Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague

On Estimation of Gravity Equation: A Cluster Analysis

Božena Bobková

IES Working Paper: 37/2014

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Bobková B. (2014). "On Estimation of Gravity Equation: A Cluster Analysis" IES Working Paper 37/2014. IES FSV. Charles University.

On Estimation of Gravity Equation: A Cluster Analysis

Božena Bobková[a]

[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Smetanovo nábreží 6, 111 01 Prague 1, Czech Republic
Email: bozena.bobkova@seznam.cz

December 2014

Abstract:

This article questions the slope homogeneity in a gravity equation and proposes a partially heterogeneous framework for its estimation using panel data. We suggest to employ K-mean clustering to group countries according to the gravity equation variables. Further, the gravity model is estimated on these created homogeneous groups. We apply this procedure on analysis of German trade data and confirm the slope heterogeneity in the model. When we estimate the model on each cluster separately, the estimated coefficients and their standard errors vary sufficiently. Moreover, we show that the pooled estimation technique severely under- or overestimate the effect of given variables.

Keywords: gravity model, K-mean clustering, Germany, slope heterogeneity
JEL: C23, C45, F10, F14

Acknowledgements: This research was supported by the Grant Agency of the Czech Republic, grant no. P402/12/0982.

I. Introduction

Since Tinbergen (1962), considerable amount of literature has been published on gravity equation and its application for explanation the trade flows between countries. The common approach to estimate the model resides in employing panel data (cf., Camarero et al. (2014) and Yang and Martinez-Zarzoso (2014) for recent surveys) to avoid the estimation bias caused by the heterogeneity among countries. However, the research to date has tended to allow only for intercept heterogeneity. Far too little attention has been paid wheatear this treatment of heterogeneity is sufficient. This study is an attempt to question the slope heterogeneity in estimated gravity equation as well.

Several studies has revealed the presence of parameter heterogeneity in economy. For instance, Hineline (2008) rejects the hypothesis of homogeneous growth model coefficients. He performs sufficient number of regressions on diverse sub-samples of countries and shows that the coefficients and their significance differ substantially. Clark et al. (2005) estimate he relationship between income and reported well-being using latent class techniques using panel data from twelve European countries. They strongly reject the hypothesis that individuals transform income into well-being in the same way.

In many recent studies, the authors address slope heterogeneity by understanding panel data from the time series perspective. For instance, Pesaran (1995) and Pesaran et al. (1999) propose mean group (MG) and pooled mean group (PMG) estimators which aggregate (MG) or pool and aggregate (PMG) the coefficients from separate time series estimates. Further, Maddala et al. (1994) and Maddala et al. (1997) suggest Bayesian shrinkage estimators that shrink the individual heterogeneous estimates towards the pooled homogeneous estimates. The main limitation of these methods is that they rely on the assumption of large time dimension and number of crossectional units. Pesaran et al. (1999) demonstrate that for short time period the mean group estimator is likely to be severely biased.

Sarafidis and Weber (2009) and Sarafidis and Weber (2014), Chang-Ching and Ng (2012) and Kapetanios (2006) propose different approach. They argue for partial pooling by estimating the model on heterogeneous clusters. They cluster the individuals according to information criteria or residual sum of squares from the estimated model. However, in our opinion, these type of clustering rely overly on the estimated model and its specification. In the presence of some misspecification errors of the estimated model, the clusters would be misspecified as well.

This study offer a simpler and more data-based clustering-estimation strategy. We propose to split the countries according to K-means clustering based on time means of explanatory variables of gravity equation. The time means of variables are applied to employ the main information in time for given countries. Further, the proper number of clusters is then set according to Calinski and Harabasz (1974) pseudo-F index. This method models the heterogeneity among countries only based on the data and not on the model. We applied this strategy on the dataset for German export. According to our methodology, we grouped the countries into nine clusters and estimated the gravity equation on them. We discovered the slope coefficients to differ seriously for each groups of countries.

The remainder of the paper is organized as follows. Section II introduces the clustering and estimation methodology. Section III specifies data. Section IV describes the suggested groups and estimation results. Section V is the conclusion.

II. Methodology

I. Clustering

To group the countries into clusters, we perform k-means clustering. K-means clustering, originally introduced by MacQueen (1967) , is a popular unsupervised learning algorithm which aims to partition n observations into k clusters. More specifically, given a set of observed vectors $\{X_1, X_2, .., X_n\}$, the algorithm splits n observations into k sets $C = \{C_1, C_2, .., C_k\}$ to minimize the square error objective function:

$$\sum_{j=1}^{k} \sum_{i=1}^{n} (d_i^{(j)})^2, \tag{1}$$

where $d_i^{(j)}$ is a chosen distance measure between a data point and the mean of cluster C_j. The distance measure is selected by the nature of variables used for clustering. Our analyzed data consist of both continuous and discrete variables, so we apply Gower's distance measure in accordance with Gower (1971).

K-means clustering procedure requires to fix the number of groups in which the individuals supposed to be divided. The clustering literature provides several stopping rules to establish the optimal number of clusters. Milligan and Cooper (1985) evaluate a large number of these rules and find out Calinski and Harabasz pseudo-F index to be one of the best. Calinski and Harabasz (1974) propose pseudo-F index for g groups and N observations as:

$$CH = \frac{\frac{tr(B)}{g-1}}{\frac{tr(W)}{N-g}}, \tag{2}$$

where B is the between-cluster sum of squares and cross-products matrix, and W stands for the within-cluster sum of squares and cross-products matrix.

In addition, we apply the clustering procedure to split our dataset consisting of 177 German trading partners. More specifically, we run several clusterings to group the countries into 2, 3,...,20 clusters according to the explanatory variables[1] of gravity model and then apply the CH stopping rule to establish proper number of groups. Further, we decide to employ the mean values of variables for given countries to group them according to their average (typical) behavior and characteristics in time. Moreover, we aim to overcome the problem of countries switching from one to other cluster in time.

II. Gravity equation estimation and variables

The traditional approach of estimation gravity equation is based on multi-country models analysing huge trade panel data sets. However, these data sets also incorporate lots of noise information. We argue for a different framework; we target our analysis on the one-way trade flow of a home country to its trading partners. This enables us to study the relationships in the gravity model in

[1] the variables are described in detail in Appendix A

more specific way. In accordance with Anderson and Van Wincoop (2001), the export function of our "home country" i to its trading partners j is defined as:

$$X_{ijt} = \beta_0 Y_{it}^{\beta_1} Y_{jt}^{\beta_2} T_{ijt}^{\beta_3} e^{\theta_j} \epsilon_{ijt}, \tag{3}$$

where X_{ijt} stands for home country export, Y_{it} and Y_{jt} are gross domestic products of given country and T_{ijt} is the vector of trade cost variables. Particularly, we define in this vector the traditional gravity variables like distance from the home country to its trading partner and population of the partner country. Moreover, we also include exchange rate deviation index (ERDI), currency union dummy variable, trade barriers and institutional variables in partner country. As institutional variables, we assume the government effectiveness index, quality of education index and the average institutional quality index based on Heritage foundation indexes. We also cover the dummy variables for years 2008, 2009 and 2010 as a proxy for the recent economic crisis. Last but not least, θ_j stands for the multilateral resistance term and ϵ_{ijt} are the disturbances. All of the variables are precisely defined in Appendix A.

We estimate the log-linearized version of equation 3. According to Egger (2004) and Egger and Pfaffermayr (2004) , we employ the Hausman-Taylor type of estimator. Firstly, we aim to control for the likely endogeneity of the GDP's and institutional variables. We assume that export influences domestic GDP and the trading partner's GDP. Moreover, we also presume that if Germany exports to a country it would probably also affect the trading partner's institutional environment. Secondly, HT estimator allows for consistent estimation of time invariant variables like distance in our case. Furthermore, Brun et al. (2005) argue for HT for gravity equation estimation to explain the gravity distance puzzle. To control for the non-randomness of our observation in each cluster, we provide bootstrapped standard errors with 1000 replications.

III. Data

We analyse a strongly balanced panel data set of Germany and its 177 trading partners in time period 1995-2011. The data were collected from the databases of Eurostat, the World Bank, IMF, European Commission, Geobytes, CEPII, Heritage Foundation, Mannheim enterprise database, WTO, database on Tenders of Public Procurement in the EU and national statistical offices.

IV. Results

I. Results of clustering

After clustering countries into two to twenty groups, the Calinski Habanasz pseudo - F index suggests 9 groups as optimal[2]. The optimal grouping schema is summarized in Tables 1 and 2. Moreover, Table 3 presents the summary statistics of the gravity variables and also of the German export for each group.

We can find that the first cluster comprises mainly from the non-European countries which are the third most distant from Germany on average. These countries dispose of the second lowest GDP and rather above average population. These countries also impose strong tariff policy. According to their above-average ERDI, German Euro is significantly appreciated relative to their

[2]For this grouping the index reached its highest value 256.96

currencies. Moreover, these countries suffer from poorer institutional environment.

The second cluster consists of five strong economies including the USA. They are the most distant from Germany and second most populated. These countries feature the best institutional background and lowest ERDI. On the other hand, they also impose the highest tariffs for German imports. Further, the third group links the non-European and some East-European countries together. None of these countries is a member of the European union. This group pays rather for a middle group in ranking of the GDP, population, distance and ERDI. Although their tariff policy could be considered as third most benevolent, the average absolute value of their tariff index is still high. The institutional environment of this group does not evince common trend in all variables. While the education could be classified as satisfactory, the common institutional environment based on Heritage Foundation indexes is rather average. Moreover, these countries suffer from quite low government effectiveness on average.

China and India were merged into one separate cluster. These two countries could be understood as an outlier group and we do not estimate the gravity equation for them. These countries are the most populated with highest GDP. They are situated a quite long way from Germany and German Euro is strongly appreciated relatively to their national currencies. They impose tough tariff policy. Comparing to other groups, they are also burdened by a poorer institutional environment when the worst seems to be the quality of education.

Further, the fifth cluster comprises of the European countries. Furthermore, this group includes the majority of the "old European union" states like France or Belgium and also embodies the European microstates like San Marino, Andorra and Liechtenstein. Apart from the majority mentioned microstates, all of the merged countries are members of the Eurozone. San Marino pays for the currency union member as well and Andorra disposes of the issuing rights. These countries are the third richest in terms of the GDP and the second least populated on average. These countries are also on average closest to Germany. According to the nature of these countries, their tariff burden to Germany is rather negligible and ERDI is the second lowest. The quality of institutions pays for one of the highest.

The sixth cluster embodied mainly the African countries. These countries are very poor on average, their GDP is ranked as the lowest among other clusters with the third highest population. This group could be also considered as a median group in terms of distance from Germany. Further, there are imposed quiet high tariffs between these countries and Germany and also ERDI is the third highest among the clusters. Moreover, these countries suffer from inferior institutional environment. Namely, the quality of education is the lowest among clusters; and the government effectiveness and the overall institutional quality pay for the second lowest.

Last but not least, the seventh group consists mainly of former Soviet union members. Comparing to others, this cohort pays for the third poorest in terms of GDP with fourth lowest population. These countries are also situated as the third closest to Germany. However, we can identify possible trade barriers in highest ERDI and weakest government effectiveness and overall institutional index. Moreover, the there exists also significant tariffs for trade.

Almost every country of the eighth group lies in Europe. The majority of them are members of the European union accessing the union in the last two acceptation rounds. Moreover, some of the countries joined the Eurozone. These countries are also on average the second closest neighbours to Germany. Their GDP pays for the fourth highest among the clusters with the lowest population. The institutional background could be considered as very good.

Finally, the last cluster comprises of a combination of non-European countries when significant part of them is located in the Latin America. These countries are the second most distant from Germany with the median GDP and third lowest population. There are imposed second highest tariffs for trade. On the other hand, their institutional environment is more than satisfactory and also their ERDI pays for the third lowest among clusters.

In addition, we have obtained diverse groups based on "gravity characteristics". We assume that the behaviour of countries in each group is homogenous.

Table 1: *Results of k-mean clustering:*

cluster 1	cluster 2	cluster 3	cluster 4	cluster 5
BOLIVIA	AUSTRALIA	ALBANIA	CHINA	ANDORRA
CAMBODIA	CANADA	ALGERIA	INDIA	AUSTRIA
COLOMBIA	JAPAN	ARMENIA		BELGIUM
CROATIA	NEW ZEALAND	CAMEROON		DENMARK
CUBA	UNITED STATES	COTE D'IVOIRE		FINLAND
ECUADOR		DOMINICAN REP.		FRANCE
EL SALVADOR		EGYPT		GREECE
GUYANA		ETHIOPIA		IRELAND
INDONESIA		FIJI		ITALY
KENYA		GABON		LIECHTENSTEIN
LESOTHO		GEORGIA		LUXEMBOURG
MADAGASCAR		GHANA		NETHERLANDS
MALAWI		GUATEMALA		PORTUGAL
MALDIVES		GUINEA		SAN MARINO
MONGOLIA		HAITI		SPAIN
MOZAMBIQUE		HONDURAS		
NICARAGUA		JORDAN		
PAPUA N. G.		KAZAKHSTAN		
PARAGUAY		KYRGYZSTAN		
SRI LANKA		LEBANON		
SWAZILAND		MACEDONIA		
TANZANIA		MOLDOVA		
UGANDA		MOROCCO		
VENEZUELA		ROMANIA		
VIETNAM		RUSSIA		
ZAMBIA		SAUDI ARABIA		
ZIMBABWE		SERBIA		
		TUNISIA		
		UKRAINE		

II. Results of gravity equation estimation

As described in Section II, we estimated gravity equation for the full sample of countries and then for each group obtained trough the clustering procedure. The results are summarized in Tables 4and 5[3].

The theory of gravity equation suggests a trade flow between countries to be unit elastic in home and partner's GDP and in distance. We can see that the estimated partner's GDP elasticity on pooled sample is lower, specifically 0.76. Moreover, partner's GDP explains German export significantly. When we allow for partial slope heterogeneity, the estimated elasticity and its significance vary substantially among clusters. The largest, most significant and also to the theoretical value closest effect of partner's GDP was revealed in the fifth, seventh and eighth cluster, whereas the lowest and least significant estimates are found in the first and third group. The difference between their estimated coefficients is more than a half. Moreover, only the coefficient of partner's GDP in sixth cluster could be considered as similar to the full sample estimates. From these results is apparent that the pooled estimate could be highly misleading. When we were rely on full sample estimates, we would especially significantly underestimate the impact of partner's GDP in the majority of European countries and on the other hand, we would substantially overestimate the effect in many non-European and poor countries.

Looking at the export elasticity in home GDP, the problem of pooling becomes more severe. According to full sample estimates, one per cent change in domestic GDP significantly increases German export to given country by 0.5 per cent. Conversely to this result, we do not identify effect of domestic GDP as even significant in first, second, fifth and seventh cluster. The elasticity estimates of third and ninth group even substantially exceed the unit elasticity assumption whereas the elasticity estimates of eight group fulfils them exactly. Moreover, we revealed even slightly significant negative effect of German GDP on its export to countries in sixth cluster. In addition, the pooled estimation would lead us to a wrong conclusions for all countries.

Further, the pooled estimate of distance seems to be the most misleading from all coefficients. We significantly estimate the trade elasticity in distance to the value -1 (in accordance with the theory). However, the separate cluster estimates suggest that distance does not influence German trade significantly in any group, except the third cluster. Moreover, the estimated effect for these mainly non-European and poor countries is almost twice as big. This problem has been already addressed by Brun et all. 2005. This study analyzes whether the distance died in the today's globalized world. They find out that poor countries have been marginalized by the current wave of globalization while rich countries have benefited from a death of distance. Our findings partially confirm their results when we conclude that there exists only a smaller specific group of countries marginalized by globalization.

According to full sample estimates, the impact of exchange rate deviation could be classify as significant when one per cent increase in the index would lower German export by about -0.5 per cent. Similar coefficient were estimated in second and sixth cluster. However, the deviation seems to not have a significant effect on export for the majority of European countries whereas the effect for the countries from first and seventh cluster is substantially larger than suggested by pooled estimate.

[3]Standard errors are the bootstrapped standard errors with 1000 replications; p<0.01 ***, p<0.05 ** and p<0.1 *.

Population, tariff policy or being part of the currency union of partner's country do not influence German export in full sample estimates nor in particular cluster estimation. The situation is however different with the institutional variables. The pooled estimation supports only the overall institutional index as significant for explaining German export. However, allowing for heterogeneity, we find out that this variable is significant only in poor countries and its effect could be three times larger than in pooled estimates. Furthermore, quality of education significantly improves trade in eight cluster whereas government effectiveness increases significantly German export to third cluster.

Finally, we estimated a negative effect of recent economic crises in years 2009 and 2010 on the full sample. We confirm these findings in separate cluster estimates for the developed and rich countries (in second, fifth and eight cluster). German export to countries in sixth cluster was also extensively hit by the crisis with even three times higher effect than in full sample estimates.

V. Conclusion

The purpose of the present research was to address the estimation of gravity equation of international trade using panel data. We elaborated the idea of imposing higher level of heterogeneity in the estimation process. More specifically, we did not rely only on varying intercepts in terms of fixed effects, but we questioned the slope homogeneity as well. We were motivated by the fact that the countries around the world face too diverse economical, political and even natural circumstances and historical development. In our opinion, it seems rather doubtful that these varied countries would share the common patterns in their trading relationships to other countries.

Instead of easing the homogeneity assumption to full coefficient heterogeneity, we allowed the slopes to vary only among different groups of countries when within the groups the coefficients are homogenous. The members of these groups supposed to evince similar behaviour in terms of variables employed for the gravity equation estimation and by that also some homogeneity. Estimating gravity equation on these proposed homogenous groups, we are able to take the advantage of pooling. We proposed to find similar countries by clustering procedure to create the groups based purely on data. Finally, gravity equation is then estimated for each cluster separately and the results are compared with the whole sample estimates. We applied our strategy to estimate German export. In the first step, we merged the countries into nine groups according to K-mean clustering. In the second step, we estimated the gravity model employing the Hausman-Taylor estimation technique to control for possible endogen and time-invariant variables.

The results confirmed our doubts about slope homogeneity in gravity equation. We demonstrated that the pooled estimates are for the majority of countries highly misleading. More specifically, we showed that the estimated coefficients and significance of given variables vary substantially among clusters and also differ from the pooled estimates. We identified that only the partner's GDP has significant effect on German export in both pooled and all heterogeneous estimates. However, the pooled estimates severely underestimate the impact of GDP in majority of European countries, whereas the effect in many non-European and poor countries is overestimated. Further, our findings also partially support the evidence of Brun et al. (2005) about the death of distance. We determined only one specific group of rather poorer countries whose distance from Germany influences significantly German export and its effect was estimated almost twice as large as the pooled estimate. Last but not least, the estimated impact of quality of institutional variables was found significant mainly in poorer non-European countries. Compared to the full sample estimate, some countries evince this effect almost three times larger.

In addition, we addressed the problem of heterogeneity in gravity equation estimation. According to our results, we pointed out how important this strategy is and how relying on the pooled estimates could be misleading.

References

Anderson, J. E. and Van Wincoop, E. (2001). Gravity with gravitas: A solution to the border puzzle. Working Paper 8079, National Bureau of Economic Research.

Brun, J. F., Carrere, C., Guillaumont, P., and de Melo, J. (2005). Has Distance Died? Evidence from a Panel Gravity Model. *World Bank Economic Review*, 19(1):99–120.

Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3.

Camarero, M., Gomez, E., and Tamarit, C. (2014). Is the euro effect on trade so small after all? new evidence using gravity equations with panel cointegration techniques. *Economics Letters*, 124(1):140–142.

Chang-Ching, L. and Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. *Journal of Econometric Methods*, 1(1).

Clark, A., Etile, F., Postel-Vinay, F., Senik, C., and Van der Straeten, K. (2005). Heterogeneity in Reported Well-Being: Evidence from Twelve European Countries. *Economic Journal*, 115(502):C118–C132.

Egger, P. (2004). On the Problem of Endogenous Unobserved Effects in the Estimation of Gravity Models. *Journal of Economic Integration*, 19:182–191.

Egger, P. and Pfaffermayr, M. (2004). Distance, trade and FDI: a Hausman-Taylor SUR approach. *Journal of Applied Econometrics*, 19(2):227–246.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27:857–874.

Hineline, D. R. (2008). Parameter heterogeneity in growth regressions. *Economics Letters*, 101(2):126–129.

Kapetanios, G. (2006). Cluster analysis of panel data sets using non-standard optimisation of information criteria. *Journal of Economic Dynamics and Control*, 30(8).

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297. University of California Press, Berkeley, Calif.

Maddala, G. S., Srivastava, V. K., and Li, H. (1994). Shrinkage estimators for the estimation of short-run and long-run parameters from panel data models. *Working paper Ohio State University, Ohio*.

Maddala, G. S., Trost, R. P., Li, H., and Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics*, 15:90–100.

Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a dataset. *Psychmetrika*, 50:159–179.

Pesaran, M. H., Shin, Y., and Smith, R. P. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association*, 94(446):621–634.

Pesaran, M. H.and Smith, R. P. (1995). Estimating Long-run Relationships from Dynamic Heterogeneous Panels. *Journal of Econometrics*, 68(1):79–113.

Sarafidis, V. and Weber, N. (2009). To pool or not to pool: a partially heterogeneous framework. *MPRA paper:36155*.

Sarafidis, V. and Weber, N. (2014). A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics*.

Tinbergen, J. (1962). An analysis of world trade flows. *Shaping the World Economy: Suggestions for an International Economic Policy, New York: Twentieth Century Fund*.

Yang, S. and Martinez-Zarzoso, I. (2014). A panel data analysis of trade creation and trade diversion effects: The case of asean-china free trade area. *China Economic Review*, 29(0):138 – 151.

Table 2: *Results of k-mean clustering - cont.*

cluster 6	cluster 7	cluster 8	cluster 9
AFGHANISTAN	AZERBAIJAN	BULGARIA	ANTIGUA AND BARBUDA
ANGOLA	BELARUS	CYPRUS	ARGENTINA
BANGLADESH	BOSNIA AND HER.	CZECH REPUBLIC	BAHAMAS
BENIN	ERITREA	ESTONIA	BAHRAIN
BURKINA FASO	IRAN	FAROE IS.	BARBADOS
BURUNDI	IRAQ	HUNGARY	BELIZE
CHAD	LIBYA	ICELAND	BERMUDA
CONGO	NORTH KOREA	ISRAEL	BOTSWANA
CONGO, DR	TAJIKISTAN	LATVIA	BRAZIL
CTL AFRICAN R.	TURKMENISTAN	LITHUANIA	BRUNEI
EQ. GUINEA	UZBEKISTAN	MALTA	CAYMAN IS.
LAOS		NORWAY	CHILE
LIBERIA		POLAND	COSTA RICA
MALI		SLOVAKIA	FR. POLYNESIA
MAURITANIA		SLOVENIA	GREENLAND
MYANMAR		SWEDEN	GRENADA
NEPAL		SWITZERLAND	HONG KONG
NIGER		TURKEY	JAMAICA
NIGERIA		UNITED KINGDOM	KUWAIT
PAKISTAN			MACAO
RWANDA			MALAYSIA
SENEGAL			MARSHALL IS.
SIERRA LEONE			MAURITIUS
SUDAN			MEXICO
SYRIA			NAMIBIA
TOGO			NL. ANTILLES
YEMEN			OMAN
			PANAMA
			PERU
			PHILIPPINES
			QATAR
			SEYCHELLES
			SINGAPORE
			SOUTH AFRICA
			SOUTH KOREA
			ST VINCENT
			TAIWAN
			THAILAND
			TRINIDAD AND TOBAGO
			UNITED ARAB EMIRATES
			URUGUAY

Table 3: *Statistics in clusters*

		x_{ijt}	y_{jt}	d_{ij}	l_{jt}	c_{ijt}	t_{ijt}	g_{jt}	er_{ijt}	edu_{jt}	$inst_{jt}$
c 1	min	0.6	514.3	853.2	0.2	1.0	5.0	2.9	1.0	13.2	18.5
	mean	257.0	60760.7	8443.1	23.5	1.0	7.5	36.1	2.5	50.6	54.7
	max	2920.2	808956.9	13779.2	241.0	1.0	8.0	81.0	5.9	74.9	76.3
c 2	min	384.0	46485.4	6541.7	3.7	1.0	9.0	81.5	0.5	80.6	64.3
	mean	16251.6	2818143.0	11475.6	94.6	1.0	9.0	92.4	1.0	92.7	76.1
	max	76153.1	11700000.0	18219.9	311.9	1.0	9.0	98.1	1.5	100.0	81.7
c 3	min	2.4	1751.4	1159.3	0.8	1.0	1.0	1.0	0.6	31.0	33.0
	mean	1250.4	105847.7	4379.0	20.8	1.0	6.8	36.1	2.0	73.9	56.0
	max	34378.3	1716662.0	16157.9	148.3	1.0	8.0	73.7	5.2	92.4	71.2
c 4	min	1847.3	819563.1	6565.8	937.0	1.0	8.0	43.6	1.7	34.8	45.1
	mean	13247.6	3016466.0	7298.7	1182.8	1.0	8.0	54.4	2.6	42.0	51.6
	max	64712.2	8127605.0	8031.7	1347.4	1.0	8.0	59.7	3.5	52.0	56.4
c 5	min	9.4	821.9	377.7	0.0	1.0	1.0	62.1	0.6	15.2	57.4
	mean	20632.1	356217.4	985.5	15.5	2.4	1.3	88.8	1.4	74.4	68.9
	max	101444.3	1782411.0	2021.6	63.1	3.0	3.0	100.0	5.8	96.3	82.8
c 6	min	0.6	903.0	2842.6	0.4	1.0	6.0	0.0	1.2	11.0	23.7
	mean	127.5	34743.0	5506.6	28.5	1.0	7.4	18.0	2.6	33.9	48.2
	max	1273.0	351085.8	8725.1	175.3	1.0	8.0	58.0	5.4	53.5	65.1
c 7	min	6.8	1865.2	1020.2	3.4	1.0	5.0	0.5	1.0	39.1	1.1
	mean	542.1	76859.8	3756.6	16.9	1.1	7.6	14.2	2.7	65.3	37.0
	max	4360.8	719174.7	8198.4	75.2	2.0	9.0	41.7	5.9	82.7	57.3
c 8	min	1.9	724.1	483.4	0.0	1.0	1.0	43.8	0.6	33.0	45.7
	mean	9466.1	197814.6	1326.7	12.4	1.4	2.8	81.3	1.8	76.1	66.9
	max	69658.7	1776838.0	2972.2	74.0	3.0	6.0	100.0	4.5	99.3	81.3
c 9	min	0.5	81.4	3702.1	0.0	1.0	7.0	3.8	0.7	26.5	35.7
	mean	1235.6	145882.5	8604.8	16.5	1.0	8.1	70.3	1.7	68.0	68.8
	max	11654.2	1649262.0	16305.0	194.9	3.0	9.0	100.0	3.6	93.4	90.5

Table 4: *Estimation of gravity equation in clusters*

	Full		s.e.	**c 1**		s.e.	**c 2**		s.e.	**c 3**		s.e.	**c 5**		s.e.
ly_{jt}	0.763	***	0.11	0.409	*	0.23	0.858	**	0.42	0.426	*	0.22	1.064	***	0.30
ly_{it}	0.532	**	0.21	0.428		0.57	0.375		0.71	1.381	**	0.57	0.460		0.40
ld_{ij}	-1.081	***	0.09	-1.298		1.97	2.246		8.23	-1.812	***	0.54	-0.289		0.74
ler_{ijt}	-0.462	***	0.07	-0.777	***	0.21	-0.554	***	0.21	-0.294	*	0.16	0.075		0.10
ll_{jt}	0.150		0.10	0.481		0.33	0.809		0.61	-0.662		0.48	-0.522		0.61
t_{ijt}	0.005		0.02	-0.037		0.14				0.045		0.04			
c_{ijt}	-0.067		0.05										0.017		0.04
$inst_{jt}$	0.018	***	0.00	0.018	*	0.01	0.007		0.01	0.003		0.01	0.000		0.01
edu_{jt}	0.003		0.01	-0.004		0.02	-0.003		0.04	0.011		0.02	0.012		0.01
g_{jt}	0.000		0.00	0.001		0.00	0.011		0.01	0.009	***	0.00	-0.004		0.01
y2008	0.013		0.03	0.116	*	0.07	-0.046	*	0.03	0.045		0.04	0.006		0.03
y2009	-0.111	***	0.03	-0.014		0.08	-0.166	**	0.07	-0.006		0.08	-0.117	***	0.04
y2010	-0.054	*	0.03	-0.039		0.08	-0.084	*	0.05	-0.015		0.08	-0.052		0.04
const.	-2.303		2.45	4.540		19.68	-33.667		76.17	-4.113		7.55	-5.735		4.45
rho	0.840			0.897			0.996			0.986			0.982		
nobs	2824			435			85			490			244		
$P > \chi^2$	0			0			0			0			0		

Table 5: *Estimation of the gravity equation in clusters - cont.*

	c 6		s.e.	**c 7**		s.e.	**c 8**		s.e.	**c 9**		s.e.
ly_{jt}	0.797	**	0.33	0.985	***	0.18	1.018	***	0.19	0.611	*	0.35
ly_{it}	-1.758	*	1.02	-1.732		1.16	0.998	**	0.45	1.335	**	0.52
ld_{ij}	-3.276		2.57	-0.867		1.62	-0.863		0.58	-0.530		1.94
ler_{ijt}	-0.529	***	0.21	-0.775	***	0.28	-0.037		0.15	-0.396	**	0.18
ll_{jt}	0.072		0.80	0.260		1.46	-0.050		0.51	0.425		0.48
t_{ijt}	-0.102		0.11	-0.001		0.21	-0.044		0.03	0.245		0.36
c_{ijt}							-0.028		0.08			
$inst_{jt}$	0.017	**	0.01	0.032	***	0.01	0.006		0.01	0.035	***	0.01
edu_{jt}	0.085	***	0.03	0.049		0.03	-0.022	**	0.01	-0.019		0.02
g_{jt}	-0.002		0.00	0.001		0.01	0.006		0.00	-0.017		0.01
y2008	0.060		0.07	-0.206	*	0.13	-0.025		0.04	-0.025		0.10
y2009	-0.321	**	0.14	-0.236		0.17	-0.177	***	0.06	-0.115		0.09
y2010	-0.225	**	0.10	0.014		0.09	-0.117	***	0.04	0.019		0.09
const.	47.429		29.57	22.884		20.34	-10.714		6.26	-17.635		19.88
rho	0.906			0.858			0.912			0.890		
nobs	340			175			322			699		
$P > \chi^2$	0			0			0			0		

Appendix A: List of variables

	Variable	Description	Source
x_{ijt}	German export	mil. EUR	Eurostat
y_{jt}	Partner's GDP at PPS	mil. EUR	Eurostat and IMF
y_{it}	German GDP at PPS	mil. EUR	Eurostat
d_{ij}	Distance	km	CEPII
er_{ijt}	ERDI		IMF, World Bank andown estimate
l_{jt}	Partner's Population	mil. of inhabitans	IMF
t_{ijt}	Tariff index	dummy 1 - 9	WTO, UNCTAD and own estimate
c_{ijt}	Currency union	Euro, dummy 1 - 3	own estimate
$inst_{jt}$	Overall institutional index	Average of all indexes, per cent	Heritage foundation
edu_{jt}	Quality of education index	per cent	United Nations (HD Reports)
g_{jt}	Government effectivness index	per cent	World bank
y2008	Dummy of year 2008	0 -1	own estimate
y2009	Dummy of year 2009	0-1	own estimate
y2010	Dummy of year 2010	0 -1	own estimate

IES Working Paper Series

2014

1. Diana Zigraiova, Petr Jakubik: *Systemic Event Prediction by Early Warning System*
2. Pavla Blahova, Karel Janda, Ladislav Kristoufek: *The Perspectives for Genetically Modified Cellulosic Ethanol in the Czech Republic*
3. Oxana Babecká Kucharčuková: *The impact of exchange rate volatility on trade: Evidence for the Czech Republic*
4. Oliver Polyák: *The Impact of Euro Adoption on Export Performance: Comparison of the Czech Republic and Slovakia*
5. Lucie Davidová, Vladimír Benáček: *Determinants of Austrian International Trade: Analysis Based on the Gravity Model*
6. Julie Chytilová, Václav Korbel: *Individual and Group Cheating Behavior: A Field Experiment with Adolescents*
7. Petr Janský: *Consumer Demand System Estimation and Value Added Tax Reforms in the Czech Republic*
8. Pavel Hait, Petr Janský: *Inflation Differentials among Czech Households*
9. Hana Džmuráňová, Petr Teplý: *Risk management of savings accounts*
10. Tomáš Fiala, Tomáš Havránek: *Ailing Mothers, Healthy Daughters? Contagion in the Central European Banking Sector*
11. Petra Andrlíková: *Is Barrier version of Merton model more realistic? Evidence from Europe*
12. Martin Pospíšil, Jiří Schwarz: *Bankruptcy, Investment, and Financial Constraints: Evidence from a Post-Transition Economy*
13. Matt Hunt, Goran Serdarević, Tom Ovington, Clive Kenny: *Evidence for a ladder of investment in Central and Eastern European countries*
14. Petra Andrlíková: *Bayesian default probability models*
15. Petra Buzková: *Has the Relationship Between Market and Model CDS Price Changed during the EMU Debt Crisis?*
16. Jakub Černý, Jiří Witzany: *Interest Rate Swap Credit Valuation Adjustment*
17. Martin Gregor: *Receiver's access fee for a single sender*
18. Pavel Ryska: *Deflation and Economic Growth in Long-Term Perspective*
19. Michal Paulus, Eva Michalíková, Vladimír Benáček: *German International Trade: Interpreting Export Flows According to the Gravity Model*
20. Michal Bauer, Nathan Fiala, Ian Levely: *Trusting Former Rebels: An Experimental Approach to Understanding Reintegration after Civil War*
21. Michal Skořepa: *Family Matters: Concurrent Capital Buffers in a Banking Group*
22. Martin Gregor: *Access Fees for Competing Lobbies*
23. Jozef Baruník, František Čech: *On the Modelling and Forecasting Multivariate Realized Volatility: Generalized Heterogeneous Autoregressivee (GHAR) Model*
24. Roman Horváth, Ayaz Zeynalov: *The Natural Resource Curse and Institutions in Post-Soviet Countries*
25. Petr Janský: *Impact of Excise Duties on Households in the Czech Republic*
26. Petr Jakubík, Thomas Reininger: *What Are the Key Determinants of Nonperforming Loans in CESEE?*

27. Jaromír Baxa, Tomáš Olešňaník: *Current Accounts in the Eurozone Countries: The Role of Euro, Fiscal Policies and Financial Developments*
28. Petr Polák: *The Productivity Paradox: A Meta-analysis*
29. Tomáš Havránek, Zuzana Iršová, Karel Janda, David Zilberman: *Selective Reporting and the Social Cost of Carbon*
30. Jozef Baruník, Tomáš Křehlík: *Coupling High-Frequency Data with Nonlinear Models in Multiple-Step-ahead Forecasting of Energy Markets' Volatility*
31. Lenka Šťastná, Jana Votápková: *Efficiency of Hospitals in the Czech Republic: Conditional Efficiency Approach*
32. Eva Michalíková, Michal Paulus: *Gravity Model Analysis: Robust Evidence from the Czech Republic and Corruption Matching*
33. Jozef Baruník, Lucie Kraicová: *Estimation of Long Memory in Volatility Using Wavelets*
34. Daniel Benčík: *Range-based Volatility Estimation and Forecasting*
35. Roman Horváth, Júlia Jonášová: *Central Banks' Voting Records, Financial Crisis and Future Monetary Policy*
36. Arshad Hayat: *FDI and Economic Growth: The Role of Natural Resources*
37. Božena Bobková: *On Estimation of Gravity Equation: A Cluster Analysis*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz