Appendices
An empirical dynamic model of trade with consumer accumulation

A Constructions of the samples

The dataset used in the paper is initially disaggregated at the monthly level. From this raw dataset, a number of steps are implemented to improve the reliability and consistency of the data. First, I describe the operations implemented for the first empirical exercise, that uses a wide set of products. Then, I describe the procedures implemented to obtain the final sample used in the structural estimation.

A Data appendix for the reduced-form exercise

I implement two important steps to prepare the data for the regressions displayed in the reduced-form exercise. First, I clean outliers and product categories that do not provide a meaningful and consistent unit of count across years. Second, I correct for the partial-year bias.

Cleaning and harmonization I make three different operations to clean the dataset from potential outliers or measurement errors.

- First, I use the algorithm from Pierce and Schott (2012) and Van Beveren, Bernard, and Vandenbussche (2012) to account for changes in product categories at the eight digit level. This algorithm allows me to obtain categories that are consistent across the sample years (1996-2010).
- Second, I drop product categories that meet one of the following criteria:
  - the counting unit is changing across years.
  - the counting unit is not identical within the category (because of the previous step, the current product category can contain eight digit categories with different units).
- Finally, because unit values, constructed as export values divided by quantities, are a source of measurement errors, I winsorize them at the eight-digit product category$\times$country$\times$year level. Specifically, I set at the values of the 5th and 95th percentiles the prices that are beyond these two thresholds.

Correction for partial-year bias As described in Berthou and Vicard (2015) and Bernard, Boler, Massari, Reyes, and Taglioni (2017), a firm will sell less in average during its first calendar year as exporter. This is because calendar years do not necessarily match the beginning of the exporting activity. In order to correct for this potential bias, I reconstruct the dataset to align calendar exporting years of each exporter. The idea is to define a new year for each spell of export, setting the first month of this year as representative of a regular year, and constructing exporting spells based on this new starting month.

Specifically, the following procedure is applied to each firm-destination-product triplet: for the earliest observation in 1996, if no observation is seen in 1995, a new spell is defined: the month of this first flow is probabilistically drawn based on the number of flows observed during the following 12 months. Then, the year is set to 1996 or 1997 depending on whether the initial month is earlier or later than July. The following observations are adjusted accordingly to preserve the duration between monthly export flows, as long as there is no discontinuity in the exporting activity according to the newly defined calendar years. In case of discontinuity, the next observation becomes a new reference point, and the same procedure is applied for this observation and the following ones.

Once this adjustment is implemented, I aggregate the data at the yearly-level. Specifically, I sum values exported within each newly created calendar year at the firm-product-category level. Moreover, I obtain yearly prices using an export-weighted average of monthly prices. In case of missing prices, I assume a weight of zero for this observation.[69] If this observation is the only observation within a firm-destination-product- year combination, I drop all the observations within the firm-destination-product triplet.

This procedure leaves me with sales and prices measured at the firm-product-destination-year level, with no missing observation in prices, and adjusted for the existence of partial-year of exporting.

B Data appendix for the structural estimation

The procedure to clean the data for the structural estimation is different than the reduced-form exercise. I describe in this subsection the choice of the wine industry and the set of destinations I use for implementing my estimation. Then, I describe the cleaning procedure implemented on the wine producers and provide summary statistics on the final sample of firms used in the estimation.

B.1 Wine industry

The decision to implement this estimation on wine exporters relies on two constraints. First of all, I study the entry decision made at the firm level. This level of analysis is explained by the fact that brands and reputation are often defined by the firm that produces the good. Therefore, this requires to study firms that display a small level of heterogeneity in terms of goods. A car producer for instance, that also exports car pieces, or engines for other vehicles, is difficult to analyze as a single-product firm. However, a wine producer mostly export wines, and specifically bottles of wine, whose prices are easy to define, and aggregate at the firm level. For these reasons when defining my sample, I exclusively use wine producers that do not export any other goods outside of wine. A large share of the trade in wine is made by wholesalers who export other types of items, and for which the study at the level of the firm is irrelevant. In addition to this homogeneity constraint, my estimation procedure requires enough firms which export to several destinations. As a major exporting industry from France, the wine industry meets both of these conditions: a large number of exporters, exporting a precisely defined good.

[69] Since quantities are sometimes missing, I compute an average price rather than computing the price from the ratio of yearly values and quantities.

In addition to imposing restrictions on the set of firms included in the final sample, I only use a restricted set of destinations.

B.2 Selection of destinations

I select 15 different destinations on which I analyze the behaviors of French exporters. These destinations have been selected among the 20 most popular destinations for wine exports from France, excluding countries with large import/export platforms such as Denmark and Singapore, while reflecting some heterogeneity in terms of location. Moreover, I divide these destinations in three groups, for which I estimate different entry and fixed costs of exporting.. The list of these destinations can be found in table 4.

TABLE 4: List of destination countries included in the structural sample

	Group 1 *Europe*		**Group 2** *Americas*	**Group 3** *Asia/Oceania*
Great-Britain	Germany	Belgium	(Brazil)	Australia
Netherlands	Italy	Spain	Canada	China
Ireland	Sweden	Switzerland	United States	Japan

Note that I do not include Brazil in the estimation sample. The observations related to this destination are used in the out-of-sample exercise and are excluded so that it does not affect the estimation procedure.

B.3 Aggregation

Because the estimation is conducted at the firm-destination-year level, it is necessary to aggregate the sales and quantities exported across products exported by the firm. The choice of the wine industry is crucial here since bottles of wines are quantities that can be easily aggregated. An industry producing differentiated goods would have made this aggregation less straightforward. The aggregation of prices and sales are the following:

$$
\begin{aligned}
p_{fdt} &= \sum_{h=1}^{H_{fdt}} w_{fhdt} \frac{s_{fhdt}}{q_{fhdt}} \qquad \text{with} \quad w_{fhdt} \equiv \frac{s_{fhdt}}{\sum_h s_{fhdt}} \\
s_{fdt} &= \sum_{h=1}^{H_{fdt}} s_{fhdt}
\end{aligned}
$$

where H_{fdt} is the number of 8-digit observations for each firm-destination-year triplet. Moreover, note that there is a certain number of missing quantities in the data. Therefore, I assign a weight w_{fhdt} equal to zero to the observations that have quantities or values exported equal to one or zero. When this observation is the only one at the firm-destination-year level (no other product is sent to this market by this firm this year), I dropped all the observations related to this firm from the sample.

B.4 Partial-year bias

Similar to the sample used in the reduced form exercise, I will correct for the partial-year bias, by redefining the entry months of all entering exporters. As a consequence, I shift all the subsequent flows to maintain the same sequence in the exports of the firm. Therefore, exports during the first year will look similar to the subsequent years of exporting.

B.5 Cleaning

I clean the data to avoid the potential existence of outliers in prices. In particular, I exclude observations that display extreme prices along two dimensions. First, I flag observations which log difference is larger than two, or lower than -2 relative to the previous year. Second, I flag prices which log value is larger than two or lower than two relative to the average price of the firm-year pair. After having flagged those observations, I dropped all observations of a firm that contains at least one flagged observation.

Finally, a last criterion for a firm to be included in the final sample is based on the number of observations. Many firms export one year to one market during the sample period, and this does not provide enough information to analyze their exporting behavior. Therefore, I only keep firms that recorded at least 10 exporting events, and have exporting activity in at least two destinations. Note that with 14 destinations and 14 years of data, the maximum number of observations by a given firm is 196. This selection process could present a problem as it is likely to affect the estimates of entry and fixed costs of exporting, by only looking at successful firms. However, this procedure will tend to select firms that survive several years, rather than short-lived exporters: as a consequence, it tends to go against the theory of consumer accumulation that can accommodate small and short-lived exporters relative to the standard model. Finally, I only keep firms that have some exporting activity to Brazil during the sample period.

B.6 Final sample

Once these cleaning steps were implemented, I obtain a sample of 236 firms. The following tables present summary statistics regarding this sample.

TABLE 5: Description of the sample used in the structural estimation

Statistics:	*pc5*	*median*	*pc95*	*mean*	N
# observations per firm	11	29	107	41.1	236
av. # destinations per firm-year	1.4	3.16	8.36	3.92	2 394
av. # years per firm-destination	2.1	4.77	9.09	5.02	1 836

Table 5 provides information regarding the number of observations provided by the sampled firms, as well as the number of destinations they export to in an average year. One can see that the firms selected are relatively large, with a minimum number of export episodes equal to 10 by the sampling procedure. However, the median firm only records 29 export episodes, while the maximum number of episodes in the dataset is 196 (14×14). Moreover, they are relatively diversified in terms of destinations since the median firm exports to 3.16 destinations in an average year.

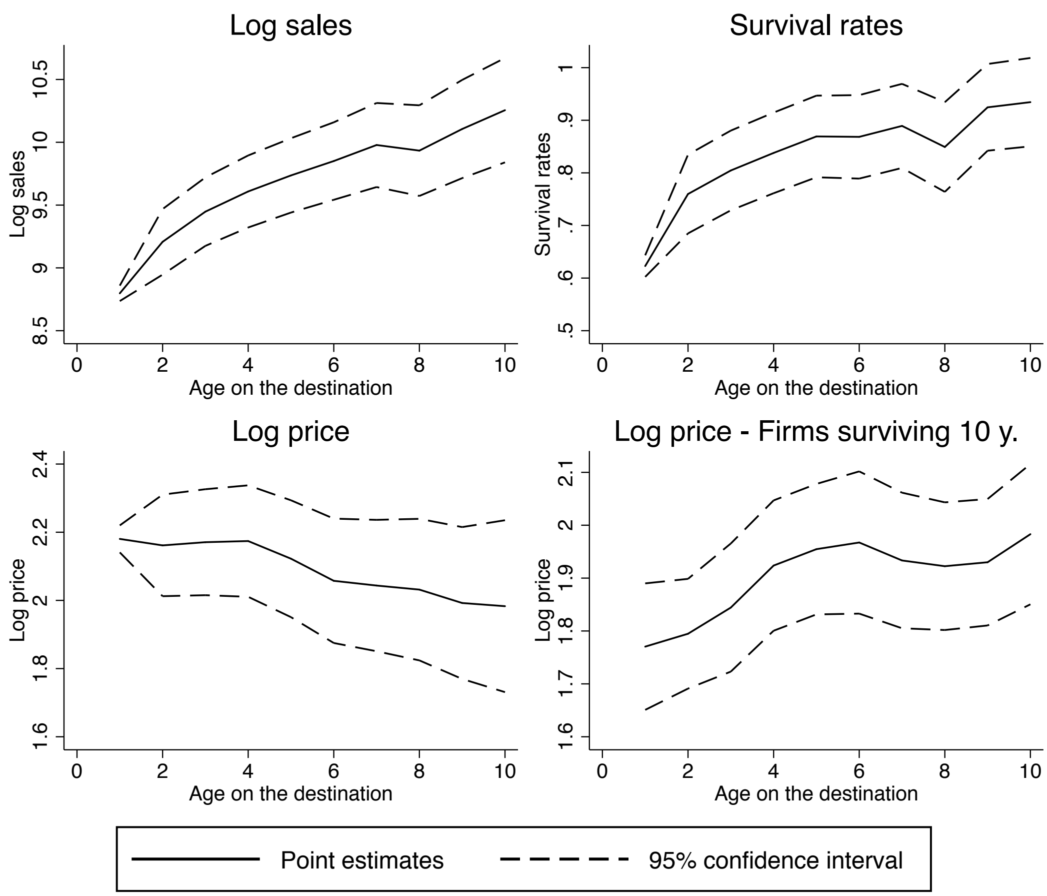

FIGURE 12: Sales, prices and survival rates across ages (Wine producers)

Notes: The figure reports the average log sales, log prices and survival rates of wine producers in a destination at different ages. The estimates are obtained from the regression of these dependent variables on a set of age dummies. The age in a destination is defined as the number of years a firm has been successively exporting to this country. 95 percent confidence intervals are constructed using standard errors estimates clustered at the firm-destination level.

In order to inspect how this sampling procedure affects the trajectories of the exporters, I replicate the regressions on age dummies I perform in section I. Figure 12 reports the results of these regressions for sales, prices and survival rates. The patterns of sales and prices are very similar to the ones observed using the comprehensive sample: sales appear to increase in the early years, with the an average growth rate of 30 percent the first year. However, we can see that the survival rates in the structural sample are larger than the ones displayed in the exhaustive data. While the survival rate was close to 45 percent in the full sample, it is around 60 percent in this restricted sample. This arises because of the requirement made during the selection of exporters: because the estimation procedure requires firms with several observations, this tends to eliminate firms with very large attrition rates that do not records many episodes of exporting activity. Note that this difference in survival rates between exhaustive and restricted samples will play against the story I develop in this paper. Large attrition rates will be consistent with a story that emphasizes strong dependence in demand rather than an important role for sunk costs of entry. Finally, prices tend to decrease with ages when we do not control for selection. However, they are increasing when we restrict the sample to firms surviving 10 years.

B Additional age regressions

In this section, I provide additional results and describe alternative specifications to look at the correlation between sales or prices and age in an export market.

A Additional specification

The specification shown in the main text of the paper uses variations within firm across destinations, to identify sales and prices dynamics. However, another source of identification to capture these dynamics would be to identify trends within specific exporting spells. More specifically, a natural specification would include firm-destination-product-spell and destination-product-year fixed effects to capture the heterogeneity across firms, products and markets. However, including this set of fixed effects makes it impossible to identify a trend in prices or sales across ages: it can capture deviations from the trend, but not the trend itself. To understand why, consider a sample of firms on a given market *pdt*. Because of the market-level fixed effect, their average price is normalized to zero. Now consider the same set of firms one year later, when all firms aged by one year. The market-level fixed effect means that their average price will be normalized to zero as well. Therefore, a trend in prices cannot be identified. Intuitively, the fact that all firms are getting one year older every year implies the absence of control groups. Therefore, this specification does not allow to test for a trend in sales or prices, but can identify for which ages the growth is more pronounced.

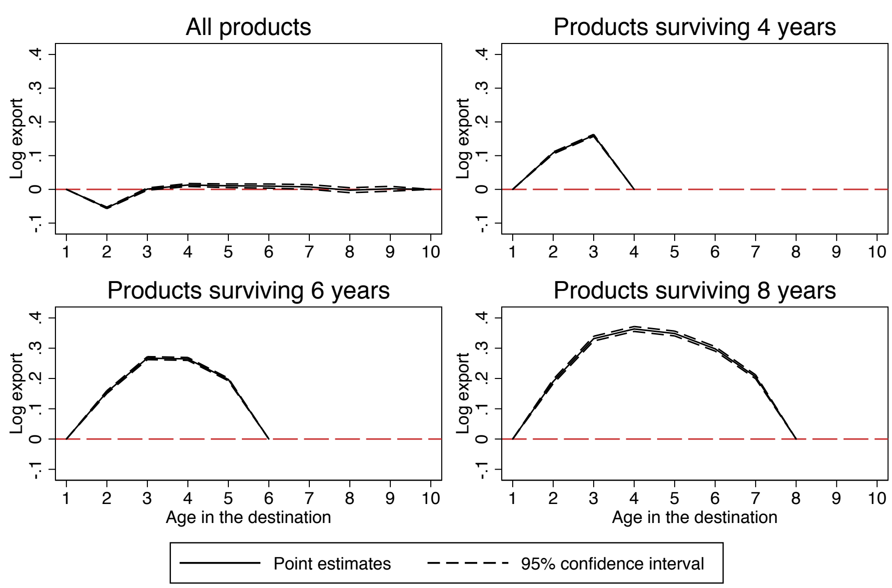

FIGURE 13: Sales across export ages, within variation

Notes: The figure reports the cumulative growth of sales compared to age one, of a firm-product category pair in a destination at different ages. The regression uses logarithm of sales as dependent variable, and includes product category×destination×year and firm×product category×destination×spell fixed effects. 95 percent confidence intervals are constructed using standard errors clustered at the firm-product-destination-spell level.

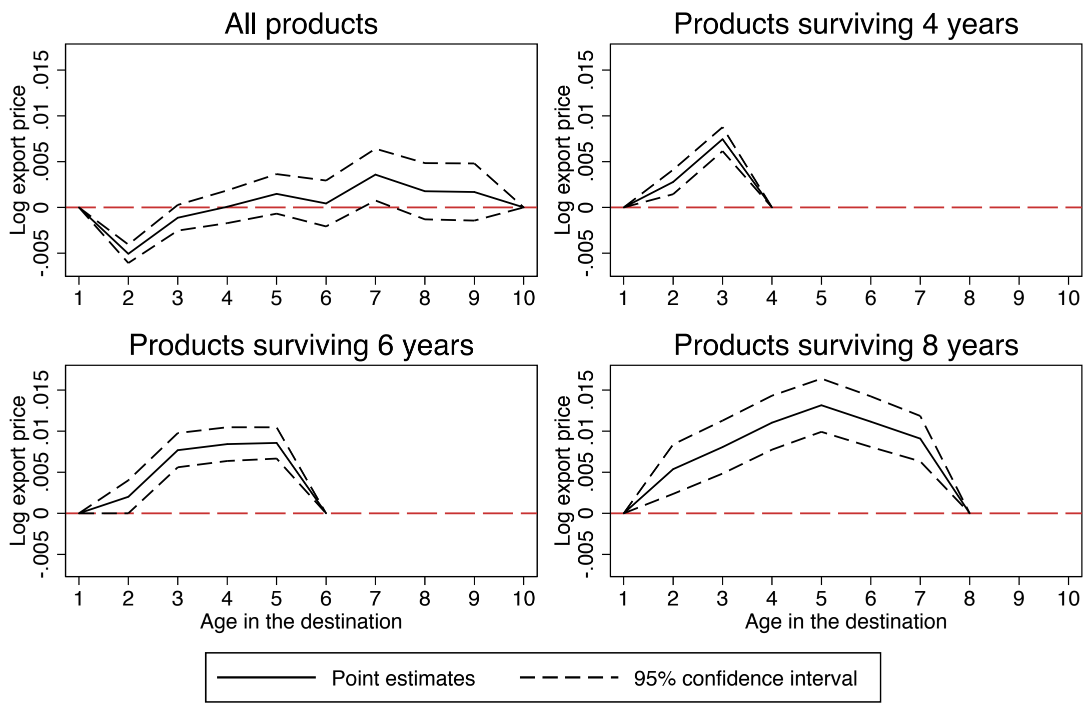

FIGURE 14: Prices across export ages, within variation

Notes: The figure reports the cumulative growth of prices compared to age one, of a firm-product category pair in a destination at different ages. The regression uses logarithm of price as dependent variable, and includes product category×destination×year and firm×product category×destination×spell fixed effects. 95 percent confidence intervals are constructed using standard errors clustered at the firm-product-destination-spell level.

Figures 13 and 14 report the results of this specification for sales and prices. As we can see, even sales are not increasing with age in this specification: in fact, the coefficients are negative for young firms, which means that sales increase less between years 1 and 2. Similar patterns appear for prices. However, regressions including a constant set of firms are more informative. We can see that both prices and sales tend to increase faster in the first years of exporting, which is consistent with the findings from the two main specifications used in the main text.

B Additional tables

I summarize regression results in tables 6 and 7. In addition to the specification displayed in the main text (columns (2), (4) and (8), I report coefficients using the identification within-spell highlighted above (columns (5) and (9)), and replicate the result found in Berman et al. (Forth.) using French custom data (columns (6) and (10) of table 6). Moreover, this table also shows that the trends displayed in the main text also hold in the cross-section of firms (columns (1), (3) and (7)).

TABLE 6: Age regressions (full sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Survival		Log sales				Log price			
age 2	0.24***	0.15***	0.59***	0.39***	-0.055***	0.42***	0.022***	0.012***	-0.0051***	-0.014***
	(0.0003)	(0.0005)	(0.001)	(0.002)	(0.001)	(0.002)	(0.0006)	(0.0008)	(0.0005)	(0.0007)
age 3	0.33***	0.23***	1.02***	0.76***	0.00090	0.83***	0.038***	0.025***	-0.0011	-0.015***
	(0.0004)	(0.0006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.0009)	(0.0010)	(0.0007)	(0.0009)
age 4	0.38***	0.27***	1.32***	1.03***	0.013***	1.12***	0.048***	0.034***	0.000069	-0.016***
	(0.0005)	(0.0007)	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.0009)	(0.001)
age 5	0.40***	0.29***	1.54***	1.23***	0.011***	1.35***	0.057***	0.039***	0.0015	-0.017***
	(0.0006)	(0.0008)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
age 6	0.42***	0.30***	1.74***	1.40***	0.0098**	1.54***	0.059***	0.042***	0.00043	-0.019***
	(0.0007)	(0.0009)	(0.003)	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)
age 7	0.44***	0.30***	1.92***	1.55***	0.0073*	1.72***	0.068***	0.044***	0.0036*	-0.022***
	(0.0008)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)	(0.002)
age 8	0.45***	0.31***	2.08***	1.67***	-0.0025	1.89***	0.072***	0.041***	0.0018	-0.025***
	(0.0009)	(0.001)	(0.005)	(0.005)	(0.004)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
age 9	0.46***	0.31***	2.22***	1.78***	0.0018	2.02***	0.072***	0.043***	0.0017	-0.027***
	(0.001)	(0.001)	(0.006)	(0.006)	(0.004)	(0.006)	(0.003)	(0.002)	(0.002)	(0.002)
age 10	0.47***	0.32***	2.33***	1.88***	.	2.14***	0.074***	0.041***	.	-0.031***
	(0.001)	(0.002)	(0.007)	(0.007)		(0.007)	(0.003)	(0.003)		(0.003)
N	17 503 978	11 543 384	19 311 381	12 871 145	10 780 577	14 484 211	19 311 381	12 871 145	10780577	14484211
R^2	0.29	0.59	0.37	0.74	0.86	0.66	0.78	0.92	0.96	0.90
Dest-Prod-Year FE	Y	Y	Y	Y	Y	N	Y	Y	Y	N
Firm-Prod-Year FE	N	Y	N	Y	N	Y	N	Y	N	Y
Firm-Prod-Dest-Spell FE	N	N	N	N	Y	N	N	N	Y	N

Note: * p<0.05, ** p<0.01, *** p<0.001

Table 7: Age regressions (sample of surviving firms)

	(1)	(2) Log sales	(3)	(4)	(5) Log price	(6)
Products surviving 4 years						
age 2	0.21***	0.31***	0.11***	0.0024**	0.0026	0.0028***
	(0.002)	(0.005)	(0.002)	(0.0009)	(0.002)	(0.0007)
age 3	0.36***	0.61***	0.16***	0.0067***	0.012***	0.0075***
	(0.003)	(0.005)	(0.001)	(0.001)	(0.002)	(0.0007)
age 4	0.29***	0.75***	.	-0.00057	0.014***	.
	(0.004)	(0.006)		(0.002)	(0.002)	
N	3 886 412	2 497 364	3 851 329	3 886 412	2 497 364	3 851 329
R^2	0.37	0.79	0.88	0.81	0.95	0.96
Products surviving 6 years						
age 2	0.27***	0.35***	0.16***	0.0020	0.0011	0.0020*
	(0.003)	(0.007)	(0.002)	(0.001)	(0.003)	(0.001)
age 3	0.48***	0.68***	0.27***	0.0071***	0.0070*	0.0077***
	(0.004)	(0.008)	(0.003)	(0.002)	(0.003)	(0.001)
age 4	0.59***	0.88***	0.26***	0.0072***	0.013***	0.0084***
	(0.005)	(0.008)	(0.002)	(0.002)	(0.003)	(0.001)
age 5	0.63***	1.02***	0.20***	0.0061*	0.012***	0.0086***
	(0.006)	(0.009)	(0.002)	(0.003)	(0.003)	(0.0010)
age 6	0.55***	1.11***	.	-0.0043	0.013***	.
	(0.007)	(0.009)		(0.003)	(0.003)	
N	2 686 575	1 763 426	2 678 763	2 686 575	1 763 426	2 678 763
R^2	0.37	0.80	0.86	0.81	0.95	0.96
Products surviving 8 years						
age 2	0.28***	0.39***	0.19***	0.0048*	0.0083	0.0054***
	(0.005)	(0.01)	(0.004)	(0.002)	(0.004)	(0.002)
age 3	0.51***	0.75***	0.33***	0.0075**	0.0099*	0.0081***
	(0.007)	(0.01)	(0.004)	(0.003)	(0.005)	(0.002)
age 4	0.63***	0.98***	0.36***	0.0093**	0.018***	0.011***
	(0.009)	(0.01)	(0.004)	(0.004)	(0.005)	(0.002)
age 5e	0.70***	1.15***	0.35***	0.0100*	0.016**	0.013***
	(0.01)	(0.01)	(0.004)	(0.004)	(0.005)	(0.002)
age 6	0.73***	1.31***	0.30***	0.0058	0.017**	0.011***
	(0.01)	(0.01)	(0.004)	(0.005)	(0.005)	(0.002)
age 7	0.72***	1.41***	0.21***	0.00020	0.016**	0.0091***
	(0.02)	(0.02)	(0.003)	(0.006)	(0.006)	(0.001)
age 8	0.60***	1.50***	.	-0.014	0.016**	.
	(0.02)	(0.02)		(0.007)	(0.006)	
N	1 690 783	1 073 287	1 690 783	1 690 783	1 073 287	1 690 783
R^2	0.38	0.82	0.86	0.81	0.95	0.96
Firm-Prod-Year FE	N	Y	N	N	Y	N
Firm-Prod-Dest-Spell FE	N	N	Y	N	N	Y

Note: Year x product x destinations included in all regressions. * p<0.05, ** p<0.01, *** p<0.001

C Derivations

Derivation of the law of motion of $n_{t+1}(s_t, n_t)$

Starting from the system of differential equations:

$$\frac{\partial n_{t+1}}{\partial s_t} = \eta_1(1-n_{t+1})^{\psi},$$
$$\frac{\partial n_{t+1}}{\partial n_t} = \eta_2(1-n_{t+1})^{\psi}.$$

We can rewrite this system as

$$\frac{1}{\eta_1}(1-n_{t+1})^{-\psi}\partial n_{t+1} = \partial s_t$$
$$\frac{1}{\eta_2}(1-n_{t+1})^{-\psi}\partial n_{t+1} = \partial n_t$$

Integrating on both sides, we have

$$-\frac{(1-n_{t+1})^{1-\psi}}{(1-\psi)\eta_1} = s_t + C_1(n_t)$$
$$-\frac{(1-n_{t+1})^{1-\psi}}{(1-\psi)\eta_2} = n_t + C_2(s_t)$$

where $C_1(n_t)$ and $C_2(s_t)$ are some functions of n_t and s_t. This system of equations can be rewritten wlog:

$$n_{t+1}(s_t, n_t) = 1 - (C_1(n_t) - \eta_1(1-\psi)s_t)^{\frac{1}{1-\psi}}$$
$$n_{t+1}(s_t, n_t) = 1 - (C_2(s_t) - \eta_2(1-\psi)n_t)^{\frac{1}{1-\psi}}$$

Imposing the initial condition $n_{t+1}(0,0) = 0$, we find the solution of these differential equations

$$n_{t+1}(s_t, n_t) = 1 - (1 - \eta_1(1-\psi)s_t - \eta_2(1-\psi)n_t)^{\frac{1}{1-\psi}}.$$

Optimal markup

The firm chooses the optimal markup μ to maximize the value of exporting:

$$\begin{aligned}\mu &= \operatorname{argmax} V_I(\xi, n, \mu) \\ &= \operatorname{argmax} E_\varepsilon\Big\{\pi(\xi, n, \mu, \varepsilon) + \beta EV'(\xi, n'(\xi, n, \varepsilon, \mu), 1)\Big\} \\ &= \operatorname{argmax} \int_\varepsilon \pi(\xi, n, \mu, \varepsilon) + \beta EV'(\xi, n'(\xi, n, \varepsilon, \mu), 1) dF(\varepsilon)\end{aligned}$$

Therefore, the first order condition of the problem is

$$\int_\varepsilon \frac{\partial \pi(\xi, n, \mu, \varepsilon)}{\partial \mu} + \beta\frac{\partial EV'(\xi, n'(\xi, n, \varepsilon, \mu), 1)}{\partial \mu} dF(\varepsilon) = 0$$

First, profit function is

$$\pi(\xi,n,\mu,\varepsilon) = n\exp(\lambda+X+\varepsilon^D)\mu^{-\sigma}(\mu-1)c(\xi,n,\varepsilon^S)^{1-\sigma}$$
$$\Rightarrow \frac{\partial\pi(\xi,n,\mu,\varepsilon)}{\partial\mu} = \left[(1-\sigma)\mu^{-\sigma}+\sigma\mu^{-\sigma-1}\right] n\exp(\lambda+X+\varepsilon^D)c(\xi,n,\varepsilon^S)^{1-\sigma}$$

Second, the continuation value can be rewritten $EV'(\xi,n'(\xi,n,\varepsilon,\mu),1) = EV'(\xi,n'(s,n),1)$ where s are the sales of the firm. Therefore,

$$\begin{aligned}\frac{\partial EV'(\xi,n'(\xi,n,\varepsilon,\mu),1)}{\partial\mu} &= \frac{\partial EV'(\xi,n'(s,n),1)}{\partial\mu}\\ &= \frac{\partial s}{\partial\mu}\frac{\partial n'}{\partial s}\frac{\partial EV'(\xi,n',1)}{\partial n'}\\ &= (1-\sigma)\mu^{-\sigma}n\exp(\lambda+X+\varepsilon^D)c(\xi,n,\varepsilon^S)^{1-\sigma}\frac{\partial n'}{\partial s}\frac{\partial EV'(\xi,n',1)}{\partial n'}\end{aligned}$$

Therefore, the first order condition can be rewritten

$$\begin{aligned}&\int_\varepsilon n\exp(\lambda+X+\varepsilon^D)c(\xi,n,\varepsilon^S)^{1-\sigma}\Big[(1-\sigma)\mu^{-\sigma}+\sigma\mu^{-\sigma-1}\\ &\qquad\qquad +(1-\sigma)\mu^{-\sigma}\beta\frac{\partial n'}{\partial s}\frac{\partial EV'(\xi,n',1)}{\partial n'}\Big]dF(\varepsilon)=0\\ \Leftrightarrow\quad &\int_\varepsilon \exp(\varepsilon^D+(1-\sigma)\varepsilon^S)\Big[\sigma+(1-\sigma)\mu\Big(1+\beta\frac{\partial n'}{\partial s}\frac{\partial EV'(\xi,n',1)}{\partial n'}\Big)\Big]dF(\varepsilon)=0\\ \Leftrightarrow\quad &\mu(\sigma-1)\int_\varepsilon \exp(\varepsilon^D+(1-\sigma)\varepsilon^S)\Big(1+\beta\frac{\partial n'}{\partial s}\frac{\partial EV'(\xi,n',1)}{\partial n'}\Big)dF(\varepsilon)\\ &\qquad\qquad =\sigma\int_\varepsilon \exp(\varepsilon^D+(1-\sigma)\varepsilon^S)dF(\varepsilon)\\ \Leftrightarrow\quad &\mu=\frac{\sigma}{\sigma-1}\frac{\int_\varepsilon \exp(\varepsilon^D+(1-\sigma)\varepsilon^S)dF(\varepsilon)}{\int_\varepsilon \exp(\varepsilon^D+(1-\sigma)\varepsilon^S)\Big(1+\beta\frac{\partial n'}{\partial s}\frac{\partial EV'(\xi,n',1)}{\partial n'}\Big)dF(\varepsilon)}\\ \Leftrightarrow\quad &\mu=\frac{\sigma}{\sigma-1}\frac{1}{\int_\varepsilon \omega(\varepsilon)\Big(1+\beta\frac{\partial n'}{\partial s}\frac{\partial EV'(\xi,n',1)}{\partial n'}\Big)dF(\varepsilon)}\end{aligned}$$

with $\omega(\varepsilon)=\frac{\exp(\varepsilon^D+(1-\sigma)\varepsilon^S)}{\int_\varepsilon \exp(\varepsilon^D+(1-\sigma)\varepsilon^S)dF(\varepsilon)}$.

D Estimation method

I describe in this section of the appendix the MCMC algorithm I implement. I start by describing how the Markov chain is initialized, before describing a given iteration of the chain, involving the update of the unobservables and parameters.

A Initial values

I start by describing how the unobservables are obtained, before describing the initial parameters. From the value of the price elasticity of demand, I can obtain $\log s_{fdt}+\sigma p_{fdt} = \log n_{fdt}+X_{dt}+\lambda_{ft}$. I can then decompose this term using firm-year and destination-year fixed effect to obtain $\lambda_{ft}^{(0)}$ and $X_{dt}^{(0)}$. In order to obtain $\phi_{ft}^{(0)}$, I run the regression $\log p_{fdt} - \frac{\sigma}{\sigma-1}$ on $\lambda_{ft}^{(0)}$. This allows me to obtain $\alpha^{(0)}$, and the residual is regressed on firm-year fixed effects to obtain $\phi_{ft}^{(0)}$. Having in hand initial values for the unobservables, I can use linear regressions to obtain the AR(1) coefficients for the unobservables, and use nonlinear least square to estimate $\underline{n}^{(0)}$, $n_0^{(0)}$, $\eta_1^{(0)}$ and $\eta_2^{(0)}$ after arbitrarily setting $\psi^{(0)} = 0.5$. Finally, I set values for the fixed costs parameters, and the variance parameter of the fixed cost shocks. I arbitrary set $f^{(0)} = s_v^{(0)} = delta^{(0)} = 10\,000$ and $f^{(0)} = 30\,000$ for the three different groups of countries.

After setting these initial values, I implement 500 iterations that does not account for the dynamic problem of the firm. Therefore, I sample unobservables and parameters assuming a constant mark-up and only taking advantage of the realized sales and prices. This step allows me to obtain initial conditions for the parameters and unobservables that are closer to their true values, although biased because they do not account for the dynamic problem.

Given this initial set of parameters and unobservables, I can start the iterative procedure described below.

B Creation of the grid

In order to solve for the value function as a function of Θ, I need to create a grid describing the state space of the problem. Note that the state space is made of (λ, ϕ, n, X). Consequently, I need a grid that is relatively more precise for values of the unobservables that are more prevalent. Consequently, I create the four-dimensional grid as following

- $\lambda_g \sim N(0, 3\,\text{std}(\lambda_{ft}))$
- $\phi_g \sim N(\text{mean}(\phi_{ft}), 3\,\text{std}(\phi_{ft}))$
- $X_g \sim N(\text{mean}(X_{ft}), 3\,\text{std}(X_{ft}))$
- $n_g \sim U[0\ ;\ 1]$

Note that this grid is updated when the standard deviations or averages of the current unobservables are 20 percent larger or smaller than the ones used for the current grid, such that the grid will follow the potential change in the distribution of the unobservables. I set the size of the grid to be 20 on each dimension, such that the value function will be iterated at 20^4 different grid points.

Moreover, in order to solve the optimal mark-up of the firm, I also need to specify a set of grid points for the optimal mark-up term. I create a set of grid points mk_g of size $g_m = 20$, such that $\mu = \frac{\sigma}{\sigma-1}\frac{1}{mk_g}$, with $mk_g \equiv \{1, 1.05, 1.1, ..., 2\}$.

C Iteration

Three different objects are updated at each iteration of the Markov Chain:

- the value function $V^{(s)}(\Theta^{(s)})$,
- the set of unobservables $\xi^{(s)}_{fdt} = (\lambda^{(s)}_{ft}, \phi^{(s)}_{ft}, X^{(s)}_{dt})$,
- the parameter vector $\Theta^{(s)}$.

I perform 60,000 iterations of the Markov chain, discarding the first 30,000 iterations. In the next paragraphs, I describe each of these following steps. I start by describing the step that aims to compute the value functions since they define objects that are used in the other steps. I then turn to the sampling of unobservables, and the sampling of parameters.

Update of the value function The value functions are obtained from the Bellman equation, iterated from the previous iteration of the value functions. From section II, we have

$$V_I(\xi, n) = \max_{\mu}\left\{E_\varepsilon\left\{\pi(\xi, n, \mu, \varepsilon) + \beta EV'(\xi, n'(\xi, n, \varepsilon, \mu), 1)\right\}\right\}$$

Therefore, the value function is updated the following way:

$$\begin{aligned} V^{(s+1)}(\xi_g, n_g, \mathcal{I}, \Theta^{(s)}) = s_v^{(s)} \log \Bigg[&\exp\left(\frac{1}{s_v^{(s)}} EV_O(\xi_g)\right) \\ &+ \exp\left(\frac{1}{s_v^{(s)}} \max_{mk \in mk_g} \left\{E_\varepsilon \pi(\xi_g, n_g, mk_g, \Theta^{(s)}) - f^{(s)} + EV_I(\xi_g, mk_g)\right\}\right)\Bigg] \end{aligned} \tag{15}$$

with

$$EV_I(\xi_g, mk_g) = \frac{\sum_{\xi \in \xi_g} \sum_{n \in n_g} V^{(s)}(\xi, n, I, \Theta^{(s)})\, P_n(n \,|\, \xi_g, mk_g)\, P_\xi(\xi \,|\, \xi_g)}{\sum_{\xi \in \xi_g} \sum_{n \in n_g} P_n(n \,|\, \xi_g, mk_g)\, P_\xi(\xi \,|\, \xi_g)},$$

$$EV_O(\xi_g) = \frac{\sum_{\xi \in \xi_g} V^{(s)}(\xi, n_0, 0, \Theta^{(s)})\, P_\xi(\xi \,|\, \xi_g)}{\sum_{\xi \in \xi_g} P_\xi(\xi \,|\, \xi_g)}$$

$P_\xi(.|.)$ being the transition probability of the unobservables at the current parameters, and $P_n(n \,|\, \xi_g, mk_g)$ the probability of obtaining a share n in the next period given the current unobservables ξ_g and the mark-up decision mk_g.[70] In practice, I iterate several times the Bellman equation, in order to reduce the error coming from the use of the previous value functions. In this case, I iterate not using the (s)-th value function anymore, but the current value function.

[70]This probability is obtained from the shock ε that makes the sales of the firms, and therefore the future share of consumers, non-deterministic.

In addition to updating the value function, I define, during this iteration, two objects that will be used in the sampling of parameters and unobservables. First, I save the optimal mark-up chosen by the firm. This object, evaluated on the grid, is defined as

$$mk_g^* \equiv \text{argmax}\Big\{E_\varepsilon\Big\{\pi(\xi, n, mk, \varepsilon) + \beta EV'(\xi, n'(\xi, n, \varepsilon, mk), 1)\Big\}\Big\}.$$

Second, I create the difference in expected value functions, $DEV()$, that is defined as

$$DEV(\xi_g, n_g) = EV_I(\xi_g, mk_g^*) - EV_O(\xi_g).$$

This object will be convenient when computing the difference in value functions for each firm.

Sampling of unobservables I sample unobservables using the particle Gibbs with ancestor sampling (PGAS) sampler described in Lindsten et al. (2014), which relies heavily on the particle MCMC introduced in Andrieu, Doucet, and Holenstein (2010). The idea of particle MCMC is to develop techniques that use particle filtering in MCMC algorithm. Specifically, the PGAS iteratively sample parameters conditional to a particle and update that particle conditional to parameters. Importantly, the particle filter allows the sampling of the particle, proportional to the likelihood function, with good mixing properties. An important point of this sampler is that the current unobservables need to survive all the resampling steps of the filter. Moreover, I use ancestor sampling when choosing the specific set of particles, in order to further improve the mixing of the sampler.

To further describe the sampling of unobservables, I take the example the sampling of the unobservables λ_{ft} and ϕ_{ft}, conditional to the current unobservables at iteration s, $X_{dt}^{(s)}$. These unobservables are proportional to their prior distribution, $F_\lambda()$ and $F_\phi()$, and the conditional likelihood $L(D_{fdt}|D_{fdt-1}, \lambda_{ft}, \phi_{ft}, X_{dt}^{(s)})$. The steps are the following:

- Starting from period 0, and for each firm, I generate r=1..20 particles $(\lambda_{f0}^r, \phi_{f0}^r)$ from their prior distribution $F_\lambda()$ and $F_\phi()$.
- I compute the likelihood of each of these particles for each firm:

  $$L_{f0}^r \equiv \prod_d L(D_{fd0}|D_{fd-1}, \lambda_{f0}^r, \phi_{f0}^r, X_{d0}^{(s)})$$

  using extrapolations from the functions $DEV(\xi_g, n_g)$ and $mk^*(\xi_g, n_g)$ to obtain the difference in value functions and the mark-up necessary to compute the likelihood.
- for each period t=1..T:
  - I resample 20 $(\lambda_{ft-1}^r, \phi_{ft-1}^r)$ proportionally to L_{ft-1}^r and replace $(\lambda_{ft-1}^{20}, \phi_{ft-1}^{20})$ by $(\lambda_{ft-1}^{(s)}, \phi_{ft-1}^{(s)})$.
  - generate 20 new particles, for each firm, from the prior distribution based on the resampled $(\lambda_{ft-1}^r, \phi_{ft-1}^r)$.
  - I compute the particle-specific likelihood $L_{ft}^r \equiv \prod_d L(D_{fdt}|D_{fdt-1}, \lambda_{ft}^r, \phi_{ft}^r, X_{dt}^{(s)})$

- I retain one specific particle $(\lambda_{ft}^{(s+1)}, \phi_{ft}^{(s+1)})$ by using the ancestor sampling from period T to period 0.
  - Sample one particle for each firm $(\lambda_{fT}^{(s+1)}, \phi_{fT}^{(s+1)})$ proportionally to L_{fT}^{r}
  - for each period t=T-1..0:
    - sample $(\lambda_{ft}^{(s+1)}, \phi_{ft}^{(s+1)})$ proportionally to $\frac{L_{ft}^{r}\, F(\phi_{ft+1}^{(s+1)}|\phi_{ft}^{r})\, F(\lambda_{ft+1}^{(s+1)}|\lambda_{ft}^{r})}{\sum_r L_{ft}^{r}\, F(\phi_{ft+1}^{(s+1)}|\phi_{ft}^{r})\, F(\lambda_{ft+1}^{(s+1)}|\lambda_{ft}^{r})}$

This procedure gives me a new set of unobservables $(\lambda_{ft}^{(s+1)}, \phi_{ft}^{(s+1)})$ that have sampled proportionally to the likelihood function. Then, I perform a similar procedure to sample $X_{dt}^{(s+1)}$ conditional to the parameters and unobservables.

Sampling of parameters The sampling of parameters is made more complicated by the fact that functions $DEV()$ and $mk()$ need to be reevaluated for a new Θ. Therefore, sampling parameters requires to perform a Metropolis-Hastings step for which it is necessary to iterate the value functions for this new parameter Θ, similarly to the step updating the value functions. Formally, the sampling of a given block of parameter Θ takes the following steps:

- A new parameter Θ^* is drawn using a proposal function.
- The value function $V(\xi_g, n_g, I, \Theta^*)$ is obtained from equation (15) and the functions $DEV(\xi_g, n_g)$ and $mk_g(\xi_g, n_g)$ are computed.
- I obtain by interpolation DV_{fdt} and μ_{fdt}, allowing me to compute the likelihood function for the parameter Θ^*.
- $\Theta^{(s+1)}$ is set to be Θ^* with probability $\max\left\{1, \frac{\prod_t \prod_d \prod_f L_{fdt}(D, \xi_{fdt}^{(s+1)}; \Theta^*)}{\prod_t \prod_d \prod_f L_{fdt}(D, \xi_{fdt}^{(s+1)}; \Theta^{(s)})}\right\}$.

All parameters of the model, with the exception of the initial mean and variance of X, enter the dynamic problem of the firm. Therefore, all parameters enter the value function and should be evaluate using this Metropolis-Hastings step. However, because this step is relative time-consuming in the algorithm, I decide to only run the full Metropolis-Hastings step for some parameter blocks. For others, that are less likely to play a significant role in the dynamic problem, I use a simple Gibbs sampler that relies on specific parts of the likelihood function. Specifically, I update the different blocks of parameters as following:

- Parameters from the supply equation (α, γ_1,γ_2) are obtained from a Gibbs sampler based on the Bayesian regression of prices on quality λ_{ft} and country group dummies.
- The parameters of the variance matrix of demand and supply shocks (Σ_{11}, Σ_{12}, Σ_{22}) are sampled from an inverse Wishart distribution, based on the demand and supply residuals.
- The parameters of the AR(1) processes (ρ_λ, σ_λ, μ_ϕ, ρ_ϕ, σ_ϕ, μ_X, ρ_X, σ_X, μ_{X_0}, σ_{X_0}) are directly sampled from the Bayesian regression of the unobservables on their lags.
- The fixed costs parameters, variances of fixed costs and exit rate (f_1^c, f_2^c, f_3^c, f_1^e, f_2^e, f_3^e, σ_ν^c, σ_ν^e, δ) are sampled from the full Metropolis Hastings step as described above, using a random walk proposal function that targets an acceptance rate of 0.2.

- The parameters of the law of motion of n (η_1, η_2, n_0, ψ) are sampled from the full Metropolis Hastings step as described above, using a random walk proposal function that targets an acceptance rate of 0.2.

Overall, this procedure is doable thanks to parallelization using GPU computing. On average, an iteration of the Markov chain takes a bit more than two seconds, which implies a total computing time of less than two days for 60 000 iterations.

D Estimation on simulated data

To test my empirical procedure, I simulate a set of data following the data generating process assumed in the model. Then, I implement my estimation procedure to test the validity of the estimation. However, because of the complexity of the estimation, I cannot perform a full Monte Carlo study of the estimation method. Therefore, I cannot test whether my estimator consistently recovers the true value of the parameters, but instead whether the true value of the parameters belongs to the confidence interval obtained from the estimation. I simulate data for 200 firms, 14 years and 14 destinations and run 60 000 iterations of my algorithm, as I do in the estimation procedure. Moreover, in order to evaluate the errors made through the interpolation procedure, I simulate data using a grid of 30 points, while only 20 points are used in the estimation. I report in figures 15 the Markov chains for all parameters, as well as the true value of the parameters displayed by the red lines. As displayed on these figures, the estimation provides posterior distribution that are very close to their true values. The approximation with a smaller grid size does not appear to create a large bias in the estimate. Even though this does not constitute a true Monte Carlo experiment, this is reassuring regarding the validity of the procedure.

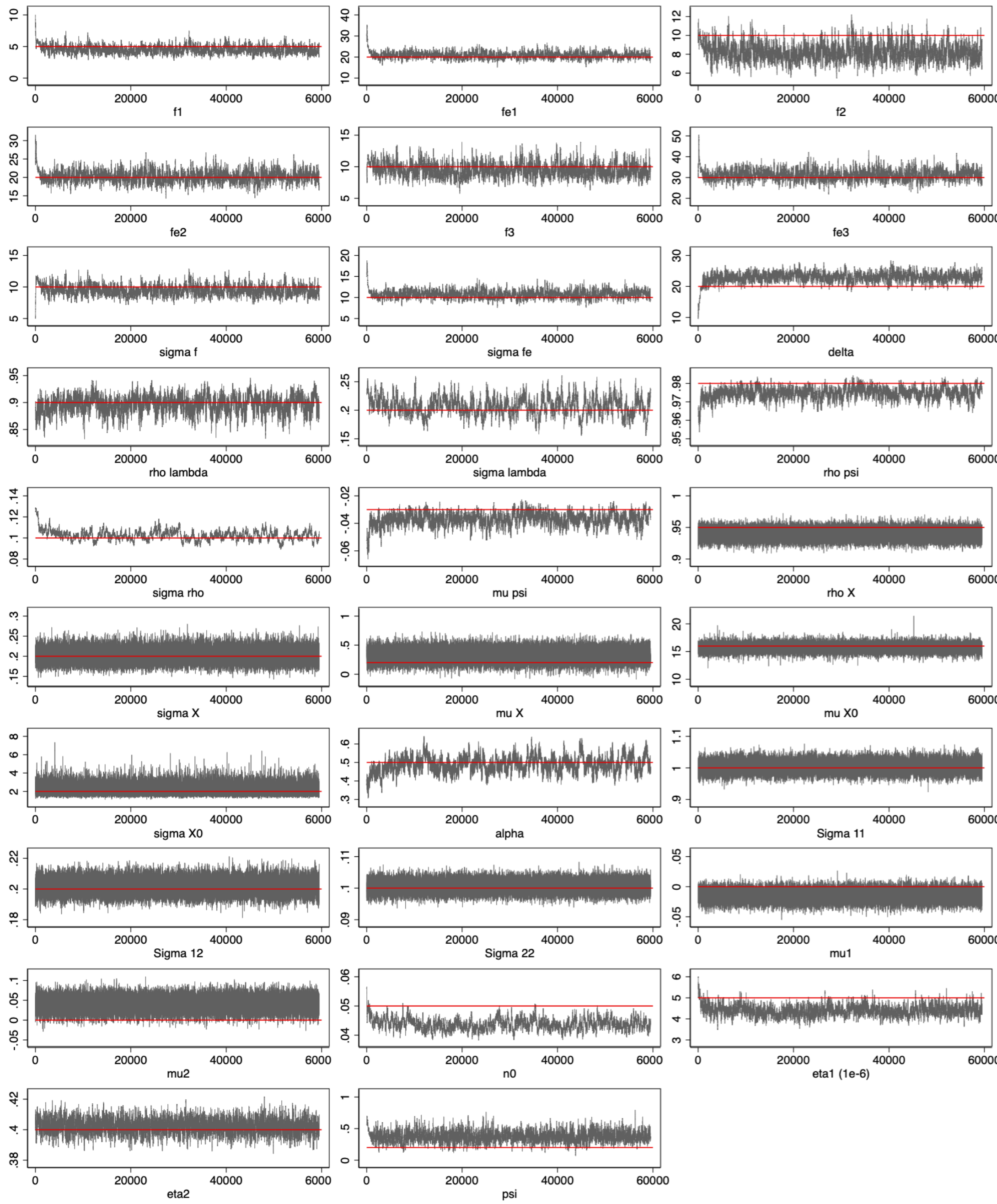

FIGURE 15: Markov Chains from the estimation on simulated data.

E MonteCarlo experiment

To further describe the validity of the estimation method, I perform a MonteCarlo experiment on a simplified version of the model. In this simplified version, I still model the entry and sales of different firms in various destination markets. However, I simplify the model developed in this paper in two ways: First, I reduce the number of serially correlated unobservables from three to two. Instead of productivity and quality components that shift sales and prices at the firm level, I only retain one firm-level unobservable, that shifts the sales in all destination markets. Moreover, I keep a market-specific and time-varying unobservable component that shift the sales and profit of all firms in a destination market. I respectively denote these two unobservables variables λ_{ft} and γ_{dt} and assume that they both follow an AR(1) process, as in the model, with the following parameters:

$$
\begin{aligned}
&\lambda_{f0} \sim N(0, \tfrac{\sigma_\lambda^2}{1-\rho_\lambda}) \\
&\lambda_{ft} \sim N(\rho_\lambda \lambda_{ft-1}, \sigma_\lambda^2) \\
&\gamma_{d0} \sim N(\mu_{\gamma 0}, \sigma_{\gamma 0}^2) \\
&\gamma_{dt} \sim N(\mu_\gamma + \rho_\gamma \gamma_{ft-1}, \sigma_\gamma^2)
\end{aligned}
$$

Second, I eliminate the dynamic pricing motive in the consumer accumulation process of the firm. As a consequence, this accumulation is exogenous and only depends on the age a_{ft} of the firm in a specific foreign market. Specifically, I model the consumer share of firm f in destination d as

$$
\begin{cases}
n(1) = n_0 \\
n(a_{fdt}) = 1 - (1-(1-\psi)\eta(a_{fdt}-1))^{\frac{1}{1-\psi}} & \text{if } a_{fdt} > 1
\end{cases}
$$

With these two simplifications relative to the model, the sales and export decisions of each firm f in each destination market d are obtained from the following equations:

$$
\begin{aligned}
&\log s_{fdt} = \lambda_{ft} + \gamma_{dt} + \log n(a_{fdt}) + \varepsilon_{fdt} \\
&\mathrm{In}_{fdt} = \begin{cases}
V(\lambda_{ft}, \gamma_{dt}, a_{fdt}) - f_e - \nu^e_{fdt} > V(\lambda_{ft}, \gamma_{dt}, 0) & \text{if } \mathrm{In}_{fdt-1} = 0 \\
V(\lambda_{ft}, \gamma_{dt}, a_{fdt}) - f_c - \nu^c_{fdt} > V(\lambda_{ft}, \gamma_{dt}, 0) & \text{if } \mathrm{In}_{fdt-1} = 1
\end{cases}
\end{aligned}
$$

with

$$
\begin{aligned}
V(\lambda_{ft}, \gamma_{dt}, a_{fdt}) &= \frac{E_\varepsilon s(\lambda_{ft}, \gamma_{dt}, a_{fdt})}{\sigma} + \beta \sigma_{\nu^c} \log\left(\exp\left(\frac{V(\lambda'_{ft}, \gamma'_{dt}, a_{fdt}+1) - f_c}{\sigma_{\nu^c}}\right) + \exp\left(\frac{V(\lambda'_{ft}, \gamma'_{dt}, 0)}{\sigma_{\nu^c}}\right)\right) \\
V(\lambda_{ft}, \gamma_{dt}, 0) &= \beta \sigma_{\nu^e} \log\left(\exp\left(\frac{V(\lambda'_{ft}, \gamma'_{dt}, 1) - f_e}{\sigma_{\nu^e}}\right) + \exp\left(\frac{V(\lambda'_{ft}, \gamma'_{dt}, 0)}{\sigma_{\nu^e}}\right)\right)
\end{aligned}
$$

assuming that $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and ν^e_{fdt} and ν^c_{fdt} follow a logistic distribution with variance parameters σ_{ν^e} and σ_{ν^c}. Following our main specification, we calibrate the values of σ and β and are left with 15 parameters to estimate: 7 related to the law of motion of the unobservables, 4 for the continuation, entry costs and their variance, 3 related to the law of motion of the consumer margin and the variance of the sales shocks, σ_ϵ.

From this simplified data generating process, we perform the following MonteCarlo experiment: we create 100 samples of 100 firms potentially exporting to 8 foreign destinations during 14 periods. For each sample, we run our MCMC estimator over 60 000 iterations of the Markov chain. We obtain an estimate of each of our 15 parameters as the median value of the Markov chain, after discarding the first 30 000 observations.

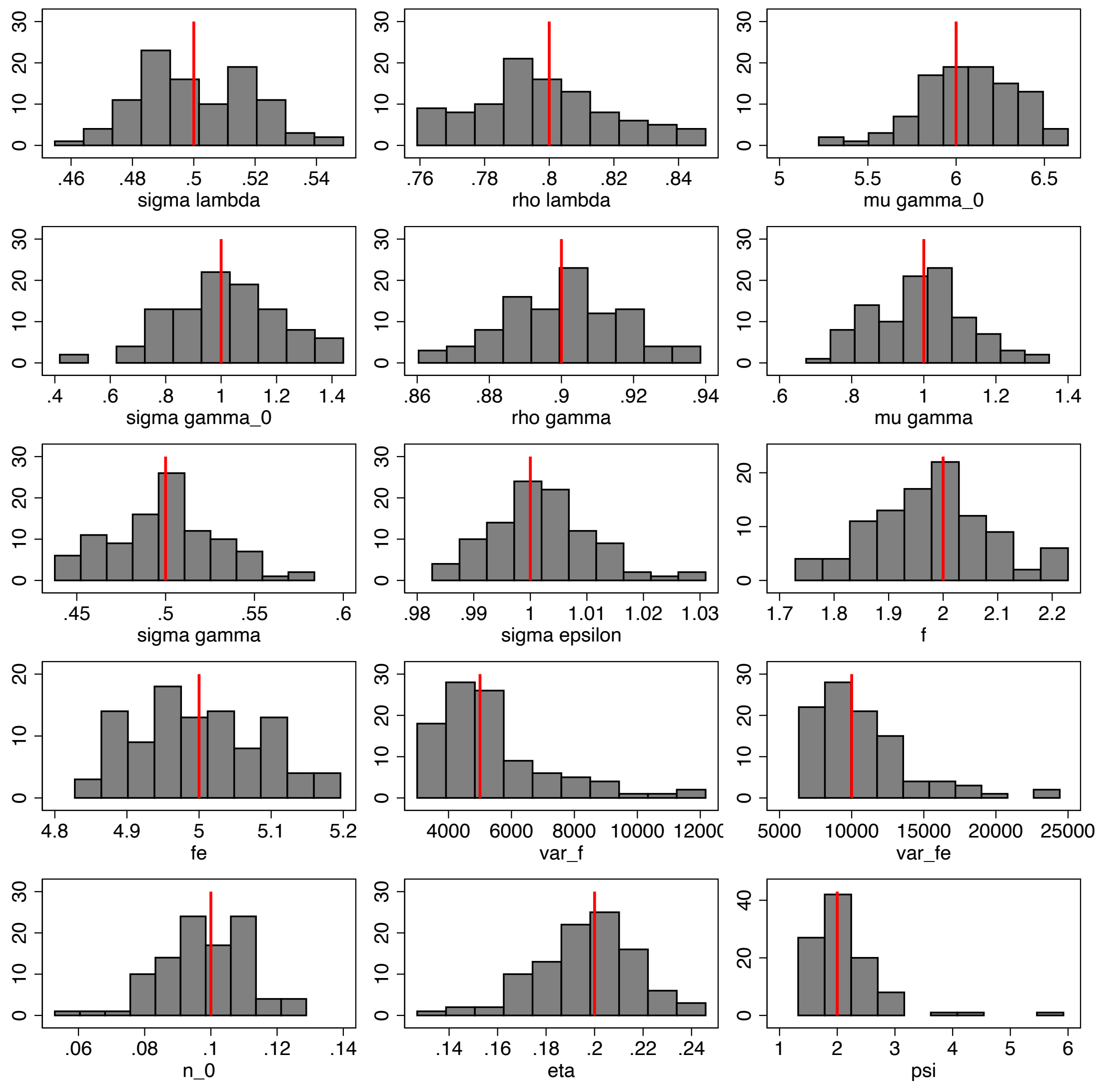

FIGURE 16: Distribution of parameter estimates from 100 MonteCarlo replications.

Figure 16 reports the distribution of the 100 different estimates obtained for each 15 parameters. Moreover, the figure reports in red the true value of each parameter, as used in the data generating process. Overall, the results of the MonteCarlo experiments validates the estimation methods. For all parameters, the distribution of estimates is well centered around the true value of the parameter. Moreover, most estimates tend to be very precise, with all estimates falling very close to the true value of the parameter. However, we do note than the estimates of the variance of the entry and continuation costs tend to be estimated with less precision. While the estimates are not biased on average, there is a small number of estimates that are much larger than the true value of the parameter.

Overall, this experiment confirms the validity of our empirical procedure: the use of particle sampling to approximate high dimensional integrals, combined with the iteration of the contraction mapping within the MCMC allows us to consistently estimate the dynamic problem described in this paper.

E Model extension

In this section, I describe an extension of the baseline model, in which I introduce heterogeneity in the initial share of consumers that firms receive when entering into a foreign market. In the model described in the main text, each firm receives a share n_0 during their first exporting year. In this section, I allow this initial share to vary for each firm-destination-year combination.

Specifically, I assume that initial shares are distributed according to a Beta distribution, with parameters n_0 and ϕ. This specification leads to a distribution of initial shares between 0 and 1, with a mean n_0 and a variance $\frac{n_0(1-n_0)}{\phi+1}$. Therefore, each firm that considers entering a foreign market draws an initial consumer share from this distribution. Having observed this draw, this firm decides whether to pay the entry cost and start exporting into this foreign destination.[71] I reestimate the model using this specification: while the baseline model imposes a consumer share equal to n_0 the first year, introducing this distribution allows the extended model to predict that some firms will start larger, and therefore are more likely to start exporting.

TABLE 8: Estimated parameters - Extended model

Parameter		**Estimate**	**90% Confidence Interval**	
			Lower bound	*Upper bound*
Continuation fixed costs	Europe	4 013	3 452	4 612
(in euros)	Americas	7 056	5 891	8 042
	Asia/Oceania	7 873	6 772	9 125
Entry fixed costs	Europe	25 028	22 056	28 098
(in euros)	Americas	18 676	16 596	20 998
	Asia/Oceania	19 063	16 911	21 486
Variance of continuation costs	σ_ν^c	12 496	10 794	14 089
Variance of entry costs	σ_ν^e	3 103	2 792	3 576
Fixed cost of n	δ	42 781	39 959	45 764
Law of motion of n	n_0	0.084	0.078	0.090
	$\eta_1 (10^{-6})$	9.14	8.13	10.29
	η_2	0.43	0.37	0.47
	ψ	0.75	0.63	0.88
	ϕ	19.27	17.27	21.72

I report the estimation results of this model in table 8. I obtain estimates for n_0 and ϕ of 0.08 and 19, which leads to an initial distribution of consumer shares with a mean of 8 percent and a standard deviation of 6 percentage points. Therefore, the model estimates a significant degree of heterogeneity in initial shares across firms. However, allowing for this heterogeneity has a small impact on the rest of the estimated parameters: specifically, only two sets of estimates appear to change with this extension. First, the entry costs and their variance decrease. This reduction is due to the fact that the model can now explain why firms enter or not with the heterogeneity in initial consumer share: when a firm is seen to not export, the standard model rationalizes this decision with high entry costs. The extended model can alternatively explain this decision by a

[71] To incorporate this heterogeneity in my estimator, I add a Metropolis-Hastings step in the MCMC iteration, in which I sample new initial shares from the Beta distribution, and accept or reject these shares based on the likelihood evaluations of the exporting spells associated with that initial share.

bad draw in terms of consumer share. More generally, because this type of heterogeneity is quite flexible, it can explain some of the observed entry decisions and reduce the need for entry costs. We also see an increase in the cost of maintaining the consumer share, δ. This increase is due to a small decrease of the average consumer share across models and the fixed costs of exporting: in order for the model to match observed exit rates, the extended model increases the unit cost of maintaining this consumer shares.

Given these small changes in parameter estimates, the outcomes of the model are relatively unaffected by the heterogeneity in initial consumer shares. I report in figure 17 the distribution of consumer shares across firms' age. In particular, the figure reports the distribution of consumer share at age 1, which was initially estimated to be 13 percent for all firms. These consumer shares are now distributed from 0 to 20 percent across firms. However, the distributions in the following years tend to be closely related to the ones obtained in the baseline model, although they tend to be smaller on average.

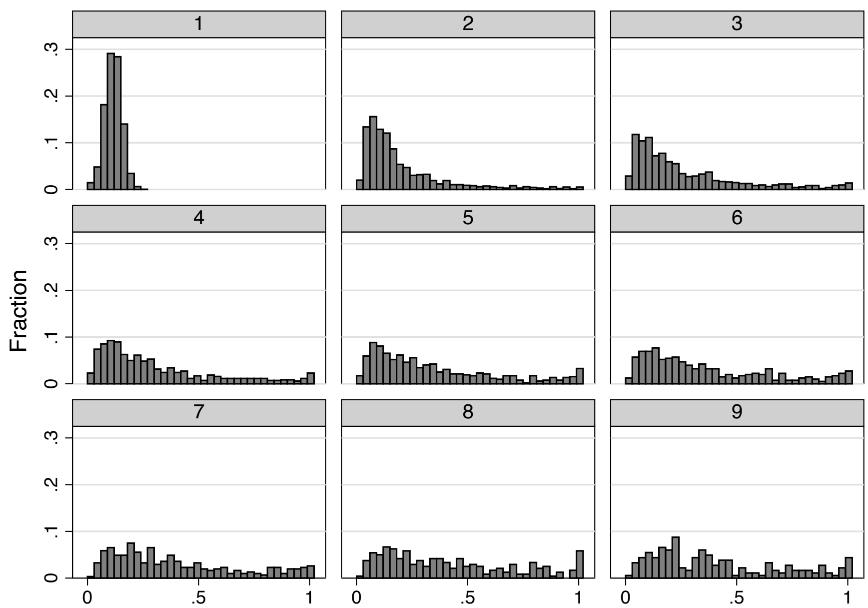

FIGURE 17: Distribution of consumer shares by firms' age (extended model)

Finally, I report in figure 18 the predicted sales, survival rates and prices across ages of the extended model, along the predictions from the model without consumer accumulation. While the model with consumer accumulation can predict the growth in sales, it actually performs poorly in terms of survival rates, similarly to the restricted model without consumer accumulation. The reason why this model does worst than the baseline in terms of survival rates and sales in the first year precisely comes from the heterogeneity in initial consumer shares. This model partly rationalizes the entry of firms by a relatively good draw in initial consumer share: since they will start with a large number of consumers, it is worth it to enter. However, this good draw makes it difficult to explain why these firms start so small, and why they exit at such a high rate. In summary, the heterogeneity in initial share helps to explain the entry of exporters, but makes it more difficult to correctly predict the sales and exit dynamics upon entry.

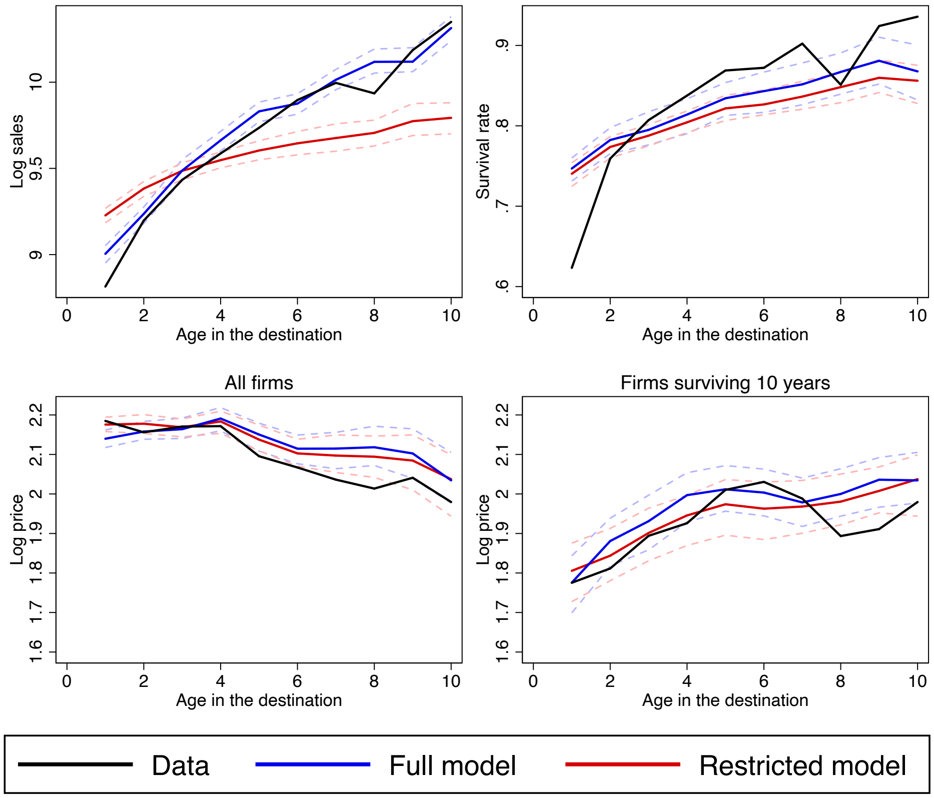

Figure 18: Predictions of survival rates, sales and prices across ages (extended model)

F Additional results

A Markov Chains

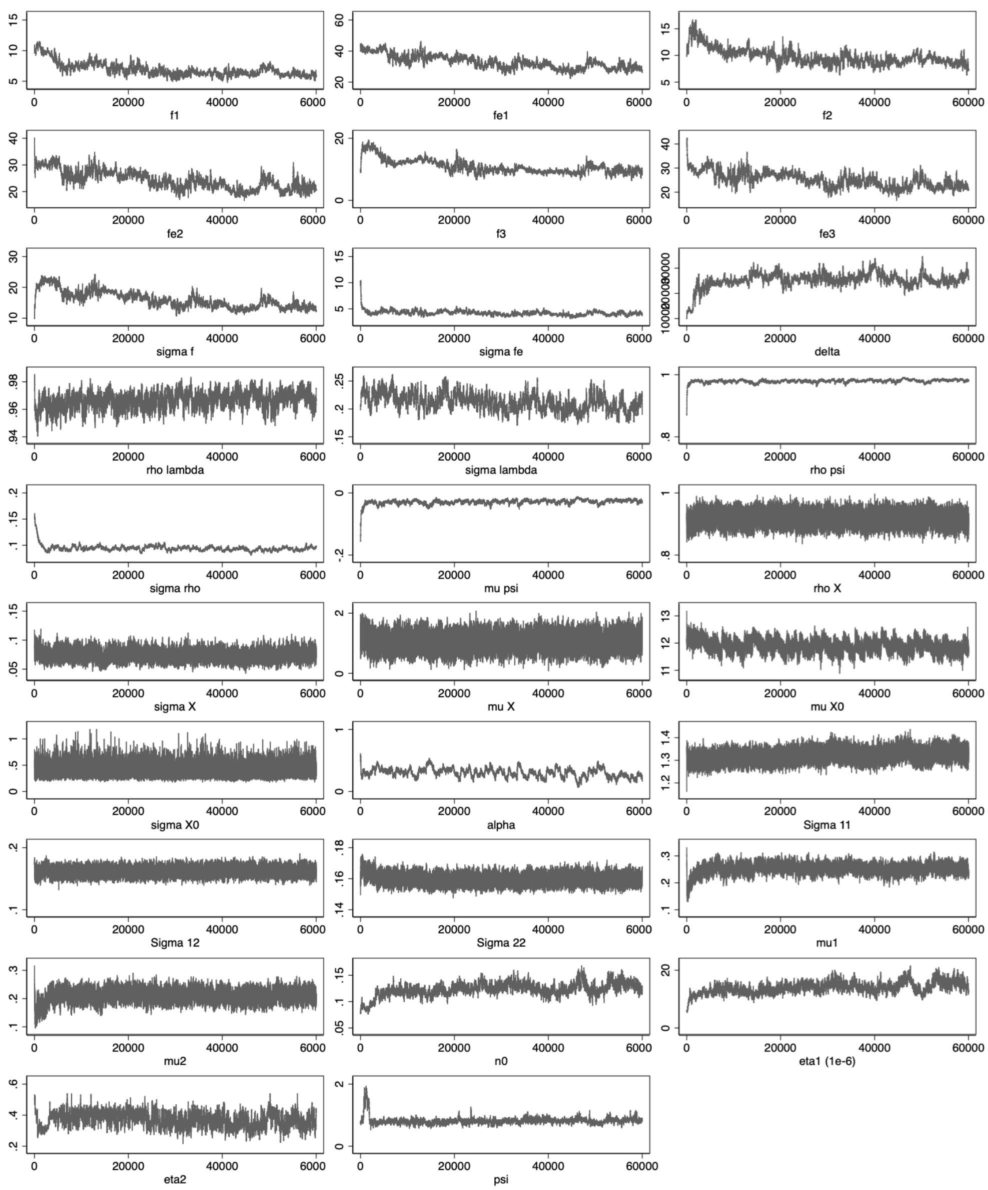

Figure 19: Markov Chains from the estimation.

Notes: 60 000 iterations are performed. Only the last 30 000 are used to compute the posterior distribution.

B Full results of the restricted model

Table 9: Estimated parameters - restricted model

Parameter		Estimate	95% Confidence Interval	
			Lower bound	*Upper bound*
Continuation fixed costs	Europe	13 036	12 307	13 867
(in euros)	Americas	15 849	14 460	17 296
	Asia/Oceania	18048	16 187	20 062
Entry fixed costs	Europe	76 579	69 242	84 690
(in euros)	Americas	61 643	55 768	68 438
	Asia/Oceania	69 304	62 399	77 144
Variance of continuation costs	σ_ν^c	22 533	19 808	25 561
Variance of entry costs	σ_ν^e	14 933	13 339	16 731
Law of motion of appeal	ρ_λ	0.97	0.96	0.97
	σ_λ	0.33	0.30	0.36
Law of motion of productivity	ρ_ψ	0.97	0.97	0.98
	σ_ϕ	0.12	0.11	0.13
	μ_ϕ	-0.03	-0.04	-0.03
Law of motion of agg. demand	ρ_X	0.95	0.92	0.98
	σ_X	0.15	0.12	0.18
	μ_X	0.60	0.27	0.94
	μ_{X_0}	10.13	9.61	10.64
	σ_{X_0}	0.84	0.57	1.26
Elasticity cost of appeal	α	0.19	0.13	0.25
Cost dummies	γ_2	0.32	0.30	0.34
	γ_3	0.30	0.28	0.33
Variance matrix	Σ_{11}	1.85	1.80	1.92
	Σ_{12}	0.19	0.18	0.20
	Σ_{22}	0.15	0.15	0.15

C Estimation results for the wood industry

In order to assess the robustness of the results found for wine exporters, I estimate my model for a sample of wood producers. I choose the wood industry because it allows me to obtain a sample of firms which export a consistently defined good: specifically, I only include in my sample exports of six HS4 product categories that are closely defined to each other, and exclude any firm that exports products outside of these categories to avoid the presence of multi-products firms.[72]

After implementing the same cleaning procedure used for the sample of wine producers, we obtain a final sample of 362 wood exporters. Moreover, because most of these exports are shipped to eight European countries, we reduce our estimation to this set of destinations.[73] Table 10 provides summary statistics on the sample of wood exporters used for the estimation.

Table 10: Description of the sample of wood exporters

Statistics:	*pc5*	*median*	*pc95*	*mean*	N
# observations per firm	11	21	61	26.7	362
av. # destinations per firm-year	1.14	2.14	4.79	2.48	3 899
av. # years per firm-destination	2.6	5.5	10.33	6.05	1 620

The main results of the estimation procedure are similar to the findings obtained with the sample of wine producers. Looking at the predictions of the model first, displayed in figure 20, we see that the model with consumer margin can predict the rise in export values and survival rates with experience in the foreign destinations: the model correctly predicts the rise of sales from 22 000 exp(10) at age 1 to 160 000 euros exp(12) at age 10. Similarly, it predicts most of the growth in survival rates with export experience. By contrast, the restricted model without consumer accumulation, cannot capture the rise in sales and therefore the rise in survival rates across ages.

However, the two models do not differ much when looking at the predictions in terms of prices. Prices tend to be relatively flat for wood exporters across ages and both models roughly match these levels. However, we notice that while the full model better predict the price in the first year for all firms, it underestimates the prices of surviving firms surviving 10 years in that first year. This result points out that wood exporters do not seem to aggressively use dynamic pricing to accumulate consumers in foreign markets.

Having discussed the predictions of both models, we now turn to the estimation of the models' parameters. Table 11 reports the estimated parameters and their confidence intervals for the full model, with consumer accumulation, and the restricted model, without accumulation, using the sample of wood exporters. Similar to the findings we reached with wine exporters, the main difference between the two models lies in the estimated fixed costs. When accounting for the

[72] I include the HS codes 4401 "Fuel wood, wood in chip...", 4403 "Wood in the rough", 4404 "Hoop wood, wooden sticks,...", 4406 "Railway or tramway sleepers of wood", 4407 "Wood sawn or chipped lengthwise", 4409 "Wood, continuously shaped".

[73] Foreign destinations are Spain, Germany, Netherlands, Belgium, Great-Britain, Portugal, Italy, Switzerland.

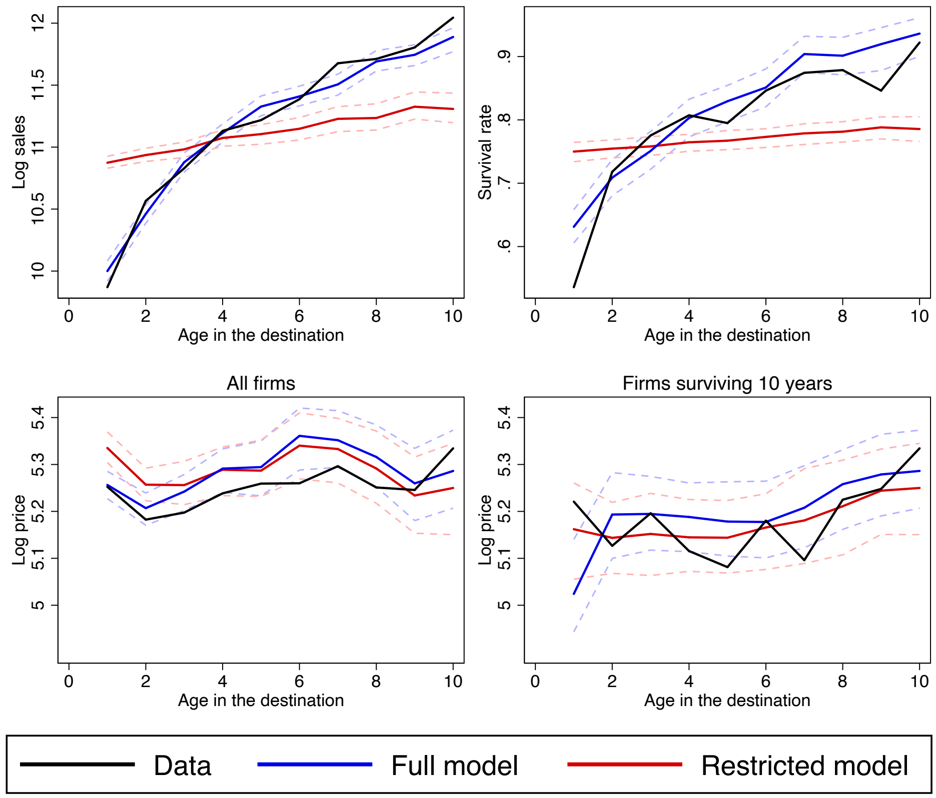

FIGURE 20: Predictions of survival rates, sales and prices across ages (sample of wood exporters)

consumer margin, entry and continuation fixed costs are estimated to be respectively 145 000 and 35 000 euros. These numbers are larger than in the case of wine exporters because the size of the trade flows are much larger: once normalized by export values, they are in the same ballpark than the numbers obtained for wine. By contrast, these numbers are much larger in the model without consumer accumulation: the fixed cost estimate is close to 10 millions euros. The reason for such a large number is that the model is unable to predict any entry or exit of firms, besides predicting that a firm will stay out if she is out, or will keep exporting if she is in. As a consequence, the model estimate a very large entry cost and a very large variance of these entry costs. With a parameter associated with the variance of entry cost at 3.3 millions, it implies that the standard deviations of the entry shocks is almost 6 millions. This large number allows the model to explain why some firms will eventually enter or exit the foreign markets. However, given these numbers, the actual characteristics of the exporters do not play a role in the entry or exit decisions, as illustrated by the flatness of the predicted survival rates across ages.

Besides this large discrepancy in estimated entry and fixed costs, we find little difference between the two models. The only element that should be noted is the larger variance obtained for the unobserved components in the restricted model. The variance of the demand, quality, productivity and aggregate demand shocks, are all larger in the restricted model relative to the model with consumer accumulation. This is further evidence that the restricted model does not fit the data well, and requires more variance in the shocks that helps the model match the data.

TABLE 11: Estimated parameters (sample of wood exporters)

Parameter		**Full model**			**Restricted model**		
		Estimate	**95% CI**		**Estimate**	**95% CI**	
Continuation fixed costs	f^c	35 382	30 337	39 917	2 188	16	7 956
Entry fixed costs	f^e	145 710	122 700	178 240	9 946 732	9 0763 50	10 879 000
Variance of continuation costs	σ_ν^c	51 927	43 616	59 856	521 772	462 340	587 740
Variance of entry costs	σ_ν^e	30 907	24 344	40 288	3 295 484	3 021 700	3 585 100
Fixed cost of n	δ	257 820	224 770	283 990			
Law of motion of n	n_0	0.029	0.026	0.033			
	$\eta_1(10^{-6})$	1.164	1.029	1.310			
	η_2	0.094	0.072	0.121			
	ψ	0.527	0.423	0.634			
Law of motion of appeal	ρ_λ	0.99	0.99	0.99	0.99	0.99	0.99
	σ_λ	0.16	0.14	0.17	0.30	0.27	0.32
Law of motion of productivity	ρ_ψ	0.94	0.93	0.95	0.99	0.98	0.99
	σ_ψ	0.05	0.05	0.07	0.09	0.09	0.09
	μ_ψ	-0.28	-0.34	-0.24	-0.06	-0.07	-0.05
Law of motion of agg. demand	ρ_X	0.96	0.89	1.00	0.97	0.93	1.00
	σ_X	0.02	0.01	0.03	0.10	0.07	0.12
	μ_X	0.71	0.08	1.96	0.46	0.00	1.14
	μ_{X_0}	18.48	18.35	18.62	17.67	17.35	17.98
	σ_{X_0}	0.06	0.04	0.12	0.42	0.27	0.65
Elasticity cost of appeal	α	0.86	0.83	0.89	0.44	0.41	0.47
Variance matrix	Σ_{11}	1.46	1.42	1.50	2.53	2.46	2.61
	Σ_{12}	0.20	0.19	0.21	0.26	0.24	0.27
	Σ_{22}	0.16	0.16	0.17	0.16	0.15	0.16

D Sensitivity to changes in the elasticity of demand

TABLE 12: Estimated parameters with different values of σ

Parameter		$\sigma = 2$	$\sigma = 3$	$\sigma = 4$
Continuation fixed costs *(in euros)*	Europe	6 258	5 791	3 481
		[5 447,7 499]	[4 841,6 499]	[3 116,3 892]
	Americas	8 938	7 745	4 186
		[7 497,10 636]	[6 651,9 012]	[3 576,4 963]
	Asia/Oceania	9 440	8 873	4 443
		[7 832,11 577]	[7 315,10 623]	[3 779,5 676]
Entry fixed costs *(in euros)*	Europe	29 730	24 641	15 596
		[25 766,35 123]	[22 694,27 655]	[13 861,17 646]
	Americas	21 633	18 081	11 366
		[18 500,26 376]	[16 168,20 379]	[9 932,13 142]
	Asia/Oceania	22 879	18 723	12 464
		[19 583,27 922]	[17 022,20 921]	[11 064,14 036]
Var. of continuation costs	σ_ν^c	13 943	12 349	6 986
		[12 063,17 379]	[10 353,13 884]	[6 001,8 507]
Var. of entry costs	σ_ν^e	4 025	2 895	2 205
		[3 488,4 662]	[2 573,3 442]	[1 928,2 419]
Fixed cost of n	δ	25 573	13 803	27 209
		[22 255,30 974]	[9 187,17 289]	[23 474,32 748]
Law of motion of n	n_0	0.13	0.09	0.05
		[0.11,0.15]	[0.080,0.11]	[0.041,0.056]
	$\eta_1(10^{-6})$	14.6	12.1	6.28
		[11.2,18.8]	[10.1,14.6]	[5.4,7.2]
	η_2	0.36	0.29	0.29
		[0.27,0.45]	[0.22,0.36]	[0.24,0.33]
	ψ	0.83	0.78	0.76
		[0.69,0.97]	[0.63,0.91]	[0.62,0.89]
Law of motion of λ	ρ_λ	0.97	0.98	0.99
		[0.96,0.98]	[0.98,0.99]	[0.99,1.00]
	σ_λ	0.20	0.25	0.31
		[0.18,0.24]	[0.18,0.29]	[0.26,0.38]
Law of motion of ψ	ρ_ψ	0.98	0.98	0.98
		[0.97,0.99]	[0.96,0.99]	[0.95,0.99]
	σ_ϕ	0.09	0.07	0.04
		[0.09,0.10]	[0.07,0.08]	[0.04,0.05]
	μ_ϕ	-0.03	-0.04	-0.03
		[-0.04,-0.02]	[-0.07,-0.02]	[-0.08,-0.02]
Law of motion of X	ρ_X	0.92	0.93	0.94
		[0.88,0.96]	[0.89,0.97]	[0.89,0.98]
	σ_X	0.07	0.08	0.06
		[0.06,0.09]	[0.06,0.10]	[0.04,0.08]
	μ_X	1.06	1.07	0.97
		[0.60,1.52]	[0.50,1.67]	[0.33,1.78]
	μ_{X_0}	11.86	14.10	16.22
		[11.50,12.26]	[13.44,14.58]	[15.50,16.89]
	σ_{X_0}	0.39	0.40	0.32
		[0.25,0.64]	[0.27,0.65]	[0.20,0.54]
Elasticity cost of λ	α	0.27	0.27	0.24
		[0.16,0.43]	[0.06,0.34]	[0.20,0.28]
Cost dummies	γ_2	0.25	0.19	0.16
		[0.22,0.28]	[0.16,0.23]	[0.12,0.20]
	γ_3	0.21	0.16	0.10
		[0.18,0.25]	[0.11,0.20]	[0.05,0.15]
Variance matrix	Σ_{11}	1.33	1.81	2.68
		[1.28,1.38]	[1.75,1.88]	[2.58,2.80]
	Σ_{12}	0.16	0.33	0.51
		[0.15,0.18]	[0.31,0.35]	[0.49,0.54]
	Σ_{22}	0.16	0.16	0.17
		[0.15,0.17]	[0.16,0.17]	[0.16,0.17]

E Additional figures

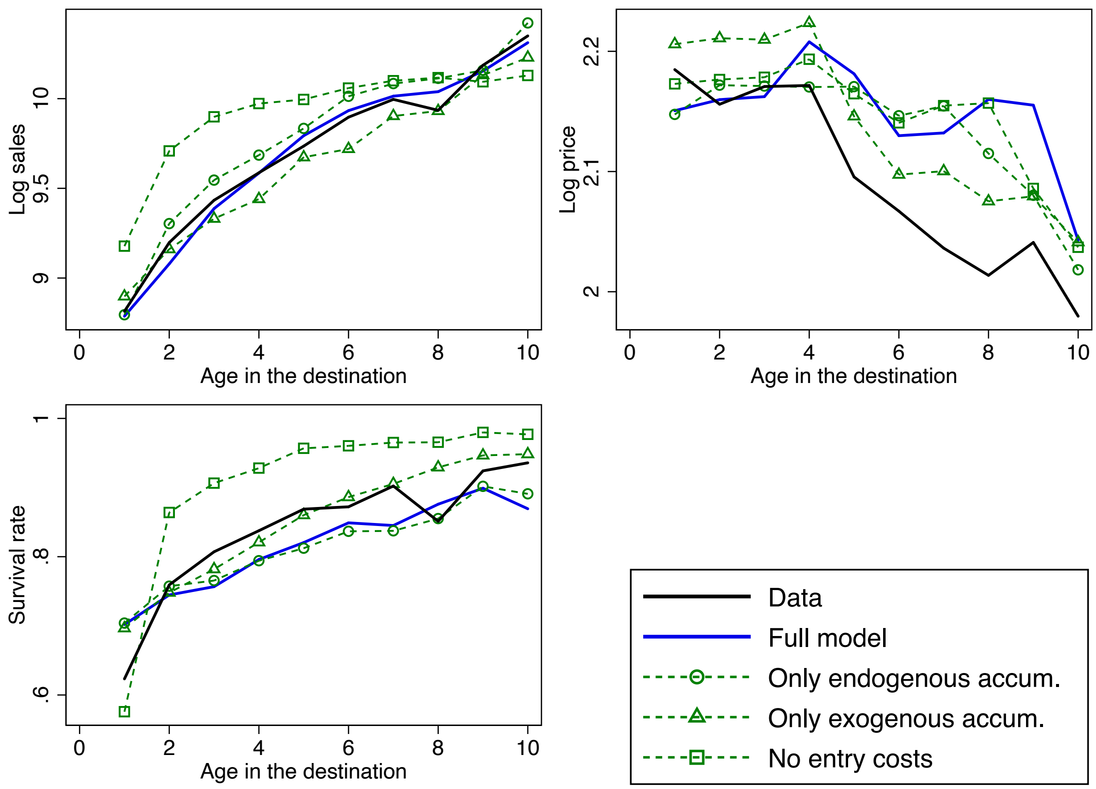

FIGURE 21: Predictions of alternative models

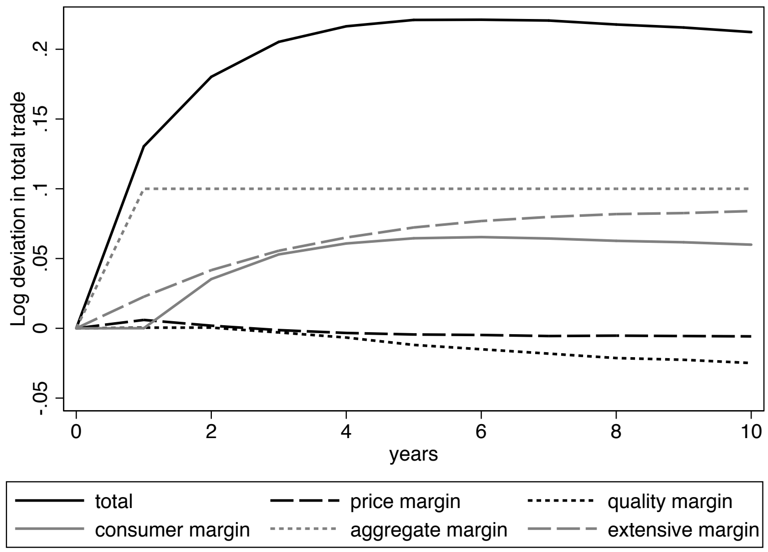

Figure 22: Effect of permanent 10 points tariffs decrease (All margins).

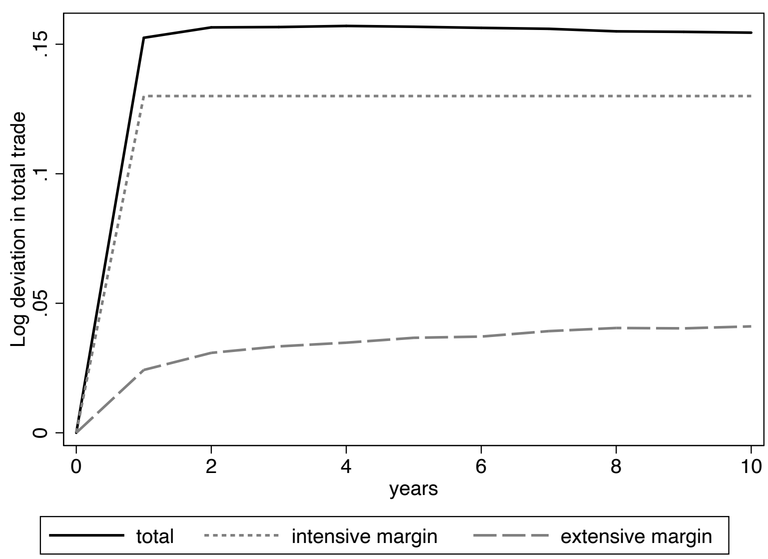

Figure 23: Effect of permanent 10 points tariffs decrease (Restricted model).

G Details on out-of-sample predictions

In order to perform out-of-sample predictions, I construct variations in the aggregate demand from Brazil. This variable is defined from the model as $X_{dt} = \log Y_{dt} - (1-\sigma)\log P_{dt} + (1-\sigma)\log(\tau_{dt} e_{dt})$. Importantly, I only need to construct variations of aggregate demand since the level will be chosen to exactly match the total exports of French firms to Brazil. Therefore, in addition to using the Brazilian GDP, the exchange rate between France and Brazil, I also need to construct a proxy for variations of the price index in the Brazilian market. In order to do so, I use variations in exchange rates from the five main countries exporting to Brazil. Table 13 describes these countries and their respective market shares.

TABLE 13: Top market shares

Country	**Average market share**
France	22.1 %
Italy	20.4 %
Chile	19.6 %
Portugal	15.6 %
Argentina	13.5 %

Note: Calculations made from BACI. Average market share is the average market share among the Brazilian imports, over the period 1997-2007, for the 4-digit category 2204 '*Wine of fresh grapes*'.

Note that the next largest wine exporter to Brazil has a market share of less than 2 percent and is therefore not included in the computation of the price index. Therefore, I construct variations in $X_{B,t}$ as following:

$$\begin{aligned} X_{B,t} - X_{B,98} &= \log Y_{B,t} - \log Y_{B,98} - (1-\sigma)\left[\log P_{B,t} - \log P_{B,98}\right] \\ &\quad + (1-\sigma)\left[\log(\tau_{F,t} e_{F,t}) - \log(\tau_{F,98} e_{F,98})\right] \\ &= \log Y_{B,t} - \log Y_{B,98} - \log\left(\sum_i \omega_{i,98}\left(\frac{e_{i,t}}{e_{i,98}}\right)^{1-\sigma} + 1 - \sum_i \omega_{i,98}\right) + (1-\sigma)\log(\frac{e_{F,t}}{e_{F,98}}) \end{aligned}$$

where the difference in Y is computed from changes in the Brazilian GDP, $\omega_{i,98}$ are the import shares of each of the five countries displayed in the table 13 and $e_{i,t}$ their exchange rates with Brazil. The obtained variations in aggregated demand for French wine is described in figure 24, which highlights the impact of the Brazilian and Argentinian devaluations.

I then perform 500 simulation of trajectories using the median product appeals and productivities obtained from the estimation, and the constructed variation in aggregated demand. I set the level of X_{dt} such that the median prediction exactly predicts the right amount of export from France to Brazil in 1998. Trajectories differ because I need to simulate demand and supply shocks (ε) and fixed costs shocks (ν) in order to obtain predictions for each firm. Predictions reported in the text are based on the median trajectories, and I report in figure 25 the 90% confidence interval of these predictions.

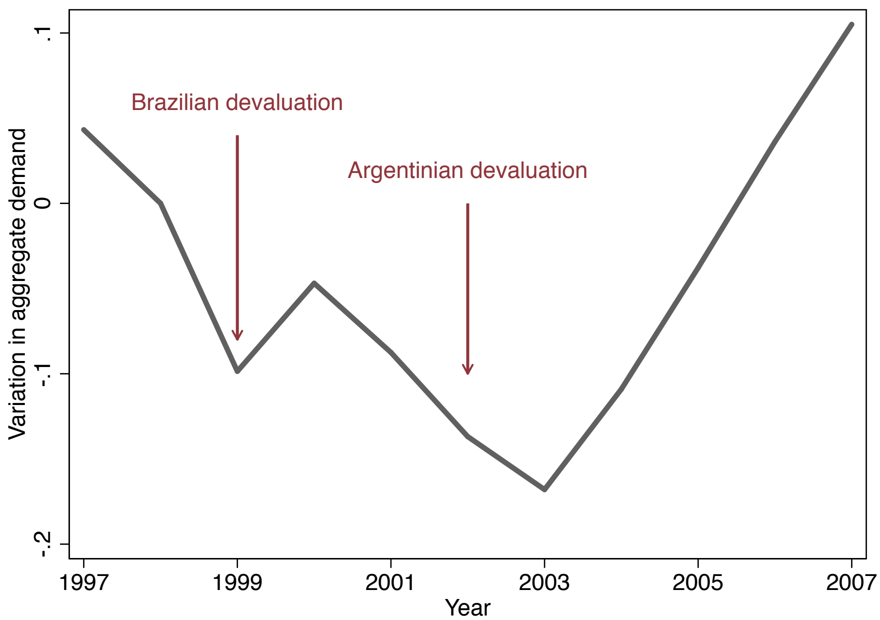

Figure 24: Computed variations in aggregate demand for French wine from Brazil.

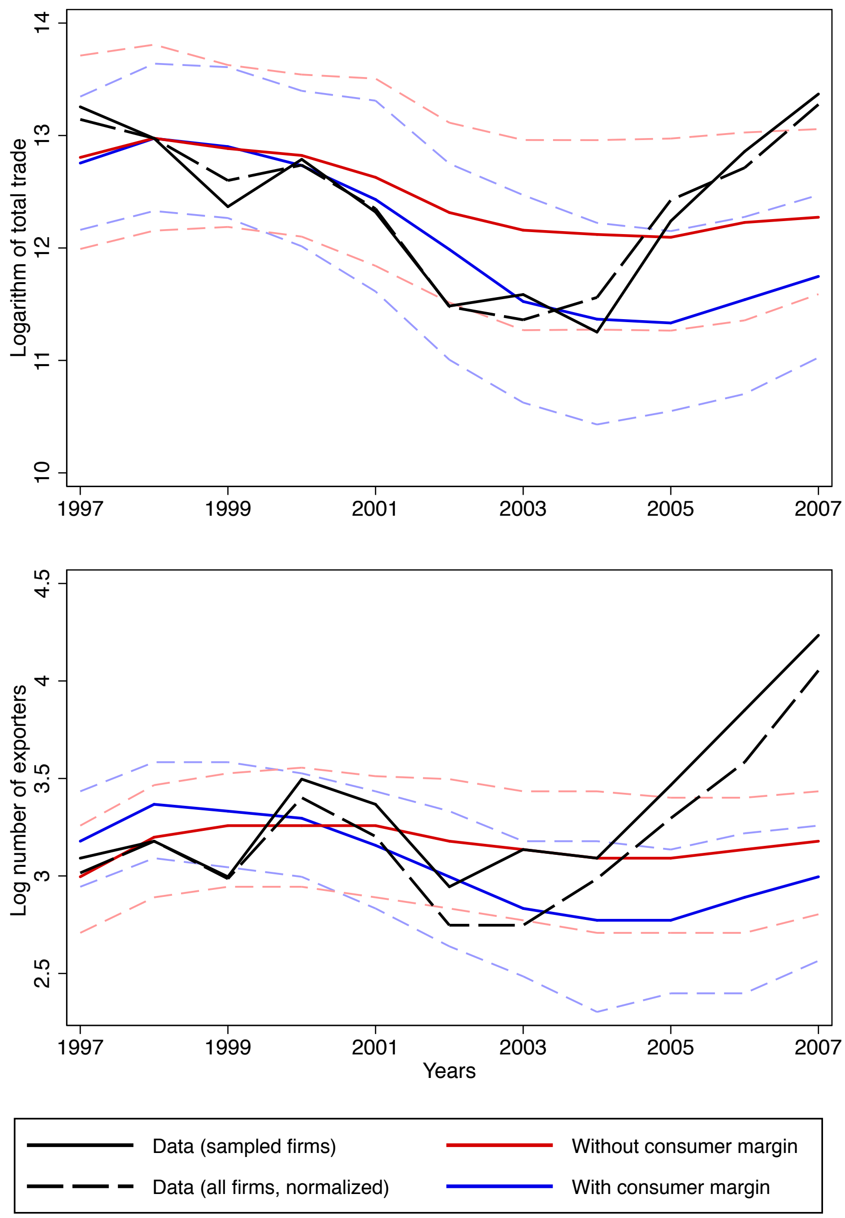

Figure 25: Total exports of wine to Brazil from selected firms

H Replication exercise

As part of the editorial process leading to the publication of this paper, A French organization, named Cascad, was tasked to perform a complete replication of the manuscript. This exercise required to gain access to the French confidential data used in the paper, and the dataset provided by the French customs was slightly different than the version received by the author in 2014. This small discrepancy in the structure of the data leads to small differences in the full sample used for section 2, and in the sample of shoe producers used in the estimation of the dynamic model. Table 14 highlights the (small) differences between samples.

Table 14: Samples comparison

Full sample	*# firm-year pairs*	*# obs. per firm-year (average)*
Original sample	1 510 030	16.5
Replication sample	1 509 898	16.5
Sample of shoe exporters	*# firms*	*# obs. per firm (average)*
Original sample	236	41.1
Replication sample	238	40.8

Despite those discrepancies, the results and conclusions of the paper were replicated. Some differences arise in tables because of slightly different samples, but none of these differences impact the results and conclusions of their respective table.[74] However, this replication exercise highlighted some numerical fragilities in the algorithm developed to estimate the dynamic model in the paper. In particular, the estimation of the model on the sample of wood exporters (appendix C) could not be performed: the initialization of the algorithm on the replication sample generates starting conditions that makes the MCMC estimator fails after a few iterations. Similarly, the estimation of the model with $\sigma = 4$ (appendix D) is very slow and could not be completed by the replication team. These problems were not present in the initial sample and does not appear for all the other programs that can correctly replicate the paper. This issue calls for some tuning in the initialization phase of the MCMC algorithm, and implies that the reliability and flexibility of the estimator could be improved with a better method to choose the initial starting value of the Markov Chain.

[74] All replicated figures and tables can be found in the Replication folder located in the IPCSR repository associated with the manuscript.