Firm Innovation, Productivity, and Trade

Bruno Cassiman *
IESE Business School and
University of Leuven

Elena Golovko †
Tilburg University

David Greenstreet ‡
University of Oxford and
Centre for Economic Performance

January 4, 2008
Preliminary and incomplete. Comments welcome.

Abstract

Firm innovation effort, productivity, and export propensity are found to be positively correlated with one another. Two alternative explanations have been offered for this — innovation increases productivity and more productive firms self-select into exporting or, alternatively, exporting increases learning and innovation, which in turn affects productivity. We use differences in timing implied by these hypotheses to disentangle their importance in a representative panel of Spanish manufacturing firms from 1990-1999. Innovation effort and exporting are allowed to affect estimated firm productivity dynamics. Firms forecast own productivity from past production experience and previous innovation and exporting. Productivity levels and dynamics are identified using errors in the firms' productivity forecasts. Then estimates of the forecastable component of productivity are related to firm innovation and export decisions.

1 Introduction

Firm productivity and export propensity have been found to be positively correlated in many settings. At least two alternative explanations are suggested in the literature for

*BCassiman@iese.edu
†e.golovko@uvt.nl
‡david.greenstreet@economics.ox.ac.uk

this positive association. On the one hand, the positive export-productivity association is explained through a selection of more productive firms into exporting. Sunk start-up costs associated with becoming an exporter and intense competition in the foreign markets lead to the self-selection of more productive firms into exporting. On the other hand, an alternative explanation for this positive correlation is the reverse effect by which firms actually learn by exporting, i.e. exporters learn from their foreign contacts adopting new products and production technologies and thereby increasing productivity.

Empirical studies by and large support the selection hypothesis for the export-productivity link (Roberts and Tybout 1997; Clerides, Lach and Tybout 1998; Bernard and Jensen 1999; Greenaway and Kneller 2000; Aw, Chen and Roberts 2001; Girma et al. 2002; Delgado, Farinas et al. 2002; Fafchamps, El Hamine et al. 2007). The general finding is that exporting firms have higher productivity than non-exporters before taking up exports, and the productivity advantage of the continuous exporters does not increase over time compared to non-exporters. Yet, a few studies find evidence in accordance with the learning-by-exporting hypothesis (Kraay 1999; Bigsten et al. 2002; Castellani 2001; Girma et al. 2002), showing that firms gain productivity advantages after they start exporting.

With both mechanisms being plausible, the observed heterogeneity in productivity raises an important question about the drivers of the high productivity of the exporting firms. One important source of productivity differences seems to be related to R&D and innovation activities (See Griliches, 1998). A number of empirical studies have documented the positive and significant effect of R&D and innovation on firm productivity and productivity growth. Crepon et al. (1998) estimating a structural model that links productivity, innovation output and innovation inputs, find that firm productivity correlates positively with higher innovation output. In line with their result, Jefferson et al. (2004) show for Chinese firms that new product sales are positively associated with productivity. Huergo and Jaumandreu (2004) using the panel of Spanish firms find that process innovation is an important determinant of productivity growth at the firm level. Investigating the relationship between innovation and productivity in four European countries, Griffith et al. (2006) find — consistent with the previous studies — that both product and process innovations have a significant positive effect on firm-level productivity in three out of the four countries. Finally, Doraszelski and Jaumandreu (2007) revisit the knowledge capital framework within an extension of Erickson and Pakes (1995) and find important effects of R&D investments on productivity.

At the same time, R&D and innovation activities seem to play an important role in explaining a firm's decision to export and export volumes. In particular, recent studies find that innovation is an important driver of the export decision. Basile (2001) for a sample of Italian manufacturing firms shows that firms introducing product and/or process innovations either through R&D or through investments in new capital are more likely to export. Bernard and Jensen (2004) find that firms switching primary SIC code — which could indicate new product introductions — significantly increase the probability of entering the export markets. In a related paper, Cassiman and Martinez-Ros (2007) find that product innovation is an important determinant of the firm's exporting activities.

Moreover, exporting may also affect firm innovation output through learning. A number of studies document a positive effect of exporting on innovation. Alvarez and Robertson (2004) for samples of Chilean and Mexican plants find a positive relationship between exporting and probability of innovating for various measures of innovation activities such as investments in improving the design, production processes, or product quality, the presence of R&D activities, and purchases of foreign technical licenses. Salomon and Shaver (2005), examining the effect of exports on the number of product innovations and patent applications, find that exporting is positively associated with the subsequent increase in the number of product innovations and the number of patent applications.

Taken together, prior empirical findings suggest that innovation activity may be responsible for productivity enhancements and at the same time lead to and be influenced by exporting. Hence, firm innovation decisions may be an important factor in explaining the correlation between exports and productivity. Recent studies investigating the relation between exports, productivity, and innovation find empirical support for this argument. Cassiman and Martinez-Ros (2007) find a strong positive effect of product innovation on the decision of a firm to export. Firms with product innovation are significantly more likely to become exporters that non-innovators. In a related paper, Cassiman and Golovko (2007) show that product innovation at least partially accounts for the positive link between exports and productivity. Once controlling for the fact that the firm realized a product innovation, this firm's productivity turns out to depend less on whether or not a firm participates in the export market. In particular, only a small difference is observed in productivity levels between exporters and non-exporters among firms that carry out product innovation.

In this paper, we aim to disentangle the importance of selection versus learning hypotheses behind the export-productivity link. Innovation activity is introduced as a possible explanatory variable driving the export-productivity correlation. Innovation increases productivity and more productive firms self-select into exporting, or, alternatively, learning-by-exporting influences innovation, which in turn positively affects productivity. Using a panel of Spanish manufacturing firms, we exploit differences in timing between innovation, exporting and their effect on productivity dynamics to sort out selection and learning.

2 Data

The data that we use in this paper come from a survey of Spanish manufacturing firms (ESEE) covering 1990-2002. This project was conducted by the Fundacin Empresa Pblica with financial support of the Ministry of Science and Technology. The sample includes the population of Spanish manufacturing firms with more than 200 employees and a stratified sample of small firms comprising 4% of the population of small firms with more than 10 and less than 200 employees. Small firms that exit the original sample during the period of 1990-2002 are replaced by firms with similar characteristics drawn from the population.[1] The original sample includes 2188 firms in 1990 and 3462 firms in 2002 from 20 distinct industries.

Due to missing values for the analysis variables and gaps in time coverage of individual firms, the resulting dataset is an unbalanced panel of 315 firms and 2934 firm-year observations. In our analysis, we focus on three industrial sectors: Textiles, Leather and footwear; Chemical products, Plastic and rubber products; Metallurgy and Metallic products. The choice is based on their importance in the Spanish trade balance as well as on the number of observations.

The ESEE data is an appropriate setting to test the relationship between innovation, exports and productivity we intend to investigate. Exporting firms constitute a large proportion of the sample and very few firms (less than 0.3%), especially among small and medium firms, engage in foreign direct investment. In this way we are able to focus particularly on exporters. Also, during the 90s Spain has experienced an entire business cycle, with growth in the beginning of the 90s, sharp recession in 1993, and a recovery during the last years of the sample period. We can usefully exploit such variation to examine the productivity dynamics during 1990-2002.

Table 1 provides some descriptive statistics for our sample computed for 1990-2002. We can see that the industries are quite different in terms of export and innovation activities, with the highest numbers for the Chemical products, Rubber and plastics sector. Also, the average firm size differs significantly across the sectors, from 122 employees in textiles to over 250 in the chemical products sector. Similar variation can be observed for the values of output, labor, capital and material inputs, with the highest values in the chemical products industry.

3 Empirical Approach

Our empirical approach comprises five equations — a production function with firm-specific productivity residual, determinants of productivity dynamics, innovation effort, hazard of innovation realization, and exporting hazard.

First the production function and productivity dynamics are estimated using a modification of the sequential learning productivity estimator proposed in Greenstreet (2007). Estimation is based on the difference between realized firm productivity and its forecastable component under information and timing assumptions. This technique is tantamount to an empirical implementation of Jovanovic (1982), in which firms use prior production experience to fore-

[1] Proportion of the firms in the year t that continue in the survey in the year t+1 is approximately 90% for 1990-2001. Among the firms that exited the sample, approximately 2.2% disappeared and approximately 7.7% stopped collaborating.

cast productivity when making decisions for the next period. The model and estimation both rely heavily on Kalman filter theory. The most significant feature for this study is that the effects of endogenous innovation and export status on productivity can be estimated without regard to their determinants. Estimated firm productivity can then be used as a generated regressor in the three remaining equations.

The next pair of equations are innovation effort and the innovation hazard. Innovation effort is a firm choice made in anticipation of future profitability in the context of a dynamic game among competing firms (e.g. as in Ericson and Pakes 1995). Consequently, innovation effort may be a function of firm state variables known at the beginning of a period — the forecastable portion of productivity, export status, the stock of previous innovations,[2] capital, age, ownership, etc. — as well as external conditions — such as market conditions and industry state — which may be controlled with a set of year dummies. In the absence of an explicit specification of the dynamic industry game, we use a linear approximation of this innovation effort policy function. The probability of a new innovation arriving is a function of prior innovation effort, prior innovations, and lagged firm state variables, excluding productivity. In particular, we allow previous export status to affect innovation probability. Thus, in addition to the direct channel from exporting to productivity, there are two potential indirect channels via innovation effort and innovation hazard. Separate identification of these three channels depends upon both the observability of innovations and differences in the assumed timing. The data record both product and process innovations.

Finally, exporting itself is a recurring binary decision with persistence. This is modeled as a function of lagged export status, the forecastable portion of productivity, prior innovations, lagged capital,[3] and exogenous conditions. The role of productivity in this equation captures the selection explanation for the association between productivity and exporting. The restriction to the forecastable portion of productivity prevents simultaneity bias with the productivity dynamics equation. This will be estimated as a dynamic probit.

Note that the timing assumptions built into each of these equations are crucial in avoiding or resolving issues of endogeneity and selection. The present specification also relies on absence of serially correlated errors in the innovation effort, innovation hazard, and export hazard equations. We intend to explore ways of relaxing this assumption.

The current iteration of this paper only reports preliminary results for the first pair (production and productivity dynamics) of equations. This allows us to examine the direct channel of learning-by-exporting's impact on productivity. We also capture the portion of the indirect channels due to the response of productivity to innovations.

[2]We assume that the arrival of new innovations is known at the beginning of the period and may therefore affect current innovation effort.

[3]Current capital is determined by the firm simultaneously with export status. If firms invest in anticipation of becoming exporters in the future (see Iacovone and Javorcik 2007), either the lags on both capital and (the predictable portion of) innovation would have to be increased or the estimates would have to be interpreted as partially reduced forms.

3.1 Estimating Productivity Dynamics

Let i be an index of firms and t indicate year. Let $t = 1$ be the first year in the panel data set, and let $t = T$ be the last period in the data. Define T_i = min[last year of establishment i's operation, T] as the last year of observed data for establishment i, and c_i = max[first year of establishment i's operation, 1] the first.

Each active firm produces output,[4] q_{it}, in year t using a vector of inputs, $\mathbf{x}_{it}$, consisting of capital,[5] labor,[6] and materials.[7] In logarithms the Cobb-Douglas production function is

$$q_{it} = \alpha_t + \beta \mathbf{x}_{it} + \mu_{it} + \xi_{it} \qquad \xi_{it} \sim \text{NIID}(0, \sigma^2) \tag{1}$$

The sequence of constants, α_t, will absorb changes in productivity common across firms, including those that may be attributable to macroeconomic forces.[8] The total residual, $\epsilon_{it} \equiv \mu_{it} + \xi_{it}$, includes a persistent, firm-specific productivity component, μ_{it}, and a normal iid noise term, ξ_{it}. μ_{it} is not directly observable by either the firm or the econometrician. The ϵ_{it} are observable at the end of each period, conditional on parameter values for the econometrician, and are noisy signals of the latent μ_{it}s. In general, μ_{it} may be a vector (times an aggregation vector) if idiosyncratic productivity has several components following separate dynamics. We currently model it as a scalar. Due to the dynamics of the μ_{it} as described below, ϵ_{it} will not be iid. In Kalman filter terminology this is known as the measurement equation.

Each period the persistent portion of firm productivity is hit with a shock, η_{it}. It will also be shifted by a set of variables, $\mathbf{e}_{it}$, which may contain other firm characteristics, innovations, or lagged innovation effort directly. We include dummy variables for arrival of a product innovation, a new process innovation, export status, and change in export status (start or stop). Thus, μ_{it} evolves according to the transition equation

$$\begin{aligned} &\mu_{it} = R\mu_{it-1} + \gamma \mathbf{e}_{it} + \eta_{it} \qquad \text{for } t > c_i \\ &\eta_{it} \sim \text{NIID}(0, Q) \end{aligned} \tag{2}$$

R captures the degree of period-to-period persistence in μ. γ is a vector of parameters representing the change in productivity due to the variables in $\mathbf{e}_{it}$. So, for example, the impact of a value of one for the new exporter dummy would be γ the first period, $R\gamma$ the second, then $R^2\gamma$ and so forth. The total first period impact of becoming an exporter would be $\gamma_j + \gamma_k$ where j and k represent the vector indices of the new exporter and exporter dummies. When interpreting the coefficient for ongoing exporter status, the role of R in reducing persistence of the prior impact must be taken into account.

[4]Sales plus change in inventory deflated with a *firm-specific* price index. Thus, we avoid biases arising from residual cross-sectional price variation as discussed in Klette and Griliches (1996).

[5]Constructed from a perpetual inventory of investment.

[6]Number of workers times average hours worked. Not quality adjusted.

[7]Raw materials, fuel and electricity costs, and purchased services deflated with a firm-specific price index.

[8]However, we do not account for potential interactions between macroeconomic phenomena and firm characteristics in the determination of productivity.

The objective of the sequential learning estimator is to remove endogeneity and selection biases from the estimation of Equations 1 and 2 by modeling firms' forecasts of productivity and using the errors in those forecasts for estimation. If the forecast errors are not in a firm's information set at the time it makes decisions about production inputs, innovation effort, export status, or exit, then they must be uncorrelated with those decision. This holds regardless of how firms make decisions with the information that is known to them. The same logic applies to consequences of previous actions and any predetermined firm characteristics. We make assumptions sufficient to recover these productivity forecast errors conditional on the parameter values:

Information: All establishments update their productivity beliefs (as if) their information is the same as that revealed from production experience through $t-1$. In other words, the sequence of ϵ_{it}s from a firm's founding through $t-1$ are sufficient statistics for its information at the beginning of t about μ_{is} for $s > t-1$.

Decision Timing: Establishments' decisions on exit and investment between $t-1$ and t, and on exporting, innovation effort, and all variable inputs for production during t, are taken before any information about ϵ_{it} is revealed. In econometric terminology, these decisions are predetermined.

Linearity: The stochastic components of Equations 1 and 2 enter linearly.

Under these conditions, the Kalman filter representation of firm learning described below generates the best[9] linear forecast of μ_{it}. When normality is assumed as well, the Kalman filter forecast becomes the minimum mean square estimator of μ_{it} and conditionally Gaussian. Thus the forecast errors are Gaussian and independent of all predetermined firm decisions and characteristics.[10,11]

To apply the Kalman filter to an establishment's productivity learning problem, define $E_t[\mu_{it}] \equiv u_{it}$ and $E_{t-1}[\mu_{it}] \equiv u_{it|t-1}$ where the subscript on the expectation operator indicates conditioning on information available up to and including t or $t-1$, respectively. Also, let P_{it} and $P_{it|t-1}$ denote corresponding variances.

Deferring the question of initial conditions to below, each iteration of the Kalman filter from $t-1$ to t, separately for each establishment i, begins with prediction equations

$$u_{it|t-1} = Ru_{it-1} + \gamma \mathbf{e}_{it} \tag{3}$$

With variance

$$\text{var}(\mu_{it} - u_{it|t-1}) \equiv P_{it|t-1} = RP_{it-1}R' + Q \tag{4}$$

[9] In the minimum mean square error sense.

[10] Greenstreet (2007) has further discussion. For the underlying statistical theory see e.g. Harvey (1989) or Lipster and Shiryaev (2001).

[11] Also note that the process of learning about a firm's own productivity described here has nothing to do with experimentation, nor with strategic interactions and hidden types. In our model, barring exit, nothing the firm does changes the information content of the signals, ϵ_{it}.

Compare this to Equation 2.

Once production experience in period t is observed, the establishment can compute its prediction error

$$\begin{aligned} v_{it} &\equiv q_{it} - \mathrm{E}_{t-1}[q_{it}] \\ &= [\alpha_t + \beta \mathbf{x}_{it} + \mu_{it} + \xi_{it})] - [\alpha_t + \beta \mathbf{x}_{it} + u_{it|t-1}] \\ &= (\mu_{it} - u_{it|t-1}) + \xi_{it} \end{aligned} \tag{5}$$

Viewed as a random variable, v_{it} has a variance

$$f_{it} = P_{it|t-1} + \sigma^2 = RP_{it-1}R' + Q + \sigma^2 \tag{6}$$

Given the normality assumptions, $v_{it} \sim \mathrm{N}(0, f_{it})$.

Finally, the prediction error is used to update beliefs about the current value of the productivity state variable

$$u_{it} = u_{it|t-1} + (P_{it|t-1}/f_{it})v_{it} \tag{7}$$

With new variance

$$P_{it} = P_{it|t-1} - P^2_{it|t-1}/f_{it} \tag{8}$$

Under normality, Equation 7 is equivalent to Bayes updating. This can also be interpreted as apportioning the prediction error to $u_{it|t-1}$ and the noise term ξ_{it} in proportion to their variances. To see this, compare the $P_{it|t-1}$ factor in the second term to the formula for f_{it} in Equation 6. Finally, even though true productivity, μ_{it}, is unobserved, Equation 7 is a linear projection from the forecast error, v_{it}, onto the productivity.

From the v_{it} and f_{it} one can construct a prediction error decomposition[12] likelihood.

$$\log L(\Gamma : Y_{iT_i}) = -\frac{1}{2}\sum_{t=c_i}^{T_i}\left[\log 2\pi + \log f_{ct} + \frac{v^2_{it}}{f_{ct}}\right] \tag{9}$$

Where Γ is the set of all parameters and Y_{iT_i} is firm i's data in the panel. Implicitly via Equations 3 through 8, the v_{it} are functions of the parameters and data and f_{it} are functions of the variance and persistence parameters. Since there are common parameters across firms, these likelihoods are pooled.

$$\log L(\Gamma : Y_t) = -\frac{1}{2}\sum_{i=1}^{I}\sum_{t=c_i}^{T_i}\left[\log 2\pi + \log f_{ct} + \frac{v^2_{it}}{f_{ct}}\right] \tag{10}$$

Where $Y_T \equiv \{Y_{1T_1}, Y_{2T_2}, \ldots, Y_{IT_I}\}$ refers to the entire panel data set. We estimate the model parameters by maximizing this pooled likelihood. It is also possible to produce analytic

[12] So named in the Kalman filter literature because it is constructed from the sequence of prediction errors, v_{it}, each constructed conditional on its predecessors.

expressions for the gradient by differentiating the filter equations. See Greenstreet (2007) for details. Unlike Greenstreet (2007) we use a numeric Hessian in the optimization algorithm and for the empirical information matrix because the asymptotic simplifications used there are not applicable once initial-condition nuisance parameters are concentrated from the likelihood.[13]

Finally, consider the initial conditions for each firm's productivity learning process. Starting simulation of the filter requires each firm's prior belief at the beginning of the first period they are observed in the panel. Greenstreet's (2007) cohort-based solution that exploited cross-sectional variation among entrants cannot be applied to the ESEE panel. We do not observe the firms from the beginning of their production experience, and thus learning and selection have begun prior to inclusion in the panel. As an alternative we maintain the normality assumption and treat the mean and variance of each firm's prior as nuisance parameters.

$$u_{ic_i|c_i-1} = w_i \qquad\qquad P_{ic_i|c_i-1} = W_i \tag{11}$$

De Jong (1988) derives a modification of the Kalman algorithm that can be used to concentrate priors out of likelihoods such as Equation 10. He exploits the fact that Kalman filter estimates of μ_{it} can be expressed as weighted sums of the total residuals, ϵ_{it}. This is true not only for the period-ahead forecast and update, but also for the extension to "smoothed" estimates which also efficiently exploit future realizations of ϵ_{it}. In fact, the resulting $\hat{w}_i$ is equivalent to the smoothed estimate that would result from a diffuse prior at c_i. It is intuitive that the econometrician's estimate of $\hat{w}_i$ will have to exploit these future realizations of ϵ_{it} because the panel contains no information about the ϵ_{it} for $t < c_i$. In effect, both the firm and the econometrician are estimating the latent μ_{ic_i}, but with conditionally independent information sets. Consequently, with this nuisance parameter initialization the sequential learning estimates are bias-reducing relative to a naive estimator using conditions on ϵ, but they are no longer consistent.[14] Although endogeneity and selection bias are still reduced, they are no longer eliminated. Let $\epsilon^f_{it} = \mu_{it} - w_i$ at $t = c_i$ be the firm's error in predicting μ_{it}, $\epsilon^e_{i1} = \mu_{i1} - \widehat{w}_i$ at $t = c_i$ be the econometrician's error, and y_{it} be regressors. Then bias in a naive estimator is proportional to $\text{cov}(u_{it|t-1}, y_{it})$ whereas bias in the sequential learning estimator is proportional to $\text{cov}((\epsilon^e_{it} - \epsilon^f_{it}), y_{it})$. In words, only the portion of the firm's forecast that the econometrician cannot recover contributes to bias. Also, as t increases $(\epsilon^e_{it} - \epsilon^f_{it})$ decreases[15] as the firm's and econometrician's information sets increasingly overlap.

To implement de Jong's (1988) solution the standard filter is run as if $w_i = u_{ic_i|c_i-1} = 0$ and $W_i = P_{ic_i|c_i-1} = 0$ and label the resulting pseudo-forecast errors and variances $v_{it}(0,0)$ and

[13]In the original likelihood, applying iterated expectations to the prediction error decomposition's iterative structure conveniently shows that all terms involving second-order derivatives of v_{it} and f_{it} are zero in expectation. The Hessian of the concentrated likelihood includes a chain-rule term multiplying a block from the original Hessian, preventing this simplification.

[14]This is typical of estimation with nuisance parameters, e.g. fixed effects in the fixed T, large N setting. The standard fixed effects estimator is, in fact, a nested special case of our model with $Q = 0$, $R = 1$, and Equation 2 substituted iteratively into Equation 1

[15]For $R \leq 1$.

$f_{it}(0,0)$, respectively. Second, augment the filter with the following equations.

$$\begin{aligned} J_{it} &= (R - R(P_{it|t-1}/f_{it}(0,0)))J_{it-1} \\ J_{ic_i-1} &= I \end{aligned} \tag{12}$$

$$\begin{aligned} h_{it} &= h_{it-1} + J'_{it-1}v_{it}(0,0)/f_{it}(0,0) \\ h_{ic_i-1} &= 0 \end{aligned} \tag{13}$$

$$\begin{aligned} H_{it} &= H_{it-1} + J'_{it-1}J_{it-1}/f_{it}(0,0) \\ H_{ic_i-1} &= 0 \end{aligned} \tag{14}$$

Before concentration the likelihood in Equation 10 can be rewritten in terms of $v_{it}(0,0)$, $f_{it}(0,0)$, w_i, W_i, h_{iT_i}, and H_{iT_i}.[16] Maximizing this re-expressed likelihood with respect to w_i gives $w_i = (H_{iT_i})^{-1}h_{iT_i}$. Substituting for w_i gives the profiled likelihood actually programmed.

$$\log L(\Gamma : \{Y_T\}) = -\frac{1}{2}\sum_{i=1}^{I}\Bigg[\log|I + W_iH_{iT_i}| - h'_{iT_i}H^{-1}_{iT_i}h_{iT_i} + \sum_{t=1}^{T_i}\left[\log 2\pi + \log f_{it}(0,0) + \frac{v_{it}(0,0)^2}{f_{it}(0,0)}\right]\Bigg] \tag{15}$$

This leaves the prior variances, W_i, to be profiled. These are important in the sequential learning estimator because they affect the rate of belief adjustment ("gain"), and thus the productivity forecasts, via the P_{it}s and f_{it}s. If the implicit weighting[17] of the ϵ_{it}s is not efficient, $(\epsilon^e_{it} - \epsilon^f_{it})$ will be unnecessarily large and bias increase. However, the likelihood in Equation 15 is completely uninformative about W_i. As a consequence of over-fitting of the $\hat{w}_i$ the maximizing $\hat{W}_i$ is 0 regardless of the true value of W_i.

The formula for $\hat{w}_i$ is already a weighted average of the future residuals. Starting from $\hat{w}_i$ our model has the firm incorporating newly observed residuals by re-weighting. But these newly observed residuals are already weighted in $\hat{w}_i$. The extent of re-weighting depends on $\hat{W}_i$. So what is the most desirable actual weighting of the residuals? In setting $\hat{W}_i$ the most important thing in reducing bias is to make the most efficient use possible of the past residuals known to both the firm and the econometrician. Appropriate actual weight on future residuals also helps by matching as closely as possible the information from $t < c_i$ possessed by the firm but not by the econometrician.

Conveniently, $\hat{W}_i = 0$ actually produces the most efficient implicit weighting in at least some

[16]See Greenstreet (2007) or de Jong (1988).

[17]Durbin and Koopman (2001 pp 82-3) have explicit expressions for these weights.

cases.[18] So we would like to fix $\hat{W}_i = 0$. Unfortunately, at $\hat{W}_i = 0$ the sequential learning estimate of our model is nonconvergent. Specifically, one of the variance components, usually σ^2 but Q in one instance, approaches arbitrarily close to 0.[19] Other than the usual downward bias of variance estimates when estimated means are overfitted to random realizations of error terms, we do not yet completely understand this phenomenon. As a temporary expedient, we fix $\hat{W}_i = 0.04$, a modest value that produces convergence and is roughly half the magnitudes found in Greenstreet (2007).[20]

4 Productivity Results

Table 2, using our multifactor productivity estimates, shows that firms in the three Spanish manufacturing industries analyzed have the typical positive association between exporting and productivity, which we seek to analyze. The difference is especially notable among textile and leather goods firms.[21] However, there is a great deal of productivity variation among firms within each group in all three industries — some non-exporters are more productive than some exporters. The last two columns focus on the firms whose export status changes at some point in the panel. New exporters are more productive on average than non-exporters, but less so than the typical exporter. The first relation is consistent with selection, an immediate productivity effect of exporting, or parallel impacts of innovation. Productivity less than ongoing exporters could also be due to prior selection, the cumulative effect of learning-by-exporting, or increased innovation induced by access to the additional export markets. Firms that stop exporting tend to be less productive than exporters and in the case of leather and textiles are also less productive than the typical non-exporter. This points to reverse selection and may also be indicative of sunk costs in becoming an exporter.

The parameter estimates are reported in Table 3. The production function parameters, especially for textiles and leather products, indicate decreasing returns to scale. This is typical of estimates based exclusively on within-firm variation due to attenuation bias (Griliches and Mairesse, 1995). The capital intensity of chemical, plastic, and rubber products firms stands out, but in other respects the production functions themselves are unremarkable. Estimates of the noise variance σ^2 exhibit the strong attraction towards 0 discussed in the last section, although now they are convergent. In compensation, the persistent shock variance Q

[18]This is most readily apparent in the fixed effects special case (see footnote 14). In that case $\hat{W}_i = 0$ leads to $P_{it|t-1} = 0$ and $u_{it|t-1} = \hat{w}_i$ for all t.

[19]Nothing in the likelihoods of Equations 10 or 15 guarantee non-negative estimates of σ^2 or Q in finite samples or for a misspecified model. The implementing program enforces nonnegativity by actually estimating the corresponding standard deviations. This introduces a noncontinuity in the gradient at 0. Furthermore, with $\hat{W}_i = 0$ at exactly $\hat{\sigma}^2 = 0$ Equation 7 would require division by 0 when $t = c_i$ (via Equation 6).

[20]Which, by using entry cohorts, should be expected to produce larger values for W.

[21]*Levels* cannot be usefully compared across industries because they exclude the time dummies. However, within industry differences are comparable because the data underlying Table 1 can be thought of as pooled, detrended cross-sections.

is probably overestimated to accommodate overall variability in model fit. Nevertheless, the magnitudes are instructive. The square roots (0.36, 0.46, 0.37 respectively) are not much smaller than the cross-sectional deviations reported in Table 2. Whether as transient noise or persistent shocks, random variation[22] is a major source of heterogeneity in firm productivity. Estimates of R indicate that the shocks are not permanent, but they are quite persistent in textile and leather products and somewhat in chemical, plastic, and rubber products.

The results on determinants of productivity transition are weak or insignificant, but hint at important differences across the three industries. Limited variation may account for the insignificant results for firms that began or halted exporting. Our data includes just 22, 29, and 34 instances of firms (by industry respectively) beginning[23] export while in the panel and 16, 23, and 24 instances when export halted. On the other hand, colinearity is unlikely to be a major issue. Among the productivity transition regressors, in all three industries the largest (in absolute value) bivariate correlation is between product and process innovation. These correlation coefficients are 0.33, 0.39, and 0.34, respectively.

The coefficients on process innovation are economically small — a 1-3% average change in productivity — as well as statistically insignificant. There could be several explanations. Adopting a new or modified process may have a temporary negative effect, masking a positive long-run impact on productivity. Including lagged process innovations in the transition equation might detect this. Second, heterogeneity in the magnitude of these innovations may be inducing a downward measurement bias. The next iteration of this paper will examine these possibilities. Finally, it could be that adopting specific process innovations simply doesn't have large productivity effects in these industries.

The coefficient on product innovation is positive in all three industries and is marginally significant in the chemical, plastic, and rubber products industry. This is consistent with the studies of productivity and innovation discussed in the introduction. Our results are conditional on exporter status, suggesting a direct channel from product innovation to productivity irrespective of a firm's export status. This is one half of a potential explanation for the productivity-exporter association. If in continuing this research we confirm Cassiman and Martinez-Ros' (2007) finding that product innovation increases the probability of becoming an exporter, then innovation motivated by non-trade considerations may induce a portion of the positive relation between exporting and productivity.

Status as an exporter is only (marginally) significant in textiles and leather, and is even slightly negative in chemical, plastic, and rubber products. So we have modest evidence for learning-by-exporting in textiles and leather. Furthermore this is an effect that does not operate through innovation, but is direct or through some channel not represented by our regressors. Using point estimates and taking account of the coefficients for Began Export and R, the productivity impact in the first year of exporting is 12.8% (6.1+6.7). In the second year of exporting this increases to 17.3% (12.86*0.867+6.14). In the long run, the

[22]Or forces not accounted for in our model.
[23]Or resuming

estimated effect of exporting on productivity of a textile or leather firm approaches 46.2% (computed from $6.14/(1-R)$).[24] This is substantial, but doesn't fully account for the 0.87 difference between exporters and non-exporters reported in Table 2. It is also less than the within-group standard deviations in that table.

Thus far, we have found statistically modest evidence for three conclusions: 1) A direct channel for learning-by-exporting in textiles and leather products; 2) A positive effect of product innovation on productivity, which may constitute half of an innovation-driven explanation of the export-productivity association, in chemical, plastic, and rubber products; and 3) Heterogeneity across industries in the relationships among exporting, innovation, and productivity. This last result suggests that it may be fruitful to consider aspects of industry structure and technology more closely when studying the relative importance of selection into exporting, learning-by-exporting, and joint causation by product innovation.

[24]This is actually an understatement. Strictly speaking we should exponentiate the log scale this far from 0.

5 References

Alvarez, R. and R. Robertson 2004. "Exposure to Foreign Markets and Firm-Level Innovation: Evidence from Chile and Mexico." Journal of International Trade and Economic Development 13 (1): 57-87

Aw, B.Y., Chen, X., and Roberts, M. J., 2001. Firm-level Evidence on Productivity Differentials and Turnover in Taiwanese Manufacturing. Journal of Development Economics 66, 51-86.

Basile, R., 2001. Export behaviour of Italian manufacturing firms over the nineties: the role of innovation. Research Policy 30 (8), 1185-1201.

Bernard, A., Jensen, J., 1999. Exceptional exporter performance: cause, effect or both?. Journal of International Economics 47, 1-25.

Bernard, A., Jensen, J., 2004. Why some firms export. The Review of Economics and Statistics 86 (2).

Bigsten, A., Collier, P., Dercon, S., Fafchamps, M., Gauthier, B., Gunning, J.W., Habarurema, J., Oduro, A., Oostendorp, R., Pattillo, C., Soderbom, M., Teal,F. and Zeufack, A. (2002). Do African Manufacturing Firms Learn from Exporting?, Oxford University, Centre for the Study of African Economies Working Paper Series, WPS/2002-09.

Cassiman, B., Golovko, E., 2007. Innovation and the export-productivity link. IESE working paper, mimeo.

Cassiman, B., Martinez-Ros, E., 2007. Product Innovation and Exports: Evidence from Spanish Manufacturing. IESE working paper, mimeo.

Castellani, D., 2001. Export Behavior and Productivity Growth: Evidence from Italian Manufacturing Firms. Mimeo, ISE-Universit di Urbino.

Clerides, S., Lach, S. and Tybout, J., 1998. Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco. The Quarterly Journal of Economics 113(3).

Crepon, B., Duguet, E., and Mairesse, J., 1998. Research and Development, Innovation and Productivity: An Econometric Analysis at the Firm Level. Economics of Innovation and New Technology 7(2), 115-158.

de Jong, P. 1988. "The Likelihood for a State Space Model." *Biometrika*, 75(1):165-169.

Delgado, M., Farinas, J., and Ruano, S., 2002. Firm productivity and export markets: a non-parametric approach. Journal of International Economics 57(2), 397-422.

Doraszelski, U., Jaumandreu, J., 2007. R&D and Productivity: Estimating Production Functions when Productivity is Endogenous. Working paper, Harvard University.

Durbin, J. and S.J. Koopman, 2001. Time Series Analysis by State Space Methods. Oxford University Press.

Ericson, R. and A. Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62(1):53-82.

Fafchamps, M., El Hamine, S., and Zeufack, A., 2007. Learning to Export: Evidence from Moroccan Manufacturing. Journal of African Economies, June.

Girma, S., D. Greenaway and R. Kneller 2002. "Does exporting lead to better performance? A microeconometric analysis of matched firms". GEP Research Paper 02/09. University of Nottingham.

Greenstreet, D. 2007. "Exploiting Sequential Learning to Estimate Establishment-Level Productivity Dynamics and Decision Rules." Department of Economics, University of Oxford, WP.

Griffith, R., Huergo, E., Mairesse, J., and Peeters, B., 2006. Innovation and productivity across four European countries. NBER Working paper W12722.

Griliches, Z., 1994. R&D and Productivity: The Econometric Evidence. The University of Chicago Press: Chicago.

Griliches, Zvi and Jacques Mairesse, 1995. "Production Functions: The Search for Identification." National Bureau of Economic Research. Working Paper 5067.

Harvey, A.C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.

Huergo E., Jaumandreu, J., 2004. Firms' age, process innovation and productivity growth. International Journal of Industrial Organization 22, 541- 559.

Iacovone, L. and B. Javorcik, 2007. "Getting Ready: Preparation for Exporting." WP.

Jefferson, G., Huamao, B., Xiaojing, G. and Xiayung, Y., 2004. R&D performance in Chinese industry. Economics of Innovation and New Technology13 (1/2). Special issue "On Empirical Studies of Innovation in the knowledge driven economy", Guest editors: B.H. Hall and J. Mairesse.

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

670.

Klette, T.J. and Z. Griliches, 1996. “The Inconsistency of Common Scale Estimators when Output and Prices are Unobserved and Endogenous.” Journal of Applied Econometrics, 11:343-361.

Kraay, A., 1999. Exportations et Performances Economiques: Etude d’un Panel d’Entreprises Chinoises. Revue d’Economie du Dveloppement, 1-2/1999, 183-207.

Lipster, R. S. and A.N. Shiryaev. 2001. *Statistics of Random Processes. Vol II. Applications.* 2nd ed. Trans. A. B. Aries. Applications of Mathematics Vol. 6. Springer-Verlag.

Pakes, A., Ericson, R., 1995. Markov Perfect Industry Dynamics: A Framework for Empirical work. Review of Economic Studies 62(1), 53-82.

Roberts, M., Tybout, R., 1997. The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs. American Economic Review 87(4), 545-564.

Salomon, R., J. Shaver, 2005. Learning-by-exporting: New Insights from Examining Firm Innovation. Journal of Economics and Management Strategy 14(2): 431-461.

Table 1. Means over 1990-2002

Industry	# of obs	Exporters,%	Innovators,%	Size	Ln (output)	Ln (material input)	Ln (labor input)	Ln (capital input)
Textiles, Leather and footwear	929	52.0 (49.9)	33.4 (47.2)	122.66 (184.70)	12.81 (1.81)	11.96 (2.21)	11.34 (1.39)	12.31 (2.26)
Chemical products, Plastic and rubber products	941	74.0 (43.8)	52.3 (49.9)	254.51 (290.17)	14.68 (1.83)	14.04 (1.91)	12.27 (1.36)	14.60 (2.03)
Metallurgy, Metallic products	1064	63.2 (48.2)	48.4 (50.0)	227.59 (667.51)	13.84 (1.87)	13.14 (1.99)	11.71 (1.43)	13.52 (2.39)

Standard deviations are given in parentheses.

Table 2
Multifactor Productivity of Exporters
(ESEE Sample of Spanish Firms)

	Exporter	Non-Exporter	Began Export	Halted Export
Textiles and	2.47	1.60	1.83	1.45
Leather Products	(1.27)	(0.96)	(0.93)	(0.68)
Chemical, Plastic,	0.66	0.23	0.45	0.45
and Rubber Products	(0.84)	(0.63)	(0.58)	(0.66)
Metallic Products,	0.23	0.08	0.15	0.13
Machinery and Equip.	(0.49)	(0.33)	(0.37)	(0.30)

- Log scale.
- Based on parameter estimates reported in results section.
- Standard deviations in parantheses.

Table 3
Estimated Productivity Dynamics

	Textiles and Leather Products		Chemical, Plastic, and Rubber Products		Metallic Products, Machinery and Equip.	
Production Function:						
Capital	0.0916	*	0.2319	****	0.0268	
	(0.0484)		(0.0578)		(0.0287)	
Labor	0.1856	****	0.1867		0.3597	****
	(0.0419)		(0.0741)		(0.0525)	
Materials	0.4078	****	0.4808	****	0.6087	****
	(0.0395)		(0.0680)		(0.0371)	
Productivity Transition:						
Product Innovation	0.0333		0.0665	*	0.0201	
	(0.0360)		(0.0395)		(0.0321)	
Process Innovation	-0.0084		-0.0115		0.0297	
	(0.0340)		(0.0360)		(0.0274)	
Exporter	0.0614	*	-0.0087		0.0160	
	(0.0334)		(0.0460)		(0.0347)	
Began Export	0.0672		0.0110		0.0132	
	(0.0812)		(0.0900)		(0.0666)	
Halted Export	-0.1104		-0.0293		0.0292	
	(0.0963)		(0.1033)		(0.0790)	
Sigma^2	1.51E-16		1.84E-23		-3.41E-28	
	(4.53E-10)		(1.47E-13)		(5.90E-16)	
Q	0.1316	****	0.2113	****	0.1340	****
	(0.0066)		(0.0106)		(0.0063)	
R	0.8670	****	0.7359	****	0.5882	****
	(0.0445)		(0.0693)		(0.0558)	
Observations:	929		941		1064	

- Asymptotic standard errors in parentheses.
- Sigma^2 and Q transformed from their square roots by delta method.
- W set to 0.04.
- Year dummies not reported

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

**** Significant at 0.1%.