Online Appendix – Not For Publication

A Additional Empirical Analysis

A.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 $t+1$. In this section, we report firm-level statistics in periods when firms make forecasts, therefore up to the year $t = 2013$. In Table A-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 A-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.

In Table A-2, we report number of firms in major countries/regions in 2013. The countries/regions are consistent with our regional analysis in Section 5.4 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 C-26.

Table A-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

Table A-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 C-26 for the list of countries in each region.

facts, we show that they hold in both the whole sample and the manufacturing subsample.

Table A-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.

A.1.1 Compare with Industry Shares in the Census

In this subsection, we compare the composition of industries in the Basic Survey on Overseas Business Activities ("the Survey") and in the Economic Census for Business Frame ("the Census"). The Census is designed to summarize economic activities for all establishments and enterprises in all industries in Japan and is conducted every five years. We pick the common year of 2009 for comparison. We aggregate the Census industries (JSIC two-digit industries) to the slightly more aggregated industries in the Survey, and calculate the fraction of firms and employment in each industry. Note that a firm refers to a foreign affiliate of a Japanese multinational company in the survey and refers to an enterprise in the census.

Table A-4 reports the corresponding numbers for each industry.

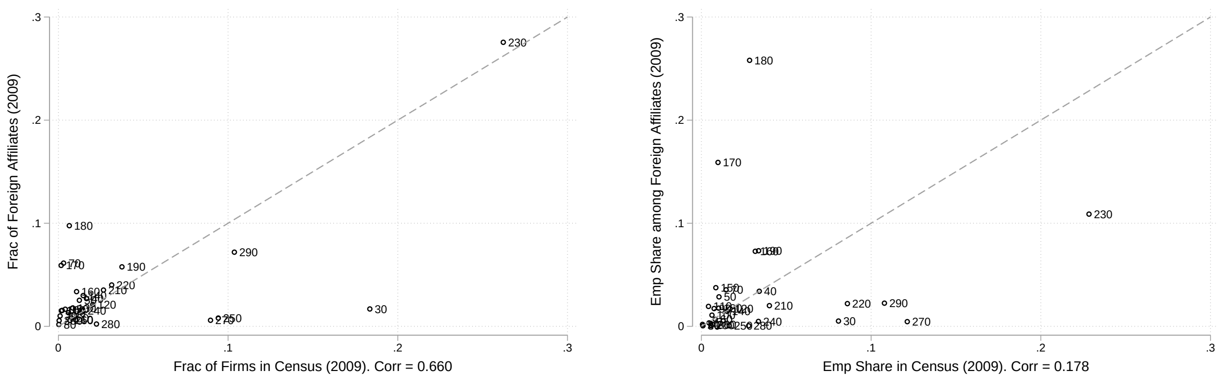
Table A-4: Industry Share, the Survey v.s. Census (2009)

Industry		Fraction of Firms		Employment Share	
Code	Description	Survey	Census	Survey	Census
10	Agriculture, forestry, and fisheries	0.006	0.010	0.003	0.005
20	Mining and quarrying of stone and gravel	0.010	0.001	0.002	0.001
30	Construction	0.017	0.184	0.005	0.081
40	Food and beverages, tobacco, and feed	0.027	0.017	0.034	0.034
50	Textile mill products	0.025	0.012	0.029	0.010
60	Lumber and wood products and of pulp, paper and paper products	0.008	0.008	0.007	0.008
70	Chemical and allied products	0.061	0.003	0.035	0.015
80	Petroleum and coal products	0.002	0.000	0.001	0.001
90	Ceramic, stone and clay products	0.013	0.006	0.017	0.007
100	Iron and steel iron industries	0.015	0.002	0.011	0.006
110	Non-ferrous metals and products	0.015	0.002	0.019	0.004
120	Fabricated metal products	0.021	0.020	0.017	0.017
130	General-purpose machinery	0.018	0.008	0.018	0.010
140	Production machinery	0.030	0.015	0.015	0.015
150	Business oriented machinery	0.017	0.004	0.037	0.009
160	Electrical machinery, equipment and supplies	0.034	0.011	0.073	0.032
170	Information and communication electronics equipment and of electronic parts and devices	0.059	0.002	0.159	0.010
180	Transportation equipment	0.098	0.006	0.258	0.028
190	Miscellaneous manufacturing industries	0.058	0.037	0.073	0.034
200	Electricity, gas, heat supply and water	0.006	0.000	0.001	0.006
210	Information and communications	0.035	0.027	0.020	0.040
220	Transport	0.040	0.031	0.022	0.086
230	Wholesale and retail trade	0.275	0.262	0.109	0.228
240	Finance and insurance	0.015	0.014	0.005	0.034
250	Real estate	0.008	0.094	0.001	0.016
260	Goods rental and leasing	0.006	0.007	0.002	0.007
270	Accommodations, Eating and drinking places	0.006	0.090	0.005	0.121
280	Education and learning support, Medical, health care and welfare, Compound services	0.002	0.022	0.001	0.028
290	Services, etc.	0.072	0.104	0.022	0.108

Notes: This table reports the fraction of firms and employment in each industry, in the survey and in the census (2009). A firm refers to a foreign affiliate of a Japanese multinational company in the survey, and refers to an enterprise in the census.

In Figure A-1, we plot the industry shares in the survey against those in the census, either measured by the number of firms or employment. The fraction of firms by industry in the two sources has a correlation of 0.66, while the employment shares have a correlation of 0.18. For some industries, their shares among all firms are similar in the two datasets. For example, 27.5% of the firms in the survey belong to "Wholesale and retail trade" (code = 270), and the corresponding number in the census is 26.2%. We see larger discrepancies in some other industries. For example, only 0.6% of firms in the census belong to "Transportation equipment" (code = 180), while 9.8% of firms in the survey are in this industries. This indicates strong comparative and absolute advantage of Japanese auto manufacturers in the international market.

Figure A-1: Industry Shares in the Survey (2009) v.s. Census (2009)



(a) Number of Firms (b) Employment

Notes: This figure plots the industry shares in the survey against those in the census (2009). The dashed line indicates the 45 degree line. The label of each dot is the industry code, and the corresponding industry description can be found in Table A-4.

A.2 Validation of Sales Forecasts

In this section, we present evidence that firms' 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 A-5, 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 A-5, 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 Tables A-15 and A-20 by dropping all observations with zero expected growth rates. Our empirical results remain largely unchanged.

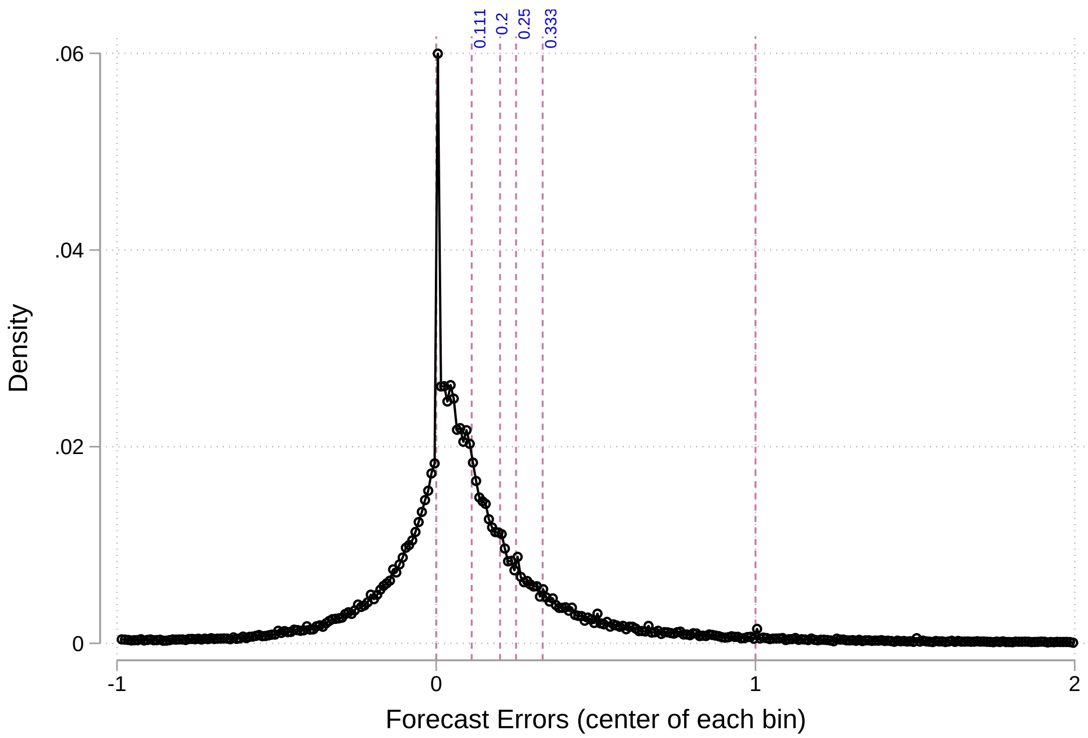
Table A-5: The Most Frequent Values of Expected Growth Rates

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 A-6. The first three columns of Table A-6 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 predictive 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. Moreover, the relatively high within-R^2s reported in the table are quite remarkable

Figure A-2: Density of expected growth, $\frac{E_t(R_{i,t+1})}{R_{i,t}} - 1$



Notes: Each circle represents the density of forecasting errors in a symmetric neighborhood around the center of the bin. Each bin has equal width 0.01, with the left boundary closed and the right boundary open (e.g., [-0.02, -0.01), [-0.01, 0), [0, 0.01), etc). We drop observations with expected growth above 2, which accounts for 3.3% of the sample.

and support the argument that the sales forecasts are economically meaningful factors that affect firm decisions. By contrast, firms make these forecasts seriously, and the forecasts contain more information on the firms' future conditions than realized outcomes in the past.

Table A-6: 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.

A.3 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

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

and obtain the residual forecast error $\hat{\epsilon}_{t,t+1}^{FE,\text{log}}$. 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 A-7 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, $\hat{\epsilon}_{t,t+1}^{FE,\text{log}}$ and $\hat{\epsilon}_{t,t+1}^{FE,\text{pct}}$, is zero by construction, the mean and median of $FE_{t,t+1}^{\text{log}}$ 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 $|\hat{\epsilon}_{t,t+1}^{FE,\text{log}}|$ (and $|\hat{\epsilon}_{t,t+1}^{FE,\text{pct}}|$) are similar to those of $\left|FE_{t,t+1}^{\text{log}}\right|$ (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.

A.4 Robustness Checks for Fact 1: Affiliate Age on Uncertainty

A.4.1 Detailed Baseline Regression Results

We now examine how age affects a firm i's absolute forecast error in year t using OLS regressions:

$$|FE_{it,t+1}^{\text{log}}| = \delta_n + \beta X_{it} + \delta_{ct} + \delta_{st} + \varepsilon_{it}, \tag{A-1}$$

Table A-7: 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.

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.

Column 1 in Table A-8 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, we control for the size of the firms and their parent companies in Japan (measured by log employment). In Column 3, we further control for firm-level fixed effects. Although larger firms tend to have smaller absolute forecast errors, the age effects survive.[29]

To evaluate the robustness of our results, we restrict our sample to (1) surviving entrants and (2) firms in manufacturing. Column 4 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 5 focuses on

[29] Tanaka, Bloom, David, and Koga (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.

Table A-8: Age effects on the absolute log forecast errors, $|FE^{\log}_{t,t+1}|$

Sample:	All Firms			Survivors	Manufacturing
Dep.Var: $\lvert FE^{\log}_{t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t = 2)$	-0.066^a	-0.059^a	-0.063^a	-0.068^a	-0.072^a
	(0.007)	(0.007)	(0.008)	(0.010)	(0.011)
$\mathbb{1}(\text{Age}_t = 3)$	-0.102^a	-0.089^a	-0.088^a	-0.093^a	-0.104^a
	(0.007)	(0.007)	(0.008)	(0.010)	(0.011)
$\mathbb{1}(\text{Age}_t = 4)$	-0.128^a	-0.113^a	-0.110^a	-0.108^a	-0.127^a
	(0.007)	(0.007)	(0.008)	(0.011)	(0.011)
$\mathbb{1}(\text{Age}_t = 5)$	-0.142^a	-0.125^a	-0.116^a	-0.121^a	-0.128^a
	(0.007)	(0.007)	(0.008)	(0.012)	(0.011)
$\mathbb{1}(\text{Age}_t = 6)$	-0.142^a	-0.124^a	-0.114^a	-0.120^a	-0.131^a
	(0.007)	(0.007)	(0.008)	(0.013)	(0.011)
$\mathbb{1}(\text{Age}_t = 7)$	-0.152^a	-0.131^a	-0.120^a	-0.134^a	-0.138^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	-0.125^a	-0.140^a
	(0.007)	(0.007)	(0.009)	(0.016)	(0.012)
$\mathbb{1}(\text{Age}_t = 9)$	-0.160^a	-0.135^a	-0.122^a	-0.126^a	-0.143^a
	(0.007)	(0.007)	(0.008)	(0.017)	(0.012)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.172^a	-0.137^a	-0.121^a	-0.129^a	-0.137^a
	(0.007)	(0.007)	(0.009)	(0.019)	(0.012)
$\log(\text{Emp})_t$		-0.021^a	-0.024^a	-0.035^a	-0.025^a
		(0.001)	(0.002)	(0.005)	(0.002)
$\log(\text{Parent Emp})_t$		0.001	0.001	0.010	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	0.104	0.122	0.366	0.357	0.363

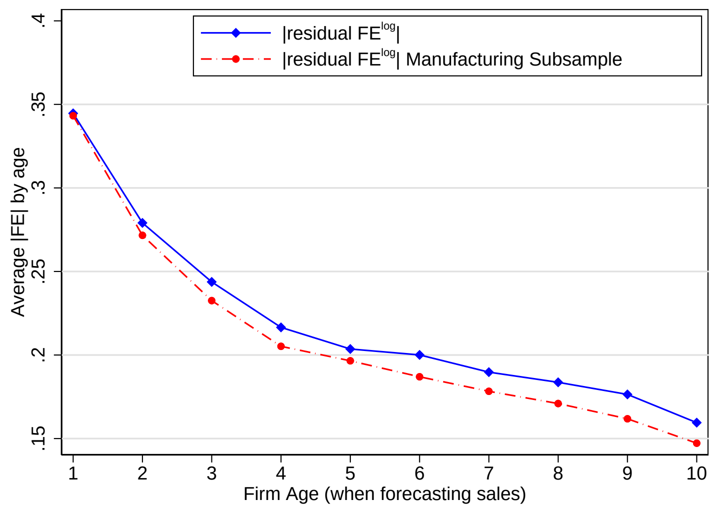
Notes: Standard errors are clustered at the business group level. Significance levels: c: 0.10, b: 0.05, a: 0.01. The dependent variable is the absolute value of forecast errors in all regressions. Age refers to the age of the firm when making the forecasts. Regressions in Columns 1–3 include all firms. Survivors (Column 4) refer to firms that have continuously appeared in the sample from age one to seven. Column 5 focuses on firms in manufacturing.

the manufacturing subsample, and the results are similar.

A.4.2 Alternative Measures of FE

We first show that our baseline results in Figure 2 and Table A-8 of the paper are robust to alternative measures of forecast errors. Figure A-3 plots the average absolute value of the residual forecast errors $\hat{\epsilon}^{FE,\log}_{t,t+1}$, for the entire sample and for the manufacturing subsample, respectively. We see a clear pattern that older firms make more precise forecasts.

Figure A-3: $|\hat{\epsilon}^{FE,\log}_{t,t+1}|$ declines with firm age



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

Tables A-9 and A-10 use the absolute value of percentage forecast errors and residual log forecast errors, respectively.

Table A-9: 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 A-10: 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.

A.4.3 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, Schmid, and Zeke (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

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

where $\delta_b^{\text{labor}} \times \delta_{ct}$ 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 $\delta_b^{\text{markup}} \times \delta_{ct}$ and $\delta_b^{\text{quality}} \times \delta_{ct}$ 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, Eslava, and Xu (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 A-11, we replicate regressions in Table A-8 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 A-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.

A.4.4 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 A-12, 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.[30] 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, Oldenski, and Ramondo (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 A-8 of the paper. However, the age coefficients are still negatively significant and maintain 80% of the magnitude of those in Table A-8.

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 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

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

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 A-12: 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.

A.4.5 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:

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

We then project $\hat{\upsilon}^2_{FE}$ onto X in the second-stage regression.[31] 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 A-13, 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 A-8 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 forecast errors.

[31] We use $\hat{\upsilon}$ to denote the residual term here to distinguish from the residual forecast errors defined in equation A-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.

Table A-13: Age effects on the variance forecast errors: conditional variance regressions

Dep. Var.	$\hat{\upsilon}^2_{FE,\log}(t,t+1)$				$\hat{\upsilon}^2_{FE,\text{pct}}(t,t+1)$
Sample:	All Firms		Survivors	Manufacturing	All Firms
	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t = 2)$	-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)
$\mathbb{1}(\text{Age}_t = 3)$	-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)
$\mathbb{1}(\text{Age}_t = 4)$	-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)
$\mathbb{1}(\text{Age}_t = 5)$	-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)
$\mathbb{1}(\text{Age}_t = 6)$	-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)
$\mathbb{1}(\text{Age}_t = 7)$	-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)
$\mathbb{1}(\text{Age}_t = 8)$	-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)
$\mathbb{1}(\text{Age}_t = 9)$	-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)
$\mathbb{1}(\text{Age}_t \geq 10)$	-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)
$\log(\text{Emp})_t$	-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)
$\log(\text{Parent Emp})_t$	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.

A.4.6 Partial Year Effects

In Table A-14, 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 A-14: 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.

A.4.7 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 A-8 of the paper. The results are almost identical (see Table A-15).

Table A-15: 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.

A.4.8 Excluding Observations with High Expected Growth

In our data, some firms report sales forecasts that imply very high expected growth ($\geq$ 100%). Among the firms that expect at least 100% sales growth, 71% of them have realized growth rates of 100% and above, and 84% of them have realized growth rates of 50% and above. Therefore, we argue that these forecasts are serious and contain useful information.

However, to be conservative, we perform robustness checks of regressions in Table A-8 of the paper by dropping observations with such high expected growth. The results are reported in Table A-16 and are similar to the baseline.

Table A-16: Age effects on the absolute value of forecast errors: drop observations with expected growth rates $\geq 100\%$

Sample:	All Firms			Survivors	Manufacturing
Dep.Var: $\lvert FE^{\log}_{t,t+1}\rvert$	(1)	(2)	(3)	(4)	(5)
$\mathbb{1}(\text{Age}_t = 2)$	-0.095^a	-0.086^a	-0.079^a	-0.094^a	-0.100^a
	(0.010)	(0.010)	(0.011)	(0.015)	(0.016)
$\mathbb{1}(\text{Age}_t = 3)$	-0.126^a	-0.113^a	-0.100^a	-0.117^a	-0.125^a
	(0.010)	(0.010)	(0.011)	(0.015)	(0.015)
$\mathbb{1}(\text{Age}_t = 4)$	-0.146^a	-0.132^a	-0.118^a	-0.127^a	-0.146^a
	(0.010)	(0.010)	(0.012)	(0.016)	(0.015)
$\mathbb{1}(\text{Age}_t = 5)$	-0.156^a	-0.140^a	-0.122^a	-0.139^a	-0.145^a
	(0.010)	(0.010)	(0.011)	(0.016)	(0.015)
$\mathbb{1}(\text{Age}_t = 6)$	-0.155^a	-0.138^a	-0.121^a	-0.141^a	-0.151^a
	(0.010)	(0.010)	(0.012)	(0.017)	(0.016)
$\mathbb{1}(\text{Age}_t = 7)$	-0.164^a	-0.144^a	-0.125^a	-0.148^a	-0.154^a
	(0.010)	(0.010)	(0.012)	(0.017)	(0.016)
$\mathbb{1}(\text{Age}_t = 8)$	-0.167^a	-0.146^a	-0.126^a	-0.140^a	-0.157^a
	(0.010)	(0.010)	(0.012)	(0.018)	(0.016)
$\mathbb{1}(\text{Age}_t = 9)$	-0.171^a	-0.148^a	-0.126^a	-0.143^a	-0.159^a
	(0.010)	(0.010)	(0.012)	(0.020)	(0.016)
$\mathbb{1}(\text{Age}_t \geq 10)$	-0.182^a	-0.150^a	-0.125^a	-0.143^a	-0.152^a
	(0.009)	(0.009)	(0.012)	(0.021)	(0.016)
$\log(\text{Emp})_t$		-0.019^a	-0.021^a	-0.030^a	-0.023^a
		(0.001)	(0.002)	(0.006)	(0.003)
$\log(\text{Parent Emp})_t$		0.000	0.001	0.008	0.000
		(0.001)	(0.003)	(0.008)	(0.003)
Industry-year FE	Y	Y	Y	Y	Y
Country-year FE	Y	Y	Y	Y	Y
Firm FE			Y	Y	Y
N	123316	120701	115334	19462	72299
R^2	0.099	0.117	0.356	0.366	0.352

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.

A.5 Robustness Checks for Fact 2

A.5.1 Auto-correlations using alternative forecast errors

Table A-17: 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
<u>All industries</u>				
$corr(FE^{\text{log}}_{t,t+1}, FE^{\text{log}}_{t-1,t})$	0.137	0.170	0.152	0.120
	[96452]	[10410]	[13801]	[72241]
$corr(FE^{\text{pct}}_{t,t+1}, FE^{\text{pct}}_{t-1,t})$	0.105	0.137	0.120	0.093
	[96967]	[10578]	[13875]	[72514]
$corr(\hat{\epsilon}^{FE,\text{log}}_{t,t+1}, \hat{\epsilon}^{FE,\text{log}}_{t-1,t})$	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]
<u>Manufacturing</u>				
$corr(FE^{\text{log}}_{t,t+1}, FE^{\text{log}}_{t-1,t})$	0.139	0.193	0.151	0.116
	[60123]	[5828]	[8591]	[45704]
$corr(FE^{\text{pct}}_{t,t+1}, FE^{\text{pct}}_{t-1,t})$	0.108	0.172	0.116	0.089
	[60364]	[5906]	[8623]	[45835]
$corr(\hat{\epsilon}^{FE,\text{log}}_{t,t+1}, \hat{\epsilon}^{FE,\text{log}}_{t-1,t})$	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: $FE^{\text{log}}_{t,t+1}$ is the log deviation of the realized sales in year $t+1$ from the sales forecast made in year t. $FE^{\text{pct}}_{t,t+1}$ is the percentage deviation of realized sales from expected sales. The other two measures, $\hat{\epsilon}^{FE,\text{log}}_{t,t+1}$ and $\hat{\epsilon}^{FE,\text{pct}}_{t,t+1}$, are the residual forecast errors, which we obtain by regressing $FE^{\text{log}}_{t,t+1}$ 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. Number of observations used for each correlation is shown in the brackets below. All correlation coefficients are significant at the 1% level.

A.5.2 AR(1) Models with Age Interactions

In this section, we perform several robustness checks of the regressions in Table 1 of the paper. We first replace the log forecast errors with alternative definitions of forecast errors. Table A-18 uses percentage forecast errors, while Table A-19 uses residual log forecast errors. The results in Table A-19 are almost identical to those obtained using log forecast errors, while the magnitudes of the estimates in Table A-18 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 1. The results are very similar (see Table A-20). Finally, we exclude firms with high expected growth rates ($\geq 100\%$) and find similar results as in Table 1 (see Table A-21).

Table A-18: 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{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(Age$_t$)		-0.015^a		-0.023^a		-0.024^a		-0.029^a
		(0.006)		(0.007)		(0.009)		(0.010)
log(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(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 A-19: 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{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(Age$_t$)		-0.019^a		-0.025^a		-0.030^a		-0.035^a
		(0.006)		(0.007)		(0.009)		(0.011)
log(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(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 A-20: 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{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(Age$_t$)		-0.014^b		-0.013^b		-0.025^a		-0.023^b
		(0.006)		(0.006)		(0.009)		(0.009)
log(Emp)$_t$			0.002^b	0.002^b			0.001	0.001
			(0.001)	(0.001)			(0.001)	(0.001)
log(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.

Table A-21: AR(1) regressions with Age Interactions, excluding firms with expected growth rates $\geq 100\%$

Sample:	All Affiliates				Manufacturing			
Dep.Var: $FE^{\log}_{t+1,t+2}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$FE^{\log}_{t,t+1}$	0.081^{a}	0.077^{a}	0.078^{a}	0.076^{a}	0.084^{a}	0.084^{a}	0.079^{a}	0.079^{a}
	(0.016)	(0.015)	(0.016)	(0.016)	(0.021)	(0.021)	(0.021)	(0.022)
$\times \max\{\text{Age}_t, 10\}$	-0.004^{b}		-0.004^{b}		-0.006^{b}		-0.005^{c}	
	(0.002)		(0.002)		(0.003)		(0.003)	
$\times \log(\text{Age}_t)$		-0.013^{b}		-0.012^{c}		-0.020^{b}		-0.017^{c}
		(0.006)		(0.006)		(0.009)		(0.009)
$\log(\text{Emp})_t$			0.001	0.001			-0.000	-0.000
			(0.001)	(0.001)			(0.001)	(0.001)
$\log(\text{Parent Emp})_t$			-0.008^{c}	-0.008^{c}			-0.009	-0.009
			(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	89476	89476	87868	87868	55842	55842	55267	55267
R^2	0.208	0.208	0.211	0.211	0.233	0.233	0.235	0.235

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.

A.6 Robustness Checks for Fact 3

In this section, we perform a few robustness checks for Fact 3. In Fact 2, we have shown that the AR(1) coefficients are also affected by firm age. Therefore, we examine the robustness of the above results by introducing a horse race between country characteristics and firm age in Table A-22. We find that age still significantly reduces the AR(1) coefficients, and the country characteristics have the expected effects as in Table 2 of the paper.

In columns 4 and 8 of Table A-22, 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. However, the correlation between the management score and GDP per capita in our sample is 0.94, and we 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.)

Table A-22: AR(1) coef and horse race between country characteristics

	Dep.Var: $FE^{\log}_{t+1,t+2}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$FE^{\log}_{t,t+1}$	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(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 Bloom et al. (2014). Management score, time zone differences and log GDP per capita are all standardized to faciliate interpretation of the coefficients.

B Theory Appendix

B.1 Proof for Proposition 1

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

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

Therefore, (log) forecast error of sales are

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

From equation (B-1), 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 (B-1) as

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

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 (B-2), we calculate the variance of forecast error as

$$\begin{aligned}
Var(FE^{\log}_{n-1,n}) &= \frac{\zeta(n-1,\lambda)^2\sigma_\varepsilon^2}{(n-1)\sigma^2} + \frac{(1-\zeta(n-1,\lambda))^2\lambda\sigma_\varepsilon^2}{\sigma^2} \qquad \text{(B-3)} \\
&= \frac{\sigma_\varepsilon^2}{\sigma^2}\Big(\frac{\lambda}{1+(n-1)\lambda}\Big) + \frac{(\sigma-1)^2\sigma^2_{\nu_n}}{\sigma^2}.
\end{aligned}$$

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

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

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.

B.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 (7);*

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

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 (10);*

4. *a measure function of firms $\lambda\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1, \theta\right)$ 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*
$$\lambda\left(d\varphi_0, \bar{s}_0, 0, d\theta\right) = (1-\eta) J g_\theta\left(\theta\right) d\theta \times g_{\varphi_0}\left(\varphi_0\right) d\varphi_0,$$
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 $P \equiv \left(\int_{\omega\in\Omega} e^{\theta(\omega)} p(\omega)^{1-\sigma} d\omega\right)^{1/(1-\sigma)}$. Therefore, we have:*
$$P^{1-\sigma} = \sum_{n\geq 1} \int_{\varphi_{n-1}, \bar{s}_{n-1}, \theta} \begin{array}{c} e^{\theta} \times p\left(\theta, \varphi_n, b\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right)\right)^{1-\sigma} \times (1-\eta) \times \\ o\left(\varphi_{n-1}, \bar{s}_{n-1}, n-1\right) \times \lambda\left(d\varphi_{n-1}, d\bar{s}_{n-1},, n, d\theta\right) \end{array}.$$

B.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 C.2. In this case, L should also include labor used for entry, i.e., $L = L^{prod} + L^{fixed} + L^{entry}$, where $L^{entry} = Jf^e$.

B.4 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

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

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

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

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.

B.5 Full Information Rational Expectation Models with Endogenous Selection

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

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

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 (B-1) 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

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

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 (B-2) 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.

B.6 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):

$$a_t(\omega) = \theta(\omega) + \varepsilon_t(\omega). \tag{B-1}$$

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 $N(0, \sigma_\varepsilon^2)$. 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

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

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[32]

$$l_t = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma}\left(\frac{b(\varphi_{n-1},\bar{a}_{n-1},n-1)}{w}\right)^{\sigma} YP^{\sigma-1}, \tag{B-3}$$

where

$$\begin{aligned} b(\varphi_{n-1},\bar{a}_{n-1},n-1) &\equiv E\left(e^{\frac{a_t}{\sigma}}\varphi_n^{\frac{\sigma-1}{\sigma}}|\varphi_{n-1},\bar{a}_{n-1},n\right) \\ &= \exp\left\{\frac{\mu_{n-1}}{\sigma}+\frac{\sigma_{n-1}^2+\sigma_\varepsilon^2}{2\sigma^2}+\frac{\sigma-1}{\sigma}((1-\rho)\mu_\varphi+\rho\log\varphi_{n-1})+\frac{(\sigma-1)^2\sigma_{\nu_n}^2}{2\sigma^2}\right\}, \end{aligned} \tag{B-4}$$

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

$$\log(R_n(\theta)) = \log\left(\frac{Y}{P^{1-\sigma}}\right)+\frac{a_t}{\sigma}+\frac{\sigma-1}{\sigma}\log(\varphi_n)+(\sigma-1)\log b(\varphi_{n-1},\bar{a}_{n-1},n-1)+(\sigma-1)\Big[\log\left(\frac{\sigma-1}{\sigma w}\right)\Big], \tag{B-5}$$

and

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

The resulting log forecast error of sales is

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

which can be rewritten as

$$\begin{aligned} FE_{n-1,n}^{\log} &= \frac{(\sigma-1)\nu_{n+1}}{\sigma}-\frac{(\sigma-1)^2\sigma_{\nu_{n+1}}^2}{2\sigma^2} \\ &+\frac{(1-\zeta(n-1,\lambda))(\theta-\bar{\theta})+\varepsilon_t-\zeta(n-1,\lambda)\frac{\sum_{i=t-n+1}^{t-1}\varepsilon_i}{n-1}}{\sigma}-\frac{\sigma_{n-1}^2+\sigma_\varepsilon^2}{2\sigma^2}, \end{aligned} \tag{B-6}$$

where

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

The autocovariance of (log) sales forecast errors is simply

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

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

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

For forecast errors made at ages n and $n+1$, they share two common components in equation (B-6): $\theta - \bar{\theta}$ 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.

B.6.1 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.

B.7 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:

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

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

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

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.

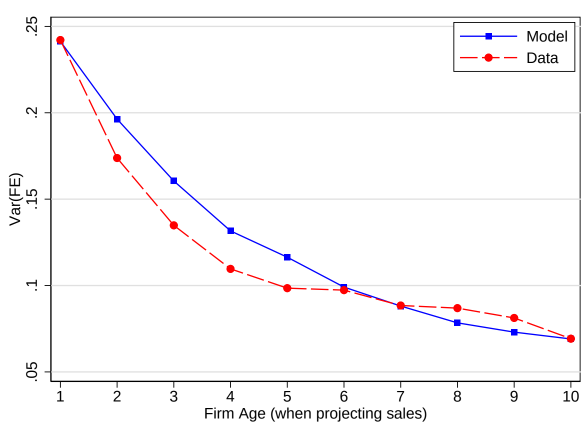
C Additional Quantitative Results

C.1 Model Fit: Dynamics of Forecast Errors and Sales

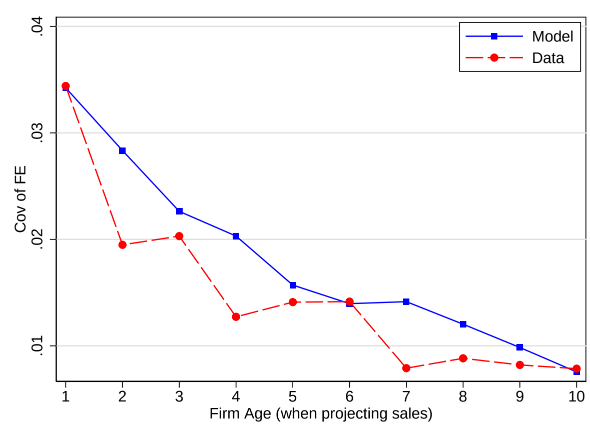
In this section, we examine how our calibrated model performs regarding untargeted moments, focusing on moments of forecast errors and sales. In Figure C-4, we plot the age profile of the variance and covariance of forecast errors. In the calibration, we match these moments for the youngest and oldest 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 3 in the paper.

In Figure C-5, 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 C-4: Moments of Forecast Errors, Model vs. Data



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

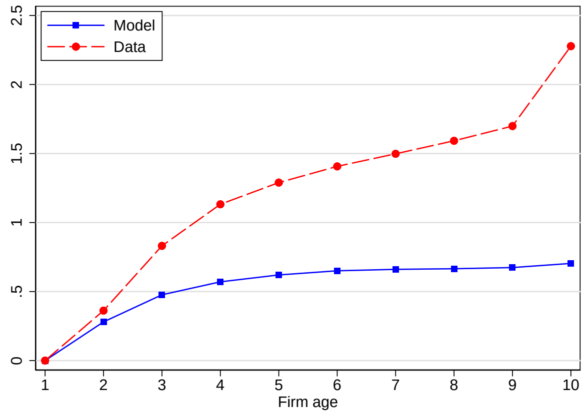


(b) $Cov(FE_{n,n+1}, FE_{n+1,n+2})$

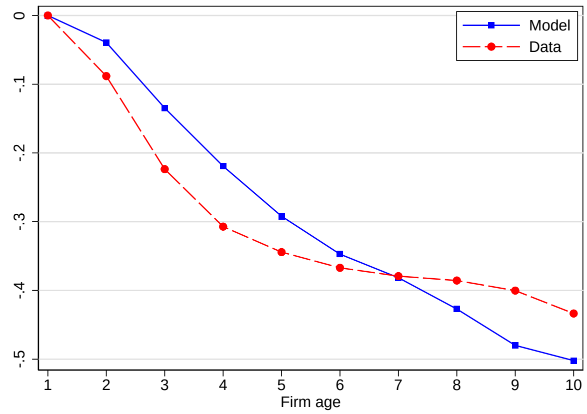
C.2 Sensitivity to General Equilibrium and Free Entry

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

Figure C-5: Moments of Sales, Model vs. Data



(a) Average Log(Sales)



(b) Change in SD(Δlog(Sales))

total (inelastic) labor supply. Columns under "Fixed J" in Table C-23 present the similar comparative statics as those in Table 5, 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 C-23 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 C-23, 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.[33] 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.

[33]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.

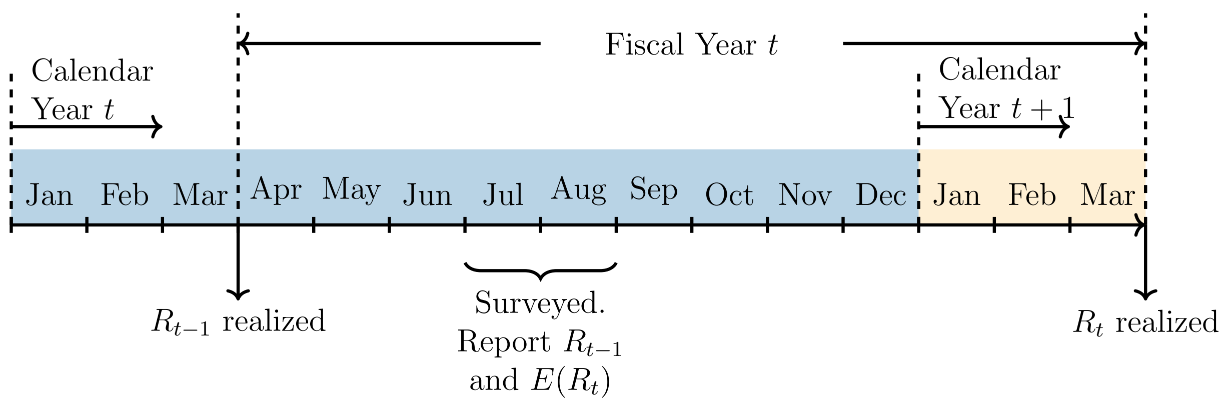
Table C-23: The Impact of σ_ε Under General Equilibrium

	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 demand equals 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.

C.3 Robustness with More Firms that Enter Early within a Fiscal Year

Figure C-6: Timeline of the Survey and Forecasts



As mentioned in Section 2.1 of the paper, firms report their expected sales not at the beginning of the fiscal year (Apr), but in the third month into the current fiscal year (Jul to Aug). We present the survey timeline graphically in Figure C-6. It is plausible that firms learn little about their signal in the current period three months into the year, but it is also possible that firms learn about their signal partially. To address this issue, we consider a robustness check in this section. We do not attempt to quantify a model at higher frequency. Instead, we keep our assumption that each period is a year, but adjust the composition of firms of a particular age in our data so that they are on average "three months older".

Recall that in our baseline calibration, we adopt a strategy to address another issue: firms enter in different months of the same fiscal year so firms of a particular age in our data may have different experiences when measured in months. As explained in paper footnote 22, we use a mix of "age n" and "age $n+1$" firms to mimic "age n" firms in the data. In particular, we assume that any entrant has an equal chance to enter at the beginning or at the end of a fiscal year, so we are effectively using "age $n+0.5$" firms in the model to mimic "age n" firms in the data. In this section, we assume instead that 75% of the entrants enter at the beginning of a fiscal year and 25% enter at the end. Therefore, we effectively use "age $n+0.75$" firms in the model to mimic "age n" firms in the data. This makes firms older by three months on average, which is consistent with the survey timeline.

We re-calibrate our baseline model using the new mixing strategy and present the calibration results in Table C-24. We find a larger σ_θ (1.01 instead of 0.96) and a slightly large σ_ε (1.35 instead of 1.34). This is because firms in the data are assumed to be three months older than in our baseline, and to match the same covariance in FEs, we need larger σ_θ and σ_ε. Another difference is that we need slightly smaller fixed costs. Firms have more uncertainty at the beginning (higher σ_θ) so we need a smaller fixed costs to match the same exit rates. To put it differently, given the new learning parameters, they tend to exit more given the same fixed costs.

Table C-24: Parameters Calibrated Without Solving the Model

Parameters	Value	Description	Source/Target	Moments Data	Moments Model
Panel A: Calibrated without solving the model					
σ	4	elasticity of substitution between different varieties	Bernard et al. (2003)		
β	0.96	discount factor	4% real interest rate		
η	0.03	exogenous death rate	exit rate of the largest 5% of firms above age ten		
Panel B: Calibrated by solving the model and matching moments					
f_m	0.0086	fixed cost	average exit rate of incumbents	0.093	0.093
σ_θ	1.01	std of θ	$Cov(FE_{t-1}, FE_t)$ at age one	0.034	0.034
σ_ε	1.35	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	0.29	$\sigma_{\nu_n} = \kappa_0 + \kappa_1(1-n/10)^2$	Var(FE) above at age one	0.242	0.245
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	0.667

In Table C-25, we replicate the counterfactual results in Table 5 in the paper using the new calibration results. We find that gains from moving to information are slightly larger (6.41% instead of 6.36%). This is intuitive because firms now have higher information friction about θ. The gains from information, measured as moving towards a higher value of $\sigma_\varepsilon = 2.50$, also become larger due to the same reason.

Table C-25: Aggregate Outcomes under Different σ_ε

Panel A: $f = 0.0086$ (benchmark)	(1)	(2)	(3)
Statistics	High Info. Friction $\sigma_\varepsilon = 2.50$	Baseline Info. Friction $\sigma_\varepsilon = 1.35$	Perfect Info.
Mass of Active Firms	11.393	10.526	9.103
Incumbents Average θ	0.641	0.822	1.105
Incumbents Average $\theta + (\sigma - 1)\log\varphi$	0.197	0.243	0.329
Q/L	3.623	3.775	4.018
Δ% Q/L	-4.03		6.41
Panel B: $f = 0$	(1)	(2)	(3)
Statistics	High Info. Friction $\sigma_\varepsilon = 2.50$	Baseline Info. Friction $\sigma_\varepsilon = 1.35$	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	4.622	4.730	4.916
Δ% Q/L	-2.28		3.93

Notes: This table reports equilibrium outcomes under a high level of information frictions ($\sigma_\varepsilon = 2.50$), baseline model ($\sigma_\varepsilon = 1.35$) and perfect information, with different values of fixed costs (baseline value, 0.0086, and alternative value, 0). As is explained in paper footnote 24, 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.

C.4 Details of Calibration by Region

Table C-26 provides the list of countries in each region analyzed in Section 5.4 of the paper. Note that China and United States are not listed here since they are single countries.

Table C-26: 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 C-27 is a longer version of Table 6 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 C-27: 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	1.61	0.08	0.70	0.34	0.0074	0.011	0.006	0.264	0.073	0.078	0.011	0.006	0.264	0.073	0.078	-2.88	3.68
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	1.61	0.08	0.70	0.34	0.0093	0.011	0.006	0.264	0.073	0.078	0.012	0.006	0.268	0.072	0.103	-3.58	4.04
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 C-26.

References

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

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

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.

David, J., L. Schmid, and D. Zeke (2019): "Risk-Adjusted Capital Allocation and Misallocation," Working Paper.

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

Fieler, A. C., M. Eslava, and D. Y. Xu (2018): "Trade, Quality Upgrading, and Input Linkages: Theory and Evidence from Colombia," *American Economic Review*, 108, 109–146.

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

Garetto, S., L. Oldenski, and N. Ramondo (2019): "Multinational Expansion in Time and Space," Working Paper 25804, National Bureau of Economic Research.

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

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

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

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