DISCUSSION PAPER SERIES

IZA

IZA DP No. 5175

From Estimation Results to Stylized Facts: Twelve Recommendations for Empirical Research in International Activities of Heterogeneous Firms

Joachim Wagner

September 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

From Estimation Results to Stylized Facts: Twelve Recommendations for Empirical Research in International Activities of Heterogeneous Firms

Joachim Wagner
Leuphana University Lueneburg
and IZA

Discussion Paper No. 5175
September 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5175
September 2010

ABSTRACT

From Estimation Results to Stylized Facts: Twelve Recommendations for Empirical Research in International Activities of Heterogeneous Firms*

Heterogeneous firms are at the heart of both the *New New International Trade Theory* and the *Micro-econometrics of International Firm Activities*. One important aim of micro-econometric studies is to uncover stylized facts that hold over space and time, and that can both inspire theoretical models that are based on "realistic" assumptions, and inform policy debates in an evidence-based way. Which results from the thousands of empirical estimates reported in the literature on the micro-econometrics of international firm activities do we consider as convincing? Based on my own experience from the last twenty years I use the opportunity of this lecture to make twelve recommendations that, hopefully, will help to find the right way on the thorny road from estimation results to stylized facts. I will deal with the following topics: comparisons of means vs. comparisons of distributions; extremely different firms, or outliers; unobserved heterogeneity; simultaneous occurrence of differences across quantiles, outliers, and unobserved heterogeneity; heterogeneous effects of international firm activities on firm performance; replication; within-study replication by international research teams; meta-analysis; and talking to practitioners.

JEL Classification: F14, C21, C23

Keywords: international firm activities, heterogeneous firms, stylized facts, robust statistics, replication, meta anaysis

Corresponding author:

Joachim Wagner
Leuphana University Lueneburg
Institute of Economics
P.O. Box 2440
D-21314 Lueneburg
Germany
E-mail: wagner@leuphana.de

* Many thanks to Vincenzo Verardi for his cooperation on the application of the robust estimator for fixed-effects models, to David Powell for his cooperation on the application of the quantile regression estimator for fixed effects models, and to Nils Braakmann and Christian Pfeifer for helpful remarks on an earlier draft. Earlier versions were presented as an invited lecture at the DIME-ISGEP International Workshop on Firm Selection and Country Competitiveness in Nice-Cote-d'Azur in March 2010, as a keynote paper at the Sixth Danish International Economics Workshop, Aarhus, June 3-4, 2010, and as a keynote lecture at the international conference "Productivity and Internationalisation – A Micro-data approach", The Hague, September 1-3, 2010. I thank participants for helpful comments. All remaining errors are mine.

1. Motivation

Some twenty years ago economists working on empirical investigations of international trade issues started to recognize that trade is performed by firms, and that these internationally active firms differ from firms that are not engaged on international markets. Furthermore, they realized that firms are heterogeneous, and that the representative firm is a myth.[1] During the following years a comprehensive literature emerged that formed the field of *Micro-econometrics of International Firm Activities*. Economists all over the world used large comprehensive sets of firm level data collected by the statistical agencies to investigate the differences between firms with different forms of international activities, and the causes and consequences of these international activities.[2]

These empirical studies inspired a number of theoretical papers that model the behavior of heterogeneous firms in open economies.[3] This literature emerged to what is now labeled the *New New International Trade Theory*. Some of these theoretical papers developed testable hypotheses that lead to micro-econometric studies with results that bounced back to theory. The mushrooming growth of this literature indicates that this is a fertile ground for both theoretical and empirical analyses.

One important aim of empirical studies in this field of economics (and in other fields as well) is to uncover stylized facts that hold over space and time, and that can

[1] Pioneering papers in this field include Bernard and Jensen (1995) and Wagner (1995). Note that this has been recognized by business economists active in empirical research in the field of international management for a long time; see the papers collected in the five volume set edited by McNaughton and Bell (2009) where the first paper included dates from 1978.

[2] For partial surveys of this empirical literature see Greenaway and Kneller (2007), López (2005), and Wagner (2007).

[3] The canonical paper in this literature is Melitz (2003) who explicitly motivates his theoretical model by referring to findings in the micro-econometric literature; see Helpman (2006) for a survey.

both inspire theoretical models that are based on “realistic” assumptions, and inform policy debates in an evidence-based way. Which results from the thousands of empirical estimates reported in the literature on the micro-econometrics of international firm activities do we consider as convincing? Based on my own experience from the last twenty years I use the opportunity of this lecture to make twelve recommendations that, hopefully, will help to find the right way on the thorny road from estimation results to stylized facts. I will deal with the following topics: comparisons of means vs. comparisons of distributions (section 2); extremely different firms, or outliers (section 3); unobserved heterogeneity (section 4); simultaneous occurrence of differences across quantiles, outliers, and unobserved heterogeneity (section 5); heterogeneous effects of international firm activities on firm performance (section 6); replication (section 7); within-study replication by international research teams (section 8); meta-analysis (section 9); and talking to practitioners of international firm activities (section 10).

2. Comparison of means vs. comparison of distributions

Heterogeneous firms are at the heart of both the *New New International Trade Theory* and the *Micro-econometrics of International Firm Activities*. The implications of firm heterogeneity for empirical analyses, however, are not always fully taken into account. One important aspect here is the frequent use of (unconditional or conditional) means for different groups of firms – say, exporters and firms selling on the national market only – as a basis for comparisons. A case in point is the well known fact that in almost all countries and periods examined exporters are significantly more productive than non-exporters on average (see Wagner 2007). This fact is usually documented by comparing the mean of productivity for the two groups of firms, using either a statistical test for the significance of differences in (unconditional) means of exporting and non-exporting firms, or by performing a test for the statistical significance of the estimated regression coefficient of a dummy-variable indicating the exporter status of a firm in an empirical model that controls for industry affiliation and firm size (i.e. the difference in the conditional mean).

As a first step this is fine. But one should not stop here. As Moshe Buchinsky (1994: 453) put it: "'On the average' has never been a satisfactory statement with which to conclude a study on heterogeneous populations." An empirical study of heterogeneous firms should look at differences in the whole distribution of the variable under investigation between groups of firms, not only at differences at the mean.

To illustrate this point I will look at productivity differences between exporters and firms selling on the national market only. The data used were collected in personal interviews with firm owners or top managers. The population covered encompasses all manufacturing establishments with at least 5 employees in the German state of Lower Saxony. From this population a random sample (stratified by industry and size classes) was interviewed. Detailed information on the data set and how it can be accessed by researchers is given in Gerlach, Hübler and Meyer (2003). The survey has information on whether or not a plant was an exporter in 1995. One great advantage of this survey for the exercise performed here is that in using these data I am allowed to do things that are impossible for me to do using the confidential firm level data from official statistics that are usually used for empirical investigations on international activities of German firms. For instance, I can report results for minimum and maximum values of variables that are by definition values for single firms and, therefore, cannot be revealed when using the data from official statistics.[4]

Table 1 reports figures for labour productivity (defined as sales per employee[5]) for exporters and non-exporters.[6] As expected, exporters are on average more productive than non-exporters. This difference in the unconditional mean is statistically highly significant (with a p-value of 0.007), and it is large from an economic point of view (66,178 Deutschmark, or 30.22 percent, in favour of the exporters). However, the mean value of a variable might be heavily influenced by a small number of extremely large or small observations, especially if the number of firms is fairly small like in the exercise performed here.

[Table 1 near here]

A look at selected percentiles of the productivity distribution for exporters and for non-exporters reported in table 1 reveals that both exporters and non-exporters are highly heterogeneous with regard to their productivity. The largest value is 160 (115) times the smallest value for exporters (non-exporters). Exporters do have a higher productivity than non-exporters at each percentile, and not only at the mean. The hypothesis that the productivity distribution of exporters stochastically dominates the productivity distribution of non-exporters can be tested by the Kolmogorov-

[4] These data have been used before for empirical investigations of international firm activities; see Wagner (1998, 2001, 2006a, 2006b).

[5] Note that the figures are in Deutschmark because they refer to 1995; it might help to get an impression about what is behind these figures if a rule of thumb is applied and all values in Deutschmark are divided by two to get the approximate amount in Euro.

[6] All computations were performed using Stata 11.0; see StataCorp (2009).

Smirnov test. This non-parametric test for first order stochastic dominance of one distribution over another was introduced into the empirical literature on exports and productivity by Delgado, Farinas and Ruano (2002), and it has been applied in a number of papers on the micro-econometrics of international firm activities. Let F and G denote the cumulative distribution functions of productivity for two groups of firms (say, exporters and firms that serve the national market only). First order stochastic dominance of F relative to G is given if F(z) – G(z) is less or equal zero for all z with strict inequality for some z. Given two independent random samples of plants from each group, the hypothesis that F is to the right of G can be tested by the Kolmogorov-Smirnov test based on the empirical distribution functions for F and G in the samples (for details, see Conover 1999, p. 456ff.).

Results reported in the lower panel of table 1 show that according to the Kolmogorov-Smirnov test the two productivity distributions do differ, and that the null hypothesis that productivity is higher among exporters cannot be rejected at any conventional error level while the null hypothesis that the productivity is higher among non exporters has to be rejected at any error level. Therefore, the conclusion drawn from the comparison of the average values of labour productivity – exporters are more productive than non-exporters – is the same as the conclusion that follows from a comparison of the whole distributions of labour productivity for the two groups of firms. Obviously, this has not to be the case with other data sets from other times or other countries. At least in my view results based on a comparison of distributions are more convincing when firms are heterogeneous. My first recommendation, therefore, is:

R1: In a comparison of groups of heterogeneous firms do not only test for differences in mean values – look at differences at percentiles of the distributions, and test for differences between distributions, too.

Empirical studies on international firm activities usually go beyond a comparison of (unconditional) mean values for groups of firms and look at differences between firms controlling for firm characteristics by using regression analysis, i.e. they look at differences in the conditional mean value. Taking the example of productivity differences between exporting and non-exporting firms, researchers often are interested in the ceteris paribus difference in productivity between exporters and non-exporters that are of the same size and from the same industry. This exporter productivity premium is estimated by a regression of the (log of) labour productivity on a dummy variable that takes on the value of one if a firm is an exporter plus a set of control variables.[7] Results reported in the upper left corner of table 2 show that the estimated exporter premium, conditional on firm size and industry affiliation, is 33.79 percent in our example (and larger than the unconditional premium that is 30.22 percent according to the figures reported in table 1).

[Table 2 near here]

[7] See e.g. International Study Group on Exports and Productivity (ISGEP) (2008) for comparable results for 14 countries. Note that this regression equation is not meant to be an empirical model to explain labour productivity at the firm level; the data set at hand here is not rich enough for such an exercise. It is just a vehicle to test for, and estimate the size of, the exporter premium controlling for other firm characteristics. Furthermore, note that productivity differences at the firm level are notoriously difficult to explain empirically. “At the micro level, productivity remains very much a measure of our ignorance.” (Bartelsman and Doms 2000, p. 586)

If we acknowledge that firms are heterogeneous, we have reasons to suspect that the conditional difference in labour productivity between exporting and non-exporting firms does not need to be the same for all firms. For example, it might be the case that the productivity difference between exporters and non-exporters of the same size and from the same industry is higher for firms at the lower end of the productivity distribution. If we are interested in the size of the exporter premium, and if we regress the log of labour productivity on an exporter dummy variable and a set of control variables using ordinary least squares (OLS), there is no room for firm heterogeneity of this kind. OLS assumes that the conditional distribution of productivity, given the set of firm characteristics included in the regression, is homogeneous. This implies that no matter what point on the conditional distribution is analyzed, the estimates of the relationship between productivity (the dependent variable) and the firm characteristics (the independent variables) are the same.

If one wants to test the empirical validity of the assumption made by OLS, and if one is interested in the evaluation of the size of the exporter premium at different points of the conditional productivity distribution, one has to apply a different estimation technique that is tailor-made for this – quantile regression. A discussion of technical details of quantile regression is beyond the scope of this paper; canonical references are the pioneering paper by Koenker and Bassett (1978), the survey by Buchinsky (1998) and the monograph by Koenker (2005), while Koenker and Hallock (2001) provide a non-technical introduction. Suffice it to say here that in contrast to OLS (that gives information about the effects of the regressors at the conditional mean of the dependent variable only) quantile regression can provide parameter estimates at different quantiles. Therefore, it gives information on the variation in the effect of independent variables on the dependent variable at different quantiles. The estimated regression coefficients can be interpreted as the partial derivative of the conditional quantile of the dependent variable (here: labour productivity) with respect to a particular regressor (e.g., being an exporter or not), i.e. the marginal change in productivty at the k^{th} conditional quantile due to a change in exporter status. For each quantile it can be shown whether the effect of a particular independent variable is positive or negative, and how large this effect is compared to other quantiles. This provides information about the heterogeneity of plant behavior. Note that quantile regression is not the same as applying OLS to subsets of the data produced by dividing the complete data set into different percentiles of the dependent variable. This would mean that not all of the data are being used for each estimate, and it would introduce the familiar type of sample selection bias. For each quantile regression estimate all of the data are being used; some observations, however, get more weight than others.

Estimation results for the exporter productivity premium from quantile regressions[8] are reported in the upper panel of table 2. The estimated exporter premium is statistically different from zero, positive, and large from an economic point of view for all quantiles. The premium varies across the different quantiles, and there seems to be a u-shaped pattern showing a higher premium at both ends of the conditional productivity distribution than at the median. According to the results of tests for coefficient equality between pairwise quantiles and across all quantiles reported in the lower panel of table 2, however, these differences between the estimated exporter premia are never statistically different from zero.

[8] Micro-econometric studies on international firm activities using quantile regression include Dimelis and Louri (2002), Falzoni and Grasseni (2005), Wagner (2006a), Yasar, Nelson and Rejesus (2006), Yasar and Morrison Paul (2007), Trofimenko (2008), Serti and Tomasi (2009), Bellone, Guillou and Nesta (2010), Haller (2010) and Arnold and Hussinger (2010).

The bottom line, then, is that the relationship between exporting and labour productivity is the same at each point of the conditional productivity distribution. Bellone, Guillou and Nesta (2010) report a similar finding for a sample of firms from France. Obviously, this has not to be the case with other data sets from other times[9] or other countries. Cases in point are the study by Yasar, Nelson and Rejesus (2006) who find for Turkish firm level data that the exporter productivity premium increases as one moves from the lower tail to the upper tail of the distribution, and the study by Serti and Tomasi (2009) who report that the respective coefficients are much larger at the lower quintiles, especially for firms selling goods to European and low income countries. At least in my view, therefore, results based on a comparison across different quantiles of the conditional distribution are more convincing when firms are heterogeneous. My second recommendation, therefore, is:

R2: In a comparison of groups of heterogeneous firms do not only test for differences in conditional mean values estimated by OLS (*or any other econometric method that focuses on the conditional mean of a dependent variable) – look at differences at quantiles of the conditional distribution using quantile regression, and test for differences between quantiles, too.*

3. Extremely different firms (outliers)

If one investigates a sample of heterogeneous firms it often happens that some variables for some firms are far away from the other observations in the sample. For

[9] Arnold and Hussinger (2010) report results for pooled data from a panel of German firms covering the years 1996 to 2002. They find an inverted u-shaped pattern of productivity premia for exporters; however, they do not test for the statistical significance of the differences between the quantiles.

example, in the sample of exporting and non-exporting firms that is analyzed here according to table 1 there are a few firms with labour productivity values that are extremely low or extremely high compared to the mean values. These extreme values might be the result of reporting errors (and, therefore, wrong), or due to idiosyncratic events (like in the case of a shipyard that produces a ship over a long time and that reports the sales in the year when the ship is completed and delivered), or due to firm behavior that is vastly different from the behavior of the majority of firms in the sample. Observations of this kind are termed outliers. Whatever the reason may be, extreme values of labour productivity may have a large influence on the mean value of labour productivity computed for the exporters and non-exporters in the sample, on the tails of the distribution of labour productivity, and on the estimates of the exporter premium. Conclusions with regard to the productivity differences between exporters and non-exporters, therefore, might be influenced by a small number of firms with extremely high or low values of productivity, and the same is true for any other empirical investigation using data for a sample of heterogeneous firms.

Researchers from the field of micro-economics of international firm activities usually are aware of all of this. Given that due to confidentiality of the firm level data single observations as a rule cannot be inspected closely enough to detect and correct reporting errors, or to understand the idiosyncratic events that lead to extreme values, a widely used procedure to keep these extreme observations from shaping the results is to drop the observations from the top and bottom one percent of the distribution of the variable under investigation. A case in point is the international comparison study on the exporter productivity premium by the International Study Group on Exports and Productivity (ISGEP) (2008, p. 610).

To illustrate the effects of trimming the sample this way all computations for table 1 and table 2 were repeated for a sample without the observations from the top and bottom one percent of the productivity distribution. Results are reported in table 3 and table 4.

[Table 3 and Table 4 near here]

Table 3 reveals that the share of exporters and non-exporters is the same in both samples. Mean values of labour productivity are lower for both groups of firms, and the same holds for the standard deviations. A t-test for differences in productivity means again rejects the null-hypothesis of no difference. The difference in the unconditional mean of productivity is large from an economic point of view (55,016 Deutschmark, or 26.62 percent, in favour of the exporters), but is considerably smaller than for the whole sample (66,178 Deutschmark, or 30.22 percent). This illustrates that the mean value of a variable might be heavily influenced by a small number of extremely large or small observations.

Exporters do have a higher productivity than non-exporters at each percentile, and not only at the mean, after trimming the sample, too. Furthermore, results of the Kolmogorov-Smirnov test are again in line with the hypothesis that the productivity distribution of exporters stochastically dominates the productivity distribution of non-exporters.

The exporter productivity premium that is estimated by an OLS-regression of the (log of) labour productivity on a dummy variable that takes on the value of one if a firm is an exporter plus a set of control variables is reported in the upper left corner of table 4 for the trimmed sample. The estimated premium, conditional on firm size and industry affiliation, is 28.51 percent – considerably lower than the 33.79 percent premium estimated using the full sample.

Estimation results for the exporter productivity premium from quantile regressions are reported in the upper panel of table 4. The estimated exporter premium is statistically different from zero, positive, and large from an economic point of view for all quantiles. The premium varies across the different quantiles, and as for the complete sample there seems to be a u-shaped pattern showing a higher premium at both ends of the conditional productivity distribution than at the median. Again, according to the results of tests for coefficient equality between pairwise quantiles and across all quantiles reported in the lower panel of table 4. However, these differences between the estimated exporter premia are never statistically different from zero.

The bottom line, then, is that the big picture of the relationship between exporting and labour productivity that is sketched using the complete sample and the trimmed sample without the top and bottom one percent observations from the distribution of labour productivity is the same. The unconditional and the conditional exporter premium, however, is smaller for the firms in the trimmed sample. It should be pointed out that, obviously, this has not to be the case with other data sets from other times or other countries.

Dropping the firms from the top and the bottom one percent of the productivity distribution and comparing the results of empirical investigations with and without these firms with extremely high or extremely low values of labour productivity might be considered as a first and useful step to check the sensitivity of results. However, although this approach seems to be rather popular it is in some sense arbitrary. Why the top and bottom one percent? Why not choose a larger or smaller cut-off point?

There are alternative approaches to deal with extreme observations (outliers) that are substantiated in statistics, and we will turn to these methods now.

One approach that is advocated in the literature has already been applied in our exercise. Quantile regression is often used to deal with outliers. As Yasar, Nelson and Rejesus (2006, p. 682) put it: "Quantile regression estimates are considered robust relative to least squares estimates. In contrast to the least squares estimator, the quantile regression estimates place less weight on outliers and are found to be robust to departures from normality." Quantile regression at the median is identical to least absolute deviation (LAD) regression that minimizes the sum of the absolute values of the residuals rather than the sum of their squares (as in OLS). This estimator is also known as the L_1, or median regression, estimator. Results reported in table 2 and table 4 demonstrate that this estimator is indeed robust with respect to the inclusion of observations with extreme values of labour productivity – the estimated exporter premium at the median of the conditional labour productivity distribution is 20.06 percent for the full sample and 19.17 percent for the sample without the observations from the top and the bottom one percent of the productivity distribution.

LAD regression, however, is not a panacea against outliers. To see why, following Rousseeuw and Leroy (1987) we distinguish three types of outliers that influence the OLS estimator: vertical outliers, bad leverage points, and good leverage points. Verardi and Croux (2009, p. 440) illustrate this terminology in a simple linear regression framework (the generalization to higher dimensions is straightforward) as follows: "Vertical outliers are those observations that have outlying values for the corresponding error term (the *y* dimension) but are not outlying in the space of explanatory variables (the *x* dimension). Their presence affects the OLS estimation and, in particular, the estimated intercept. Good leverage points are observations that are outlying in the space of explanatory variables but that are located close to the regression line. Their presence does not affect the OLS estimation, but it affects statistical inference because they do deflate the estimated standard errors. Finally, bad leverage points are observations that are both outlying in the space of explanatory variables and located far from the true regression line. Their presence significantly affects the OLS estimation of both the intercept and the slope."

Using this terminology one can state that the median regression estimator protects against vertical outliers but not against bad leverage points (Verardi and Croux 2009, p. 441; Koenker 2005, p. 268). Another quite popular robust estimator is the M-estimator proposed by Huber (1964) that generalizes median regression to a wider class of estimators; it is implemented in Stata via the command `rreg`. However, as pointed out by Verardi and Croux (2009, p. 442), rreg can only identify isolated outliers and is inappropriate when clusters of outliers exist where one outlier can mask the presence of another, and the initial values for the algorithm is not robust to bad leverage points.

Full robustness can be achieved by using the so-called MM-estimator that can resist contamination of the data set of up to 50% of outliers (i.e., that has a breakdown point[10] of 50 % compared to zero percent for OLS). A discussion of the details of this estimator is beyond the scope of this paper (see Verardi and Croux (2009) for this estimator and for Stata commands to compute it; Maronna, Martin and Yohai (2006) is a comprehensive textbook treatment of robust statistics). Suffice it to say here that this estimator combines a breakdown point of 50 percent with a high

[10] The breakdown point of an estimator is the highest fraction of outliers that an estimator can withstand, and it is a popular measure of robustness.

efficiency (the degree of which can be chosen by the researcher). Explicit formulas for the estimator are not available, they are computed by numerical optimization.

Table 5 reports results for the exporter premium computed for the sample of 618 firms described above using OLS and LAD-regression (already reported in table 2) plus results estimated using the Huber M-estimator (via rreg) and the MM-estimator (via mmregress).[11] Furthermore, given that the MM-estimator (like LAD and rreg) uses all 618 observations but a weighting scheme, results for an OLS-estimation for a sample without the outliers detected by mmregress are reported in the last column, too.[12]

[Table 5 near here]

The estimated labour productivity premium is statistically highly significant and large from an economic point of view for all estimators applied. The estimated size, however, differs considerably. The estimated premium from the fully robust MM-estimator is considerably lower than the values from both OLS and LAD applied to the full or the trimmed sample (see table 4).[13] In my view it is important to document the extent to which estimation results are influenced by extreme observations. Given

[11] Computations were done using the ado-files provided by Verardi and Croux (2009) with the efficiency parameter set at 0.7 as suggested there based on a simulation study; details are available on request.

[12] Note that it is not possible to state the efficiency of the estimator when these outliers are dropped; results are reported to document the effect on the point estimate only.

[13] For another example showing that outliers can shape a result in a study on exporting see Verardi and Wagner (2010b). Using the MM-estimator to test for self-selection of more productive firms into exporting beyond countries of the Euro-zone they find an ex-ante differential that is statistically significant and large only for enterprises that exported to the Euro-zone already and start to export to countries outside the Euro-zone – a conclusion that differs considerably from the one based on the use of non-robust OLS regression analysis.

that the presence of outliers can be expected to be the rule in data sets for heterogeneous firms my third recommendation is:

R3: Carefully check how your results are influenced by firms that are very much different from the mass of firms in your sample. Report results based on OLS and on the MM-estimator.

4. Unobserved heterogeneity

The last two sections dealt with the consequences of observed firm heterogeneity for micro-econometric studies of international firm activities. Firm heterogeneity, however, might be caused by factors that are either not observed by the researcher and that, therefore, are not included in the empirical model, or that are unobservable to a researcher. A case in point with regard to the exporter productivity premium is management quality. Although management quality has been considered as an important source of performance differences between firms for a very long time – Syverson (2010, p. 14) mentions a study published in 1887 that made this point – empirical evidence on this is scarce due to data limitations. As Syverson (2010, p. 14) puts it, "(t)he identity, much less the characteristics, practices, or time allocation of individual managers are rarely known. Furthermore, managerial inputs can be very abstract. It's not just time allocation that matters, but what the manager does with their time, like how they incentivize workers or deal with suppliers." A recent study by Bloom and Van Reenen (2010) that relates management practices to productivity shows, among others, that firms that export (but do not produce) overseas are better-managed than domestic non-exporters, but are worse-managed than multinationals.

In the data sets used to empirically investigate international firm activities variables that measure management quality are missing.[14] This would not pose a big problem if management quality would be uncorrelated with the other variables included in the empirical model (e.g., the exporter status) – of course it would not be possible to investigate the role of management quality for productivity differences between firms empirically, but the estimated coefficient for the exporter dummy variable would be an unbiased estimate of the exporter productivity premium (given all other assumptions for the applicability of OLS are fulfilled). However, one would not expect that management quality is uncorrelated with either the exporter status or other variables like firm size. Not controlling for management quality then leads to biased estimates for the exporter premium.

A standard solution for this problem that is widely used in the literature on the micro-econometrics of international firm activities is the estimation of fixed effects[15] models for panel data. Using pooled cross-section time-series data for firms and including fixed firm effects in the empirical model allows to control for time invariant unobserved firm heterogeneity, and to estimate the coefficients for the time variant variables that are included in the models without any bias caused by the non-inclusion of the unobserved variables that are correlated with these included variables. A case in point is the paper by the International Study Group on Exports and Productivity (ISGEP) (2008), were in table 4 exporter productivity premia are

[14] If you are aware of data sets that include this information please let me know!

[15] As an aside, note that although in the theoretical models from the New New International Trade Theory productivity differentials between firms are modeled as the results of a random draw from a productivity distribution (see e.g. Melitz 2003) it is not appropriate to use random effects models instead of fixed effects models in the empirical investigations. Random effects models assume that the observed variables in the empirical model and the unobserved variables not included in the model are uncorrelated – an assumption that makes no sense here.

reported based on empirical models with and without fixed effects. If fixed firm effects are added to control for time invariant unobserved heterogeneity the point estimates of the exporter productivity premia are much smaller compared to the results based on pooled data only.

Thus, unobserved firm heterogeneity does matter. However, before dropping the estimates based on pooled data without fixed effects one should remember that by construction the estimated coefficients of the exporter status variable from the empirical models with fixed firm effects are identified by observations only that change their exporter status (at least once) during the period under investigation. This exporter status of a firm tends to be highly persistent over adjacent years. Table 6 illustrates this with evidence from German manufacturing enterprises over the two-year periods from 1995/96 to 2005/06. On average, some 93 percent of all enterprises that did not export in a year did not export in the following year, and the share of firms that exported in year t+1 among the exporting enterprises in year t is even higher.

[Table 6 near here]

Furthermore, we know that firms that enter or exit the export market are different from firms that persistently stay in or out of it. Using a panel of German manufacturing establishments Wagner (2008a) finds that firms that stop exporting in year t were in t-1 less productive than firms that continue to export in t, and that firms that start to export in year t are less productive than firms that export both in year t-1 and in year t. This means that the coefficient of the exporter status variable that gives us the estimate for the exporter productivity premium is in a sense estimated for quite different samples when models with and without firm fixed effects are used. Therefore, both estimates with and without fixed effects should be reported, and they should be amended by a table that documents the share of different patterns of export participation in the whole sample over time: How many firms exported in each year? How many firms did not export in a single year? How many firms entered the export market in year two and persistently exported until the final year under consideration? How many firms entered the export market in year two, exported in year three and year four, and stopped to export in year five without exporting again until the final year? Etc. A table like that might help to interpret the difference between the estimated exporter productivity premium from empirical models with and without fixed firm effects.

A similar argument holds when the share of exports in total sales instead of the exporter status is included in an empirical model to estimate the productivity premium of higher export intensity.[16] A case in point is again the paper by the International Study Group on Exports and Productivity (ISGEP) (2008), were in table 5 such exporter productivity premia are reported based on empirical models with and without fixed effects. If fixed firm effects are added to control for time invariant unobserved heterogeneity the point estimates of the exporter productivity premia are much smaller compared to the results based on pooled data only.

[16] If the share of exports in total sales is not used as a right-hand side variable but as the endogenous variable the fact that this variable can by definition only take values between zero and one (or zero and one hundred percent) and that there is a probability mass at zero (due to a high share of firms with no exports at all) has to be taken into account. In a cross section framework this leads to a fractional logit model (see Papke and Wooldridge (1996) for the method and Wagner (2001) for an application to exports) that easily generalizes to a fixed-effects panel data model in the case of data for a population only (see Wagner 2003). Recently Papke and Wooldridge (2008) published a fixed-effects estimator for fractional response variables in a balanced panel that has been applied to model the share of exports in total sales in Wagner (2008b).

Again, unobserved firm heterogeneity does matter. However, before dropping the estimates based on pooled data without fixed effects one should keep in mind that by construction the estimated coefficients of the export share variable in the model with fixed firm effects uses only the information on the change of the share of exports in total sales within the firms over the time span included in the panel. While we have good reasons to suspect that this share tends to vary over time, it is an open question how large this variation over time within firms is compared to the variation of the share of exports between firms. Information on the relative importance of the within variation and the between variation is helpful to put the estimates with and without fixed firm effects into perspective, because in the fixed effects model the coefficient of a regressor with little within variation will be imprecisely estimated (and will be not identified if there is no within variation at all). A table reporting overall, between, and within variance for any variable used in the empirical model should be included and discussed (see Cameron and Trivedi 2009, p. 239 for an example).

It seems appropriate to mention a dilemma here that is well known to applied researchers. Usually, the within variation that is needed to identify the coefficient of a regressor tends to increase with the length of the panel used in the estimation. A long panel covering many years, therefore, might be considered to be a better basis for empirical investigations than a short panel for only some years. However, the fixed effects that control for unobserved firm heterogeneity are by assumption time invariant, and this assumption seems more appropriate in shorter than in longer panels. A case in point is the quality of the management of a firm mentioned at the beginning of this section to motivate the application of fixed effects regression methods when investigating heterogeneous firms econometrically. The assumption that management quality does not vary over time seems more convincing in the short run than in the long run (if only because bad management quality can be expected to lead to either market exit or to a new and better (or at least, different) group of in the longer run). The empirical researcher, therefore, is facing a trade-off – usually, the longer the panel, the larger is the within variation in the regressors, but the less appropriate is the assumption of time invariant unobserved heterogeneity.

A case in point is my study on the determinants of the share of exports in total sales that uses panel data for a sample of East and West German manufacturing firms for 1999 to 2002 and that applies both a fractional logit estimator for the pooled data and a fractional probit panel estimator to control for unobserved heterogeneity via fixed firm effects (Wagner 2008b). While firm size, human capital intensity and R&D intensity turn out to be statistically significantly positively related to export performance (as expected) in the empirical models without fixed effects, this is no longer the case (except for firm size in West Germany) when fixed firm effects are included. Is it correct to argue that this demonstrates that in German manufacturing industries it is neither human capital intensity, nor R&D intensity *per se* that make a successful exporter, but that unobserved time-invariant characteristics that are correlated with these observed characteristics are all that matter? The panel is short enough to justify the assumption that important unobserved firm characteristics can indeed be considered as time-invariant. But the within variation of the regressors might well be too small over this short time period to estimate their effect precisely.

The only solution for this dilemma seems to be an exogenous event that leads to large enough variation in the observed regressors in a short time span. Take one of your favourite drinks if this happens in a data set you use!

This leads to my fourth recommendation:

R4: If panel data are used, report results for pooled data with and without fixed effects, and document time patterns for discrete variables plus the within and between variation of continuous variables.

5. Differences across quantiles, outliers, and unobserved heterogeneity: Three challenges at a time

As a next step we will discuss the different aspects of working with data for heterogeneous firms that have been discussed separately in section two to four simultaneously. Is it possible to tackle all three problems – different effects at different (conditional) quantiles of the distribution of a variable under consideration, outliers, and unobserved heterogeneity - simultaneously? Is there such a thing as a highly robust quantile regression estimator for fixed effects models? To the best of my knowledge, the answer as of today is "unfortunately, no" – but there is progress to report.

As regards outliers and unobserved heterogeneity, Verardi and Croux (2009, p. 452) point out that "(i)n particular, development of robust procedures for panel-data and time-series models would be of major interest for applied economic research." A highly robust MM-estimator for panel data with fixed effects has been proposed recently by Bramati and Croux (2007). While a discussion of details of this estimator is beyond the scope of this paper the underlying idea is to center the series of observations for a firm in a similar way to what is generally done when applying the within transformation that is used to estimate a fixed effects model. The difference here is that the series are centered by removing the median instead of demeaning because the mean is largely distorted by outliers. Having centered the series, a robust estimator can be applied to deal with atypical individuals. The outcoming results will be comparable to those of a fixed effects estimator but will not be distorted by the presence of atypical individuals.

Verardi and Wagner (2010a) apply this newly developed method to the estimation of exporter productivity premia for firms from manufacturing industries in West Germany and compare the results to those from using the standard fixed effects estimator. The empirical study uses pooled data for the years 1995 to 2006. The dependent variable is the log of labor productivity (defined as sales per employee; in Euro). Two empirical models are estimated that differ in the way exports are measured – either as a dummy variable that takes the value of one if an enterprise is an exporter in a year (model 1), or as the share of exports in total sales in a year and its squared valued (model 2). Both empirical models include the number of employees and its squared value plus year dummy variables and a constant.

For both models 3.07 percent of the enterprises are identified to be outliers (1,060 in case of model 1 and 1,052 in model 2), and this holds for 12.42 percent (or 37,666) observations in the case of model 1 and for 12.36 percent (or 37,497) observations in the case of model 2. Dropping these outliers leads to a drastic change in the estimation results for the exporter productivity premium and to a dramatic change in the conclusions drawn: While the estimated exporter premium is statistically highly significant and large from an economic point of view, taking on a value of 13.43 percent, this estimate (while still statistically highly significant) drops to 0.997 percent when the same model is estimated using the robust fixed effects method. According to the results from the robust fixed effects regression there is no such thing as a large exporter productivity premium! Comparing the results for model

2, the conclusions drawn do differ between the standard and the robust fixed effects regression, too: While productivity is rising at a decreasing rate with an increase in the share of exports according to the results from the standard fixed effects estimation there is no such pattern revealed from the robust fixed effects regression, and the increase of productivity with an increase in the share of exports in total sales is much less pronounced. This demonstrates that outliers can drive results from an empirical study with heterogeneous firms.[17]

As regards quantile regression (that takes care of differences at the different quantiles of the conditional distribution, and that is robust to outliers in the dependent variable) Koenker (2004) suggests that unobserved fixed effects can be controlled by including firm dummy variables in the regression. With a large number of firms, however, this approach becomes technically not feasible due to convergence problems (Yasar, Nelson and Rejesus 2006, p. 682).[18]

Recently David Powell (2009) developed a method for unconditional quantile regression for panel data with exogenous or endogenous regressors.[19] Powell and Wagner (2010) apply this approach to estimate the exporter productivity premium at quantiles of the productivity distribution for manufacturing enterprises in Germany.

Results for West Germany show that the exporter productivity premium declines over the productivity distribution. The premium is highly statistically significant, and very large from an economic point of view, at the lower end. The

[17] Verardi and Wagner (2010b) report a similar result in a study on the exporter premia by destination of exports.

[18] Chen and Khan (2008) suggest an alternative fixed effects estimator for quantile regression models that, however, can deal with pooled data for two periods only. For an application of this approach that is not related to international firm activities see Gamper-Rabindran, Khan and Timmins (2010). For a random effects quantile regression estimator see Geraci and Bottai (2007).

[19] A discussion of the details of this estimator is far beyond the scope of this paper; see Powell (2009) for details.

estimated coefficient for the ten percent quantile shows a productivity premium of exporting over non-exporting firms of 15.6 percent. The premium is statistically significantly different from zero at a conventional level in the first two thirds of the productivity distribution only. This clearly demonstrates that the premium is not constant among enterprises from different parts of the productivity distribution. The estimated coefficient from an OLS fixed-effects regression using the same empirical model and the same sample of enterprises is 0.118 (which translates to a productivity premium of 12.5 percent) – this premium at the conditional mean, therefore, is much less informative with regard to the relation between productivity and exporting than the results for the various quantiles.

Results for East Germany show a higher premium at the lower end of the productivity distribution, too. The estimated coefficient for the ten percent quantile indicates a productivity premium of exporting over non-exporting firms of 15.6 percent, identical to the results for West Germany. Contrary to what is found for West Germany, in East Germany the exporter productivity premium is statistically different from zero at a conventional error level over nearly the complete productivity distribution, and the estimated premia do not differ significantly between the 20 percent quantile and the 80 percent quantile. The estimated coefficient from an OLS fixed-effects regression using the same empirical model and the same sample of enterprises is 0.120 (which translates to a productivity premium of 12.7 percent) – like in the case of West Germany this premium at the conditional mean is much less informative with regard to the relation between productivity and exporting than the results for the various quantiles.

Note that the point estimate of the exporter productivity premium at the conditional mean is virtually identical for West and East Germany, and the difference between the two estimated premia is not statistically significant. Looking at the conditional mean only, therefore, leads to the wrong conclusion that the relation between productivity and exports is identical in enterprises from West and East Germany when unobserved time invariant firm heterogeneity is controlled for by fixed enterprise effects – a statement that is clearly demonstrated to be wrong by applying quantile regression with fixed effects.

This leads to my fifth recommendation:

R5: If panel data are used, report results for pooled data with fixed effects using the standard fixed effects estimator plus both the quantile regression fixed effects estimators and the highly robust fixed-effects estimator.

6. Heterogeneous effects of international firm activities on firm performance

When firms are heterogeneous we have reasons to suspect that any effect of international firm activities on firm performance – like the effect of exporting on productivity growth or profitability – will vary across the firms under consideration. This is illustrated by results I found in a survey of exporting firms from the region I live in (see Wagner 2009): For example, strong positive effects of exports on profits are reported by half of the firms only, while more than one third argued that there is a small positive effect only, and thirteen percent reported no positive effect at all. Similar results are found for other dimensions of firm performance. While the probability to find a strong positive effect increases with an increasing share of exports in total sales, other firm characteristics (including size, years of export experience, and research and development activities) are not related to positive export effects.[20]

In the literature on the causal effects of international firm activities on firm performance this heterogeneity of effects is often ignored. The standard approach – that has been pioneered by Wagner (2002) and Girma, Greenaway and Kneller (2003) and that has been applied in a large number of studies ever since – uses propensity score matching to uncover what is known as the *average treatment effect on the treated* (ATT). In a nutshell[21] this ATT is the average difference in a performance dimension (say, productivity growth) that can be observed between firms that started to engage in an activity (say, exporting) and firms that did not, while the firms from the two groups of starters and non-starters are made of matched pairs, or statistical twins, that were identical (or at least as similar as possible) in all characteristics that are relevant for the probability to start exporting and for the outcome variable in the year before some of the firms started to export.

This standard approach looks at the average treatment effect on the treated. To repeat an argument put forward in section 2 above, as a first step this is fine. But one should not stop here. To quote Moshe Buchinsky (1994: 453) again: "'On the average' has never been a satisfactory statement with which to conclude a study on heterogeneous populations." An empirical study of heterogeneous firms should also consider the variation in the effect of exporting (or any other form of international firm activity) over the treated firms.

[20] Kneller and Pisu (2010) report results from a survey of UK firms on the returns to exporting and report differences between firms regarding these effects, too.

[21] A discussion of the details of the method used to estimate the ATT is beyond the scope of this paper; see Wagner (2002) and the literature cited therein.

One approach that makes a step in this direction, and that has been applied to the analysis of causal effects of international firm activities on firm performance, is not to consider the effects of a discrete treatment (like exporting or not) but of a continuous treatment (like exporting a different share of the total product of a firm). In this approach not an average treatment effect on the treated is estimated but a dose-response function is computed that shows the effect of a specific dose of a treatment (a given share of exports in total sales) on an indicator of firm performance (like productivity growth or profitability). Pioneering studies using this continuous treatment approach in the field of micro-econometrics of international firm activities are Fryges (2009), Fryges and Wagner (2008, 2010), and Vogel and Wagner (2010).[22] This approach turns out to be a powerful tool to uncover heterogeneous effects of international firm activities on firm performance.[23] For example, Fryges and Wagner (2008) find that exporting improves labor productivity growth only within a sub-interval of the range of firms's export sales ratios.

These considerations lead to my sixth recommendation:

R6: A study of the causal effects of international firm activities on firm performance should not only look at the average effect - it should consider the variation of this effect over the range of the extent of this activity.

[22] A discussion of the details of the method used to estimate a dose-response function is beyond the scope of this paper; see Fryges and Wagner (2008) and the literature cited therein.

[23] A different approach looks at the heterogeneous effects of a treatment that is the same for all subjects. This approach focuses on different effects of the same treatment on subjects from different strata, see Brand and Xie (2007) for details. I am not aware of any application of this approach from the micro-econometrics of international firm activities.

7. The role of replication

Everybody who ever produced estimation results is aware of the fact that only a tiny fraction of these results is documented in tables of working papers, and often an even smaller fraction is published in the final version of a paper. The published results are not a random sample of all the variants that came out of the printer, they are a highly selected sample made of those results we as researchers consider to be the most suitable – be it because they are the ones that are most in line with our priors, or the most statistically significant, or whatever. Given that we all are aware of this publication bias, we usually tend to be reluctant to trust published results.

If a paper carefully performed an empirical investigation using a sample of heterogeneous firms to look at an aspect of international firm activity, taking all the recommendations given above into account, does it show more than that the results reported are possible results? Can the results be taken as a reliable guide for evidence-based policy recommendations? In most of the times (if not at any time) in my opinion the answer should be "no". Besides the publication bias mentioned above, errors can occur at every stage of an empirical investigation (not to mention that it is easy to fake numbers reported in tables), and the data used can be non-representative in many ways (that are often not obvious for the researcher and even less for the reader), taken from a specific part of the population of all firms at a specific point in time sampled in a specific way. Therefore, my seventh recommendation is:

R7: Never consider results based on one sample of firms from one country and from one period of time as a stylized fact.

Replication can be most helpful on the long way from isolated estimation results to stylized facts. In this context, replication can take on two different forms. The first one is related to the reproducibility of a set of published results, and it is termed *pure replication* by Daniel Hamermesh (2007, p. 716). To qualify as a scientific contribution a published result must be reproducible. As Hrishikesh Vinod and Bruce McCullough (2003, p. 888) put it, “(r)eplication is the cornerstone of science. Research that cannot be replicated is not science, and cannot be trusted either as part of the profession’s accumulated body of knowledge or as a basis for policy.”

Results that can be trusted must be reproducible by the original investigator (and by any other person with the technical skills needed) even after a considerable amount of time. This might seem to be self-evident, but many of us know that it is not (see Hamermesh (2007) for examples). Try to reproduce results from a paper you published ten years ago, and you will often see that this is not a trivial task. Did you keep all the raw data and all the program files to process them? Did you carefully document which program file produced which line of the table of results in the paper? Is the (version of) the program you used still working on your current computer, and are you able to read the data from the storage medium you have it on? Let’s hope for the best. Helpful hints for housekeeping to safeguard against any negative surprise are for example given by Koenker and Zeileis (2009) and Long (2009).

Moreover, more and more journals demand that the material used to produce results in a published paper is submitted to a data archive, and this should make replication of results an easy task. If the data are confidential – which is usually the case with the firm level data used in micro-econometric studies of international firm activities – and can only be accessed inside the statistical offices, this is somewhat more complicated, but not impossible. Usually, these data are confidential but not exclusive, and replication is possible for all researchers that qualify as users of the data. In my view this contributes to the confidence we can have in published results. However, as Frey, Eichenberger and Frey (2009, p. 154) argue "(in) economics, to replicate the work published by others, and to reveal mistakes, usually is not a promising strategy, because the scholars exposed are likely to be among the referees". But this depends on the editors, and the mere possibility that it is easy to replicate published results, and to demonstrate the degree of fragility of it, should work as an incentive to publish only robust and reliable results. Making all details of the computations done for an empirical study easily available to other researchers, therefore, is highly recommended:

R8: Always make the program files and the data used in an empirical study available for replication and extensions. If the data are confidential, give detailed information how they can be accessed by other researchers.

The second type of replication besides *pure replication* is named *scientific replication* by Hamermesh (2007, p. 727); this approach means "re-examining an idea in some published research by studying it using a different data set chosen from a different population from that used in the original paper". Results generated from data for one economy in one period cannot generally be expected to hold for another economy or the same economy in another period due to differences in institutions or its changes over time, or to time and region specific shocks. "If our theories are intended to be general, to describe the behavior of consumers, firms, or markets independent of the social or broader economic context, they should be tested using data from more than just one economy” (Hamermesh 2007, p. 728). To put it differently, and again quoting Hamermesh (2000, p. 376), “the credibility of a new finding that is based on carefully analyzing two data sets is far more than twice that of a result based only on one.”

The literature on the micro-econometrics of international firm activities is a case in point. Much of it started with the Brookings paper by Bernard and Jensen (1995) that used data from the U.S., and the empirical approach introduced in this paper (and in later papers by the authors) has been applied by others using similar data from many other countries ever since (see the survey by Wagner 2007). These replication studies that often gave rather similar results over space and time all contributed to the big picture we now have on the causes and consequences of international activities of heterogeneous firms. Scientific replication studies of this type provide valuable steps on the way from estimation results to stylized facts.

A problem with this type of replication studies, however, is that the incentives for performing them are often quite low. As Hamermesh (2000, p. 376) puts it, “(t)here is little or no reward to replication”, not least because “(n)o editor of a major journal is likely to publish replications of previous original pieces” (Hamermesh 2007, p. 731). But my experience shows that more and more referees and editors of good field journals and general journals are willing to recognize the value added provided by papers that report evidence on the validity or not of an empirical result reported in other studies. This leads to my recommendation number nine:

R9: Recognize the important role of scientific replication studies that re-examine ideas from published research using different data sets from different countries and periods – and do this as an author, as a referee, and as an editor.

8. Within-study replication by international research teams

One way scientific replication is performed is by a number of authors in different studies investigating the same topic. Another way to perform it is that one author in one study analyzes different data sets from different periods of time and/or different countries, and it is named *within-study replication* in this case (Hamermesh 2007, p. 730). This approach of within-study replication is especially attractive. First of all, it provides evidence on the empirical validity of a finding beyond a specific point in time and space from the outset of the publication of the original study, and not after an often large time span due to lags in recognition, preparation and publication related to replication studies. Second, if work is done by a single researcher (or a single research team) the chances that all the details of the empirical study are identical (or at least very similar) across the data sets tends to be quite high.

Cases in point that use a large number of data sets from various countries include the study on the wage curve – a non-linear relationship between the level of unemployment and the wage level in a region - by Blanchflower and Oswald (1994), the study on the determinants of union membership in 18 EU countries by Schnabel and Wagner (2007), and the study on the relationship between age and the probability of being a union-member by Blanchflower (2007). These and other studies following this within-study replication approach use micro data at the individual level that are either taken from surveys that are designed as international comparative surveys from the outset (like the data from the International Social Survey Programme – ISSP (Uher 2000) and from the European Social Survey – ESS (http://ess.nsd.uib.no), or that are made comparable across countries ex post (like the data from the Luxembourg Income Study - LIS (Smeeding, Jesuit and Alkemade 2002)). These individual level data can easily be investigated by a single researcher after downloading the data on his own computer (e.g., data from ISSP and ESS), or via remote data access (like LIS).

Unfortunately, the situation tends to be completely different when it comes to the field of micro-econometric investigations of international firm activities. In most cases the firm level data are strictly confidential, and as a rule these data can only be used on computers located inside the statistical agencies that are in charge of collecting the data. The data cannot cross borders (not even in a united Europe where borders are hard to recognize when you cross them by car or train), and often they cannot be accessed by citizens of a foreign country (who are not liable to jurisdiction in case of violation of privacy in the country where the data are located). Within-study replication using firm level data from various countries, therefore, usually cannot be performed by one author (or a team of authors) from one country.

A way out is to form a team of researchers who are located in different countries, each of whom does have access to firm level data from her or his country, to agree on a unified empirical approach, and to perform a within-study replication where strictly comparable results for each country are produced by the author(s) from that country. The pioneering study of this type of within-study replication by an international research team in the field of micro-econometrics of international firm activities is Bernard, Jensen and Wagner (1997). This study reports directly comparable results on the differences between exporters and non-exporters for the US and Germany. Recently, teams of researchers active in this field from some 15 countries joined to form the *International Study Group on Exports and Productivity* (ISGEP) and to apply the approach of within-study replication using confidential firm level longitudinal data from various countries. The study looks at cross-country differences in exporter productivity premia estimated by using comparable data and a unified empirical approach (International Study Group on Exports and Productivity (ISGEP) 2008). Another example is a study on self-selection of services firms into exporting that uses identical empirical models and comparable data for France, Germany and the UK (Temouri, Vogel and Wagner 2010).

In my view results from this type of replication studies demonstrate that it surely pays to coordinate the approach used in empirical research ex-ante among the researchers involved. Therefore, my recommendation number ten is:

R10: If within-study replication is not possible for you due to limitations of access to firm level data from several countries join with teams from various countries to perform ex-ante coordinated studies using a unified approach.

9. Meta-analysis – A tool to uncover stylized facts

Ideas from an emerging literature often tend to be analyzed empirically in a large number of scientific replication studies over the years after the original study has been published. A case in point from the micro-econometrics of international firm activities is the literature on the relationship between exports and productivity that grew out of the original paper by Bernard and Jensen (1995). One way to tease out general facts about this and any other topic (assuming such facts exist) is to collect all the results and prepare a survey article (see Hamermesh 2007, p. 731). I did this for the literature on exports and productivity that was published until the end of 2006 and summarized the core results in a table (see Wagner 2007). A bird's-eye view on this collection of results revealed both striking similarities and differences.

If a comprehensive survey of empirical studies on a topic reveals that the effect under investigation tends to be rather similar across space and time, this might be viewed as evidence for the existence of a stylized fact that can be used to guide both economic theory and economic policy. If results differ considerably among the various studies using data from different periods and countries, the next task on the agenda is to find out why they differ.

Obvious suspects that can cause differences in results between empirical studies include, among others, different sampling frames (Are establishments or enterprises used as the unit of analysis? Is the lower bound of the number of employees in a unit identical across the samples?), different definitions of variables (Is productivity measured as sales per employee, value added per employee, or one of several estimates for total factor productivity?), and different specification of the empirical model (Is firm size controlled for by including the squared value of employees? Is unobserved heterogeneity controlled for in a fixed-effects model?). The within-study approach to replication recommended in the section above takes care of these causes for cross-study differences in results.

If results differ across space and time when a unified approach in the form of a within-study replication has been conducted a promising way to investigate these differences is to perform a meta-analysis. This approach regresses the variable under consideration – say, the exporter premium found for a country j – on characteristics of that country j that are considered as possible determinants of the premium (e.g., the average tariff rate, or distance to trading partners). A pioneering application of this approach in the literature on the micro-econometrics of international firm activities can be found in International Study Group on Exports and Productivity (ISGEP) (2008).[24] This study demonstrates that a meta-analysis can be

[24] Meta-analyses from this literature that are not based on estimation results from a within-study replication include Görg and Strobl (2001), Martins and Yang (2008) and Tingvall and Ljungwall (2010). A recent example from the international management literature is Bausch, Fritz and Boesecke (2007). Meta-analyses are routinely used in, for example, medicine, and they are more and more

a powerful tool to summarize what we can learn from all the replication studies, and that it can help to uncover the stylized facts that are needed to inform both theoretical modeling and policy debates.

This leads to my recommendation number eleven:

R11: Use a meta-analysis to understand why empirical results differ across space and time, and to uncover stylized facts that can help to inform theoretical modeling and policy debates.

10. Talking to practitioners

While all my recommendations given so far deal with aspects of how to take care of heterogeneity in the analysis of large sets of firm data I would like to add one final recommendation that aims at a different point:

R12: Leave your office and your computer, and go out to talk to decision makers that care about international firm activities in their day-to-day work.

Richard Freeman (1989, p. xi) pointed out that "(g)etting the opinions of the subjects of our research is also the only advantage we have over physicists. Quarks and gluons do not talk about what they do or why …". While many economist doubt that we can learn much from field work of this type, others state that "learning by asking those who are doing" (Blinder 1990) can be important. George Akerlof (2007,

---

popular in management (under the heading of evidence-based management); see Frese et al. (2008). For a collection of Stata ado-files to perform meta-analyses see Sterne, Newton and Cox (2009). Note that starting in 2010 a new journal, *Research Synthesis Methods*, is published by Wiley Interscience that has focus on meta- analyses.

p. 28) argued in his Nobel-lecture: “In contrast to reliance on statistical testing, disciplines other than economics typically put much greater weight on a naturalistic approach. This approach involves detailed case studies. Such observation of the small often has been the key to understanding of the large.”

This is illustrated by Pack (2006) who gives a recent discussion of a number of case studies dealing with international technology transfer. He points out that case studies provide a rich source of evidence on the details of the transfer and absorption process and offer important clues to the type of microeconomic detail that would contribute to deeper understanding of the process. And he argues that insights from case studies could help frame relevant questions and point out the limitations of econometric studies that are due to the absence of information in censuses that has been shown to be relevant in extensive case study interviews.

The combination of qualitative case studies and quantitative analysis that amalgamates micro-econometric investigations of firm level data with discussions with actors inside the firms has recently been labeled *pin factory* or *insider econometrics* (see Freeman and Shaw 2009, p. 3) – where the term pin factory refers to Adam Smith’s famous example of the division of labor, and insider econometrics refers to the use of information that is only available to persons with detailed knowledge of the firm. While following this approach is quite far from common practice in economics, the list of insider econometric studies is long and growing (see Ichniowski and Shaw 2009). As Martin Feldstein (2000, p. iii) argued non-economists are “invariably surprised that the process of visiting companies, looking at production, and asking questions is an unusual part of economic research. It seems like such a natural thing to do. But as economists all know, it is unusual. … (W)e rarely go and look and ask. I think this is a pity. Looking and asking provide insights and suggest hypotheses – and can shoot-down wrong ideas – in ways that go beyond introspection and reading."

From my own (unfortunately, still rather limited) experience I can tell that it is true that you can learn a lot from talking to people who are active in international business. However, I learned, too, that it is not easy to find owners or managers who are willing to talk about these topics. Time constraints, and fear of talking about topics that are considered to be "trade secrets", are barriers that are hard to overcome. Therefore, if you happen to meet with an owner or a manager of a firm the other day, talk to him about your latest theoretical or empirical attempt to uncover the causes and consequences of international firm activities. The increase of your knowledge from a discussion like this might well be larger than the increase from reading yet another paper on the subject – although it should be kept in mind that not only firms but both managers and papers, too, are vastly heterogeneous!

11. Concluding remark

In his Nobel-lecture James Heckman (2001, p. 674 and p. 732) pointed out that "(t)he most important discovery [from micro-econometric investigations, *J.W.*] was the evidence on the pervasiveness of heterogeneity and diversity in economic life. … The evidence from microeconomic data has already had a substantial effect on the development of macroeconomic theory, which is slowly abandoning the representative agent paradigm." International economics is a case in point. Over the past decade the *New New International Trade Theory* and the *Micro-econometrics of International Firm Activities* considerably changed the way economist think about firms that are active on international markets. In this paper I offer recommendations that, hopefully, will contribute to a more complete recognition of firm heterogeneity in the empirical part of this venture.

References

Akerlof, George A. (2007), The Missing Motivation in Macroeconomics. *American Economic Review* 97 (1), 5-36.

Arnold, Jens M. and Katrin Hussinger (2010), Exports versus FDI in German Manufacturing: Firm Performance and Participation in International Markets. *Review of International Economics* 18 (4), 595-606.

Bartelsman, Eric J. and Mark Doms (2000). 'Understanding Productivity: Lessons from Longitudinal Micro Data.' *Journal of Economic Literature* XXXVIII (3): 569-594.

Bausch, Andreas, Thomas Fritz and Kathrin Boesecke (2007), Performance Effects of Internationalization Strategies: A Meta-Analysis. *Reserach in Global Strategic Management* 13, 143-176.

Bellone, Flora, Sarah Guillou and Linel Nesta (2010), To what extent Innovation accounts for firm export premia? University of Nice – Sophia Antipolis, mimeo, February.

Bernard, Andrew B. and J. Bradford Jensen (1995), Exporters, Jobs, and Wages in U.S. Manufacturing: 1976 – 1987. *Brookings Papers on Economic Activity, Microeconomics* 1, 67-119.

Bernard, Andrew B., J. Bradford Jensen and Joachim Wagner (1997), The good go abroad: Evidence from longitudinal micro data on German and U.S. exporers. In: Seppo Laaksonen (ed.), *The Evolution of Firms and Industries*. Helsinki: Statistics Finland, p. 489-500.

Blanchflower, David G. (2007), International Patterns of Union Membership. *British Journal of Industrial Relations* 45 (1), 1-28.

Blanchflower, David G. and Andrew J. Oswald (1994), *The Wage Curve*. Cambridge, MA and London, England: MIT Press.

Blinder, Alan S. (1990), Learning By Asking Those Who Are Doing. *Eastern Economic Journal* XVI (4), 297-306.

Bloom, Nicholas and John Van Reenen (2010), Why Do Management Practices Differ across Firms and Countries? *Journal of Economic Perspectives* 24 (1), 203-224.

Bramati, Caterina Maria and Christophe Croux (2007), Robust estimators for the fixed effects panel data model. *Econometrics Journal* 10 (3), 521-540.

Brand, Jennie E. and Yu Xie (2007), Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. University of California, Los Angeles, Paper CCPR-033-07, December.

Buchinsky, Moshe (1994), Changes in the U.S. Wage Structure 1963 – 1987: Application of Quantile Regression. *Econometrica* 62 (2), 405-458.

Buchinsky, Moshe (1998), Recent advances in quantile regression models. A practical guideline for empirical research. *Journal of Human Ressources* 33 (1), 88-126.

Chen, Songnian and Shakeeb Khan (2008), Semiparametric estimation of nonstationay censored panel data models with time varying factor loads. *Econometric Theory* 24 (5), 1149-1173.

Conover, W. J. (1999), *Practical Nonparametric Statistics*. Third edition. New York etc.: John Wiley.

Delgado, Miguel A., Jose C. Farinas and Sonia Ruano (2002), Firm productivity and export markets: a non-parametric approach. *Journal on International Economics* 57 (2), 397-422.

Dimelis, Sophia and Helen Louri (2005), Foreign ownership and production efficiency: a quantile regression analysis. *Oxford Economic Papers* 54 (3), 449-469.

Falzoni, Anna Maria and Mara Grasseni (2005), Home Country Effects of Investing Abroad: Evidence from Quantile Regressions. Centre for Research on Innovation and Internationalisation, Universitá Bocconi, Milano, CESPRI Working Paper 170.

Feldstein, Martin (2000), Preface: The NBER-Sloan Project on Productivity Change. NBER/Sloan Project Report, Industrial Technology and Productivity: Incorporating Learning from Plant Visits and Interviews into Economic Reserach. Cambridge, MA: National Bureau of Economic Research.

Freeman, Richard B. (1989), *Labour Markets in Action*. New York etc.: Harvester Wheatsheaf.

Freeman, Richard B. and Kathryn L. Shaw (2009), Introduction. In: Richard B. Freeman and Kathryn L. Shaw (Ed.), *International Differences in the Business Practices and Productivity of Firms*. Chicago and London: University of Chicago Press.

Frese, Michael, Andreas Bausch, Peter Schmidt, Andreas Rauch and Rüdiger Kabst (2008), Evidence-based Entrepreneurship (EBE): A systematic approach to cumulative science. Mimeo, University of Giessen.

Frey, Bruno S., Reiner Eichenberger and René L. Frey (2009), Editorial Ruminations: Publishing Kyklos. *Kyklos* 62 (2), 151-160.

Fryges, Helmut (2009), The Export-Growth Relationship: Estimating a Dose-Response Function. *Applied Economics Letters* 16, 1855-1859.

Fryges, Helmut and Joachim Wagner (2008), Exports and Productivity Growth: First Evidence from a Continous Treatment Approach. *Review of World Economics* 144 (4), 695-722.

Fryges, Helmut and Joachim Wagner (2010), Exports and Profitability – First Evidence for German Manufacturing Firms. *The World Economy* 33 (3), 399-423.

Gamper-Rabindran, Shanti, Shakeeb Khan and Christopher Timmins (2010), The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach. *Journal of Development Economics* 92 (2), 188-200.

Geraci, Marco and Matteo Bottai (2007), Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 8 (1), 140-154.

Gerlach, Knut, Olaf Hübler and Wolfgang Meyer (2003), The Hannover Firm Panel (HFP). *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 123 (3), 463-470.

Girma, Sourafel, David Greenaway and Richard Kneller (2003), Export Market Exit and Performance Dynamics: A Causality Analysis of Matched Firms. *Economics Letters* 83 (3), 317-324.

Görg, Holger and Eric Strobl (2001), Multinational Companies and Productivity Spillovers: A Meta-Analysis. *Economic Journal* 111 (November), F723-F739.

Greenaway, David and Richard Kneller (2007), Firm Heterogeneity, Exporting and Foreign Direct Investment. *Economic Journal* 117 (February), F134-F161.

Haller, Stefanie (2010), Exporting, Importing, Intra-Firm Trade and Firm Performance. ESRI Dublin, mimeo, March.

Hamermesh, Daniel S. (2000), The craft of labormetrics. *Industrial and Labor Relations Review* 53 (3), 363-380.

Hamermesh, Daniel S. (2007), Viewpoint: Replication in economics. *Canadian Journal of Economics* 40 (3), 715-733.

Heckman, James.J. (2001), Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy* 109 (4), 673-748.

Helpman, Elhanan (2006), Trade, FDI, and the Organization of Firms. *Journal of Economic Literature* XLIV (September), 589-630.

Huber, Peter J. (1964), Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35 (1), 73-101.

Ichniowski, Casey and Kathryn L. Shaw (2009), Insider Econometrics: Empirical Studies of how Management Matters. National Bureau of Economic Research NBER Working Paper 15618, December.

International Study Group on Exports and Productivity (ISGEP) (2008), Understanding Cross-Country Differences in Exporter Premia: Comparable Evidence for 14 Countries. *Review of World Economics* 144 (4), 596-635.

Kneller, Richard and Mauro Pisu (2010), The returns to exporting: evidence from UK firms. *Canadian Journal of Economics* 43 (2), 494-519.

Koenker, Roger (2004), Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91 (1), 74-89.

Koenker, Roger (2005), *Quantile Regression*. Cambridge etc.: Cambridge University Press.

Koenker, Roger and Gilbert Bassett (1978), Regression quantiles. *Econometrica* 46 (1), 33-50.

Koenker, Roger and Kevin F. Hallock (2001), Quantile regression. *Journal of Economic Perspectives* 15, Fall, 143-156.

Koenker, Roger and Achim Zeileis (2009), On reproducible econometric research. *Journal of Applied Econometrics* 24 (5), 834-847.

Long, J. Scott (2009), *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press.

López, Ricardo A. (2005), Trade and Growth: Reconciling the Macroeconomic and Microeconomic Evidence. *Journal of Economic Surveys* 19 (4), 623-648.

Maronna, Ricardo A., R. Douglas Martin and Víctor J. Yohai (2006), *Robust Statistics. Theory and Methods*. Chichester etc.: Wiley.

Martins, Pedro S. and Yong Yang (2008), The impact of exporting on firm productivity: a meta-analysis of the learning-by-exporting hypothesis. *Review of World Economics* 145 (3), 431-445.

McNaughton, Rod B. and Jim Bell (Ed.) (2009), *Entrepreneurship and Globalization*, Volume 1 – 5, Los Angeles etc.: Sage.

Melitz, Mark J. (2003), The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* 71 (6), 1695-1725.

Pack, Howard (2006), Econometric versus case study approaches to technology transfer. In: Bernhard Hoekman and Beata Smarzynska Javorcik (Eds.), *Global Integration and Technology Transfer*. Washington, D.C.: Palgrave Macmillan and The World Bank, p. 29-50.

Papke, Leslie and Jeffrey M. Wooldridge (1996), Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Economics* 11, 619-632.

Papke, Leslie and Jeffrey M. Wooldridge (2008), Panel Data Models for Fractional Response Variables with an Application to Test Pass Rates. *Journal of Econometrics* 145, 121-133.

Powell, David (2009), Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors. RAND Working Paper WR-710, October.

Powell, David and Joachim Wagner (2010), The Exporter Productivity Premium along the Productivity Distribution: First Evidence from a Quantile Regression Approach for Fixed Effects Panel Data Models. University of Lueneburg Working Paper Series in Economics No. 182, August.

Rousseeuw, Peter J. and Annick M. Leroy (1987), *Robust Regression and Outlier Detection.* New York etc.: John Wiley and Sons.

Schnabel, Claus and Joachim Wagner (2007), Union density and determinants of union membership in 18 EU countries: evidence from micro data, 2002/03. *Industrial Relations Journal* 38 (1), 5-32.

Serti, Francesco and Chiara Tomasi (2009), Self-selection along different export and import markets. Laboratory of Economics and Management Sant'Anna School of Advanced Studies, Pisa, LEM Working Paper Series 2009/18, November.

Smeeding, Timothy M., David K. Jesuit and Paul Alkemade (2002), The LIS/LES Project Databank: Introduction and Overview. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 122 (3), 497-517.

StataCorp (2009), *Stata. Release 11. Statistical Software*. College Station, TX: StataCorp.

Sterne, Jonathan A. C., H. Joseph Newton and Nicholas J. Cox (Eds.) (2009), *Meta-Analysis in Stata: An Updated Collection from the Stata Journal.* College Station, TX: StataCorp.

Syverson, Chad (2010), What determines productivity? National Bureau of Economic Research NBER Working Paper 15712, January.

Temouri, Yama, Alexander Vogel and Joachim Wagner (2010), Self-selection into export markets by business services firms – Evidence from France, Germany and the United Kingdom. University of Lueneburg Working Paper Series in Economics No. 183, August.

Tingvall, Patrik Gustavsson and Christer Ljungwall (2010), Is China different? A Meta-Analysis of Export-led Growth. Stockholm School of Economics, CERC Working Paper 15, April.

Trofimenko, Natalia (2008), Learning by Exporting: Does It Matter Where One Learns? Evidence from Colombian Manufacturing Firms. *Economic Development and Cultural Change* 56 (July), 871-894.

Uher, Rolf (2000), The International Social Survey Programme (ISSP). *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 120 (4), 663-672.

Verardi, Vincenzo and Christophe Croux (2009), Robust regression in Stata. *The Stata Journal* 9 (3), 439-453.

Verardi, Vincenzo and Joachim Wagner (2010a), Robust Estimation of Linear Fixed Effects Panel Data Models with an Application to the Exporter Productivity Premium. University of Lueneburg Working Paper Series in Economics No. 168, April.

Verardi, Vincenzo and Joachim Wagner (2010b), Productivity premia for German manufacturing firms exporting to the Euro-area and beyond: First evidence from robust fixed effects estimations. University of Lueneburg Working Paper Series in Economics No. 172, May.

Vinod, Hrishikesh D. and Bruce D. McCullough (2003), Verifying the Solution from a Nonlinear Solver: A Case Study. *American Economic Review* 93 (3), 873-892.

Vogel, Alexander and Joachim Wagner (2010), Exports and Profitability – First Evidence for German Business Services Enterprises. *Applied Economics Quarterly* 56 (1), 7-30.

Wagner, Joachim (1995), Exports, Firm Size, and Firm Dynamics. *Small Business Economics* 7 (1), 29 – 39.

Wagner, Joachim (1998), Bestimmungsgründe internationaler Firmentätigkeit. *Jahrbücher für Nationalökonomie und Statistik* 217 (5), 613-627.

Wagner, Joachim (2001), A Note on the Firm Size – Export Relationship. *Small Business Economics* 17 (4), 229-237.

Wagner, Joachim (2002), The Causal Effects of Exports on Firm Size and Labor Productivity: First Evidence from a Matching Approach. *Economics Letters* 77 (2), 287-292.

Wagner, Joachim (2003), Unobserved Firm Heterogeneity and the Size-Exports Nexus: Evidence from German Panel Data. *Review of World Economics* 139 (1), 161-172.

Wagner, Joachim (2006a), Export Intensity and Plant Characteristics: What Can We Learn From Quantile Regression? *Review of World Economics* 142 (1), 195-203.

Wagner, Joachim (2006b), Exports, foreign direct investment, and productivity: evidence from German firm level data. *Applied Economics Letters* 13, 347-349.

Wagner, Joachim (2007), Exports and Productivity: A Survey of the Evidence from Firm-Level Data. *The World Economy* 30 (1), 60-82.

Wagner, Joachim (2008a), Export Entry, Export Exit and Productivity in German Manufacturing Industries. *International Journal of the Economics of Business* 15 (2), 169-180.

Wagner, Joachim (2008b), Exports and Firm Characteristics – First Evidence from Fractional Probit Panel Estimates. University of Lueneburg Working Papers in Economics No. 97, August.

Wagner, Joachim (2009), Exporte und Firmenerfolg: Welche Firmen profitieren wie vom internationalen Handel? *AStA – Wirtschafts- und Sozialstatistisches Archiv* 3 (2), 109-122.

Yasar, Mahmut and Catherine J. Morrison Paul (2007), International linkages and productivity at the plant level: Foreign direct investment, exports, imports and licensing. *Journal of International Economics* 71 (2), 373-388.

Yasar, Mahmut, Carl H. Nelson and Roderick Rejesus (2006), Productivity and Exporting Status of Manufacturing Firms: Evidence from Quantile Regressions. *Review of World Economics* 142 (4), 675-694.

Table 1: Productivity differences between exporters and non-exporters – Full sample

	Exporters	Non-Exporters
Number of firms (share in all firms)	377 (61%)	241 (39 %)
Labour productivity (Sales in Deutschmark per employee)		
Mean	285,163	218,985
(Std. Dev.)	(327,911)	(272,110)
Test for difference in means (p-value)	0.007	
Percentiles		
1%	60,000	50,000
5%	100,000	74,783
10%	120,000	85,714
25%	154,429	120,000
50%	216,216	164,286
75%	307,692	234,286
90%	500,000	366,667
95%	666,667	462,500
99%	1,166,667	928,000
Smallest values		
	31,250	33,333
	53,763	35,294
	58,333	50,000
	60,000	50,000
Largest values		
	1,166,667	750,000
	1,634,021	928,000
	2,542,373	1,037,500
	5,000,000	3,837,209
Skewness	9.224	10.116
Kurtosis	121.311	131.568

Kolmogorov – Smirnov test for equality of distribution functions

	p-value
H_0: Equality of distributions	0.000
H_0: Productivity of non-exporters is larger	0.000
H_0: Productivity of exporters is larger	0.998

Table 2: Exporter productivity premium (percent) – Full sample

		OLS-estimates	Quantile regression estimates 0.10	0.25	0.50	0.75	0.90
Exporter premium (percent)		33.79	24.52	24.37	20.06	31.54	28.45
	p	0.000	0.008	0.001	0.008	0.000	0.000

Note: The exporter productivity premium is calculated from the estimated regression coefficients of the exporter dummy variable ß as 100*($e^{ß}$ – 1). The p-values for quantile regression estimates are based on standard errors bootstrapped with 500 replications. All regressions include the number of employees (also in squares) and 17 two-digit industry dummy variables.

Tests for coefficient equality between pairwise quantiles and across all quantiles

A: Pairwise

	p-value		p-value		p-value		p-value
0.10 vs. 0.25	0.987	0.25 vs. 0.50	0.580	0.50 vs. 0.75	0.220	0.75 vs. 0.90	0.777
0.10 vs. 0.50	0.699	0.25 vs. 0.75	0.512	0.50 vs. 0.90	0.460		
0.10 vs. 0.75	0.597	0.25 vs. 0.90	0.746				
0.10 vs. 0.90	0.787						

B: Joint test for all quantiles p-value = 0.810

Note: The null hypothesis is that the coefficients are equal between pairwise quantiles (panel A) and across all quantiles (panel B). Test statistics are based on the variance-covariance matrix of the coefficients of the system of quantile regressions used to compute the estimates for the exporter productivity premia reported in the upper panel of this table. P-values for the F-values are reported.

Table 3: Productivity differences between exporters and non-exporters – Sample without top and bottom one percent of productivity distribution

	Exporters	Non-Exporters
Number of firms (share in all firms)	371 (61 %)	236 (39 %)
Labour productivity (Sales in Deutschmark per employee)		
Mean	261,667	206,651
(Std. Dev.)	(165,912)	(138,654)
Test for difference in means (p-value)	0.000	
Percentiles		
1%	67,283	62,500
5%	100,000	75,000
10%	120,000	93,333
25%	154,429	125,540
50%	215,517	165,686
75%	307,190	235,000
90%	461,794	366,667
95%	608,273	462,500
99%	878,378	750,000
Smallest values		
	58,333	53,846
	60,000	38,824
	66,667	62,500
	67,283	64,000
Largest values		
	878,378	723,077
	909,091	750,000
	1,066,667	928,000
	1,066,667	1,037,500
Skewness	2.033	2.616
Kurtosis	7.928	12.544

Kolmogorov – Smirnov test for equality of distribution functions

	p-value
H_0: Equality of distributions	0.000
H_0: Productivity of non-exporters is larger	0.000
H_0: Productivity of exporters is larger	0.997

Table 4: Exporter productivity premium (percent) – Sample without top and bottom one percent of productivity distribution

		OLS-estimate	Quantile regression estimates 0.10	0.25	0.50	0.75	0.90
Exporter premium (percent)		28.51	22.29	18.95	19.17	29.35	28.74
	p	0.000	0.003	0.005	0.004	0.001	0.003

Note: The exporter productivity premium is calculated from the estimated regression coefficients of the exporter dummy variable ß as 100*($e^{ß}$ – 1). The p-values for quantile regression estimates are based on standard errors bootstrapped with 500 replications. All regressions include the number of employees (also in squares) and 17 two-digit industry dummy variables.

Tests for coefficient equality between pairwise quantiles and across all quantiles

A: Pairwise

	p-value		p-value		p-value		p-value
0.10 vs. 0.25	0.672	0.25 vs. 0.50	0.976	0.50 vs. 0.75	0.200	0.75 vs. 0.90	0.952
0.10 vs. 0.50	0.739	0.25 vs. 0.75	0.297	0.50 vs. 0.90	0.390		
0.10 vs. 0.75	0.532	0.25 vs. 0.90	0.432				
0.10 vs. 0.90	0.620						

B: Joint test for all quantiles p-value = 0.771

Note: The null hypothesis is that the coefficients are equal between pairwise quantiles (panel A) and across all quantiles (panel B). Test statistics are based on the variance-covariance matrix of the coefficients of the system of quantile regressions used to compute the estimates for the exporter productivity premia reported in the upper panel of this table. P-values for the F-values are reported.

Table 5: A comparison of OLs and robust estimates of the exporter productivity premium

	OLS-estimate	Median regression estimate	Huber M-estimate	MM-estimate	OLS-estimate without outliers
Exporter premium (percent)	33.79	20.06	25.85	15.53	19.52
p	0.000	0.008	0.000	0.008	0.000

Note: The exporter productivity premium is calculated from the estimated regression coefficient of the exporter dummy variable ß as $100*(e^{ß} – 1)$. Median regression is the quantile regression for the 0.50 quantile; the p-value is based on standard errors bootstrapped with 500 replications. The Huber M-estimate uses the rreg-command implemented in Stata. The MM-estimate uses mmregress (Verardi and Croux 2009) with the efficiency set at 0.7. The OLS-estimate without outliers uses the sample without the 53 outliers (31 vertical outliers that are not outlying in the space of explanatory variables and 22 bad leverage points) identified by the MM-estimator. All regressions include the number of employees (also in squares) and 17 two-digit industry dummy variables.

Table 6: Firm transitions in the German export market, 1995 - 2006

	20 – 49 workers				50 – 249 workers				250 – 499 workers				+499 workers			
Year t status	No exports		Exports		No exports		Exports		No exports		Exports		No exports		Exports	
Year t+1 Status	No Ex-Ports	Exports	No Ex-ports	Exports	No Ex-ports	Exports	No Ex-ports	Exports	No Ex-ports	Exports	No Ex-ports	Exports	No Ex-ports	Exports	No Ex-ports	Exports
Average 1995-2006	0.9346	0.0654	0.0448	0.9552	0.9267	0.0733	0.0234	0.9766	0.9280	0.0720	0.0119	0.9881	0.9321	0.0679	0.0110	0.9890

Note: The value of each figure is the proportion of firms in each year t statuses (exports – no exports) that chooses each of the two possible statuses in year t+1