A robust version of the FGLS estimator for panel data

Anabela Rocha* and M. Cristina Miranda**

XXV Congresso
Sociedade Portuguesa
de Estatística
2021 Évora

ISCA e CIDMA, Universidade de Aveiro
CEAUL, Universidade de Lisboa

Outline

1. Motivation.
2. Panel data model.
3. Robust estimation of the panel data model.
4. An illustration with Grunfeld data.
5. Final comments.
6. References.

Panel data (PD) or longitudinal data:

- Suitable for describing economic and financial data.
- Panel data models (PDM) may consider a time or firm effect.
- Real data contain atypical observations (AO).
- AO affects PD classic estimation.
- Robust PD model estimation less affected by AO rarely used in econometric studies.

Panel data

- The records of a set of variables for a period of time (panel data) may reveal more information then the observation of the same variables in a single moment (cross sectional data).
- Take a set of firms with records of some variables of interest in each firm for 10 years e.g.
- Typically a PD set consists in a number of observations of different units taken repeatedly for a number of times, usually small when compared with the number of units under study.

Panel data

- The units may be firms, **individuals**, countries, families, etc., depending on the object of study - macro or micro economics.
- One of the main motivations of this approach is the presence of heterogeneity among individuals.
- Panel data models allow for detection of effects that would be imperceptible with cross sectional data or with time series data.

A panel data model may be expressed by the following equation:

$$y_{it} = \mathrm{x}_{it}\boldsymbol{\beta} + u_{it}, \tag{1}$$

with disturbance term $u_{it} = (\mu_i + \nu_{it})$, i - individuals, firms, coutries, etc.; t - time period; y_{it} - observation of the dependent variable; $\boldsymbol{\beta}$ - vector of parameters; x_{it} - observation of the independent variables; μ_i - unobservable individual-specific effect; ν_{it} - remainder disturbance.

Panel data model (PDM)

With DPM with random effects, we assume the following:

- μ_i iid with zero mean and variance σ_μ^2,
- ν_{it} iid with zero mean and variance σ_ν^2,
- μ_i and ν_{it} independents,
- x_{it} independents with μ_i and $\nu_{it}, \forall i, t$.

Depending on the assumed hypothesis, we'll have different models: pooled, fixed or random effects.

Classic estimation of PDM

Feasible Generalized Least Squares (FGLS) is the usual way to estimate the PDM and may be expressed by the following:

$$\hat{\beta}_{FGLS} = \left(\mathrm{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathrm{X} \right)^{-1} \mathrm{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathrm{y}, \tag{2}$$

where $\hat{\boldsymbol{\Omega}}$ represents an estimate of the covariance matrix of the errors of the model. Frequently we use the sample covariance matrix of the residuals.

- Panel data - observations of several variables for a period of some years/months and coutries/firms panels - often present atypical observations or outliers.
- In this type of data its important to detect multivariated outliers - visual observation is not easy.
- Classical methods of PDM estimation (FGLS) may be seriously affected with the presence of outliers.

Robust estimation of the PDM

- Robust estimation - is not seriously affected by the presence of outliers (observations with low probability of belonging to the same distribution of the majority of the data).
- Some robust procedures have been proposed for PDM: Bramati and Croux (2007), Aquaro and Cízek (2013, 2018), Dhaene and Zhu (2017) - They have adapted robust regression methods (Rousseeuw and Leroy, 1987) to DPM.
- There are few papers with application of robust methodologies in the fields of economics and finance.

Robust estimation of the PDM

Proposal: to robustify the *FGLS* estimator, recalling that its implementation includes three steps:

1. estimate the parameters of the PLS (Pooled Least Squares) and collect the residuals(error estimates),
2. estimate the covariance matrix of the errores $\boldsymbol{\Omega}$ with the residuals of the former step,
3. estimate the FGLS parameters with the estimated covariance matrix obtained in the former step.

RFGLS (Robust Feasible Generalized Least Squares)

1. estimate the PLS parameters and computaion of the residuals,
2. **robust estimation** of the covariance matrix of the errros - **CovOGK**,
3. FGLS parameters estimation using the estimated covariance matrix covOGK.

Robust estimation of the PDM

Robust covariance matrix estimator - *OGK* (Orthogonalized Gnanadesikan-Kettenring) (Maronna and Zamar, 2002):

- allow to obtain robust location and scale estimators,
- estimates each pair of variables covariancees in a robust way,
- it is the result of a transformation of a previously proposed estimator (Gnanadesikan and Kettenring, 1972):

$$\sigma_{XY} = cov(X, Y) = \frac{\sigma^2(X + Y) - \sigma^2(X - Y)}{4},$$

where $\sigma^2(\cdot)$ is a variance functional of $(\cdot)$. A robust estimator is obtained when $\sigma^2(\cdot)$ is a robust functional.

Grunfeld data

Grunfeld data consist in annual records from some american firms from 1935 to 1954.
There is a total of 220 observations - 20 years and 11 firms - 5 variables:

- ***invest*** - investment in dolares - **dependent variable**;
- ***value*** - market value- **independent variable**;
- ***capital*** - capital - **independent variable**;
- ***firm*** - General Motors, US Steel, General Electric, Chrysler, Atlantic Refining, IBM, Union Oil, Westinghouse, Goodyear, Diamond Match, American Steel;
- ***year*** - years from 1935 to 1954.

There are multiple versions of these data. We used the later version available at
https://eeecon.uibk.ac.at/ zeileis/grunfeld/Grunfeld.csv.

Grunfeld data - model fit

The original *Grunfeld* data were first analysed in the frame of Grunfeld's PHd in Economics,where he studied *"The Determinants of Corporate Investment"* (Grunfeld, 1958). This set of data continues to be used to illustrate different panel data studies for research and educational purpose.

Grunfeld formulated a model that describes the dependence of the investment from (y) the value (x_1) and the capital (x_2) (Gujarati, 2003).

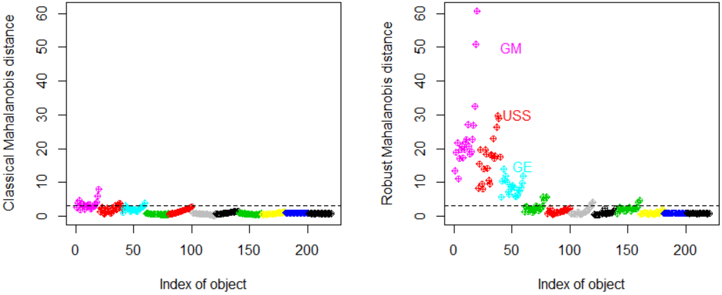
investment equation:

$$invest_{it} = \beta_0 + \beta_1 value_{it} + \beta_2 capital_{it} + u_{it}, \tag{3}$$

com $i = 1, ..., 11$ e $t = 1, ..., 20$.

Grunfeld data - outliers detection

Multivariate outliers detection (Rousseeuw and Van Zomeren, 1990) with robust Mahalanobis distance - *Moutlier* function from R *package chemometrics*.



The robust method reveals three firms: **GM, USS, GE**.

Grunfeld data - parameter estimation

The presence of outliers - 3 firms: **GM, USS, GE** - justifies the use of robust estimation methods.

The estimation was made for the fixed effects model and for the random effects model, based in FGLS and RFGLS, respectively.

We used two R *packages*:
FGLS - function *pggls* from *plm* (panel data analysis)
RFGLS - function *covOGK* from *robustbase* (robust methods) and we adapted the *pggls* function to include the robust covariance matrix.

Model parameters estimates:

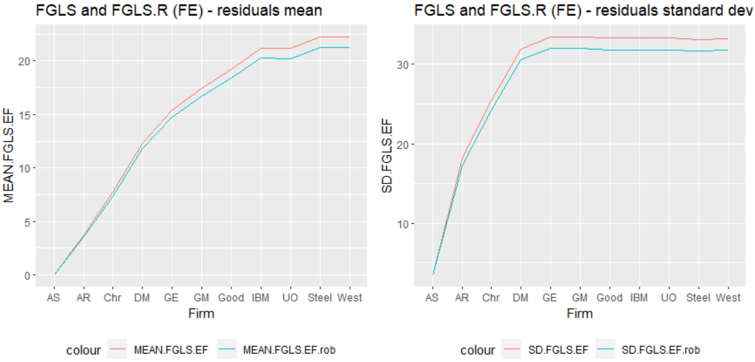
There are some differences between the estimates obtained with the two estimators - FGLS e RFGLS.

	EF		EA	
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
FGLS	0.110	0.309	0.114	0.228
RFGLS	0.113	0.295	0.124	0.191

We performed a residual analysis in order to evaluate the performance of the two estimators.

XXV Congresso Sociedade Portuguesa de Estatística 2021 Évora

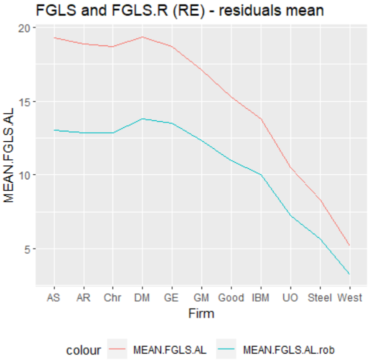
Performance of the robust estimator RFGL - residual analysis with fitted model FE.

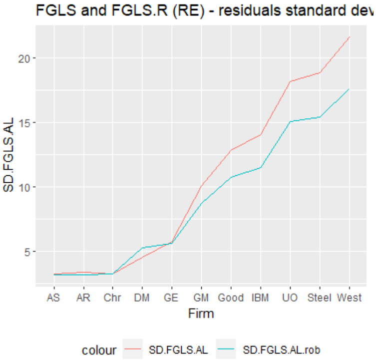


FE model - the residuals obtained after applying robust method **RFGLS** present **less mean values and less standard deviation for almost every companies**.

XXV Congresso Sociedade Portuguesa de Estatística 2021 Évora

Performance of the robust estimator RFGL -residual analysis of the model RE.





RE model - The residuals produced with robust method **RFGLS** present **less mean values and less standard deviations for almost every companies**.

Outcomes

- **Residuals mean and standard deviation were lowers** for the robust method.
- **For both type of models**, random or fixed effects, the residuals obtained with the robust method presented lower mean and lower standard deviation.
- **The robust estimated model is less affected by the identified outliers of the three firms: GM, USS, GE.**

Final comments

- Panel data - suitable representation for economical and financial data.
- Financial and economic real data often contain outliers.
- Robust methods are recommended for this type of data analysis.
- **Robust Mahalanobis distance** - turned possible to detect outliers that were present in the *Grunfeld* data set.
- **The robust estimator was the one with better performance**: residuals with less mean and less standard deviation.

Future work

- simulation study - RFGLS estimator properties;
- inserting robust procedures in the remainder steps of FGLS estimator.

Thank you.

This work is supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020.

REPÚBLICA
PORTUGUESA

CIDMA CENTRO DE I&D EM MATEMÁTICA E APLICAÇÕES CENTER FOR R&D IN MATHEMATICS AND APPLICATIONS

FCT Fundação para a Ciência e a Tecnologia

References

[1] Aquaro, M. and Cizek, P., One-step robust estimation of fixed-effects panel data models, *Statistics and Data Analysis*, Vol. 57, No. 1, pp. 536-548, 2013.

[2] Aquaro, M. and Cizek, P., Robust estimation and moment selection in dynamic fixed-effects panel data models, *Computational Statistics*, Vol. 33, No. 10, pp. 675–708, 2018.

[3] Baltagi, B.H., Econometric analysis of panel data, Wiley, 2013.

[4] Bramati, M. and Croux, C., Robust estimators for the fixed effects panel data model, *The Econometrics Journal*, Vol. 10, No. 3, pp. 521-540, 2007.

[5] Dhaene, G. and Zhu, Y., Median-based estimation of dynamic panel models with fixed effects, *Computational Statistics and Data Analysis*, Vol. 113(C), pp. 398-423, 2017.

[6] Greene, W., Econometric Analysis, Seventh edition, Prentice Hall, 2011.

[7] Grunfeld, Y., The Determinants of Corporate Investment, Ph.D. thesis, Department of Economics, University of Chicago, 1958.

[8] Maronna, R. and Zamar, R., Robust estimates of location and dispersion of high-dimensional datasets; *Technometrics*, Vol. 44 No.4, pp. 307–317, 2002.

[9] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). https://www.R-project.org/.

[10] Rousseew P. e Leroy, A., Robust regression and outlier detection. John Wiley & Sons, New York, 1987.