Productivity and R&D as Drivers of Exports and Domestic Sales: Semiparametric Evidence from French Firm-level Data*

Peter H. Egger, Katharina Erhardt, and Andrea Lassmann

June 27, 2013

Abstract

Earlier work suggests that the relationship between exports and trade costs and exports and foreign market size is well described by a log-linear functional form. A log-linear relationship between exports and productivity is underlying workhorse models of trade as well. The same pattern is assumed to hold for domestic sales. How the intensity of R&D affects quality or preferences and productivity, and hence, exports and domestic sales, is less well established. This paper is devoted to estimating the relationship between productivity, R&D, and sales to different markets in a flexible, nonparametric way which is embedded in a semiparametric estimation approach. Using French firm-level data, we can confirm a standard log-linear relationship for exports. However, it takes a rather nonlinear functional form for domestic sales, suggesting that the data are generated by a potentially different class of models than the one considered traditionally.

Keywords: firm-level exports; productivity; semiparametric methods

JEL Classification: F14; D24; C14; O3

1 Introduction

Firm-level productivity (see Melitz, 2003; Arkolakis, 2008; Chaney, 2008; Melitz and Ottaviano, 2008) and quality or product features (see Hummels and Klenow, 2005; Hummels and Skiba, 2004; Lugovskyy and Skiba, 2010) are key drivers of demand in new trade theory models. Such models assume that the impact of both productivity

*Acknowledgements: We would like to thank Ron Davies and participants at the CESifo-Delphi conference at the ifo Institute of Munich in 2013.
and quality or preferences on firm-level domestic sales and exports is log-linear and positive (see Arkolakis and Muennder, 2011; Baldwin and Harrigan, 2011; Corcos, Del Gatto, Mion, and Ottaviano, 2011; Egger, Egger, and Kreickemeier, 2012). It has never been shown to date whether these assumptions hold empirically or not when considering firm-level data. The aim of this paper is to shed light on the matter by allowing for a flexible relationship between sales to different markets and quality or preferences measured by R&D as well as total factor productivity (TFP) in a semiparametric approach, and to contrast the evidence with an assumed log-linear form using firm-level data.

The paper bases its evidence on firm-level data for France as provided by Bureau van Dijk. The data contain information on a firm’s exports along with balance sheet data on total sales and value added, profits, employment, and labor costs, material costs, fixed assets and capital costs, and intangible as well as tangible assets over time. These are the ingredients required to estimate firm-level total factor productivity as well as marginal production costs, permitting an analysis of the relationship between total factor productivity and R&D on firm-level domestic sales (of exporters and non-exporters) as well as on exports. The latter analysis is carried out by way of nonparametric model estimation in a cross section of firms.

This analysis reveals that, for a large proportion but not all of the exporters, total factor productivity as well as the intangibility ratio are related to export sales in close-to a log-linear fashion as is commonly hypothesized. The same is true for domestic sales of exporters. However, total factor productivity affects sales of non-exporters in a hump-shaped fashion, while the impact of the intangibility ratio is still relatively linear. In general, both total factor productivity and asset intangibility are related significantly to sales of exporters as well as sales of domestic sellers. The additional analysis of total sales yields a different pattern, which indicates that focusing on aggregate sales results in a biased relationship and points to the need to distinguish between destination markets. These findings suggest a number of conclusions. For instance, an arbitrary domestic seller’s log ratio of total factor productivity to any other domestic seller (e.g., the marginal firm in the market at the lower productivity cutoff) is not proportional to the same two firms’ ratio of domestic sales. The latter is at odds with the assumption of a Pareto distribution in conjunction with monopolistically competitive domestic suppliers. Adopting this assumption is not harmful in samples of exporters but it is for domestic suppliers. In any case, this evidence for France points to a potential need of rather a different class of models than the one usually employed in international economics in recent years.

The remainder of the paper is organized as follows. The next section relates our work to existing knowledge about the impact of productivity and fixed costs on
exporting. Section 3 provides details about the identification strategy adopted in this paper. Section 4 introduces the French firm-level data-set used in the empirical analysis and summarizes the estimation results, and the last section concludes.

2 Determinants of firm-level exports and domestic sales

In a variety of isomorphic trade models, firm i’s sales volume to market $\ell = \{H, F\}$ – where $H$ and $F$ refer to Home and Foreign – can be generically formulated as

$$X_i^\ell = \frac{(B_i\Phi_i^{-1} C_i T_i^\ell)^\theta Y_i^\ell}{(P_i^\ell)^\theta} \equiv (B_i\Phi_i^{-1} C_i)^\theta D_i^\ell,$$

(1)

where $B_i$ is an inverse preference parameter for firm i’s output, $\Phi_i$ is the firm’s total factor productivity, $C_i$ are marginal costs per unit of factor input, $T_i^\ell$ are (multilateral) trade costs which are unity whenever $\ell = F$, $Y_i^\ell$ are expenditures in market $\ell$ on goods as the ones of i, and $P_i^\ell$ is the consumer price index for goods of the same type as i’s in market $\ell$. By convention, we may define $D_i^\ell \equiv (T_i^\ell/P_i^\ell)^\theta Y_i^\ell$. It is customary to introduce index s to refer to sector $s = 1, ..., S$ along with $N_s$ and $N_s$ to refer to the number and the set of firms in sector s. In that case, $N = \sum_{s=1}^{S} N_s$ would denote the total number and $N = N_1 \cup ... \cup N_S$ would denote the overall set of firms in all sectors together. In line with the literature, let us assume that trade costs, expenditures, and price indices are sector specific so that we can replace the index i on $D_i^\ell$ by s. Then, we may reformulate (1) as

$$X_i^\ell = (B_i\Phi_i^{-1})^\theta C_i^\theta D_s^\ell.$$

(2)

What is important at this point is that $B_i$, $\Phi_i$, and $C_i$ are assumed to have a log-linear impact on the log-transformed $X_i^\ell$, say, $x_i^\ell \equiv \ln X_i^\ell$. In this paper, we will assume that $(B_i\Phi_i^{-1})^\theta$ can be modeled as a flexible, possibly non-multiplicative function of total factor productivity, $\Phi_i$, and firm-level research and development (R&D) or asset intangibility, $R_i$, so that $(B_i\Phi_i^{-1})^\theta = g^\ell(\Phi_i, R_i)$. The latter is consistent with the assumption taken in this paper that research input ($R_i$) as part of $B_i$ and research output ($\Phi_i$) together affect not only firm-level efficiency but also consumer preferences. It is this paper’s task to ask whether and to which extent this is the case.
3 Semiparametric identification strategy

For estimation, it is convenient to log-transform equation (2) and to add a stochastic term. For this, let us use the convention to refer to log-transformed variables by corresponding lower-case letters, and let us replace the sector-specific variable $d_s$ by a fixed effect $\delta_s(i)$ which pertains to the sector firm $i$ belongs in, and refer to the stochastic term by $u^\ell_i$ to obtain

$$x^\ell_i = g^\ell(\phi_i, r_i) + \theta c_i + \delta^\ell_s(i) + u^\ell_i,$$

where $g^\ell(\cdot)$ denotes a flexible, smooth, nonparametric function of its arguments.

Let us sort the data first by $\phi_i$ and then by $r_i$ so that firms $i - 1$ and $i + 1$ are the nearest neighbors to $i$ in terms of $\{\phi_i, r_i\}$. Then, with densely-enough populated data and a smooth function $g^\ell(\cdot)$, $g^\ell(\phi_i, r_i) - g^\ell(\phi_{i-1}, r_{i-1}) \approx 0$ (see Yatchew, 2003) so that

$$(x^\ell_i - x^\ell_{i-1}) \approx \theta(c_i - c_{i-1}) + (\delta^\ell_s(i) - \delta^\ell_s(i-1)) + (u^\ell_i - u^\ell_{i-1}).$$

Hence, $\theta$ and $\delta_s(i)$ can be estimated from a linear model on the differenced data as in (4). With estimates $\hat{\theta}$ and $\hat{\delta}_s$ at hand, one may estimate $g^\ell(\cdot)$ nonparametrically from a model involving the transformed log market-specific sales of firm $i$ as

$$\xi^\ell_i \equiv x^\ell_i - \hat{\theta} c_i - \hat{\delta}^\ell_s = g^\ell(\phi_i, r_i) + \nu^\ell_i.$$

For empirical analysis, we will measure $r_i$ by a function of the ratio of intangible assets and intangible plus tangible assets, and we will use estimates of $\hat{\phi}_i$ obtained from a semiparametric model following Levinsohn and Petrin (2003), $\hat{\phi}_i$.

Unit production costs can be derived from a Cobb-Douglas production function of the general form $Y = \Phi L^\alpha K^\beta$, with parameters obtained from estimating $\hat{\phi}_i$ as in Levinsohn and Petrin (2003) and adjusting output by productivity, $Y_i = Y_i/\Phi_i$:

$$C_i = \frac{\Omega_i + M_i}{Y_i},$$

where $\Omega_i$ denotes total wages and $M_i$ denotes interest paid.

For the nonparametric estimation of (5), we will pick each firm $v \in \mathcal{N}$ one after another as a reference point and calculate the distance between $i$ and $v$ for any generic variable $z \in \{\phi, r\}$ as $\Delta z_{iv} \equiv z_i - z_v$ in order to estimate $g^\ell(\cdot)$ by a bivariate product kernel-weighted local linear regression model of the form

$$\sum_{i=1}^{N} \left\{ \xi^\ell_i - \alpha^\ell_v - \beta^\ell_{rv} \Delta r_{iv} - \beta^\ell_{\hat{\phi}v} \Delta \hat{\phi}_{iv} \right\} K^\ell_h(\Delta r_{iv}) K^\ell_h(\Delta \hat{\phi}_{iv}),$$

where $\mathcal{N}$ is the set of all firms in the sample.
where $K_{h,z}$ is a kernel function about argument $\Delta z$ using bandwidth $h_z$, with $z \in \{r, \hat{\phi}\}$. In general, we employ an Epanechnikov kernel function and determine the optimal bandwidth by least-squares cross validation (see Fan and Gijbels, 1996).

4 Empirical analysis using French firm-level data

For the empirical analysis in this paper, we use balance sheet information from a panel of French firms as published by Bureau van Dijk in its Amadeus Database. The advantage of using data for France is that a large number of enterprises are covered, and that, unlike for most other economies covered by Amadeus, information about exports is available beyond balance sheet data. Between 1998 and 2007, the database provides information on 994,560 observations, representing 99,456 firms. We will use the time variation in the data to estimate $\phi_{it}$ in the next subsection and, for the sake of feasibility of the nonparametric regressions and because variation over time is not essential in understanding the relationship between the variables of interest in this paper, we will create a cross section in the following subsections of this paper.

4.1 Estimating log total factor productivity, $\phi_{it}$

We estimate total factor productivity in line with Levinsohn and Petrin (2003), using the following variables for firms in the manufacturing sector that our data allows us to use as factors of production: log profit-adjusted output $y_{it}$ (value added—profit or loss in a given period); log labor input $l_{it}$ (number of employees); log capital input $k_{it}$ (total assets); and log intermediate input $m_{it}$ (materials). We deflate the variables except for $l_{it}$ using price indices from French annual national accounts for 40 NACE Rev. 2 Divisions provided by INSEE. Discarding missing observations leads to summary statistics of the variables as shown in Table 1.

While the number of employees in manufacturing decreases over time in our sample, the remaining variables exhibit some volatility across the years but are generally quite stable. In order for $m_{it}$ to be a valid proxy for technology, it has to be increasing in $\phi_{it}$, conditional on capital. Even though this assumption is sometimes violated for low values of capital in the smoothed graph presented in Figure 1, a vast majority of observations lies in those parts of the graph where this

---

1We are aware of the fact that Amadeus does not provide information on the universe of all firms in France. However, the first and second moments of the data match relatively well between the (economically somewhat larger) sub-set in Amadeus and the universe of the data (Farid Toubal provided evidence on this on the occasion of a discussion of Egger, Egger, and Kreickemeier, 2012, at the “Globalization and Labor Market Outcomes: Recent Advance” conference at Banque de France on May 16-17, 2013.

5
Table 1: Descriptive statistics for estimation of φ_{it}

<table>
<thead>
<tr>
<th>Year</th>
<th>Obs.</th>
<th>Output</th>
<th>Labor</th>
<th>Capital</th>
<th>Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>15910</td>
<td>4616</td>
<td>90</td>
<td>15205</td>
<td>9341</td>
</tr>
<tr>
<td>1999</td>
<td>23293</td>
<td>3712</td>
<td>75</td>
<td>12467</td>
<td>8097</td>
</tr>
<tr>
<td>2000</td>
<td>24658</td>
<td>3818</td>
<td>72</td>
<td>13791</td>
<td>8305</td>
</tr>
<tr>
<td>2001</td>
<td>26162</td>
<td>4075</td>
<td>75</td>
<td>14128</td>
<td>9465</td>
</tr>
<tr>
<td>2002</td>
<td>28128</td>
<td>4446</td>
<td>75</td>
<td>14484</td>
<td>9762</td>
</tr>
<tr>
<td>2003</td>
<td>29259</td>
<td>4005</td>
<td>73</td>
<td>13299</td>
<td>9402</td>
</tr>
<tr>
<td>2004</td>
<td>30651</td>
<td>4127</td>
<td>69</td>
<td>14238</td>
<td>10196</td>
</tr>
<tr>
<td>2005</td>
<td>31688</td>
<td>3950</td>
<td>69</td>
<td>14922</td>
<td>10731</td>
</tr>
<tr>
<td>2006</td>
<td>31035</td>
<td>3545</td>
<td>61</td>
<td>12847</td>
<td>8155</td>
</tr>
<tr>
<td>2007</td>
<td>36664</td>
<td>3608</td>
<td>57</td>
<td>14437</td>
<td>10280</td>
</tr>
</tbody>
</table>

Source: Bureau van Dijk Amadeus database. Values represent total number of non-missing observations (Obs.) and mean values of value added (output), number of employees (labor), total assets (capital) and intermediate input costs (materials) by year.

Condition is fulfilled. The estimation of the total factor productivity parameter φ_{it} is based on a Cobb-Douglas production function (in logs) of the following form:

\[ y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \phi_{it} + \eta_{it}, \]

where \( \eta_{it} \) is an error term. Assuming common input and output prices across firms, no measurement error in the input demand function, and \( l_{it} \) being uncorrelated with productivity shocks in \( t+1 \), \( \beta_m \) and \( \beta_k \) can be identified.\(^2\) The estimated elasticities are presented in Table 2. They are quantitatively comparable to the OLS estimates. This could suggest that the bias from running OLS is small. Moreover, the results are relatively robust to using (profit-)unadjusted value added instead of the adjusted measure. The latter suggests that profitability does not vary systematically with productivity (as would be the case with monopolistic competition).

\(^2\)In a first stage, \( y_{it} = \beta_l l_{it} + f_{it}(m_{it}, k_{it}) + \eta_{it} \) is estimated using a locally weighted quadratic least squares approximation, where \( f_{it}(m_{it}, k_{it}) = \beta_0 + \beta_k k_{it} + \beta_m m_{it} + \phi_{it}(m_{it}, k_{it}) \). Then, the conditional moments are estimated and the expectation conditional on \( (k_{it}, m_{it}) \) is subtracted. Using two moment conditions to identify \( \beta_m \) and \( \beta_k \), and six over-identifying restrictions, \( y_{it} = \beta_0 + \beta_k k_{it} + \beta_m m_{it} + E[\phi_{it}|\phi_{it-1}] + \eta_{it} \) is estimated in a second stage.
Figure 1: Productivity ($\hat{\phi}_{it}$) as a function of capital ($k_{it}$) and material costs ($m_{it}$).

Table 2: Levinsohn and Petrin (2003) productivity estimation

<table>
<thead>
<tr>
<th></th>
<th>Coeff.</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log real value added</td>
<td>0.566</td>
<td>0.003***</td>
</tr>
<tr>
<td>Log number of employees</td>
<td>0.410</td>
<td>0.013***</td>
</tr>
<tr>
<td>Log real total assets</td>
<td>0.410</td>
<td>0.013***</td>
</tr>
<tr>
<td>Observations</td>
<td>274215</td>
<td></td>
</tr>
</tbody>
</table>
4.2 Estimating the trade elasticity, $\theta$, and the sector-specific fixed effects, $\delta^t_s$

With estimates of total factor productivity at hand, we drop outliers in the 99th percentile of $\hat{\phi}_{it}$ and the first and 99th percentile of $r_{it}$ and create a cross section of firm-level data by averaging the data over the years 2005-2007. This is a period short enough to reduce bias from averaging over full business cycles yet long enough to alleviate bias from using annual figures. We then employ these data in the semiparametric identification strategy described in the previous chapter. The choice of variables used is based on the parameters in equation (1) and includes age, age squared $d$, log unit costs, $c_i = \ln C_i$, and industry-specific factors, $\delta_{s(i)}$. We summarize the descriptive statistics for these variables in Table 3.

In line with much of the literature and the assumptions in Chapter 2, we estimate a constant $\theta$ across markets using $c_i$ derived from the aggregate of firms. Since this paper focuses on differences across markets by separately analyzing the sales of exporters versus domestic sales, the variables included in the table are differentiated according to destination markets for illustrative purposes.

What stands out in the table is the following: (i) the dispersion regarding total factor productivity is high, accruing to a few outliers as shown by the percentiles displayed in the table, and (ii) a large number of firms does not engage in R&D activity as measured by the intangibility ratio $r_i$ (i.e., $r_i = 0$) at all or do not report any tangibles in their balance sheet. Moreover, domestic sales as well as total sales exhibit a higher mean and a smaller standard deviation than export sales. Export sales have a lower mean than domestic sales of exporters as well as of non-exporters. However, the 95th percentiles of the foreign and domestic sales of exporters are larger than the ones of domestic sales of non-exporters, which is consistent with previously found patterns (e.g., Bernard, Jensen, Redding, and Schott, 2007). Histograms of the four outcomes shown in Figure 2 provide further evidence: most firms do not export (or they export very little), and of those who export, a relatively small share of their sales is exported.

In addition to the variables shown in Table 3, we employ binary indicator variables for 23 industries at the NACE Rev.2 2-digit level. Dividing eq. (4) by $\sqrt{2}$, in order to maintain the scale of the variance of the undifferenced model, yields

$$\frac{(x_i - x_{i-1})}{\sqrt{2}} \approx \theta \frac{(c_i - c_{i-1})}{\sqrt{2}} + \frac{(\delta^t_{s(i)} - \delta^t_{s(i-1)})}{\sqrt{2}} + \frac{(u_i - u_{i-1})}{\sqrt{2}}.$$

We estimate $\theta$ and the coefficients on the sector indicators using a control function approach that seeks to overcome the endogeneity problem associated with unit production costs. In a first stage, we regress $(c_i - c_{i-1})$ on the differenced age, age...
Table 3: Descriptive statistics for semiparametric identification strategy

<table>
<thead>
<tr>
<th></th>
<th>Obs.</th>
<th>Mean</th>
<th>SD</th>
<th>p5</th>
<th>p95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log total sales</td>
<td>55038</td>
<td>2.76</td>
<td>1.56</td>
<td>0.61</td>
<td>5.65</td>
</tr>
<tr>
<td>Productivity</td>
<td>55038</td>
<td>0.61</td>
<td>3.19</td>
<td>0.31</td>
<td>0.95</td>
</tr>
<tr>
<td>Intangibility ratio</td>
<td>55038</td>
<td>0.22</td>
<td>0.30</td>
<td>0</td>
<td>0.87</td>
</tr>
<tr>
<td>Age</td>
<td>55000</td>
<td>21</td>
<td>18</td>
<td>3</td>
<td>53</td>
</tr>
<tr>
<td>Log unit cost</td>
<td>55038</td>
<td>3.84</td>
<td>0.39</td>
<td>3.23</td>
<td>4.43</td>
</tr>
<tr>
<td>Log exports</td>
<td>27846</td>
<td>0.38</td>
<td>2.86</td>
<td>-4.60</td>
<td>5.11</td>
</tr>
<tr>
<td>Productivity</td>
<td>27846</td>
<td>0.61</td>
<td>4.42</td>
<td>0.31</td>
<td>0.89</td>
</tr>
<tr>
<td>Intangibility ratio</td>
<td>27846</td>
<td>0.19</td>
<td>0.27</td>
<td>0</td>
<td>0.83</td>
</tr>
<tr>
<td>Age</td>
<td>27829</td>
<td>25</td>
<td>20</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>Log unit cost</td>
<td>27846</td>
<td>3.82</td>
<td>0.36</td>
<td>3.27</td>
<td>4.34</td>
</tr>
<tr>
<td>Log domestic sales of exporters</td>
<td>27862</td>
<td>3.11</td>
<td>1.56</td>
<td>0.83</td>
<td>5.84</td>
</tr>
<tr>
<td>Productivity</td>
<td>27862</td>
<td>0.61</td>
<td>4.42</td>
<td>0.31</td>
<td>0.89</td>
</tr>
<tr>
<td>Intangibility ratio</td>
<td>27862</td>
<td>0.19</td>
<td>0.27</td>
<td>0</td>
<td>0.83</td>
</tr>
<tr>
<td>Age</td>
<td>27845</td>
<td>25</td>
<td>20</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>Log unit cost</td>
<td>27862</td>
<td>3.82</td>
<td>0.36</td>
<td>3.27</td>
<td>4.34</td>
</tr>
<tr>
<td>Log domestic sales of non-exporters</td>
<td>35308</td>
<td>2.24</td>
<td>1.33</td>
<td>0.41</td>
<td>4.52</td>
</tr>
<tr>
<td>Productivity</td>
<td>35308</td>
<td>0.61</td>
<td>2.45</td>
<td>0.30</td>
<td>0.99</td>
</tr>
<tr>
<td>Intangibility ratio</td>
<td>35308</td>
<td>0.24</td>
<td>0.31</td>
<td>0</td>
<td>0.88</td>
</tr>
<tr>
<td>Age</td>
<td>35282</td>
<td>19</td>
<td>16</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td>Log unit cost</td>
<td>35308</td>
<td>3.86</td>
<td>0.41</td>
<td>3.20</td>
<td>4.47</td>
</tr>
</tbody>
</table>

Source: Bureau van Dijk Amadeus database. Variables measured in real terms. (Obs.=number of non-missing observations, SD=standard deviation, p5=5th percentile, p95=95th percentile). All variables are averages over the period 2005-2007.
Figure 2: Histogram of dependent variables

(a) Log export sales
(b) Log domestic sales of exporters
(c) Log domestic sales of non-exporters
(d) Log total sales
Table 4: Difference equation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i - x_{i-1}$</td>
<td></td>
</tr>
<tr>
<td>$c_i - c_{i-1}$</td>
<td>$-122.786$</td>
</tr>
<tr>
<td></td>
<td>$(28.601)***$</td>
</tr>
</tbody>
</table>

Observations 1227
Bootstrap replications 100

Notes: The parameter is based on a control function approach using data averages for 2005-2007. All variables are differenced with respect to the nearest neighbor in terms of both total factor productivity and the intangibility ratio and divided by $\sqrt{2}$. Results from the first stage and coefficients of 23 (jointly significant) differenced industry indicators are suppressed for the sake of brevity. The second stage is estimated in a seemingly unrelated regression, where the (theory-consistent) coefficient of $(c_i - c_{i-1})$ is constrained to be equal across destination markets $\ell$. Bootstrapped standard errors reported.

squared, and the sector indicators. This obtains a differenced residual which is included in the second stage together with $(c_i - c_{i-1})$ and the sector indicators. The second stage is estimated in a seemingly unrelated regression, where the theory-consistent coefficient of $(c_i - c_{i-1})$ is constrained to be equal across destination markets $\ell$. Since first differencing requires densely-enough populated data and a smooth function $g^t(\cdot)$ to remove the nonparametric effect, we drop observations in the 99th percentile regarding $\hat{\phi}_i$, and in the 10th and 90th percentile regarding $r_i$ from our sample. Results are reported in Table 4. As theory would suggest, $\hat{\theta}$ carries a negative sign, being significant at the 1%-level, and amounts to about $-123$. Essentially, this suggests that firms behave more or less in a perfectly competitive way. However, part of the effect might be captured by the sector-specific effects, and, at this point, we are not interested in discerning the two but mainly in estimating residual demand. With $\hat{\theta}$ and estimated fixed sector-specific effects at hand, we can move on to remove the estimated parametric effect and estimate $g^t(\hat{\phi}_i, r_i)$ on exports consistently in the next subsection, which is at the heart of this paper’s interest.

4.3 Estimating the nonparametric impact of total factor productivity and the intangibility ratio on exports, $g^F(\hat{\phi}_i, r_i)$

As outlined in Section 2, R&D and TFP might affect not only firm-level efficiency but also consumer (product/firm specific) preferences and as a consequence, $\hat{\xi}_F^t$ may depend on both $\hat{\phi}_i$ and $r_i$. To study the structural relationship between the vari-
ables of interest, estimating $g^F(\cdot)$ nonparametrically from a bivariate product kernel-weighted local linear regression as in equ. (7), we remove the estimated parametric effect as obtained in the previous subsection from $x^F_i$ to obtain $\xi^F_i \equiv x^F_i - \hat{\theta}_c - \hat{\delta}^F_s$. Prior to that, we eliminate base-period prices in $\hat{\phi}_i$ accordingly by using the residual of a regression of $\hat{\phi}_i$ on the sector indicators in the remaining part of this paper instead of $\hat{\phi}_i$.\(^3\) We smooth the data in both directions (total factor productivity and asset intangibility) using cross-validation and perform a bivariate product kernel local linear regression (see Fan and Gijbels, 1996). Results from estimating $g^F(\cdot)$ for firm-level exports at an optimal bandwidth $h^F_\hat{\phi} \approx 2.0$ and of $h^F_\hat{\phi} \approx 0.2$ are illustrated in Panel (a) of Figure 3.

Figure 3 suggests a more or less log-linear impact of total factor productivity as well as the intangibility ratio on firm-level exports. However, the figure suggests a nonmonotonicity at relatively high levels of $\hat{\phi}_i$.\(^4\) To that end, this is good news for economic theory. The slope of log exports with regard to $\hat{\phi}_i$ is about 12.11. The latter is a measure of $\hat{\theta}$, which is much different from the parameter on $(c_i - c_{i-1})$ in the previous subsection. However, that could be due to the multi-collinearity of $(c_i - c_{i-1})$ with the differenced sector-specific indicator variables, as said before. The slope with regard to the intangibility ratio is about 0.59.\(^5\)

### 4.4 Estimating the nonparametric impact of total factor productivity and the intangibility ratio on domestic sales, $g^H(\hat{\phi}_i, r_i)$, and total sales, $g^T(\hat{\phi}_i, r_i)$

Heterogenous firms literature suggests $g^f(\cdot)$ to behave similarly regarding exporters as well as domestic firms. We shed light on this prediction by accounting for three additional outcomes: the domestic sales of exporters, domestic sales of non-exporters, and total sales. With data summarized in Tables 3 and 4 at hand, we use the same

---

\(^3\)Note that we assume sector-specific prices in general. While we have deflated all relevant variables, $\hat{\phi}_i$ still exhibits prices of the base period. The variable $r_i$ is measured by a ratio and prices thus cancel out. The notation regarding the adjusted $\hat{\phi}_i$ will remain unchanged.

\(^4\)We estimate confidence bands around the estimates in Figure 3 by following Yatchew (2003). Specifically, we perform oversmoothing using the generic bandwidth $h_z \times 1.1$ to obtain fitted values $\xi^+_i$ and undersmoothing using $h_z \times 0.9$ to obtain residuals $\nu^-_i$. We then apply a bootstrap procedure with heteroskedastic residuals (wild bootstrap) by resampling 100 times and estimating $g^f(\cdot)$ with $\hat{\xi}_i = \xi^+_i + \nu^-_i$.

\(^5\)Notice that, in the data, $\sigma_{\text{export}} = \sigma = 1 - \theta \approx 19$ according to a Melitz-type model when considering all firms. This is close to slope of positive log exports with regard to $\hat{\phi}_i$, which is 19.7. Hence, taking theory seriously, whereby the parameter on preferences or quality would be $\hat{\theta}$ as well, we would say that the intangibility ratio is related to quality or preferences at a parameter of about $0.59/\hat{\theta} \approx 0.005$. 

12
Figure 3: Fitted residual exports ($\hat{\xi}_i$)

(a) Fitted residual exports

(b) 95% confidence intervals

Plots are based on an optimal (cross-validation) bandwidth of 2 for $\hat{\phi}_i$ and 0.2 for $r_i$. Estimated surfaces based on 50×50 grid points. Individual observations are shown in the upper panel.
approach as described previously and illustrate the results in Figures 4–6. The figures reveal a number of novel empirical patterns. First of all Figures 6 and 5 are less smooth (based on optimal bandwidths) and therefore, exhibit spikes that are more pronounced. Second, in line with the literature, the density of observations suggests a log-linear impact of TFP and R&D in all Figures. Third, Figure 4 regarding domestic sales of exporters is qualitatively identical to export sales as shown in the previous chapter, confirming the predictions of a number of models. The average gradient is somewhat lower than for exports, amounting to 12.11 regarding $\hat{\phi}_i$ and to 0.60 regarding $r_i$. Fourth, Figure 5 illustrates that this is not the case with respect to the sales of non-exporting firms, where $\xi_i$ is neither linear in TFP nor in R&D, and the relationship is not necessarily positive either, which is clearly in contrast to standard model predictions. Panel (b) of Figure 5 suggests that the relationship is non-monotonic at increasing $\hat{\phi}_i$. In addition, the effect of R&D on $\xi_i$ is not clearly linear either, but $\hat{\xi}_i$ lies within the confidence bounds at lower levels of R&D. The average gradient of domestic sales of non-exporters is much lower compared to export sales; it is 5.25 regarding $\hat{\phi}_i$ and negative, amounting to $-1.04$ regarding $r_i$. Finally, the inspection of Figure 6 regarding total sales suggests that in contrast to the relationship regarding productivity, the relationship between asset intangibility (R&D) and domestic sales is more or less linear. In addition, the confidence bands in the lower panel of the figure suggest a different pattern than is found for exports and domestic sales, pointing to a spurious relationship between the variables of interest. Taking the different patterns for exporters and non-exporters into account, it is not surprising that an aggregate consideration can not reveal an unambiguous and significant relationship. This points to the need of distinguishing between destination markets in general.

Overall, the separate analysis of exports versus domestic sales of exporters and non-exporters supports the following. While the results for exports are well in line with Melitz (2003) the results for non-exporters suggest a different conclusion. Our results seem to indicate a non-monotonic relationship between TFP and sales for non-exporters inconsistent with the implications of Melitz (2003) where non-exporters are less productive but still produce output that is positively and logarithmically dependent on productivity. At first glance, the similarity between domestic sales of exporters and their exports is striking, especially when comparing these patterns to the sales of non-exporters. The obvious difference for the same destination market suggests a structural difference between exporters and non-exporters beyond productivity. Our results confirm that productivity explains which firms among exporters yield higher sales irrespective of the destination market, however, productivity alone cannot explain which firms export. In order to identify additional determinants of exports it is worthwhile to have a closer look at lower panels of Fig-
ures 4 and 5. Domestic sales of exporters are clearly increasing in productivity in a log-linear way for relatively high levels of R&D. Sales of non-exporters are, on the other hand, non-linear especially for lower levels of R&D. Hence, the role of R&D might be an important ingredient in determining which firms become exporters and which firms do not in a new theoretical framework. In general, trade theory would suggest that the relationship between sales and product-related preferences, or quality features, exhibits a functional form similar to the one between productivity and sales. Taking our results into consideration, such a relationship and its link to R&D could be examined in more detail.

5 Conclusions

This paper examines whether the standard assumption from trade models indicating a linear relationship between exports and domestic sales, and productivity as well as quality, or preferences, holds. We have used French firm-level data to allow for a flexible relationship between productivity, R&D, and sales to different destination markets within a semiparametric framework. While our results confirm that in line with the literature, a log-linear pattern holds for both exports and domestic sales of exporters, this is not the case for the sales of non-exporters. The latter are characterized by a non-monotonic relationship between TFP and sales, hence, we find evidence against the prediction about a similar pattern across destination markets. Furthermore, domestic firms differ from exporters in the sense that the relationship between TFP and sales is non-linear for lower levels of R&D. Our results suggest that these differences should be taken into account in future models of trade.

References


Figure 4: Fitted residual domestic sales of exporters ($\hat{\xi}_i$)

(a) Fitted residual domestic sales of exporters

(b) 95% confidence intervals

Plots are based on an optimal (cross-validation) bandwidth of 2 for $\hat{\phi}_i$ and 0.2 for $r_i$. Estimated surfaces based on $50 \times 50$ grid points. Individual observations are shown in the upper panel.
Figure 5: Fitted residual domestic sales of non-exporters ($\hat{\xi}_i$)

(a) Fitted residual domestic sales of non-exporters

(b) 95% confidence intervals

Plots are based on an optimal (cross-validation) bandwidth of 0.4 for $\hat{\phi}_i$ and 0.2 for $r_i$. Estimated surfaces based on 50x50 grid points. Individual observations are shown in the upper panel.
Figure 6: Fitted residual total sales ($\hat{\xi}_i$)

(a) Fitted residual total sales

(b) 95% confidence intervals

Plots are based on an optimal (cross-validation) bandwidth of 0.6 for $\hat{\phi}_i$ and 0.2 for $r_i$. Estimated surfaces based on 50×50 grid points. Individual observations are shown in the upper panel.


