

Adjusting the Internationalization Productivity Premium for Firms Heterogeneity by using the Quantile Regression Decomposition

Maria Rosaria Ferrante and Marzia Freo

Abstract The relationship between firm's internationalization and productivity, widely analyzed in the literature, predicts the presence of a productivity premium for internationally involved firms with respect to the domestic ones. This finding is alternatively supported either by comparing the means of productivities over the two groups controlled for the relevant covariates or by the analysis of the stochastic dominance of the unconditional productivity distributions. In the present paper we unify the two approaches by estimating the entire distribution of the internationalization premium adjusted from the effects of different composition of groups. The statistical tool we adopt is based on the quantile regression decomposition approach. The main result we obtain shows that: i) by controlling for effects of firms heterogeneity the productivity premium is substantially reduced; ii) the adjusted productivity premium is significant only for less productive firms.

Key words: TFP, quantile regression, statistics for business and industry

1 Introduction and positioning within the literature

A number of studies in the last decade focused on the relationship between the internationalization of firms and total or partial productivity, prompted by the seminal empirical work of Bernard and Jensen (1995). The main results of these works is that productivity is positively related to international involvement (for a complete and up-to-date review see Greenaway and Kneller, 2007, and Wagner, 2007).

From an applied perspective a number of different statistical and econometric tools have been adopted to evaluate the direction and extent of this relationship in the

Maria Rosaria Ferrante, University of Bologna; email: maria.ferrante@unibo.it

Marzia Freo, University of Bologna; email: marzia.freo@unibo.it

literature on the internationalization of firms. To briefly illustrate the main features of the different approaches, the international status of firms will be defined by considering two groups of firms: those operating in the international market (INT group), and those operating only in the domestic market (DOM group). Bernard and Jensen (1995) tested the difference in productivity between INT and DOM by roughly comparing group productivity averages, so focusing on an “unconditional” comparison based on a synthetic indicator of productivity. In order to consider the whole productivity distribution rather than simply the averages, Delgado et al. (2002) evaluated the hypothesis of the first order stochastic dominance between group distributions. However, this approach is “unconditional” in the sense that it neglects that the relationship between productivity and the TFP can also be affected by a number of other firm characteristics. A way to control for firm’s heterogeneity consists of estimating a regression model that adopts productivity as an outcome variable, the dummy INT/DOM being one of the covariates describing firm’s characteristics (see Wagner (2007) for a review). Unfortunately this approach is based on the comparison of (conditional) averages which neglects the form of the productivity distribution.

This paper contributes to the understanding of the relationship between firm productivity and internationalization from more perspectives. In fact it unifies the two main streams of the literature on the relationship between productivity and international status: the one based on (conditional) linear models and the one based on comparing distribution functions using the concept of the first order stochastic dominance.

First, we single out a “conditional” comparison between groups, that is we compare the “net” TFP premium. In fact, the comparison of the TFP between groups can lead to biased conclusions if it is singled out without taking into account that compared groups can differ with respect to many characteristics defining heterogeneity of firms, including, for example, size, location and type of industry. Since each of these characteristics could have some association with the level of firm productivity, the evaluation of the productivity gap between groups (the so called “unadjusted” or “raw” comparison) would consist of a comparison neglecting the *ceteris paribus* condition. Then, we refer the comparison to the whole productivity distribution, that is, along its different quantiles, in order to not limiting the comparison to synthetic indicators (i.e. to the average). The strategy above outlined allows answering to the following research questions:

- a) Is the TFP gap significantly different from zero? And if yes, does the gap amount depend on the TFP level? For example, is it lower or higher for firms with a lower or higher TFP?
- b) Is the relationship between TFP and variables describing firms’ heterogeneity different (in sign and amount) at different TFP distribution quantiles?
- c) Does the TFP gap “survive” conditionally for a number of characteristics of firms and, if the answer is affirmative, does the extent of this “surviving” gap depend on the amount of TFP?
- d) How much of the TFP gap is explained by characteristics of firms (self-selection hypothesis, as in Wagner (2007)) and how much by the net (so-called “adjusted”) productivity premium?

The statistical approach we adopt to deal with this multi-object task was proposed by Author et al. (2005), Machado and Mata (2005) and Melly (2005) in the framework of the comparison of wage distributions in reference to different racial groups or genders. Based on the idea that a full understanding of the raw productivity premium requires a disentangling of the effects of differences in the structural characteristics of firms from the effects of differences in returns to the characteristics, this decomposition

approach permits to evaluate what proportion of the productivity gap between INT and DOM is due to different characteristics of the firms and what proportion is due to their internationalization status.

The paper is organized as follows. In Section 2 we illustrate the methodology and show how differences in productivity distribution can be decomposed into three factors: coefficients, covariates and residuals. In Sections 3 and 4 we respectively describe the data and analyze the results obtained. In Section 5 some conclusions are drawn.

2 Methodology

The quantile regression approach (Koenker and Basset, 1978) is a methodology to estimate the regression coefficients at different quantiles of the distribution of the response variable. The coefficients of a specific covariate refer to different quantiles of the dependent variable allowing the evaluation of if and how the covariate impacts on the response variable at different quantiles. Then the regions of the support of the dependent variable where these effects are especially weak or strong or not present at all can be identified.

The quantile decomposition technique uses the quantile regression model to provide an estimator of distribution function of the dependent variable in presence of covariates, which enables a calculation of counterfactual density. By comparing actual and counterfactual densities, it partitions the observed distribution of an outcome (usually wages, here TFP) into return components (coefficients), quantity components (group composition) and residuals and computes their impact on the overall outcome distribution. In other words, this technique provides information on differences between distributions evaluated moving from the lower to the upper tail of the conditional productivity distribution.

The rationale of the quantile decomposition technique develops as follows. Let $\{(y_i, x_i)\}_{i=1}^N$ be an independent sample from a population with y_i the outcome variable and x_i a $K \times 1$ vector of covariates, then

$$y_i = Q_\tau(y_i | x_i) + u_i = x_i \beta(\tau) + u_i, \quad \forall \tau \in (0,1) \quad (1)$$

where $Q_\tau(y_i | x_i)$ is the τ^{th} quantile of the distribution of y_i conditionally on the vector of K covariates $x_i = (x_{i1}, x_{i2}, \dots, x_{iK})$. Quantile regression enables the estimation of the relationship between a dependent variable and regressors along the whole distribution, imposing a minimal number of assumptions concerning the shape of the distribution function. For logged dependent variables the quantile regression coefficients may be interpreted as rates of returns to different characteristics at the specified quantile of the conditional distribution. If equation (1) is correctly specified, $Q_\tau(y_i | x_i)$, as a function of $\tau \in (0,1)$, provides a full characterization of the conditional distribution of y given x . On the other hand, by exploiting the knowledge of the dependence on the covariates, it is possible to construct counterfactual conditional distributions corresponding to different scenarios for covariates and coefficients. Since the conditional distribution does not reflect the variability of covariates in the population, the following step is required to marginalize the conditional distribution using alternative scenarios for the distribution of firm attributes, by integrating the estimated conditional quantile function

over the distribution of $\mathbf{x}$ and τ in order to obtain an estimate of the counterfactual density, as follows:

$$\hat{f}(y) = \hat{f}(g(x), y) = \iint_{\mathbf{x}, \tau} \hat{f}(y|x) g(x) d\mathbf{x} d\tau$$

So far, by applying the covariates composition at state 0, $g_0(x)$, to the coefficients at state 1, $\beta_1(\tau)$, it is possible to estimate the outcome counterfactual density $\hat{f}(g_0(x), y)$ which would happen if covariate composition were as at state $s = 0$ and returned as at state $s = 1$. In other words, the density of output may be adjusted by group composition.

After adjusting for compositional effects, the residual difference is partly due to the difference in returns between groups and partly to the within group inequality due to residuals. Since the $\tau = 0.5$ quantile of the residual distribution of u given $\mathbf{x}$ is consistently estimated by $x[\beta(\tau) - \beta(0.5)]$, according to Melly (2005), the within inequality component may be obtained by computing the distribution that would prevail if the median return to characteristics are as at state 1 but the residuals are at state 0, by using the return $\beta^{1,0}(\tau) = \beta^1(0.5) + \beta^0(\tau) - \beta^0(0.5)$.

To estimate the unconditional quantile at operational level, the problem of lack of monotonicity, that is $\tau_j \leq \tau_k \Rightarrow x_j \beta(\tau_j) \leq x_k \beta(\tau_k)$, may emerge. Here, the consistent and asymptotically normally distributed estimator of the population q^{th} quantile in (Melly, 2005)

$$\hat{Q}_\theta[f(g(x), \beta)] = \inf \left\{ q: \sum_{i=1}^N \sum_{j=1}^{N_j} (\tau_j - \tau_{j-1}) I(x_i \beta(\tau_j) \leq q) \geq \theta \right\} \quad (2)$$

is used to this purpose. This method allows for the simulation of counterfactual distributions that can be used to decompose differences in distribution. By using the equation (2), posing a regular grid of sufficiently small quantiles, for example $(\tau_j - \tau_{j-1}) = 0.01$, conditional distribution may be integrated to obtain marginal distribution.

Summing up, the observed difference between unconditional θ quantiles of the outcome densities over two groups are decomposed as

$$\begin{aligned} \hat{Q}_\theta[f(g_1(x), \beta_1)] - \hat{Q}_\theta[f(g_0(x), \beta_0)] &= \left[\hat{Q}_\theta[f(g_1(x), \beta_0)] - \hat{Q}_\theta[f(g_0(x), \beta_0)] \right] \\ &+ \left[\hat{Q}_\theta[f(g_1(x), \beta_{1,r0})] - \hat{Q}_\theta[f(g_1(x), \beta_0)] \right] \\ &+ \left[\hat{Q}_\theta[f(g_1(x), \beta_1)] - \hat{Q}_\theta[f(g_1(x), \beta_{1,r0})] \right] \end{aligned}$$

where the first component is explained by changes in the composition of characteristics, the second difference measures the between-group inequality and is explained by the difference in median coefficients, while the last difference measures the within-group inequality and is explained by residuals as covariates and coefficients are kept fixed.

For the purposes of this paper, the component due to characteristics may be interpreted as the observable effect of self-selection factors on the choice of openness to international markets, that due to coefficients, measuring the between-group difference may be interpreted as the net internationalization productivity premium, and the last component measures the residual within-group difference.

As a further interpretation, if the total differences are considered an Average Treatment Effect of the openness to international market on productivity, the coefficient component is the Average Treatment Effect on the Treated.

When interpreting results a caveat should be kept in mind. By construction, the method provides an accounting decomposition conditional on a given model and may only detect the influence of modeled covariates.

3 The TFPs heterogeneity

The analysis is conducted on firm level data from the ninth Capitalia survey, which contains accounting information from balance-sheets as well as information on geographical location, export, sub-contracting, and innovation. It includes all Italian firms with more than 500 employees and a stratified sample of those with less than 500 employees. In this analysis, after a cleaning procedure aiming at the exclusion of observations with missing data on outcome variable or on covariates, 3,431 manufacturing firms were selected.

Manufacturers involved in export during the three year period were about 75.3% of the whole, 20% of firms are involved in commercial penetration, 30.3% in agreements while other internationalization forms are less frequent. As a whole 76.7% of the firms are involved in at least one internationalization choice while the other 23.3% are purely domestic firms not involved in any form of internationalization. We consider firms as internationalized (INT) if they utilize at least one of the modes mentioned, whereas the remaining firms are domestic (DOM).

Firm heterogeneity may be found in compositional characteristics of INT versus DOM groups. Within the DOM group more than one third of firms (34%) operate within Mineral and Metal products industries, while only 19% in the INT group. Firms open to international markets more frequently operate in Textiles and Clothing, Non-electric Machinery, and Furniture industries. With respect to regional aspects, INT are more often found in district areas and in Northern or Central Italy. Other features associated with international openness are the skill intensity of employees, the presence of innovation or investments and group membership. Finally, international firms tend to be larger than domestic firms.

The Total Factor Productivity (TFP) is measured at firm level by estimating nine Cobb-Douglas production functions by industry, with value added as output, total costs of labor as labor input and the book value of fixed and intangible assets as capital input by aggregation of industry. All nominal variables are deflated by proper index numbers. To take into account the simultaneity problem between input decisions and productivity shock, it is adopted the semi-parametric Levinson and Petrin (2003) estimator, in which deflated intermediate costs are assumed as proxy of capital. Since the level of TFPs cannot be measured in any meaningful units, movements relative to other firms or to a representative firm need to be computed. At the end firm-year specific TFPs, scaled respect industry means and log transformed, provide relative measures of how firm specific TFP diverges from the average during the three year period.

The main evidence emerging from the unconditional comparison (Tab. 1) is that a relevant productivity premium exists for INT firms with respect to DOM firms. The premium is quite uniform along the distribution, that is, of the same entity for more productive and less productive units in each group. DOM firms are found to be less

productive than INT firms of 11.2% in mean. Stochastic dominance of TFP distribution for international firms is found as well. The TFP gap between INT and DOM groups is quite stable along the distributions, from 12.1% at the first quartile to 13.1% at the third quartile.

Table 1: Summary statistics of TFP distribution by international status

<i>ln(TFP)</i>	mean	1 st quartile	median	3rd quartile
All firms	-0.070	-0.274	-0.086	0.122
Domestic	-0.156	-0.367	-0.183	0.015
International	-0.044	-0.245	-0.053	0.146
Int. productivity premium	0.112	0.121	0.130	0.131

In order to adjust for the compositional effects observed for INT and DOM groups and to identify how much of the TFP gap is explained by characteristics of firms we run quantile regressions. For each international status, the log of the TFP is regressed on a set of influent attributes. To the aim of singling out the role of the different covariates, we run different model specifications in terms of regressors. Here the results of the specification with the covariates most used in the empirical literature on firm heterogeneity are presented (geographical, manufacturing district and sectoral dummies, features describing firm behavior such as labor skill intensity, innovations or investments implemented in the period and membership to groups and firm size). In Table 2 the estimates of median and inter-decile regressions of the $\ln(TFP)$ are reported. A significant coefficient in the median regression means that a significant effect of the specific variable on the median level of response is detected for that group. A significant coefficient in the inter-decile regression means that a significant different impact between the ninth and first decile, that is between high and low values of response variable for the specific covariate, is detected in that group.

Table 2. Median regression and interdecile ranges coefficients by international status.

<i>Dependent: ln(TFP)</i>	$\beta(0.5)$		$\beta(0.9) - \beta(0.1)$	
	DOM	INT	DOM	INT
Constant	-0.691	-0.652	0.871	0.691
Food and beverages	0.144	0.041	-0.052	0.135
Textile, clothing and leather	-0.009	0.025	-0.265	0.031
Wood, paper, printer and publishing	0.032	0.069	-0.274	-0.101
Chemicals, rubber and plastic	-0.016	-0.035	0.074	-0.008
Non-electric machinery	0.001	-0.020	-0.191	-0.097
Electric electronic medical instruments	-0.060	-0.042	-0.153	0.013
Vehicles and transportation	-0.141	-0.166	0.030	0.404
Furniture	0.023	0.024	-0.236	-0.062
South	-0.031	-0.028	0.159	0.045
District	0.010	0.021	-0.050	-0.018
Skill intensity	0.147	0.216	0.211	0.367
Innovation	-0.035	0.019	-0.025	-0.042
Investment	0.011	0.016	-0.069	0.071
Group	0.123	0.008	0.307	0.107
Ln employment	0.121	0.114	-0.031	-0.042

In bold significant coefficients.

All variables have expected signs. Location in the south of Italy has a negative association with productivity, while in manufacturing district area the impact is positive. Internationalized TFP increases in the presence of investment spending or innovations. As the skill intensity of labour increases, both median level of total productivity and heterogeneity in productivity distribution within the groups significantly increase. The same happens for membership to a company group.

4 Decomposition results

Basing on the results of previous quantile regressions, differences in TFP distributions are decomposed following the quantile decomposition methodology. Results are presented in Table 3. Export productivity premium ranges from 11.2% at the second decile in the bottom of distribution to 12.1% at the eighth decile in the higher part of distribution. When the covariates are considered, the firms' attributes explain an amount of gap of 7.1% at the second decile of the distribution, of 9.1% at the eighth decile of productivity distribution. The composition effect is then relevant along the whole TFP distribution. The component explained by the coefficient, which may be considered the adjusted internationalization productivity premium is around 5% for firms which belong to the less productive part of populations and is absent from the bottom half of distribution. The component explained by residuals is quite uniformly below 2% along the distribution.

Table 3. *Components of total productivity premium due to residual, covariates and coefficients at different quantiles.*

Quantile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Total premium	.099	.112	.123	.135	.137	.137	.131	.121	.072
due to residuals	-.018	-.017	-.006	.002	.008	.012	.016	.016	-.018
covariates	.070	.071	.076	.084	.086	.086	.086	.091	.093
coefficients	.047	.057	.054	.049	.043	.038	.029	.015	-.003

As in any method comparing two states (as for example index-numbers), the decomposition may change if the reference state changes. To control for these effects, results have been obtained by reversing the order of the decomposition and computing the productivity loss for domestic versus international firms. Changes observable in quantitative absolute results are negligible with no effects on the overall picture.

5 Conclusions

Many papers in the empirical literature on the heterogeneity of firms find a productivity premium associated with international involvement. Two main questions are still open on this topic. The first entails the amount of the productivity premium which, although in relative terms, is rarely measured. The second entails the direction of the causality link between productivity and international openness. Does a self-selection mechanism induce the more productive firms to enter the international market

or do internationalized firms, under the pressure of the global competition, become more productive by means of a learning-by-exporting type process?

The present paper does not provide a definitive answer to the above questions but certainly provides some useful insights. The analysis is conducted on micro-level data of a representative sample of Italian manufacturing firms. At first, as expected, the main evidence emerging from the unconditional comparison is that a relevant productivity premium for INT firms with respect to DOM firms exist and that it is uniform along the entire distribution, i.e. for more productive and less productive units. When the effect of firms heterogeneity is controlled for, by using quantile decomposition, firms involved in internationalization are found still differ from domestic firms with respect to TFP with some variants: the adjusted internationalization premium is estimated positive and around 5% for the less productive firms, while it is negligible and not significant for the most productive ones. Clearly this finding does not give a definitive answer to the causality direction, but it restates the open question about the level and direction of the TFP gap, as we estimate that the TFP gap is 5% and it only regards the less productive firms.

In this framework, the self-selection hypothesis is only confirmed for the less productive firms. From this point of view, the internationalization policies should only target firms with a low level of TFP and they should foster the firms' performances and their productivity instead of promoting internationalization per se. Considering the learning-by-exporting hypothesis, our results may support previous evidence that this process mainly entails less productive (and often smaller) firms (Lileeva and Trafler, 2007). According to this hypothesis less productive firms seem to engage in a catching-up type behaviour and take more competitive advantage from openness to international markets. In this framework, policies targeting internationalization could result in an increase of firm productivity.

To further understand the puzzle of causality, the time dimension in the data may no longer be neglected and future research will be focused on evaluating the pre-entry and post-entry (in the international markets) TFP differences.

References

1. Author D.H., Katz L.F., Kearney M.S.: Rising wage inequality: the role of composition and prices. NBER Working Paper 11628 (2005).
2. Bernard A. B., Jensen J.: Exporters, jobs and wages in US manufacturing: 1976-87, Brookings Pap. Econ. Act., Microeconomics, 67--112 (1995).
3. Delgado M. A., Farinas J. C., Ruano S.: Firm productivity and export markets: a non-parametric approach, *J Int. Ec.* 57, 397--422 (2002).
4. Greenaway D., Kneller R.: Firm heterogeneity, exporting and foreign direct investment: a survey. *Econ. J.*, 117, 134--161 (2007).
5. Koenker R., Bassett, G.: Regression Quantiles. *Econometrica*, 46(1), 33--50 (1978).
6. Levinsohn J., Petrin A.: Estimating production functions using inputs to control for unobservables, *Rev. Econ. Stud.*, 70, 317--342 (2003).
7. Lileeva A., Treffler D.: Improved access to foreign markets raises plant-level productivity. NBER Working Paper. 13297. (2007)
8. Machado J.A.F., Mata J.: Counterfactual decomposition of changes in wage distributions using quantile regression. *J. Appl. Econometrics*. 20, 445--465 (2005).
9. Melly B.: Decomposition of differences in distribution using quantile regression, SIAW Working Paper. (2005).
10. Wagner, J.: Exports and productivity: A survey of the evidence from firm-level data, *The World Economy*, 30(1), 60--82 (2007).