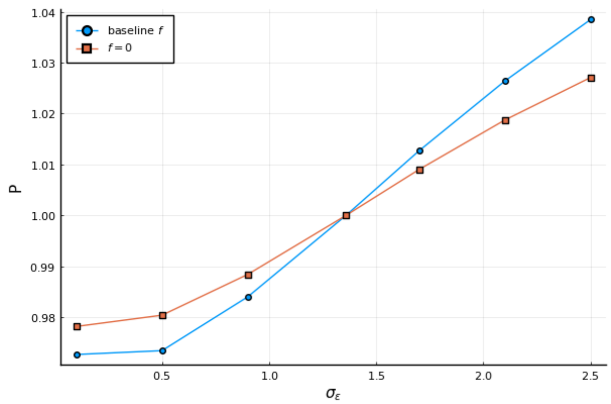
Sims, C. A. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.

Tanaka, M., N. Bloom, J. M. David, and M. Koga (2019): "Firm performance and macro forecast accuracy," *Journal of Monetary Economics.*

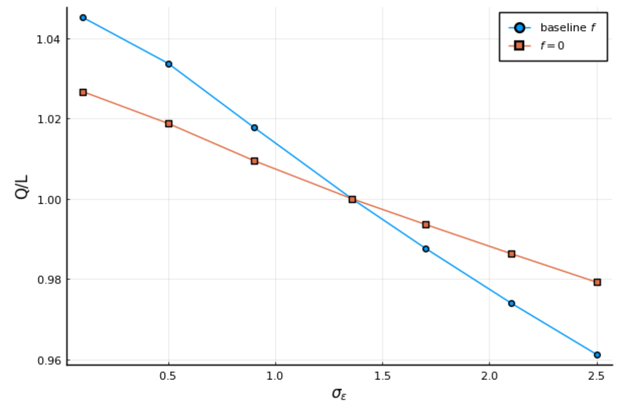
Triebs, T. P. and J. Tumlinson (2013): "Learning to Forecast the Hard Way—Evidence from German Reunification," Working Paper 19209, National Bureau of Economic Research.

Vavra, J. (2014): "Inflation dynamics and time-varying volatility: New evidence and an Ss interpretation," *Quarterly Journal of Economics*, 129, 215–258.

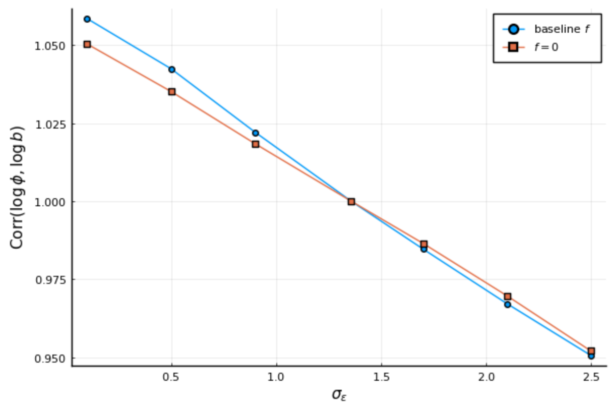
Figure 5: The Impact of σ_ε on Aggregate Outcomes



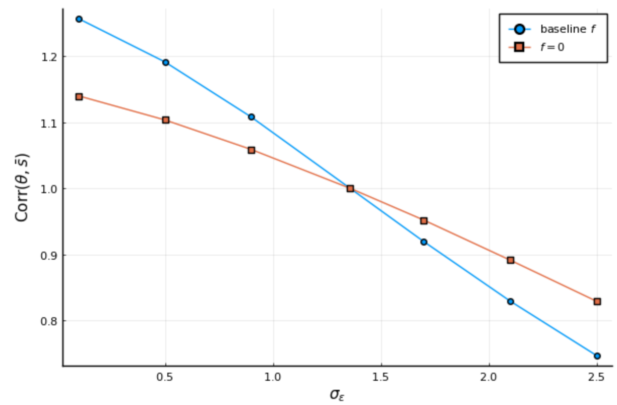
(a) P



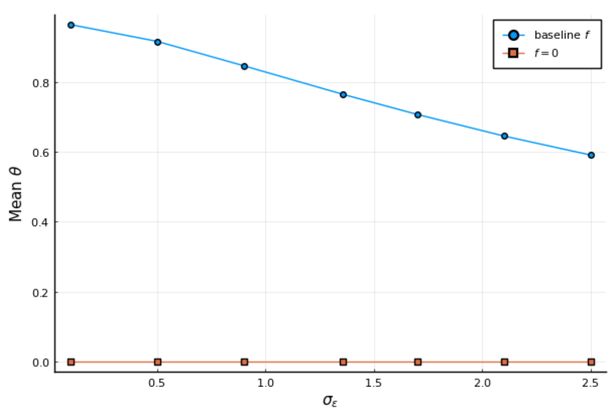
(b) Q/L



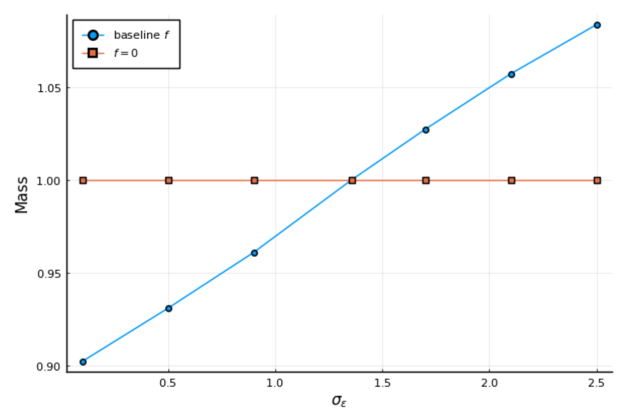
(c) Corr $(\log \phi, \log b)$



(d) Corr $(\theta, \bar{s})$



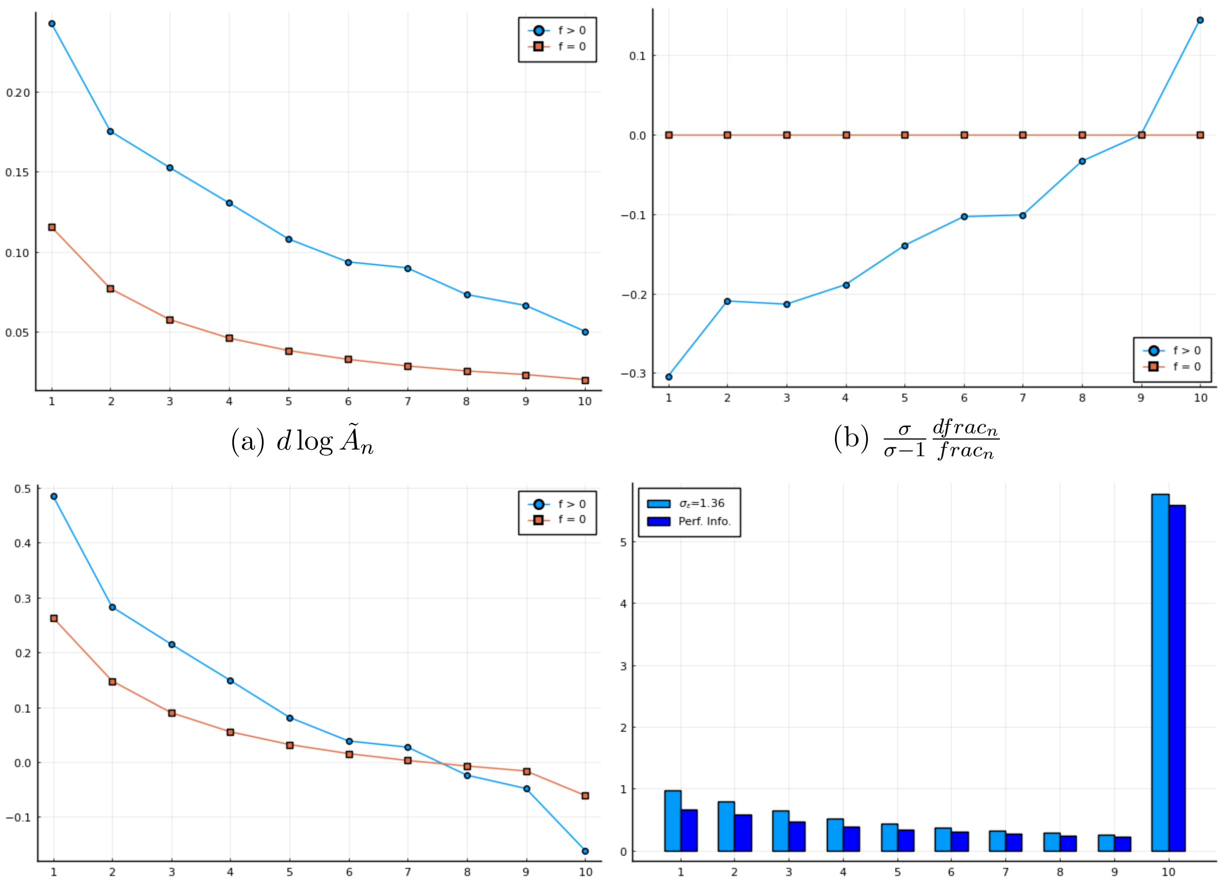
(e) Average Incumbent θ



(f) Relative Mass of Firms: M

Notes: Panels (a) to (f) shows the aggregate industry price index, labor productivity, correlation between $\log(\phi)$ and $\log(b)$ among incumbents, correlation between θ and $\bar{s}$ among incumbents, average incumbent θ and equilibrium mass of firms under different fixed costs f and different values of σ_ε. All variables are normalized to one in the case $\sigma_\varepsilon = 1.36$, the calibrated value in our baseline model. $\log \phi$ is defined as the combination of labor productivity and demand, $(\sigma - 1) \log \varphi + \theta$, which determines the size of the firm in a static model. b is defined as in equation (9). The blue dotted line indicates a model with fixed costs f at the value in the baseline calibration, while the red line with squares indicates a model with zero fixed costs (no endogenous entry/exit).

Figure 6: Decomposing the Impact of σ_ε Across Age Groups: $\sigma_\varepsilon = 1.36 \to$ Perfect Information



(a) $d \log \tilde{A}_n$

(b) $\frac{\sigma}{\sigma-1} \frac{dfrac_n}{frac_n}$

(c) $d \log(\bar{L}_n^{prod} / \bar{L}^{prod})$

(d) Firm Mass

Notes: Panels (a) to (c) plot the three key components in the change in normalized industry labor productivity according to equation (21), contributed by firms of different ages n (capped by 10 years), when changing the model from the baseline imperfect information ($\sigma_\varepsilon = 1.36$) to a dynamic model in which firms have perfect information about θ. The blue dotted line represents the case in which the fixed costs f are kept at the baseline value, 0.0093. The red line with squares represents the case where $f = 0$, i.e., without endogenous entry/exit. Panel (d) shows the mass of firms at different ages in the imperfect and perfect information model, respectively.

Online Appendix for "Uncertainty, Imperfect Information, and Expectation Formation over the Firm's Life Cycle"

Cheng Chen Tatsuro Senga Chang Sun Hongyong Zhang*

July, 2022

Abstract

In the online appendix, we first provide robustness checks for the stylized facts documented in the paper. Second, we provide a theory appendix that considers alternative setups of the model and their implications on the forecast errors. In particular, we show that the perfect information benchmark and a Jovanovic-type learning model with payoff-relevant noises cannot yield serially correlated forecast errors. Finally, we present additional results from our quantitative analysis.

*Chen: Clemson University, cchen9@clemson.edu. Senga: Queen Mary University of London, ESCoE, and RIETI, t.senga@qmul.ac.uk. Sun: University of Hong Kong, sunc@hku.hk. Zhang: RIETI, zhang-hong-yong@rieti.go.jp.

1 Additional Empirical Analysis

1.1 Basic Data Description

Our data include firms belonging to Japanese multinational corporations (MNCs) in various countries and industries in 1995-2014. Our baseline regression sample requires a well-defined forecast error (FE) from period t to $t+1$, so in this section we report firm-level statistics up to the year 2013. In Table OA.1, we report the average number of business groups (groups of firms belonging to the same parent firm) and number of firms in a typical year in four periods, 1995-2000, 2001-2005, 2006-2010 and 2011-2013. These numbers gradually increase over time, while the average/median firm size measured by employment remains stable. On average, we have 6922 firms belonging to 1781 unique parent firms in a typical year during the entire sample period.

Table OA.1: Descriptive Statistics by Time Periods

	Annual Average # of		Employment Statistics			
Year range	Business Groups	Firms	Mean	25th Perc.	50th Perc.	75th Perc.
1995-2000	1059	4701	300.1	20	75	248
2001-2005	1503	6612	335.5	21	79	270
2006-2010	2244	8295	333.6	19	71	244
2011-2013	2919	9592	297.8	16	62	218
1995-2013	1781	6922	319.0	19	71	247

Notes: This table reports the average number of firms/business groups in our baseline regression sample and the corresponding employment statistics in each period. Source: Basic Survey on Overseas Business Activities, Ministry of Economy, Trade and Industry (METI).

In Table OA.2, we report number of firms in major countries/regions in 2013. The countries/regions are consistent with our regional analysis in Section 5.3.3 of the paper. A large number of firms in our sample are in major markets of Japanese MNCs such as China, ASEAN countries and the United States. A small number of firms operate in regions such as Africa, Middle East and Eastern Europe. For the list of countries in each region, see Table OA.17.

Table OA.3 reports the number of firms in the top 10 industries in 2013. Our data contains both manufacturing and services firms. Not surprisingly, the industry that contains the largest number of firms is “wholesale and retail trade”, followed by “manufacturing of transportation equipment”, an industry that is well-known for Japanese firms’ overseas footprint. It is clear from the table that our sample covers a wide range of industries. For all our key facts, we show that they hold in both

Table OA.2: Number of Firms in Major Countries/Regions, 2013

Major Country/Region	# of Firms
Africa	41
Middle East	70
Eastern Europe	142
Latin America	307
ASEAN	2556
China	3430
Western Europe	920
United States	1287

Notes: This table reports the number of firms in major countries/industries in 2013. See Table OA.17 for the list of countries in each region. Source: Basic Survey on Overseas Business Activities, Ministry of Economy, Trade and Industry (METI).

the whole sample and the manufacturing subsample.

Table OA.3: Number of Firms in Top 10 Industries, 2013

Industry	# of Firms
Wholesale and retail trade	3001
Transportation equipment	1119
Miscellaneous manufacturing industries	622
Other Business Services	611
Chemical and allied products	547
Information and communications equipment	496
Transport	434
Production machinery	385
Electrical machinery, equipment and supplies	347
Information and communications	331

Notes: This table reports the number of firms in the top 10 industries in 2013. Source: Basic Survey on Overseas Business Activities, Ministry of Economy, Trade and Industry (METI).

1.2 Alternative Definitions of Forecast Errors and Summary Statistics

We introduce two alternative definitions of forecast errors, which are used for robustness checks later.

First, we define the percentage deviation of the realized sales from the sales forecasts as

$$FE_{t,t+1}^{\text{pct}} = \frac{R_{t+1}}{E_t\left(R_{t+1}\right)} - 1.$$

Second, we construct a measure for the "residual forecast error" measure in an effort to isolate the firm-level idiosyncratic components reflected in the forecast errors. To exclude systemic components, such as business cycles, from the forecast errors, we project the raw forecast error onto country-year and industry-year fixed effects

$$FE_{t,t+1}^{\log} = \delta_{ct} + \delta_{st} + \hat{\epsilon}_{t,t+1}^{FE,\log}, \tag{1}$$

and obtain the residual forecast error $\hat{\epsilon}_{t,t+1}^{FE,\log}$. As it turns out, the fixed effects only account for about 11% of the variation, which indicates that firm-level uncertainty plays a dominant role in generating the firms' forecast errors. We obtain $\hat{\epsilon}_{t,t+1}^{FE,\text{pct}}$ based on the percentage forecast errors for additional robustness checks using the same approach.

The first four rows of Table OA.4 report summary statistics of our main forecast error definition (log deviation, raw) as well as the alternative forecast errors. While the mean of the residual forecast errors, $\hat{\epsilon}_{t,t+1}^{FE,\log}$ and $\hat{\epsilon}_{t,t+1}^{FE,\text{pct}}$, is zero by construction, the mean and median of $FE_{t,t+1}^{\log}$ and $FE_{t,t+1}^{\text{pct}}$ are also close to zero. In the middle four rows, we report the summary statistics of the absolute value of various constructed forecast errors. Since the country-year and industry-year fixed effects account for a small fraction of the variation, the mean, median, and standard deviation of $|\hat{\epsilon}_{t,t+1}^{FE,\log}|$ (and $|\hat{\epsilon}_{t,t+1}^{FE,\text{pct}}|$) are similar to those of $\left|FE_{t,t+1}^{\log}\right|$ (and $\left|FE_{t,t+1}^{\text{pct}}\right|$). The patterns of manufacturing firms' forecast errors are similar to the overall patterns, as shown by the last four rows of the table.

Table OA.4: Summary statistics of the forecast errors

	Obs.	mean	std. dev.	median
$FE^{\text{log}}_{t,t+1}$	131834	-0.024	0.298	-0.005
$FE^{\text{pct}}_{t,t+1}$	132373	0.017	0.332	-0.006
$\hat{\epsilon}^{FE,\text{log}}_{t,t+1}$	131550	-0.000	0.280	0.011
$\hat{\epsilon}^{FE,\text{pct}}_{t,t+1}$	132090	0.000	0.314	-0.022
$\lvert FE^{\text{log}}_{t,t+1}\rvert$	131834	0.200	0.222	0.130
$\lvert FE^{\text{pct}}_{t,t+1}\rvert$	132373	0.203	0.263	0.130
$\lvert\hat{\epsilon}^{FE,\text{log}}_{t,t+1}\rvert$	131550	0.184	0.211	0.115
$\lvert\hat{\epsilon}^{FE,\text{pct}}_{t,t+1}\rvert$	132090	0.189	0.251	0.117
$FE^{\text{log}}_{t,t+1}$ - Manufacturing	80987	-0.022	0.278	-0.004
$FE^{\text{pct}}_{t,t+1}$ - Manufacturing	81244	0.014	0.307	-0.004
$\lvert FE^{\text{log}}_{t,t+1}\rvert$ - Manufacturing	80987	0.186	0.208	0.123
$\lvert FE^{\text{pct}}_{t,t+1}\rvert$ - Manufacturing	81244	0.188	0.242	0.124

Notes: $FE^{\text{log}}_{t,t+1}$ is the log deviation of the realized sales from the sales forecasts, while $FE^{\text{pct}}_{t,t+1}$ is the percentage deviation of the realized sales from the sales forecasts. $\hat{\epsilon}^{FE,\text{log}}_{t,t+1}$ is the residual log forecast error, which we obtain by regressing $FE^{\text{log}}_{t,t+1}$ on a set of industry-year and country-year fixed effects. Similarly, $\hat{\epsilon}^{FE,\text{pct}}_{t,t+1}$ is the residual percentage forecast error, which we obtain by regressing $FE^{\text{pct}}_{t,t+1}$ on a set of industry-year and country-year fixed effects.

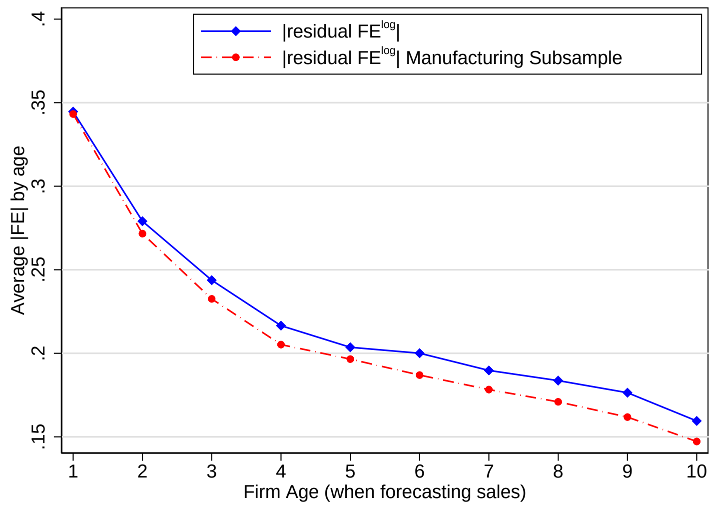
1.3 Robustness Checks for Fact 1: Affiliate Age on Uncertainty

1.3.1 Alternative Measures of FE

We first show that our baseline results in Figure 2 and Table 3 of the paper are robust to alternative measures of forecast errors. Figure OA.1 plots the average absolute value of the residual forecast errors $\hat{\epsilon}^{FE,\text{log}}_{t,t+1}$, for the entire sample and for the manufacturing subsample, respectively. We see a clear pattern that older firms make more precise forecasts.

Tables OA.5 and OA.6 use the absolute value of percentage forecast errors and residual log forecast errors, respectively.

Figure OA.1: $|\hat{\epsilon}^{FE,\log}_{t,t+1}|$ declines with firm age



Note: Average absolute value of residual $FE^{\log}$ by age cohorts.

Table OA.5: Age effects on the absolute percentage forecast errors, $|FE^{\text{pct}}_{t,t+1}|$

Sample:	All Firms			Survivors	Manufacturing
Dep.Var: $\lvert FE^{\text{pct}}_{t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t = 2)$	-0.068^a	-0.061^a	-0.066^a	-0.065^a	-0.063^a
	(0.008)	(0.008)	(0.009)	(0.012)	(0.012)
$\mathbb{1}(\text{Age}_t = 3)$	-0.104^a	-0.091^a	-0.090^a	-0.086^a	-0.088^a
	(0.008)	(0.008)	(0.009)	(0.012)	(0.011)
$\mathbb{1}(\text{Age}_t = 4)$	-0.131^a	-0.116^a	-0.112^a	-0.101^a	-0.114^a
	(0.009)	(0.009)	(0.009)	(0.013)	(0.012)
$\mathbb{1}(\text{Age}_t = 5)$	-0.142^a	-0.125^a	-0.115^a	-0.110^a	-0.108^a
	(0.008)	(0.008)	(0.009)	(0.015)	(0.012)
$\mathbb{1}(\text{Age}_t = 6)$	-0.145^a	-0.126^a	-0.116^a	-0.114^a	-0.116^a
	(0.008)	(0.008)	(0.009)	(0.015)	(0.012)
$\mathbb{1}(\text{Age}_t = 7)$	-0.157^a	-0.135^a	-0.122^a	-0.131^a	-0.121^a
	(0.008)	(0.008)	(0.010)	(0.016)	(0.012)
$\mathbb{1}(\text{Age}_t = 8)$	-0.159^a	-0.135^a	-0.120^a	-0.117^a	-0.121^a
	(0.008)	(0.008)	(0.010)	(0.018)	(0.012)
$\mathbb{1}(\text{Age}_t = 9)$	-0.161^a	-0.136^a	-0.120^a	-0.118^a	-0.125^a
	(0.009)	(0.009)	(0.010)	(0.020)	(0.012)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.175^a	-0.139^a	-0.120^a	-0.122^a	-0.119^a
	(0.008)	(0.008)	(0.010)	(0.022)	(0.013)
$\log(\text{Emp})_t$		-0.022^a	-0.033^a	-0.043^a	-0.032^a
		(0.001)	(0.003)	(0.006)	(0.003)
$\log(\text{Parent Emp})_t$		-0.000	0.000	0.005	0.000
		(0.001)	(0.003)	(0.007)	(0.003)
Industry-year FE	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y
Firm FE			Y	Y	Y
N	131757	128931	123609	22090	77062
R^2	0.094	0.110	0.339	0.318	0.338

Notes: Standard errors are clustered at the business group level, c 0.10 b 0.05 a 0.01. The dependent variable is the absolute value of forecast errors in all regressions. Age is the age of the firm when making the forecasts. Regressions in columns 1, 2 and 3 include all firms, while the regression in column 4 only includes firms that have continuously appeared in the sample from age 1 to age 7.

Table OA.6: Age effects on the absolute residual log forecast errors, $\left|\hat{\epsilon}^{\log}_{FE,t,t+1}\right|$

Sample:	All Firms			Survivors	Manufacturing
Dep.Var: $\left|\hat{\epsilon}^{\log}_{FE,t,t+1}\right|$	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t = 2)$	-0.066^a	-0.059^a	-0.065^a	-0.073^a	-0.071^a
	(0.007)	(0.007)	(0.007)	(0.009)	(0.011)
$\mathbb{1}(\text{Age}_t = 3)$	-0.100^a	-0.087^a	-0.087^a	-0.093^a	-0.098^a
	(0.007)	(0.007)	(0.007)	(0.010)	(0.010)
$\mathbb{1}(\text{Age}_t = 4)$	-0.126^a	-0.111^a	-0.110^a	-0.110^a	-0.124^a
	(0.007)	(0.007)	(0.008)	(0.011)	(0.011)
$\mathbb{1}(\text{Age}_t = 5)$	-0.138^a	-0.121^a	-0.115^a	-0.121^a	-0.124^a
	(0.007)	(0.007)	(0.008)	(0.012)	(0.011)
$\mathbb{1}(\text{Age}_t = 6)$	-0.141^a	-0.123^a	-0.115^a	-0.123^a	-0.127^a
	(0.007)	(0.007)	(0.008)	(0.012)	(0.011)
$\mathbb{1}(\text{Age}_t = 7)$	-0.151^a	-0.129^a	-0.122^a	-0.135^a	-0.134^a
	(0.007)	(0.007)	(0.008)	(0.013)	(0.011)
$\mathbb{1}(\text{Age}_t = 8)$	-0.155^a	-0.132^a	-0.122^a	-0.128^a	-0.136^a
	(0.007)	(0.007)	(0.008)	(0.015)	(0.011)
$\mathbb{1}(\text{Age}_t = 9)$	-0.161^a	-0.136^a	-0.127^a	-0.130^a	-0.141^a
	(0.007)	(0.007)	(0.008)	(0.017)	(0.011)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.173^a	-0.138^a	-0.125^a	-0.132^a	-0.137^a
	(0.007)	(0.006)	(0.008)	(0.018)	(0.011)
$\log(\text{Emp})_t$		-0.022^a	-0.027^a	-0.037^a	-0.027^a
		(0.001)	(0.002)	(0.005)	(0.002)
$\log(\text{Parent Emp})_t$		-0.000	0.001	0.011	0.001
		(0.001)	(0.003)	(0.007)	(0.003)
Industry-year FE	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y
Firm FE			Y	Y	Y
N	131230	128429	123111	21982	76823
R^2	0.082	0.104	0.361	0.365	0.352

Notes: Standard errors are clustered at the business group level, c 0.10 b 0.05 a 0.01. The dependent variable is the absolute value of forecast errors in all regressions. Age is the age of the firm when making the forecasts. Regressions in columns 1, 2 and 3 include all firms, while the regression in column 4 only includes firms that have continuously appeared in the sample from age 1 to age 7.

1.3.2 Conditional Variance: a Two-step Approach

Second, we address the concern that the decline in $|FE|$ may reflect a reduction in firms' biases in the level of FEs rather than a reduction in the variance of FEs. We do so by characterizing the conditional variance of FEs using a two step procedure and test whether it depends on the firm's age. To derive this, we first assume that the conditional expectation of forecast errors is linear in the independent variables (including fixed effects)

$$E(FE|X) = \beta X.$$

Therefore, the conditional variance becomes

$$V(FE|X) = E((FE - \beta X)^2|X).$$

To test whether $V(FE|X)$ depends on firm age and other independent variables, we first regress FE on all the regressors and obtain the squared residual term:

$$\hat{v}_{FE}^2 \equiv (FE - \hat{\beta}X)^2.$$

We then project $\hat{v}_{FE}^2$ onto X in the second-stage regression.[1] When we include firm age as an independent variable, the coefficient of age in the second-stage regression is informative about whether the variance of firm-level forecast errors is affected by firm age. One can test other potential determinants of the variance in the same way.

In Table OA.7, we perform the two-step procedure, using the log forecast error as the key dependent variable (FE in the derivation above). Though the age coefficients here are not directly comparable to regressions with absolute forecast errors as the dependent variable, this procedure reveals similar patterns as Table 3 of the paper: firm-level uncertainty declines as firms gain more experience. In Column 5, we define forecast errors using percentage deviations, and the effects of age on conditional variance of these errors are similar to that in Column 2 where we use the log

[1] We use $\hat{v}$ to denote the residual term here to distinguish from the residual forecast errors defined in equation 1. The latter is obtained by purging the country-year and industry-year fixed effects only, while the former purges all regressors that we believe may affect the conditional variance of forecast errors, including the age dummies and other controls.

forecast errors.

Table OA.7: Age effects on the variance forecast errors: conditional variance regressions

Dep. Var.	$\hat{v}^2_{FE,\log}(t,t+1)$				$\hat{v}^2_{FE,\mathrm{pct}}(t,t+1)$
Sample:	All Firms		Survivors	Manufacturing	All Firms
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\mathrm{Age}_t = 2)$	-0.063^a	-0.036^a	-0.047^a	-0.037^a	-0.043^a
	(0.008)	(0.006)	(0.008)	(0.008)	(0.012)
$\mathbb{1}(\mathrm{Age}_t = 3)$	-0.094^a	-0.052^a	-0.057^a	-0.056^a	-0.064^a
	(0.008)	(0.006)	(0.008)	(0.008)	(0.012)
$\mathbb{1}(\mathrm{Age}_t = 4)$	-0.118^a	-0.066^a	-0.068^a	-0.071^a	-0.082^a
	(0.008)	(0.006)	(0.009)	(0.008)	(0.012)
$\mathbb{1}(\mathrm{Age}_t = 5)$	-0.124^a	-0.068^a	-0.073^a	-0.070^a	-0.081^a
	(0.008)	(0.006)	(0.010)	(0.008)	(0.013)
$\mathbb{1}(\mathrm{Age}_t = 6)$	-0.125^a	-0.069^a	-0.074^a	-0.072^a	-0.086^a
	(0.008)	(0.006)	(0.010)	(0.009)	(0.013)
$\mathbb{1}(\mathrm{Age}_t = 7)$	-0.131^a	-0.073^a	-0.081^a	-0.077^a	-0.092^a
	(0.008)	(0.006)	(0.011)	(0.009)	(0.013)
$\mathbb{1}(\mathrm{Age}_t = 8)$	-0.131^a	-0.071^a	-0.072^a	-0.077^a	-0.083^a
	(0.008)	(0.006)	(0.012)	(0.009)	(0.013)
$\mathbb{1}(\mathrm{Age}_t = 9)$	-0.134^a	-0.074^a	-0.073^a	-0.081^a	-0.083^a
	(0.008)	(0.006)	(0.013)	(0.009)	(0.014)
$\mathbb{1}(\mathrm{Age}_t \geq 10)$	-0.135^a	-0.072^a	-0.080^a	-0.077^a	-0.082^a
	(0.007)	(0.006)	(0.014)	(0.009)	(0.013)
$\log(\mathrm{Emp})_t$	-0.017^a	-0.016^a	-0.022^a	-0.015^a	-0.028^a
	(0.001)	(0.002)	(0.004)	(0.002)	(0.004)
$\log(\mathrm{Parent\ Emp})_t$	0.001	0.002	0.004	0.002	-0.001
	(0.001)	(0.002)	(0.007)	(0.002)	(0.003)
Industry-year FE	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y
Firm FE		Y	Y	Y	Y
N	128429	123111	21982	76823	123609
R^2	0.071	0.317	0.307	0.300	0.261

Notes: Standard errors are clustered at the business group level. Significance levels: c 0.1, b 0.05, a 0.01. Age is the age of the firm when making the forecasts. Regressions in columns 1, 2 and 5 include all firms. Column 3 includes firms that continuously show up in the data from age one to age seven. Column 4 focuses on the manufacturing subsample.

1.3.3 Excluding Naive Forecasts

Third, we show that our results are not driven by firms that use simple forecasting rules. In our data, about 3.4% of the firms use their current sales as their sales forecasts for the next year. Though it is impossible to gauge what fraction of these firms misreport their forecasts, we try to be conservative and drop all of them from our dataset and run the regressions in Table 3 of the paper. The results are almost identical (see Table OA.8).

Table OA.8: Age effects on the absolute value of forecast errors: no naive forecasting rule

Sample:	All Firms			Survivors	Manufacturing
Dep.Var: $\lvert FE^{\log}_{t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t = 2)$	-0.068^a	-0.061^a	-0.064^a	-0.071^a	-0.072^a
	(0.007)	(0.007)	(0.008)	(0.010)	(0.011)
$\mathbb{1}(\text{Age}_t = 3)$	-0.102^a	-0.088^a	-0.087^a	-0.095^a	-0.103^a
	(0.007)	(0.007)	(0.008)	(0.010)	(0.011)
$\mathbb{1}(\text{Age}_t = 4)$	-0.126^a	-0.112^a	-0.108^a	-0.108^a	-0.124^a
	(0.007)	(0.007)	(0.008)	(0.011)	(0.011)
$\mathbb{1}(\text{Age}_t = 5)$	-0.142^a	-0.124^a	-0.115^a	-0.122^a	-0.127^a
	(0.007)	(0.007)	(0.008)	(0.012)	(0.011)
$\mathbb{1}(\text{Age}_t = 6)$	-0.143^a	-0.124^a	-0.114^a	-0.123^a	-0.131^a
	(0.007)	(0.007)	(0.008)	(0.013)	(0.011)
$\mathbb{1}(\text{Age}_t = 7)$	-0.152^a	-0.130^a	-0.120^a	-0.138^a	-0.136^a
	(0.007)	(0.007)	(0.008)	(0.014)	(0.011)
$\mathbb{1}(\text{Age}_t = 8)$	-0.156^a	-0.133^a	-0.121^a	-0.130^a	-0.140^a
	(0.007)	(0.007)	(0.009)	(0.015)	(0.012)
$\mathbb{1}(\text{Age}_t = 9)$	-0.160^a	-0.135^a	-0.122^a	-0.133^a	-0.141^a
	(0.007)	(0.007)	(0.009)	(0.017)	(0.012)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.172^a	-0.137^a	-0.121^a	-0.138^a	-0.135^a
	(0.007)	(0.007)	(0.009)	(0.019)	(0.012)
$\log(\text{Emp})_t$		-0.021^a	-0.023^a	-0.034^a	-0.024^a
		(0.001)	(0.002)	(0.005)	(0.002)
$\log(\text{Parent Emp})_t$		0.001	0.001	0.008	-0.000
		(0.001)	(0.003)	(0.007)	(0.003)
Industry-year FE	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y
Firm FE			Y	Y	Y
N	127278	124872	119615	21481	75179
R^2	0.107	0.124	0.368	0.361	0.365

Notes: Standard errors are clustered at the business group level. Age is the age of the firm when making the forecasts. Regressions in columns 1, 2 and 3 include all firms, while the regression in column 4 only includes firms that have continuously appeared in the sample from age 1 to age 7.

1.3.4 Controlling Market/Product Diversification

As firm ages, it is possible that they enhance their capabilities and diversify their businesses by selling to more markets and selling more products . This diversification argument implies that firm demand becomes less volatile when the firm becomes older and provides an alternative interpretation of the age effects on the absolute forecast errors. To evaluate the relevance of this alternative explanation, we construct various measures of market/product diversification for firms, and show that including them in the regressions does not eliminate the impact of age on the decline in variance of forecast errors. Therefore, we argue that learning about demand provides a good explanation for the patterns documented in the paper.

In Columns 1 and 2 of Table OA.9, we use the number of destination markets as a measure of market diversification and the Herfindahl-Hirschman Index (HHI) as an inverse measure of market diversification, respectively. In our data, we observe the firms' sales up to six markets: the host country (local market), Japan, Asia, North America, Europe and the rest of the world.[2] We therefore define the HHI of firm i as

$$HHI_i^{markets} = \sum_{m=1}^{6} s_{im}^2,$$

where s_{im} is the share of market m sales in firm i's total sales. Consistent with the findings in Garetto et al. (2019), we find that firms grow by diversifying their destination markets (results available upon request). Columns 1 and 2 show that market diversification has a negative impact on the absolute value of forecast errors and reduced the impact of age compared to Column 3 in Table 3 of the paper. However, the age coefficients are still negatively significant and maintain 80% of the magnitude of those in Table 3.

The Japanese foreign activities survey provides limited information on sales by market, and does not break down affiliated firms' sales by product. To construct finer measures of market/product diversification, we merge the subset of firms operating in China with the China customs data (2000 - 2009). This involves translating the firms' names to Chinese (most of them are in English in the foreign activities survey) and matching them with the exporter names in the customs data. We

[2] Affiliates' sales to the four continents exclude the sales in the local market, if they are located in any of these continents.

were able to match 3925 out of the 7317 affiliated firms in China to the customs data between 2000 and 2009. Among the matched firms, the median number of exporting destinations is two (maximum = 149), and the median number of HS 6-digit products is four (maximum = 461).

In Columns 3 and 4, we calculate a refined measure of market diversification by combining the customs data with the six-market diversification measures in Columns 1 and 2. In particular, if the firm can be matched to the customs data, the number of markets it serves equals to the number of export destinations or the number of export destinations plus one, depending on whether it sells locally in China. The HHI of market sales is also calculated by combing the local sales and sales to each export destination. To increase the sample size, we use the six-market diversification measures, if the firm cannot be matched to the customs data. To capture the potential non-linear effects of the number of markets, we use the logarithm of this variable instead of its level. As is shown in the table, these market diversification measures have a negative but insignificant effect on firm-level uncertainty of the affiliated firms in China, while the age effects remain large and significant.

Finally, in Columns 5 and 6, we examine the impact of product diversification. For each given year, we calculate the number of export products at the HS 6-digit level, and also the HHI using product level sales of a firm i in China

$$HHI_i^{products} = \sum_{p=1}^{N_i} s_{ip}^2,$$

where N_i is the total number of products and s_{ip} is the export share of product p in firm i's total exports. One caveat is that we only observe exports by products from the China customs data but do not observe sales by product in the local market, so the product diversification variables inevitably contain measurement errors. However, we believe they still capture the extent to which firms diversify their product portfolio. Similar to Columns 3-4, we see a negative and insignificant impact of product diversification on firm-level uncertainty, while the age effects remain significant and large.

Table OA.9: Age effects on the absolute value of forecast errors: controlling for market/product diversification

Sample:	All Affiliates		All Chinese Affiliates		Matched with China Customs	
Dep.Var: $\lvert FE^{\log}_{t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}(\text{Age}_t = 2)$	-0.047^a	-0.049^a	-0.059^a	-0.061^a	-0.047	-0.048
	(0.009)	(0.009)	(0.015)	(0.015)	(0.031)	(0.031)
$\mathbb{1}(\text{Age}_t = 3)$	-0.064^a	-0.067^a	-0.079^a	-0.080^a	-0.073^b	-0.073^b
	(0.008)	(0.008)	(0.015)	(0.015)	(0.030)	(0.030)
$\mathbb{1}(\text{Age}_t = 4)$	-0.082^a	-0.085^a	-0.102^a	-0.104^a	-0.077^b	-0.077^a
	(0.008)	(0.009)	(0.016)	(0.016)	(0.030)	(0.030)
$\mathbb{1}(\text{Age}_t = 5)$	-0.091^a	-0.094^a	-0.119^a	-0.122^a	-0.088^a	-0.089^a
	(0.008)	(0.009)	(0.016)	(0.016)	(0.030)	(0.030)
$\mathbb{1}(\text{Age}_t = 6)$	-0.089^a	-0.092^a	-0.116^a	-0.120^a	-0.074^b	-0.074^b
	(0.009)	(0.009)	(0.016)	(0.016)	(0.032)	(0.032)
$\mathbb{1}(\text{Age}_t = 7)$	-0.096^a	-0.100^a	-0.123^a	-0.127^a	-0.074^b	-0.075^b
	(0.009)	(0.009)	(0.017)	(0.017)	(0.032)	(0.032)
$\mathbb{1}(\text{Age}_t = 8)$	-0.097^a	-0.100^a	-0.123^a	-0.126^a	-0.085^b	-0.085^b
	(0.009)	(0.009)	(0.017)	(0.017)	(0.033)	(0.033)
$\mathbb{1}(\text{Age}_t = 9)$	-0.100^a	-0.104^a	-0.128^a	-0.132^a	-0.085^b	-0.086^b
	(0.009)	(0.009)	(0.017)	(0.017)	(0.034)	(0.034)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.100^a	-0.103^a	-0.127^a	-0.131^a	-0.078^b	-0.079^b
	(0.009)	(0.009)	(0.018)	(0.018)	(0.037)	(0.037)
# of Markets at t	-0.003^a					
	(0.001)					
HHI Market Sales at t		0.015^a		0.010		
		(0.005)		(0.009)		
log # of Markets at t			-0.002			
			(0.003)			
log # of HS6 Products at t					-0.002	
					(0.004)	
HHI HS6 Product Exports at t						0.006
						(0.014)
$\log(\text{Emp})_t$	-0.022^a	-0.023^a	-0.026^a	-0.026^a	-0.030^a	-0.030^a
	(0.002)	(0.002)	(0.004)	(0.004)	(0.011)	(0.011)
$\log(\text{Parent Emp})_t$	-0.001	-0.001	0.000	-0.001	0.001	0.001
	(0.003)	(0.003)	(0.006)	(0.006)	(0.010)	(0.010)
Industry-year FE	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
N	109102	104598	27103	26514	8066	8177
R^2	0.372	0.376	0.376	0.378	0.396	0.393

Notes: Standard errors are clustered at the business group level, c 0.10 b 0.05 a 0.01.. Age is the age of the firm when making the forecasts. Columns 1-2 include all firms, columns 3-4 include all firms operating in China, while columns 5-6 include firms that can be matched to the China Customs data. In columns 1-2, we calculate # of markets and HHI of market sales using information on firms' sales in six markets: the host country, Japan, Asia, North America, Europe and Latin America, where the sales to the four continents exclude those in the host country if the firm locates in one of the continents. In columns 3-4, one market refers to one country if the firm can be found in the customs data, while the market is defined in the same way as columns 1-2 if the match is unsuccessful. In columns 5-6, we only focus on the firms that can be found in the customs data. The number of products and the HHI index are calculated at the HS 6-digit product level that the firm exports.

1.3.5 Partial Year Effects

In Table OA.10, we show that the age effects, especially the difference between age one and age two firms, are not driven by the "partial year effects". The partial year effects are potentially relevant here since some age one firms entered relatively late in its founding year. As a result, they may not have enough information to make a precise forecast at the time of the survey. To investigate this issue, we use the information on the firms' founding months and split the age one firms into two groups: those that entered in the first half of the founding year and those that entered in the second half of the founding year.

In Columns 1 and 2, we treat the age one firms that entered in the second half of the year as the base group. These firms have less than six months of experience at the time of survey (age $\in (0, 0.5)$), and should arguably have the highest forecast error. We then include the other age dummies, including one dummy indicating age one firms that entered in the first half of the year (age $\in (0.5, 1)$). We find some suggestive evidence that an additional six month of experience reduces the absolute forecast errors, though the effect is not significant when we include firm fixed effects. On the other hand, age two firms have significantly smaller forecast errors than both groups of age one firms.

In Columns 3-4, we provide additional robustness checks by excluding age one firms that entered in the second half of the founding year. In Column 5, we exclude age one firms and show that the decline in forecast errors is still significant after age two, though at a smaller scale. All these results are consistent with learning and cannot be totally driven by the partial year effect of age one firms.

Table OA.10: Age effects on the absolute residual forecast errors: robustness to partial year effects.

Sample:	All Affiliates		Excluding Age 0-0.5		Excluding Age 0-1
Dep.Var: $\lvert FE^{\log}_{t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t \in (0.5,1))$	-0.022^c	-0.011			
	(0.013)	(0.015)			
$\mathbb{1}(\text{Age}_t = 2)$	-0.069^a	-0.068^a	-0.048^a	-0.058^a	
	(0.010)	(0.010)	(0.009)	(0.011)	
$\mathbb{1}(\text{Age}_t = 3)$	-0.100^a	-0.093^a	-0.079^a	-0.084^a	-0.027^a
	(0.009)	(0.010)	(0.010)	(0.011)	(0.005)
$\mathbb{1}(\text{Age}_t = 4)$	-0.124^a	-0.115^a	-0.103^a	-0.106^a	-0.049^a
	(0.009)	(0.010)	(0.010)	(0.011)	(0.006)
$\mathbb{1}(\text{Age}_t = 5)$	-0.136^a	-0.122^a	-0.115^a	-0.113^a	-0.056^a
	(0.009)	(0.010)	(0.009)	(0.011)	(0.006)
$\mathbb{1}(\text{Age}_t = 6)$	-0.135^a	-0.119^a	-0.114^a	-0.110^a	-0.053^a
	(0.009)	(0.010)	(0.009)	(0.011)	(0.006)
$\mathbb{1}(\text{Age}_t = 7)$	-0.142^a	-0.126^a	-0.121^a	-0.116^a	-0.060^a
	(0.009)	(0.010)	(0.009)	(0.011)	(0.006)
$\mathbb{1}(\text{Age}_t = 8)$	-0.144^a	-0.126^a	-0.123^a	-0.116^a	-0.060^a
	(0.010)	(0.011)	(0.010)	(0.012)	(0.006)
$\mathbb{1}(\text{Age}_t = 9)$	-0.146^a	-0.128^a	-0.125^a	-0.118^a	-0.062^a
	(0.009)	(0.011)	(0.009)	(0.012)	(0.006)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.148^a	-0.126^a	-0.127^a	-0.118^a	-0.062^a
	(0.009)	(0.011)	(0.009)	(0.012)	(0.007)
$\log(\text{Emp})_t$	-0.021^a	-0.024^a	-0.020^a	-0.023^a	-0.020^a
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
$\log(\text{Parent Emp})_t$	0.001	0.001	0.001	0.001	-0.000
	(0.001)	(0.003)	(0.001)	(0.003)	(0.003)
Industry-year FE	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y
Firm FE		Y		Y	Y
N	128429	123111	126914	121671	120217
R^2	0.122	0.366	0.118	0.362	0.361

Notes: Standard errors are clustered at the business group level, c 0.10 b 0.05 a 0.01. In columns 1-2, we use age one firms that entered in the second half of the founding year as the base group (age $\in (0, 0.5)$) and include an additional dummy variable indicating whether the age one firms entered in the first half of the founding year (age $\in (0.5, 1)$). In column 3-4, we exclude age one affiliated firms that entered in the second half of the founding year. Column 5 excludes all age one firms.

1.3.6 Idiosyncratic Shocks or Heterogeneous Exposure to Aggregate Shocks?

Though we have shown that our results are robust to using the "residual forecast errors", which arguably tease out the systematic forecast errors due to aggregate shocks, they may still be affected by the aggregate economy since firms may have heterogeneous exposure to aggregate shocks (David et al., 2019). In this subsection, we construct alternative measures of residual forecast errors to tease out such heterogeneous exposure.

There are multiple mechanisms through which firms have heterogeneous exposure to aggregate shocks. David et al. (2019) show that, all else equal, (1) labor intensive firms are more exposed to cyclical movements in wages (2) firms facing a high demand elasticity (setting a lower markup) respond more strongly to aggregate shocks, and (3) high-quality products are more cyclical since households tend to consume higher quality goods in booms due to non-homothetic preferences. To account for such heterogeneous exposure, we construct an alternative residual forecast error by running the following regression

$$FE^{\log}_{it,t+1} = \delta^{\text{labor}}_b \times \delta_{ct} + \delta^{\text{markup}}_b \times \delta_{ct} + \delta^{\text{quality}}_b \times \delta_{ct} + \delta_{st} + \hat{\epsilon}^{FE,\log}_{it,t+1}$$

where $\delta^{\text{labor}}_b \times \delta_{ct}$ indicates a set of labor-share-bin-country-year fixed effects. The labor share bins are obtained by dividing our sample into ten equally-sized bins based on the firms' labor share (wage bill divided by total sales). We define $\delta^{\text{markup}}_b \times \delta_{ct}$ and $\delta^{\text{quality}}_b \times \delta_{ct}$ in similar ways. We use the ratio of total sales to material costs as a measure of the markup and workers' average wage as a measure of output quality. The markup measure is proportional to price over marginal cost as long as (1) the output elasticity with respect to materials is constant and (2) materials are a flexible input, i.e., not subject to adjustment frictions (de Loecker and Warzynski, 2012). We use workers' wage to approximate firm output quality, as previous studies show that firms producing high-quality output tend to be more skill intensive. (see, for example, Verhoogen (2008); Fieler et al. (2018)) Finally, δ_{st} is a set of industry-year fixed effects, which we also include when calculating the baseline residual forecast errors.

This specification captures heterogeneous responses to aggregate shocks (country-year fixed effects) based on firm characteristics such as labor share, markup and output quality. It includes substantially more fixed effects compared to the regression we use to obtain the baseline residual forecast errors (only country-year and industry-year fixed effects). The expanded set of fixed effects explains 23% of the variation in the raw forecast errors. The residuals, capturing forecast errors due to idiosyncratic shocks, still maintain 77% of the variation in the raw forecast errors.

In Table OA.11, we replicate regressions in Table 3 of the paper. Though the age coefficients are smaller, they are still significantly negative and are about 85% of those estimated with raw forecast errors. Note that the number of observations are smaller than in the paper, as much more singletons are dropped when we estimate the residual forecast errors due to the added fixed effects.

Table OA.11: Age effects on the absolute value of alternative residual forecast errors, where we have purged an expanded set of fixed effects.

Sample:	All Firms			Survivors	Manufacturing
Dep.Var: $\lvert\hat{\epsilon}^{\log}_{FE,t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t = 2)$	-0.034^a	-0.035^a	-0.037^a	-0.048^a	-0.019
	(0.008)	(0.008)	(0.009)	(0.012)	(0.012)
$\mathbb{1}(\text{Age}_t = 3)$	-0.051^a	-0.047^a	-0.043^a	-0.053^a	-0.028^b
	(0.008)	(0.008)	(0.009)	(0.012)	(0.012)
$\mathbb{1}(\text{Age}_t = 4)$	-0.081^a	-0.075^a	-0.065^a	-0.067^a	-0.052^a
	(0.008)	(0.008)	(0.009)	(0.012)	(0.012)
$\mathbb{1}(\text{Age}_t = 5)$	-0.090^a	-0.082^a	-0.069^a	-0.081^a	-0.052^a
	(0.008)	(0.008)	(0.009)	(0.013)	(0.012)
$\mathbb{1}(\text{Age}_t = 6)$	-0.096^a	-0.087^a	-0.071^a	-0.078^a	-0.056^a
	(0.008)	(0.008)	(0.009)	(0.014)	(0.012)
$\mathbb{1}(\text{Age}_t = 7)$	-0.105^a	-0.093^a	-0.078^a	-0.098^a	-0.065^a
	(0.008)	(0.008)	(0.009)	(0.015)	(0.012)
$\mathbb{1}(\text{Age}_t = 8)$	-0.107^a	-0.095^a	-0.077^a	-0.090^a	-0.064^a
	(0.008)	(0.008)	(0.009)	(0.016)	(0.012)
$\mathbb{1}(\text{Age}_t = 9)$	-0.114^a	-0.099^a	-0.081^a	-0.093^a	-0.071^a
	(0.008)	(0.008)	(0.009)	(0.017)	(0.013)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.123^a	-0.099^a	-0.078^a	-0.090^a	-0.065^a
	(0.008)	(0.008)	(0.009)	(0.019)	(0.013)
$\log(\text{Emp})_t$		-0.019^a	-0.022^a	-0.028^a	-0.023^a
		(0.001)	(0.002)	(0.005)	(0.003)
$\log(\text{Parent Emp})_t$		-0.001	0.006^b	0.009	0.007^a
		(0.001)	(0.003)	(0.008)	(0.002)
Industry-year FE	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y
Firm FE			Y	Y	Y
N	98102	97968	93145	16494	61000
R^2	0.075	0.094	0.352	0.369	0.340

Notes: Standard errors are clustered at the business group level, c 0.10 b 0.05 a 0.01.

1.4 Robustness Checks for Fact 2

1.4.1 Auto-correlations using alternative forecast errors

Table OA.12: Correlation of $FE_{t,t+1}$ and $FE_{t-1,t}$, overall and by age group

Sample	All ages	Age 2-4	Age 5-7	Age $\geq$ 8
All industries				
$corr(FE^{\text{pct}}_{t,t+1}, FE^{\text{pct}}_{t-1,t})$	0.105	0.137	0.120	0.093
	[96967]	[10578]	[13875]	[72514]
$corr(\hat{\epsilon}^{FE,\log}_{t,t+1}, \hat{\epsilon}^{FE,\log}_{t-1,t})$	0.113	0.154	0.141	0.092
	[96194]	[10373]	[13764]	[72057]
$corr(\hat{\epsilon}^{FE,\text{pct}}_{t,t+1}, \hat{\epsilon}^{FE,\text{pct}}_{t-1,t})$	0.087	0.122	0.111	0.070
	[96707]	[10541]	[13838]	[72328]
Manufacturing				
$corr(FE^{\text{pct}}_{t,t+1}, FE^{\text{pct}}_{t-1,t})$	0.108	0.172	0.116	0.089
	[60364]	[5906]	[8623]	[45835]
$corr(\hat{\epsilon}^{FE,\log}_{t,t+1}, \hat{\epsilon}^{FE,\log}_{t-1,t})$	0.118	0.177	0.139	0.092
	[60049]	[5817]	[8580]	[45652]
$corr(\hat{\epsilon}^{FE,\text{pct}}_{t,t+1}, \hat{\epsilon}^{FE,\text{pct}}_{t-1,t})$	0.092	0.160	0.103	0.070
	[60289]	[5895]	[8612]	[45782]

Notes: $FE^{\text{pct}}_{t,t+1}$ is the percentage deviation of realized sales from expected sales. The other two measures, $\hat{\epsilon}^{FE,\log}_{t,t+1}$ and $\hat{\epsilon}^{FE,\text{pct}}_{t,t+1}$, are the residual forecast errors, which we obtain by regressing $FE^{\log}_{t,t+1}$ and $FE^{\text{pct}}_{t,t+1}$ on a set of industry-year and country-year fixed effects.. Age is measured at the end of year t. The manufacturing subsample is constructed in the same way as the previous section. Number of observations used for each correlation is shown in the brackets below.

1.4.2 AR(1) Models with Age Interactions

In this section, we perform several robustness checks of the regressions in Table 5 of the paper. We first replace the log forecast errors with alternative definitions of forecast errors. Table OA.13 uses percentage forecast errors, while Table OA.14 uses residual log forecast errors. The results in Table OA.14 are almost identical to those obtained using log forecast errors, while the magnitudes of the estimates in Table OA.13 are slightly smaller. Next, we exclude firms that use current sales as their sales forecasts for the next year and re-run the regressions in Table 5. The results are very similar (see Table OA.15).

Table OA.13: AR(1) regressions with Age Interactions, Percentage Forecast Errors

Sample:	All Affiliates				Manufacturing			
Dep.Var: $FE^{\text{pct}}_{t+1,t+2}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$FE^{\text{pct}}_{t,t+1}$	0.071^a	0.068^a	0.106^a	0.098^a	0.085^a	0.080^a	0.112^a	0.103^a
	(0.014)	(0.013)	(0.018)	(0.017)	(0.020)	(0.020)	(0.024)	(0.024)
$\times \max\{\text{Age}_t, 10\}$	-0.005^a		-0.008^a		-0.007^a		-0.009^a	
	(0.002)		(0.002)		(0.003)		(0.003)	
$\times \log(\text{Age}_t)$		-0.015^a		-0.023^a		-0.024^a		-0.029^a
		(0.006)		(0.007)		(0.009)		(0.010)
$\log(\text{Emp})_t$	-0.003^a	-0.003^a	-0.004^a	-0.004^a	-0.004^a	-0.004^a	-0.005^a	-0.005^a
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\log(\text{Parent Emp})_t$	-0.010^b	-0.010^b	-0.009^b	-0.009^b	-0.009	-0.009	-0.012^c	-0.012^c
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Business Group FE	Y	Y			Y	Y		
Busi.Group-Age FE			Y	Y			Y	Y
N	93971	93971	85278	85278	58862	58862	52720	52720
R^2	0.181	0.180	0.250	0.250	0.198	0.198	0.266	0.266

Notes: Standard errors are clustered at the business group level. c 0.10 b 0.05 a 0.01. The “manufacturing” subsample refers to affiliated firms that are in manufacturing sectors.

Table OA.14: AR(1) regressions with Age Interactions, Residual Log Forecast Errors

Sample:	All Affiliates				Manufacturing			
Dep.Var: $\hat{\epsilon}^{FE,\log}_{t+1,t+2}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\epsilon}^{FE,\log}_{t,t+1}$	0.106^a	0.101^a	0.138^a	0.128^a	0.118^a	0.116^a	0.147^a	0.144^a
	(0.014)	(0.014)	(0.019)	(0.018)	(0.020)	(0.020)	(0.026)	(0.026)
$\times \max\{\text{Age}_t, 10\}$	-0.006^a		-0.009^a		-0.009^a		-0.011^a	
	(0.002)		(0.002)		(0.003)		(0.003)	
$\times \log(\text{Age}_t)$		-0.019^a		-0.025^a		-0.030^a		-0.035^a
		(0.006)		(0.007)		(0.009)		(0.011)
$\log(\text{Emp})_t$	0.003^a	0.003^a	0.002^c	0.002^c	0.002	0.002	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\log(\text{Parent Emp})_t$	-0.010^b	-0.010^b	-0.010^b	-0.010^b	-0.011^c	-0.011^c	-0.014^b	-0.014^b
	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Business Group FE	Y	Y			Y	Y		
Busi.Group-Age FE			Y	Y			Y	Y
N	93478	93478	84839	84839	58630	58630	52510	52510
R^2	0.097	0.097	0.168	0.168	0.111	0.111	0.182	0.182

Notes: Standard errors are clustered at the business group level. c 0.10 b 0.05 a 0.01. The “manufacturing” subsample refers to affiliated firms that are in manufacturing sectors.

Table OA.15: AR(1) regressions with Age Interactions, excluding firms with zero expected growth rates

Sample:	All Affiliates				Manufacturing			
Dep.Var: $FE^{\log}_{t+1,t+2}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$FE^{\log}_{t,t+1}$	0.101^{a}	0.096^{a}	0.098^{a}	0.093^{a}	0.111^{a}	0.109^{a}	0.105^{a}	0.104^{a}
	(0.014)	(0.014)	(0.014)	(0.014)	(0.019)	(0.019)	(0.019)	(0.020)
$\times \max\{\text{Age}_t, 10\}$	-0.005^{a}		-0.004^{a}		-0.007^{a}		-0.007^{a}	
	(0.002)		(0.002)		(0.002)		(0.002)	
$\times \log(\text{Age}_t)$		-0.014^{b}		-0.013^{b}		-0.025^{a}		-0.023^{b}
		(0.006)		(0.006)		(0.009)		(0.009)
$\log(\text{Emp})_t$			0.002^{b}	0.002^{b}			0.001	0.001
			(0.001)	(0.001)			(0.001)	(0.001)
$\log(\text{Parent Emp})_t$			-0.011^{a}	-0.011^{a}			-0.011^{c}	-0.011^{c}
			(0.004)	(0.004)			(0.006)	(0.006)
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Business Group FE	Y	Y	Y	Y	Y	Y	Y	Y
Age FE	Y	Y	Y	Y	Y	Y	Y	Y
N	92871	92871	91378	91378	58214	58214	57646	57646
R^2	0.206	0.206	0.208	0.208	0.231	0.231	0.233	0.233

Notes: Standard errors are clustered at the business group level. c 0.10 b 0.05 a 0.01. The “manufacturing” subsample refers to affiliated firms that are in manufacturing sectors.

1.5 Robustness Checks for Fact 3

Table OA.16: AR(1) coef and horse race between country characteristics

	Dep.Var: $FE_{t+1,t+2}$							
$FE^{\log}_{t,t+1}$	0.1488^a	0.1338^a	0.1175^a	0.1160^a	0.1632^a	0.1410^a	0.1272^a	0.1237^a
	(0.0222)	(0.0174)	(0.0176)	(0.0177)	(0.0221)	(0.0180)	(0.0182)	(0.0181)
× Management Score (WMS 2015)	-0.0130				-0.0127			
	(0.0084)				(0.0082)			
× Time Diff from Japan		0.0163^b		0.0301^a		0.0151^b		0.0291^a
		(0.0073)		(0.0087)		(0.0072)		(0.0087)
× log GDP p.c. 1995			-0.0110^c	-0.0272^a			-0.0110^c	-0.0269^a
			(0.0067)	(0.0080)			(0.0066)	(0.0080)
× $\log(\text{Age})_t$	-0.0291^a	-0.0282^a	-0.0220^a	-0.0220^a				
	(0.0087)	(0.0070)	(0.0070)	(0.0070)				
× $\min\{Age, 10\}$					-0.0103^a	-0.0091^a	-0.0076^a	-0.0073^a
					(0.0025)	(0.0021)	(0.0021)	(0.0021)
Industry-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Busi.Group-Age FE	Y	Y	Y	Y	Y	Y	Y	Y
N	53433	86271	86271	86271	53433	86271	86271	86271
R^2	0.284	0.270	0.270	0.271	0.284	0.270	0.270	0.271

Notes: Standard errors are clustered at the business group level. Significance levels: c 0.1, b 0.05, a 0.01. Management score is from the World Management Survey up to 2015 . Management score, time zone differences and log GDP per capita are all standardized to faciliate interpretation of the coefficients.

2 Theory Appendix

2.1 Full Information Rational Expectation Models

In this subsection, we derive the expression of the forecast error in the full information rational expectation (FIRE) model. We calculate the logarithm of *realized* sales in period t as

$$\log(R_n(\theta, \varphi_{n-1})) = (\sigma - 1)\left[\log(\sigma - 1) - \log(\sigma)\right] + \log(Y) + (\sigma - 1)\log(P)$$
$$+\theta + (\sigma - 1)\left[b(\varphi_{n-1}, n) - \log(w)\right] + \frac{\sigma - 1}{\sigma}\log(\varphi_n),$$

where

$$b(\varphi_{n-1}, n) \equiv E\left(\varphi_n^{\frac{\sigma-1}{\sigma}} | \varphi_{n-1}, n\right).$$

Since the firm knows θ in the FIRE model, the logarithm of forecasted sales is

$$\log(R_n(\theta, \varphi_{n-1})) = (\sigma - 1)\left[\log(\sigma - 1) - \log(\sigma)\right] + \log(Y) + (\sigma - 1)\log(P)$$
$$+\theta + (\sigma - 1)\left[b(\varphi_{n-1}, n) - \log(w)\right] + b(\varphi_{n-1}, n),$$

which leads to

$$FE^{\log}_{n-1,n} = \frac{(\sigma-1)\nu_n}{\sigma} - \frac{(\sigma-1)^2\sigma^2_{\nu_n}}{2\sigma^2}. \tag{2}$$

Thus, we have

$$Cov\left(FE^{\log}_{n-1,n}, FE^{\log}_{n,n+1}\right) = Cov\left(\frac{(\sigma-1)\nu_n}{\sigma}, \frac{(\sigma-1)\nu_{n+1}}{\sigma}\right) = 0.$$

Therefore, forecast errors are serially uncorrelated in FIRE models.

2.2 Full Information Rational Expectation Models with Endogenous Exits

In this subsection, we consider the case in which incumbent firms can choose to exit after observing the its productivity shock and the demand draw. There are two sub-cases to discuss. First, following the same timing assumption adopted in the paper, we assume that the firm observes its productivity shock at age $n-1$ when choosing to stay at age n. In this case, there is an exit cutoff on the productivity shock $\bar{\varphi}_{n-1}(\theta)$ (depending on θ) below which incumbent firms exit. Thus, incumbents that have survived at both ages n and $n+1$ must satisfy

$$\log\varphi_{n-1} = \mu_\varphi + \rho\log\varphi_{n-2} + \nu_{n-1} \geq \log\left(\bar{\varphi}_{n-1}(\theta)\right), \quad \log\varphi_n = \mu_\varphi + \rho\log\varphi_{n-1} + \nu_n \geq \log\left(\bar{\varphi}_n(\theta)\right), \tag{3}$$

Conditioning on $\log\varphi_{n-2}$ and survival at both ages n and $n+1$, there is a negative correlation between ν_{n-1} and ν_n implied by equation (3) as $\log\varphi_{n-1} = \mu_\varphi + \rho\log\varphi_{n-2} + \nu_{n-1}$. The intuition is that a better contemporaneous productivity innovation at age $n-1$ (that pushes up the productivity realization at age $n-1$) makes survival at age $n+1$ (that depends on the productivity realization at age n) easier, which implies worse productivity innovations at age n on average. This leads to a negative correlation between the contemporaneous productivity innovations at ages $n-1$ and n, conditioning on survival. However, the autocovariance of the forecast errors at ages n and $n+1$ is still zero, as the productivity innovation at age $n+1$ that enter into the forecast error at age $n+1$

is still random conditioning on the survival at ages $n-1$ and n:

$$\begin{aligned}
& Cov(FE^{\log}_{n-1,n}, FE^{\log}_{n,n+1}|\text{surviving at both ages } n \text{ and } n+1) \\
= \; & Cov\Big(\frac{(\sigma-1)\nu_n}{\sigma}, \frac{(\sigma-1)\nu_{n+1}}{\sigma}\Big| \log\varphi_{n-1} \geq \log\left(\bar{\varphi}_{n-1}(\theta)\right), \mu_\varphi + \rho\log\varphi_{n-1} + \nu_n \geq \log\left(\bar{\varphi}_n(\theta)\right)\Big) \\
= \; & 0.
\end{aligned}$$

Second, we consider the sub-case that the firm observes its productivity shock at age n when choosing to stay at age n which is different from the assumption used in the paper but common in most firm dynamics models (e.g., Hopenhayn (1992)). In this case, then exit cutoff at age n is related to the productivity shock at age n or $\bar{\varphi}_n(\theta)$ below which incumbent firms exit. Again, a better contemporaneous productivity innovation at age n makes survival at age $n+1$ easier, which implies worse productivity innovations at age $n+1$ on average. This leads to a negative correlation between the contemporaneous productivity innovations at ages $n-1$ and n, conditioning on survival. Thus, survivors at ages n and $n+1$ must satisfy

$$\mu_\varphi + \rho\log\varphi_{n-1} + \nu_n \geq \log\left(\bar{\varphi}_n(\theta)\right), \quad \mu_\varphi + \rho\log\varphi_n + \nu_{n+1} \geq \log\left(\bar{\varphi}_{n+1}(\theta)\right), \tag{4}$$

Conditioning on $\log\varphi_{n-1}$ and survival at both ages n and $n+1$, there is a negative correlation between ν_n and ν_{n+1} implied by equation (4) as $\log\varphi_n = \mu_\varphi + \rho\log\varphi_{n-1} + \nu_n$. Therefore, the correlation of forecast errors becomes

$$\begin{aligned}
& Cov(FE^{\log}_{n-1,n}, FE^{\log}_{n,n+1}|\text{surviving at both ages } n \text{ and } n+1) \\
= \; & Cov\Big(\frac{(\sigma-1)\nu_n}{\sigma}, \frac{(\sigma-1)\nu_{n+1}}{\sigma}\Big| \mu_\varphi + \rho\log\varphi_{n-1} + \nu_n \geq \log\left(\bar{\varphi}_n(\theta)\right), \mu_\varphi + \rho\log\varphi_n + \nu_{n+1} \geq \log\left(\bar{\varphi}_{n+1}(\theta)\right)\Big) \\
< \; & 0.
\end{aligned}$$

Finally, the proof would be the same (with changes in notations), if we assume that the demand shifter θ follows an AR(1) process and the productivity shock is time-invariant. In total, the FIRE model cannot be used to rationalize forecast errors made in two consecutive periods are positively correlated.

2.3 ε is a Real Shock as in Jovanovic (1982)

In this section, we show that forecast errors made by firms in Jovanovic (1982) are serially uncorrelated—a property that rational expectations models with *full information* also inherit. In order to show this property, we modify our model presented in the paper in the following way. We assume that the firm-specific demand shifter, $a_t(\omega)$, is the sum of a time-invariant permanent demand draw $\theta(\omega)$ and a transitory demand shock $\varepsilon_t(\omega)$ as in Arkolakis et al. (2018):

$$a_t(\omega) = \theta(\omega) + \varepsilon_t(\omega). \tag{5}$$

Firms understand that $\theta(\omega)$ is drawn from a normal distribution $N\left(\bar{\theta}, \sigma_\theta^2\right)$, and the independently and identically distributed (i.i.d.) transitory demand shock, $\varepsilon_t(\omega)$, is drawn from another normal distribution $N\left(0, \sigma_\varepsilon^2\right)$. We assume that the firm observes the sum of the two demand components, $a_t(\omega)$, at the end of each period, not each of them separately. Thus, the firm needs to learn about its permanent demand every period by forming an posterior belief about the distribution of θ. In summary, we drop the "pure" informational noise from the model and assume that the firm cannot differentiate the permanent demand draw from the transitory demand shock. As a result, the realized overall demand shifters, $a_1, a_2, \ldots, a_t$, become the noisy signals for the permanent demand draw $\theta(\omega)$. The crucial difference here is that the transitory demand shock now acts as both an informational noise *and* as a "real" shock that directly affects the firm's overall demand.

We modify the firm's belief updating process as follows. Since both the prior and the realized demand shifters are normally distributed, the posterior belief is also normally distributed. A firm that is $n+1$ years old has observed the realized demand shifters in the past n periods: $a_1, a_2, \ldots, a_n$, the Bayes' rule implies that the posterior belief about θ is normally distributed with mean μ_n and variance σ_n^2 where

$$\mu_n = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + n\sigma_\theta^2}\bar{\theta} + \frac{n\sigma_\theta^2}{\sigma_\varepsilon^2 + n\sigma_\theta^2}\bar{a}_n,\ \sigma_n^2 = \frac{\sigma_\varepsilon^2 \sigma_\theta^2}{\sigma_\varepsilon^2 + n\sigma_\theta^2}. \tag{6}$$

The history of signals $(a_1, a_{2,} \ldots, a_n)$ is summarized by age n and the average demand shifter:

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

Therefore, the firm believes that the overall demand shifter in period $t+1$, $a_{n+1} = \theta + \varepsilon_{n+1}$, has a normal distribution with mean μ_n and variance $\sigma_n^2 + \sigma_\varepsilon^2$. The difference from the paper is that it is the average demand shifter $\bar{a}_n$ (not $\bar{s}_n$) that is the firm's state variable.

We study the firm's static optimization problem under the modified assumptions now. Given the belief about a_n, an age-n firm chooses employment level l_n to maximize its expected per-period profit at age n, $E_{a_n,\varphi_n|\bar{a}_{n-1},\varphi_{n-1},n}(\pi_n)$. The realized per-period profit at age n is

$$\pi_n = p_n(a_n)\varphi_n l_n - w \times l_n - wf.$$

Firms set the price after observing the realized demand a_n and the productivity shock φ_n to sell all the output. Maximizing $E_{a_n,\varphi_n|\bar{a}_{n-1},\varphi_{n-1},n}(\pi_n)$, the optimal employment of an age-n in period t is[3]

$$l_t = \left(\frac{\sigma-1}{\sigma}\right)^\sigma \left(\frac{b(\varphi_{n-1}, \bar{a}_{n-1}, n-1)}{w}\right)^\sigma Y P^{\sigma-1}, \tag{7}$$

where

$$\begin{aligned} b(\varphi_{n-1}, \bar{a}_{n-1}, n-1) &\equiv E\left(e^{\frac{a_t}{\sigma}} \varphi_n^{\frac{\sigma-1}{\sigma}} | \varphi_{n-1}, \bar{a}_{n-1}, n\right) \\ &= \exp\left\{\frac{\mu_{n-1}}{\sigma} + \frac{\sigma_{n-1}^2 + \sigma_\varepsilon^2}{2\sigma^2} + \frac{\sigma-1}{\sigma}((1-\rho)\mu_\varphi + \rho\log\varphi_{n-1}) + \frac{(\sigma-1)^2\sigma_{\nu_n}^2}{2\sigma^2}\right\}, \end{aligned} \tag{8}$$

and n is the firm's age. As a result, the logarithm of realized sales and the logarithm of forecasted sales of an age-n firm are

$$\log(R_n(\theta)) = \log\left(\frac{Y}{P^{1-\sigma}}\right) + \frac{a_t}{\sigma} + \frac{\sigma-1}{\sigma}\log(\varphi_n) + (\sigma-1)\log b(\varphi_{n-1}, \bar{a}_{n-1}, n-1) + (\sigma-1)\left[\log\left(\frac{\sigma-1}{\sigma w}\right)\right], \tag{9}$$

and

$$\log\left(E_{n-1}(R_n)\right) = \log\left(\frac{Y}{P^{1-\sigma}}\right) + \sigma\log b(\varphi_{n-1}, \bar{a}_{n-1}, n-1) + (\sigma-1)\left[\log\left(\frac{\sigma-1}{\sigma w}\right)\right].$$

The resulting log forecast error of sales is

$$FE_{n-1,n}^{\log} = \frac{(\sigma-1)\nu_{n+1}}{\sigma} - \frac{(\sigma-1)^2\sigma_{\nu_{n+1}}^2}{2\sigma^2} + \frac{\varepsilon_t + (\theta - \mu_{n-1})}{\sigma} - \frac{\sigma_{n-1}^2 + \sigma_\varepsilon^2}{2\sigma^2},$$

[3] Since we always consider the steady state, time script t does not play a role in the optimization problem.

which can be rewritten as

$$FE^{\log}_{n-1,n} = \frac{(\sigma-1)\nu_{n+1}}{\sigma} - \frac{(\sigma-1)^2\sigma^2_{\nu_{n+1}}}{2\sigma^2}$$
$$+\frac{(1-\zeta(n-1,\lambda))(\theta-\bar{\theta})+\varepsilon_t-\zeta(n-1,\lambda)\frac{\sum_{i=t-n+1}^{t-1}\varepsilon_i}{n-1}}{\sigma} - \frac{\sigma^2_{n-1}+\sigma^2_{\varepsilon}}{2\sigma^2}, \quad (10)$$

where

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

The autocovariance of (log) sales forecast errors is simply

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

Therefore, the "real" demand shock that also acts as an informational noise cannot generate non-zero autocorrelation of forecast errors.

For forecast errors made at ages n and $n+1$, they share two common components in equation (10): $\theta - \bar{\theta}$ and $\sum_{i=t-n+1}^{t-1}\varepsilon_i$. Thus, if the prior mean of θ is below (or above) the actual permanent demand shifter, the firm would make positive (or negative) forecast errors at ages n and $n+1$. Similarly, if the sum of the past transitory shocks (up to age $n-1$) is negative (or positive), the firm would make positive (or negative) forecast errors at ages n and $n+1$. In any case, the forecast errors are positively autocorrelated. This is exactly the reason why forecast errors are positively autocorrelated in the paper, as the transitory (information) shocks do not enter into the realization of overall demand shifter. However, as the transitory demand shock, ε_t also enter into the realized demand shifter, there is the third term ε_t which enters into $FE^{\log}_{n-1,n}$ positively but into $FE^{\log}_{n,n+1}$ negatively. The existence of the payoff-relevant noise in Jovanovic (1982), ε_t, causes the negative autocorrelation of forecast errors. And, this additional force perfectly offsets the two forces that cause the positive autocorrelation of forecast errors discussed above.

2.3.1 Alternative intuition

Another way to gain some intuition about the uncorrelated forecast errors in Jovanovic (1982) is that the Bayesian updating with an unbiased prior yields the best linear unbiased estimator

(BLUE) for the overall demand shifter at age n, $a_n = \theta + \varepsilon_n$. To see this, recall that

$$E(\theta | a_{n-1}, a_{n-2}, \ldots, a_1) = \mu_{n-1}.$$

According to Hayashi (2000) Proposition 2.7, the conditional expectation is the "best predictor" (i.e., minimizes mean squared error). Since μ_{n-1} is a weighted average of the prior $\bar{\theta}$ and previous signals $a_{n-1}, \ldots, a_1$, it must be the "best linear predictor". Note that this property also holds if the goal is to predict $a_n = \theta + \varepsilon_n$, since ε_n is independent of past shocks.

In Jovanovic (1982), the (log) forecast error of sales will be proportional to $a_n - \mu_{n-1} = \theta + \varepsilon_n - \mu_{n-1}$. The previous forecast error is proportional $a_{n-1} - \mu_{n-2}$, a linear combination of $a_{n-1}, \ldots, a_1$. Since $E(a_n - \mu_{n-1} | a_{n-1}, \ldots, a_1) = 0$, we must have $E(a_n - \mu_{n-1} | a_{n-1} - \mu_{n-2}) = 0$.

When ε_n is *payoff-irrelevant* as in our model, the forecast errors are defined as $\theta - \mu_{n-1}$ instead of $a_n - \mu_{n-1}$. Therefore, we do not have $E(\theta - \mu_{n-1} | \theta - \mu_{n-2}) = 0$, though from the previous discussion we know that $E(\theta - \mu_{n-1} | a_{n-1} - \mu_{n-2}) = 0$. Consider regressing $\theta - \mu_{n-1}$ on $\theta - \mu_{n-2}$. If we use $a_{n-1} - \mu_{n-2} = \theta + \varepsilon_{n-1} - \mu_{n-2}$, then we will obtain a zero coefficient. Regressing the current forecast error on the previous forecast error defined in our model, $\theta - \mu_{n-2}$, creates a "non-classic measurement error" in the regressor. The direction of the "bias" can be seen from the covariance below:

$$Cov(\theta - \mu_{n-1}, \theta - \mu_{n-2}) = Cov(\theta - \mu_{n-1}, a_{n-1} - \mu_{n-2} - \varepsilon_{n-1}) = Cov(\mu_{n-1}, \varepsilon_{n-1}).$$

Since ε_{n-1} enters μ_{n-1} positively, the covariance is positive. Therefore the auto-covariance and the AR(1) coefficient of the forecast errors will be positive.

2.4 ε is a Real Shock and θ is time-varying

In this subsection, we show that the sales forecast errors are still uncorrelated over time, even when we assume that the permanent demand draw, θ, is time-varying. In particular, we assume that θ_t follows an AR(1) structure:

$$\theta_t = \rho\theta_{t-1} + \zeta_t$$

and

$$a_t = \theta_t + \varepsilon_t.$$

In addition, we make the assumption an age-n firm only observes $a_{t-n+1}, ..., a_{t-1}$ up to the beginning of period t (i.e., $n-1$ signals).

The forecast error of firm sales still consists of two parts: the demand-side error and the supply-error:

$$\begin{aligned} FE^{log}_{t,t+1} &\equiv \log R_{t+1} - \log E_t R_{t+1} = \frac{a_{t+1}}{\sigma} + \frac{\sigma-1}{\sigma}\log\varphi_{t+1} - \log E_t(e^{a_{t+1}/\sigma}\varphi_{t+1}^{\frac{\sigma-1}{\sigma}}) \\ &= \underbrace{\frac{a_{t+1}}{\sigma} - \log E_t(e^{a_{t+1}/\sigma})}_{FE^d_{t,t+1}} + \underbrace{\frac{\sigma-1}{\sigma}\log\varphi_{t+1} - \log E_t(\varphi_{t+1}^{\frac{\sigma-1}{\sigma}})}_{FE^s_{t,t+1}} \\ &= \frac{\theta_{t+1} - \mu_{t+1} + \varepsilon_{t+1}}{\sigma} - \frac{\sigma_t^2}{2\sigma^2} + \frac{(\sigma-1)\nu_{t+1}}{\sigma} - \frac{(\sigma-1)^2\sigma^2_{\nu_{t+1}}}{2\sigma^2} \qquad (11) \\ &= \frac{e_{t+1} + \varepsilon_{t+1}}{\sigma} - \frac{\sigma_t^2}{2\sigma^2} + \frac{(\sigma-1)\nu_{t+1}}{\sigma} - \frac{(\sigma-1)^2\sigma^2_{\nu_{t+1}}}{2\sigma^2}, \qquad (12) \end{aligned}$$

where $\mu_{t+1} \equiv E_t\theta_{t+1}$ is the forecast of θ_{t+1} made in period t and e_{t+1} is the forecast error of θ_{t+1}. The term of σ_t^2 is the variance of forecast errors in period $t+1$. Variable ν_{t+1} and the term of $\sigma^2_{\nu_{t+1}}$ are the productivity innovation and its variance in period $t+1$. Note that both σ_t^2 and $\sigma^2_{\nu_{t+1}}$ are non-stochastic terms and thus uncorrelated over time. Moreover, the productivity innovation is i.i.d. both over time and across firms (and independent of demand innovations), thus we have

$$Cov\left(\frac{(\sigma-1)\nu_{t+1}}{\sigma}, \frac{(\sigma-1)\nu_t}{\sigma}\right) = 0,$$

and

$$Cov\left(FE^{log}_{t-1,t}, FE^{log}_{t,t+1}\right) = Cov\left(\frac{e_t + \varepsilon_t}{\sigma}, \frac{e_{t+1} + \varepsilon_{t+1}}{\sigma}\right) = Cov\left(\frac{a_t - \mu_t}{\sigma}, \frac{e_{t+1} + \varepsilon_{t+1}}{\sigma}\right).$$

Now, we calculate the correlation between the forecast error of θ_{t+1} (i.e., e_{t+1}) and the realized demand shock a_i where $i \in t-n+2, t-n+3, ..., t$ of age-n firms. The case discussed in the subsection is a variant of Muth (1960)'s model. The optimal forecasting rule concerning θ_t follows:

$$\mu_t = (\rho - K_{t-1})\mu_{t-1} + K_{t-1}a_{t-1},$$

where K_{t-1} is is the Kalman gain. Note that $forecast_t$ is the belief formed at the beginning of period t without observing a_t. To be consistent with the notation in earlier sections, we denote forecast for θ_t using μ_t.

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

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

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

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

Given Σ_t, we can use the 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 (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

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

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 $s \leq t$)

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

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:

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

where $\sigma_\theta^2 = \equiv \frac{\sigma_\zeta^2}{1-\rho^2}$ is the variance of the demand shocks in the steady state. 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. In particular, we have

$$
\begin{aligned}
& \sigma^2 Cov\left(FE_{t-1,t}^{log}, FE_{t,t+1}^{log}\right) \\
= \;& Cov(a_{t+1} - \mu_{t+1}, a_t - \mu_t) \\
= \;& Cov(e_{t+1} + \varepsilon_{t+1}, a_t - \mu_t) \\
= \;& Cov(e_{t+1}, a_t - K\sum_{j=0}^{\infty}(\rho-K)^j a_{t-1-j}) = 0,
\end{aligned}
$$

as $cov(e_{t+1}, a_s) = 0$ for any $s \leq t$ and the transitory shock ε_{t+1} is independent of any shock that has been relied up to period t. Therefore, the (log) forecast errors of sales are serially uncorrelated, even if the demand shock follows an AR(1) process.

3 Additional Quantitative Results

3.1 Details of Calibration by Region

Table OA.17 provides the list of countries in each region analyzed in Section 5.3.3 of the paper. Note that China and United States are not listed here since they are single countries.

Table OA.17: List of countries by region

Region	Countries
Africa	Cote d'Ivoire; Egypt, Arab Rep.; Kenya; Nigeria; South Africa; Swaziland; Tanzania; Tunisia; Zimbabwe;
Middle East	Iran, Islamic Rep.; Israel; Kuwait; Saudi Arabia; United Arab Emirates;
Eastern Europe	Czech Republic; Hungary; Poland; Romania; Russian Federation; Slovak Republic; Slovenia; Ukraine;
Latin America	Argentina; Bolivia; Brazil; Chile; Colombia; Ecuador; El Salvador; Guatemala; Honduras; Mexico; Nicaragua; Peru; Puerto Rico; Trinidad and Tobago; Uruguay; Venezuela, RB;
ASEAN	Brunei Darussalam; Cambodia; Indonesia; Lao PDR; Malaysia; Myanmar; Philippines; Thailand; Vietnam;
Western Europe	Austria; Belgium; Croatia; Denmark; Finland; France; Germany; Greece; Italy; Netherlands; Norway; Portugal; Spain; Sweden; United Kingdom;

Table OA.18 is a longer version of Table 12 in the paper. It presents the model moments together with the data moments that are targeted. It also shows the change in the price indices when we consider perfect information in each region.

Table OA.18: Calibration by region, data and model moments

	Parameters						Data Moments					Model Moments					Gains from Perfect Info.	
Region	σ_θ	σ_ε	$\sigma_\theta^2/\sigma_\varepsilon^2$	σ_{ν_1}	$\sigma_{\nu_{10}}$	f	Cov_1	Cov_{10}	Var_1	Var_{10}	exit rate	Cov_1	Cov_{10}	Var_1	Var_{10}	exit rate	% Δ P	% Δ Q/L
Panel A: change five parameters																		
Africa	0.86	2.57	0.11	0.51	0.37	0.0152	0.040	0.019	0.186	0.100	0.106	0.040	0.020	0.186	0.100	0.105	-7.82	12.56
Middle East	0.83	2.64	0.10	0.58	0.45	0.0142	0.038	0.019	0.226	0.134	0.104	0.038	0.019	0.226	0.134	0.102	-7.81	12.16
Eastern Europe	1.41	1.80	0.62	0.58	0.32	0.0079	0.065	0.014	0.283	0.072	0.103	0.068	0.014	0.283	0.072	0.101	-6.32	9.44
Latin America	1.62	1.66	0.95	0.39	0.39	0.0070	0.069	0.013	0.218	0.097	0.106	0.073	0.013	0.218	0.097	0.103	-3.94	7.13
ASEAN	0.44	1.50	0.09	0.70	0.34	0.0075	0.011	0.006	0.264	0.073	0.078	0.008	0.006	0.264	0.072	0.078	-2.69	3.54
China	1.12	1.48	0.57	0.64	0.31	0.0074	0.044	0.010	0.275	0.065	0.089	0.044	0.010	0.276	0.065	0.089	-4.82	7.16
Western Europe	0.91	1.47	0.39	0.50	0.31	0.0131	0.033	0.009	0.178	0.065	0.107	0.034	0.009	0.179	0.065	0.106	-4.25	6.95
United States	0.78	1.49	0.27	0.52	0.31	0.0147	0.028	0.009	0.180	0.063	0.112	0.028	0.009	0.180	0.063	0.110	-4.49	7.11
Panel B: change four parameters, fix f																		
Africa	0.86	2.57	0.11	0.51	0.37	0.0093	0.040	0.019	0.186	0.100	0.106	0.039	0.020	0.184	0.098	0.073	-5.88	9.77
Middle East	0.83	2.64	0.10	0.58	0.45	0.0093	0.038	0.019	0.226	0.134	0.104	0.037	0.018	0.225	0.130	0.074	-6.11	9.83
Eastern Europe	1.41	1.80	0.62	0.58	0.32	0.0093	0.065	0.014	0.283	0.072	0.103	0.068	0.015	0.281	0.072	0.112	-6.15	9.86
Latin America	1.62	1.66	0.95	0.39	0.39	0.0093	0.069	0.013	0.218	0.097	0.106	0.071	0.014	0.219	0.098	0.125	-4.63	7.84
ASEAN	0.44	1.50	0.09	0.70	0.34	0.0093	0.011	0.006	0.264	0.073	0.078	0.011	0.006	0.268	0.072	0.102	-3.42	3.87
China	1.12	1.48	0.57	0.64	0.31	0.0093	0.044	0.010	0.275	0.065	0.089	0.042	0.010	0.280	0.065	0.104	-4.98	7.70
Western Europe	0.91	1.47	0.39	0.50	0.31	0.0093	0.033	0.009	0.178	0.065	0.107	0.034	0.010	0.177	0.066	0.081	-3.58	5.93
United States	0.78	1.49	0.27	0.52	0.31	0.0093	0.028	0.009	0.180	0.063	0.112	0.029	0.008	0.177	0.063	0.077	-3.27	5.55

Notes: Panel (A) shows the results when we re-calibrate five parameters for each region $(\sigma_\theta, \sigma_\varepsilon, \kappa_1, \kappa_0, f)$. We present age-dependent volatility $\sigma_{\nu_1}, \sigma_{\nu_{10}}$ instead of κ_1, κ_0 to facilitate interpretation. We target five moments in this calibration, $Cov(FE_{n-1,n}, FE_{n,n+1})$ for $n = 1$ and $n \geq 10$, $Var(FE_{n,n+1})$ for $n = 1$ and $n \geq 10$ and incumbent exit rates, respectively. $\%\Delta P$ and $\%\Delta Q/L$ are the percentage changes in the price index and the labor productivity when we change the model from the calibrated imperfect information case to perfect information. Panel (B) reports the results when we re-calibrate the learning and uncertainty related parameters but keep the fixed costs at the baseline value $f = 0.0093$. We target the first four moments but do not attempt to match the exit rates in the data. The model matches the data moments well (other than the untargeted exit rates in Panel B). A full list of countries in each region can be found in Online Appendix Table OA.17.

References

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

David, Joel, Lukas Schmid, and David Zeke, "Risk-Adjusted Capital Allocation and Misallocation," Working Paper 2019.

de Loecker, Jan and Frederic Warzynski, "Markups and Firm-Level Export Status," *American Economic Review*, 2012, *102* (6), 2437–71.

Fieler, Ana Cecília, Marcela Eslava, and Daniel Yi Xu, "Trade, Quality Upgrading, and Input Linkages: Theory and Evidence from Colombia," *American Economic Review*, January 2018, *108* (1), 109–146.

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

Hopenhayn, Hugo A., "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 1992, *60* (5), 1127–1150.

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

Muth, John F., "Optimal properties of exponentially weighted forecasts," *Journal of the american statistical association*, 1960, *55* (290), 299–306.

Verhoogen, Eric A., "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, May 2008, *123* (2), 489–530.