DOI: 10.20472/IAC.2019.052.039

**TATIANA MEDKOVÁ**
**University of Economics, Prague, Czech Republic**

# ROBUST STATISTICAL INFERENCE IN PANEL DATA INCLUDING A PRACTICAL APPLICATION IN THE MIGRATION ANALYSIS


## Abstract:

The need for robust statistical inference is well-documented even in the elementary case of a regression with a randomly sampled cross section. The usual ordinary least square standard errors are generally biased under the presence of heteroskedasticity; a phenomenon that seems to be a rule rather than an exception in applied anaylsis. The article describes several methods to deal with the biased standard errors grouping them in two categories: sandwich variance estimators and multi-way clustering. Moreover, the empirical application is included. An analysis of migration in European countries using the theory of gravity model is done applying several standard errors correction methods.



## Keywords:

migration, robust, standard errors, heteroskedasticity, panel data

**JEL Classification:** C23, C51, F22

## Introduction

The paper describes general methods that can be combined with panel data in order to correct estimations of standard errors where presence of clustering in the data set leads to heteroscedasticity and unrobust estimations. The need for robust statistical inference is well-documented even in the elementary case of a regression with a randomly sampled cross section. The usual OLS standard errors are generally biased under the presence of heteroskedasticity. The biased standard errors is an often problem that researchers deal with in applied empirical analysis. Angrist and Pischke (2008) note that in large samples where bias is not likely to be a problem and heteroscedasticity is detected, we typically see standard errors increase by about 25 percent when moving from the conventional estimator to the estimator with degrees of freedom correction.

Cameron and Trivedi (2005) presented an example where they generated data with conditionally heteroscedastic errors $\mathrm{Var}(\mathrm{u}|\mathrm{x}) = \mathrm{x}$ and estimated the model $\mathbf{y} = \mathbf{J} + \mathbf{x} + \mathbf{u},$ where $\mathbf{J}$is a vector of ones,$\mathrm{x}_i \sim \mathrm{N}(0{,}25), \mathrm{u}_i = \mathrm{x}_i \epsilon_i$with$\epsilon_i \sim \mathrm{N}(0{,}4)$. The difference between conventional OLSstandard errors and heteroscedasticity-robust ones is much wider than that quoted by Angristand Pischke. In the limit, the robust standard errors are$\sqrt{3}$ times larger.

This demonstrates that a failure to use robust standard errors can lead to quite different conclusions in statistical inference. The issue can be even much more pronounced in the case ofpanel data. This data typeposes challenges regarding valid statistical inference. In a paneldataset, the following issues invalidate the conventional standard errors:

- heteroscedasticity of the “usual type”, where the variance of random errors is affected bythe values of explanatory variables;
- between-cluster heteroskedasticity, where different clusters exhibit different variance ofrandom errors,
- within-cluster correlation of disturbances (often called intra-class correlation in thiscontext) due to the presence of unobserved factors or in theform of serial correlation over time.

The latter issue is especially important and may lead to conventional errors that are severely biased; Cameron and Trivedi (2009) show an example with real-life large-sample data where therobust standard errors are 2–3 times larger than conventional.

The remainder of the article is structured as follows. In the next section, we theoretically describe how to deal with the adverse effects of clustering on statistical inference in two categories: sandwich variance estimators and multi-way clustering. The empirical section shortly shows empirical application in the migration analysis using panel data about 8 European countries between 2011 -2014.

## 1 Robust statistical inference

As a starting point, consider a simple regression where the error term is expected to have aclustered structure of the short panel model with unobserved heterogeneityfor individual $i$ in group (cluster) $g$

$$y_{ig} = \beta_0 + \beta_1 x_{ig} + u_{ig}, \tag{1}$$

$$u_{ig} = c_g + v_{ig}, \tag{2}$$

where $c_g$ is a random component specific to class $g$ and $v_{ig}$ is the left-over disturbance; both $c_g$and $v_{ig}$are assumed to be iid errors. Moulton (1986) pointed out that when regressors vary onlyat the group level, an error structure like above can increase standard errors sharply. Angristand Pischke (2008) introduce the Moulton factor that expresses how much we over-estimateprecision by ignoring intra-class correlation. The Moulton factor is defined as

$$\frac{V(\hat{\beta}_1)}{V_{OLS}\,(\hat{\beta}_1)} = 1 + (n-1)\rho \tag{3}$$

where $V_{OLS}(\hat{\beta}_1)$ is the conventional OLS variance formula for the regression slope parameter,$V(\hat{\beta}_1)$is the correct sampling variance formula given the error structure (2), $\rho$ is called theintra-class correlation coefficient and is defined as

$$\rho = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_v^2} \tag{4}$$

where $\sigma_c^2$ is the variance of $c_g$ and $\sigma_v^2$ is the variance of $v_{ig}$. The Moulton factor tells us thatconventional standard errors become increasingly misleading as $n$ and $\rho$ increase. Angrist andPischke (2008) explain that the Moulton factor increases with group size because with a fixedoverall sample size, larger groups mean fewer clusters. Since observations are assumed to beindependent across clusters but not within, there is less independent information in the sample.

A straightforward way to avoid the Moulton factor issue is to apply a regression model thatdirectly accounts for the disturbance structure. For (1), one could employ the RE estimator, along with its conventional estimator of the coefficients' variance matrix. However, this conventional estimator is only consistent as long as the both disturbancecomponents are iid, or at least homoscedastic and non-correlated.

The usual way to deal with this issue is to use a robust variance matrix estimator, producing theso-called cluster-robust standard errors. In addition to protecting the intra-class correlation,this estimator is also consistent under the presence of both heteroscedasticity and within-clusterserial correlation (in the case of panel data).

## 1.1 Sandwich variance estimators

The theory of robust variance estimators dates back to the seminal papers of Huber (1967),Eicker (1967) and White (1980). None of these papers actually deals with the problem ofclustering, but all of them point towards a sandwich structure of the variance matrix that hasfavourable robustness properties (e.g. against heteroskedasticity). As noted by Williams (2000),the adjustment of the Huber-Eicker-White results to the clustering problem had been aroundshortly after the seminal papers, but was poorly documented; see Froot (1989) for one of theearly references and Wooldridge (2010) for a detailed textbook exposition.

A well-known result (e.g. White, 1980) regarding the OLS estimator of $\beta$ in the canonical linearregression model, $\mathbf{y} = \mathbf{X\beta} + \mathbf{u}$, states that as the sample size$\mathrm{N}$increases$\sqrt{N}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta})$converges in distribution to$\mathrm{N}(\mathbf{0}, \Omega)$, where

$$\Omega = \; [\mathrm{E}(\mathbf{X}' \;\; \mathbf{X})]^{-1} \, \mathrm{Var}(\mathbf{X}' \;\; \mathbf{u})[\mathrm{E}(\mathbf{X}' \;\; \mathbf{X})]^{-1}. \tag{5}$$

The reason for the term sandwich variance matrix is obvious from (5). The Huber-Eicker-White estimator uses straightforward estimators for individual parts of the sandwich,which do not rely on the homoscedasticity assumption for their consistency. The Huber-Eicker-White estimator itself has several versions, differing in minor small-sample corrections, with thesimplest one being

$$\hat{\boldsymbol{\vartheta}}_{HCO} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\textstyle\sum_{i=1}^{N} \hat{u}_i^2 \, x_i' x_i)(\boldsymbol{X}'\boldsymbol{X})^{-1}, \tag{6}$$

where $x_i$ is the ith row of $\mathbf{X}$and $\hat{u}_i$ is the ith OLS residual. This version is referred to as

HC0 in most statistical packages (HC stands for heteroscedasticity corrected); see (Davidson and MacKinnon, 2004) for other variants.

Recall that the conventional estimator of the OLS variance matrix has the form

$\hat{\boldsymbol{V}} = \hat{\sigma}^2 \; (\mathbf{X}' \;\; \mathbf{X})^{-1}$,where $\hat{\sigma}^2$ is the residual standard error. Thus, the sandwich estimator(6) can be rewrittenas

$$\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{V}} M \hat{\boldsymbol{V}} \tag{7}$$

where$\widehat{\boldsymbol{V}}$ is the conventional estimator of variance, referred to as the bread of the sandwich,and $M$ is the sandwich's meat. This turns out to be the overall structure of most sandwichestimators.

The heteroscedasticity-consistent estimator does not correct for typical panel/multilevel dataproblem of intraclass correlation due to cluster-specific random errors, nor for the problem ofserial correlation of random errors in short panels. For panel and multilevel data, a cluster-robust varianceestimator (CRVE) is more appropriate. In order to turn the HC0 estimator into a cluster-robust one, we merely change the meat to

$$\boldsymbol{M} = \sum_{g=1}^{G} \boldsymbol{X}_g{}'\widehat{\boldsymbol{u}}_g\widehat{\boldsymbol{u}}_g{}'\boldsymbol{X}_g \tag{8}$$

where $G$ is the number of clusters, $\boldsymbol{X}_g$ is the usual matrix $\mathbf{X}$ with the selection of rows restrictedto individuals from cluster $g$ only, and $\hat{u}_g$is a vector of OLS residuals for individuals from cluster$g$. The square roots of the diagonal elements of the CRVE matrix are called cluster-robuststandard errors.

The sandwich variance matrices can also be computed for non-linear models estimated by maximum likelihood (such as logistic regression). Typically, the form (7) is again used, where$\widehat{\boldsymbol{V}}$is the usual Hessian-based variance matrix (, and the meat is obtained from the rows ofthe $\boldsymbol{X}$ matrix weighted by likelihood scores; for more details, see (Cameron and Trivedi, 2005).

## 1.2 Multi-way clustering and robust inference

In a multilevel dataset, multiple levels of hierarchy can be present. The natural question thenarises, which levels to cluster over. If the levels of hierarchy are nested within one another,e.g. regions → schools → classes → students, the general consensus is that one should clusterover the topmost level, e.g. regions (Cameron and Miller, 2015).

The levels of hierarchy do not have to be nested, however. In our empirical analyses, we studya gravity model of migration where observations are clustered by the source country and thedestination country – two grouping dimensions that intersect. The practical solution that issometimes used in empiric applications is to cluster by the intersection of these two groupings,i.e. to use the cluster-robust standard errors clustered by source-destination pairs. Cameronand Miller (2015) warn that this approach is inadequate since it imposes that observations areindependent within the same destination country but in different origin countries. It makes senseto assume that all observations that share e.g. a concrete source country may be correlated toa certain extent (even if the destination country varies).

The approach proposed by Cameron and Miller (2015) is to calculate a multi-way cluster-robustvariance matrix estimate. The method relies on asymptotics that are in the number of clusters ofthe dimension with the fewest number of clusters. In statistical packages, the two-way clusteringrobust variance estimators have not been implemented yet. However, the estimates can bemanually calculated from one-way cluster-robust estimates in the following manner:

$$\widehat{\boldsymbol{\vartheta}_{2w}} = \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{1}} + \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{2}} - \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{12}} \ , \tag{9}$$

where $\widehat{\boldsymbol{\vartheta}_{2w}}$ is the two-way clustering robust variance estimate;$\widehat{\boldsymbol{\vartheta}}_{\boldsymbol{1}}\widehat{\boldsymbol{\vartheta}}_{\boldsymbol{2}}$) is an estimate of thevariance matrix robust to one-way clustering by variable 1 (variable 2), in our case the source(destination) country;$\widehat{\boldsymbol{\vartheta}}_{\boldsymbol{12}}$ is an estimate of the variance matrix robust to one-way clustering byan intersection of variables 1 and 2 (all distinct source-destination pairs).

It needs to be noted that (9) is not guaranteed to be positive semi-definite. This can be thecase mostly when clustering is done over the same groups as the fixed effects. A fix recommendedby Cameron and Miller (2015) consists in obtaining an eigedecomposition of the matrix andconverting any negative eigenvalues to zero. The procedure is carried out in three steps:

1. the variance matrix $\widehat{\boldsymbol{\vartheta}_{2w}}$ is decomposed into the product of its eigenvectors and eigenvalues:

$$\widehat{\boldsymbol{\vartheta}_{2w}}[\boldsymbol{\beta}] = \boldsymbol{U}\boldsymbol{\Lambda}\mathbf{U}\,', \tag{10}$$

   where $\mathbf{U}$ contains the eigenvectors of $\widehat{\boldsymbol{\vartheta}_{2w}}$, and $\boldsymbol{\Lambda} = \mathrm{diag}[\lambda_1, \ldots, \lambda_d]$ contains the eigenvaluesof $\widehat{\boldsymbol{\vartheta}_{2w}}$ on its diagonal.

2. Next, $\boldsymbol{\Lambda}^{+}$is created as $\boldsymbol{\Lambda}^{+} = \mathrm{diag}[\lambda_1^{+}, \ldots, \lambda_d^{+}]$with $\lambda_j^{+} = \max(0, \lambda_j)$.

3. The final estimated variance matrix becomes

$$\widehat{\boldsymbol{\vartheta}_{2w}}^{+}[\boldsymbol{\beta}] = \boldsymbol{U}\boldsymbol{\Lambda}^{+}\mathbf{U}\,', \tag{11}$$

## 2 The empirical application in the migration analysis

The migration analysis of Polonyankina (2017) that uses the random effects model for panel data is extended bycorrections of clustered error terms using the introduced in the previous chapter.

The empirical part tests the validity of the gravity model for intra-European immigration since: European immigrants originally from other European country than they currently stay create a significant portion of foreigners living in EU countries.

### 2.1 Motivation

The gravity model (described e.g. in Anderson, 1979, Metulini, 2013) is in a contradiction with the push and pull factors model, the basic one for migration, in the expected impact of the economic development of a source country.

The push and pull factors model (described inBorjas, 2010) assumes negative impact of distance on immigration. Immigration is expected to increase with a gap in economic development between source and destination countries where: an economic development of a source country is expected to have a negative impact on immigration, and an economic development of destination country is expected to have a positive impact.The push and pull factors model expect that workers are more motivated to move the higher is a gap between economic developments of countries.

The gravity model gives other possibility to look at the intra-European immigration assuming that migration works on the same mechanism as trade of goods, expecting a higher share of workers between more developed countries. The gravity model supposes that size (in relative or absolute terms) has a positive effect on spatial interaction. It's modification for migration analysis assumes that the GDP (as a relative size measurement) of both the host and source country has a positive impact on migration.The gravity model expects that migration works on principles of trade exchanged, assuming to be driven by an economic growth of both, destination and source country, and is reduces by their distance.

One can argue the push and pull factors theory for international migration could not be valid for European intra-immigration. The European union is specific since countries are geographically close to each other, there are no significant differences in economic development and lifestyle, there are no military conflicts plus there is free movement of labor and individuals. Intra-European immigration can be driven by different mechanisms than immigration from third non-European countries since there are no significant differences in human rights, culture, economic development and lifestyle between country of origin and destination country, both members of the European union. Moreover, free movement of labor and human capital inside the European union reduce some pull factors and costs of immigration significantly. Given that the theory of the push and pull factors, assumed to be valid for international migration, could not describe migration impacts inside the European Union.

The goal of the empirical part is to estimate the dependence of migration, within eight European countries, on their GDP and the distances to each other in order to examine the validity of the gravity model mechanism.

The empirical application stress standard errors robust estimations since several clusters are expected in the data. The panel data model regressions report random effects and correlated random effects which are combined with robust standard errors, clustered standard errors by country of origin and two-level clustering by country of origin and destination country.

## 2.2 Data and methods

The main data source is the Eurostat web page. The GDP per capita source is the Czech Statistical Office database. Countries were chosen based on Eurostat data availability since the model requires information about migration flows for all countries. Regression analysis is based on panel data for the years 2011-2014 and countries: Belgium, Italy, Hungary, the Netherlands, Finland, Sweden, Norway and Switzerland.

The same regression model as in Polonyankina (2017) is used where $n = 8, k = 1$ and $t = 2011 - 2014$:

$$y_t = \alpha J + X_{it}\beta_o + X_{jt}\beta_d + D\gamma + \varepsilon_t, \quad (6)$$

where:

- $y$ is an $n^2 x\, 1$ vector of logarithms of dependent variables which defined migration between countries in year $t$,
- $J$ is an$n^2 x\, 1$ intercept vector,
- $X_{it}$is an$n^2 x\, 1$ vector of logarithms of GDP per capita of source countries in year$t$,
- $X_{jt}$ is an$n^2 x\, 1$ vector of logarithms of GDP per capita of host countries in year $t$,
- $D$ is an$n^2 x\, 1$ vector of logarithms of distances between host and source countries (in km),
- $\varepsilon \sim N(0, \sigma^2 I_n)$ ,
- $\alpha, \beta_d, \beta_o, \gamma$ are parameters to be estimated.

The model works with macroeconomic time series, given that variables of the model are tested for stationary applying the Harris-Tzavalis unit-root test. The unit-root test concludes that both GDP variables $X_{it}$ and $X_{jt}$ do not meet the stacionarity assumption. However, our panel data set is short, including 4 years and 8 countries (56 combinations of migration flows), given that:

(i) In general, for short panels non-stacionarity does not have essential impacts.

(ii) (Results of the unit-root test need to be taken with a discretion. Harris and Tzavalis (1999) shows on simulation experiments (table 2a, Harris and Tzavalis, 1999) that the test power is low for approximately the same short data set.

Finally, we decided to include two types of regression sets where $X_{it}$ and $X_{jt}$ are included once in first difference difference transformation and once originally, which is the first extension from Polonyankina (2017).

The panel data estimations are done in Stata using the random effects (RE) and correlated random effects models (CRE) to deal with a potential unobserved heterogenity following the methodology ofWooldridge (2010), which is the second extension of Polonyankina (2017).

Since distance is time-invariant and FE is not applicable. We deal with the limitation by inclusion of CRE, which allows us to include the time-constant variables and at the same time delivers the FE estimates on the time-varying covariates. CRE relaxes the RE assumption by assuming a specific form of dependence between the unobserved heterogenityand the explanatory variables. Given that RE could be more efficient for some cases and we include both, CRE and RE, in the presentation.

For both regression sets we calculate 3 types of standard errors (the third extension from Polonynakina, 2017):

1. The Huber-Eicker-White estimator where standard errors corrected for heteroscedasticity, the basic commonly used correction;
2. The clustered (Huber-White-sandwich) errors to correct errors for dependencies within a region;
3. It is reasonable to assume that all observations that share e.g. a concrete source country may be correlated to certain extent (even if the destination country varies). In particular, we use a variance estimator for the random-effects estimator that is robust to two-way clustering, i.e. clustering across (i) the source country and (ii) the destination country.

The corrected two-way clustering variance calculation is not incorporated in STATA and was programmed manually including correction for non-positive semi-definite variance matrix, applying the procedure described in section Multi-way clustering and robust inference.

## 2.3 Results

The results on table 1 present the estimation of random effects panel data model (RE) and correlated random effects models (CRE) with GDP in the first difference. Table 2 presents the estimation of random effects panel data model (RE) and correlated random effects models (CRE) with GDP without the first difference correction and with time dummies.

During our analysis we estimated also the first regression set (table 1) with time dummies, with no impact on the conclusions. The three types of standard errors are presented:

- *Robust SE* robust standard errors;
- *Clustered SE* clustered standard errors by a destination country;
- *Two-way SE* the two-way cluster-standard errors dealing with possible two-way clustering by a destination country and a source country.

The conservative p-values (*p-value*) are used for the final significance evaluation following the conservative approach of Angrist and Pischke (2008), where the p-value used for our result evaluation is based on the largest of the three calculated standard errors.

The estimated parameters show that distance has a negative and statistically significant impact on migration, which is in line with both the push and pull theory and the gravity model.

For GDP of a destination country a positive impact was estimated, not significant for models on table 1 and significant for models on table 2. The positive impact is expected by both the push and pull theory and the gravity model.

**Table 1: Regression results set 1**

*Source: Own calculation is Stata.*

| *explanatory variables* | *RE* | *CRE* |
|---|---|---|
| **GDP source country (log-diff.)** | -0.038 | -0.042 |
| **p-value** | 0.531 | 0.690 |
| (Robust SE) | **(0.061)** | (0.061) |
| (Clustered SE) | (0.057) | (0.011) |
| (Two-way SE) | (0.057) | **(0.105)** |
| **GDP destination country (log-diff.)** | 1.575 | 1.595 |
| **p-value** | 0.354 | 0.361 |
| (Robust SE) | (0.873) | (0.888) |

| | | |
|---|---|---|
| (Clustered SE) | (1.578) | (1.065) |
| (Two-way SE) | **(1.695)** | **(1.743)** |
| **Distance (log)** | -0.864*** | -0.861*** |
| **p-value** | 0.003 | 0.005 |
| (Robust SE) | **(0.293)** | **(0.304)** |
| (Clustered SE) | (0.173) | (0.275) |
| (Two-way SE) | (0.131) | (0.133) |
| Year dummies | no | no |

Footnotes – 9 point, ARIAL

**Table 2: Regression results set 2**

*Source: Own calculation is Stata.*

| *explanatory variables* | *RE* | *CRE* |
|---|---|---|
| **GDP source country (log)** | -0.091 | -0.067 |
| **p-value** | 0.195 | 0.516 |
| (Robust SE) | (0.060) | (0.056) |
| (Clustered SE) | (0.054) | (0.047) |
| (Two-way SE) | **(0.070)** | **(0.103)** |
| **GDP destination country (log)** | 1.042*** | 2.543*** |
| **p-value** | 0.001 | 0.002 |
| (Robust SE) | (0.190) | **(0.760)** |
| (Clustered SE) | (0.164) | (0.729) |
| (Two-way SE) | **(0.299)** | (0.710) |
| **Distance (log)** | -0.849*** | -0.870*** |
| **p-value** | 0.002 | 0.002 |
| (Robust SE) | **(0.268)** | **(0.286)** |
| (Clustered SE) | (0.241) | (0.262) |
| (Two-way SE) | (0.132) | (0.195) |
| Year dummies | yes | yes |

Footnotes – 9 point, ARIAL

The impact of the GDP of a source country on migration is estimated negative, but also not statistically significant. The gravity model assumptions do not fit the migration analysis results since the gravity model expects a positive and statistical significant impact. The analysis did not find an evidence of non-rejection of the hypothesis about the gravity model validation for intra-European immigration.

The results tend to support the push and pull factors theory regarding the negative impact estimated, however estimates are not significant. This can be caused by the choice of the European countries for the analysis,where the countries with stable and similarly developed economies are included. For these countries the migration push factors might not as strong as in the case of mostly economically weaker countries outside of the Europe.

It should be noted that the practical application discovered difficulties to meet data requirements the gravity model testing: Information about immigrant nationality is necessary in a data set, which is not listed for all countries and years on the Eurostat web page.

## Conclusion

The paper summarised the methods usually used in order to correct estimations of standard errors where presence of clustering in the data set leads to heteroscedasticity and unrobust estimations. The need for robust statistical inference is well-documented even in the elementary case of a regression with a randomly sampled cross section data where the usual ordinary least square standard errors are generally biased under the presence of heteroskedasticity. We introduce several basic sandwich variance estimators and a multi-way cluster-robust variance matrix estimate.

In empirical application the methods are combined with panel data random effects and correlated random effects models. The validity of the gravity model for the intra-European immigration wasanalyzedfor eight European countries between 2011 and 2014.

The expected negative impact of distance on migration was estimated to be significant in all models.As mentioned before, the gravity model assumes a positive effect of the GDP of the source country. However, the push and pull theory assumes the opposite effect, i.e., emigration is higher in countries with lower GDP since individuals expect that immigration to economically stronger countries will improve their economic situation. The analysis did not find any evidence of a positive impact of the GDP of the source country, as is expected by the gravity model, but rather, a negative, but not significant impact was estimated.The empirical analysis did not confirm the validity of the gravity model. The estimated parameters appear to corroborate the validity of the push and pull factors theory.

The empirical analysis suffered from the lack of data since information about labor force interactions between the countries is needed for the gravity model.

## Acknowledgements

The work on this paper was supported byIGS F4/60/2018 of the University of Economics, Prague.

## Reference

Anderson, J. A. (1979). A Theoretical Foundation for the Gravity Model, The American economic Review, Vol. 69, p. 106-116.

Angrist, J. D. and Pischke Js (2008) Mostly Harmless Econometrics: An Empiricist's Companion, Princeton: Princeton University Press.

Borjas G. J. (2010). Labour Economics, Boston et al.: McGraw-Hill, ISBN: 9780073511368.

Cameron, A. C. andMiller D. G. (2015) A Practitioneres Guide to Cluster-Robust Inference. Journal ofHuman Resources*, 50(2): 317-372.

Cameron, A. C.,Miller D. G. and J. B.Gelbach (2011) Robust Inference With Multiway Clustering. Journal of Business & Economic Statistics, 29(2): 238-249.

Cameron A. C. and P. K. Trivedi (2005) MICROECONOMETRICS: Methods and Applications, Cambridge University Press. New York May 2005.

Cameron A. C. and P. K. Trivedi (2009) Microeconometrics Using Stata, Stata Press books.

Davidson R. and MacKinnon J. G. (2004)Econometric Theory and Methods. New York: Oxford University Press.

Eicker F. (1967) Limit theorems for regression with unequal and dependent errors. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley, CA: University of California Press, 1: 59-82.

Froot KA (1989) Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *Journal of Financial and Quantitative Analysis,* 24: 333?355.

Harris R. D. F. and Tzavalis E. (1999) Inference for unit roots in dynamic panels where the time dimension is fixed. Journal of Econometrics, 91: 201-226.

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, CA: University of California Press, 1: 221?233.

Metulini R. (2013). Spatial gravity models for international trade: a panel analysis among OECD countries, European Regional Science Association conference paper, ersa13p522.

Polonyankina T. (2017) Does the Gravity Model Work for the Modelling of Migration Between European Countries from 2011 to 2014? *Proceedings of the 7th Economics & Finance Conference, Tel Aviv.* Praha: International Institute of Social and Economic Sciences (IISES), 2017: 287-295.

White H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48: 817?830.

Williams R. L. (2000) A note on robust variance estimation for cluster-correlated data. Biometrics, 56: 645?646.

Wooldridge J. M. (2010) Econometric Analysis Of Cross Section and Panel Data. MA: MIT Press.