Verardi, Vincenzo; Wagner, Joachim

Working Paper
Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium

Working Paper Series in Economics, No. 168

Provided in Cooperation with:
Institute of Economics, Leuphana Universität Lüneburg

Suggested Citation: Verardi, Vincenzo; Wagner, Joachim (2010) : Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium, Working Paper Series in Economics, No. 168, Leuphana Universität Lüneburg, Institut für Volkswirtschaftslehre, Lüneburg

This Version is available at:
http://hdl.handle.net/10419/57109

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Robust Estimation of Linear Fixed Effects Panel Data Models with an Application to the Exporter Productivity Premium

by
Vincenzo Verardi and Joachim Wagner

University of Lüneburg
Working Paper Series in Economics

No. 168
April 2010

www.leuphana.de/institute/ivwl/publikationen/working-papers.html

ISSN 1860 - 5508
Robust Estimation of Linear Fixed Effects Panel Data Models with an Application to the Exporter Productivity Premium

Vincenzo Verardi and Joachim Wagner

[This version: April 28, 2010]

Abstract:

In empirical studies it often happens that some variables for some units are far away from the other observations in the sample. These extreme observations, or outliers, often have a large impact on the results of statistical analyses – conclusions based on a sample with and without these units may differ drastically. While applied researchers tend to be aware of this, the detection of outliers and their appropriate treatment is often dealt with in a rather sloppy manner. One reason for this habit seems to be the lack of availability of appropriate canned programs for robust methods that can be used in the presence of outliers. Our paper intents to improve on this situation by presenting a highly robust method for estimation of the popular linear fixed effects panel data model, and to supply Stata code for it. In an application from the field of the micro-econometrics of international firm activities we demonstrate that outliers can indeed drive results.

Keywords: Robust estimation, panel data, outliers, Stata, exporter productivity premium

JEL Classification: C23, C81, C87, F14

* All computations were done in the research data centre of the Statistical Office in Berlin. The data used are confidential but not exclusive; information how to access the data is provided in Zühlke et al. (2004). To facilitate replication and extensions Stata code for the robust estimation of fixed effects linear panel data models is available from the first author, and the Stata do-files used to compute the empirical results in the application are available from the second author on request.

Vincenzo Verardi
University of Namur (CRED) and
Université Libre de Bruxelles (ECARES and CKE)
Rempart de la Vierge 8
B-5000 Namur, Belgium
e-mail: vverardi@fundp.ac.be

Joachim Wagner (corresponding author)
Leuphana University Lueneburg
Institute of Economic
P. O. Box 2440
D-21314 Lüneburg, Germany
e-mail: wagner@leuphana.de
1. Motivation

In his Nobel-lecture James Heckman (2001, p. 674 and p. 732) pointed out that “(t)he most important discovery [from micro-econometric investigations] was the evidence on the pervasiveness of heterogeneity and diversity in economic life.” Everybody who ever worked with individual or firm level data will strongly agree that if one investigates a sample of heterogeneous economic units it often happens that some variables for some individuals or firms are far away from the other observations in the sample. These extreme observations, or outliers, often have a large impact on the results of statistical analyses – conclusions based on a sample with and without these units may differ drastically.

While applied researchers tend to be aware of this, the detection of outliers and their appropriate treatment is usually not considered as an important issue. Often the distribution of some variables with extreme values is trimmed by dropping the top or bottom one percent of observations or so, or other ad hoc procedures are used. Given the large literature on statistical methods that are robust to outliers¹ and the (at least, potentially) detrimental consequences of ignoring them this habit should change.

One reason for the usually sloppy habit towards outliers seems to lie in the lack of availability of appropriate canned programs in the popular software used by applied economists. At least with regard to Stata this changed recently due to the publication of code for highly robust methods in Verardi and Croux (2009) where, however, methods for the robust analysis of cross section data are dealt with only. Fixed effects models for panel data that are very popular in applied economics are

¹ For a recent comprehensive textbook treatment see Maronna, Martin and Yohai (2006)
not covered. In this paper we will close this gap by presenting a highly robust procedure for the estimation of linear fixed effects panel data models.

To demonstrate that an appropriate treatment of outliers does make a difference we report results from an empirical application. The example that will be used for motivation and illustration is taken from the field of international firm activities. This topic is selected for the sole reason that one of us is active in this area for a long time – any other field from applied economics could serve as an illustrative example, too.

That said, note that 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.²

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.³ These empirical studies inspired a number of theoretical papers that model the behavior of heterogeneous firms in open economies.⁴ 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-

---

² Pioneering papers in this field include Bernard and Jensen (1995) and Wagner (1995).

³ For partial surveys of this empirical literature see Greenaway and Kneller (2007), López (2005), and Wagner (2007).

⁴ 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.
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.

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 example, in a sample of exporting and non-exporting firms there usually 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 – the ceteris paribus productivity difference between exporting and non-exporting firms. 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).

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. Section 2 will present such a highly robust procedure for the estimation of linear fixed effects panel data models. In section 3 this method will be used to estimate the exporter productivity premium - the ceteris paribus productivity difference between exporting and non-exporting firms – in Germany, and the results will be compared to the results from using the standard fixed effects estimator to demonstrate that outliers do make a large difference. Section 4 concludes.

2. Robust Estimation of Linear Fixed Effects Panel Data Models

In cross-sectional regression analysis, three types of outliers can cause least squares to breakdown. Rousseeuw and Leroy (1987) define them as vertical outliers, bad leverage points and good leverage points. Vertical outliers are observations that are outlying in the y-dimension but not in the space of the explanatory variables (x-variables). Their existence affects both the estimation of the intercept and of the regression coefficients, but the effect on the latter is milder. Bad leverage points are observations that are both outlying in the space of the explanatory variables and located far from the regression line. They severely affect the estimation of both the intercept and the slope coefficients. Finally, good leverage points are observations that are outlying in the space of the explanatory variables but are located close to the
regression line. Their existence only marginally influences the estimation of both the intercept and the regression coefficients but does affect inference. When working with Panel data, a fourth category of outliers should be considered, namely block concentrated outliers that correspond to a situation in which most of outlying observations are concentrated in a limited number of time series (see Bramati and Croux, 2007).

To deal with the presence of any of these types of outliers, Bramati and Croux (2007) propose two equally well performing estimators, the Within Groups Generalized M-estimator (WGM) and the Within Groups MS-estimator (WMS). The idea underlying both, is to center the series in a similar way to what is generally done when applying the within transformation. The difference here is that 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.

In this paper, we use exactly the same logic to robustly estimate a fixed effect model. We first center the entire series to remove individual fixed effects and then run a robust estimator to identify the outliers. Outlying individuals are then awarded a weight zero and a standard fixed effect model is fitted to the remaining observations. The robust estimator we use for the outlier identification step is an S-estimator which is known to be particularly robust to outliers. The logic behind this estimator is that, instead of minimizing the variance of the residuals as in OLS, another measure of dispersion of the residuals, less sensitive to outliers, is minimized. The measure of spread minimized here is an M-estimator of scale (see Verardi and Croux (2009) for further details).
Technically speaking, consider the general formulation of the fixed effects linear panel data model.

\[ y_{it} = \alpha_i + x_{it}^T \beta + e_{it} \quad t = 1, \ldots, T \]

where subscript \( i \) denotes the cross-section dimension, whereas \( t \) denotes the time series dimension. The \( y_{it} \) term denotes the dependent variable, \( x_{it} \) is the \( K \times 1 \) column vector of explanatory variables, \( \beta \) is a \( K \times 1 \) vector of the regression parameters and the \( u_i \)'s are the unobservable time-invariant individual fixed effects. Finally, the \( e_{it} \)'s denote the disturbance terms which are assumed to be uncorrelated through time and cross-sections.

The first step is therefore to center the variables. This leads to a set of new variables defined as \( \bar{y}_{it} = y_{it} - \text{med}_i y_{it} \) and \( \bar{x}_{it}^{(j)} = x_{it}^{(j)} - \text{med}_i x_{it}^{(j)} \) where \( x_{it}^{(j)} \) (for \( j = 1, \ldots, K \)) is the \( j \)th explanatory variable measured for individual \( i \) at time \( t \).

The second step consists in regressing \( \bar{y}_{it} \) on the \( \bar{x}_{it}^{(j)} \)'s using an S-estimator and thereby obtaining the estimated parameters:

\[ \hat{\beta} = \arg \min_\beta \sigma(r_1(\beta), \ldots, r_T(\beta)) \]

where \( r \) are the estimated residuals and \( \sigma \) is an M-estimator of scale (i.e. a measure of dispersion that withstands the presence of outlying individuals, see Dehon et al., 2009).

Having obtained the residuals and the estimated measure of dispersion, by relying on the assumed normality of the residuals, we can easily identify the outlying observations by flagging those individuals that have robust standardized residuals (i.e. residuals obtained by the S-estimator divided by \( \hat{\beta} \)) that are larger than 2. The final step is then to run a standard fixed-effect estimation awarding a weight zero to the outliers.
3. **Application: The Exporter Productivity Premium in Germany**

Next, the method proposed in section 2 is used to estimate the exporter productivity premium - the ceteris paribus productivity difference between exporting and non-exporting firms\(^5\) – for firms from manufacturing industries in West Germany, and the results will be compared to the results from using the standard fixed effects estimator to demonstrate that outliers do make a large difference. The empirical study uses pooled data for the years 1995 to 2006.\(^6\) 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.

Results are reported in table 1. 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\(^7\), this estimate (while still

---

\(^5\) Note that this is a core topic in the literature on the micro-econometrics of international firm activities; see Wagner (2007) for a survey and International Study Group on Exports and Productivity (ISGEP) (2008) for an international comparison.

\(^6\) For a description of the data see Malchin and Voshage (2009).

\(^7\) Note that the estimated coefficient \(\beta\) of the dummy variable from the semi-logarithmic model has to be transformed by \((e^{\beta} - 1) \times 100\) to get the percentage difference between exporters and non-exporters.
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\(^8\) 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.

[Table 1 near here]

4. Concluding remarks

Researchers active in applied microeconomics are often aware of the fact that extreme observations, or outliers, can have a large impact on the results of statistical analyses, and that conclusions based on a sample with and without these units may differ drastically. To our experience, however, the detection of outliers and their appropriate treatment is often dealt with in a rather sloppy manner, not least due to the lack of availability of appropriate canned programs for methods that are robust against these extreme observations. The highly popular linear fixed effects panel data estimator is a case in point.

\(^8\) Note that the estimated maximum of the inversely u-shaped relationship between the share of exports in total sales and labour productivity is 168.84 percent, and, therefore, the negative sign of the estimated coefficient for the squared value of the share of exports in total sales does not indicate that there is a point after which a further increase in exports in total sales is related to a lower value of labour productivity.
Our paper intents to help to improve this situation by presenting a highly robust method for the estimation of linear fixed effects panel data models and to supply Stata code for it. An application from the field of the micro-econometrics of international firm activities demonstrates how a small fraction of outliers drives the results of an empirical estimation. We hope that these results motivate the routine application of robust methods in future micro-econometric investigations.

References


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


Table 1: Exporter Productivity Premia for West German Manufacturing Enterprises: Results for Standard and Robust Fixed-Effects Estimation

<table>
<thead>
<tr>
<th></th>
<th>Standard Fixed Effects Regression</th>
<th>Robust Fixed Effects Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exporter (Dummy; 1 = yes)</td>
<td>$\beta$ 0.126, $p$ 0.000</td>
<td>$\beta$ 0.00992, $p$ 0.000</td>
</tr>
<tr>
<td>Number of observations</td>
<td>303,294</td>
<td>265,628</td>
</tr>
<tr>
<td>Number of enterprises</td>
<td>34,570</td>
<td>33,510</td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of exports in total sales</td>
<td>$\beta$ 0.00466, $p$ 0.000</td>
<td>$\beta$ 0.000728, $p$ 0.000</td>
</tr>
<tr>
<td>Share of exports in total sales (squared)</td>
<td>$\beta$ -0.0000138, $p$ 0.008</td>
<td>$\beta$ 0.0000182, $p$ 0.000</td>
</tr>
<tr>
<td>Number of observations</td>
<td>303,294</td>
<td>265,797</td>
</tr>
<tr>
<td>Number of enterprises</td>
<td>34,570</td>
<td>33,508</td>
</tr>
</tbody>
</table>

Note: Results are for pooled enterprise data from 1995 – 2006. The dependent variable is the log of labour productivity (defined as sales per employee; in Euro). Both empirical models include the number of employees and its squared value plus year dummy variables and a constant. $\beta$ is the estimated regression coefficient, $p$ is the prob-value.
Working Paper Series in Economics

(recent issues)

No.167:  
Stephan Humpert: Machen Kinder doch glücklich? April 2010

No.166:  
Joachim Wagner: Produktivität und Rentabilität in der niedersächsischen Industrie im Bundesvergleich. Eine Benchmarking-Studie auf der Basis vertraulicher Firmendaten aus Erhebungen der amtlichen Statistik, April 2010

No.165:  
Nils Braakmann: Neo-Nazism and discrimination against foreigners: A direct test of taste discrimination, March 2010

No.164:  
Amelie Boje, Ingrid Ott and Silvia Stiller: Metropolitan Cities under Transition: The Example of Hamburg/ Germany, February 2010

No.163:  
Christian Pfeifer and Stefan Schneck: Relative Wage Positions and Quit Behavior: New Evidence from Linked Employer-Employee-Data, February 2010

No.162:  
Anja Klaubert: “Striving for Savings” – religion and individual economic behavior, January 2010

No.161:  
Nils Braakmann: The consequences of own and spousal disability on labor market outcomes and objective well-being: Evidence from Germany, January 2010

No.160:  
Norbert Olah, Thomas Huth und Dirk Löhr: Geldpolitik mit optimaler Zinsstruktur, Januar 2010

No.159:  

No.158:  

No.157:  
Institut für Volkswirtschaftslehre: Forschungsbericht 2009, Januar 2010

No.156:  

No.155:  
Nils Braakmann and Joachim Wagner: Labor market adjustments after a great import shock: Evidence from the German clothing industry and the Multi-Fibre Arrangement, December 2009

No.154:  
Joachim Wagner: Zehn Jahre European Data Watch: Dokumentation von Datensätzen für die empirische Wirtschafts- und Sozialforschung und Zugangswegen zu den Daten, Dezember 2009

No.153:  
Joachim Wagner: Offshoring and work performance: Self-Selection, effects on performance, or both? December 2009

No.152:  
Christian Pfeifer: Effective Working Hours and Wages: The Case of Downward Adjustment via Paid Absenteeism, November 2009

No.151:  

No.150:  
Franziska Boneberg: Recht und Realität von Mitbestimmung im westdeutschen Dienstleistungssektor: 11 Fallstudien, November 2009
No.149: Birgit Müller, Martin Quaas, Karin Frank and Stefan Baumgärtner: Pitfalls and potential of institutional change: Rain-index insurance and the sustainability of rangeland management, November 2009


No.147: Markus Groth: Potentiale und Risiken der Nutzung von Methan aus Methanhydraten als Energieträger, Oktober 2009

No.146: Sandra Derissen, Martin Quaas and Stefan Baumgärtner: The relationship between resilience and sustainable development of ecological-economic systems, October 2009


No.143: Nils Braakmann: Are there social returns to both firm-level and regional human capital? – Evidence from German social security data. September 2009

No.142: Nils Braakmann and Alexander Vogel: How does economic integration influence employment and wages in border regions? The case of the EU-enlargement 2004 and Germany’s eastern border, September 2009


No.137: Roland Olbrich, Martin F. Quaas and Stefan Baumgärtner: Sustainable use of ecosystem services under multiple risks – a survey of commercial cattle farmers in semi-arid rangelands in Namibia, September 2009

No.136: Joachim Wagner: One-third codetermination at company supervisory boards and firm performance in German manufacturing industries: First direct evidence from a new type of enterprise data, August 2009


No.132: **Stefan Baumgärtner and Sebastian Strunz:** The economic insurance value of ecosystem resilience, July 2009

No.131: **Matthias Schröter, Oliver Jakoby, Roland Olbrich, Marcus Eichhorn and Stefan Baumgärtner:** Remote sensing of bush encroachment on commercial cattle farms in semi-arid rangelands in Namibia, July 2009

No.130: **Nils Braakmann:** Other-regarding preferences, spousal disability and happiness: Evidence for German Couples, May 2009


No.128: **Sebastian Troch:** Drittelbeteiligung im Aufsichtsrat – Gesetzliche Regelung versus Unternehmenspraxis. Ausmaß und Bestimmungsgründe der Umgehung des Drittelbeteiligungs gesetzes in Industrieunternehmen, Mai 2009


No.126: **Nils Braakmann:** The role of firm-level and regional human capital for he social returns to education – Evidence from German social security data, April 2009


No.124: **Nils Braakmann and Alexander Vogel:** The impact of the 2004 EU-enlargement on enterprise performance and exports of service enterprises in the German eastern border region, April 2009 [revised version forthcoming in: Review of World Economics]

No.123: **Alexander Eickelpasch and Alexander Vogel:** Determinants of Export Behaviour of German Business Services Companies, March 2009

No.122: **Maik Heinemann:** Stability under Learning of Equilibria in Financial Markets with Supply Information, March 2009

No.121: **Thomas Wein:** Auf der Speisekarte der DPAG: Rechtliche oder ökonomische Marktzutritts schranken? März 2009

No.120: **Nils Braakmann und Joachim Wagner:** Product Diversification and Stability of Employment and Sales: First Evidence from German Manufacturing Firms, February 2009

No.119: **Markus Groth:** The transferability and performance of payment-by-results biodiversity conservation procurement auctions: empirical evidence from northernmost Germany, February 2009

No.118: **Anja Klaubert:** Being religious – A Question of Incentives? February 2009


No.116: **Alexander Vogel und Joachim Wagner:** Import, Export und Produktivität in niedersächsischen Unternehmen des Verarbeitenden Gewerbes, Januar 2009

No.115: **Nils Braakmann and Joachim Wagner:** Product Differentiation and Profitability in German Manufacturing Firms, January 2009

(see www.leuphana.de/institute/ivwl/publikationen/working-papers.html for a complete list)