# Demand learning and firm dynamics: evidence from exporters

Online Appendix

January 8, 2018

Nicolas Berman[1] Vincent Rebeyrol[2] Vincent Vicard[3]

[1]Aix-Marseille University (Aix-Marseille School of Economics), CNRS, EHESS, Graduate Institute Geneva and CEPR. E-mail: nicolas.berman@univ-amu.fr.
[2]Toulouse School of Economics. E-mail: vincent.rebeyrol@ut-capitole.fr
[3]CEPII and Banque de France. E-mail: vincent.vicard@cepii.fr.

# Contents

# A Detailed derivations and proofs

**Optimal quantities and prices.** Firms choose quantities by maximizing expected profits subject to demand. Using (1), we get:

$$\max_q \int \pi_{ijkt} dG_{t-1}(a_{ijkt}) = \max_q q_{ijkt}^{1-\frac{1}{\sigma_k}} \left(\frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}}\right)^{\frac{1}{\sigma_k}} \mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] - \frac{w_{it}}{\varphi_{ikt}} q_{ijkt} - F_{ijk}$$

The FOC writes:

$$\left(1-\frac{1}{\sigma_k}\right) q_{ijkt}^{-\frac{1}{\sigma_k}} \left(\frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}}\right)^{\frac{1}{\sigma_k}} \mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] = \frac{w_{it}}{\varphi_{ikt}}$$

$$\Leftrightarrow q_{ijkt}^{*} = \left(\frac{\sigma_k}{\sigma_k - 1}\frac{w_{it}}{\varphi_{ikt}}\right)^{-\sigma_k} \left(\frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}}\right) \mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]^{\sigma_k}$$

And from the constraint, we get $p_{ijkt}^{*} = \left(\frac{\sigma_k}{\sigma_k-1}\frac{w_{it}}{\varphi_{ikt}}\right)\left(\frac{e^{\frac{a_{ijkt}}{\sigma_k}}}{E_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]}\right)$

**Updating of firm's beliefs about expected demand.** First note that firm $i$ has a prior about the demand shock given by $a_{ijkt} \sim \mathcal{N}(\widetilde{\theta}_{ijkt-1}, \widetilde{\sigma}^2_{ijkt-1}+\sigma^2_{\varepsilon})$ and thus $e^{\frac{a_{ijkt}}{\sigma_k}} \sim L\mathcal{N}(\frac{\widetilde{\theta}_{ijkt-1}}{\sigma_k}, \frac{\widetilde{\sigma}^2_{ijkt-1}+\sigma^2_{\varepsilon}}{\sigma_k^2})$.

It follows that $\mathbb{E}_{t-1}[e^{\frac{a_{ijkt}}{\sigma_k}}] = \int \left(e^{\frac{a_{ijkt}}{\sigma_k}}\right) dG_{t-1}(a_{ijkt}) = e^{\frac{1}{\sigma_k}\left(\widetilde{\theta}_{ijkt-1}+\frac{\widetilde{\sigma}^2_{ijkt-1}+\sigma^2_{\varepsilon}}{2\sigma_k}\right)}$. Hence:

$$\Delta \ln \mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right] = \frac{1}{\sigma_k}\left(\Delta\widetilde{\theta}_{ijkt} + \frac{\widetilde{\sigma}^2_{ijkt} - \widetilde{\sigma}^2_{ijkt-1}}{2\sigma_k}\right)$$

Using the the definition of $\Delta\widetilde{\theta}_{ijkt}$, $g_t$, $\widetilde{\sigma}^2_{ijkt-1}$ and $\widetilde{\sigma}^2_{ijkt}$ (see (3) and (4)), it is easy to show that $\frac{\widetilde{\sigma}^2_{ijkt-1}-\widetilde{\sigma}^2_{ijkt}}{g_t} = \widetilde{\sigma}^2_{ijkt-1}$. It follows the expression in the text (11):

$$\Delta \ln \mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right] = \frac{g_t}{\sigma_k}\left(a_{ijkt} - \widetilde{\theta}_{ijkt-1}\right) - \frac{g_t}{\sigma_k}\frac{\widetilde{\sigma}^2_{ijkt-1}}{2\sigma_k}$$

But we only observe $\Delta\varepsilon^q_{ijkt+1} = \sigma_k \Delta \ln \mathbb{E}_t[e^{\frac{a_{ijkt+1}}{\sigma_k}}]$.

It follows that $\Delta\varepsilon^q_{ijkt+1} = g_t\left(a_{ijkt} - \widetilde{\theta}_{t-1}\right) - g_t\frac{\widetilde{\sigma}^2_{ijkt-1}}{2\sigma_k}$. As $\varepsilon^q_{ijkt} = \sigma_k \ln \mathbb{E}_{t-1}[e^{\frac{a_{ijkt}}{\sigma_k}}]$, we get $\varepsilon^q_{ijkt} = \widetilde{\theta}_{ijkt-1} + \frac{\widetilde{\sigma}^2_{ijkt-1}+\sigma^2_{\varepsilon}}{2\sigma_k}$, or $\widetilde{\theta}_{ijkt-1} = \varepsilon^q_{ijkt} - \frac{\widetilde{\sigma}^2_{t-1}+\sigma^2_{\varepsilon}}{2\sigma_k}$. This leads to the equation we test in the main text:

$$\Delta\varepsilon^q_{ijkt+1} = g_t\left(a_{ijkt} - \varepsilon^q_{ijkt}\right) + g_t\frac{\sigma^2_{\varepsilon}}{2\sigma_k} \tag{21}$$

**Prediction 1.** Prediction 1 states that $a_{ijkt} - \varepsilon^q_{ijkt}$, has a larger impact on firms' updating, the younger the firms are. Using (21), we immediately get:

$$\frac{\partial \Delta \varepsilon^q_{ijkt+1}}{\partial \left(a_{ijkt} - \varepsilon^q_{ijkt}\right)} = g_t > 0$$

Updating is larger for younger firms, as $g_t$ decreases with $t$.

**Prediction 2: Impact of market uncertainty.** Moreover, the updating process is also affected by the level of market uncertainty $\sigma^2_\epsilon$. Formally:

$$\frac{\partial^2 \left(\Delta \varepsilon^q_{ijkt+1}\right)}{\partial \left(a_{ijkt} - \varepsilon^q_{ijkt}\right) \partial \sigma^2_\epsilon} = -\frac{g_t^2}{\sigma^2_{jk0}} < 0$$

Updating decreases with uncertainty, as a signal is less informative when market uncertainty is larger. As a consequence, market uncertainty dampens the speed of learning. In other words, updating decreases less with age, the more uncertain the market. This can be seen noting that:

$$\frac{\partial^2 \left(\Delta \varepsilon^q_{ijkt+1}\right)}{\partial \left(a_{ijkt} - \varepsilon^q_{ijkt}\right) \partial t} = -\frac{1}{\left(\frac{\sigma^2_\epsilon}{\sigma^2_{jk0}} + t\right)^2}$$

which is larger (less negative) in more uncertain markets (with larger $\sigma^2_\epsilon$).

**Dynamics of prices and quantities**. The model predicts expected growth rates of opposite signs for quantities and prices. This result comes from (14) and (15). Taking the first difference of these equations in expected terms, we directly get the expected growth rates. We find:

$$\mathbb{E}\left[\Delta \ln Z^q_{ijkt+1}\right] = -\frac{1}{\sigma_k}\mathbb{E}\left[\Delta \ln Z^p_{ijkt+1}\right]$$

Given that firms that decrease in size will on average be more likely to exit, the expected growth rate of quantities must be positive for survivors. Hence, the expected growth rate of prices for these firms should be negative and smaller by a factor $-\frac{1}{\sigma_k}$. Quantitatively, this is very close to what we find in table A.23.

**Prediction 3.** Prediction 3 states that the variance of growth rates within cohort decrease with cohort age. The variance of these growth rates can be expressed as:

$$\mathbb{V}\left[\Delta \ln Z^q_{ijkt+1}\right] = \sigma_k^2 \mathbb{V}\left(\Delta \ln \mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]\right) \tag{22}$$

$$\begin{aligned}\mathbb{V}\left[\Delta \ln Z^p_{ijkt+1}\right] &= \left(\frac{1}{\sigma_k}\right)^2 \mathbb{V}\left(\Delta a_{ijkt+1}\right) + \mathbb{V}\left(\Delta \ln \mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]\right) \\ &\quad -\frac{2}{\sigma_k} Cov\left(\Delta \ln \mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right], \Delta a_{ijkt+1}\right)\end{aligned} \tag{23}$$

First, $a_{ijkt+1}$ and $a_{ijkt}$ being drawn from the same distribution, $\mathbb{V}\left[\Delta a_{ijkt+1}\right] = 2\sigma^2_\epsilon$. Second, using (11), we get:

$$\mathbb{V}\left(\Delta \ln \mathbb{E}_t \left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]\right) = \left(\frac{\sigma_\epsilon}{\sigma_k\left(\frac{\sigma_\epsilon^2}{\sigma_{jk0}^2} + t\right)}\right)^2$$

As $\mathbb{E}\left[\Delta a_{ijkt+1}\right] = 0$, we get:

$$Cov\left(\Delta \ln \mathbb{E}_t \left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right], \Delta a_{ijkt+1}\right) = \mathbb{E}\left[\Delta \ln \mathbb{E}_t \left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right] \Delta a_{ijkt+1}\right]$$

Expanding this expression and using the fact that $a_{ijkt}$ and $a_{ijkt+1}$ are independent and that $\mathbb{E}\left[a_{ijkt}\right] = \mathbb{E}\left[a_{ijkt+1}\right] = \overline{a}_{ijkt-1}$, we get:

$$Cov\left(\Delta \ln \mathbb{E}_t \left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right], \Delta a_{ijkt+1}\right) = -\frac{\sigma_\epsilon^2}{\sigma_k\left(\frac{\sigma_\epsilon^2}{\sigma_{jk0}^2} + t\right)}$$

Finally, plugging this term into (22) and (23) and after rearranging, we get the following expressions which are both strictly decreasing with $t$:

$$\mathbb{V}\left[\Delta \ln Z_{ijkt+1}^q\right] = \left(\frac{\sigma_\epsilon}{\left(\frac{\sigma_\epsilon^2}{\sigma_{jk0}^2} + t\right)}\right)^2 \tag{24}$$

$$\mathbb{V}\left[\Delta \ln Z_{ijkt+1}^p\right] = \left(\frac{\sigma_\epsilon}{\sigma_k}\right)^2 \left(\left(\frac{1}{\left(\frac{\sigma_\epsilon^2}{\sigma_{jk0}^2} + t\right)} + 1\right)^2 + 1\right) \tag{25}$$

# B Data and descriptive evidence on firm dynamics

## B.1 Dataset construction

We use data on values and quantities sold by French firms, by destination, HS6 product and year, over the period 1994-2005. We focus on the subset of HS6 product categories that remain stable in the HS classification over the period in order to be able to track firms over time on specific markets.[4] As we use the first two years to define entry, we concentrate on the years 1996-2005. Note that firms-products-destinations that already export at the beginning of the period (in 1994 or 1995) are not considered, as we are interested in post-entry dynamics.

Because intra-EU and extra-EU flows are treated differently by the French Customs, we harmonize the data in several ways. The declaration of extra-EU export flows is mandatory when a transaction exceeds 1,000 euros or 1,000 kg. For shipments to EU countries, firms have to report their detailed expeditions when their total exports to all EU countries exceed a threshold over the year of 38,100 euros before 2001, 99,100 euros in 2001 and 100,000 euros between 2002 and 2005. Firms below the reporting threshold are required to fill a simplified form without the details on the product exported and the destination market. In order to harmonize the data requirement over the different destinations, we drop all intra-EU export flows below 1,000 euros, as well as firms that report at least once under the simplified procedure (as for these firms, we do not observe their flows in all markets). We also check that all our results are unchanged when removing EU destinations from the sample.

## B.2 Additional descriptive evidence

This section provides further details on the computation of the stylized facts presented in section 2 of the main text.

**Contribution to aggregate sales growth.** The literature has documented the essential contribution of young firms to industry dynamics, either in terms of aggregate output, employment or trade. Haltiwanger et al. (2013) show for instance that US start-ups display substantially higher rates of job creation and destruction in their first ten years, and that these firms represent a large share of total employment after a decade of existence. These patterns are also found for other countries (see Criscuolo *et al.*, 2014 for evidence on 18 OECD countries; Lawless (2014) on Irish firms, Ayyagari *et al.*, 2011 for developing countries). Similar facts characterize trade dynamics: Eaton et al. (2008) and Bernard et al. (2009) show that exporters start small but that, conditional on survival, they account for large shares of total export growth after a few years.

Our exporter-level data exhibit comparable features. We compute the contribution of the intensive margin (incumbent firm×product×destination) and the firm- and firm-market extensive margins to the growth of total French exports on a year-on-year basis or over the entire time frame of the sample (between 1996 and 2005). We use mid-point growth rates to account for entries and exits Bricongne *et al.* (2012). Initial size relates to firms' sales the first year of entry on a specific market on which they export up to 2005.

[4] The frequent changes in the combined nomenclature (CN8) prevents us to use this further degree of disaggregation of the customs' product classification.

Over the 1996-2005 period, we find that, on average, new firm-destination-product triplets represent only 12.3% of total export value after a year, but their share reaches 53.5% after a decade (27.3% due to new markets served by incumbents and 26.2% by new firms exporting, see Table A.1). The contribution of the extensive margin to aggregate exports is determined by three components of firm dynamics: entry, survival and post entry growth on new markets. Since new exporters typically do not survive more than a few years in export markets,[5] firm selection and growth are important drivers of aggregate trade growth over longer horizons, besides the size at entry. Column (2) of Table A.1 shows that pure growth after entry accounts for around 40% of the end-of-period share of newly created firm-destination-product triplets. The objective of our paper is precisely to understand how learning about demand can explain this post-entry dynamics.

Table A.1: Shares in end-of-period French aggregate exports

| | (1) Average yoy 1996/2005 | (2) Overall 1996/2005 |
|---|---|---|
| New exporters | 2.4% | 25.9% |
| *Initial size* | - | *16.4%* |
| *Growth since entry* | - | *9.6%* |
| New product-destination | 9.9% | 27.7% |
| *Initial size* | - | *16.7%* |
| *Growth since entry* | - | *11.0%* |
| Incumbent exporter-product-destination | 87.7% | 46.4% |
| Total | 100% | 100% |

Note: Source: French Customs. Column (1) presents the average contribution to year-on-year growth rates, i.e. the contribution of each subcomponent to the yearly growth rates, observed for each year of our sample, then averaged across years. Column (2) reports the contributions of each subcomponent to the total growth of French exports between 1996 and 2005. Initial entry measures firms' sales the year of first entry on a specific market on which they still export in 2005; and growth since entry measures the contribution of sales growth between the first entry and 2005.

**Firm-product-destination specific factors are a key component of sales' growth.** We decompose the variance of sales growth, in a way similar in spirit to Eaton et al. (2011).[6] We first regress firm-market specific sales growth on a set of destination-product-time dummies. The $R^2$ of such a regression is 0.14: market-specific dynamics play a limited role. Adding firm-product-time fixed effects increases the $R^2$ to 0.46, suggesting that supply-side factors such as productivity do a good job at explaining variations of firms' sales over time. However, it

[5]For French exporters, the average survival rate at the firm-product-destination level is 32% between the first and second year, and 9% over a five-year horizon.

[6]Eaton et al. (2011) show, using firm-destination data, that firm-specific effects explain well the probability of serving a market (57%), but less so sales variations conditional on selling in a market (39%). Munch and Nguyen (2014) find that the mean contribution of the firm component to unconditional sales variations is 49%. They also show that the firm-specific effects are more important for firms already established in a product-destination market. Lawless and Whelan (2014) find an adjusted pseudo-$R^2$ of 45% on a sample of Irish exporters.

appears clearly that sales growth remains largely driven by firm-market specific factors. Our paper concentrates on this part of firm dynamics, with the objective of understanding the extent to which it is consistent with firms learning about their demand.

Table A.2: Decomposition of the variance of sales

| Dependent var. | (1) Growth of exports | (2) | (3) Value of exports | (4) |
|---|---|---|---|---|
| Product-destination-time FE | Yes | Yes | - | - |
| Firm-product-time FE | - | Yes | - | - |
| Product-destination FE | - | - | Yes | - |
| Firm-product FE | - | - | Yes | - |
| Firm-product-destination FE | - | - | - | Yes |
| $R^2$ | 0.14 | 0.46 | 0.57 | 0.80 |

Note: OLS estimations based on French customs data. Each column contains the $R^2$ of a separate regression of the dependent variable on a specific set of fixed effects.

**Firm-market growth and its variance decline with age, conditional on size.** Columns (1), (4) and (5) of Table A.3 below shows the coefficients used to plot Figure 1 of the main text. Column (2) shows that similar results for firm growth are obtained when including in the estimations firm×product×year fixed effects. Specifications reported in columns (1)-(3) include dummies by decile of firms size computed by HS4-product×destination, and HS2 sector and year fixed effects. Firms size is defined as average firm×product×destination sales over $t$ and $t-1$ (Haltiwanger et al., 2013). Standard errors are clustered at the firm level. The exit probability in column (3) is estimated using a linear probability model. Column (4) includes year fixed effects and controls for average size, computed as the mean of average sales over $t$ and $t-1$ across firms by cohort. Standard errors are clustered at the market level.

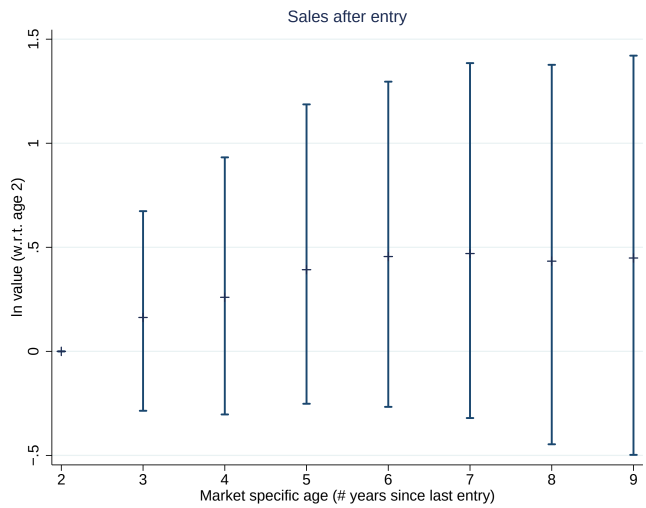
**Post-entry growth dynamics are heterogenous across survivors.** Finally, Figures A.1 and A.2 show that the heterogenous growth dynamics of quantities that we discuss in the main text also hold for the value of sales, and for different cohorts of firm-markets.

Table A.3: Age, growth and volatility of sales and exit rates

| Dep. var. | (1) Growth value (log) | (2) | (3) Exit probability | (4) Variance of growth value (log) |
|---|---|---|---|---|
| $\text{Age}_{ijkt} = 1$ | | | $0.296^{a}$ | |
| | | | (0.009) | |
| $\text{Age}_{ijkt} = 2$ | $0.437^{a}$ | $0.398^{a}$ | $0.169^{a}$ | $0.389^{a}$ |
| | (0.015) | (0.015) | (0.009) | (0.015) |
| $\text{Age}_{ijkt} = 3$ | $0.132^{a}$ | $0.174^{a}$ | $0.093^{a}$ | $0.258^{a}$ |
| | (0.014) | (0.015) | (0.008) | (0.015) |
| $\text{Age}_{ijkt} = 4$ | $0.079^{a}$ | $0.105^{a}$ | $0.052^{a}$ | $0.189^{a}$ |
| | (0.014) | (0.015) | (0.008) | (0.015) |
| $\text{Age}_{ijkt} = 5$ | $0.055^{a}$ | $0.069^{a}$ | $0.025^{a}$ | $0.138^{a}$ |
| | (0.014) | (0.015) | (0.008) | (0.015) |
| $\text{Age}_{ijkt} = 6$ | $0.047^{a}$ | $0.049^{a}$ | 0.008 | $0.093^{a}$ |
| | (0.014) | (0.015) | (0.008) | (0.015) |
| $\text{Age}_{ijkt} = 7$ | $0.032^{b}$ | $0.031^{b}$ | -0.002 | $0.054^{a}$ |
| | (0.014) | (0.015) | (0.008) | (0.015) |
| $\text{Age}_{ijkt} = 8$ | $0.033^{b}$ | $0.031^{b}$ | -0.007 | $0.040^{b}$ |
| | (0.014) | (0.015) | (0.007) | (0.016) |
| $\text{Age}_{ijkt} = 9$ | 0.018 | 0.011 | | $0.038^{b}$ |
| | (0.016) | (0.017) | | (0.016) |
| $\text{Size}_{ijkt/t-1}$ - decile 1 | $-0.251^{a}$ | $-0.156^{a}$ | $0.326^{a}$ | |
| | (0.012) | (0.021) | (0.004) | |
| $\text{Size}_{ijkt/t-1}$ - decile 2 | $-0.219^{a}$ | $-0.142^{a}$ | $0.290^{a}$ | |
| | (0.006) | (0.008) | (0.004) | |
| $\text{Size}_{ijkt/t-1}$ - decile 3 | $-0.210^{a}$ | $-0.193^{a}$ | $0.258^{a}$ | |
| | (0.005) | (0.006) | (0.004) | |
| $\text{Size}_{ijkt/t-1}$ - decile 4 | $-0.189^{a}$ | $-0.175^{a}$ | $0.228^{a}$ | |
| | (0.005) | (0.005) | (0.003) | |
| $\text{Size}_{ijkt/t-1}$ - decile 5 | $-0.169^{a}$ | $-0.156^{a}$ | $0.197^{a}$ | |
| | (0.005) | (0.005) | (0.003) | |
| $\text{Size}_{ijkt/t-1}$ - decile 6 | $-0.143^{a}$ | $-0.130^{a}$ | $0.163^{a}$ | |
| | (0.005) | (0.004) | (0.003) | |
| $\text{Size}_{ijkt/t-1}$ - decile 7 | $-0.120^{a}$ | $-0.105^{a}$ | $0.133^{a}$ | |
| | (0.004) | (0.004) | (0.003) | |
| $\text{Size}_{ijkt/t-1}$ - decile 8 | $-0.090^{a}$ | $-0.077^{a}$ | $0.097^{a}$ | |
| | (0.004) | (0.004) | (0.002) | |
| $\text{Size}_{ijkt/t-1}$ - decile 9 | $-0.051^{a}$ | $-0.039^{a}$ | $0.055^{a}$ | |
| | (0.004) | (0.004) | (0.002) | |
| Average $\text{Size}_{cjkt/t-1}$ | | | | $0.022^{a}$ |
| | | | | (0.001) |
| Observations | 1,666,317 | 1,456,113 | 3,061,865 | 348,536 |
| Year FE | Yes | Yes | Yes | Yes |
| Sector (HS2) FE | Yes | Yes | Yes | - |
| Firm-product-year FE | - | Yes | - | - |

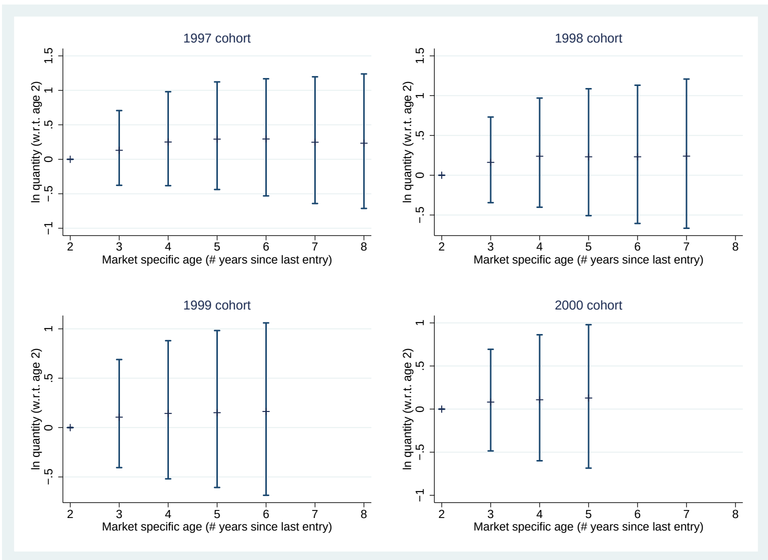
Robust standard errors clustered by firm (respectively destination-product in columns (4)) in parentheses. $^{c}$ significant at 10%; $^{b}$ significant at 5%; $^{a}$ significant at 1%. Size computed as average size in $t$ and $t+1$; size bins by decile are computed at the destination-product(HS4) level. The omitted age category is 10 years.

Figure A.1: Sales dynamics over time for surviving firms




Note: This figure plots statistics about market-specific firm sales values with respect to age. Values are normalized to 1 in age 2. The upper and lower limits of the boxes represent the first and last quartiles of the variable, with the median in between.

Figure A.2: Quantity dynamics over time for surviving firms: robustness




Note: This figure plots statistics about market-specific firm quantities with respect to age. Quantities are normalized to 1 in age 2. The upper and lower limits of the boxes represent the first and last quartiles of the variable, with the median in between.

# C Elasticity of substitution estimates

Table A.4 reports the descriptive statistics on the elasticities of substitution estimated from equation 18, for the full estimation sample (upper panel) and across HS6 products (lower panel). We first report the overall estimates, then the statistics obtained when products with insignificant $\beta$ coefficients (at the 5% level) are removed from the sample, and finally the statistics obtained when products with $\sigma_k < 1$ are excluded. As can be seen from the upper panel, insignificant coefficients only represent 1.6% of the observations in the final sample, while dropping theory inconsistent elasticities (lower than 1) further eliminates only 0.1% of the observations. This clearly show that, whenever we can precisely estimate these elasticities, we get plausible coefficients.

Across products, our estimates yield a mean (resp. median) $\sigma_k$ of 7.17 (resp. 5.51) after dropping insignificant or theory-inconsistent ones (lower panel). The median is largely unaffected by our cleaning rules. Except in the case in which insignificant estimates are kept, the distribution of $\sigma_k$ does not contain extreme value: the 99% is equal to 28.7. The last three rows report $\sigma_k$ for different categories of products according to the Rauch (1999)'s liberal classification. As expected, differentiated goods exhibit a mean (resp. median) of 6.2 (resp. 5.2), lower than referenced priced goods (9.1 and 7.2 respectively) and homogenous goods (11.1 and 9.0 resp.). Those means are statistically different at the 1% level, with t-stat of -12.4 (differentiated vs. referenced), -13.3 (differentiated vs. homogenous), and -3.1 (referenced vs. homogenous).

Table A.4: Statistics on elasticities of substitution

| | Obs. | Mean | S.D. | 1% | 25% | Median | 75% | 99% |
|---|---|---|---|---|---|---|---|---|
| | | | | Full sample | | | | |
| $\sigma_k$, all estimates | 1883748 | 7.30 | 60.42 | 2.21 | 3.58 | 5.11 | 6.70 | 33.06 |
| $\sigma_k$, if $\beta$ significant | 1854359 | 6.20 | 4.79 | 2.24 | 3.57 | 5.09 | 6.59 | 26.16 |
| $\sigma_k$, if $\beta$ significant and $\sigma_k > 1$ | 1854141 | 6.20 | 4.78 | 2.24 | 3.57 | 5.09 | 6.59 | 26.16 |
| | | | | Across HS6 products | | | | |
| $\sigma_k$, all estimates | 3542 | 13.91 | 221.62 | -69.55 | 3.82 | 5.83 | 10.27 | 116.10 |
| $\sigma_k$, if $\beta$ significant | 2780 | 7.10 | 5.55 | 1.68 | 3.86 | 5.49 | 8.41 | 28.73 |
| $\sigma_k$, if $\beta$ significant and $\sigma_k > 1$ | 2767 | 7.17 | 5.47 | 1.89 | 3.88 | 5.51 | 8.42 | 28.73 |
| $\sigma_k$, if $> 1$, differentiated goods | 1778 | 6.24 | 4.21 | 2.00 | 3.77 | 5.17 | 7.20 | 21.82 |
| $\sigma_k$, if $> 1$, referenced priced goods | 670 | 9.09 | 6.78 | 1.80 | 4.46 | 7.24 | 11.61 | 32.98 |
| $\sigma_k$, if $> 1$, homogenous goods | 159 | 11.14 | 8.71 | 1.63 | 5.23 | 8.97 | 15.07 | 50.82 |

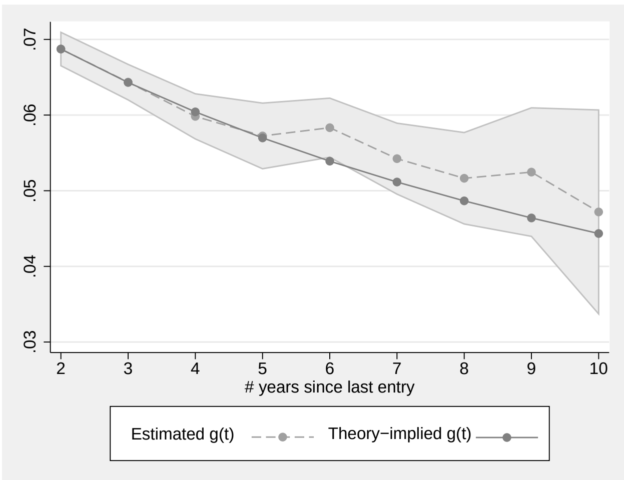
Source: Authors computations from French Customs data. Elasticities of substitution estimated from equation (17). $\sigma_k$, if significant $\beta$ means that we keep only the estimate when $\beta$ estimated from equation (17) is statistically different from zero at the 5% level. $\sigma_k$, if $> 1$ means that we further drop the observation if $\sigma_k > 1$. The classification of goods into differentiated, referenced priced and homogeneous comes from Rauch (1999).

# D Main results – Graphical representation

The figures below depicts the coefficients of 2, column (4). Equation (18) predicts that these coefficient should follow the following shape: $g_t = \frac{1}{\sigma_\varepsilon^2/\sigma_{jk0}^2+t}$ with $g_t$ measuring the speed of learning. To determine whether our set of estimated coefficients significantly differ from this shape, in Figure we have taken year 2 (the first coefficient) as a benchmark; from this coefficient we can infer the value of $\sigma_\varepsilon^2/\sigma_{jk0}^2 \approx 12$. Assuming the coefficient of year 2 is indeed correct, the shape of our coefficients is quite similar to the one implied by our functional form assumption.

We can go further and test this restriction: does our model perform significantly better than a model in which we would constrain the coefficients (from year 3 onwards) to follow the shape of $g_t$? When we test these restrictions, they are rejected at the 5% level: the shape implied by our coefficient is different from the one implied by the normality assumption (the p-value associated with the hypothesis that the models are the same is 0.02). However, as is apparent in Figure R3.1 below, this is mostly due to a difference in the coefficients after age 5. In fact, when we consider all coefficients but the one of year 6, the restrictions are no longer rejected (the p-value of 0.17). When we concentrate on the first four years (for which we have more observations and therefore more precise estimates), the p-value of the F-test of the restrictions is as high as 0.66.

Figure A.3: Functional form assumption

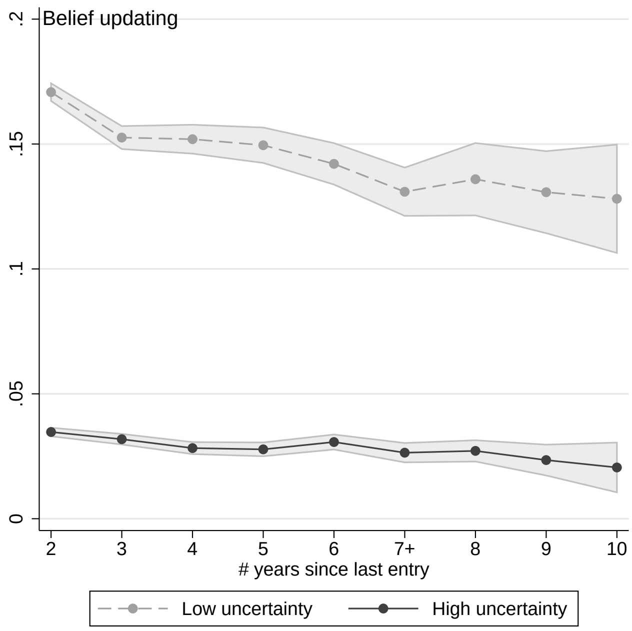



Note: This figure plot the coefficient and 95% confidence intervals from Table 2, column (4), and the coefficient implied by our functional form assumption $g_t = 1/(\sigma_\varepsilon^2/\sigma_{jk0}^2 + t)$. At age 2 we assume that the coefficient is correctly estimated and infer the rest of the theory-based coefficient from the expression of $g_t$.

# E Belief updating and uncertainty

To illustrate the impact of uncertainty on the updating process and how it evolves with age, we split our sample into markets with low (below the first quartile of uncertainty) and high (above the third quartile) uncertainty in Table A.5 and Figure A.4. We still find evidence for the updating process on both sub-samples but the average level of belief updating following a demand shock is larger on less uncertain markets (0.171 versus 0.035 for firms of age two). As expected from prediction 2, the profile of learning is also flatter on more uncertain markets.

Figure A.4: Uncertainty and belief updating




The figure plots the coefficients of regressions similar to Table 2, column (4), ran on two different sub-samples defined according to the market level of uncertainty (below the first quartile and above the third quartile). Uncertainty is computed as the standard deviation of the demand shocks $a_{ijkt}$, computed by market. Grey areas represent 90% confidence bands.

Table A.5: Prediction 1: the role of uncertainty (subsamples)

| | (1) | (2) |
|---|---|---|
| Dep. var. | $\Delta\varepsilon^q_{ijk,t+1}$ | |
| Uncertainty | High | Low |
| $(a_{ijkt} - \varepsilon^q_{ijkt}) \times \text{Age}_{ijkt} = 2$ | $0.035^a$ | $0.171^a$ |
| | (0.001) | (0.002) |
| $\times\ \text{Age}_{ijkt} = 3$ | $0.032^a$ | $0.153^a$ |
| | (0.001) | (0.003) |
| $\times\ \text{Age}_{ijkt} = 4$ | $0.028^a$ | $0.152^a$ |
| | (0.001) | (0.004) |
| $\times\ \text{Age}_{ijkt} = 5$ | $0.028^a$ | $0.150^a$ |
| | (0.002) | (0.004) |
| $\times\ \text{Age}_{ijkt} = 6$ | $0.031^a$ | $0.142^a$ |
| | (0.002) | (0.005) |
| $\times\ \text{Age}_{ijkt} = 7$ | $0.026^a$ | $0.131^a$ |
| | (0.002) | (0.006) |
| $\times\ \text{Age}_{ijkt} = 8$ | $0.027^a$ | $0.136^a$ |
| | (0.003) | (0.009) |
| $\times\ \text{Age}_{ijkt} = 9$ | $0.023^a$ | $0.131^a$ |
| | (0.004) | (0.010) |
| $\times\ \text{Age}_{ijkt} = 10$ | $0.021^a$ | $0.128^a$ |
| | (0.006) | (0.013) |
| Observations | 454040 | 438324 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. Age dummies included alone but coefficients not reported. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). $\text{Age}_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). Uncertainty is the standard deviation of $a_{ijkt}$, computed by market. High and low mean above the third quartile and below the first quartile of the uncertainty variable.

# F Extensions of the model

We consider in this section alternative versions of the model and discuss their implications for our identification strategy.

## F.1 Firms set price first, monopolistic competition

Let us first consider the opposite of our baseline assumption: prices are set first, before demand shocks are realized. Once the demand shock is observed, firms then choose quantities. The maximization problem becomes:

$$\max_p \int \pi_{ijkt} dG_{t-1}(a_{ijkt}) \qquad \text{s.t.} \qquad q_{ijkt} = e^{a_{ijkt}} p_{ijkt}^{-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}}$$

$$\max_p p_{ijkt}^{1-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}} \mathbb{E}_{t-1}\left[e^{a_{ijkt}}\right] - \frac{w_{it}}{\varphi_{ikt}} \mathbb{E}_{t-1}\left[e^{a_{ijkt}}\right] p_{ijkt}^{-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}} - F_{ijk}$$

From the FOC and the constraint we get:

$$\begin{aligned} p^*_{ijkt} &= \frac{\sigma_k}{\sigma_k - 1} \frac{w_{it}}{\varphi_{ikt}} \\ q^*_{ijkt} &= e^{a_{ijkt}} \left( \frac{\sigma_k}{\sigma_k - 1} \frac{w_{it}}{\varphi_{ikt}} \right)^{-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}} \end{aligned}$$

With constant price elasticity, firms choose prices as constant mark-ups over marginal costs: prices do not depend on sales, but solely on supply side characteristics. Quantities then adjust to the demand level. Therefore, if prices are determined before observing the demand shocks, while quantities can fully adjust to it, neither prices nor quantities depend on firm beliefs. We would get:

$$\begin{aligned} \varepsilon^q_{ijkt} &= \ln Z^q_{ijkt} = a_{ijkt} \\ \varepsilon^p_{ijkt} &= \ln Z^p_{ijkt} = 0 \end{aligned}$$

Regressing $\varepsilon^p_{ijkt}$ on $\varepsilon^q_{ijkt}$ should generate insignificant $\widehat{\beta}$ coefficients and the absolute value of $\varepsilon^q_{ijkt}$ should not decrease with age.

## F.2 Firms set price first, oligopolistic competition

Second, we consider the case of an oligopolistic market structure with Bertrand competition (so still price first), to allow for variable markups. Formally, we assume that consumers in country $j$ maximize utility derived from the consumption of goods from $K$ sectors. Each sector is composed of a small enough set of differentiated varieties of product $k$:

$$\begin{aligned} U_j &= \mathbb{E}\sum_{t=0}^{+\infty}\beta^t \ln\left(C_{jt}\right), \text{ with } C_{jt} = \prod_{k=0}^{K} C_{jkt}^{\mu_k} \\ \text{and } C_{jkt} &= \left(\sum_{\Omega_{kt}} \left(e^{a_{ijkt}}\right)^{\frac{1}{\sigma_k}} q_{ijkt}(\omega)^{\frac{\sigma_k-1}{\sigma_k}} d\omega\right)^{\frac{\sigma_k}{(\sigma_k-1)}} \end{aligned}$$

with $\Omega_{kt}$ the (small enough) set of varieties of product $k$ available at time $t$. We assume that firms take income $Y_{jt}$ as constant, i.e. we assume that $K$ is large enough.

The upper tier utility maximization implies $C_{jkt} = \frac{\mu_k Y_{jt}}{P_{jkt}}$. It follows the demand in market $j$ at time $t$ for a variety of product $k$:

$$q_{ijkt} = e^{a_{ijkt}} p_{ijkt}^{-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}} = C_{jkt} e^{a_{ijkt}} \frac{p_{ijkt}^{-\sigma_k}}{P_{jkt}^{-\sigma_k}}$$

with the price index of sector $k$ in country $j$ defined as:

$$P_{jkt} = \left(\sum_{\Omega_{kt}} e^{a_{ijkt}} p_{ijkt}^{1-\sigma_k} di\right)^{\frac{1}{1-\sigma_k}}$$

The firm maximization program writes:

$$\max_p \int \pi_{ijkt} dG_{t-1}(a_{ijkt}) \qquad \text{s.t.} \qquad q_{ijkt} = e^{a_{ijkt}} p_{ijkt}^{-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}}$$

It follows:

$$\begin{aligned} p_{ijkt}^* &= \frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1} \frac{w_{it}}{\varphi_{ikt}} \\ q_{ijkt}^* &= e^{a_{ijkt}} \left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1} \frac{w_{it}}{\varphi_{ikt}}\right)^{-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}} \end{aligned}$$

with

$$\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right] = \sigma_k - (\sigma_k - 1)\,\mathbb{E}_{t-1}\left[s_{ijkt}\right]$$

where $\mathbb{E}_{t-1}\left[s_{ijkt}\right]$ is the expected market share at the beginning of period $t$. The residuals from the estimation in logs with fixed effects are:

$$\begin{aligned} \varepsilon_{ijkt}^q &= a_{ijkt} - \sigma_k \ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{E_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right) \\ \varepsilon_{ijkt}^p &= \ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right) \end{aligned}$$

As the demand shock now appears in the residual quantities, we regress $\varepsilon_{ijkt}^q$ on $\varepsilon_{ijkt}^p$:

$$a_{ijkt} - \sigma_k \ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right) = \beta\left(\ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right)\right) + \lambda_{ijk} + v_{ijkt}$$

We obtain:

$$\widehat{\beta} = -\sigma_k \text{ and } \widehat{v}_{ijkt} = \varepsilon_{ijkt}$$

and

$$\widehat{\lambda_{ijk}} + \widehat{v}_{ijkt} = a_{ijkt}$$

To test this alternative specification, we look at the dynamics of prices, that reflect the evolution of firms beliefs:

$$\Delta\varepsilon^{p}_{ijkt+1} = \Delta \ln\left(\frac{E_t\left[\varepsilon(s_{ijkt})\right]}{E_t\left[\varepsilon(s_{ijkt})\right]-1}\right)$$

As $\frac{\partial\left(\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]\right)}{\partial\left(\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)} < 0$, a positive shock (i.e. generating a positive updating) implies a decrease in the expected price elasticity and an increase in markup. In Table A.6, we assess the empirical relevance of this alternative model. We do not find evidence of a positive relationship between prices and demand shocks. Overall, our data are therefore not consistent with the assumption of firms choosing their price first.

Table A.6: Prediction 1: demand shocks and beliefs updating (assuming Bertrand)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^{p}_{ijkt+1}$ | | |
| $a_{ijkt}$ | -0.004$^{a}$ | -0.005$^{a}$ | -0.005$^{a}$ | |
| | (0.001) | (0.001) | (0.001) | |
| $\times$ Age$_{ijkt}$ | | 0.000 | 0.000 | |
| | | (0.000) | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | | | -0.003$^{a}$ |
| | | | | (0.001) |
| $\times$ Age$_{ijkt} = 3$ | | | | -0.001$^{a}$ |
| | | | | (0.000) |
| $\times$ Age$_{ijkt} = 4$ | | | | -0.001$^{a}$ |
| | | | | (0.000) |
| $\times$ Age$_{ijkt} = 5$ | | | | -0.000 |
| | | | | (0.000) |
| $\times$ Age$_{ijkt} = 6$ | | | | -0.001$^{a}$ |
| | | | | (0.000) |
| $\times$ Age$_{ijkt} = 7$ | | | | -0.001$^{a}$ |
| | | | | (0.000) |
| $\times$ Age$_{ijkt} = 8$ | | | | -0.000 |
| | | | | (0.000) |
| $\times$ Age$_{ijkt} = 9$ | | | | -0.000 |
| | | | | (0.000) |
| $\times$ Age$_{ijkt} = 10$ | | | | 0.001 |
| | | | | (0.000) |
| Age$_{ijkt}$ | 0.001$^{a}$ | 0.001$^{a}$ | 0.001$^{a}$ | |
| | (0.000) | (0.000) | (0.000) | |
| Observations | 1883748 | 1883748 | 1883748 | 1883748 |

Robust standard errors clustered by firm in parentheses (bootstrapped in column (3)). $^{c}$ significant at 10%; $^{b}$ significant at 5%; $^{a}$ significant at 1%. Age dummies included alone in columns (4) but coefficients not reported. Shocks $a_{ijkt}$ are computed assuming Bertrand competition, i.e. by regressing $\varepsilon^{q}_{ijkt}$ on $\varepsilon^{p}_{ijkt}$ instead of the opposite. See text for more details. Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit).

## F.3 Partial quantity adjustment

Here, we maintain our assumption that quantities are set first, but allow firms to observe part of the demand shock before taking their quantity decision. Prices then fully adjust once the other part of the demand shock is observed.

Suppose that the demand shock $a_{ijkt}$ can be decomposed into 2 components: $a_{ijkt} = a^1_{ijkt} + a^2_{ijkt}$, with $a^1_{ijkt} \sim \mathcal{N}\left(\overline{a}^1_{ijk}, \varsigma\sigma^2_\varepsilon\right)$, $a^2_{ijkt} \sim \mathcal{N}\left(\overline{a}^2_{ijk}, (1-\varsigma)\,\sigma^2_\varepsilon\right)$ and $\overline{a}^1_{ijk} + \overline{a}^2_{ijk} = \overline{a}_{ijk}$. Firms can observe $a^1_{ijkt}$ before taking their quantity decision. $a^2_{ijkt}$ is then realized and firms fully adjust their prices. For simplicity, we assume that $a^1_{ijkt}$ does not bring additional information, i.e. $Cov(a^1_{ijkt}, a^2_{ijkt}) = 0$.

$\overline{a}^1_{ijk}$ and $\varsigma$ capture the relative importance of the first (observed) shock and therefore the importance of the learning process for firms: if $a^1_{ijkt}$ captures the entire demand shock ($\overline{a}_{ijk} = \overline{a}^1_{ijk}$ and $\varsigma = 1$), there is nothing to learn about. Beliefs are only related to $a^2_{ijkt}$, the part of the demand shock which is not observed at the time of the quantity decision. The distribution of beliefs is now described by $G_{t-1}(a^2_{ijkt})$.

After having observed $a^1_{ijkt}$, firms choose quantities by maximizing expected profits subject to demand. We get:

$$\max_q \int \pi_{ijkt} dG_{t-1}(a^2_{ijkt}) = \max_q q_{ijkt}^{1-\frac{1}{\sigma_k}} \left(\frac{\mu_k Y_{jt}}{P^{1-\sigma_k}_{jkt}}\right)^{\frac{1}{\sigma_k}} e^{\frac{a^1_{ijkt}}{\sigma_k}} \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right] - \frac{w_{it}}{\varphi_{ikt}} q_{ijkt} - F_{ijk}.$$

The constraint can now be written $p_{ijkt} = \left(\frac{\mu_k Y_{jt} e^{a^1_{ijkt}} e^{a^2_{ijkt}}}{q_{ijkt} P^{1-\sigma_k}_{jkt}}\right)^{\frac{1}{\sigma_k}}$. From the FOC and the constraint we get:

$$\begin{aligned} p^*_{ijkt} &= \left(\frac{\sigma_k}{\sigma_k - 1}\frac{w_{it}}{\varphi_{ikt}}\right)\left(\frac{e^{\frac{a^2_{ijkt}}{\sigma_k}}}{\mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]}\right) \\ q^*_{ijkt} &= \left(\frac{\sigma_k}{\sigma_k - 1}\frac{w_{it}}{\varphi_{ikt}}\right)^{-\sigma_k}\left(\frac{\mu_k Y_{jt}}{P^{1-\sigma_k}_{jkt}}\right) e^{a^1_{ijkt}} \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]^{\sigma_k}. \end{aligned}$$

As before, quantities depend on firms' beliefs while prices are still a constant markup over marginal cost in expected terms. We get:

$$\begin{aligned} \varepsilon^q_{ijkt} &= \ln Z^q_{ijkt} = a^1_{ijkt} + \sigma_k \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right] \\ \varepsilon^p_{ijkt} &= \ln Z^p_{ijkt} = \frac{1}{\sigma_k} a^2_{ijkt} - \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]. \end{aligned}$$

Note that if $\overline{a}^1_{ijk} = \overline{a}_{ijk}$ and $\varsigma = 1$, all the demand shock is observed and $\varepsilon^q_{ijkt}$ captures the demand shock only while $\varepsilon^p_{ijkt}$ does not depend neither on the demand shock, nor on firm beliefs (which are irrelevant in that case). This case is equivalent to the one where prices are set first. If on the other hand $\overline{a}^1_{ijk} = \varsigma = 0$, we are back to our baseline assumption of fixed quantities.

Importantly, all our theoretical predictions still hold in the intermediate case. In particular, equation (11) still describes the evolution of beliefs, which are now related to the distribution of $a^2_{ijkt}$.

**Identification**. If quantities can partly adjust, $\varepsilon^q_{ijkt}$ captures both the firm beliefs and part of the demand shock, i.e. our measure of beliefs becomes noisy. This is innocuous when looking at the dynamics of $\varepsilon^q_{ijkt}$ (see 6.1) or when looking at the relationship between the variance of growth rates and age cohorts (see 6.2), but it has implications for the identification of the demand shocks $v_{ijkt}$. Regressing $\varepsilon^p_{ijkt}$ on $\varepsilon^q_{ijkt}$ gives:

$$\left(\frac{1}{\sigma_k}a^2_{ijkt} - \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]\right) = \beta\left(a^1_{ijkt} + \sigma_k \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]\right) + \lambda_{ijk} + v_{ijkt}.$$

It follows:

$$\widehat{\beta} = -\frac{1}{\sigma_k}\Lambda_P \text{ with } \Lambda_P = \frac{\mathbb{V}\left(\widetilde{\sigma_k \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]}\right)}{\mathbb{V}\left(\widetilde{\sigma_k \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]}\right) + \mathbb{V}\left(\widetilde{a^1_{ijkt}}\right)}$$

where variables with the sign$^\sim$ are demeaned in the $ijk$ dimension. We get $0 < \Lambda_P < 1$: $\widehat{\beta}$ is underestimated due to the attenuation bias introduced by the noisy measure of firms' beliefs.

Hence, the estimated shock $\widehat{\lambda_{ijk}} + \widehat{v}_{ijkt}$ may be biased, but the direction of this bias is unclear as we would like now to isolate $a^2_{ijkt}$ and not $a_{ijkt} = a^1_{ijkt} + a^2_{ijkt}$. Indeed, firms now form beliefs about the part of the demand shock which is not observed at the time of the quantity decision. $\widehat{v}_{ijkt}$ may thus be larger or smaller than $a^2_{ijkt}$.

Suppose for instance that $\Lambda_P = 1/2$. This implies that $\mathbb{V}\left(\widetilde{\sigma_k \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]}\right) = \mathbb{V}\left(\widetilde{a^1_{ijkt}}\right)$. In this case our estimated demand shock would be:

$$\widehat{\lambda_{ijk}} + \widehat{v}_{ijkt} = \frac{1}{\sigma_k}a^2_{ijkt} + 2\sigma_k\left(a^1_{ijkt} - \sigma_k \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]\right)$$

The direction of the bias depends on, among $\sigma_k \ln \mathbb{E}_{t-1}\left[e^{\frac{a^2_{ijkt}}{\sigma_k}}\right]$ and $a^1_{ijkt}$, which one is the most important component of $\varepsilon^q_{ijkt}$.

**Equation under test**. We obtain:

$$\Delta\varepsilon^q_{ijkt+1} = \sigma_k \Delta \ln \mathrm{E}_t\left[e^{\frac{a^2_{ijkt+1}}{\sigma_k}}\right] + \Delta a^1_{ijkt+1}$$

It is worth noting that $\Delta\varepsilon^q_{ijkt}$ still fully captures the updating process, as $\Delta a^1_{ijkt+1} = 0$ in expected terms. It is now about the true value of $\overline{a}^2_{ijk}$. We get: $\Delta\widetilde{\theta}_t = g_t\left(a^2_{ijkt} - \widetilde{\theta}_{t-1}\right)$, and thus $\Delta \ln \mathrm{E}_t \left[ e^{\frac{a^2_{ijkt+1}}{\sigma_k}} \right] = \frac{1}{\sigma_k} \left( \Delta \widetilde{\theta}_t + \frac{\widetilde{\sigma}^2_t - \widetilde{\sigma}^2_{t-1}}{2\sigma_k} \right)$.

It follows:

$$\Delta \varepsilon^q_{ijkt+1} = g_t \left( a^2_{ijkt} - \widetilde{\theta}_{t-1} - \frac{\widetilde{\sigma}^2_{t-1} - \widetilde{\sigma}^2_t}{g_t 2\sigma_k} \right) + \Delta a^1_{ijkt+1}$$

Further,

$$\widetilde{\theta}_{ijkt-1} = \varepsilon^q_{ijkt} - \frac{\widetilde{\sigma}^2_{t-1} + \sigma^2_\varepsilon}{2\sigma_k} - a^1_{ijkt}$$

Hence:

$$\Delta \varepsilon^q_{ijkt+1} = g_t \left( a^2_{ijkt} - \varepsilon^q_{ijkt} \right) + g_t \left( \frac{\sigma^2_\varepsilon}{2\sigma_k} + a^1_{ijkt} \right) + \Delta a^1_{ijkt+1}$$

The possibility of some partial quantity adjustment may just generate some extra unconditional growth at the firm-market level (see the second term), as $\Delta a^1_{ijkt+1} = 0$ in expected terms. So, beyond the fact that $a^2_{ijkt}$ may be biased upwards or downwards, our strategy is left unaffected.

How this potential bias may affect our results on beliefs updating (prediction 1)? Consider our baseline specification, equation (19). There are two distinct issues here. First, $\widehat{\alpha}_1$ – the average extent of belief updating – might be upward or downward biased, depending on the direction of the bias of our estimated demand shocks. Second, if this bias depends on firm-market age, this may affect how $\widehat{\alpha}_1$ evolves with age, which is key for our findings.

As discussed in the main text, a simple way to gauge the importance of this issue is to focus on sectors or destinations for which quantities are more likely to be rigid (those for which $\widehat{\lambda_{ijk}} + \widehat{v}_{ijkt}$ is more likely to be correctly estimated) and to compare the results with our baseline estimates of Table 2. We expect less quantity adjustment for complex goods (in which many different relationship-specific inputs are used in the production process) and in destinations characterized by longer time-to-ship. In Table A.7, we restrict our sample to sectors or destinations which are above the sample median in terms of time-to-ship or input complexity. Data on sector-specific complexity comes from Nunn (2007), and data on time-to-ship between France's main port (Le Havre) and each of the destinations' main port from Berman *et al.* (2013).

Results in Table A.7 show that the updating of the firms' beliefs following a demand shock is quantitatively close to our baseline estimates (columns (1) and (4)), which suggests that the bias of our estimated demand shocks, if any, is limited. Further, the coefficient on the interaction term between demand shocks and age (columns (2)-(3) and (5)-(6)), is also similar our baseline estimates. The coefficient on the interaction term between demand shocks and age is slightly lower than our baseline in the case of complex goods (col. (5) of Table A.7). In column (6), however, we see that this result is only driven by effect of the last age category, 10 years of experience, which is itself quite imprecisely estimated.

Altogether, these results suggest that our assumption of fixed quantities should not be rejected in light of our data.

Table A.7: Prediction 1: robustness (fixed quantities)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^q_{ijk,t+1}$ | | | $\Delta\varepsilon^q_{ijk,t+1}$ | |
| Sample | | Long time-to-ship | | | Complex goods | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | $0.072^a$ | $0.081^a$ | | $0.077^a$ | $0.084^a$ | |
| | (0.002) | (0.003) | | (0.002) | (0.003) | |
| $\times$ Age$_{ijkt}$ | | $-0.003^b$ | | | $-0.002^a$ | |
| | | (0.001) | | | (0.001) | |
| $\times$ Age$_{ijkt} = 2$ | | | $0.075^a$ | | | $0.080^a$ |
| | | | (0.002) | | | (0.002) |
| $\times$ Age$_{ijkt} = 3$ | | | $0.071^a$ | | | $0.078^a$ |
| | | | (0.003) | | | (0.002) |
| $\times$ Age$_{ijkt} = 4$ | | | $0.068^a$ | | | $0.075^a$ |
| | | | (0.004) | | | (0.003) |
| $\times$ Age$_{ijkt} = 5$ | | | $0.062^a$ | | | $0.070^a$ |
| | | | (0.005) | | | (0.004) |
| $\times$ Age$_{ijkt} = 6$ | | | $0.069^a$ | | | $0.074^a$ |
| | | | (0.005) | | | (0.004) |
| $\times$ Age$_{ijkt} = 7$ | | | $0.068^a$ | | | $0.071^a$ |
| | | | (0.008) | | | (0.004) |
| $\times$ Age$_{ijkt} = 8$ | | | $0.062^a$ | | | $0.072^a$ |
| | | | (0.010) | | | (0.005) |
| $\times$ Age$_{ijkt} = 9$ | | | $0.063^a$ | | | $0.065^a$ |
| | | | (0.013) | | | (0.008) |
| $\times$ Age$_{ijkt} = 10$ | | | $0.040^b$ | | | $0.079^a$ |
| | | | (0.018) | | | (0.012) |
| Age$_{ijkt}$ | $-0.029^a$ | $-0.029^a$ | | $-0.031^a$ | $-0.031^a$ | |
| | (0.001) | (0.001) | | (0.001) | (0.001) | |
| Observations | 358418 | 358418 | 358418 | 800015 | 800015 | 800015 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). Age dummies included alone in columns (3) and (6) but coefficients not reported. Complex goods means in the above the sample median of the variable, and large time-to-ship above the median for extra-EU observations. Data on sector-specific complexity comes from Nunn (2007), and data on time-to-ship between France's main port (Le Havre) and each of the destinations' main port from Berman *et al.* (2013).

## F.4 Oligopolistic Competition - Cournot

We investigate here the possibility that firms markups are variable. To do so, we consider the same model as before, but assume that competition is oligopolistic within sectors and not monopolistic. Formally, we assume that consumers in country $j$ maximize utility derived from the consumption of goods from $K$ sectors. Each sector is composed of a small enough set of differentiated varieties of product $k$:

$$\begin{aligned} U_j &= E\sum_{t=0}^{+\infty}\beta^t \ln\left(C_{jt}\right), \text{ with } C_{jt} = \prod_{k=0}^{K} C_{jkt}^{\mu_k} \\ \text{and } C_{jkt} &= \left(\sum_{\Omega_{kt}} \left(e^{a_{ijkt}}\right)^{\frac{1}{\sigma_k}} q_{ijkt}(\omega)^{\frac{\sigma_k-1}{\sigma_k}} d\omega\right)^{\frac{\sigma_k}{(\sigma_k-1)}} \end{aligned}$$

with $\rho$ the discount factor. $\Omega_{kt}$ the (small enough) set of varieties of product $k$ available at time $t$, and $\sum^K \mu_k = 1$. We assume that firms take income $Y_{jt}$ as constant, i.e. we assume that $K$ is large enough.

The upper tier utility maximization implies $C_{jkt} = \frac{\mu_k Y_{jt}}{P_{jkt}}$. It follows the demand in market $j$ at time $t$ for a variety of product $k$:

$$q_{ijkt} = e^{a_{ijkt}} p_{ijkt}^{-\sigma_k} \frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}} = C_{jkt} e^{a_{ijkt}} \frac{p_{ijkt}^{-\sigma_k}}{P_{jkt}^{-\sigma_k}}$$

with the price index of sector $k$ in country $j$ defined as:

$$P_{jkt} = \left(\sum_{\Omega_{kt}} e^{a_{ijkt}} p_{ijkt}^{1-\sigma_k} di\right)^{\frac{1}{1-\sigma_k}}$$

**Equilibrium**. Firms maximize profits, given the demand they face. They maximize:

$$\max_q \int \pi_{ijkt} dG_{t-1}(a_{ijkt}) \qquad \text{s.t.} \qquad p_{ijkt} = \left(\frac{C_{jkt} e^{a_{ijkt}}}{q_{ijkt}}\right)^{\frac{1}{\sigma_k}} \frac{\mu_k Y_{jt}}{C_{jkt}}$$

We get:

$$\frac{\partial \int \pi_{ijkt} dG_{t-1}(a_{ijkt})}{\partial q_{ijkt}} = \left(\left(C_{jkt}\right)^{\frac{1}{\sigma_k}} q_{ijkt}^{-\frac{1}{\sigma}} \frac{\mu_k Y_{jt}}{C_{jkt}} \mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma}}\right]\right)\left(1 - \frac{1}{\sigma_k}\right)\left(1 - \mathbb{E}_{t-1}\left[s_{ijkt}\right]\right) - \frac{w_{it}}{\varphi_{ikt}}$$

where $\mathbb{E}_{t-1}\left[s_{ijkt}\right]$ is the expected market share at the beginning of period $t$:

$$\mathbb{E}_{t-1}\left[s_{ijkt}\right] = \frac{\mathbb{E}_{t-1}\left[q_{ijkt} p_{ijkt}\right]}{C_{jkt} P_{jkt}} = \frac{\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma}}\right] \mathbb{E}_{t-1}\left[p_{ijkt}^{1-\sigma_k}\right]}{P_{jkt}^{1-\sigma_k}}$$

It follows:

$$
\begin{aligned}
q^*_{ijkt} &= \left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\frac{w_{it}}{\varphi_{ikt}}\right)^{-\sigma_k}\frac{\mu_k Y_{jt}}{P^{1-\sigma_k}_{jkt}}\left(\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]\right)^{\sigma_k} \\
p^*_{ijkt} &= \frac{e^{\frac{a_{ijkt}}{\sigma_k}}}{\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]}\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\frac{w_{it}}{\varphi_{ikt}}\right)
\end{aligned}
$$

with

$$
\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right] = \frac{1}{\frac{1}{\sigma_k}+\left(1-\frac{1}{\sigma_k}\right)\mathbb{E}_{t-1}\left[s_{ijkt}\right]}
$$

**Identification**. Purged quantities and prices:

$$
\begin{aligned}
\varepsilon^q_{ijkt} &= \sigma_k\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]-\ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right)\right) \\
\varepsilon^p_{ijkt} &= \frac{a_{ijkt}}{\sigma_k}-\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]-\ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right)\right)
\end{aligned}
$$

We regress $\varepsilon^p_{ijkt}$ on $\varepsilon^q_{ijkt}$:

$$
\frac{a_{ijkt}}{\sigma_k}-\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]-\ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right)\right)
$$

$$
=\beta\left(\sigma_k\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]-\ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right)\right)\right)+\lambda_{ijk}+v_{ijkt}
$$

We get:

$$
\widehat{\beta}=-\frac{1}{\sigma_k} \quad \text{and} \quad \widehat{v}_{ijkt}=\frac{1}{\sigma_k}\varepsilon_{ijkt}
$$

And

$$
\widehat{\lambda_{ijk}}+\widehat{v}_{ijkt}=\frac{a_{ijkt}}{\sigma_k}
$$

Put differently, our strategy to identify demand signals is still valid if firms markups are variable. This is because firms' markups affect purged prices and quantities in the same way as beliefs do.

**Equation under test**. We obtain:

$$
\Delta\varepsilon^q_{ijkt+1}=\sigma_k\left[\Delta\ln \mathrm{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]-\Delta\ln\left(\frac{E_t\left[\varepsilon(s_{ijkt})\right]}{E_t\left[\varepsilon(s_{ijkt})\right]-1}\right)\right]
$$

$\Delta\varepsilon^q_{ijkt}$ does not only capture the updating process, but is also impacted by changes in expected mark-up.

The updating process itself does not change however, we still get $\Delta\widetilde{\theta}_t = g_t\left(a_{ijkt}-\widetilde{\theta}_{t-1}\right)$, and $\Delta \ln \mathrm{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]=\frac{1}{\sigma_k}\left(\Delta\widetilde{\theta}_t+\frac{\widetilde{\sigma}^2_t-\widetilde{\sigma}^2_{t-1}}{2\sigma_k}\right)$.

It follows

$$\Delta\varepsilon^q_{ijkt+1} = g_t\left(\left(a_{ijkt} - \widetilde{\theta}_{t-1}\right) - \frac{\widetilde{\sigma}^2_{t-1} - \widetilde{\sigma}^2_t}{g_t 2\sigma_k}\right) - \sigma_k \Delta \ln\left(\frac{E_t\left[\varepsilon(s_{ijkt})\right]}{E_t\left[\varepsilon(s_{ijkt})\right] - 1}\right)$$

Further,

$$\widetilde{\theta}_{ijkt-1} = \varepsilon^q_{ijkt} - \frac{\widetilde{\sigma}^2_{t-1} + \sigma^2_\varepsilon}{2\sigma_k} + \sigma_k \ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right] - 1}\right)$$

And we obtain:

$$\Delta\varepsilon^q_{ijkt+1} = g_t\left(a_{ijkt} - \varepsilon^q_{ijkt}\right) + g_t\frac{\sigma^2_\varepsilon}{2\sigma_k} - \sigma_k\left(g_t \ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right] - 1}\right) + \Delta \ln\left(\frac{E_t\left[\varepsilon(s_{ijkt})\right]}{E_t\left[\varepsilon(s_{ijkt})\right] - 1}\right)\right)$$

With variable mark-ups, our main equation includes two new terms.

The first term is the level of the expected mark-ups, $\ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right)$. This term comes from the fact that the expected mark-up also affects our measure of beliefs, $\varepsilon^q_{ijkt}$. As it is a component of $\varepsilon^q_{ijkt}$ and therefore of $\left(a_{ijkt} - \varepsilon^q_{ijkt}\right)$, we need to control for it to avoid a standard omitted variable bias. Hence, we need to control for $\varepsilon^q_{ijkt}$ in the estimation.

The second term captures the change in expected mark-ups $\Delta \ln\left(\frac{E_t\left[\varepsilon(s_{ijkt})\right]}{E_t\left[\varepsilon(s_{ijkt})\right]-1}\right)$, and it depends on the updating process through the change in the expected market share. It follows that our measure of beliefs updating is now underestimated as the quantity reaction to a demand shock is dampened by the mark-up reaction: when firms update positively, they tend to increase their quantities but also their prices, which dampens their overall quantity reaction. Under very weak conditions however, the quantity reaction to beliefs updating is still positive (see the proof below). It means that in the case of variable mark-ups, what we interpret quantitatively as beliefs updating becomes the overall reaction of purged quantities $\varepsilon^q_{ijkt}$ to belief updating. $\varepsilon^q_{ijkt}$ becomes an increasing function of firm's beliefs, but cannot be seen as identical to firm's beliefs. Thus, our results still provide evidence for the updating process, but in a qualitative sense.

Importantly, the relation that goes from beliefs to expected markups (through the expected market share) is not log linear. Put differently, two firms of different sizes, but updating in the exact same way, will not have the same mark-up reaction to this updating. This means that we need again to control for firm size, to be able to compare the beliefs updating of firms with the same initial market share.

**Proof**. Quantity increase after a positive updating: $\frac{\partial\left(\Delta\varepsilon^q_{ijkt}\right)}{\partial\left(\Delta \ln \mathrm{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]\right)} > 0$

First, we express $\varepsilon^q_{ijkt}$ in terms of beliefs and market share. We have:

$$\varepsilon^q_{ijkt} = \sigma_k\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] - \ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right] - 1}\right)\right)$$

Note that:

$$\ln\left(\frac{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]}{\mathbb{E}_{t-1}\left[\varepsilon(s_{ijkt})\right]-1}\right)=-\ln\left(1-\frac{1}{\sigma_k}\right)-\ln\left(1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)$$

As we have purged $\varepsilon^q_{ijkt}$ of its $ikt$ components, we get:

$$\varepsilon^q_{ijkt}=\sigma_k\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]+\ln\left(1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)\right)$$

Second, let's find the relation between beliefs and market share. Market share is given by:

$$\mathbb{E}_{t-1}\left[s_{ijkt}\right]=\frac{\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]\mathbb{E}_{t-1}\left[p^{1-\sigma_k}_{ijkt}\right]}{P^{1-\sigma_k}_{jkt}}$$

And expected price:

$$\mathbb{E}_{t-1}\left[p^{1-\sigma_k}_{ijkt}\right]=\mathbb{E}_{t-1}\left[\left(\frac{w_{it}}{\varphi_{ikt}}\frac{\sigma_k}{\sigma_k-1}\right)^{1-\sigma_k}\left(1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)^{\sigma_k-1}\right]$$

Given that we work with purged prices and quantities, we obtain:

$$\begin{aligned}\mathbb{E}_{t-1}\left[s_{ijkt}\right] &= \mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]\mathbb{E}_{t-1}\left[p^{1-\sigma_k}_{ijkt}\right]\\ \mathbb{E}_{t-1}\left[p^{1-\sigma_k}_{ijkt}\right] &= \mathbb{E}_{t-1}\left[\left(1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)^{\sigma_k-1}\right]\end{aligned}$$

It follows that $\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma}}\right]=\frac{\mathbb{E}_{t-1}\left[s_{ijkt}\right]}{\left(1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)^{\sigma-1}}$ and we obtain:

$$\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]=\ln \mathbb{E}_{t-1}\left[s_{ijkt}\right]-(\sigma_k-1)\ln\left(1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)$$

As the expected market share is an increasing function of the beliefs, we only need to show that $\Delta\varepsilon^q_{ijkt+1}$ is increasing in firm's expected market share.

Third, we can now express $\varepsilon^q_{ijkt}$ as a function of the expected market share only:

$$\varepsilon^q_{ijkt}=\sigma_k\left(\ln \mathbb{E}_{t-1}\left[s_{ijkt}\right]-(\sigma_k-2)\ln\left(1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]\right)\right)$$

We get:

$$\frac{\partial\varepsilon^q_{ijkt}}{\partial\mathbb{E}_{t-1}\left[s_{ijkt}\right]}=\sigma_k\left(\frac{1}{\mathbb{E}_{t-1}\left[s_{ijkt}\right]}+(\sigma_k-2)\frac{1}{1-\mathbb{E}_{t-1}\left[s_{ijkt}\right]}\right)$$

It follows that $\varepsilon^q_{ijkt}$ is an increasing function of the expected market share if:

$$\frac{\partial\varepsilon^q_{ijkt}}{\partial\mathbb{E}_{t-1}\left[s_{ijkt}\right]}>0\Leftrightarrow 1+(\sigma_k-3)\,\mathbb{E}_{t-1}\left[s_{ijkt}\right]>0$$

This condition is necessarily fulfilled if $\sigma_k>2$. If $\sigma_k<2$, we can concentrate on the limiting case $\sigma_k=1$. $\mathbb{E}_{t-1}\left[s_{ijkt}\right]<1/2$ provides another sufficient condition for the above condition to hold.

## F.5 Product-destination specific productivity

Here, we introduce a product-destination component to productivity. Specifically, we assume that the unit cost of producing good $k$ for market $j$ at time $t$ is:

$$\frac{w_{it}}{\varphi_{ikt}}\frac{1}{\varphi_{ijkt}}$$

with $\varphi_{ijkt} > 0$ and where $\frac{1}{\varphi_{ijkt}}$ can be understood as a cost wedge for market $jk$ with respect to the average cost of this good. Further, it could also capture differences in product quality for the same good across markets. Finally, it could capture differences in transportation costs between French competitors in market $jk$ at time $t$.

**Equilibrium**. The optimal price and quantities are given by:

$$\begin{aligned}
q^*_{ijkt} &= \left(\frac{\sigma_k}{\sigma_k - 1}\frac{w_{it}}{\varphi_{ikt}\varphi_{ijkt}}\right)^{-\sigma_k}\left(\frac{\mu_k Y_{jt}}{P_{jkt}^{1-\sigma_k}}\right)\left(\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]\right)^{\sigma_k} \\
p^*_{ijkt} &= \left(\frac{\sigma_k}{\sigma_k - 1}\frac{w_{it}}{\varphi_{ikt}\varphi_{ijkt}}\right)\left(\frac{e^{\frac{a_{ijkt}}{\sigma_k}}}{\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]}\right)
\end{aligned}$$

**Identification**. Purged quantities and prices are:

$$\begin{aligned}
\varepsilon^q_{ijkt} &= \sigma_k\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] + \ln\left(\varphi_{ijkt}\right)\right) \\
\varepsilon^p_{ijkt} &= \frac{1}{\sigma_k}a_{ijkt} - \left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] + \ln\varphi_{ijkt}\right)
\end{aligned}$$

We regress $\varepsilon^p_{ijkt}$ on $\varepsilon^q_{ijkt}$:

$$\frac{1}{\sigma_k}a_{ijkt} - \left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] + \ln\varphi_{ijkt}\right) = \beta\left(\sigma_k\left(\ln \mathrm{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] + \ln\left(\varphi_{ijkt}\right)\right)\right) + \lambda_{ijk} + v_{ijkt}$$

We obtain:

$$\widehat{\beta} = -\frac{1}{\sigma_k} \quad \text{and} \quad \widehat{v}_{ijkt} = \frac{1}{\sigma_k}\varepsilon_{ijkt}$$

And

$$\widehat{\lambda_{ijk}} + \widehat{v}_{ijkt} = \frac{1}{\sigma_k}a_{ijkt}$$

Our identification strategy is still valid if productivity incorporates a $ijk$ component.

**Equation under test**. We now get:

$$\Delta\varepsilon^q_{ijkt+1} = \sigma_k\left[\Delta\ln \mathrm{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right] + \Delta\ln\left(\varphi_{ijkt+1}\right)\right]$$

And $\Delta\varepsilon^q_{ijkt+1}$ cannot be seen as updating only. The updating process itself does not change, we still get $\Delta\widetilde{\theta}_t = g_t\left(a_{ijkt} - \widetilde{\theta}_{t-1}\right)$, and thus $\Delta\ln\mathrm{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right] = \frac{1}{\sigma_k}\left(\Delta\widetilde{\theta}_t + \frac{\widetilde{\sigma}^2_t - \widetilde{\sigma}^2_{t-1}}{2\sigma_k}\right)$.

It follows

$$\Delta\varepsilon^q_{ijkt+1} = g_t\left(a_{ijkt} - \widetilde{\theta}_{t-1} - \frac{\widetilde{\sigma}^2_{t-1} - \widetilde{\sigma}^2_t}{g_t 2\sigma_k}\right) + \sigma_k\Delta\ln\left(\varphi_{ijkt+1}\right)$$

Further,

$$\widetilde{\theta}_{ijkt-1} = \varepsilon^q_{ijkt} - \frac{\widetilde{\sigma}^2_{t-1} + \sigma^2_\varepsilon}{2\sigma_k} - \sigma_k\ln\left(\varphi_{ijkt}\right)$$

Hence:

$$\Delta\varepsilon^q_{ijkt+1} = g_t\left(a_{ijkt} - \varepsilon^q_{ijkt}\right) + g_t\frac{\sigma^2_\varepsilon}{2\sigma_k} + \sigma_k\left[g_t\ln\left(\varphi_{ijkt}\right) + \Delta\ln\left(\varphi_{ijkt+1}\right)\right]$$

As for the case of variable mark-ups, our equation includes two new terms.

This term comes from the fact that $\ln\left(\varphi_{ijkt}\right)$ affects our measure of beliefs, $\varepsilon^q_{ijkt}$. As it is a component of $\varepsilon^q_{ijkt}$ and thus of $\left(a_{ijkt} - \varepsilon^q_{ijkt}\right)$, we need to control for firm size to avoid a standard omitted variable bias.

The presence of the second term, $\Delta\ln\left(\varphi_{ijkt+1}\right)$ comes from the fact that $\Delta\varepsilon^q_{ijkt+1}$ also reflects the dynamics of productivity.

If $\Delta\ln\left(\varphi_{ijkt+1}\right)$ is uncorrelated with the updating process, the interpretation of our results should be unaffected. If however $\Delta\ln\left(\varphi_{ijkt+1}\right)$ is positively affected by the updating process, because a positive updating would lead firms to invest to improve $\varphi_{ijkt}$, then our measure of updating becomes a measure of the overall impact of the updating process on $\Delta\varepsilon^q_{ijkt+1}$: it would not only capture the updating process itself but also how the quantity response is magnified by a change in productivity. $\varepsilon^q_{ijkt}$ would become an increasing function of firm's beliefs, and our evidence of the updating process would become qualitative as we would not identify firms' beliefs, but only a function of it. As this productivity response could be size dependent, we need again to control for firm size. The decline of the overall reaction of $\Delta\varepsilon^q_{ijkt+1}$ to demand shocks over time, conditional on size, however still provides evidence for an updating process.

Finally, note that $\Delta\varepsilon^q_{ijkt+1}$ will also capture the dynamics of $\ln\left(\varphi_{ijkt}\right)$ that is uncorrelated with the updating process. In turn, it could introduce some noise into our measure of the updating process. But if this dynamics was important, it should be observed in the dynamics of $\varepsilon^q_{ijkt}$ and $\varepsilon^p_{ijkt}$. Results in section 6.1 show however that, when we concentrate on the within firm-market dynamics of these elements (Figure 2.c), no important pattern emerge: $\varepsilon^q_{ijkt}$ and $\varepsilon^p_{ijkt}$ are roughly constant over time, which suggests that there should not be any important dynamics in $\ln\left(\varphi_{ijkt}\right)$, beyond the one possibly driven by the updating process.

## F.6 Controlling for size – robustness

This section presents additional results controlling for size when testing prediction 1. Our main results appear in Table A.8. Columns (1) and (2) are similar to our baseline regressions (Table 2, columns (2) and (4)), except that $jkt$ fixed effects are introduced in the estimation of the price residuals $\varepsilon^{p}_{ijkt}$, as predicted by models with variable markups. The average level of belief updating is slightly larger than in our baseline estimates, but the effect of age is similar. In columns (3) to (6) we additionally control for firm size, as measured by the value sold by firm $i$ on market $jk$ during year $t-1$ divided by the total value exported by French firms in market $jk$ during year $t-1$. Size is introduced either linearly in columns (3) and (4) or through bins computed using market-specific deciles in columns (5) and (6). Our coefficients of interest are extremely stable across specifications.[7]

Table A.9 shows that our results are not sensitive to the measurement of firm size. In columns (1)-(4), we measure firm size as market shares in quantity and introduce it either linearly or through bins as in Table A.8. Alternatively, in columns (5)-(8) firm size is measured as the log of the value exported by firm $i$ to market $jk$ in year $t-1$.

Finally, Table A.10 includes an interaction term between size and $a_{ijkt} - \varepsilon^{q}_{ijk}$ to account for the fact that age and size are correlated. We report results using as a measure of firm size either the market share in value (columns (1)-(4)) and quantity (columns (5)-(8)) introduced linearly (odd columns) or through bins (even columns).

In all cases, the coefficients on $a_{ijkt} - \varepsilon^{q}_{ijk}$ and its interaction with age remain close to our benchmark results in Table A.8.

[7] Note that the positive coefficient on age in column (5) cannot be directly interpreted as this estimation also includes a full set of interaction terms between age and size.

Table A.8: Prediction 1: controlling for size

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dep. var. | | | $\Delta\varepsilon^q_{ijk,t+1}$ | | | |
| Robustness | Controlling for $\text{FE}_{jkt}$ in prices | | Controlling for $\text{FE}_{jkt}$ in prices and size | | | |
| Size | | | Linear | | Bins | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | 0.103$^a$ | | 0.103$^a$ | | 0.102$^a$ | |
| | (0.002) | | (0.002) | | (0.002) | |
| $\times$ $\text{Age}_{ijkt}$ | -0.003$^a$ | | -0.003$^a$ | | -0.003$^a$ | |
| | (0.000) | | (0.000) | | (0.000) | |
| $\times$ $\text{Age}_{ijkt} = 2$ | | 0.096$^a$ | | 0.096$^a$ | | 0.096$^a$ |
| | | (0.002) | | (0.002) | | (0.002) |
| $\times$ $\text{Age}_{ijkt} = 3$ | | 0.093$^a$ | | 0.093$^a$ | | 0.093$^a$ |
| | | (0.002) | | (0.002) | | (0.002) |
| $\times$ $\text{Age}_{ijkt} = 4$ | | 0.087$^a$ | | 0.087$^a$ | | 0.087$^a$ |
| | | (0.002) | | (0.002) | | (0.002) |
| $\times$ $\text{Age}_{ijkt} = 5$ | | 0.086$^a$ | | 0.086$^a$ | | 0.087$^a$ |
| | | (0.003) | | (0.003) | | (0.002) |
| $\times$ $\text{Age}_{ijkt} = 6$ | | 0.082$^a$ | | 0.082$^a$ | | 0.081$^a$ |
| | | (0.003) | | (0.002) | | (0.003) |
| $\times$ $\text{Age}_{ijkt} = 7$ | | 0.079$^a$ | | 0.079$^a$ | | 0.078$^a$ |
| | | (0.003) | | (0.003) | | (0.003) |
| $\times$ $\text{Age}_{ijkt} = 8$ | | 0.076$^a$ | | 0.076$^a$ | | 0.076$^a$ |
| | | (0.004) | | (0.004) | | (0.004) |
| $\times$ $\text{Age}_{ijkt} = 9$ | | 0.077$^a$ | | 0.076$^a$ | | 0.077$^a$ |
| | | (0.005) | | (0.005) | | (0.005) |
| $\times$ $\text{Age}_{ijkt} = 10$ | | 0.074$^a$ | | 0.074$^a$ | | 0.075$^a$ |
| | | (0.009) | | (0.009) | | (0.009) |
| $\text{Age}_{ijkt}$ | -0.034$^a$ | | -0.040$^a$ | | 0.019$^a$ | |
| | (0.001) | | (0.001) | | (0.002) | |
| $\text{Size}_{ijkt}$ | | | -1.053$^a$ | -1.015$^a$ | | |
| | | | (0.016) | (0.017) | | |
| $\times$ $\text{Age}_{ijkt}$ | | | 0.109$^a$ | 0.101$^a$ | | |
| | | | (0.003) | (0.003) | | |
| Observations | 1870377 | 1870377 | 1870377 | 1870377 | 1501840 | 1501840 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%. $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17). Compared to our baseline methodology, in this table we include $jkt$ fixed effects in the estimation of the price residuals $\varepsilon^p_{ijkt}$ used to identify demand shocks. $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). $\text{Age}_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). $\text{Size}_{ijkt}$ is proxied by the value sold by firm $i$ on market $jk$ during year $t$ divided by the total value exported by French firms in market $jk$ during year $t$. Columns (5) and (6) include size bins corresponding to the ten deciles of size variable, computed by market-year. Age dummies included alone in columns (2), (4) and (6) but coefficients not reported.

Table A.9: Prediction 1: controlling for size (robustness 1/2)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Dep. var. | $\Delta\varepsilon^q_{ijk,t+1}$ | | | | | | | |
| Size | Market share (quantity) | | | | log values | | | |
| Functional form | Linear | | Bins | | Linear | | Bins | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | 0.098$^a$ | | 0.082$^a$ | | 0.104$^a$ | | 0.103$^a$ | |
| | (0.002) | | (0.002) | | (0.002) | | (0.002) | |
| × Age$_{ijkt}$ | -0.003$^a$ | | -0.002$^a$ | | -0.004$^a$ | | -0.003$^a$ | |
| | (0.000) | | (0.000) | | (0.000) | | (0.000) | |
| × Age$_{ijkt} = 2$ | | 0.092$^a$ | | 0.079$^a$ | | 0.098$^a$ | | 0.097$^a$ |
| | | (0.002) | | (0.002) | | (0.002) | | (0.002) |
| × Age$_{ijkt} = 3$ | | 0.089$^a$ | | 0.077$^a$ | | 0.094$^a$ | | 0.093$^a$ |
| | | (0.002) | | (0.002) | | (0.002) | | (0.002) |
| × Age$_{ijkt} = 4$ | | 0.083$^a$ | | 0.071$^a$ | | 0.088$^a$ | | 0.087$^a$ |
| | | (0.002) | | (0.002) | | (0.002) | | (0.002) |
| × Age$_{ijkt} = 5$ | | 0.083$^a$ | | 0.072$^a$ | | 0.087$^a$ | | 0.087$^a$ |
| | | (0.002) | | (0.002) | | (0.002) | | (0.002) |
| × Age$_{ijkt} = 6$ | | 0.079$^a$ | | 0.068$^a$ | | 0.083$^a$ | | 0.082$^a$ |
| | | (0.002) | | (0.002) | | (0.002) | | (0.003) |
| × Age$_{ijkt} = 7$ | | 0.077$^a$ | | 0.066$^a$ | | 0.080$^a$ | | 0.078$^a$ |
| | | (0.003) | | (0.003) | | (0.003) | | (0.003) |
| × Age$_{ijkt} = 8$ | | 0.075$^a$ | | 0.065$^a$ | | 0.077$^a$ | | 0.076$^a$ |
| | | (0.004) | | (0.004) | | (0.004) | | (0.004) |
| × Age$_{ijkt} = 9$ | | 0.076$^a$ | | 0.068$^a$ | | 0.077$^a$ | | 0.078$^a$ |
| | | (0.005) | | (0.005) | | (0.005) | | (0.005) |
| × Age$_{ijkt} = 10$ | | 0.073$^a$ | | 0.065$^a$ | | 0.074$^a$ | | 0.074$^a$ |
| | | (0.009) | | (0.009) | | (0.009) | | (0.009) |
| Age$_{ijkt}$ | -0.039$^a$ | | 0.016$^a$ | | -0.148$^a$ | | 0.024$^a$ | |
| | (0.001) | | (0.002) | | (0.003) | | (0.002) | |
| Size$_{ijkt}$ | -0.891$^a$ | -0.857$^a$ | | | -0.184$^a$ | -0.180$^a$ | | |
| | (0.015) | (0.015) | | | (0.002) | (0.003) | | |
| × Age$_{ijkt}$ | 0.090$^a$ | 0.083$^a$ | | | 0.015$^a$ | 0.014$^a$ | | |
| | (0.002) | (0.002) | | | (0.000) | (0.000) | | |
| Observations | 1870377 | 1870377 | 1501840 | 1501840 | 1870377 | 1870377 | 1501840 | 1501840 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%. $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17). Compared to our baseline methodology, in this table we include $jkt$ fixed effects in the estimation of the price residuals $\varepsilon^p_{ijkt}$ used to identify demand shocks. $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). In columns (1)-(4), Size$_{ijkt}$ is proxied by the quantity sold by firm $i$ on market $jk$ during year $t$ divided by the total quantity exported by French firms in market $jk$ during year $t$. In columns (5)-(8), Size$_{ijkt}$ is proxied by the log of the value sold by firm $i$ on market $jk$ during year $t$. Columns (3), (4), (7) and (8) include size bins corresponding to the ten deciles of size variable, computed by market-year. Age dummies included alone in columns (2), (4), (6) and (8) but coefficients not reported.

Table A.10: Prediction 1: controlling for size (robustness 2/2)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Dep. var. | | | | $\Delta\varepsilon^q_{ijk,t+1}$ | | | | |
| Size | Market share (value) | | | | Market share (quantity) | | | |
| Functional form | Linear | | Bins | | Linear | | Bins | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | 0.102$^a$ | | 0.095$^a$ | | 0.097$^a$ | | 0.111$^a$ | |
| | (0.002) | | (0.003) | | (0.002) | | (0.003) | |
| $\times$ Age$_{ijkt}$ | -0.004$^a$ | | -0.003$^a$ | | -0.003$^a$ | | -0.002$^a$ | |
| | (0.000) | | (0.000) | | (0.000) | | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | 0.096$^a$ | | 0.100$^a$ | | 0.092$^a$ | | 0.097$^a$ |
| | | (0.002) | | (0.003) | | (0.002) | | (0.003) |
| $\times$ Age$_{ijkt} = 3$ | | 0.092$^a$ | | 0.098$^a$ | | 0.088$^a$ | | 0.095$^a$ |
| | | (0.002) | | (0.003) | | (0.002) | | (0.003) |
| $\times$ Age$_{ijkt} = 4$ | | 0.086$^a$ | | 0.092$^a$ | | 0.082$^a$ | | 0.090$^a$ |
| | | (0.002) | | (0.003) | | (0.002) | | (0.003) |
| $\times$ Age$_{ijkt} = 5$ | | 0.085$^a$ | | 0.092$^a$ | | 0.082$^a$ | | 0.090$^a$ |
| | | (0.003) | | (0.003) | | (0.002) | | (0.003) |
| $\times$ Age$_{ijkt} = 6$ | | 0.081$^a$ | | 0.087$^a$ | | 0.079$^a$ | | 0.085$^a$ |
| | | (0.003) | | (0.003) | | (0.003) | | (0.003) |
| $\times$ Age$_{ijkt} = 7$ | | 0.078$^a$ | | 0.084$^a$ | | 0.076$^a$ | | 0.082$^a$ |
| | | (0.003) | | (0.004) | | (0.003) | | (0.004) |
| $\times$ Age$_{ijkt} = 8$ | | 0.075$^a$ | | 0.082$^a$ | | 0.074$^a$ | | 0.080$^a$ |
| | | (0.004) | | (0.005) | | (0.004) | | (0.005) |
| $\times$ Age$_{ijkt} = 9$ | | 0.075$^a$ | | 0.084$^a$ | | 0.075$^a$ | | 0.084$^a$ |
| | | (0.005) | | (0.006) | | (0.005) | | (0.006) |
| $\times$ Age$_{ijkt} = 10$ | | 0.072$^a$ | | 0.081$^a$ | | 0.072$^a$ | | 0.079$^a$ |
| | | (0.009) | | (0.010) | | (0.009) | | (0.010) |
| $\times$ Size$_{ijk,t-1}$ | 0.011$^c$ | 0.011$^c$ | | | 0.008 | 0.009$^c$ | | |
| | (0.006) | (0.006) | | | (0.005) | (0.005) | | |
| Age$_{ijkt}$ | -0.040$^a$ | | 0.019$^a$ | | -0.039$^a$ | | 0.015$^a$ | |
| | (0.001) | | (0.002) | | (0.001) | | (0.002) | |
| Size$_{ijkt}$ | -1.053$^a$ | -1.014$^a$ | | | -0.884$^a$ | -0.849$^a$ | | |
| | (0.016) | (0.017) | | | (0.015) | (0.015) | | |
| $\times$ Age$_{ijkt}$ | 0.109$^a$ | 0.101$^a$ | | | 0.090$^a$ | 0.083$^a$ | | |
| | (0.003) | (0.003) | | | (0.002) | (0.002) | | |
| Observations | 1870377 | 1870377 | 1501840 | 1501840 | 1870377 | 1870377 | 1501840 | 1501840 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%. $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17). Compared to our baseline methodology, in this table we include $jkt$ fixed effects in the estimation of the price residuals $\varepsilon^p_{ijkt}$ used to identify demand shocks. $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). In columns (1)-(4), Size$_{ijkt}$ is proxied by the value sold by firm $i$ on market $jk$ during year $t$ divided by the total value exported by French firms in market $jk$ during year $t$. In columns (1)-(4), Size$_{ijkt}$ is proxied by the quantity sold by firm $i$ on market $jk$ during year $t$ divided by the total quantity exported by French firms in market $jk$ during year $t$. Age dummies included alone in columns (2), (4), (6) and (8) but coefficients not reported. Compared to our baseline estimates, these regressions include additional interaction terms between $a_{ijkt} - \varepsilon^q_{ijk,t}$ and age.

# G Firm survival

This section develops the predictions of the learning model regarding firms' survival and provides evidence that the exit behavior of firms on specific markets is in line with the demand learning process.

A firm decides to stop exporting a particular product to a given destination whenever the expected value of the profits stream associated with this activity becomes negative. At the beginning of period $t$ (after having received $t-1$ signals), expected profits for period $t$ are given by:

$$E_{t-1}\left[\pi_{ijkt}\right] = \frac{C^S_{ikt}C^S_{jkt}}{\sigma_k} e^{\left(\widetilde{\theta}_{ijkt-1} + \frac{\widetilde{\sigma}^2_{ijkt-1}+\sigma^2_\varepsilon}{2\sigma_k}\right)} - F_{ijk}.$$

Of course, the exit decision also depends on the expected future stream of profits, which depends on the evolution of $C^S_{ikt}$, $C^S_{jkt}$, $\widetilde{\theta}_{ijkt-1}$ and $\widetilde{\sigma}^2_{ijkt-1}$ over time. Our assumption of normal prior beliefs provides the conditional distribution of $\widetilde{\theta}_{ijkt}$ given $\widetilde{\theta}_{ijkt-1}$ while the distribution of $\widetilde{\sigma}^2_{ijkt-1}$ is deterministic. So, the evolution of firms' beliefs can be summarized by $\widetilde{\theta}_{ijkt-1}$ and $t$. Up to now, we have made no assumption regarding the dynamics of the $C^S_{ikt}$ and $C^S_{jkt}$ terms. Here, to proceed further, we follow Hopenhayn (1992) and introduce some (mild) assumptions on their dynamics. We label $A_{ijkt} \equiv C^S_{ikt}C^S_{jkt}$ and we assume that: i) $A_{ijkt}$ follows a Markov process, ii) $A_{ijkt}$ is bounded and iii) the conditional distribution $F\left(A_{ijkt+1} \mid A_{ijkt}\right)$ is continuous in $A_{ijkt}$ and $A_{ijkt+1}$, and $F(.)$ is strictly decreasing in $A_{ijkt}$.[8]

The set of firm state variables at time $t$ can thus be summarized by $\Omega_{ijkt} = \left\{A_{ijkt}, \widetilde{\theta}_{ijkt-1}, t\right\}$. The value function of the firm $V_{ijk}\left(\Omega_{ijkt}\right)$ satisfies the following Bellman equation:

$$V_{ijk}\left(\Omega_{ijkt}\right) = \max\left\{\mathbb{E}\left[\pi_{ijkt}\left(\Omega_{ijkt}\right)\right] + \beta\mathbb{E}\left[V_{ijk}\left(\Omega_{ijkt+1} \mid \Omega_{ijkt}\right)\right], 0\right\} \tag{26}$$

where $\beta$ is the rate at which firms discount profits and where we have normalized the value of exiting to zero.[9] The value function $V_{ijk}$ is monotonically increasing in $A_{ijkt}$ and $\widetilde{\theta}_{ijkt-1}$.[10] Intuitively, the flow of future expected profits inherits the properties of expected profits at time $t$. It follows that there exists a threshold value $\underline{\widetilde{\theta}_{ijkt-1}}(A_{ijkt}, t)$ such that a firm exits market $jk$ at time $t$ if $\widetilde{\theta}_{ijkt-1} < \underline{\widetilde{\theta}_{ijkt-1}}(A_{ijkt}, t)$. This implies:

**Prediction # 4** (firm exit): *Given $A_{ijkt}$ and $t$ (firm age), (a) the probability to exit decreases with $\widetilde{\theta}_{ijkt-1}$ and (b) a given negative difference between realized and expected demand triggers less exit for older firms.*

The literature has usually associated learning with exit rates declining with age, and we indeed find this to be the case in our estimations. However, this relation may not necessarily be monotonic (see Pakes and Ericson, 1998 for a discussion). The decision to exit not only depends

[8] While not very demanding, these assumptions restrict the set of possible dynamics for firm productivity. In that sense, our results on firm exit decision are somewhat weaker than those about firm growth, which are robust to any dynamics of firm productivity.

[9] Here, we assume that an exiting firm loses all the information accumulated in the past. If the firm enters again market $jk$ in the future, new initial beliefs will be drawn. In consequence, we treat the exit decision as irreversible.

[10] See Hopenhayn (1992) and Jovanovic (1982).

on the extent of firm updating (which indeed declines with age) but also on how $\underline{\widetilde{\theta}_{ijkt-1}}(A_{ijkt}, t)$ evolves over time. If this threshold increases very rapidly for some $t$, the exit rate could actually increase temporarily. For old firms however, i.e. when beliefs become accurate, and conditional on $A_{ijkt}$ *and* $t$, the exit rate should tend to 0.

On the other hand, an important and general implication of our demand learning model is that negative demand shocks should trigger less exits for older firms (prediction 4.b). The reason is simply that firms' posterior beliefs $\widetilde{\theta}_{ijkt-1}$ depend less and less on demand shocks as firms age. Hence, the exit rate may not always be decreasing with age, but demand shocks should always have a lower impact on the exit decision in older cohorts, because they imply less updating. Note that this prediction can also be understood as another robustness check for our formulation of a passive learning model: in an active learning model, no matter the age of the firm, demand shocks may trigger new investments. Their impact on future expected profits stream should not be weakened for older firms (see Ericson and Pakes, 1995). This prediction is not directly tested in Pakes and Ericson (1998) because they use a much less parametric model than ours which prevents them to back out demand shocks and firms' beliefs. Their test is solely based on actual firm size.

To test prediction 4, note that from equation (5), $\widetilde{\theta}_{ijkt-1}$ depends positively on $\widetilde{\theta}_{ijkt-2}$ and $a_{ijkt-1}$. We want to test if, conditional on $A_{ijkt}$ and firm age, the probability to exit at the end of period $t-1$ (i.e. beginning of period $t$) decreases with $\widetilde{\theta}_{ijkt-2}$ and $a_{ijkt-1}$.

We estimate the following probabilistic model:

$$S_{ijkt} = \alpha \text{AGE}_{ijkt-1} + \beta(a_{ijk,t-1} - \varepsilon^q_{ijk,t-1}) + \gamma \varepsilon^q_{ijkt-1} + \delta(a_{ijk,t-1} - \varepsilon^q_{ijk,t-1}) \times \text{AGE}_{ijkt-1} + \mathbf{FE} + u_{ijkt} > 0$$

Where $S_{ijkt} = 0$ is a dummy that takes the value 1 if firm $i$ exits market $jk$ in year $t$. We expect $\beta$ and $\gamma$ to be negative, and $\delta$ to be positive. **FE** include the two sets of fixed effects $\mathbf{FE}_{ikt}$ and $\mathbf{FE}_{jkt}$, which capture $C^S_{ikt}$ and $C^S_{jkt}$. We estimate this equation using a linear probability model which does not suffer from incidental parameters problems, an issue that might be important here given the two large dimensions of fixed effects we need to include.

The results for prediction 4.a are shown in Table A.11, columns (1) to (3), and are largely in line with the model: conditional on age, the exit probability decreases with the value of demand shocks $\widehat{v}$ and firm's belief (columns (1) to (3)).

Columns (4) and (5) of Table A.11 test for prediction 4.b. We simply add to our baseline specification of column (3) an interaction term between age and demand shock in $t-1$.[11] We indeed find that the coefficient on this interaction term is positive: Young firms react more to a given demand shock than mature exporters on the market. In column (5), a negative demand shock of 10% increases exit probability by 3.3 percentage points for a young firm (2 years after entry), but by only 1.3 percentage points after 7 years.

[11] Given our need to control for all $jkt$-determinants here, we use the version of $\widehat{v}_{ijk,t-1}$ computed using $jkt$-specific fixed effects, as in Table 4. This has no importance in columns (1) to (3) as the vector of fixed effects includes $\mathbf{FE}_{jkt}$, but it does in columns (4) and (5) as the coefficient on the interaction between $\widehat{v}_{ijkt-1}$ and age might reflect differences in $\widehat{v}_{ijkt-1}$ along the $jkt$ dimension (as we focus on an interaction term in this case).

Table A.11: Firm exit

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. var. | | $\Pr(S_{ijkt} = 0 \mid S_{ijkt-1} = 1)$ | | |
| $\text{Age}_{ijk,t-1}$ | -0.024$^{a}$ | -0.028$^{a}$ | -0.022$^{a}$ | -0.022$^{a}$ |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| $\varepsilon^{q}_{ijkt-1}$ | -0.043$^{a}$ | | -0.080$^{a}$ | -0.097$^{a}$ |
| | (0.000) | | (0.000) | (0.001) |
| $\times\ \text{Age}_{ijk,t-1}$ | | | | 0.004$^{a}$ |
| | | | | (0.000) |
| $(a_{ijk,t-1} - \varepsilon^{q}_{ijkt-1})$ | | 0.033$^{a}$ | -0.039$^{a}$ | -0.044$^{a}$ |
| | | (0.000) | (0.000) | (0.001) |
| $\times\ \text{Age}_{ijk,t-1}$ | | | | 0.001$^{a}$ |
| | | | | (0.000) |
| Observations | 4885284 | 4885284 | 4885284 | 4885284 |

Robust standard errors clustered by firm-product-destination in parentheses. Estimator: LPM. All estimations include $jkt$ and $ikt$ fixed effects. $^{c}$ significant at 10%; $^{b}$ significant at 5%; $^{a}$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^{q}_{ijkt-1}$ is the belief of the firm about future demand from equation (14). $\text{Age}_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit).

# H Firm growth

Our first stylized fact shows that the growth rates of quantities decline with age, conditional on size. This decline comes from two different mechanisms in the passive learning model: (i) selection (ii) larger growth rates for younger firms, *unconditional* on survival.

The impact of selection on growth rates is due to the fact that younger firms have greater variance in their growth rates, which comes from their larger updating. Firms that decrease in size are more likely to exit. Hence, the distribution of growth rates is truncated from below. As younger firms may experience more negative growth rates due to the larger variance, this truncation leads to larger growth rates for younger firms, conditional on survival. Note that this mechanism holds only if exit rates are not increasing with age, which is clearly the case in our data (see Figure 1.a in the main text and the results of the previous section on firm survival).

Second, the passive learning model is also consistent with larger growth rates for younger firms, even if we do not condition on firm survival. This unconditional growth is quite limited in our data (see figure 2.c), but is not in contradiction with the model. It should be noted however that this result is driven by the assumption that $exp(\frac{a_{ijkt}}{\sigma_k})$ is log-normally distributed and is thus sensitive to the functional form assumption. In the rest of this section we detail the proof of this result.

**Expected growth rate, conditional on size, *unconditional* on survival.** The expected (quantity) growth rate of firm $i$ at time $t$, conditional on its size, and non conditional on survival, is given by:

$$\frac{\mathbb{E}_{t-1}\left[q^*_{ijkt+1}\right]}{q^*_{ijkt}}$$

where $\mathbb{E}_{t-1}\left[q^*_{ijkt+1}\right]$ is the expected quantity at time $t+1$, conditional on the information available at time $t-1$, i.e. conditional on the information received from $t-1$ signals: $\overline{a}_{ijkt-1}$. In words, this is the expected value of $q^*_{ijkt+1}$, given that the shock in period $t$, $a_{ijkt}$, is not observed yet, and will lead to an updating of firm beliefs between $t$ and $t+1$.

Given the optimal quantity choice (see equation (7)), we get:

$$\begin{aligned}
\frac{\mathbb{E}_{t-1}\left[q^*_{ijkt+1}\right]}{q^*_{ijkt}} &= \frac{\mathbb{E}_{t-1}\left[\left(\frac{\sigma_k}{\sigma_k-1}\frac{w_{it+1}}{\varphi_{ikt+1}}\right)^{-\sigma_k}\left(\frac{\mu_k Y_{jt+1}}{P^{1-\sigma_k}_{jkt+1}}\right)\mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]^{\sigma_k}\right]}{\left(\frac{\sigma_k}{\sigma_k-1}\frac{w_{it}}{\varphi_{ikt}}\right)^{-\sigma_k}\left(\frac{\mu_k Y_{jt}}{P^{1-\sigma_k}_{jkt}}\right)\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]^{\sigma_k}} \\
&= \frac{\left(\frac{\sigma_k}{\sigma_k-1}\frac{w_{it+1}}{\varphi_{ikt+1}}\right)^{-\sigma_k}\left(\frac{\mu_k Y_{jt+1}}{P^{1-\sigma_k}_{jkt+1}}\right)\mathbb{E}_{t-1}\left[\mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]^{\sigma_k}\right]}{\left(\frac{\sigma_k}{\sigma_k-1}\frac{w_{it}}{\varphi_{ikt}}\right)^{-\sigma_k}\left(\frac{\mu_k Y_{jt}}{P^{1-\sigma_k}_{jkt}}\right)\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]^{\sigma_k}}
\end{aligned}$$

As we work with purged quantities, we label $\mathbb{E}_{t-1}\left[g_q\right]$ the expected growth rate of purged quantities:

$$\mathbb{E}_{t-1}\left[g_q\right] = \frac{\mathbb{E}_{t-1}\left[\mathbb{E}_t\left[e^{\frac{a_{ijkt+1}}{\sigma_k}}\right]^{\sigma_k}\right]}{\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right]^{\sigma_k}}$$

As $\mathbb{E}_{t-1}\left[e^{\frac{a_{ijkt}}{\sigma_k}}\right] = e^{\frac{1}{\sigma_k}\left(\widetilde{\theta}_{ijkt-1}+\frac{\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2}{2\sigma_k}\right)}$ (see appendix main text), we get:

$$\mathbb{E}_{t-1}\left[g_q\right] = \frac{\mathbb{E}_{t-1}\left[e^{\left(\widetilde{\theta}_{ijkt}+\frac{\widetilde{\sigma}_t^2+\sigma_\varepsilon^2}{2\sigma_k}\right)}\right]}{e^{\left(\widetilde{\theta}_{ijkt-1}+\frac{\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2}{2\sigma_k}\right)}}$$

Note that, from $t-1$ perspective, $\widetilde{\theta}_{ijkt}$ is a random variable as $a_{ijkt}$ is not observed. We may rewrite $\mathbb{E}_{t-1}\left[g_q\right]$ as:

$$\mathbb{E}_{t-1}\left[g_q\right] = \frac{\mathbb{E}_{t-1}\left[e^{\widetilde{\theta}_{ijkt}}\right] e^{\left(\frac{\widetilde{\sigma}_t^2+\sigma_\varepsilon^2}{2\sigma_k}\right)}}{e^{\left(\widetilde{\theta}_{ijkt-1}+\frac{\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2}{2\sigma_k}\right)}}$$

We next rewrite $\widetilde{\theta}_{ijkt}$ to explicit $a_{ijkt}$:

$$\begin{aligned}
\widetilde{\theta}_{ijkt} &= \theta_0 \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}} + \frac{1}{t}\left(\left(t-1\right)\overline{a}_{ijkt-1}+a_{ijkt}\right)\frac{\frac{t}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}} \\
&= \theta_0 \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}} + \frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}}\left(t-1\right)\overline{a}_{ijkt-1} + \frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}} a_{ijkt}
\end{aligned}$$

$\widetilde{\theta}_{ijkt}$ being linear in $a_{ijkt}$, it is also normally distributed. From $t-1$ perspective we get $\mathbb{E}\left[\widetilde{\theta}_{ijkt} \mid \widetilde{\theta}_{ijkt-1}\right] = \widetilde{\theta}_{ijkt-1}$.

Second, remind that $\mathbb{V}(a_{ijkt}) = \widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2$, so $\mathbb{V}\left(\widetilde{\theta}_{ijkt}\right) = \left(\frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}}\right)^2\left(\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2\right)$. Since $\widetilde{\theta}_{ijkt}$ is normally distributed, $e^{\widetilde{\theta}_{ijkt}}$ is lognormally distributed. We thus obtain:

$$\begin{aligned}
\mathbb{E}_{t-1}\left[g_q\right] &= \frac{e^{\left(\widetilde{\theta}_{ijkt-1}+\frac{1}{2}\left(\frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}}\right)^2\left(\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2\right)\right)} e^{\left(\frac{\widetilde{\sigma}_t^2+\sigma_\varepsilon^2}{2\sigma_k}\right)}}{e^{\left(\widetilde{\theta}_{ijkt-1}+\frac{\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2}{2\sigma_k}\right)}} \\
&= e^{\frac{1}{2}\left(\frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}}\right)^2\left(\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2\right)+\frac{\widetilde{\sigma}_t^2-\widetilde{\sigma}_{t-1}^2}{2\sigma_k}}
\end{aligned}$$

As we work with log (purged) quantities, let's take the log:

$$\ln \mathbb{E}_{t-1}\left[g_q\right] = \frac{1}{2}\left(\frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}}\right)^2\left(\widetilde{\sigma}_{t-1}^2+\sigma_\varepsilon^2\right) - \frac{\widetilde{\sigma}_{t-1}^2-\widetilde{\sigma}_t^2}{2\sigma_k}$$

Given the definitions of $\widetilde{\sigma}_t^2$ and $\widetilde{\sigma}_{t-1}^2$ (see equation (4)), we get:

$$\ln \mathbb{E}_{t-1}\left[g_q\right] = \frac{1}{2}\left(\frac{\frac{1}{\sigma_\epsilon^2}}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}}\right)^2\left(\frac{1}{\frac{1}{\sigma_0^2}+\frac{t-1}{\sigma_\epsilon^2}}+\sigma_\varepsilon^2\right)-\frac{1}{2\sigma_k}\left(\frac{1}{\frac{1}{\sigma_0^2}+\frac{t-1}{\sigma_\epsilon^2}}-\frac{1}{\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}}\right)$$

$$\ln \mathbb{E}_{t-1}\left[g_q\right] = \left(\frac{1}{2}-\frac{1}{2\sigma_k}\right)\frac{\frac{1}{\sigma_\epsilon^2}}{\left(\frac{1}{\sigma_0^2}+\frac{t}{\sigma_\epsilon^2}\right)\left(\frac{1}{\sigma_0^2}+\frac{t-1}{\sigma_\epsilon^2}\right)}$$

Note that $\frac{1}{2}-\frac{1}{2\sigma_k}>0$, so $\ln \mathbb{E}_{t-1}\left[g_q\right]$ is always positive. Moreover, $t$ appears in the denominator only, so this expression is strictly decreasing with $t$: Expected growth rates decline with firm age in market $jk$.

The source of this result comes from the functional form assumption: the profit function depends on the *exponential* of the demand shock $a_{ijkt}$. Given that $a_{ijkt}$ is normally distributed, $exp(\frac{a_{ijkt}}{\sigma_k})$ is log-normally distributed, its expectation thus depends on its variance. Without this effect, expected growth rate (non conditional on survival) of purged quantities should always be 0, no matter firm age. Second, note that expectation is taken over $exp(\frac{a_{ijkt}}{\sigma_k})$ and its variance is reduced by $\sigma_k$. But expectation is also taken over $exp(\widetilde{\theta}_{ijkt})$, which does not depend on $\sigma_k$. This is generating the result.

# I Belief updating and age: endogenous selection

This section presents the detailed results discussed in section 5.2 of the main text on survival and selection bias. All these specifications draw on the predictions of our model regarding firms' exit decision detailed in section G and Table A.11. In particular, exit probabilities depend on $a_{ijkt}$, $\varepsilon^q_{ijkt}$, $\text{Age}_{ijkt}$ and fixed effects in the $ikt$ and $jkt$ dimensions and can be estimated using a linear probability model.

We start by documenting whether the firms' updating process identified in Table 2 varies depending on their survival probability. This is application of the "identification-at-infinity" method (Chamberlain, 1986; Mulligan and Rubinstein, 2008). We expect the potential selection bias related to endogenous exit decisions to be lower on sub-samples of firms, selected on observable characteristics, most likely to survive. We first estimate equation (G) and compute the predicted probability of exit by firm×market×year. Equation (19) is then estimated on four sub-samples including respectively firms above the 20th, 40th, 60th and 80th percentiles of survival probability (i.e. below the 80th, 60th, 40th and 20th of exit probability). Table A.12 presents the results when firms are allocated in quintiles depending on their raw probability of exit. Alternatively, in Table A.13 we allocate firms in quintile of exit probability by firm-market size. In both specifications, both the coefficient on $(a_{ijkt} - \varepsilon^q_{ijkt-1})$ and its interaction with age are stable across sub-samples of firms and the results on the sample of firms most likely to survive (column (5)) is very close to the full sample (column(1)).

Table A.12: Demand shocks and beliefs updating, by exit probability

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dep. var. | | | $\Delta\varepsilon^q_{ijk,t+1}$ | | |
| Exit prob. | All | Bottom 80% | Bottom 60% | Bottom 40% | Bottom 20% |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | $0.075^a$ | $0.075^a$ | $0.074^a$ | $0.071^a$ | $0.070^a$ |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) |
| $\times$ $\text{Age}_{ijkt}$ | $-0.003^a$ | $-0.003^a$ | $-0.003^a$ | $-0.003^a$ | $-0.003^a$ |
| | (0.000) | (0.000) | (0.001) | (0.001) | (0.001) |
| $\text{Age}_{ijkt}$ | $-0.038^a$ | $-0.047^a$ | $-0.054^a$ | $-0.057^a$ | $-0.069^a$ |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) |
| Observations | 1501766 | 1154290 | 839245 | 531182 | 248194 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). $\text{Age}_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). Predicted exit probabilities are obtained by from the estimation of Table A.11, column (4).

These results suggest that endogenous exit does not bias our results. We can go further and try to account for a potential selection bias by including a correction term in our estimations. Tables A.14 and A.15 directly account for the potential selection bias by including a correction term in our estimation of equation (19). The high dimensionally of the fixed effects implied by prediction 4 in Section G (see equation (G)) for the selection equation prevents us from using a probit or other maximum likelihood estimator and implementing the standard Heckman procedure. In his review of the literature on endogenous sample selection Vella (1998) however

Table A.13: Demand shocks and beliefs updating, by exit probability (robustness)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Dep. var. | | | $\Delta\varepsilon^q_{ijk,t+1}$ | | |
| Exit prob. | All | Bottom 80% | Bottom 60% | Bottom 40% | Bottom 20% |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | 0.090$^a$ | 0.093$^a$ | 0.092$^a$ | 0.093$^a$ | 0.091$^a$ |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.004) |
| $\times$ Age$_{ijkt}$ | -0.004$^a$ | -0.005$^a$ | -0.005$^a$ | -0.005$^a$ | -0.005$^a$ |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Age$_{ijkt}$ | -0.041$^a$ | -0.047$^a$ | -0.049$^a$ | -0.053$^a$ | -0.060$^a$ |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) |
| Observations | 753646 | 552923 | 392223 | 248980 | 120723 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). Predicted exit probabilities are obtained by from the estimation of Table A.11, column (4). Samples of exit probabilities are constructed by quintiles of firm size.

proposes a number of alternative procedures based on linear (Olsen, 1980), semi-parametric (Cosslett, 1991), or polynominal estimations of correction terms. We report results of these three alternative procedures as well as a standard Heckman estimator ignoring the $ikt$ and $jkt$ fixed effects in the selection equation in Tables A.14 and A.15. Vella (1998) shows that the assumption of normality in the Heckman procedure can be relaxed to allow for consistent two step estimation using methods based on alternative distributional assumptions than probit in the selection equation. In particular, Vella (1998) argues that Olsen's procedure generally produces results similar to a Heckman two-step procedure. Instead of assuming Normality of the selection equation's error term, Olsen assumes that it follows a uniform distribution. Exclusion of at least one variable from the first step is required in Olsen, not in Heckman, as the Heckman estimator includes as a correction term the Inverse Mills ratio which maps the prediction of the selection equation into a correction term in a nonlinear fashion (hence the correction term is never perfectly collinear with the second-step regressors). The $ikt$ and $jkt$ fixed effects included in equation (G) can serve as exclusion variables in a linear procedure. The complete set of results is reported in columns (1)-(4) of table A.14. Alternatively Cosslett (1991) proposes a semi-parametric estimator in which the selection correction is approximated through indicator variables. In columns (5)-(8) of Table A.14, we use 100 bins corresponding to each centile of the predicted exit probabilities as correction terms. Finally, in columns (1)-(4) of Table A.15 the predicted probability of exit is introduced directly when estimating equation (19) in the form of a 10 degree polynomial. The last three columns of Table A.15 report the results of a standard two-step Heckman procedure excluding the $ikt$ and $jkt$ fixed effects in the probit estimation of the selection equation and using the nonlinearity of the Inverse Mills Ratio to identify its coefficient. Overall, all these alternative treatments of the sample selection bias leave our coefficients of interest largely unaffected, suggesting that endogenous selection is not driving our results.

Table A.14: Demand shocks and beliefs updating: controlling for endogenous exit

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^q_{ijkt+1}$ | | | | $\Delta\varepsilon^q_{ijkt+1}$ | | |
| Selection correction | | Linear | | | | Semi-parametric | | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | $0.065^a$ | $0.075^a$ | $0.075^a$ | | $0.065^a$ | $0.075^a$ | $0.075^a$ | |
| | (0.001) | (0.002) | (0.001) | | (0.001) | (0.002) | (0.001) | |
| $\times$ Age$_{ijkt}$ | | $-0.003^a$ | $-0.003^a$ | | | $-0.003^a$ | $-0.003^a$ | |
| | | (0.000) | (0.000) | | | (0.000) | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | | | $0.069^a$ | | | | $0.069^a$ |
| | | | | (0.001) | | | | (0.001) |
| $\times$ Age$_{ijkt} = 3$ | | | | $0.064^a$ | | | | $0.064^a$ |
| | | | | (0.001) | | | | (0.001) |
| $\times$ Age$_{ijkt} = 4$ | | | | $0.060^a$ | | | | $0.060^a$ |
| | | | | (0.002) | | | | (0.002) |
| $\times$ Age$_{ijkt} = 5$ | | | | $0.056^a$ | | | | $0.056^a$ |
| | | | | (0.002) | | | | (0.002) |
| $\times$ Age$_{ijkt} = 6$ | | | | $0.059^a$ | | | | $0.059^a$ |
| | | | | (0.002) | | | | (0.002) |
| $\times$ Age$_{ijkt} = 7$ | | | | $0.055^a$ | | | | $0.055^a$ |
| | | | | (0.003) | | | | (0.003) |
| $\times$ Age$_{ijkt} = 8$ | | | | $0.051^a$ | | | | $0.051^a$ |
| | | | | (0.004) | | | | (0.004) |
| $\times$ Age$_{ijkt} = 9$ | | | | $0.054^a$ | | | | $0.054^a$ |
| | | | | (0.007) | | | | (0.007) |
| $\Pr(\widehat{\text{exit}_{ijkt}})$ | $-0.409^a$ | $-0.409^a$ | $-0.409^a$ | $-0.417^a$ | | | | |
| | (0.005) | (0.005) | (0.003) | (0.005) | | | | |
| Age$_{ijkt}$ | $-0.054^a$ | $-0.054^a$ | $-0.054^a$ | | $-0.057^a$ | $-0.057^a$ | $-0.057^a$ | |
| | (0.001) | (0.001) | (0.000) | | (0.001) | (0.001) | (0.001) | |
| Observations | 1501766 | 1501766 | 1501766 | 1501766 | 1501766 | 1501766 | 1501766 | 1501766 |

Robust standard errors clustered by firm in parentheses (bootstrapped in columns (3) and (7)). [c] significant at 10%; [b] significant at 5%; [a] significant at 1%. Age dummies included alone in columns (4) and (8) but coefficients not reported. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). In columns (1)-(4), predicted exit probabilities are obtained from the estimation of Table A.11, column (4) and introduced directly in equation (19). In columns (5)-(8), they are introduced semi-parametrically in the second step, i.e. we included 100 bins corresponding to each percentile of the variable.

Table A.15: Demand shocks and beliefs updating: controlling for endogenous exit (robustness)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^q_{ijk,t+1}$ | | | | $\Delta\varepsilon^q_{ijk,t+1}$ | |
| Selection correction | | Polynomial | | | | Heckman | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | 0.065$^a$ | 0.075$^a$ | 0.075$^a$ | | 0.078$^a$ | 0.086$^a$ | |
| | (0.001) | (0.002) | (0.001) | | (0.002) | (0.004) | |
| $\times$ Age$_{ijkt}$ | | -0.003$^a$ | -0.003$^a$ | | | -0.003$^a$ | |
| | | (0.000) | (0.000) | | | (0.001) | |
| $\times$ Age$_{ijkt} = 2$ | | | | 0.069$^a$ | | | 0.080$^a$ |
| | | | | (0.001) | | | (0.002) |
| $\times$ Age$_{ijkt} = 3$ | | | | 0.064$^a$ | | | 0.078$^a$ |
| | | | | (0.001) | | | (0.003) |
| $\times$ Age$_{ijkt} = 4$ | | | | 0.060$^a$ | | | 0.076$^a$ |
| | | | | (0.002) | | | (0.004) |
| $\times$ Age$_{ijkt} = 5$ | | | | 0.056$^a$ | | | 0.062$^a$ |
| | | | | (0.002) | | | (0.005) |
| $\times$ Age$_{ijkt} = 6$ | | | | 0.059$^a$ | | | 0.072$^a$ |
| | | | | (0.002) | | | (0.006) |
| $\times$ Age$_{ijkt} = 7$ | | | | 0.055$^a$ | | | 0.070$^a$ |
| | | | | (0.003) | | | (0.008) |
| $\times$ Age$_{ijkt} = 8$ | | | | 0.051$^a$ | | | 0.065$^a$ |
| | | | | (0.004) | | | (0.011) |
| $\times$ Age$_{ijkt} = 9$ | | | | 0.055$^a$ | | | 0.058$^a$ |
| | | | | (0.007) | | | (0.016) |
| Age$_{ijkt}$ | -0.057$^a$ | -0.057$^a$ | -0.057$^a$ | | -0.594$^a$ | -0.595$^a$ | |
| | (0.001) | (0.001) | (0.001) | | (0.006) | (0.006) | |
| $\lambda$ | | | | | 4.922$^a$ | 4.922$^a$ | 4.918$^a$ |
| | | | | | (0.039) | (0.039) | (0.039) |
| Observations | 1501766 | 1501766 | 1501766 | 1501766 | 1550474 | 1550474 | 1550474 |

Robust standard errors clustered by firm in parentheses (bootstrapped in columns (3). $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. Age dummies included alone in columns (4) but coefficients not reported. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). In columns (1)-(4), predicted exit probabilities are obtained by from the estimation of Table A.11, column (4) and introduced directly in equation (19) in the form of a 10-degree polynomial. In columns (5)-(6), we use a Heckman estimator which estimates a probit in the first step (omitting the $ikt$ and $jkt$fixed effects) and introduces the inverse mills ratio ($\lambda$) in the second step.

# J Belief updating and age: additional robustness

## J.1 Extra-EU results

In Table A.16, we restrict our sample to extra-EU destination countries to check that the different declaration thresholds applying to intra-EU expeditions and extra-EU exports (as explained in footnote 11 of the main text) do not affect our results. Focusing on extra-EU countries reduces the number of observations by 40%, but does not alter our coefficients of interest compared to the baseline results in Table 2.

Table A.16: Prediction 1: demand shocks and beliefs updating (extra EU)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^q_{ijkt+1}$ | | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | $0.061^a$ | $0.072^a$ | $0.072^a$ | |
| | (0.001) | (0.002) | (0.001) | |
| $\times$ Age$_{ijkt}$ | | $-0.003^a$ | $-0.003^a$ | |
| | | (0.000) | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | | | $0.067^a$ |
| | | | | (0.001) |
| $\times$ Age$_{ijkt} = 3$ | | | | $0.062^a$ |
| | | | | (0.001) |
| $\times$ Age$_{ijkt} = 4$ | | | | $0.056^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 5$ | | | | $0.055^a$ |
| | | | | (0.003) |
| $\times$ Age$_{ijkt} = 6$ | | | | $0.055^a$ |
| | | | | (0.003) |
| $\times$ Age$_{ijkt} = 7$ | | | | $0.050^a$ |
| | | | | (0.003) |
| $\times$ Age$_{ijkt} = 8$ | | | | $0.048^a$ |
| | | | | (0.003) |
| $\times$ Age$_{ijkt} = 9$ | | | | $0.048^a$ |
| | | | | (0.004) |
| $\times$ Age$_{ijkt} = 10$ | | | | $0.045^a$ |
| | | | | (0.007) |
| Age$_{ijkt}$ | $-0.033^a$ | $-0.033^a$ | $-0.033^a$ | |
| | (0.001) | (0.001) | (0.001) | |
| Observations | 1109761 | 1109761 | 1109761 | 1109761 |

Robust standard errors clustered by firm in parentheses (bootstrapped in columns (3)). [c] significant at 10%; [b] significant at 5%; [a] significant at 1%. Sample extra EU destinations only. Age dummies included alone in columns (4) but coefficients not reported. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit).

### J.2 Alternative age definitions

So far we have treated each entry into a market as a new one: age was reset to zero in case of exit. We now check the sensitivity of our results to alternative definitions of age. We define two alternative measures of age. We first assumes that information on local demand is not forgotten by the firm when it does not serve a product-destination only one year and accordingly reset age to zero only after two consecutive years of exit. In the second definition, we assume that firms keep entirely their knowledge about local demand when they exit, regardless of the number of exit years; this third age variable is simply the number of exporting years since the first entry of the firm.

Table A.17 shows that the results using these alternative definitions are qualitatively similar to our baseline estimates. However, the effects of age – its direct effect and its effect on firms' reactions to demand signals – are slightly lower than in our baseline table.

Table A.17: Prediction 1: alternative age definitions

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^q_{ijk,t+1}$ | | | $\Delta\varepsilon^q_{ijk,t+1}$ | |
| Age definition | | # years since last entry (reset after 2 years exit) | | | # years exporting since first entry | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | $0.064^a$ | $0.073^a$ | | $0.064^a$ | $0.072^a$ | |
| | (0.001) | (0.001) | | (0.001) | (0.001) | |
| $\times$ Age$_{ijkt}$ | | $-0.002^a$ | | | $-0.002^a$ | |
| | | (0.000) | | | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | | $0.068^a$ | | | $0.068^a$ |
| | | | (0.001) | | | (0.001) |
| $\times$ Age$_{ijkt} = 3$ | | | $0.065^a$ | | | $0.065^a$ |
| | | | (0.001) | | | (0.001) |
| $\times$ Age$_{ijkt} = 4$ | | | $0.061^a$ | | | $0.063^a$ |
| | | | (0.002) | | | (0.002) |
| $\times$ Age$_{ijkt} = 5$ | | | $0.059^a$ | | | $0.061^a$ |
| | | | (0.002) | | | (0.002) |
| $\times$ Age$_{ijkt} = 6$ | | | $0.060^a$ | | | $0.061^a$ |
| | | | (0.002) | | | (0.002) |
| $\times$ Age$_{ijkt} = 7$ | | | $0.057^a$ | | | $0.058^a$ |
| | | | (0.002) | | | (0.003) |
| $\times$ Age$_{ijkt} = 8$ | | | $0.056^a$ | | | $0.057^a$ |
| | | | (0.003) | | | (0.003) |
| $\times$ Age$_{ijkt} = 9$ | | | $0.054^a$ | | | $0.054^a$ |
| | | | (0.004) | | | (0.004) |
| $\times$ Age$_{ijkt} = 10$ | | | $0.047^a$ | | | $0.047^a$ |
| | | | (0.007) | | | (0.007) |
| Age$_{ijkt}$ | $-0.030^a$ | $-0.030^a$ | | $-0.029^a$ | $-0.029^a$ | |
| | (0.001) | (0.001) | | (0.001) | (0.001) | |
| Observations | 1854141 | 1854141 | 1854141 | 1854141 | 1854141 | 1854141 |

Robust standard errors clustered by firm in parentheses. [c] significant at 10%; [b] significant at 5%; [a] significant at 1%. Age dummies included alone in columns (3) and (6) but coefficients not reported. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14).

### J.3 Reconstructed years

The usual aggregation of export sales by calendar year is likely to bias downward the average sales of new exporters because some enter a given market late in the year (Berthou and Vicard, 2015). The average growth rate of quantities would in turn be inflated between the first, potentially incomplete, and the second (full) year of export. When estimating equation (19), the dummy for age two picks the average bias related to the incompleteness of the first year of export. In Table A.18, we go one step further and address this issue directly by performing our estimation strategy on reconstructed years beginning the month of first entry at the firm-product-destination level. The results shows that both the average updating of the firms' beliefs and its interaction with age are quantitatively similar to our baseline in Table 2.

The drawback of using such reconstructed yearly data is the inability to control consistently for market-year fixed effects in equations (14) and (15): introducing market$\times$year fixed effects specific by firms' month of entry reduces dramatically the number of observations for which we can identify beliefs and demand shocks. We therefore stick to the usual calendar year dataset in the main text.

Table A.18: Prediction 1: reconstructed years

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^q_{ijkt+1}$ | | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | $0.067^a$ | $0.078^a$ | $0.078^a$ | |
| | (0.001) | (0.001) | (0.001) | |
| $\times$ Age$_{ijkt}$ | | $-0.003^a$ | $-0.003^a$ | |
| | | (0.000) | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | | | $0.072^a$ |
| | | | | (0.001) |
| $\times$ Age$_{ijkt} = 3$ | | | | $0.066^a$ |
| | | | | (0.001) |
| $\times$ Age$_{ijkt} = 4$ | | | | $0.066^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 5$ | | | | $0.058^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 6$ | | | | $0.061^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 7$ | | | | $0.054^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 8$ | | | | $0.056^a$ |
| | | | | (0.004) |
| $\times$ Age$_{ijkt} = 9$ | | | | $0.056^a$ |
| | | | | (0.005) |
| Age$_{ijkt}$ | $-0.010^a$ | $-0.010^a$ | $-0.010^a$ | |
| | (0.001) | (0.001) | (0.001) | |
| Observations | 1495774 | 1495774 | 1495774 | 1495774 |

Robust standard errors clustered by firm in parentheses (bootstrapped in column (3)). $^c$ significant at 10%; $^b$ significant at 5%. $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). Age dummies included alone in column (4) but coefficients not reported. In this Table years are reconstructed beginning the month of first entry at the firm-product-destination level

### J.4 $\sigma_k$ computed at 4-digit (HS4) level

In Table A.19, we use demand shocks obtained by estimating equation (17) by 4-digit product instead of 6-digit product of the Harmonized System classification in order to allow for a larger number of observations when estimating $\sigma_k$. As expected, our estimates of $\sigma_k$ are slightly lower in this case than in the baseline 6-digit case (a median of 4.98 and a mean of 5.83). The results shown in Table A.19 are close to our baseline results.

Table A.19: Prediction 1: $\sigma_k$ computed at 4-digit (HS4) level

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. var. | $\Delta\varepsilon^q_{ijkt+1}$ | | | |
| Robustness | $\sigma_k$ at HS4 level | | | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | 0.073$^a$ | 0.084$^a$ | 0.084$^a$ | |
| | (0.001) | (0.001) | (0.001) | |
| $\times$ Age$_{ijkt}$ | | -0.004$^a$ | -0.004$^a$ | |
| | | (0.000) | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | | | 0.078$^a$ |
| | | | | (0.001) |
| $\times$ Age$_{ijkt} = 3$ | | | | 0.072$^a$ |
| | | | | (0.001) |
| $\times$ Age$_{ijkt} = 4$ | | | | 0.067$^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 5$ | | | | 0.065$^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 6$ | | | | 0.064$^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 7$ | | | | 0.059$^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 8$ | | | | 0.060$^a$ |
| | | | | (0.003) |
| $\times$ Age$_{ijkt} = 9$ | | | | 0.059$^a$ |
| | | | | (0.004) |
| $\times$ Age$_{ijkt} = 10$ | | | | 0.057$^a$ |
| | | | | (0.006) |
| Age$_{ijkt}$ | -0.032$^a$ | -0.033$^a$ | -0.033$^a$ | |
| | (0.001) | (0.001) | (0.000) | |
| Observations | 1877732 | 1877732 | 1877732 | 1877732 |

Robust standard errors clustered by firm in parentheses (bootstrapped in column (3)). $^c$ significant at 10%; $^b$ significant at 5%. $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17), estimated by HS4 products instead of HS6; $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand from equation (14). Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). Age dummies included alone in column (4) but coefficients not reported.

### J.5 Controlling for $ijt$ fixed effects

The theoretical framework developed in section 3 assumes no informational spillovers, considering $\theta_{ijk0}$ as exogenous. While our identification strategy controls *de facto* for several sources of informational spillovers – the firm×product×year fixed effects included in equations (14) and (15) account for past experience gathered from selling the same product on the domestic or other markets –, it does not take into those from selling other products in the same destination. To this end, we extend our identification strategy by including $ijt$ fixed effects in equations (12) and (13) and re-estimate $a_{ijkt}$ from these alternative $\varepsilon^q_{ijkt}$ and $\varepsilon^p_{ijkt}$ to test prediction 1. Table A.20 reports the results and show that our conclusion remain robust qualitatively as well as quantitatively. This lends support to our assumption that information is indeed mostly product-market specific. If shocks and beliefs were correlated across products within destinations, the firms' response to a demand shock would partly reflect its belief updating behavior on other products: including $ijt$ fixed effects should dampen the extent of estimated belief updating.

Table A.20: Prediction 1: controlling for $ijt$ fixed effects

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Dep. var. | | $\Delta\varepsilon^q_{ijkt+1}$ | | |
| $a_{ijkt} - \varepsilon^q_{ijkt}$ | $0.091^a$ | $0.102^a$ | $0.102^a$ | |
| | (0.001) | (0.002) | (0.001) | |
| $\times$ Age$_{ijkt}$ | | $-0.003^a$ | $-0.003^a$ | |
| | | (0.001) | (0.000) | |
| $\times$ Age$_{ijkt} = 2$ | | | | $0.096^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 3$ | | | | $0.092^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 4$ | | | | $0.087^a$ |
| | | | | (0.002) |
| $\times$ Age$_{ijkt} = 5$ | | | | $0.085^a$ |
| | | | | (0.003) |
| $\times$ Age$_{ijkt} = 6$ | | | | $0.082^a$ |
| | | | | (0.003) |
| $\times$ Age$_{ijkt} = 7$ | | | | $0.076^a$ |
| | | | | (0.004) |
| $\times$ Age$_{ijkt} = 8$ | | | | $0.078^a$ |
| | | | | (0.005) |
| $\times$ Age$_{ijkt} = 9$ | | | | $0.079^a$ |
| | | | | (0.005) |
| Age$_{ijkt}$ | $-0.013^a$ | $-0.013^a$ | $-0.013^a$ | |
| | (0.000) | (0.000) | (0.001) | |
| Observations | 1217810 | 1217810 | 1217810 | 1217810 |

Robust standard errors clustered by firm in parentheses (bootstrapped in column (3)). $^c$ significant at 10%; $^b$ significant at 5%. $^a$ significant at 1%. $a_{ijkt}$ is our estimate of the demand shock from equation (17); $\varepsilon^q_{ijkt}$ is the belief of the firm about future demand. $\varepsilon^q_{ijkt-1}$ and $\varepsilon^p_{ijkt-1}$ are respectively estimated from equation (14) and equation (15) including additionally fixed effects in the $ijt$ dimension. Age$_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). Age dummies included alone in column (4) but coefficients not reported.

# K Test of stationary demand

In the learning model, firms learn about an idiosyncratic demand parameter which is assumed to be constant over time. The initial size of a firm (i.e. its initial belief) should be a useful predictor of its beliefs and sales throughout its life, even controlling for past beliefs. In other words, the evolution of firms' beliefs should not be Markov. Such a prediction would not arise in models with "active learning" where firms invest to increase their profitability, possibly through demand accumulation. To discriminate between these two classes of models, Pakes and Ericson (1998) (see also Abbring and Campbell, 2005 for an application) propose to regress current firms beliefs on their immediate past beliefs and their initial prior beliefs. In Table A.21, we regress the beliefs of the firms after $x$ years, $x = 3, ..., 8$, on their belief at the time of entry, controlling for the immediate lag of the belief. We restrict our sample to firms present at least 8 years to avoid composition effects.[12] Two results are worth mentioning. First, initial beliefs have a positive and significant effect on future beliefs, and this effect remains highly significant even 8 years after entry. Second, the immediate lag of the belief becomes a better predictor of the current belief as the firm gets older, suggesting that firms indeed converge to their demand parameter. Both results are consistent with our assumption on $\overline{a}_{ijk}$. Note that these results are not sensitive to the number of lags used: Table A.22 focuses on firms aged 6 to 8 years for which we can include up to four lags of the belief (we find a similar pattern for firms aged 5 to 8 years for which we can include up to 3 lags). We find that the initial belief remains a significant predictor of current belief after 6, 7 or 8 years when increasing the number of lags of beliefs included as explanatory variables.

[12]Similar results are obtained when restricting the sample to firms present $j$ years, $j = 5, ..., 9$.

Table A.21: Passive versus active learning

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dep. var. | $\varepsilon^q_{ijkt}$ | | | | | |
| Age definition | # years since last entry (reset after 1 year of exit) | | | | | |
| Age | 3 | 4 | 5 | 6 | 7 | 8 |
| $\varepsilon^q_{ijkt-1}$ | $0.541^a$ | $0.587^a$ | $0.632^a$ | $0.645^a$ | $0.659^a$ | $0.668^a$ |
| | (0.008) | (0.007) | (0.007) | (0.006) | (0.007) | (0.007) |
| $\varepsilon^q_{ijk0}$ | $0.144^a$ | $0.135^a$ | $0.097^a$ | $0.091^a$ | $0.079^a$ | $0.075^a$ |
| | (0.006) | (0.006) | (0.005) | (0.005) | (0.005) | (0.005) |
| Observations | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. $\varepsilon^q_{ijkt-1}$ and $\varepsilon^q_{ijk0}$ are respectively the beliefs of the firm in market $jk$ in period $t-1$ and in the first period. Beliefs given by equation (14). Sample of firms-markets present at least 8 years.

Table A.22: Passive versus active learning: robustness

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dep. var. | $\varepsilon^q_{ijkt}$ | | | | | | | | | | | |
| Age definition | # years since last entry (reset after 1 year of exit) | | | | | | | | | | | |
| Age | 6 | | | | 7 | | | | 8 | | | |
| $\varepsilon^q_{ijk0}$ | $0.091^a$ | $0.050^a$ | $0.031^a$ | $0.020^a$ | $0.079^a$ | $0.045^a$ | $0.029^a$ | $0.019^a$ | $0.075^a$ | $0.046^a$ | $0.032^a$ | $0.023^a$ |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.004) | (0.005) | (0.005) |
| $\varepsilon^q_{ijkt-1}$ | $0.645^a$ | $0.521^a$ | $0.504^a$ | $0.503^a$ | $0.659^a$ | $0.520^a$ | $0.503^a$ | $0.498^a$ | $0.668^a$ | $0.530^a$ | $0.509^a$ | $0.505^a$ |
| | (0.006) | (0.007) | (0.008) | (0.008) | (0.007) | (0.007) | (0.008) | (0.008) | (0.007) | (0.008) | (0.009) | (0.009) |
| $\varepsilon^q_{ijkt-2}$ | | $0.210^a$ | $0.172^a$ | $0.168^a$ | | $0.223^a$ | $0.181^a$ | $0.174^a$ | | $0.219^a$ | $0.171^a$ | $0.163^a$ |
| | | (0.007) | (0.008) | (0.008) | | (0.007) | (0.008) | (0.008) | | (0.007) | (0.008) | (0.008) |
| $\varepsilon^q_{ijkt-3}$ | | | $0.089^a$ | $0.076^a$ | | | $0.090^a$ | $0.069^a$ | | | $0.099^a$ | $0.075^a$ |
| | | | (0.006) | (0.007) | | | (0.007) | (0.008) | | | (0.007) | (0.008) |
| $\varepsilon^q_{ijkt-4}$ | | | | $0.034^a$ | | | | $0.050^a$ | | | | $0.054^a$ |
| | | | | (0.007) | | | | (0.006) | | | | (0.007) |
| Observations | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 | 41034 |

Robust standard errors clustered by firm in parentheses. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. $\varepsilon^q_{ijkt-1}$ and $\varepsilon^q_{ijk0}$ are respectively the beliefs of the firm in market $jk$ in period $t-1$ and in the first period. Beliefs given by equation (14).

# L Profiles of prices and quantities

Table A.23 provides the full set of results used in figure 2 of the main text. In columns (9) and (10) we additionally include firms' size as an explaining variable when regressing $\varepsilon^p_{ijkt}$ on age, to control for the fact that size would affect firms' pricing decisions in a non-CES framework. It confirms that $\varepsilon^p_{ijkt}$ decreases with age (column (9)) but not when we account for composition effects through the inclusion of firm×market fixed effects (column (10)). Table A.24 reproduces estimations in Table A.23 on our dataset with reconstructed years. As expected, reconstructing years from the month of first entry by firm×product×destination dampens the initial increase in quantities sold between the first and second year but leaves unchanged the quantity profile thereafter.

Table A.23: Dynamics of quantity and prices

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Dep. var. | | | $\varepsilon^q_{ijkt}$ | | | | $\varepsilon^p_{ijkt}$ | | | |
| Sample | | | Surv. | | | | Surv. | | | |
| $\text{Age}_{ijkt}$ | 0.095$^a$ | | | | -0.012$^a$ | | | | | |
| | (0.001) | | | | (0.001) | | | | | |
| $\text{Age}_{ijkt} = 2$ | | 0.192$^a$ | 0.317$^a$ | 0.150$^a$ | | -0.027$^a$ | -0.038$^a$ | -0.008$^a$ | | |
| | | (0.002) | (0.015) | (0.003) | | (0.001) | (0.014) | (0.001) | | |
| $\text{Age}_{ijkt} = 3$ | | 0.297$^a$ | 0.447$^a$ | 0.193$^a$ | | -0.039$^a$ | -0.047$^a$ | -0.011$^a$ | -0.007$^a$ | -0.002 |
| | | (0.003) | (0.018) | (0.004) | | (0.002) | (0.014) | (0.002) | (0.001) | (0.002) |
| $\text{Age}_{ijkt} = 4$ | | 0.378$^a$ | 0.511$^a$ | 0.212$^a$ | | -0.050$^a$ | -0.042$^a$ | -0.014$^a$ | -0.015$^a$ | -0.003 |
| | | (0.004) | (0.019) | (0.005) | | (0.002) | (0.014) | (0.002) | (0.002) | (0.002) |
| $\text{Age}_{ijkt} = 5$ | | 0.433$^a$ | 0.563$^a$ | 0.219$^a$ | | -0.053$^a$ | -0.047$^a$ | -0.014$^a$ | -0.018$^a$ | -0.002 |
| | | (0.005) | (0.021) | (0.007) | | (0.003) | (0.015) | (0.003) | (0.002) | (0.003) |
| $\text{Age}_{ijkt} = 6$ | | 0.477$^a$ | 0.594$^a$ | 0.219$^a$ | | -0.058$^a$ | -0.054$^a$ | -0.014$^a$ | -0.021$^a$ | -0.003 |
| | | (0.006) | (0.023) | (0.009) | | (0.003) | (0.015) | (0.004) | (0.003) | (0.003) |
| $\text{Age}_{ijkt} = 7$ | | 0.526$^a$ | 0.595$^a$ | 0.221$^a$ | | -0.065$^a$ | -0.050$^a$ | -0.018$^a$ | -0.028$^a$ | -0.006 |
| | | (0.008) | (0.022) | (0.011) | | (0.004) | (0.015) | (0.005) | (0.004) | (0.004) |
| $\text{Age}_{ijkt} = 8$ | | 0.556$^a$ | 0.571$^a$ | 0.219$^a$ | | -0.064$^a$ | -0.049$^a$ | -0.017$^b$ | -0.027$^a$ | -0.006 |
| | | (0.010) | (0.022) | (0.013) | | (0.006) | (0.015) | (0.007) | (0.006) | (0.007) |
| $\text{Age}_{ijkt} = 9$ | | 0.593$^a$ | 0.563$^a$ | 0.221$^a$ | | -0.077$^a$ | -0.043$^a$ | -0.024 | -0.040$^a$ | -0.012 |
| | | (0.014) | (0.022) | (0.018) | | (0.013) | (0.015) | (0.016) | (0.013) | (0.015) |
| $\text{Age}_{ijkt} = 10$ | | 0.614$^a$ | 0.496$^a$ | 0.210$^a$ | | -0.070$^a$ | -0.047$^a$ | -0.016$^c$ | -0.031$^a$ | -0.004 |
| | | (0.016) | (0.021) | (0.019) | | (0.006) | (0.015) | (0.009) | (0.007) | (0.008) |
| $\text{Size}_{ijk,t-1}$ | | | | | | | | | -0.016$^a$ | 0.006$^a$ |
| | | | | | | | | | (0.001) | (0.001) |
| Observations | 4382989 | 4382989 | 121775 | 4382989 | 4382989 | 4382989 | 121775 | 4382989 | 1883888 | 1883888 |
| Firm×Destination×Product FE | No | No | No | Yes | No | No | No | Yes | No | Yes |

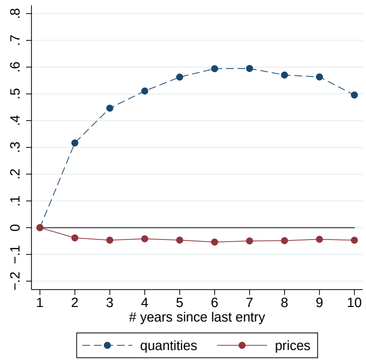
Standard errors clustered by firm in parentheses. [c] significant at 10%; [b] significant at 5%; [a] significant at 1%. $\varepsilon^q_{ijkt}$ and $\varepsilon^p_{ijkt}$ and the residuals from the quantity and prices equations (14) and (15). $\text{Size}_{ijk,t-1}$ is defined at the total quantity sold in market $jk$ by firm $i$ in year $t-1$. $\text{Age}_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). $\text{Size}_{ijkt}$ is proxied by the value sold by firm $i$ on market $jk$ during year $t$ divided by the total value exported by French firms in market $jk$ during year $t$. Similar price profiles are obtained when using alternative measures of firm size in columns (9)-(10). "Surv." means that we restrict the sample to firms-markets surviving the entire period.

Table A.24: Dynamics of quantity and prices (reconstructed years)

| Dep. var. | (1) | (2) $\varepsilon^q_{ijkt}$ | (3) | (4) | (5) | (6) $\varepsilon^p_{ijkt}$ | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| $\text{Age}_{ijkt}$ | 0.085$^a$ | | | -0.012$^a$ | | | | |
| | (0.001) | | | (0.001) | | | | |
| $\text{Age}_{ijkt} = 2$ | | 0.109$^a$ | 0.025$^a$ | | -0.022$^a$ | -0.002 | | |
| | | (0.002) | (0.003) | | (0.001) | (0.002) | | |
| $\text{Age}_{ijkt} = 3$ | | 0.229$^a$ | 0.067$^a$ | | -0.037$^a$ | -0.006$^b$ | -0.013$^a$ | -0.003 |
| | | (0.004) | (0.005) | | (0.002) | (0.003) | (0.002) | (0.002) |
| $\text{Age}_{ijkt} = 4$ | | 0.312$^a$ | 0.087$^a$ | | -0.045$^a$ | -0.008$^a$ | -0.019$^a$ | -0.004$^b$ |
| | | (0.004) | (0.006) | | (0.002) | (0.003) | (0.002) | (0.002) |
| $\text{Age}_{ijkt} = 5$ | | 0.364$^a$ | 0.089$^a$ | | -0.049$^a$ | -0.009$^a$ | -0.023$^a$ | -0.005$^c$ |
| | | (0.005) | (0.007) | | (0.003) | (0.003) | (0.002) | (0.003) |
| $\text{Age}_{ijkt} = 6$ | | 0.419$^a$ | 0.088$^a$ | | -0.057$^a$ | -0.014$^a$ | -0.029$^a$ | -0.009$^a$ |
| | | (0.007) | (0.009) | | (0.004) | (0.004) | (0.003) | (0.003) |
| $\text{Age}_{ijkt} = 7$ | | 0.461$^a$ | 0.087$^a$ | | -0.059$^a$ | -0.014$^b$ | -0.030$^a$ | -0.009$^b$ |
| | | (0.009) | (0.012) | | (0.004) | (0.005) | (0.004) | (0.005) |
| $\text{Age}_{ijkt} = 8$ | | 0.491$^a$ | 0.079$^a$ | | -0.066$^a$ | -0.015$^c$ | -0.037$^a$ | -0.010 |
| | | (0.011) | (0.014) | | (0.007) | (0.009) | (0.007) | (0.008) |
| $\text{Age}_{ijkt} = 9$ | | 0.515$^a$ | 0.080$^a$ | | -0.063$^a$ | -0.012 | -0.032$^a$ | -0.006 |
| | | (0.016) | (0.019) | | (0.006) | (0.008) | (0.007) | (0.007) |
| $\text{Size}_{ijk,t-1}$ | | | | | | | -0.015$^a$ | 0.011$^a$ |
| | | | | | | | (0.001) | (0.001) |
| Observations | 3741140 | 3741140 | 3741140 | 3741140 | 3741140 | 3741140 | 1524261 | 1524261 |
| Firm×Destination×Product FE | No | No | Yes | No | No | Yes | No | Yes |

Standard errors clustered by firm in parentheses. [c] significant at 10%; [b] significant at 5%; [a] significant at 1%. $\varepsilon^q_{ijkt}$ and $\varepsilon^p_{ijkt}$ and the residuals from the quantity and prices equations (14) and (15). $\text{Size}_{ijk,t-1}$ is defined at the total quantity sold in market $jk$ by firm $i$ in year $t-1$. $\text{Age}_{ijkt}$ is the number of years since the last entry of the firm on market $jk$ (reset to zero after one year of exit). $\text{Size}_{ijkt}$ is proxied by the value sold by firm $i$ on market $jk$ during year $t$ divided by the total value exported by French firms in market $jk$ during year $t$. "Surv." means that we restrict the sample to firms-markets surviving the entire period. This table contains estimations similar to table A.23, except that they are ran on a sample in which years have been reconstructed by firm-market, starting from the first month of export.

Figure A.5: Dynamics of prices and quantities residuals: surviving firms (1996-2005)




Note: This figure plots the coefficients obtained when regressing the prices and quantities residuals $\varepsilon^p_{ijkt}$ and $\varepsilon^q_{ijkt}$ on a set of age dummies and restricting the sample to firms-markets surviving the entire period. The complete set of coefficients and standard errors are shown in Table A.23 (columns (4) and (8)).

# M Variance of growth rates: robustness

Tables A.25 and A.26 report the full set of results used to draw figure 3 in the main text. Table A.25 shows that the variance of both $\varepsilon^{q}_{ijkt}$ and $\varepsilon^{p}_{ijkt}$ decreases with age. As expected, the decline in the variance is larger for the quantity residuals. These results are robust to controlling for the number of observations (columns (3) and (7)), focusing on permanent exporters that survive throughout our time span (columns (4) and (8)), controlling for the average firm size in the cohort (columns (3)-(6) of Table A.26) or using our alternative definitions of age (columns (7)-(14) of Table A.26). Finally, columns (1) and (2) of Table A.26 confirm that the variance of $\varepsilon^{value}_{ijkt}$ decreases sharply with age as well.

Table A.25: Prediction 2.b: age and variance of growth rates

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Dep. var. | $\mathbb{V}(\Delta\varepsilon^q_{ijkt})$ | | | | $\mathbb{V}(\Delta\varepsilon^p_{ijkt})$ | | | |
| Age definition | # years since last entry (reset after 1 year of exit) | | | | # years since last entry (reset after 1 year of exit) | | | |
| Sample | All | | | Permanent exporters[1] | All | | | Permanent exporters[1] |
| $\text{Age}_{ijkt}$ | -0.051$^a$ | | -0.045$^a$ | -0.018$^a$ | -0.029$^a$ | | -0.024$^a$ | -0.007$^a$ |
| | (0.001) | | (0.001) | (0.002) | (0.001) | | (0.001) | (0.001) |
| $\text{Age}_{ijkt} = 3$ | | -0.110$^a$ | | | | -0.066$^a$ | | |
| | | (0.004) | | | | (0.002) | | |
| $\text{Age}_{ijkt} = 4$ | | -0.169$^a$ | | | | -0.098$^a$ | | |
| | | (0.005) | | | | (0.003) | | |
| $\text{Age}_{ijkt} = 5$ | | -0.211$^a$ | | | | -0.120$^a$ | | |
| | | (0.005) | | | | (0.003) | | |
| $\text{Age}_{ijkt} = 6$ | | -0.236$^a$ | | | | -0.134$^a$ | | |
| | | (0.006) | | | | (0.004) | | |
| $\text{Age}_{ijkt} = 7$ | | -0.269$^a$ | | | | -0.148$^a$ | | |
| | | (0.008) | | | | (0.004) | | |
| $\text{Age}_{ijkt} = 8$ | | -0.303$^a$ | | | | -0.164$^a$ | | |
| | | (0.009) | | | | (0.005) | | |
| $\text{Age}_{ijkt} = 9$ | | -0.295$^a$ | | | | -0.171$^a$ | | |
| | | (0.012) | | | | (0.007) | | |
| $\text{Age}_{ijkt} = 10$ | | -0.338$^a$ | | | | -0.185$^a$ | | |
| | | (0.018) | | | | (0.011) | | |
| # observations | | | 0.008$^a$ | 0.006 | | | 0.005$^a$ | 0.005 |
| | | | (0.001) | (0.006) | | | (0.000) | (0.004) |
| Observations | 434593 | 434593 | 434593 | 44421 | 434593 | 434593 | 434593 | 44421 |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Standard errors clustered by cohort in parentheses. A cohort of exporters in a product-destination market includes all firms starting to export to that market in a given year. Cohort fixed effects included in all estimations. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. [1] firms present all years on market $jk$. "# observations " is the number of observations of the cohort in the current year.

Table A.26: Prediction 2: variance of growth rates: robustness

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dep. var. | $\mathbb{V}(\Delta\varepsilon^{value}_{ijkt})$ | | $\mathbb{V}(\Delta\varepsilon^{q}_{ijkt})$ | | $\mathbb{V}(\Delta\varepsilon^{p}_{ijkt})$ | | $\mathbb{V}(\Delta\varepsilon^{q}_{ijkt})$ | | $\mathbb{V}(\Delta\varepsilon^{p}_{ijkt})$ | | $\mathbb{V}(\Delta\varepsilon^{q}_{ijkt})$ | | $\mathbb{V}(\Delta\varepsilon^{p}_{ijkt})$ | |
| Age definition | | | # years since last entry (reset after 1 year of exit) | | | | # years since last entry (reset after 2 years exit) | | | | # years exporting since first entry | | | |
| $\text{Age}_{ijkt}$ | -0.044$^a$ | | -0.049$^a$ | | -0.026$^a$ | | -0.053$^a$ | | -0.038$^a$ | | -0.050$^a$ | | -0.035$^a$ | |
| | (0.001) | | (0.001) | | (0.001) | | (0.001) | | (0.001) | | (0.001) | | (0.001) | |
| $\text{Age}_{ijkt} = 3$ | | -0.093$^a$ | | -0.107$^a$ | | -0.060$^a$ | | -0.082$^a$ | | -0.069$^a$ | | -0.063$^a$ | | -0.054$^a$ |
| | | (0.003) | | (0.004) | | (0.002) | | (0.004) | | (0.002) | | (0.004) | | (0.003) |
| $\text{Age}_{ijkt} = 4$ | | -0.139$^a$ | | -0.165$^a$ | | -0.089$^a$ | | -0.138$^a$ | | -0.111$^a$ | | -0.116$^a$ | | -0.090$^a$ |
| | | (0.004) | | (0.005) | | (0.003) | | (0.004) | | (0.003) | | (0.004) | | (0.003) |
| $\text{Age}_{ijkt} = 5$ | | -0.179$^a$ | | -0.206$^a$ | | -0.111$^a$ | | -0.188$^a$ | | -0.148$^a$ | | -0.165$^a$ | | -0.128$^a$ |
| | | (0.005) | | (0.006) | | (0.003) | | (0.005) | | (0.003) | | (0.005) | | (0.003) |
| $\text{Age}_{ijkt} = 6$ | | -0.201$^a$ | | -0.231$^a$ | | -0.124$^a$ | | -0.230$^a$ | | -0.173$^a$ | | -0.211$^a$ | | -0.154$^a$ |
| | | (0.005) | | (0.007) | | (0.004) | | (0.006) | | (0.004) | | (0.005) | | (0.004) |
| $\text{Age}_{ijkt} = 7$ | | -0.230$^a$ | | -0.264$^a$ | | -0.138$^a$ | | -0.279$^a$ | | -0.197$^a$ | | -0.258$^a$ | | -0.182$^a$ |
| | | (0.007) | | (0.008) | | (0.005) | | (0.006) | | (0.004) | | (0.006) | | (0.004) |
| $\text{Age}_{ijkt} = 8$ | | -0.252$^a$ | | -0.298$^a$ | | -0.153$^a$ | | -0.311$^a$ | | -0.223$^a$ | | -0.297$^a$ | | -0.209$^a$ |
| | | (0.008) | | (0.009) | | (0.005) | | (0.008) | | (0.005) | | (0.008) | | (0.005) |
| $\text{Age}_{ijkt} = 9$ | | -0.258$^a$ | | -0.290$^a$ | | -0.160$^a$ | | -0.346$^a$ | | -0.251$^a$ | | -0.336$^a$ | | -0.242$^a$ |
| | | (0.010) | | (0.012) | | (0.007) | | (0.010) | | (0.006) | | (0.010) | | (0.006) |
| $\text{Age}_{ijkt} = 10$ | | -0.287$^a$ | | -0.332$^a$ | | -0.170$^a$ | | -0.433$^a$ | | -0.293$^a$ | | -0.423$^a$ | | -0.283$^a$ |
| | | (0.015) | | (0.018) | | (0.011) | | (0.018) | | (0.010) | | (0.017) | | (0.010) |
| $\text{Size}_{t-1}$ | | | -0.015$^a$ | -0.008$^a$ | -0.013$^a$ | -0.009$^a$ | | | | | | | | |
| | | | (0.002) | (0.002) | (0.001) | (0.001) | | | | | | | | |
| Observations | 434593 | 434593 | 434593 | 434593 | 434593 | 434593 | 246772 | 246772 | 246772 | 246772 | 251236 | 251236 | 251236 | 251236 |
| Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

Standard errors clustered by cohort in parentheses. A cohort of exporters in a product-destination market includes all firms starting to export to that market in a given year. Cohort fixed effects included in all estimations. $^c$ significant at 10%; $^b$ significant at 5%; $^a$ significant at 1%. [1] firms present all years on market $jk$. "# observations " is the number of observations of the cohort in the current year.

# N Additional references


Ayyagari, M., Demirguc-Kunt, A. and Maksimovic, V. (2011), "Small vs. young firms across the world : contribution to employment, job creation, and growth", Policy Research Working Paper Series 5631, The World Bank.

Berman, N., de Sousa, J., Martin P. and T. Mayer (2013), "Time to Ship during Financial Crises.", *NBER International Seminar on Macroeconomics*, 9, 225-260.

Bricongne, J.-C., Fontagne, L., Gaulier, G., Taglioni, D. and Vicard, V. (2012), "Firms and the global crisis: French exports in the turmoil", *Journal of International Economics*, vol. 87 no 1: pp. 134-146.

Criscuolo, C., Gal, P. N. and Menon, C. (2014), "The Dynamics of Employment Growth: New Evidence from 18 Countries", CEP Discussion Papers 1274, Centre for Economic Performance, LSE.

Lawless, M. (2014), "Age or size? Contributions to job creation", *Small Business Economics*, vol. 42 no 4: pp. 815-830.

Lawless, M. and Whelan, K. (2014), "Where do firms export, how much, and why?", *World Economy*, vol. 37.

Munch, J. and D. Nguyen (2014), "Decomposing firm-level variations", *Journal of Economic Behavior and Organization*, vol. 106 pp. 317-334.

Nunn, N. (2007), "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade.", *The Quarterly Journal of Economics*, 122, 569-600.

Vella, Francis (1998), "Estimating Models with Sample Selection Bias: A Survey." *Journal of Human Resources*, 33, 127-169.