

ICASE 2017

International Conference on
Applied Statistics and Econometrics

Proceedings Book

Tirana 2017

ICASE 2017
INTERNATIONAL CONFERENCE ON APPLIED STATISTICS
AND ECONOMETRICS
27-28 APRIL 2017

PROCEEDINGS

EPOKA UNIVERSITY
2017

PROCEEDINGS

International Conference on Applied Statistics and Econometrics (ICASE 2017)

27-28 April 2017

Tirana, Albania

Publisher:

Epoka University

Editors:

Uğur Ergün

Jonada Tafa

Conference Partners:

Epoka University

Institute of Statistics Republic of Albania

Economic Society of Albania

Universiteti Marin Barleti

Design:

Epoka University

Printed by: **Epoka University Press**

Place of Publication: **Tirana**

Copyright: **Epoka University, 2017**

ISBN 978-9928-135-20-9

1. Ekonomia matematike 2.Ekonometria 3.Konferenca

330.4 (062)

Reproduction of the publication for educational or other non-commercial purpose is authorized without prior permission from the copyright holder. Reproduction for resale or other commercial purposes prohibited without prior written permission of the copywrite holder.

Disclaimer: While every effort has been made to ensure the accuracy of the information, contained in this publication, Epoka University will not assume liability for writing and any use made of the proceedings, and the presentation of the participating organizations concerning the legal status of any country, territory, or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Call for Papers:

Epoka University, in collaboration with Institute of Statistics Republic of Albania and Economic Society of Albania, organizes the 1st International Conference on Applied Statistics and Econometrics (ICASE 2017). The conference will be held on 27th and 28th of April 2017, in Tirana, Albania. Academicians, scholars, professionals and students are invited to submit their papers/posters, projects, proposals, and demonstrations for ICES'15. The conference will be held in English.

Aim:

- To bring together leading academic scientists, researchers and research scholars to exchange and share their experiences and research results regarding Applied Statistics and Econometrics.
- To provide special forums for researchers, practitioners and educators to present and discuss the most recent innovations, trends, and concerns, practical challenges encountered and the solutions adopted in the field of Applied Statistics and Econometrics.

Scope and Topics:

International Conference on Applied Statistics and Econometrics will provide an excellent international academic forum for sharing knowledge and results in methodology and applications of Applied Statistics and Applied Econometrics. The main objective is to bring together leading academic scientists, researchers and research scholars to exchange and share their experiences and research results regarding Applied Statistics and Econometrics. Authors are requested to contribute to the conference by submitting papers that illustrate research results, projects, surveying works and industrial experiences.

The participation in this international conference may be under following categories:

Original Research Articles

Published Articles

Research Poster

Dissertation/ PhD Synopsis

Research Abstract

Listener/ Co-author

Prospective authors are invited to submit Full Papers/ Abstract of Original Research work or, Synopsis of PhD/Dissertation, Published work, View-points or Way Forward/ Poster by online in conference webpage.

Topics of interest include, but not limited to, the followings:

Applied Statistics

Biostatistics and Bioinformatics

Data Collection

Data Mining

Design of experiments

Mathematical Statistics

Measurement

Modeling and Simulation
Network Analysis
Sampling Techniques
Social Science Methodology
Statistical Applications
Statistics Education
Other Areas of Statistics

Applied Econometrics

Deterministic and Stochastic Assumptions
Simultaneous Equation Models
Endogenous Explanatory Variables
Distributed Lag Models
ARIMA Modelling
Volatility
Co-integration
Panel Data
Non Linear Models and Structural Breaks
Limited Dependent Variables
Vector Autoregression
Time Series Filters
Other Areas of Econometrics

Foreword

International Conference on Applied Statistics and Econometrics provides a scientific platform that helps researchers to meet, discuss and exchange ideas. It reinforces bilateral scientific collaborations among researchers from different countries.

The main purpose of the conference is to serve as a platform for researchers, practitioners and educators to present and discuss the most recent innovations, trends, concerns, practical challenges encountered and the solutions adopted in the field of Applied Statistics and Econometrics.

It is a great honor for us to welcome as Keynote Speakers two prominent professors in this conference. We deeply thank Prof. Dr. Soren Johansen and Prof. Dr. Duo Qin for their very valuable contributions to ICASE 2017.

We would also like to thank all participants, partners, organizing committee and "Finance Club" members for sharing their insights with us.

Uğur Ergün
Jonada Tafa
Epoka University
Tirana, Albania

TABLE OF CONTENTS

1- Agricultural Policies as one of Determinants of Slovak Wine Export: A Gravity Model Approach	8
<i>Eva Judinová (Slovak University of Agriculture in Nitra)</i>	
<i>Dimuth Nambuge(Slovak University of Agriculture in Nitra)</i>	
2- Albanian Institute of Statistics Approach to R Language	17
<i>Endri Raço (Tirana Polytechnic University)</i>	
<i>Alma Kondi(Tirana Polytechnic University)</i>	
3- An Empirical Analysis on the Long Run Relation between Unemployment and Higher Education in Turkey	22
<i>Gungor Turan (Epoka University)</i>	
4- Application of Gravity Model: the Albanian Agricultural Export	30
<i>Kushtrim Braha (Slovak University of Agriculture in Nitra)</i>	
<i>Ema Lazorčáková (Slovak University of Agriculture in Nitra)</i>	
<i>Miroslava Rajčániová (Slovak University of Agriculture in Nitra)</i>	
<i>Artan Qineti (Slovak University of Agriculture in Nitra)</i>	
<i>Andrej Cupák (National Bank of Slovakia)</i>	
5- Comparison of Integrated Variance Forecast	52
<i>Vladimir Holy (University of Economics, Prague)</i>	
6- Determinants of Trade and the Gravity Model of Trade: the Case of Western Balkan Countries	59
<i>Visar Malaj (Faculty of Economics, University of Tirana)</i>	
7- Exchange Rate Pass-Through Effect In Albania: A Structural VAR Approach (SVAR)	71
<i>Aida Salko (Faculty of Economics, University of Lubljana)</i>	
<i>Ardit Gjerci (Faculty of Economics, University of Lubljana)</i>	
8- Guided Kernel Density Estimator and the Gamma Kernel Estimator	85
<i>Lule Hallaci (Faculty of Natural Science)</i>	
<i>Llukan Puka (Faculty of Natural Science)</i>	
9- How do Accounting Professionals Perceive Whistleblowing Reasons and Whistleblowing Preferences?	92
<i>M.Sait Dinc (International Burch University)</i>	
<i>Cemil Kuzey (Arthur J. Bauernfeind College of Business Murray State University)</i>	
<i>AliHaydar Gungormus (Independent Scholar)</i>	
<i>Bedia Atalay (Independent Scholar)</i>	
10- Impact of Attention Driven Investments on Agricultural Commodity Prices	105
<i>Miroslava Rajcaniova (Slovak University of Agriculture in Nitra)</i>	
<i>Tomas Misecka (Slovak University of Agriculture in Nitra)</i>	
<i>Jan Pokrivcak (Slovak University of Agriculture in Nitra)</i>	
<i>Pavel Ciaian (European Commission)</i>	
11- Inflation Dynamics in Albania: A Markov Regime-Switching Approach	128
<i>Anisa Plepi (Faculty of Economics, University of Tirana)</i>	
12- Land Cover Statistics as a Measure of Natural Capital Distribution Fairness among altered Administrative Territorial Divisions	140
<i>Artan Hysa (Epoka University, Istanbul Technical University)</i>	

13- Macro-Economic Factors Affecting Ease of Business in Balkan Countries	151
Bora Kokalari (Epoka University)	
Emanuela Buci (Epoka University)	
14- Mathematical Simulation of an Accident Situation at Intersection	162
<i>Erjola Cenaj (Polytechnic University of Tirana)</i>	
<i>Raimonda Dervishi (Polytechnic University of Tirana)</i>	
<i>Shkelqim Kuka (Polytechnic University of Tirana)</i>	
15- Metadata, the DNA of Statistical Data	166
<i>Ertugrela Curumi (Albanian Institute of Statistics)</i>	
<i>Ilda Shabani (Albanian Institute of Statistics)</i>	
<i>Olta Kodra (Albanian Institute of Statistics)</i>	
16- Statistical Indicators as Potential Early Signals of Transitions in Time Series Obtained by a Statistical Model: Geomagnetic Field Case	175
<i>Klaudio Peqini (Faculty of Natural Sciences, University of Tirana)</i>	
<i>Bejo Duka (Faculty of Natural Sciences, University of Tirana)</i>	
17- The Importance of Inference Bayesian in Telecommunication Industry-Albanian Market	187
<i>Eralda Caush (Università Cattolica "Nostra Signora del Buon Consiglio")</i>	
18- Time scale regression analysis of oil and interest rate on the exchange rate: A case study for the Czech Republic	193
<i>Lukas Fryd (University of Economics, Prague)</i>	
19- Using Bayesian Methods for Categorical Data Analysis	200
<i>Erjola Cenaj (Polytechnic University of Tirana)</i>	
<i>Raimonda Dervishi (Polytechnic University of Tirana)</i>	

Agricultural Policies as one of Determinants of Slovak Wine Exports: A Gravity Model Approach¹

Eva Judinová, Dimuth Nambuge

Department of Economic Policy, FEM, Slovak University of Agriculture in Nitra, Slovakia
eva.judinova@uniag.sk, dimuth.nambuge@uniag.sk

Abstract

Slovak wine export is growing each year. Not only the amount of transactions is bigger but also new markets are approached. In general, the following factors are considered to be significant in influencing the amount of exported goods: GDP per capita and number of inhabitants in importing countries, distance between trade partners, amount of domestic production of the good and its consumption in receiving countries. According to studies dealing with identification of bilateral trade determinants, there are also other factors that can have an impact on export of a specific good: agricultural policies, membership in international and trade organizations, common characteristics and similarities between the domestic country and business partners. The objective of this article is to identify the determinants of Slovak wine export in period of 2004-2013 using the gravity model approach, and to determine the effect of agricultural policies on exports.

The results show that foreign consumers consider the Slovak wine to be an inferior good. No significant effect of EU, EU monetary union and OECD membership was identified. We also found out that Slovak wine exporters tended to export more to non-WTO countries. No evidence of impact of the free trade agreements on the wine export value was found. The coefficient of the similarity index was estimated to be significant but negative. This implies that Slovak wine export was realized prevalent to partner countries with different GDP similarity index, e.g. Germany, US, United Kingdom, China or Japan.

Keywords: foreign trade, export determinants, wine, gravity model, agriculture policies

JEL Classification: Q17, F14, C23

Literature review

Slovakia is a relatively small producer of wine; domestic production represents only about 0.2% of production of the European Union. The tradition of wine production, however, dates to the 9th century AD. Slovak wine sector is currently characterized by reduction of domestic production, which fell from 515 000 hl in 2004 to 370 000 hl in 2014. It is expected that this trend will continue. This situation opens the door to foreign producers. The share of imported wines in the Slovak total supply of wine increased from 20% (2004) to around 65% (2012-2014). Slovak consumers prefer particularly table wines, but domestic production is mainly focused on high quality wines. As demand for domestic production is a relatively small, part of the production is exported abroad. The value of exported wine is growing on average.

The decision to sell wine to foreign markets should be made with respect to the characteristics of these markets. One way of identifying the factors stimulating foreign trade is the gravity model. Most studies using gravity model deal with the simulation of total foreign trade of countries. A smaller number of studies is focused on analysing the foreign trade of specific commodities such as wine. For example, Carlucci, De Blasi, Santeramo, & Seccia (2008) use the gravity model results to formulate recommendations for the orientation of Italian wine production. They recommend that Italy should increase the production of high quality wines,

¹ This work was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0552.

because there are favourable conditions for increasing exports of this segment, and decrease the production of table wines, which international demand is falling. Koutroupi, Natos, & Karelakis (2014) analysed the competitiveness of Greek wines in the European market. According to them, the key factors of business success are the level of consumption per capita in the EU countries, the existence of common borders and the use of a common language among trading countries, and geographical range of mutual trading partners. One of the gravity model's basic variables, the distance between trade partners, is considered to be a trade barrier. Dal Bianco, Boatto, Caracciolo, & Santeramo (2014) found that the effect of distance is not as strong in the wine sector as in other sectors. It is because the wine has a long shelf life, and therefore, it does not create additional variable costs related to product's delivery speed. The authors also state that the impact of transport costs and their proxy variable (distance) on development of trade relations is limited because it plays an increasingly significant role in product differentiation. Imported wines cannot be fully substituted, and consequently, the distant importers do not substituted wines imported from distant markets with wines of close business partners.

According OECD (2016), the main goal of current agricultural policies is to achieve a productive, sustainable and resilient global food system to be able to provide all consumers with reliable access to safe, healthy and nutritious food; to enable producers to operate in an open and transparent global trading system; and to contribute to sustainable productivity and to inclusive growth and development within and across countries. The EU has been heavily criticised for its agricultural policy that induced overproduction, export dumping and distorting of markets due to employing of trade tariffs. But on the other hand, EU members are signatories of many free (FTA) and regional (RTA) trade agreements. The FTAs can be used to negotiate a reduction of tariff and non-tariff barriers, they can be employed by countries to create competitive advantages for their export of goods.

In this paper, the effect of agricultural policies on Slovak wine exports is determined by estimating the relationship between free trade agreements (of which the Slovak Republic is a signatory) and the Slovak wine export value. As was published by Baier & Bergstrand (2004), the empirical results suggest that effects of FTAs on trade flow estimated using standard cross-section gravity equation are biased. The best method to estimate the FTA effect is to employ differenced panel data, moreover, the authors proposed to add fixed and random effects into the model. However, the recent studies do not provide clear evidence on positive effects of the free trade agreements on trade stimulation; as in case of Soloaga & Winters (2001), who investigated the potential of FTA between EU and EFTA. Hatab, Romstad, & Huo (2010) estimated the effect of RTA on Egypt's agricultural export and found out that the RTA variable was not significant but positive. The fact that a country is a member of RTA with Egypt did not influence its export volume. Eger (2004) states that FTAs are not expected to have a short-term effect on volumes of trade, but in the long run, he predicted a 15% increase for NAFTA member countries.

Methodology

The gravity model can be used also for modelling the trading allocation of goods transmitted from the export country (i) to the destination importing countries (j). The objective of this article is to identify the determinants of Slovak wine export in period of 2004-2013 using the gravity model approach, and to determine the effect of agricultural policies on wine exports. In this period, Slovak wine was globally exported into 52 countries. To eliminate outliers, we have excluded observations for countries, where export occurred only once. The final data set consists of observations for 42 countries.

As a base, we used the model developed by Carlucci, De Blasi, Santeramo, & Seccia (2008):

$$\ln \text{Exp}_{jt} = \alpha_0 + \alpha \cdot \ln \text{Prod}_{it} + \beta \cdot \ln \text{PcGDP}_{jt} + \gamma \cdot \ln \text{Pop}_{jt} + \delta \cdot \ln \text{Dist}_j + \lambda \cdot \text{kGroup}_k + \varepsilon_{jt}, \quad (1)$$

where:

Exp_{jt} – value of Italian wine exports to country j in year t , in EUR (constant prices)

α_0 – constant term

Prod_{it} – production of Italian quality wine in year t , in hl

PcGDP_{jt} – GDP (gross domestic product) per capita of importing country j in year t , in USD (constant prices)

Pop_{jt} – population of importing country j in year t , in mil. of inhabitants

Dist_j – distance between importing country j and the exporting country i (Italy), in km

Group_k – dummy variable which takes the value 1 if country j belongs to group k

ε_{jt} – error term

We extended the model (1) by including other variables, which are expected to influence Slovak wine exports. To identify the relationship of depended variable and independent variables we estimate several models. Then, the selection of the best fitting model is conducted. The criteria for models' evaluation are described further in this paper.

The gravity equation has logarithmic form. Depended variable (value of Slovak wine export) can reach also zero values in a given year, but the logarithm of 0 is not mathematically defined. One way to solve this problem is to add the constant 1 ($\text{Exp}_{jt}+1$) to all values of depended variable; such model remains balanced. The second method assumes omitting all observations with zero dependent variable, $\text{Exp}_{jt} \neq 0$, (Koren & Tenreyro, 2005). Hence, an unbalanced model is created.

The first estimated model (balanced model A) is a simple extension of classic linear regression analysis to a panel data model, i.e. pooled regression model. It is an estimation method, where the heterogeneity of countries is not identified. The equation for model A is following:

$$\ln \text{Exp}_{jt} = \alpha_0 + \alpha \cdot \ln \text{Prod}_{it} + \beta \cdot \ln \text{PcGDP}_{jt} + \gamma \cdot \ln \text{Pop}_{jt} + \delta \cdot \ln \text{Cons}_{jt} + \zeta \cdot \ln \text{Dist}_j + \eta \cdot \ln \text{RFE}_{jt} + \theta \cdot \ln \text{SIM}_{jt} + \lambda_1 \cdot \text{EU}_{jt} + \lambda_2 \cdot \text{OECD}_{jt} + \lambda_3 \cdot \text{WTO}_{jt} + \lambda_4 \cdot \text{Curr}_{jt} + \lambda_5 \cdot \text{FTA}_{jt} + \lambda_6 \cdot \text{Hist}_j + \lambda_7 \cdot \text{Bord}_j + \lambda_8 \cdot \text{Lang}_j + \varepsilon_{jt} \quad (2)$$

Where:

Exp_{jt} – value of Slovak wine exports to importing country j in year t , in EUR (constant prices)

α_0 – constant term

Prod_{it} – production of Slovak wine in year t , in 1000 hl

PcGDP_{jt} – GDP per capita of importing country j in year t , in USD (constant prices)

Pop_{jt} – population of importing country j in year t , in mil. of inhabitants

Cons_{jt} – consumption per capita of importing country j in year t , in litres

Dist_j – distance between importing country j and the exporting country i , in km

RFE_{jt} – relative factor endowments between the trading countries i and j

SIM_{jt} – similarity index of the trading countries i and j

EU_{jt} , OECD_{jt} , WTO_{jt} – dummy variable which takes the value 1 if a country pair ij belongs to these organizations

Curr_{jt} – dummy variable which takes the value 1 if a country pair ij has a common currency

FTA_{jt} – dummy variable which takes the value 1 if the country pair ij has a signed free trade agreement

Hist_j – dummy variable which takes the value 1 if a country pair ij has a common territorial history

Bord_j – dummy variable which takes the value 1 if a country pair ij has a common state border

Lang_{ij} – dummy variable which takes the value 1 if a country pair *ij* has a common language base

ε_{ijt} – error term

α – η; λ₁ – λ₈ – the sensitivity change of the dependent variable to the changes in independent variables.

The second estimated model (mode B) is unbalanced pooled regression model with the same equation as for model A (2).

According to studies done on the topic of international trade, e.g. De Blasi, Seccia, Carlucci, & Santeramo (2007), to capture the unobserved heterogeneity, it is suggested to consider adding fixed effects into the panel model. Here, country-specific effects or time effects can be taken into account. These effects could have a fixed or a random characteristic. Therefore, a Hausman test was performed to define whether the supposed effects are random or fixed. A presence of fixed effects in the panel data was determined. For this reason, we estimate also models C-F with fixed effects.

The balanced model C and the unbalanced model D include country-specific fixed effects and are explained in equation 3:

$$\ln \text{Exp}_{ijt} = \alpha_0 + \alpha \cdot \ln \text{Prod}_{it} + \beta \cdot \ln \text{PcGDP}_{jt} + \gamma \cdot \ln \text{Pop}_{jt} + \delta \ln \text{Cons}_{jt} + \zeta \ln \text{RFE}_{jt} + \eta \cdot \ln \text{SIM}_{ijt} + \lambda_1 \cdot \text{EU}_{ijt} + \lambda_2 \cdot \text{OECD}_{jt} + \lambda_3 \cdot \text{WTO}_{jt} + \lambda_4 \cdot \text{Curr}_{jt} + \lambda_5 \cdot \text{FTA}_{ijt} + \varepsilon_{ijt}, \quad (3)$$

It should be noted that non-time varying variables cannot be estimated in model with the country-specific fixed effects. Because of that, the variables common language base, common territorial history, common state borders, and distance of the traded partners had to be excluded from the model.

Models E (balanced) a F (unbalanced) include both country-specific and time fixed effects. Because of presence of time-specific fixed effects, also non-country varying variables as is the production of country *i* had to be excluded from the model. Finally, models E and F are defined by the equation 4:

$$\ln \text{Exp}_{ijt} = \alpha_0 + \alpha \cdot \ln \text{PcGDP}_{jt} + \beta \cdot \ln \text{Pop}_{jt} + \gamma \ln \text{Cons}_{jt} + \delta \ln \text{RFE}_{jt} + \zeta \ln \text{SIM}_{ijt} + \lambda_1 \cdot \text{EU}_{ijt} + \lambda_2 \cdot \text{OECD}_{jt} + \lambda_3 \cdot \text{WTO}_{jt} + \lambda_4 \cdot \text{Curr}_{jt} + \lambda_5 \cdot \text{FTA}_{ijt} + \varepsilon_{ijt}. \quad (4)$$

All fixed-effects models were estimating by OLS and dummy variables for all partner countries and years (LSDV) were included. The best fitted model was selected by comparing the following characteristics (König & Schulze, 2008):

- measure of adjusted R-squared coefficient, i.e. the higher is the coefficient, the more variability in dependent variable is explained through the model,
- Mean square error (MSE), where the better model is the one with lower MSE,
- Akaike information criterion (AIC), where the better is the model with the lower AIC.

Description of variables

In this chapter, we characterised variables selected to explain the development of Slovak wine exports. Independent variables were selected in accordance with the results of related studies considering the current situation in the Slovak wine market.

Due to the orientation of Slovak consumers on table wines, which are mostly imported, we assume that changes in domestic production of wine affect the size of its export. Therefore, we estimate the impact of variable *Prod* in this paper.

GDP per capita of importing country represents the income elasticity of foreign demand for Slovak wine. We expect that the increase in income of importing countries affects the size of export positively. We also expect a positive effect of increase in population and increase in wine consumption of these countries on Slovak wine exports.

Regarding the variable *Dist*, trade theory largely assumes that the distance between business partners influences trade among countries negatively. On the other hand, the strength of this relation may be limited due to the type of commodity traded, as reported by some studies.

Slovakia is a member of international organizations that have formed a legislative framework within which the implementation of business activities occurs. These rules can act stimulating on trade when simplifying business operations between the member states. Additionally, the membership creates economic connections that can lead to strengthening of trade relations. However, we can assume that trade with non-member countries is to some extent restricted.

Other variables which influence on the Slovak wine export we want to estimate are common territorial history, common national borders and common language elements of trading countries, common currency and country membership in international organizations. We want to determine whether these factors influence Slovak wine export positively, and therefore, which countries it is advantageous for Slovak exporters to focus on.

Indexes *RFE* and *SIM* represent the rate of economic similarity between the export country and import countries. *RFE* coefficient is a proxy for the level of country's equipment with production factors (PF). If *RFE* has the value of 0, country *i* and a country *j* show the same level of equipment with PF. The higher the *RFE*, the greater is also the differences in of country's equipment with PF. We assume that the differences in production factors' equipment motivate countries to mutual trade. For calculating *RFE*, we use the equation by Baltagi, Egger, & Pfaffermayr, 2003:

$$RFE_{ijt} = |\ln PcGDP_{it} - \ln PcGDP_{jt}| \quad (5)$$

SIM index determines the similarity between *i* a *j* in size of their economy measured by GDP (Kabir & Salim, 2010):

$$SIM_{ijt} = 1 - \left(\frac{\ln GDP_i}{\ln GDP_i + GDP_j} \right)^2 - \left(\frac{\ln GDP_j}{\ln GDP_i + GDP_j} \right)^2 \quad (6)$$

SIM index takes values from 0 to 0.5, where the value of 0.5 means that the size of the trading countries' economy is the same. If *SIM* = 0, it indicates the absolute difference in the size of the economy.

FTA represents the free trade agreements between countries that signed such agreement. Considering the results of some studies, we expect the *FTA* coefficient will have a slightly positive sign, which would mean that Slovak wine exports between the member states improved.

The source of data on the population of each country and its GDP is the World bank database. Wine consumption per capita of importing countries is drawn from data portal Wineinstitute.org. Distance between Slovakia and importing countries was calculated based on the air within distance their capital cities. Data on Slovak wine production is obtained from

Eurostat and data on the value of Slovak wine exports from the INTRASTAT database of Slovak Republic. The list of FTAs is obtained from RTA database of World trade organization.

Results

As mentioned in Chapter 2, we estimated 6 models, which under different conditions describe the relationship between Slovak wine exports in the period of 2004-2013 and the factors affecting its value. In Table 1 we summarize the characteristics of estimated models and, we list the order according their suitability to explain the variability of the depend variable.

In summary, the unbalanced models are more suitable to describe the variability of the dependent variable. The best model is the unbalanced model F with fixed effects, which consist of both country-specific and time-specific effects. Based on the value of the adjusted determination coefficient's we can say that the model and selected determinants explain the variability of the dependent variable to 68.44%. Durbin-Watson statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in the data file. Since the P-value is greater than 0.05, there is no indication of serial autocorrelation in the residuals at the 95.0% confidence level.

Table 1 Comparison of estimated models according to selected criteria

model	type*	rank	AIC	adj. R ²	MSE
A	bal (pooled)	6	291,8076	25.8235	17.493
B	unbal (pooled)	3	202,2373	33.5851	7.14271
C	bal (country FE)	5	268,0959	41.5125	13.8001
D	unbal (country FE)	2	132,1128	67.0602	3.54257
E	bal (country, time FE)	4	267,4342	41.8387	13.7091
F	unbal (country, time FE)	1	127,3775	68.5837	3.37872

*bal – balanced model

unbal – unbalanced model

FE – fixed-effects model

Source: own calculation

With exception of 4 coefficients, the coefficients of country-specific fixed effects are significant at the 99.0% confidence level. Estimated equation for model F (7) and model's estimation results (Table 2) are shown below²:

$$\ln \text{Exp}_{jt} = 56.1874 - 3.7408 \ln \text{PcGDP}_{jt} - 4.0381 \ln \text{Pop}_{jt} + 2.5889 \ln \text{Cons}_{jt} - 0.2794 \ln \text{RFE}_{jt} - 6.0126 \ln \text{SIM}_{jt} + 1.7969 \text{EU} + 3.2888 \text{OECD}_{jt} - 2.071 \text{WTO}_{jt} - 0.6433 \text{FTA}_{jt} \quad (7)$$

Table 2 Regression results, model F

Parameter	Estimate	Standard Error	T Statistic	P-Value	Significant
CONSTANT	56.1874	16.6479	3.37505	0.0009	***
lnPcGDP	-3.74079	1.8232	-2.05177	0.0419	**
lnCons	2.58894	0.669156	3.86897	0.0002	***
lnPop	-4.03814	1.34866	-2.99417	0.0032	***
lnRFE	-0.279448	0.234724	-1.19054	0.2357	
lnSIM	-6.01259	1.83162	-3.28266	0.0013	***
EU	1.79693	1.57592	1.14024	0.2560	
OECD	3.2888	2.76076	1.19126	0.2354	
WTO	-2.07101	1.16432	-1.77873	0.0773	*
FTA	-0.643333	0.995533	-0.64622	0.5191	

R-squared = 77.3068 percent

²Coefficients of the fixed effects are omitted. In the equation 7, a simplified model is presented (the most insignificant variables were eliminated from the model).

R-squared (adjusted for d.f.) = 68.4399 percent
 Standard Error of Est. = 1.84233
 Mean absolute error = 1.14152
 Durbin-Watson statistic = 2.06349 (P=0.6771)
 Lag 1 residual autocorrelation = -0.0317808
 Significant: *** at 1%; ** at 5%; * at 10%

Source: own calculation

GDP per capita represents the income elasticity of foreign demand for Slovak wine. The estimated coefficient of the variable is significant at a significance level of 95%; we can say that one percent increase in GDP per capita of importing country would cause a decline in the value of Slovak wine exports by 3.74%, *ceteris paribus*. This conclusion points to the fact that foreign consumers perceive Slovak wines as inferior goods. This result could be related to further result that even the increase in the population of importing countries had not have a positive impact on the value of Slovak wine exports. Compared to our assumptions, this fact is surprising. An explanation could be that that in bigger countries, there is usually a wider range of wine products which people can choose from, and it is likely that foreign consumers preferred the other than Slovak wines more.

Based on the estimation we can say that an increase in wine consumption per capita of importing countries increased the value of wine exports from Slovakia. In 2004, there was a relatively large expansion of the European Union. The expectation was that this situation affected Slovak wine exports positively. However, the results show that the EU membership of Slovak trade partners did not affected the changes in Slovak wine exports significantly. Moreover, the variable common currency in the European Monetary Union (*Curr*) was finally eliminated from the model due to the high insignificance of its coefficient.

We can say that at the 90.0% confidence level, the fact that the trading partners (*i* and *j*) are members in the WTO, did not influenced the Slovak wine exports positively. Comparable results were determined also in the study of Lissovolik & Lissovolik (2004). According to them, some of the exporting countries tend to export more to non-WTO countries than to WTO countries. However, to be able to explain these facts better, it is needed explore the issue further and in more detail.

RFE index indicating the level of country's *i* and *j* equipment with factors of production is not significant. On the contrary, the index of similarity is highly significant; the differences in size of Slovak economy and economies of its trading partners encouraged Slovak wine exports. Countries that in terms of economy size at most differ from Slovakia are the US, Japan, Germany, China, Malta, France and the United Kingdom. Empirical data in the observed period confirm results of the estimated model, where the value of exports to mentioned countries exceeded the value of exports to countries with similarity index close to 0.5 (except for the Czech Republic, which in this case is considered an outlier).

Given the fact that the model most suitable to describe the relationship between the dependent and independent variables is the model with both country-specific and time-specific fixed effects, it was not possible to examine the effect of time and country non-varying variables: the common language base, shared territorial history, common national borders and distance between trading partners and Slovakia.

Conclusion

The aim of this paper was to identify the determinants of Slovak wine export to 42 countries worldwide in period of 2004-2013 using the gravity model approach, and to determine the effect

of agricultural policies. We found out that the most suitable models in terms of the goal of this paper are the unbalanced panel models with fixed effects.

As expected, the growth of wine consumption per capita of the importing country has resulted to an increase of Slovak wine exports. This means that for Slovak exporters it is necessary to monitor the preferences of foreign consumers and to focus on markets that have the potential to absorb the additional supply.

Surprising is the result that Slovak wines are considered inferior goods by foreign consumers. It is likely that countries, where the income per capita grew faster, would gradually reduce the consumption of Slovak wines. Therefore, it is preferable to direct the wine exports to countries with stable incomes than to faster growing economies. On the other hand, the Slovak wine producers should look for ways how to make the wine product more attractive in eyes of foreign consumers, or how to increase its added value. Moreover, with appropriate marketing tools, it is desirable to stimulate the interest of domestic wine consumers too. For example, through the organization of wine roads and wine tourism globally, through tasting and trying to win awards at national and international exhibitions, which would present the product positively.

Using the best model estimated, we were unable to identify a significant impact of membership in the EU and EMU on the value of exported wine. We also found that the Slovak wine exporters tend to trade more with countries with different sized economies such as Germany, USA, United Kingdom, China and Japan.

We did not find any evidence of impact of free trade agreements signed by the Slovak Republic on the value of wine exported to member states. The reason may be that the period was too short for the FTAs effect to manifest itself.

References

- Baier, S. L., Bergstrand, J. H. 2004. Do free trade agreements actually increase members' International trade? Working paper. Retrieved April 8, 2017, from http://www3.nd.edu/~jbergstr/Working_Papers/BaierBergstrandFTA2Oct2004.pdf
- Baltagi, B., Egger, P., Pfaffermayr, M. 2003. A generalized design for bilateral trade flow models. In *Economics Letters* 80 (2003), pp. 391–397.
- Dal Bianco, A., Boatto, V., Caracciolo, F., Santeramo, F. G. 2014. Tariffs and non-tariff frictions in the world wine trade. Retrieved August 25, 2016, from https://mpra.ub.uni-muenchen.de/61813/1/MPRA_paper_61813.pdf
- De Blasi, G., Seccia, A., Carlucci, D., Santeramo, F. G. 2007. Analysis of Italian High Quality Wine Exports using the Gravity Model Approach. Retrieved November 20, 2016, from <https://ideas.repec.org/p/ags/eea105/7901.html>
- Carlucci, D., De Blasi, G., Santeramo, F. G., Seccia, A. 2008. New challenges and opportunities for Italian exports of table wines and high quality wines. Retrieved February 2, 2017, from http://mpra.ub.uni-muenchen.de/8728/1/MPRA_paper_8728.pdf
- Egger P. 2004. Estimating regional trading bloc effects with panel data. *Review of World Economics*, 2004, 140(1), pp. 151-66.
- Hatab, A. A., Romstad, E., Huo, X. 2010. Determinants of Egyptian Agricultural Exports: A Gravity Model Approach. Retrieved April 5, 2017, from http://file.scirp.org/pdf/ME20100300002_48599092.pdf
- Kabir, M., Salim, R. 2010. Can Gravity Model Explain BIMSTEC's Trade? Retrieved February 4, 2017, from <http://www.e-jei.org/upload/W17332525G5054V7.pdf>

- Koren, M., Tenreyro, S. 2005. Volatility and Development: CEP Discussion Paper No 706. Retrieved August 23, 2016, from http://eprints.lse.ac.uk/3743/1/Volatility_and_Development.pdf
- Koutroupi, E., Natos, D., Karelakis, Ch. 2014. Assessing Exports Market Dynamics: The Case of Greek Wine Exports. Retrieved March 20, 2017, from <http://www.sciencedirect.com/science/article/pii/S2212567115000209>
- König, J., Schulze, P. M. 2008. Zur Analyse rheinland-pfälzischer Exporte mittels Gravitationsmodell. Retrieved March 22, 2017, from http://www.statoek.vwl.uni-mainz.de/Dateien/Arbeitspapier_Nr_34_Gravitationsmodell.pdf
- Lissovlik, B. and Lissovlik, Y. 2004. Russia and the WTO: The “Gravity” of Outsider Status. Retrieved February 12, 2017, from <https://www.imf.org/external/pubs/ft/wp/2004/wp04159.pdf>
- OECD. 2016. Declaration on better policies to achieve a productive, sustainable and resilient global food system. Retrieved April 7, 2017, from <https://www.oecd.org/agriculture/ministerial/statements/>
- Soloaga I, Winters A. 2001. Regionalism in the Nineties: What Effect on Trade?” The North American Journal of Economics and Finance, 2001, 12(1), pp. 1-29.

Albanian Institute of Statistics Approach to R Language

Endri Raço¹, Alma Kondi²

¹ Department of Mathematical Engineering, Tirana Polytechnic University, Albania
e.raco@fimif.edu.al

² Albanian Institute of Statistics, Albania akondi@instat.gov.al

Abstract

In 2016 R moved up to 5th place in IEEE programming language rankings. Statisticians and non-statisticians love R because R is great when it comes to statistics and data analysis. To some degree R is used by all the institutions in charge of producing official statistics. In 2015 Albanian Institute of Statistics embraced a first new step to introduce R as a future alternative tool for daily work. In this paper, we focus to the lessons learned and try to bring together any insights gained during this process. Also, based on some interesting related projects carried out in international statistical offices our work aims to emphasize the bunch of possibilities R can offer to Albanian Institute of Statistics.

Keywords: R, official statistics, INSTAT, analysis

JEL Classification: C15, C88, C63

Introduction

National statistical offices are responsible for collecting and publishing and publish empirical information regarding economic and social factors. There is a huge amount of data which continues to grow and plenty of requirements regarding analysis and reporting which led to the development of sophisticated methods to collect, process, analyze and supply information. It is a well-known fact for everybody who deals with national statistics that from time to time, there is need for customized software to perform the above-mentioned tasks. For data processing and data analysis in INSTAT, several well-established statistical software packages are commonly used:

- SAS (Institute, 2017) :sampling and weighing.
- CSPRO 6.2 (Bureau., 2012), SQL (Johnson, 2011) :data collection.
- SPSS (IBM, 2016), SAS, Office (Microsoft, 2017): data manipulation.
- SAS, RIDA (Sabanovic, 2015): logic rules and imputation.
- SPSS, SAS, Office: data tabulation.

However, if the need is for:

- reproducible research
- dynamical reports
- web-based applications
- flexible data manipulation

and last but not the least cost reduction we have to consider R as our solution to these question (Matthias Templ, 2016).

In 2016 R moved up to 5th place in IEEE programming language rankings (Smith, 2016). The number of such contributions has risen from about 100 packages in 2001 to about 10323 available packages at the time of this writing.

Another sign of R increasing popularity is the fact that several important commercial software providers such as Google, SAP, Oracle, IBM, Microsoft, ESRI are joined in a consortium to

support the worldwide community of users, maintainers and developers of R software (Consortium, 2016).

For a few years now, Albanian Institute of Statistics has made first steps toward introducing with R capabilities. There have been several activities with the main purpose of comparing R to existing tools and trying to identify the benefits of using R in everyday work.

Amongst these activities are the organization of several courses covering topics such as Data Wrangling, Data Cleaning, Regression Methods using R. Also, a very courageous step was the creation and publication of first R software in Albania, Database of Albanian Names.

The remainder of this paper is divided in two main parts. In the next section, we will have a further look to R capabilities and how R has been introduced at INSTAT. In section 3 we discuss R approach in different international statistical offices around Europe and finally, we will give some recommendations about further steps that need to be undertaken in INSTAT for making R the favorite tool and get the best of it.

R introduction in INSTAT

On 25 November 2014, INSTAT and the Embassy of Sweden in Tirana signed the Agreement for the next stage of the statistical cooperation between INSTAT and Statistics Sweden (SCB). The project follows the Swedish Governments Results Strategy for Western Balkans and Eastern Europe (Instat, 2012).

As part of this project November 2014, SCB started a pilot project to determine how to best deploy R within Albanian Institute of Statistics. The goal of this project was to select introduce INSTAT staff with R core possibilities.

It was decided to start with a 5- weeks beginner level course. This course was attended by a group of 15 persons representing all directorates inside the institution. The result was that R was introduced in three flavors, depending on its use in statistics production, statistical research, or research in methods and computation.

1. Usage of R language for visual data representation, data analysis and reporting. This part was mostly oriented to employees in charge of data analysis and reporting.
2. Usage of R language for simple regression analysis and exploratory factor analysis. This part was mostly oriented to employees in charge of methodology and sampling.
3. Integration of R language with SQL databases to perform different tasks such as reshape data frames, construct queries and return result sets.

First course resulted very successful and due to high demand a second course was organized shortly after. Another 15 employees attended the second course so in the middle of 2015 there were already 30 employees having a beginner level of familiarity with R. Besides the course, users were encouraged to compare the time spent for performing different tasks using existing software and R Language. This comparison resulted very helpful on identifying employees' real needs and creating the fundamentals for the intermediate level courses held at the end of 2015.

The second course main topics were data wrangling and data cleaning, tasks where most of time is spent in INSTAT everyday work. In two different courses, all 30 employees were presented with R packages, such as dplyr (Wickham, dplyr, 2016), lubridate (Wickham, lubridate, 2016),

tidyr (Wickham, tidyr, 2017) and stringr (Wickham, stringr, 2017). In 2016 by request of several directorates a special course of Regression Methods using R was organized. In this course were present several employees from previous courses and new employees profiled mostly in methodology and advanced analysis. Following the new trend and exploring possibilities, INSTAT with the support of UN Women created and implemented the first R software in Albania. The software presented a database of Albanian names and was intended as an interactive way of displaying interesting facts to the users (Raco, 2016). The application was a real success. Only in the first week after the presentation more than 40.000 clicks were registered in INSTAT web-page.

R in international statistical offices

The national statistical offices of Canada, Austria, Netherlands, Italy, USA, UK, the statistical offices of a few other countries and several international agencies are quite active in using R. At Statistics Austria, R is currently installed on more than 65 computers and on virtual servers. This is useful for tasks involving large memory requirements or to put content on the web, e.g. via shiny (Studio, 2017). The leading R-team at Statistics Austria consists of three methods division experts. In addition, each department has nominated one person as first contact in case of questions and problems. Beside the development of the packages laeken (Andreas Alfons, 2014), sparkTable (Alexander Kowarik, 2016), sdcTable (Meindl, 2017) and many others, R is applied in the production process for sampling, editing and imputation, estimation, analysis and output generation for several surveys (Matthias Templ, 2016).

The National Statistical Office of UK started to use R version 2.0.1. in 2004. Since 2012 an R development group has been installed, whose objective was to test if R, and related specialized packages, are ready for use in the production environment. The survey (Lumley, 2016) package is used for the Labor Force Survey (panel design). The spatstat (Baddeley, 2017) package and its kernel smoothing features are used to visualize crime data at postcode level. The MortalitySmooth package (MortalitySmooth, 2015) has been used for mortality rates estimation.

The National Statistical Institute of Romania established the Romanian R team in 2013. Various R packages play an important role for the Business Register Department, due to the need of using of administrative data in the production of business statistics. The Department of Indicators on Population and International Migration uses the package JoSAE (Breidenbach, 2015) for the application of small area estimation techniques. It is used to produce data on annual international migrant stocks. The Department of Social Statistics tests the use of R for sampling.

Particularly packages vardpoor (Juris Breidaks, 2017) and survey (Lumley, 2016) are tested for use in the production according to variance estimation for household surveys. The Statistical Office of Serbia uses the laeken package for poverty estimation. They compared the results obtained with laeken to those obtained with SAS Macros provided by Eurostat. They also use R for estimations on the monthly retail trade survey (Alexander Kowarik, 2016). Statistics Netherlands started using R in a systematic manner already in 2010. A knowledge center was built up; an internal wiki provides code examples and serves as a platform for knowledge sharing. Examples of R use in the statistical production process include the estimation of the Dutch Hospital Standardized Mortality Ratio (HSMR), the estimation of certain unemployment figures, estimation of tourist accommodations, and manipulation of supply and demand tables for National Accounts (Loo, 2012). Statistics Netherlands uses R also for data collection and web crawling with web robots, and for collecting data for compilation of price statistics. Several

packages were developed by Statistics Netherlands. The editrules package for data editing and the deducorrect package for deductive correction and deductive imputation. With the rspa package numerical records can be modified to satisfy edit rules (Loo, 2012).

Conclusions and Recommendations

Even by taking in consideration the results from the first phase of R implementation in INSTAT there is still plenty of work to do following the success stories of R implementation in statistical offices around Europe. INSTAT needs a strategy for using and distributing R software to their employees. There is still tendency of using common software and translating to R seems a difficult approach to employees used to SPSS or SAS. Moving to R seems to be slow due to the lack of strong programming knowledge and availability of many legacy code developed in other statistical software environments.

Apart from the pilot project, which ended in the beginning of 2016, SCB idea was that of a permanent R center of expertise. This center would support users, initiate upgrades develop and maintain coding standards and implements standard methodology in R. If we study carefully the approach of other statistical offices in getting R to their environment, we can create a skeleton of path INSTAT should follow in order to succeed in approaching R.

We present below this path in a number of steps:

1. SCB idea of a permanent R center of expertise has to be implemented as a key factor for further steps to be completed.
2. To assure a certain level of familiarity with R for all new users, each new R user must follow a week of training course developed by the center of expertise. Besides the course, it is advisable that all directorates assign a person in charge of R implementation inside their directorate.
3. It is crucial to start with a real-life project completely implemented in R. For example, some one or more publications have to be produced completely in R. This means the usage of R from data collection and cleaning to the creation of a reproducible reports where R really shines.
4. INSTAT has to seriously consider online reporting by using Markdown and Shiny. It is a well-known fact that old-fashioned pdfs are subject of interest only to field experts and researchers but not suitable when it comes to wide public and open government concept.
5. The organization of conferences and invitations of national and international R experts is a good practice to draw attention and grow interest in R usage benefits.

References

- Alexander Kowarik, B. M. (2016, December 13). sparkTable. *sparkTable: Sparklines and Graphical Tables for TeX and HTML*. CRAN.
- Andreas Alfons, J. H. (2014, 08 19). laeken. *laeken: Estimation of indicators on social exclusion and poverty*. CRAN.
- Baddeley, A. (2017, 03). spatstat. *spatstat analysing spatial point patterns*. GitHub.
- Breidenbach, J. (2015, 08 09). JoSAE. *JoSAE: Functions for some Unit-Level Small Area Estimators and their Variances*. CRAN.
- Bureau., D. I. (2012, May). Census and Survey Processing CSPRO. *CSPRO62*. USA: USAID.
- Consortium, R. (2016). *R Consortium*. Retrieved from R Consortium: <https://www.r-consortium.org>
- IBM. (2016, 03). SPSS Software. SPSS Software, Version 12. *SPSS Software, Version 12*. IBM.

- Instat. (2012). *Instituti i Statistikave*. Retrieved from Albanian Institute of Statistics: www.instat.gov.al
- Institute, S. (2017). *SAS Institute*. Retrieved from SAS Institute. Analytics, Business Intelligence and Data Management: https://www.sas.com/en_us/home.html
- Johnson, S. (2011). Introducing SQL Server. In Mastering. In S. Johnson, *Introducing SQL Server*. In *Mastering* (pp. 291–323). Indianapolis: Wiley Publishing.
- Juris Breidaks, M. L. (2017, 03 21). *vardpoor*. *vardpoor: Variance Estimation for Sample Surveys by the Ultimate Cluster Method*. CRAN.
- Loo, M. V. (2012). The introduction and use of R software at Statistics Netherlands. *Third International Conference of Establishment*. Montreal: American Statistical Association.
- Lumley, T. (2016, 12 01). *survey*. *survey: Analysis of Complex Survey Samples*. CRAN.
- Matthias Templ, V. T. (2016). THE SOFTWARE ENVIRONMENT R FOR OFFICIAL STATISTICS AND SURVEY METHODOLOGY. *Austrian Journal of Statistics*, 45.1, 97. doi:10.17713/ajs.v45i1.100
- Meindl, B. (2017, 02 23). *sdcTable*. *sdcTable: Methods for Statistical Disclosure Control in Tabular Data*. CRAN.
- Microsoft. (2017). *Office365 Software*. Retrieved from Microsoft. Office365 Software, Version 2017.: www.office.com
- MortalitySmooth. (2015, 08 18). *MortalitySmooth: Smoothing and Forecasting Poisson Counts with P-Splines*. CRAN.
- Raco, E. (2016). *Statistikat e Emrave*. Retrieved from Statistikat e Emrave: <https://statemra.shinyapps.io/emra>
- Sabanovic, E. (2015, September 16). *UNECE*. Retrieved from United Nations Economic Commission for Europe: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2015/mtg1/WP_8_Bosnia_and_Herzegovina_Editing_and_Imputation_in_Household_Based_Surveys.pdf
- Smith, D. (2016). *Revolution Analytics*. Retrieved from revolutionanalytics: <http://blog.revolutionanalytics.com/2016/07/r-moves-up-to-5th-place-in-ieee-language-rankings.html>
- Studio, R. (2017). *Shiny*. Retrieved from Shiny Software, Version 1.0. 250: www.shiny.rstudio.com
- Wickham, H. (2016, 06 24). *dplyr*. *dplyr: A Grammar of Data Manipulation*. CRAN.
- Wickham, H. (2016, 09 13). *lubridate*. *lubridate: Make Dealing with Dates a Little Easier*. CRAN.
- Wickham, H. (2017, 02 18). *stringr*. *stringr: Simple, Consistent Wrappers for Common String Operations*. CRAN.
- Wickham, H. (2017, 01 10). *tidyr*. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*. CRAN.

An Empirical Analysis on the Long Run Relation between Unemployment and Higher Education in Turkey

Güngör Turan

Epoka University, Department of Economics, gturan@epoka.edu.al

Abstract

In this empirical paper, a long-run co-integration between higher education and unemployment in Turkey has been investigated. ARDL bounds model which is a long-run co-integration method has been used based on the number of unemployed and higher education graduates time series in Turkey in 1961-2012 period. The results of bounds test conclude that there is no long-run co-integration and evidence between higher education and unemployment in the reference period in Turkey. As a result, this study supports to some extent the current debate on the availability of "non-qualified" higher education which does not generate adequate link between higher education and labor market to employ particularly for higher education graduates in Turkey.

Keyword: Unemployment, higher education, co-integration, bound test, Turkey

JEL Classification: I25; J60

Introduction

In economics, the long-run relationship between education and unemployment are examined within the context of neo-classical economic growth theory (Solow, 1956; Swan, 1956). Since the late 1970s, economists have started to give special attention to education through the accumulation of human capital investments in long-run economic growth (Ashenfelter & Ham, 1979; Mincer, 1991; Barro, 2001; Mankiw *et al.*, 1992; Jorgenson & Fraumeni, 1992; Hanushek & Kimko, 2000). Unlike neo-classical growth models, recently, along with the emergence of endogenous growth model that accepts as an inherent factor of human capital formation and accumulation in economic growth, the relationship between education and economic growth began to discuss again in economic growth theory and literature (Romer, 1986; Lucas, 1988; Aghion & Howitt, 1992).

Dissemination of high education by rising skill level employed of increasing graduates may lead to higher growth and lower unemployment. Higher education institutions requires human capital which will produce higher education programs. The increase in the number of programs can lead to a permanent change in the human capital stock. Supply side of labor market and differences in human capital demand are important in the determination of regional human capital stock. Higher education institutions by giving local graduates and managing research activities help to raise the level of a region's human capital (Hunter, 2013). Basic training of primary level for the production of goods and services may be sufficient, while secondary-level education in the workplace employees enables you use your technology. Higher education requires a level that will lead to new technology and inventions (Keller, 2006). College graduates generate spillover effects that facilitates the flow of knowledge from universities to company's right and contains the accumulation of human capital formation and accumulation. One of the important mechanisms that facilitate the spread of knowledge is the mobility of human capital which leads to university graduates transition from universities to companies (Audretsch *et al.*, 2005).

Research-oriented higher education institutions facilitates the spread of knowledge in the local economy. At the same time, research-intensive fields tend to have wider human capital inventory. The spreading effects of academic research and development studies to the benefits of local business depends on the economic environment and infrastructure support of a region. Increasing demand for skillful workforce by the effect of academic research and development efforts, rather than the expansion in the supply of local graduates, has a large scale of causal effect over the local human capital levels. Higher education institutions are vital functions to local economic development. The development and enrichment of local higher education institutions can trigger the spillover effect in local economies (Abel & Deitz, 2011). Thus, through accumulation and formation of human capital, long term sustainable growth with low unemployment can be accomplished together.

In literature, empirical studies are much more concentrated on the long-run relationship between levels of education and unemployment. In these studies, it has been determined that there is an inverse relationship between levels of education and unemployment. In other words, while levels of education rises, unemployment decreases in the long-run (Mincer, 1991; Wolbers, 2000; Garrouste *et al.*, 2010). Researchs over the expansion of higher education policy which are assessing the effects of the policy of enlargement of higher education among higher education graduates in the European labor markets in the 1990s are examined increasing unemployment among higher education graduates, and particularly further growth of the problem of unemployment among youth university graduates (Schomburg, 2000; Mora *et al.*, 2000; Woodley & Brennan, 2000). Plümper and Schneider (2007), founded that the rise in the unemployment rate leads to an increase in college enrollment, at the same time, however, fell significantly in expenditures per student in Germany. Nunez and Livanos (2010), in their paper on the impact of unemployment on higher education at the national level, concluded that the effects of higher education between the EU 15 countries exhibit different behaviors. Accordingly, the countries which have powerful influence of higher education on employment are Finland, Belgium and UK. On the other hand, South European countries higher education graduates Italy, Greece and Portugal as well as, are faced with the problems of unemployment because of insufficient employment creation in the labor markets. More importantly, similar negative results in the labour market are valid for France, Luxemburg, Germany and Sweden which have accepted quality and reputation in higher education in the international arena. Erdem and Tuğcu (2012), investigated co-integration and causality relationship between higher education and unemployment in Turkey, and they found a statistically significant relationship between these two variables. According to the results obtained, higher education graduates are one of the factors affecting the increase in unemployment in the long-run in Turkey. Unlike, Dongshu *et al.* (2016), investigated the effects of higher education expansion policy on unemployment, and concluded that higher education expansion policy reduced unemployment among college graduates in China.

The purpose of this article is to contribute to insufficient number of empirical studies that have been done so far over the subject of long-run relationship between higher education and unemployment in Turkey. The rest of the article composed of a brief look at the labor market and higher education in Turkey, specification of the data, building of empirical method and methodology, empirical analysis, and conclusion after evaluation of the test results.

Turkish Labor Market and Higher Education: An Overview

Turkey's economy, in the period of 1962 and 1977, under the leadership of economy policies based on planning, realized a stable and high growth process. With the support of the period of high growth in the world economy, it has been provided annual growth rates up to 10% in the manufacturing industry, and 6% for all of the economy in the 1960s. Until domestic market-

oriented through import substitution industrialization process collapsed, between the years 1962 and 1977, there was no contradiction in the economy. It had been the fastest increase of the period of employment in manufacturing. But, this growth period which was based on new investments and an increase in the total factors rather than production and productivity increases using available facilities and factories, had been ended with great depression in the economy reflected higher price increases, bottle necks in manufactory production and difficulties in foreign payments in 1977 the first signs began to emerge. This economic depression period by gaining political and social character over time has led to a drastic change of the model in Turkish economy and industrialization process after 1980. The main feature of the new period starting with the 1980s, through domestic market-oriented import substitution industrialization strategy completely abandoned, export-driven economic growth model has been introduced (Turan, 2015).

Following 1980, the biggest impact of these economic transformation, industrialization and growth model executed with outward-oriented neo-liberal economic policies has been on the labor markets. The economic and financial crisis in this process are deeply influenced the labour markets and thus the mass public. The failure in the implementation of national economic policies coupled with global economic volatility and instability, Turkey has survived the most severe successive economic crisis in its economic history in the period of 1994 and 2001. Although Turkey's economy recovering entered into a new wave of growth since 2002, rises in employment has been limited by increases in productivity in industry. For this reason, while production increases, employment levels remained low in the manufacturing industry, and the vast majority of the employment growth has been performed in the services sector. Erratic growth rates did not prevent higher non-farm unemployment which it has already high (Turan, 2015; İsmihan & Kırılıcı, 2009).

In the 1990s, although a marked decrease observed in the population growth rate which was a thousand of 19.9, due to the fact that young population structure, working age population has continuously increased. Especially in the period covering the years 2002-2007, relatively high growth has been reached, but job creation capacity of Turkey's economy has remained at limited level, and the employment rate has not exceed low level of 46 percent so far. Thus, unemployment rate has continued its high level which has over the level of 10 percent following 2002 period to the present. Non-agricultural unemployment rate is over 14 percent. Because of difficulties of job finding in the labor market by the effects of discourage of job finding expectations and extension of job search duration, the number of people out of the labor market has constant increased in the mass of over 3 million registered unemployed as of January 2015 (Turkstat, 2016). Therefore, in fact, the number of unemployed has reached larger dimensions. This number is over 6 million when combined with the the number of registered unemployed. Therefore, prolonged unemployment and being discouraged worker has also led to a deepening and mass poverty in Turkey (Turan, 2015).

If we look at the developments in higher education brief, prior to the last university reform in 1981 while there are a limited number of universities and graduates, the numbers and graduates have increased in the new era under Turkish Higher Education Council. A remarkable point in the restructuring of Turkish higher education system after 1981, private universities has began to be established also. Since the 1990s, it has been observed dramatic increases in the number of universities and graduates by the impact of new universities which were spreading to the whole country with the addition of a large number of private universities. While the total number of faculties and colleges were 55 and the number of graduates were 6,025 in the education year of 1960/1961, total number of them respectively were 1914 and 573,434 in 2011/2012 education year (Turkstat, 2015). Although saving important quantitative developments, the content, structure and quality of Turkish higher education system has been

continuously the subject of debate, and universities have exhibited ups and downs in higher education (Turan, 2016; Şen, 2012; Balaban, 2012; Sargin, 2007).

Empirical Model and Methodology

In this study, the long-run relationship between higher education and unemployment has been tested by using a log-linear function which is formulated as follows:

$$\ln U_t = \alpha + \beta \ln HE_t + \epsilon_t \quad (1)$$

Here, HE_t represents the number of graduates in higher education, α fixed term, U_t the number of unemployed, and ϵ error term.

To determine the long-run relationship between the higher education and unemployment it has been employed the long-term co-integration bound test which is known as Autoregressive Distributed Lag (ARDL). The bounds test which was developed by Pesaran *et al.* (2001) has more advantageous according to the traditional Engle and Granger (1987), and Johansen and Juselius (1990) co-integration tests. In traditional co-integration tests assumed that all variables are integrated at I(1) level. Whereas ARDL bounds test can be employable irrespective of whether the variables are integrated at I(0), I(1) or mutually co-integrated.

The following estimation of regression equation is done in bound test:

$$\Delta \ln U_t = \alpha_0 + \sum_{i=1}^p \alpha_{1i} \Delta \ln U_{t-1} + \sum_{i=0}^p \alpha_{2i} \Delta \ln HE_{t-1} + \theta_1 \ln U_{t-1} + \theta_2 \ln HE_{t-1} + u_t \quad (2)$$

Where; Δ is the difference operator, p is the lag length, and u is the serially uncorrelated error term. The ARDL test is performed in two stages: In the first, the null hypothesis of no-cointegration long-term relationship between the variables described as $H_0: \theta_1 = \theta_2 = 0$ is tested against $H_1: \theta_1 \neq 0, \theta_2 \neq 0$. F-statistic is used to test the relationship of long-term co-integration. Since the asymptotic distribution of this F-statistic is non-standard irrespective of whether the variables are I (0) or I (1), two tables of critical values are developed by Pesaran *et al.* (2001). One assumes that all variables are I (0) and the other that all variables are I (1). In this case, it contains a bound covering all possible classification of the variables. If the calculated F-statistic lies above the upper level of the bound, the H_0 is rejected supporting the existence of co-integration relationship in the long-run. If the calculated F-statistic lies below the lower level of the bounds, the H_0 cannot be rejected, and does not support the relationship of co-integration. If the calculated F-statistic falls between the bounds, then the result is inadequate and in this case, the error correction term which is known Error Correction Model (ECM) is used to determine the existence of co-integration. If obtained ECM is negative and significant, the variables are accepted to be co-integrated in the long-run.

After the determination of a long-run relationship, the next phase of the ARDL test which is ECM is formulated as follows:

$$\Delta \ln U_t = \alpha + \sum_{i=1}^p \omega_k \Delta \ln U_{t-1} + \sum_{i=0}^p \lambda_k \Delta \ln HE_{t-1} + \omega EC_{t-1} + u_t \quad (3)$$

Where; ω is the error correction parameter, and EC gives the residual.

Since long-run co-integration relationship between variables breaks the stability of the parameters, whether testing variables should be stable over time. Tests for parameter stability which are developed by Brown *et al.* (1975) are cumulative sum (CUSUM) and cumulative sum of squares (CUSUMSQ) tests which are widely used in ARDL modelling framework. These are based on the recursive regression residuals, and they have been updated against

structural breaks in the model. The existence of a co-integration relationship between variables supports at least the existence of a one-way causality relationship. As a result of ARDL test supporting the existence of co-integration relationship, the causality relationship should be tested between variables. To do this, modified Wald (MWALD) test which is developed by Donaldo and Lütkepohl (1996) is recommended in the literature.

Empirical Analysis and Findings

The first step of the ARDL procedure is to test whether all variables are stationary or not. In other words, to test irrespective of whether the variables are integrated at I (0), I (1) or mutually co-integrated. For this application, the ADF and PP unit root tests are recommended in the literature. The results of the unit root tests are given in Table 1 below. Applying the unit root tests to the first-differences of each series leads to a very clear rejection of the hypothesis that the data are I (2), which is important for the legitimate application of the bounds test below.

Table 1

Results for unit root tests

Level	Variables	Test		Type
		ADF		PP
Constant	lnHE	-0.8014		-0.8029
	lnU	-1.1720		-1.1724
Constant+Trend	lnHE	-4.1803*		-3.0512
	lnU	-1.2679		-1.4207
First Difference				
Constant	lnHE	-4.5806		-7.6235
	lnU	-5.8010		-5.7831
Constant +Trend	lnHE	-4.5504		-7.5855
	lnU	-5.7762		-5.7533

*stationary at level.

The next step in the bounds testing approach to co-integration is perform the F-test on selected ARDL model including appropriate lag lengths. The optimal lag length is imposed as 5 using Vector Auto Regressive (VAR). Applying committed regression model considering the number of lag length, F-statistic value is 3.7059. This result is compared to the bounds test critical values table lower and upper bounds values which are developed by Pesaran *et al.* (2001), and founded that the calculated F-statistic lies below the lower level of the bound, the H_0 cannot be rejected, and does not support the long-run relationship of co-integration at 1%, 5% and 10% levels as reported in Panel A at Table II. As a result, there is no long-run co-integration relationship between higher education and unemployment in Turkey. On the other hand, hence the bounds test results do not support the long-run relationship between higher education and unemployment it is not possible to make a statistical inference and analysis over the short-run dynamics of the model. At the same time, because of no long-run evidence between higher education and unemployment, it is impeded the direction or directions of the relationship of causality.

Table 2

Results for co-integration analysis

Panel A: Co-integration tests	Dependent variable:lnU
-------------------------------	------------------------

F-statistic	3.7059
Error-correction parameter	-0.8729 [0.0569]
Panel B: Long-run parameters	
Constant	-0.3070 [0.4983]
lnHE	0.0552 [0.9051]
Panel C: Diagnostic checking	
Adjusted-R ²	-0.2242
Serial correlation: Breusch-Godfrey LM test statistic	0.9880 [0.4409]
Heteroscedasticity: White test statistic	0.5299 [0.8233]
Functional form: Ramsey's Reset test statistic for regression specification error	0.9921 [0.3260]
Panel D: Stability tests	
CUSUM	stable
CUSUMQ	stable

Note that F statistic critical values are generated from Pesaran *et al.* (2001), p.300, Table CI (iii), Case III at level 10% (4.04-4.78), at 5% level (4.95-5.73), and at 1% level (6.84-7.84). Numbers in brackets are p-values.

As reported

in Panel A at Table 2, both the result of F-statistic and error-correction parameter do not support the long-run co-integration between higher education and unemployment in Turkey. According to the Panel B which gives long-run parameters, the independent variable lnHE is not statistically significant. Since the ARDL method uses the Ordinary Least Squares (OLS) to estimate the co-integration vector, it should be checked that the assumptions of the OLS estimator are not violated. To do this, the diagnostic checking must be done. The Panel C gives the diagnostic checking tests results. These results illustrated that the estimated ARDL model ensures the assumptions of no-serial correlation, homoscedasticity, and no-functional misspecification. Panel D gives the results of stability tests. The stability of long-run coefficients was tested by applying CUSUM and CUSUMSQ tests. Both the results of stability tests illustrated that the estimated ARDL model provides stable parameters in the long-run.

Conclusion

In this empirical paper, the long-run con-integration between higher education and unemployment in Turkey has been investigated. ARDL bound test which is a long-run co-integration test has been used based on the number of higher education graduates and unemployed time series in Turkey in 1961-2012 period. The results of bounds test conclude that there is no evidence of a long-run relationship between higher education and unemployment in the reference period in Turkey. In other words, higher education and unemployment are not moving together both directly and inversely in the long-run in Turkey. The results of this study supports to some extent the current debate on the availability of "non-qualified" higher education which does not generate enough link between higher education and labor market to employ particularly for higher education graduates in Turkey. On the other hand, hence the bounds test results do not support the long-run relationship between higher education and unemployment, it is not possible to make a statistical inference and analysis over the short-run dynamics of the model. At the same time, because of no long-run evidence between higher education and unemployment, it is impeded the direction or directions of the relationship of causality. Because of no long-run co-integration, and both inverse and statistically significant evidence between higher education and unemployment in Turkey, it can

be said that higher education does not effective to combat with unemployment problem, and it does not give enough and sustainable support to decrease unemployment in the long-run in Turkey. It can be observed that unemployment depends on labor and total factor productivity levels which are led by changings in working hours and capacity utility rates particularly in manufacturing industry in the short-run. Hence, the long-run unemployment maintains at its high levels in Turkey. From here, it can be determined that the contribution of higher educated human capital to the long-run economic growth is inadequate and limited compared to physical capital and primary and secondary educated human capital in Turkey.

References

- Abel, J. R. and Deitz, R., 2011. The Role of Colleges and Universities in Building Local Human Capital. *Current Issues in Economics and Finance*, 17, pp. 1–7.
- Aghion, P. and Howitt, P., 1992. A Model of Growth Through Creative Destruction. *Econometrica*, 60, pp. 323-351.
- Barro, R. J., 2001. Human Capital and Growth. *The American Economic Review*, 91, pp. 12-17.
- Brown, R. L., Durbin, J. and Evans, J. M., 1975. Techniques for Testing the Constancy of Regression Relationship over Time. *Journal of the Royal Statistical Society*, 37, pp. 149-192.
- Dolado, J. J. and Lutkepohl, H., 1996. Making Wald Tests Work for Cointegrated VAR Systems. *Econometric Reviews*, 15, pp. 369-386.
- Dongshu, O. and Zhong, Z., 2016. *Higher Education Expansion and Labor Market Outcomes for Young College Graduates*. IZA Discussion Paper, No. 9643.
- Engle, R. F. and Granger, C. W. J., 1987. Co-integration and Error-Correction: Representation, Estimation, and Testing. *Econometrica*, 55, pp. 251-276.
- Erdem, E. and Tugcu, C. T., 2012. Higher Education and Unemployment: a cointegration and causality analysis of the case of Turkey. *European Journal of Education*, Vol. 47, No. 2, pp. 299-309.
- Garrouste, C., Kozovska, K. and Perez, E. A., 2010). *Education and Long-Term Unemployment*, Paper prepared for the third edition of the workshop “Geographical Localisation, Intersectoral Reallocation of Labour and Unemployment Differentials” (GLUNLAB3), RCEF, European Union.
- Hanushek, E. A. and Kimko, D. D., 2000. Schooling, Labor-Force Quality, and The Growth of Nations, *The American Economic Review*, 90, pp. 1184-1208.
- Johansen, S. and Juselius, K., 1990. Maximum Likelihood Estimation and Inference on Co-Integration—with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, 52, pp. 169-210.
- Keller, Katarina R. I., 2006. Investment in Primary, Secondary, and Higher Education and The Effects on Economic Growth. *Contemporary Economic Policy*, 24, pp. 18–34.
- Mankiw, N.G., Romer, D. and Weil, D.N., 1992. A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107, pp. 407-437.
- Mora, J.G., Jose, G. M. and Adela, G. A., 2000. Higher Education and Graduate in Spain. *European Journal of Education*, 35, pp. 229–237.

- Ismihan, M. and Kivılcım M. Ö., 2009, Productivity and Growth in an Unstable Emerging Market Economy: The Case of Turkey, 1960-2004. *Emerging Markets Finance and Trade*, 45, pp. 4–18.
- Nunez, I. and Livanos, I., 2010. Higher Education and Unemployment in Europe : An Analysis of The Academic Subject and National Effects. *Higher Education*, 59, pp. 475-487.
- Romer, M. P., 1986. Increasing Returns to Long-Run Growth. *Journal of Political Economy*, 94, pp. 1002-1037.
- Pesaran, M. H., Shin, Y. and Smith, R. J., 2001. Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics*, 16, pp. 289-326.
- Plumper, T. and Schneider, C. J., 2007. Too Much to Die, Too Little to Live: Unemployment, Higher Education Policies and University Budgets in Germany. *Journal of European Public Policy*, 14, pp. 631–653.
- Sargın, S., 2007. Türkiye’de Üniversitelerin Gelişim Süreci ve Bölgesel Dağılımı [The Development Process of Universities and Theirs Regional Distribution in Turkey], *Süleyman Demirel Üniverstiy Social Sciences Institute Review*, 3, pp. 133-150.
- Schomburg, H., 2000. Higher Education and Graduate Employment in Germany. *European Journal of Education*, 35, pp. 189-200.
- Solow, R. M., 1956. A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, 70, pp. 65-94.
- Şen, Z., 2012. Türkiye’de Yükseköğretim Sistemi Eleştirileri ve Öneriler [The Critics and Sugestions of Turkish Higher Education System]. *Yükseköğretim Review*, 2, pp. 1-9.
- Şenses, F., 1994. Labor Market Response to Structural Adjustment and Institutional Pressures. *METU Studies in Development*, 21, pp. 405-448.
- Turan, G., 2016. Türkiye’de Yüksek Öğretim ve Ekonomik Büyüme [Higher Education and Economic Growth in Turkey]. *Çimento İşveren*, 30, pp. 8-17.
- Turan, G., 2015. Türkiye’de Büyüme ve İşsizlik [Growth and Unemployment in Turkey]. *Çimento İşveren*, 29, pp. 10-17.
- TURKSTAT, 2015. *Statistical Indicators 1923-2013*.
- TURKSTAT, 2016. *Statistical Indicators 1923-2013*.
- Wolbers, M., 2000. The Effects of Level of Education on Mobility between Employment and Unemployment in the Netherlands. *European Sociological Review*. 16, pp. 185-200
- Woodley, A. and Brennan, J., 2000. Higher Education and Graduate Employment in the United Kingdom. *European Journal of Education*, 35, pp. 239–249.

Application of Gravity model: the Albanian Agricultural Export

Kushtrim Braha¹, Ema Lazorčáková¹, Miroslava Rajčániová¹, Artan Qineti¹, Andrej Cupák²

¹Faculty of Economics and Management, Department of Economic Policy, Slovak University of Agriculture in Nitra, Slovakia

²National Bank of Slovakia, Bratislava, Slovakia

Abstract

Despite its huge agricultural endowments and productive potential, Albania has a sharp trade deficit with agricultural commodities. The main objective of this study is to analyse key determinants of Albanian agricultural export. Here we employ baseline gravity model considering conventional gravity variables for Albanian export flows for the time period 1996-2013. The Poisson Pseudo-Maximum Likelihood (PPML) regression is used for stepwise estimations of the augmented gravity model, including effects of Albanian Diaspora, exchange rate and price stability, trade liberalization and institutional distance. In the last section, we estimate agricultural export potential with the main trading partners, revealing the absolute difference between actual and predicted agricultural export. Main findings of this study suggest that agricultural export flow increases with increasing economic size, revealing higher impact of importer's absorbing potential comparatively to Albania's productive potential. On the other hand, growth in domestic demand, resulting from increase in population, leads to reduction of agricultural export. Moreover, agricultural export flows are determined by low transportation costs (distance), adjacency proximity (sharing common border) and linguistic similarities. Influence of Albanian Diaspora residing in the importing partner countries is found to have robust effect on the promotion of agricultural export. Results of this study reveal that exchange rate variability has a positive impact, while bilateral institutional distance has diminishing effects on Albanian agricultural exports. Findings of this study are important for trade and agricultural policy makers. From the trade policy perspective, one should assume that the platform of agricultural export promotion should aim market diversification in those countries (other than neighbouring countries) in which Albanian farmers can exploit their comparative advantage. On the other hand, from the agricultural policy perspective, special attention should be paid to measures that lead to improvement of the competitiveness of local farmers.

Key words: Agricultural trade, export, gravity model, panel data, Albania

JEL classification: Q17, Q18, F14

Introduction

Albania initiated transition into a market economy since the early 1990s. Transition from communism into free market system was unique and escorted with dramatic turbulences. Early period of market reforms endorsed radical model of the shock therapy, guiding Albania's economic system to drastic and profound structural changes. Price controls were lifted, markets were liberalized and privatization process initiated (McCarthy et al., 2009). Initial reforms, between 1993 and 1996, resulted with outstanding economic growth, marking highest growth rates compared to all transition economies. However, in 1997, flourishing financial pyramid schemes ruined both political and economic system. The country witnessed collapse of pyramid investment schemes, which were larger (relative to the size of the economy) than any previous schemes of this kind (Korovilas, 1999). Hence, Albania plunged into deep economic crisis. Rioting and civil unrest brought the country in the edge of civil war. Events from that period served to Albania as hardship lesson of market and institutional failure. Since then, fast and systematic recovery took place. Sustained economic growth of 2000s, among other factors, is

a merit of integration into international markets. Improvement of trade links and injection of foreign investments into domestic economy fuelled development perspective of Albania.

Albania is an agricultural economy. Agriculture employs more than a half of the population and accounts about a quarter of output (Zahariadis, 2007; EC, 2014). Hence, it has a huge potential to become engine of economic growth and competitiveness in international markets (USAID, 2012). Despite its indisputable potential, agricultural sector in Albania faces significant challenges. Predominant constraints of agriculture include small and fragmented farms (average farm size of 1.2 ha), migration from rural areas, underdeveloped irrigation system, low labour productivity, and limited technological level (USAID, 2012; EC, 2014). Interest for investment in agricultural sector remains low as well. Additional agricultural constraints are derived from the complex land reform (see Cungu and Swinnen, 1999; Deininger et al., 2012; Qineti et al., 2015). Majority of the small farms in Albania are subsistent and agricultural production serves to home consumption. Empirical studies (i.e. Mc Carthy et al., 2009) suggest that the farm households cultivating staple crops achieve to market only 4 to 8 percent of their production. The rest is used for self-consumption.

Studies utilizing aggregate trade flows in Albania (see Xhepa and Agolli, 2004; Asllani, 2013; Fetahu, 2014; Sejдини and Kraja, 2014) report unexploited trade potential. They suggest that main constraints of Albanian foreign trade rest on the limitations of domestic supply. Trade flows are determined by trade links with neighbouring countries, low transportation costs and cultural links. Moreover, they put emphasis on non-tariff trade barriers such as market access, border procedures, free movement, development and dissemination of information.

Albania has adopted a liberal trade regime since the very beginning of its economic transition. It was among the first steps of transition reforms. The process of trade liberalization has been intensified particularly after the accession of Albania in WTO in the year 2000 (Government of Albania, 2015). Membership in WTO induced deep reforms in legislation and trade policies in compliance with WTO guiding principles. The main objectives of Albania's trade policy are coherent with WTO principles and therefore guarantee the absence of quantitative restrictions on imports and exports, export subsidies, any kind of tax on exports and export bans (WTO, 2016). Further steps of trade liberalization followed Albania's involvement in the regional integration through a network of bilateral Free Trade Agreements (FTAs) with its regional countries. Later on, bulk of bilateral FTAs melted into the creation of Regional Trade Agreement (RTA), known as renewed Central Europe Free Trade Agreement (CEFTA 2006). This RTA incorporated group of countries from Southeast Europe (Albania, Bosnia and Herzegovina, Croatia, Kosovo, Macedonia, Montenegro, Moldova and Serbia) and entered in force in 2007. The map of liberalized trade agreements is further extended with the signature of FTA with Turkey in 2008. In 2008, Albania signed another FTA with European Free Trade Agreement Association (EFTA) countries (Norway, Switzerland, Iceland and Lichtenstein). FTA with EFTA countries entered in force in 2011. Most importantly, since 2009, Albania is implementing the Stabilization and Association Agreement (SAA) with the European Union (EU). Meanwhile the free trade agreement, which is integral part of SAA, is in force since 2006. However, early roots of trade liberalization with the EU date from 1999. Since then, Albania benefited from Autonomous Trade Preferences with the EU, granting duty-free access to EU market for nearly all products from Albania (excluding only wine, sugar, certain beef products and certain fisheries products, which enter the EU under preferential tariff quotas, as negotiated under the SAA). Summing up, Albania's trade is operating in free trade regime with EU, EFTA, Turkey, and its neighbouring CEFTA 2006 countries.

The main objective of this paper is to explain main determinants of agricultural export in Albania. The paper is organized as follows: the next section provides retrospective of previous

studies employing gravity model in agricultural trade. The following section describes methodology, estimation strategy as well as variables and data used in empirical estimation. Then we present and discuss results of the estimation in the subsequent section. Lastly we summarize and draw conclusions.

To our knowledge, this paper is first attempt that employs gravity model in determining key aspects of agricultural export in the case of Albania. This study estimates implications of conventional gravity variables including wide range of other factors, such as border effects, cultural links, migration, price instability and exchange rate variability, free trade agreements, quality of institutions, on the potential of agricultural export in Albania.

Retrospective of previous studies

Gravity model has been used in agricultural trade analysis as a baseline model for estimating the effect of a variety of policy issues. Country level analysis utilizing gravity model in agricultural trade analysis are scarce. Thus, Ševela (2002) applied gravity model to explain Czech agricultural export. Except of conventional variables, the study observes effects of import tariff for agricultural products, exchange rate and membership in EU and EFTA. Results of the study are consistent with theoretical framework of the gravity model.

Previous studies analyse many different trade determinants. Studies dealing with the effects of trade liberalization (FTAs, RTAs and Preferential Trade Agreements) suggest that these instruments serve as an attractive platform to promote agricultural trade. Typically, positive effects of trade liberalization are translated in elimination of trade restrictions and facilitating integration through liberalization of non-tariff barriers. With some exception, majority of the previous studies suggest net trade creating effects (Jayasinghe and Sarker, 2008; Grant and Lambert, 2008; Korinek and Melatos, 2009; Sun and Reed, 2010; Koo et al., 2006). Pishbahar and Huchet-Bourdon (2008) employ extended gravity model to estimate the impact of eleven RTAs on European agricultural imports. Their findings suggest that majority of European Union RTAs supports agricultural exports of developing countries to the EU market. On the other hand, two most important and unilateral (Generalized System of Preference and the agreement with Mexico) have negative effect on agricultural exports.

Studies dealing with effects of immigration links on trade date since the early 1990s. As Gould (1994) stresses out immigrant links have potential to decrease transaction costs resulting from knowledge of home-country markets, language, preferences, and personal contacts (see for example Genc et al., 2012; Head and Ries, 1998; Raulch and Trindade, 2002; Peri and Requena-Silvente, 2010). On the other hand, Parsons (2005) is interested in the effects of the stock of immigrants from the EU expansion countries residing within each EU-15 country. The results indicate that Eastern European immigrants exert a positive influence on both EU-15 imports and exports. It is predicted that a 10% rise in Eastern European immigration will increase EU-15 imports from these countries by 1.4% and EU-15 exports by 1.2%.

Effects of exchange rate volatility are frequently incorporated in analyses of price competitiveness in international markets (for example Maitah et al., 2016) but also in gravity models dealing with agricultural trade. Thus, Cho et al. (2002) employ panel data to estimate gravity models for ten developed country. They found out that real exchange rate uncertainty has had negative effect on agricultural trade. Moreover, the negative impact of uncertainty on agricultural trade has been more significant compared to other sectors. Extension to this study can be found in Kandilov (2008) and studies for specific countries include the work of Fertő and Fogarasi (2011), Sheldon et al. (2013), Kafle and Kennedy (2012), Koo et al. (1994), Frankel and Wei (1998).

Institutional effects on agricultural trade have received a great attention recently. Levchenko (2004) investigates quality of institutions (quality of contract enforcement and property rights). His paper studies consequences of trade when institutional differences are the source of comparative advantage among countries. Findings of the study imply that institutional differences are important determinant of trade flows. Moreover, results of the paper suggest that institutional differences diverge less developed countries to gains from trade. Similarly, Linders et al. (2005) found that institutional distance has a negative effect on bilateral trade, presumably because the transaction costs of trade between partners from dissimilar institutional settings are high. They stress out that institutional quality of both the importer and exporter increases the amount of bilateral trade.

Methodology and materials

Gravity model specification

Gravity model has become a workhorse (Eichengreen and Irwin, 1998) in international trade analysis. Bulk of empirical studies rank the gravity model among the most accurate tools in explaining and predicting bilateral trade. Conventional theory of gravity model in international trade emerged in the early 1960s with the pioneering studies of Tinbergen (1962) and Pöyhönen (1963). Later on, empirical works utilizing gravity model were initiated by Linnemann (1966). Since then, evolution of the gravity model and diversity of its application was remarkable. Theoretical framework of the gravity model is borrowed from the gravity law of physics. Isaac Newton's gravity model assumes that attraction between two heavily bodies is proportional to the product of their masses and inversely related to the distance between them (Frankel, 1997). Translated into the international trade theory, gravity model suggests that volume of trade between two countries is proportional to their economic size (national incomes) and inversely related to the distance. Therefore, gravity model predicts that economically rich and geographically close countries trade more together than with third countries (Pokrivčák and Šindlerová, 2011). Main advantages of the gravity model lay on results of empirical work. Linders and De Groot (2006) suggest that gravity model is particularly efficient in explaining a large portion of the variation in bilateral trade.

For the last fifty years, gravity equations have dominated empirical studies in international trade. In its basic form, the amount of trade between countries is assumed to be increasing in their sizes, as measured by their national incomes, and decreasing in the cost of transportation between them (Cheng and Wall, 2005). Therefore, the basic form of the gravity equation is expressed as follows:

$$T_{ij} = \beta_0 \frac{GDP_i^{\beta_1} GDP_j^{\beta_2}}{DIST_{ij}^{\beta_3}} \quad (1)$$

where T_{ij} is bilateral trade between country i and j ; GDP_i (GDP_j) is economic size of country i (j) measured by GDP; $DIST_{ij}$ is bilateral distance between the two countries; β_0 is a constant, β_1, β_2 and β_3 are parameters often estimated in a log-linear reformulation of the model.

For the purpose of this study, we employ modified gravity model used by McCallum (1995). It is adjusted for logarithmic form and allows adding supplementary variables:

$$\ln X_{ij} = \beta_0 + \beta_1 \ln GDP_i + \beta_2 \ln GDP_j + \beta_3 \ln DIST_{ij} + \beta_4 \delta_{ij} + \varepsilon_{ij} \quad (2)$$

where X_{ij} is trade flow from country i to country j (in our case export), GDP_i and GDP_j is GDP of the country i and country j , $DIST_{ij}$ is distance between country i and j , δ_{ij} is dummy variable for the other factors influencing trade flows, and ε_{ij} is error term.

We adopted the above equation to fit it to the gravity model for agricultural exports in Albania. Further we adjusted the basic form of the gravity model equation (baseline model is called Model 1 in the Results section) for agricultural exports of Albania as follows:

$$\ln X_{ij} = \beta_0 + \beta_1 \ln GDP_i + \beta_2 \ln GDP_j + \beta_3 \ln GDP_{pc_i} + \beta_4 \ln POP_j + \beta_5 \ln POP_i + \beta_5 \ln DIST_{ij} + \varepsilon_{ij} \quad (3)$$

where X_{ij} is the value of agricultural exports from country i (Albania) to country j (importer). GDP_i and GDP_j stand for real GDP of country i and j , and measure economic size of the two economies. POP_i and POP_j are market size variables indicating population of the country i and j . $DIST_{ij}$ represents distance between country i and j . ε_{ij} is a stochastic disturbance term that is assumed to be well-behaved.

In order to estimate key determinants of agricultural export, we follow a stepwise procedure. First, we estimate the baseline gravity model to determine the coefficients of Albania's agricultural export flows (hereinafter Model 1). Subsequently, we augment the baseline model with dummy variables controlling for the income effects (Model 2), effects of adjacency, linguistic similarities and cultural links (Model 3), effects Albanian Diaspora (Model 4), effects of bilateral exchange rate and price stability of the importing country (Model 5), effects of trade liberalization with CEFTA, EU, EFTA and Turkey (Model 6), and institutional effects (Model 7). Finally, we estimate pooled effects of all variables included in the model (Model 8). For this purpose, the baseline model is modified with supplementary variables, as follows:

$$\begin{aligned} \ln X_{ij} = & \beta_0 + \beta_1 \ln GDP_i + \beta_2 \ln GDP_j + \beta_3 \ln POP_i + \beta_4 \ln POP_j + \beta_5 \ln DIST_{ij} + \beta_6 GDP_{pc_{ij}} \\ & + \beta_7 ADJ_{ij} + \beta_8 LAND_j + \beta_9 LANG_{ij} + \beta_{10} COL_{ij} + \beta_{11} \ln DIA_{ij} + \beta_{12} \ln EXR_{ij} + \beta_{13} INF_j \\ & + \beta_{14} CEFTA_{ij} + \beta_{15} SAAeu_{ij} + \beta_{16} EFTA_{ij} + \beta_{17} FTAtur_{ij} + \beta_{18} INSTdist_{ij} + \varepsilon_{ij} \end{aligned} \quad (4)$$

where $GDP_{pc_{ij}}$ is income effect variable indicating income differential between Albania and importer. The next two variables determine transportation costs. ADJ_{ij} is a dummy indicating if country i and j share common land border. $LAND_j$ dummy shows whether importing country j is landlocked. Variables aiming to capture cultural and historical similarities, respectively transaction and information costs follow. $LANG_{ij}$ shows whether country i and j has a common primary language. COL_{ij} indicates whether importer was Albania's colonizer. DIA_{ij} is stock of Albanian Diaspora in partner countries. EXR_{ij} is real exchange rate variable measured by the units of the importing country's home currency per Albanian Lek (ALL) and INF_j represents inflation rate (annual CPI rate) in the importing country. $CEFTA_{ij}$, $SAAeu_{ij}$, $EFTA_{ij}$ and $FTAtur_{ij}$ stands for free trade agreements with CEFTA, European Union, EFTA and Turkey. $INSTdist_{ij}$ shows bilateral institutional distance between Albania and import partner (see Linders et al., 2005).

Model variables

The dependent variable used in this study is the volume of Albanian agricultural exports to its partner countries. In this paper, we utilize conventional income variables explaining bilateral trade flows. Exporter's GDP (Albania) explains country's productive potential, while GDP of importing partner reflects absorbing potential, respectively purchasing power (see Koo et al., 1994). Theoretical framework of the gravity model predicts positive relationship to trade for both variables. Population is another conventional variable injected in the model with the aim to explain relationship between market size and Albanian agricultural export flows. There is no *a priori* relationship between exports and the populations of either the exporting or importing country (Martinez-Zarzoso and Nowak-Lehmann, 2003; Armstrong, 2007). An estimated coefficient of population of the exporter may have negative or positive sign depending on whether the country exports less when it is big (absorption capacity) or whether a big country exports more compared to a small country (economies of scale).

In order to investigate effects of transportation costs we embrace the variable of geographical distance between the capital city of Albania (Tirana) and capitals of importing countries. Increasing distance between trading partners proxies higher transport costs and decreases Albanian export flows. Therefore, gravity model predicts negative coefficient for this variable. Similarly, trade with landlocked countries involves higher trade costs, therefore negative coefficient is expected. On the other hand, lower transport and transaction costs are associated with neighbouring countries. Hence, we expect positive coefficient for the variable explaining exports with countries that share common border with Albania (see Anderson and Van Wincoop, 2001; Jansen and Piermartini, 2009).

Further, gravity equation is augmented with dummy variables predicting effects of cultural and historical similarities between Albania and importing countries. Here we impose dummy variables explaining whether Albania's trade partners were a former Albania's colonizer or if they share common primary language. These variables have been frequently used in the literature aiming to capture information costs. In particular, our interest is extended to the effects of Albanian migrants living in importing countries. Literature suggests that migrant ties can stimulate exports by lowering transaction costs and bringing their preferences for goods produced in home country. Hence, Albanian migrants might lower information and transaction costs through knowledge of home-country markets, language, business contracts etc. Therefore, empirical studies suggest that larger migrant stocks are associated with higher trade flows (see Gould, 1994; Bryant et al., 2004; Parsons, 2005).

The effects of trade liberalization are observed by incorporating dummy variables controlling for the impact of RTA with CEFTA 2006 countries (in force since 2007), SAA with EU (in force since 2009), FTA with EFTA (in force since 2011) and FTA with Turkey (in force since 2008).

Effects of exchange rate are frequently incorporated in gravity models dealing with agricultural trade (see Koo et al., 1994; Frankel and Wei, 1998; Hatab et al., 2010). In our case, annual exchange rate is determined by the Albania's currency units (ALL/Albanian Lek) per one unit of the importing country currency. We expect that an increase in exchange rate would devalue Albanian currency, hence exports would be cheaper. In such a case, devaluation of the domestic currency should increase Albanian agricultural export. Therefore, as the result we expect a coefficient with positive sign. Another factor influencing trade flows is price stability. In order to capture effects of price stability here, we incorporate in the model inflation rate (annual CPI rate) of the importing partner. Therefore, we expect a negative sign for the coefficient of inflation.

There is common agreement that institutional quality has substantially positive impact on bilateral trade flows (De Groot et al., 2004) and reducing the level of uncertainty (Jansen and Nordås, 2004). Therefore, if trade is supported by an effective rule of law, and if government regulation is transparent, countries engage in more trade (Linders et al., 2005). Following De Groot et al. (2004) we measure effects of bilateral institutional distance between Albania and its trading partners. Institutional distance between country pairs is measured as follows:

$$INSTdist_{ij} = \frac{1}{6} \sum_{k=1}^6 (I_{ki} - I_{kj})^2 / V_k \quad (5)$$

INSTdist is institutional distance, I_{ki} indicates country i score on World Governance Indicator's k th dimension and V_k is variance of this dimension across all countries.

In the last stage this paper, we estimate Albanian export potential by comparing actual and predicted export flows with individual trading partners.

Gravity model estimation technique

The choice of gravity equation estimator has been frequently debated among the scholars dealing with performance of the gravity model. Prevalence of heteroskedasticity and zero bilateral trade flows in the standard empirical methods were the focus of criticism (see Helpman et al., 2008; Westerlund and Wilhelmsson, 2009; Silva and Tenreyro, 2006). Hence, Silva and Tenreyro (2006) argue that standard empirical methods employed in estimating gravity equations are inconsistent and lead to biased results. They suggest that the use of standard log-linear estimator suffers from the presence of heteroscedasticity, which in turn might yield biased estimates of the true elasticities. On the other hand, various approaches have been employed in dealing with zero flows. Some authors suggest dropping the zero flows from sample (Linneman, 1966) or adding a constant to all trade flows to estimate log-linear equation (Rose, 2004).

Despite controversies and existence of wide range of estimation techniques such as Heckman model (Gomez-Herrera, 2013), FGLS (Martinez-Zarzoso, 2013), Helpman model (Helpman et al., 2008), Tobit model (Martin and Pham, 2008) etc. previous studies reveal that it is difficult to advocate a sole estimation technique as the best-performing. Choice of the method should be based on both economic and econometric considerations (Linders and De Groot, 2006) including robust specification checks and tests (Martinez-Zarzoso, 2013). For the purpose of this study, we adopted econometric approach using the Poisson Pseudo-Maximum Likelihood (PPML) estimator model, as proposed by Silva and Tenreyro (2006, 2011). PPML provides a natural way to deal with zero values and is robust to different patterns of heteroskedasticity. Even the critical voices (Martin and Pham, 2008) of PPML estimator suggest that in the case of small fraction of zero values, the PPML estimator model is the best performing method for the gravity model estimation. In this study the share of zero values is relatively low (18.6 percent), which indicates that the use of PPML estimator is appropriate.

Data

Panel data used in this study comprises Albanian agricultural exports to 46 import partners, including countries from EU-28, CEFTA 2006, EFTA and BRICS, as well as USA, Japan and Turkey. Data utilized in this study cover the period 1996-2013. Trade flows observed here cover 92% of Albanian agricultural exports for the given period. Data on agricultural export flows were obtained from the UNCTAD, disaggregated according to Standard International Trade Classification (SITC, rev. 3). Data on real Gross Domestic Product (GDP), population, exchange rate and inflation were acquired from the same source. Data on distance between capital cities, together with dummies on cultural and historical links such as adjacency (sharing common land border), common primary language and Albania's former colonizer were obtained from the CEPII (Centre d'Etudes Prospectives et d'Informations Internationales) database. Data on common RTAs with trading partners were utilized from the WTO (World Trade Organization). Lastly, data for institutional distance were obtained from the World Governance Indicators (WGI) database (Kaufmann et al., 2010). Data on the stock of Albanian Diaspora residing in the importing countries were obtained from the World Bank migration database. Missing data for the given time period in the case of institutional variables and stock of Albanian migrants were interpolated. Definition of variables, expected coefficient signs and basic statistics of the employed variables are summarized in Appendix Table 1. Correlation matrix presented in the Appendix Table 2 suggests that the issues related to multicollinearity are not present in the dataset. Data processing and empirical estimations were conducted on Stata 12.

Results

Agricultural trade in Albania

Albania is endowed with natural resources, such as fertile land, and suitable climatic conditions for agricultural production. Abundance of natural resources combined with low labour costs provides good grounds for intensification of labour intensive agricultural activities. Moreover, geographical layout, proximity to the EU market, and access to sea transport, make export potential viable in terms of low transport costs. Therefore, agriculture fulfils preconditions to excel Albanian export and shrink the actual sharp trade deficit. Despite its great potential, Albania remains a country with low agricultural exports and high dependency on imports. Since the early period of transition, agricultural exports marked a significant growth. Between the period 1996 and 2013, volume of agricultural exports increased from 32.4 million USD to 171.3 million USD. Data on Albanian agricultural trade (Figure 1) reveals that since 1996 agricultural exports marked over a five-fold increase, while imports rose at slower pace (3 times). Despite such impressive growth, data from 2013 suggest that agricultural exports/import coverage rate is only 20%, meaning that import to export ratio is as high as 5:1 (Figure 2).

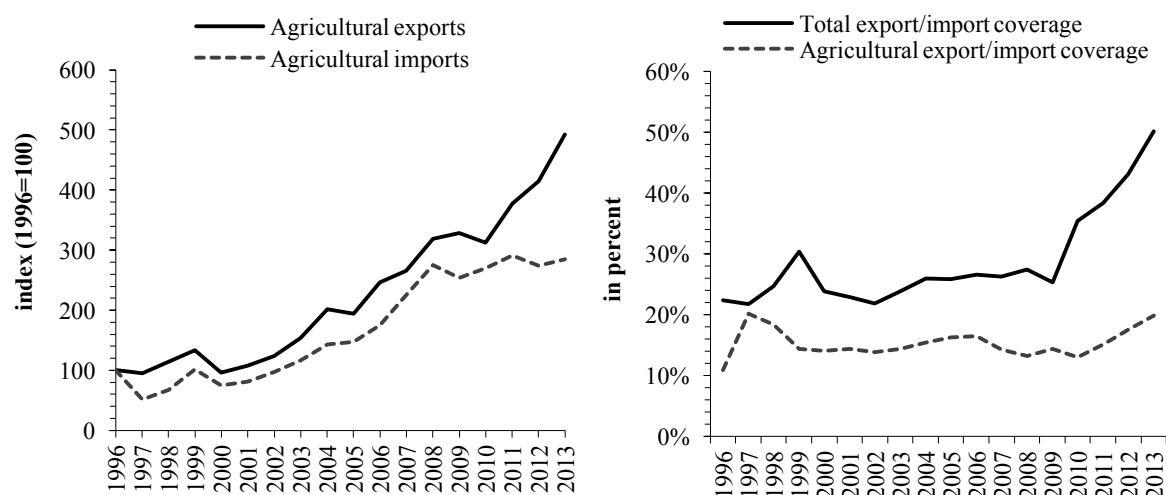


Figure 1: Growth of Albanian agricultural trade Figure 2: Agricult. import/export coverage
Source: UNCTAD, own elaboration

Destination of Albanian agricultural export

European Union is the main economic and trade partner for Albania since the beginning of transition process. Among others, strong trade linkages are reflected in the case of Albanian agricultural export destination. The share of agricultural exports to EU-28 constitutes two thirds (66.8%) of total agricultural exports for the period 2008-2013 (Figure 3). A slight decline in the share of agricultural exports to EU is directly affected by the global crisis of 2008-2009. According to ACCIT (2013) the crisis in Italy and Greece and drastic decline of domestic demand in both neighbouring countries had a direct impact in the slowdown of Albanian exports. Moreover, our estimations confirm that this is particularly true in the case of agricultural exports. Before the crisis (2007) share of agricultural exports to Italy was 40.0% while in 2013 it dropped at 35.1%. Similar outcome took place with agricultural exports to Greece, a fall from 10.5% in 2007 to 8.7% in 2013.

On the other hand, trade links with the majority of CEFTA 2006 countries have been well established even before the free trade agreement entered in force. Share of agricultural exports to the group of neighbouring SEE (South Eastern Europe) countries is 13.4%. Despite significant increase since 1996-2001, the share of agricultural exports to CEFTA 2006 countries remained relatively constant. In addition, EC (2015) suggest that Albanian export potential to

these group countries remains unexploited. Establishment of the CEFTA 2006 has particular merits in lowering technical barriers, but remains behind in releasing administrative barriers such as customs procedures, as well as dealing with barriers in the area of sanitary and phytosanitary measures. EFTA is inferior agricultural export partner to Albania. Total share of agricultural exports to EFTA countries is incremental, accounting for 0.3% of total agricultural exports. Unattractiveness of Albanian agricultural exports to this group of economies reflects high transport costs due to the large distance between EFTA members and Albania. Similarly to the trade pattern with EFTA, agricultural trade with informal trading block of BRICS countries (Brazil, Russia, India, China and South Africa) is very low. Total agricultural exports to BRICS during the period 1996-2013 were statistically insignificant (less than 1%) or 13.7 million USD.

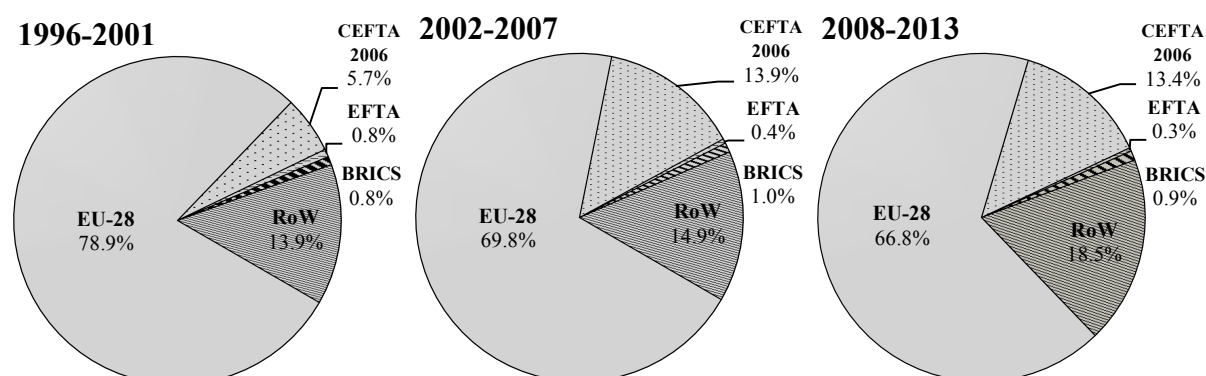


Figure 3: Agricultural exports, by trading blocs (in percentage)

Source: UNCTAD, own elaboration

Empirical results

Baseline model estimations reported in the Table 1 (Model 1) reveal that obtained results are persistent with theoretical framework. The coefficients of importer's economic size (*GDP*) and market size (*POP*) are positive and statistically significant. Importer's economic size is positive and significant in all estimated models, while the significance of the importer's market size varies over the estimated models. Results suggest that Albanian agricultural export will increase proportionally with an increase of importer's economic size. On the other hand, Albania's economic size is found to be positive but statistically insignificant, whilst the domestic market size has a robust significant negative coefficient. *Ceteris paribus*, increase in Albanian population enables domestic market to absorb a greater portion of agricultural production and reduces surpluses dedicated for export. This outcome is particularly relevant in the low income countries where agricultural and food commodities are perceived as normal goods. As expected, our results illustrate that distance has negative impact on agricultural exports in all estimated models. Such an outcome is typical for conventional gravity model analysis, since the distance is expected to affect export flows negatively. Increasing geographical distance between the capital city of Albania (Tirana) and capitals of importing countries proxies higher transport costs and decreases therefore agricultural export flows.

In addition to the traditional variables, we adjust the baseline model with the variable of bilateral income differential aiming to test for the relative strength of the Linder hypothesis vis-à-vis the Heckscher-Ohlin (HO) hypothesis. Yielded result (Model 2) implies that estimated coefficient of this variable is negative, but statistically insignificant. However, the estimates of the pooled model (Model 8) find the variable statistically significant at 5 percent. Such result implies that income disparities tend to decrease agricultural export flows, emphasizing income convergence as relevant factor in promoting export. Therefore, findings of this study support the Linder hypothesis in the case of Albania.

Results of the model augmented with effects of adjacency (sharing common border), linguistic similarities and colonial links (Model 3) confirm the common validity with theoretical foundations of the gravity model. Positive and significant coefficients obtained for these variables depict that Albanian agricultural export is strongly influenced by the transportation and transaction costs. Indeed, results predict higher agricultural export flow with countries that share common border with Albania. Similarly, common primary language and colonial links with the importing country tend to foster agricultural export flows. On the other hand, effect of landlocked importing country, despite the expected negative coefficient sign, is found statistically insignificant.

Once we extended the baseline model with the effects of Diaspora (Model 4), results revealed a strong impact of the Albanian immigrants residing in the importing country. Presence of a larger Albanian immigrant stock in the importing countries is associated with lower transaction and information costs and higher agricultural export flows. Moreover, relevance of the Albanian Diaspora, as it can be seen in the pooled model estimates (Model 8), prevails on its significance over the transaction costs (adjacency) and linguistic similarities (common language). Therefore, any trade enhancing policy aiming to promote agricultural export in the case of Albania should perceive Diaspora as irreplaceable platform for export promotion and growth.

Results of the effects of the bilateral exchange rate and price stability in the importing country are presented in Model 5. As expected, exchange rate has a significant positive coefficient, indicating that depreciation in Albanian Lek (ALL) against the currencies of importing partners facilitates agricultural exports. By contrary, coefficient of price stability (inflation) is found statistically insignificant, despite the expected negative coefficient sign.

Findings of this study yield relatively ambiguous results related to the effects of trade liberalization (Model 6). Results show that RTA with CEFTA 2006 countries had positive and significant impact on agricultural export creation, while export diversion effects prevail from the FTA with EFTA members. Accordingly, results induce negative coefficients for SAA with EU and FTA with Turkey, but statistically insignificant. This outcome should be interpreted with cautions, for at least two particular reasons. Firstly, impact of the free trade agreements in agriculture tends to produce delayed effects because of the asymmetric nature of FTAs. Actually, this outcome is persistent with previous studies indicating that it may take a several years or even longer until actual export creation effects in agriculture occur. And secondly, it might signal weak competitiveness of the Albanian farmers and their inferior position towards heavily subsidized farmers of the importing countries.

The effects of institutional environment in agricultural export are observed in Model 7. Results of the baseline model extended with bilateral institutional distance derive significant negative coefficient indicating that costs of agricultural export increase with institutional distance. Performance of Albanian agricultural export diminishes with higher institutional quality disparities between trading partners. Indeed, institutional heterogeneity induces higher transaction costs and restrictive effects on Albanian agricultural export. Therefore, the greater is the institutional quality gap with the importing country the lower are Albanian agricultural export flows.

Table 1: PPML regression results of the gravity model: Agricultural export of Albania

AGR_exp	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
ln_GDP_imp	0.855*** (0.058)	0.927*** (0.087)	1.011*** (0.070)	0.367*** (0.066)	0.579*** (0.102)	0.967*** (0.069)	1.134*** (0.100)	0.781*** (0.135)
ln_GDP_exp	0.135 (0.140)	0.150 (0.137)	0.068*** (0.134)	0.045 (0.113)	0.303** (0.144)	-0.005 (0.176)	-0.292 (0.182)	-0.471*** (0.166)
ln_POP_imp	0.303*** (0.059)	0.209** (0.094)	0.337*** (0.064)	0.464*** (0.069)	0.554*** (0.091)	0.228*** (0.072)	-0.129 (0.118)	0.037 (0.116)
ln_POP_exp	-5.595** (2.345)	-5.597** (2.334)	-5.476** (2.456)	-8.953*** (1.883)	-4.993** (2.284)	-9.579** (4.801)	-5.784*** (2.197)	-14.680*** (2.915)
ln_DIST	-2.462*** (0.100)	-2.426*** (0.097)	-2.293*** (0.107)	-1.330*** (0.097)	-2.371*** (0.108)	-2.434*** (0.095)	-2.135*** (0.135)	-1.146*** (0.113)
GDPpc_dist		-0.047 (0.029)						-0.065** (0.032)
ADJ			1.098*** (0.152)					0.016 (0.202)
LANG			0.933** (0.363)					-0.640** (0.251)
LAND			-0.043 (0.196)					0.766*** (0.210)
COL			0.394*** (0.135)					0.764*** (0.175)
ln_DIA				0.303*** (0.022)				0.275*** (0.034)
ln_EXR					0.276*** (0.072)			0.144*** (0.052)
INF					-0.009 (0.006)			-0.014*** (0.005)
CEFTA						0.561* (0.311)		0.688** (0.272)
SAA_eu						-0.291 (0.206)		-0.396*** (0.139)
FTA_efta						-1.875*** (0.370)		-1.105** (0.454)
FTA_tur						-0.005 (0.224)		-0.671*** (0.218)
INST_dist							-0.152*** (0.037)	-0.146*** (0.027)
cons	47.137** (19.777)	46.869** (19.641)	42.975** (20.504)	68.938*** (15.830)	40.106** (19.218)	79.437** (39.713)	51.457*** (18.650)	117.275*** (24.271)
R2	0.884	0.886	0.877	0.933	0.889	0.878	0.891	0.949
Observations	792	792	792	747	792	792	783	738

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Own elaboration

Potential of agricultural export

In the last section of this study we estimate Albania's export potential by comparing actual agricultural exports with predicted exports. Results presented in this section show the absolute difference between the actual and predicted level of agricultural export ($A - P$). A positive value implies the possibility of agricultural export expansion while a negative value indicates that Albania has exceeded its export potential with a trading partner. For the sake of simplicity, results of the export potential are presented in the aggregate format for the period 1996-2013.

As it is revealed in the Figure 4, Albania overexploits its agricultural export potential with its traditional EU neighbouring markets (Greece and Italy), culturally proximate trade partners (Kosovo, Turkey, Croatia, Bosnia and Herzegovina) as well as geographically distant countries (USA and Japan). On the other hand, Albania has unused agricultural export potential particularly with the Central and Eastern European Countries (CEECs) such as Bulgaria, Romania, Hungary and Poland. With this group of new EU member countries, Albania has institutional similarities and comparatively lower transport and transaction costs. Therefore, market access into these markets is significantly easier compared to the EU developed countries. Additional advantage to the market expansion in such markets is related to similarities in consumer preferences and the common status of transitional economies, such as it is the case of Albania.

On the other hand, results of this study identify untapped export potential of the Albanian agricultural exports in the group of developed European countries. This is particularly true for the Western European markets such as UK, France, Switzerland and Germany. As it is noted from the results of previous section (particularly in the case of Italy and Greece) primary advantage to market expansion in this group of countries is large presence of Albanian Diaspora. Migrant links, among other factors, might serve as a solid platform for intensification towards these export markets. On the other hand, the main barriers in exploiting export potential in these countries are related to higher transport and transaction costs, institutional dissimilarities and higher quality standards.

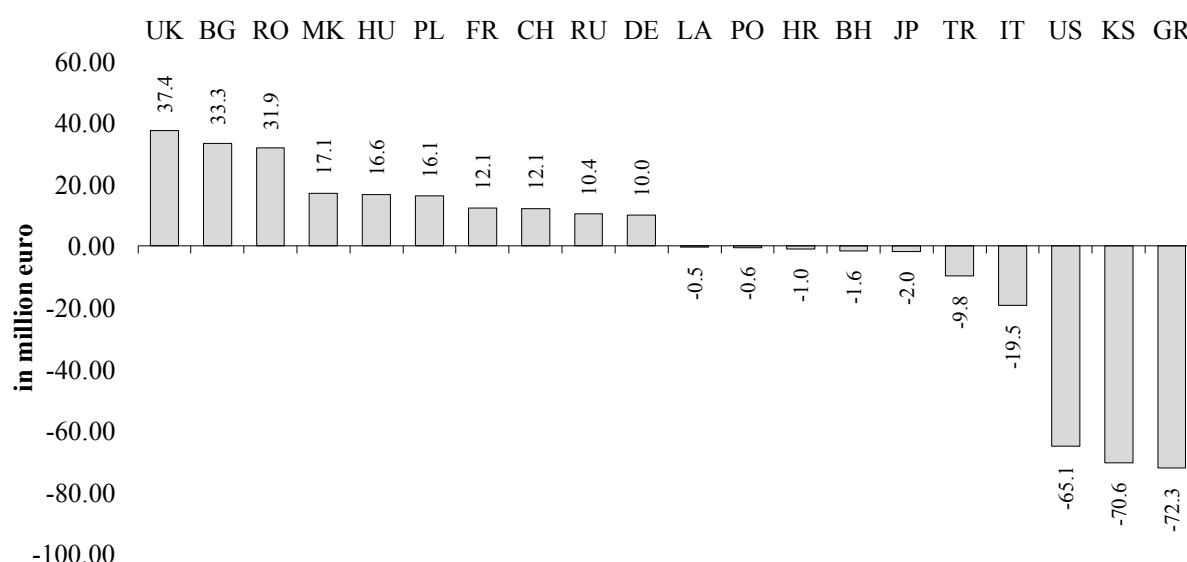


Figure 4: Potential agricultural export 1996-2013 (actual export - predicted export)

Source: UNCTAD, own elaboration

Discussion and remarks

Our gravity analysis for Albanian agricultural export leads to comparable results as models for other countries. For example, a study of determinants of Turkish agricultural exports to the European Union (Erdem and Nazlioglu, 2008) found that Turkish agricultural exports to the EU are positively correlated with the size of the economy, the importer population, the Turkish population living in the EU countries, the non-Mediterranean climatic environment, and the membership to the EU-Turkey Customs Union Agreement while they are negatively correlated with agricultural arable land of the EU countries and geographical distance between Turkey and the EU countries. Results from Albania also confirm importance of traditional gravity variables and importance of exporter's Diaspora for export of agricultural products.

Transformation of the agricultural sector is a very sensitive aspect. In many Central and Eastern European Countries it was connected to the transition process and later also with adoption of common EU rules. Experience from Central and Eastern European Countries (see Svatos and Smutka, 2010; Svatos et al., 2010) revealed, that the process of EU accession reflected positively in results of agricultural trade. Moreover, EU accession resulted in agricultural export concentration in the common internal market (Svatos and Smutka, 2010). On the other side, trade creating effect of RTAs was confirmed by Korinek and Melatos (2009). Their gravity model for members of three regional trade agreements suggests that the creation of AFTA (ASEAN Free Trade Agreement), COMESA (Common Market for Eastern and Southern Africa) and MERCOSUR (Southern Cone Common Market) has increased trade in agricultural products between the RTAs countries. They also found that in some cases, lack of transport and communications infrastructure, in addition to supply constraints, lessens the effect of the RTAs on trade flows. Besides RTAs, preferential trade policies can also help to support international trade (Cipollina et al., 2010). Most developing countries can export to the European Union and the United States with preferential market access. The results show (Cipollina et al., 2010) that preferential schemes have a significant impact on trade in terms of margins and intensity, and such effect seems to be stronger in the case of EU preferences, although with significant differences across products. In the case of Albania not all RTAs and FTAs have the same effect on agricultural trade, in our study export creating affect was confirmed for RTA with CEFTA 2006 countries and export diversion effect for FTA with EFTA countries.

According to gravity model for Egypt's agricultural exports (Hatab et al., 2010) 1% increase in Egypt's GDP generates more than 5% increase in its agricultural export flows. In contrast, the increase in Egypt's GDP per capita causes exports to decrease, similarly as in our model. Authors argue on such outcome emphasizing that economic growth increases per capita demand for all normal goods. Moreover, the exchange volatility has positive coefficient (depreciation in Egyptian Pound stimulates agricultural exports) and transportation costs have a negative influence on Egyptian agricultural exports. The same outcome of exchange rate volatility can be observed in the case of Hungarian agricultural exports (Fogarasi, 2011). Other variables, such as population and income (GDP) of export destination countries have positive sign, while distance from Hungary has a negative one.

Effects of the institutional determinants in agricultural trade were investigated by Bojnec and Fertő (2015). They focus on effects of quality of institutions and similarity of institutions in explaining variation in bilateral agricultural and food exports among OECD countries. Study finds out that good quality of institutions reduces the effects of distance. Factors influencing bilateral trade among the Western Balkan countries were identified in the work of Trivic and Klimczak (2015). They considered geographical, economic or political determinants as well as factors constituting cultural, communicational and historical proximity between countries. Their results differ from traditional results gained from gravity analysis in the way that the

strongest influence on trade values were exhibited by variables representing ease of a direct communication and similarity of religious structures. In addition, war and one-year-post-war effect showed a strong and statistically important influence. The authors therefore conclude that non-economic factors in the region of the Western Balkans play the most important role in determining trade values between countries. Our analysis for the case of Albania confirms these results to the extent that Albanian immigrants in importing countries represent a significant factor for export growth, even if the countries are geographical neighbours or have similar language. Furthermore our results indicate that more similar institutional environment of the trade partner to Albanian one has positive effect on its agricultural export.

Conclusion

The paper employs gravity model approach to analyse main determinants of agricultural export in Albania. The study utilizes econometric approach using Poisson Pseudo-Maximum Likelihood (PPML) estimation for Albanian agricultural export flows with major trading partners for the period 1996-2013. Main results of the baseline model suggest that agricultural export flow increase with increasing economic size (GDP), revealing higher impact of importer's absorbing potential comparatively to Albania's productive potential. On the other hand, increase of Albanian market size (population) has diminishing effects on agricultural export flows. *Ceteris paribus*, growth in domestic demand, resulting from population growth, leads to reduction of agricultural export. As expected, findings of this study suggest that increasing distance between trading partners is associated with reduction of Albanian agricultural export.

Albanian agricultural export is highly concentrated in a limited number of importing partners, respectively in neighbouring countries (such as Italy and Greece). It indicates that geographical proximity, low transport and transaction costs are key drivers of agricultural export. Such an outcome is supported by the results of the augmented gravity model conducted in this study. Namely, results reveal that higher agricultural export flows are associated with neighbouring countries sharing common border. Moreover, stronger linguistic similarities and cultural links with importing partners (such as Kosovo and Macedonia) tend to accelerate Albanian agricultural export. Influence of Albanian Diaspora residing in the importing partner countries is found to have robust effect on the promotion of agricultural export. Interestingly, findings of this study suggest that effects of Diaspora prevail on their importance over the transport and transaction costs.

On the other hand, devaluation of the Albanian currency has significantly positive impact on Albanian agricultural export flow, prevailing on its relevance over the price stability (inflation) in the importing countries. Concerning the effects of trade liberalization on the performance of agricultural export, our findings depict that RTA with CEFTA 2006 countries had trade creating, while FTA and EFTA trade diverse effect. Effects of SAA with EU and FTA with Turkey are found statistically insignificant. Actually, these findings should be perceived with caution due to asymmetric nature and short time lap since these trade agreements entered into force.

Lastly, bilateral institutional distance tends to diminish Albanian agricultural exports. Therefore, institutional convergence with the EU standards, based on the principles of well functioning market economy, would influence the extension of Albanian exports in those European markets (in which breakthrough of Albanian agricultural exports is limited due to institutional barriers). Moreover, improvement of institutional quality would have influence on interim institutional stability for domestic farmers, including better credit access, fight against corruption and sustainable political stability.

Findings of this study are important for trade and agricultural policy makers. From the trade policy perspective, one should assume that the platform of agricultural export promotion should aim market diversification in those countries (other than neighbouring countries) in which Albanian farmers can exploit their comparative advantage. Indeed, Albania is a small and open economy operating in the liberalized trade regime therefore any trade restrictive efforts might produce negative effects. On the other hand, from the agricultural policy perspective, special attention should be paid to measures that lead to improvement of the competitiveness of local farmers. Public investments in the rural infrastructure and irrigation system should be accompanied with direct farmer support. Notably, Albania has huge potential to become competitive actor in international markets if supportive measures are directed in increasing productivity of labour intensive agricultural sectors, such as fruits, vegetables, medical plants and fishery. Further specialisation in these sectors is supported by the present factor market endowments, natural resources and climate conditions in Albania.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0552. The authors also acknowledge financial support from the projects VEGA 1/0806/15, VEGA 1/0930/15, VEGA 1/0797/16.

References

- ACCIT (Albanian Centre for Competitiveness and International Trade). (2013) „*Albania Trade Report 2013*”. Albanian Centre for Competitiveness and International Trade. [Online]. Available: <http://acit.al/index.php/trade-reports> [Accessed: 15 Sep. 2016].
- Anderson, J. E. and Van Wincoop, E. (2001) „Borders, trade and welfare”. In Collins, S. and Rodrik, D. (Eds.), pp. 207 – 244. DOI <http://dx.doi.org/10.1353/btf.2001.0002>.
- Anderson, J. E. and Van Wincoop, E. (2001) „Gravity with gravitas: a solution to the border puzzle”, *American Economic Review*, Vol. 106, pp. 170 – 192. ISSN 0002-8282.
- Armstrong, S. (2007) „Measuring trade and trade potential: A survey“, Crawford School Asia Pacific Economic Paper No. 368. DOI <http://dx.doi.org/10.2139/ssrn.1760426>.
- Asllani, A. (2013) „Trade Gravity, Diversification and Correlation Relationship between Albania and Balkans’ Countries“. Risk and its assessment in regional economy. Proceedings of the Fifth International Conference for Risk. Ohrid: Albanian Centre for Risk.
- Bojnec, S. and Fertő, I. (2015) „Institutional Determinants of Agro-Food Trade”, *Transformations in Business and Economics*, Vol. 14, No. 2, pp. 35 – 52. ISSN 1648-4460.
- Bryant, J., Genç, M. and Law, D. (2004) „Trade and migration to New Zealand”, Working paper No. 04/18, Wellington: New Zealand Treasury.
- CEPII (2016) “*GeoDist database*”. Centre d’Etudes Prospectives et d’Informations Internationales, November 2016. [Online]. Available: http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id=6 [Accessed: 18 Dec. 2016].
- Cheng, I. H. and Wall, H. J. (2005) „Controlling for Heterogeneity in Gravity Models of Trade and Integration”, *Federal Reserve Bank of St. Louis Review*, Vol. 87, No. 1, pp. 49 – 63.

- Cho, G., Sheldon, I. M. and McCorriston, S. (2002) „Exchange rate uncertainty and agricultural trade“, *American Journal of Agricultural Economics*, Vol. 84, No. 4, pp. 931 – 942. DOI <http://dx.doi.org/10.1111/1467-8276.00044>.
- Cipollina, M., Laborde, D. and Salvatici, L. (2010) „Do Preferential Trade Policies (Actually) Increase Exports? A comparison between EU and US trade policies“, Agricultural and Applied Economics Association, Series 2013 Annual Meeting, August 4-6, 2013, Washington, D.C. No. 150177.
- Cungu, A. and Swinnen, J. F. (1999) „Albania’s radical agrarian reform“, *Economic Development and Cultural Change*, Vol. 47, No. 3, pp. 605 – 619. DOI <http://dx.doi.org/10.1086/452421>.
- De Groot, H. L., Linders, G. J., Rietveld, P. and Subramanian, U. (2004) „The institutional determinants of bilateral trade patterns“, *Kyklos*, Vol. 57, No. 1, pp. 103 – 123. DOI <http://dx.doi.org/10.1111/j.0023-5962.2004.00245.x>.
- Deininger, K., Savastano, S. and Carletto, C. (2012) „Land fragmentation, cropland abandonment, and land market operation in Albania“, *World Development*, Vol. 40, No. 10, pp. 2108 – 2122. DOI <http://dx.doi.org/10.1016/j.worlddev.2012.05.010>.
- EC (European Commission). (2014) „*Albania: Bilateral Relations in Agriculture*“. [Online]. Available: http://ec.europa.eu/agriculture/bilateral-relations/pdf/albania_en.pdf [Accessed: 18 Dec. 2016].
- Eichengreen, B. and Irwin, D. A. (1998) „The role of history in bilateral trade flows“. In *The Regionalization of the World Economy*, pp. 33 – 62. University of Chicago Press. DOI 10.7208/chicago/9780226260228.001.0001.
- Erdem, E. and Nazlioglu, S. (2008) „Gravity Model of Turkish Agricultural Exports to the European Union“, International Trade and Finance Association Conference Papers (p. 21). bepress.
- Fertő, I. and Fogarasi, J. (2011) „On Trade Impact of Exchange Rate Volatility and Institutional Quality: The Case of Central European Countries“. Paper prepared for presentation at the EAAE 2011 Congress, Zurich, Switzerland.
- Fetahu, E. (2014) „Trade Integration between Albania and European Union: A Gravity Model Based Analysis“, *European Scientific Journal*, Vol. 10, No. 7, pp. 185 – 199.
- Fogarasi, J. (2011) „The Effect of Exchange Rate Volatility upon Foreign Trade of Hungarian Agricultural Products“, Research Institute of Agricultural Economics – Hungarian Academy of Sciences: Studies in Agricultural Economics, No. 113, pp. 85 – 96.
- Frankel, J. A. (1997) „Regional Trading Blocs in the World Economic System“, Washington: Institute for International Economics.
- Frankel, J. A. and Wei, S. J. (1998) „Regionalization of world trade and currencies: Economics and politics“. In *The Regionalization of the World Economy*, pp. 189 – 226, University of Chicago Press. DOI 10.7208/chicago/9780226260228.001.0001.
- Genc, M., Gheasi, M., Nijkamp, P. and Poot, J. (2012) „The impact of immigration on international trade: a meta-analysis“, In *Migration Impact Assessment: New Horizons*. DOI <http://dx.doi.org/10.4337/9780857934581.00019>.
- Gómez-Herrera, E. (2013) „Comparing alternative methods to estimate gravity models of bilateral trade“, *Empirical Economics*, Vol. 44, No. 3, pp. 1087 – 1111. ISSN 0377-7332.

- Gould, D. M. (1994) „Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows“, *Review of Economics and Statistics*, Vol. 76, pp. 302 – 316. DOI <http://dx.doi.org/10.2307/2109884>.
- Government of Albania. (2015) „*Business and Investment Development Strategy for the period 2014-2020*“. [Online]. Available: http://www.ekonomia.gov.al/files/userfiles/Business&Investment_Dev._Strategy.pdf [Accessed: 16 Nov. 2016].
- Government of Albania. (2016) „*Albania's Reform Programme 2016-2018*“. [Online]. Available: http://www.ekonomia.gov.al/files/userfiles/Albania_s_Economic_Reform_Programme_2016-2018.pdf [Accessed: 17 Nov. 2016].
- Grant, J. H. and Lambert, D. M. (2008) „Do regional trade agreements increase members' agricultural trade?“, *American Journal of Agricultural Economics*, Vol. 90, No. 3, pp. 765 – 782. DOI <http://dx.doi.org/10.1111/j.1467-8276.2008.01134.x>.
- Hatab, A. A., Romstad, E. and Huo, X. (2010) „Determinants of Egyptian agricultural exports: a gravity model approach“, *Modern Economy*, Vol. 1, No. 3, pp. 134 – 143. DOI <http://dx.doi.org/10.4236/me.2010.13015>.
- Head, K. and Ries, J. (1998) „Immigration and trade creation: econometric evidence from Canada“, *Canadian Journal of Economics*, Vol. 31, No. 1, pp. 47 – 62. DOI <http://dx.doi.org/10.2307/136376>.
- Helpman, E., Melitz, M. J. and Rubinstein, Y. (2008) „Estimating trade flows: Trading partners and trading volumes“, *Quarterly Journal of Economics*, Vol. 73, pp. 441 – 486. DOI <http://dx.doi.org/10.1162/qjec.2008.123.2.441>.
- Jansen, M. and Nordås, H. K. (2004) „Institutions, trade policy and trade flows“. WTO Staff Working Paper, No. ERSD-2004-02. DOI <http://dx.doi.org/10.2139/ssrn.923544>.
- Jansen, M. and Piermartini, R. (2009) „Temporary migration and bilateral trade flows“, *The World Economy*, Vol. 32, No. 5, pp. 735 – 753. DOI <http://dx.doi.org/10.1111/j.1467-9701.2009.01167.x>.
- Jayasinghe, S. and Sarker, R. (2008) „Effects of regional trade agreements on trade in agrifood products: Evidence from gravity modeling using disaggregated data“, *Applied Economic Perspectives and Policy*, Vol. 30, No. 1, pp. 61 – 81. DOI <http://dx.doi.org/10.1111/j.1467-9353.2007.00392.x>.
- Kafle, K. and Kennedy, P. L. (2012) „Exchange rate volatility and bilateral agricultural trade flows: The case of the United States and OECD Countries“, LAP LAMBERT Academic Publishing.
- Kandilov, I. T. (2008) „The effects of exchange rate volatility on agricultural trade“, *American Journal of Agricultural Economics*, Vol. 90, No. 4, pp. 1028 – 1043. DOI <http://dx.doi.org/10.1111/j.1467-8276.2008.01167.x>.
- Kaufmann, D., Kraay, A. and Mastruzzi, M. (2010) „*The Worldwide Governance Indicators: a summary of methodology, data and analytical Issues*“, World Bank Policy Research Working Paper no. 5431.
- Koo, W. W., Karemera, D. and Taylor, R. (1994) „A gravity model analysis of meat trade policies“, *Agricultural Economics*, Vol. 10, No. 1, pp. 81 – 88. DOI [http://dx.doi.org/10.1016/0169-5150\(94\)90042-6](http://dx.doi.org/10.1016/0169-5150(94)90042-6).

- Koo, W. W., Kennedy, P. L. and Skripnitchenko, A. (2006) „Regional preferential trade agreements: Trade creation and diversion effects“, *Applied Economic Perspectives and Policy*, Vol. 28, No. 3, pp. 408 – 415. DOI <http://dx.doi.org/10.1111/j.1467-9353.2006.00306.x>.
- Korinek, J. and Melatos, M. (2009) „Trade Impacts of Selected Regional Trade Agreements in Agriculture“, OECD Trade Policy Working Papers, No. 87, OECD publishing.
- Korovilas, J. P. (1999) „The Albanian economy in transition: the role of remittances and pyramid investment schemes“, *Post-Communist Economies*, Vol. 11, No. 3, pp. 399 – 415. DOI <http://dx.doi.org/10.1080/14631379995940>.
- Levchenko, A. A. (2004) „Institutional quality and international trade“, International Monetary Fund, Working paper No. WP/04/231. ISSN 1018-5941.
- Linders, G. J. and De Groot, H. L. (2006) „Estimation of the gravity equation in the presence of zero flows“, Tinbergen Institute Discussion Paper, No. 06-072/3. DOI <http://dx.doi.org/10.2139/ssrn.924160>.
- Linders, G. J., Slangen, A., De Groot, H. L. and Beugelsdijk, S. (2005) „Cultural and institutional determinants of bilateral trade flows“, Tinbergen Institute Discussion Paper, No. 05-074/3. DOI <http://dx.doi.org/10.2139/ssrn.775504>.
- Linnemann, H. (1966) „An Econometric Study of International Trade Flows“, Amsterdam: North-Holland.
- Maitah, M., Kuzmenko, E. and Smutka, L. (2016) “Real Effective Exchange Rate of Rouble and Competitiveness of Russian Agrarian Producers”, *Economies*, Vol. 4, No. 3. DOI [10.3390/economies4030012](https://doi.org/10.3390/economies4030012).
- Martin, W. and Pham, C. S. (2008) „Estimating the gravity equation when zero trade flows are frequent“, MPRA Working Paper No. 9453, University Library of Munich.
- Martinez-Zarzoso, I. (2013) “The log of gravity revisited”, *Applied Economics*, Vol. 45, No. 3, pp. 311 – 327. DOI <http://dx.doi.org/10.1080/00036846.2011.599786>.
- Martinez-Zarzoso, I. and Nowak-Lehmann, G. (2003) „Augmented gravity model: An empirical application to Mercosur-European Union trade flows“, *Journal of Applied Economics*, Vol. 6, No. 2, pp. 291 – 316. ISSN 1514-0326.
- McCallum, J. (1995) „National Borders Matter: Canada–US Regional Trade Patterns“, *American Economic Review*, Vol. 85, No. 3, pp. 615 – 623.
- McCarthy, N., Carletto, C., Kilic, T. and Davis, B. (2009) „Assessing the impact of massive out-migration on Albanian agriculture“, *European Journal of Development Research*, Vol. 21, No. 3, pp. 448 – 470. DOI <http://dx.doi.org/10.1057/ejdr.2009.12>.
- Parsons, C. (2005) „Quantifying the trade-migration nexus of the enlarged EU. A comedy of errors or much ado about nothing“, Sussex Migration Working Paper no. 27.
- Peri, G. and Requena-Silvente, F. (2010) „The trade creation effect of immigrants: evidence from the remarkable case of Spain“, *Canadian Journal of Economics*, Vol. 43, No. 4, pp. 1433 – 1459. DOI <http://dx.doi.org/10.1111/j.1540-5982.2010.01620.x>.
- Pishbahar, E. and Huchet-Bourdon, M. (2008) „European Union’s preferential trade agreements in agricultural sector: a gravity approach“, *Journal of International Agricultural Trade and Development*, Vol. 5, No. 1, pp. 93 – 114. ISSN 1556-8520.

- Pokrivčák, J. and Šindlerová, K. (2011) „Gravity Model of EU's Bilateral Trade with Different Products”, *Acta Oeconomica et Informatica*, Vol. 14, pp. 33 – 37. ISSN 1336-9261.
- Pöyhönen, P. (1963). „A Tentative Model for the Volume of Trade between Countries”, *Weltwirtschaftliches Archiv*, Band 90, Heft 1, pp. 93 – 100.
- Qineti, A., Rajcaniova, M., Braha, K., Ciaian, P. and Demaj, J. (2015) „Status quo bias of agrarian land structures in rural Albania”, *Post-Communist Economies*, Vol. 27, No. 4, pp. 517 – 536. DOI <http://dx.doi.org/10.1080/14631377.2015.1084732>.
- Rauch, J. E. and Trindade, V. (2002) „Ethnic Chinese networks in international trade”, *Review of Economics and Statistics*, Vol. 84, No. 1, pp. 116 – 130. DOI <http://dx.doi.org/10.1162/003465302317331955>.
- Rose, A. K. (2004) „Do We Really Know That the WTO Increases Trade?”, *American Economic Review*, Vol. 94, pp. 98 – 114. DOI <http://dx.doi.org/10.1257/000282804322970724>.
- Sejdini, A. and Kraja, I. (2014) „International Trade of Albania. Gravity model”, *European Journal of Social Sciences*, Vol. 2, No. 1, pp. 220 – 228. ISSN 1450-2267.
- Ševela, M. (2002) “Gravity-type model of Czech agricultural export”, *Agricultural Economics*, Vol. 48, No. 10, pp. 463 – 466. ISSN 0139-570X.
- Sheldon, I., Mishra, S. K., Pick, D. and Thompson, S. R. (2013) „Exchange rate uncertainty and US bilateral fresh fruit and fresh vegetable trade: an application of the gravity model”, *Applied Economics*, Vol. 45, No. 15, pp. 2067 – 2082. DOI <http://dx.doi.org/10.1080/00036846.2011.650330>.
- Silva, J. S. and Tenreyro, S. (2006) „The log of gravity”, *The Review of Economics and Statistics*, Vol. 88, No. 4, pp. 641 – 658. DOI <http://dx.doi.org/10.1162/rest.88.4.641>.
- Silva, J. S. and Tenreyro, S. (2011) „Further simulation evidence on the performance of the Poisson Pseudo-Maximum Likelihood Estimator”, *Economics Letters*, Vol. 112, No. 2, pp. 220 – 222. ISSN 0165-1765.
- Sun, L. and Reed, M. R. (2010) „Impacts of free trade agreements on agricultural trade creation and trade diversion”, *American Journal of Agricultural Economics*, Vol. 92, No. 5, pp. 1351 – 1363. DOI <http://dx.doi.org/10.1093/ajae/aaq076>.
- Svatos, M. and Smutka, L. (2010) „Development of agricultural trade and competitiveness of the commodity structures of individual countries of the Visegrad Group”, *Agricultural Economics-Zemедelska Ekonomika*, Vol. 58, No. 5, pp. 222 – 238. ISSN 0139-570X.
- Svatos, M., Smutka, L. and Miffek, O. (2010) „Competitiveness of agrarian trade of EU-15 countries in comparison with new EU member states”, *Agricultural Economics-Zemедelska Ekonomika*, Vol. 56, No. 12, pp. 569 – 582. ISSN 0139-570X.
- Tinbergen, J. (1962) „Shaping the World Economy: Suggestions for an International Economic Policy”, New York: The Twentieth Century Fund. DOI 10.1002/tie.5060050113.
- Trivic, J. and Klimczak, L. (2015) „The determinants of intra-regional trade in the Western Balkans. Proceedings of Rijeka Faculty of Economics”, *Journal of Economics and Business*, Vol. 33, No. 1, pp. 37 – 66. ISSN 1331-8004.
- UNCTAD (United Nations Conference on Trade and Development). (2015) „Data Center”. [Online]. Available: <http://unctadstat.unctad.org/EN/> [Accessed: 20 Sep. 2016].

- USAID (2012) „*Performance Evaluation of the Albanian Agricultural Competitiveness Program*”[Online]. Available: <https://www.usaid.gov/sites/default/files/documents/1863/Final%20Albania%20Report%20July%2030.pdf> [Accessed: 16 Oct. 2016].
- Westerlund, J. and Wilhelmsson, F. (2009) „Estimating the gravity model without gravity using panel data”, *Applied Economics*, Vol. 43, pp. 641 – 649. DOI <http://dx.doi.org/10.1080/00036840802599784>.
- World Bank (2015) „*Global Bilateral Migration*”. [Online]. Available: <http://databank.worldbank.org/data/reports.aspx?source=global-bilateral-migration#> [Accessed: 16 Oct. 2016].
- World Bank (2015) „*Migration data: Bilateral Migration Matrix*”. [Online]. Available: <http://www.worldbank.org/en/topic/migrationremittancesdiasporaissues/brief/migration-remittances-data> [Accessed: 18 Oct. 2016].
- World Bank (2015) „*Worldwide Governance Indicators*”. [Online]. Available: <http://data.worldbank.org/data-catalog/worldwide-governance-indicators> [Accessed: 25 Oct. 2016].
- WTO (World Trade Organization) (2015) „*Albania: Preferential Trade Agreements*”. [Online]. Available: <http://ptadb.wto.org/Country.aspx?code=008> [Accessed: 15 Nov. 2016].
- WTO (World Trade Organization). (2015) „*Albania: Regional Trade Agreements*”. [Online]. Available: <http://rtais.wto.org/UI/PublicSearchByMemberResult.aspx?lang=1&membercode=008&redirect=1> [Accessed: 16 Nov. 2016].
- WTO (World Trade Organization). (2016) “Albania Trade Policy Review”, WTO Secretariat Report, WT/TPR/337.
- Xhepa, S. and Agolli, M. (2004) „Albania’s foreign trade through a gravity approach”, Albanian Centre for International Trade Research Paper.
- Zahariadis, Y. (2007) „The Effects of the Albania-EU Stabilization and Association Agreement: Economic Impact and Social Implications”, London: Overseas Development Institute, ESAU Working Paper 17.

Appendices

Appendix Table 1: Definition, expected sign and basic statistics of the model variables

Variable	Code	Definition	Source	Period	Expected sign	Summary statistics				
						Obs.	Mean	ST D.	Min	Max
Agricultural export	AGR_exp	Agricultural exports of Albania (in million USD)	UNCTAD	1996-2013		792	1.628	5.435	0.000	60.215
GDP importer	ln_GDP_imp	Log of real GDP of importing country (in million USD)	UNCTAD	1996-2013	+	792	11.890	2.096	7.066	16.641
GDP exporter	ln_GDP_exp	Log of real GDP of Albania (in million USD)	UNCTAD	1996-2013	+	792	8.804	0.598	7.743	9.465
Population importer	ln_POP_imp	Log of population of importing country (in thousands)	UNCTAD	1996-2013	+/-	792	9.365	1.942	5.599	14.125
Population exporter	ln_POP_exp	Log of population of exporting country (in thousands)	UNCTAD	1996-2013	+/-	792	8.016	0.031	7.966	8.047
Distance	ln_DIST	Log of Distance between capitals of Albania and importer	CEPII	1996-2013	-	792	7.233	0.962	5.050	9.159
GDP pc distance	GDPpc_dist	GDP per capita distance between Albania and importer	UNCTAD	1996-2014	+/-	792	1.869	3.554	0.000	27.698
Adjacency	ADJ	= 1 if Albania and importer share common border	CEPII	1996-2013	+	792	0.068	0.252	0.000	1.000
Language	LANG	= 1 if Albania and importer share common language	CEPII	1996-2014	+	792	0.034	0.182	0.000	1.000
Landlocked	LAND	= 1 if importer is landlocked, dummy	CEPII	1996-2015	-	792	0.182	0.386	0.000	1.000
Colony	COL	= 1 if importer was Albania's colonizer, dummy	CEPII	1996-2016	+	792	0.023	0.149	0.000	1.000
Albanian Diaspora	ln_DIA	Log of Albanian migrant stock in importing country	World Bank	1996-2016	+	747	5.904	2.934	0.000	13.425
Exchange rate	ln_EXR	Log of exchange rate between ALL/currency of importer	UNCTAD	1996-2013	+	792	3.665	1.661	-0.76	7.157
Inflation	INF	Inflation rate of the importer (CPI annual rate)	UNCTAD	1996-2013	-	792	7.086	39.37	-4.48	1058.3
CEFTA 2006	CEFTA	= 1 if RTA with CEFTA 2006 countries, in force	WTO	Since 2007	+	792	0.061	0.239	0.000	1.000
SAA with EU	SAA_eu	= 1 if SAA with EU, in force	WTO	Since 2009	+	792	0.172	0.377	0.000	1.000
EFTA	FTA_efta	= 1 if FTA with EFTA countries, in force	WTO	Since 2011	+	792	0.011	0.106	0.000	1.000
FTA Turkey	FTA_tur	= 1 if FTA with Turkey, in force	WTO	Since 2012	+	792	0.008	0.087	0.000	1.000
Institutional distance	INST_dist	Institutional distance between Albania and importer	WGI	1996-2016	+/-	783	3.662	3.228	0.000	11.938

Note: RTA (Regional Trade Agreement). FTA (Free Trade Agreement), SAA (Stabilization and Association Agreement), ALL (Albanian Lek), CPI (Consumer Price Index)

Source: Own elaboration

Appendix Table 2: Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
(1) AGR_exp	1.00 0																		
(2) ln_GDP_imp	0.25 9	1.0 00																	
(3) ln_GDP_exp	0.12 8	0.1 71	1.0 00																
(4) ln_POP_imp	0.18 2	0.8 25	- 0.0 09	1.00 0															
(5) ln_POP_exp	- 0.13 7	- 0.1 51	- 0.8 27	- 0.00 8	1.0 00														
(6) ln_DIST	- 0.17 1	0.5 63	- 0.0 41	0.47 3	0.0 37	1.0 00													
(7) GDPpc_dist	- 0.00 8	0.1 77	0.2 52	- 0.18 2	- 0.2 35	0.1 65	1.0 00												
(8) ADJ	0.13 0	- 0.2 49	0.0 40	- 0.16 9	- 0.0 37	- 0.4 75	- 0.1 18	1.0 00											
(9) LANG	0.00 3	- 0.2 58	- 0.0 04	- 0.14 1	0.0 03	- 0.3 75	- 0.0 87	0.6 21	1.0 00										
(10) LAND	- 0.07 9	- 0.1 97	0.0 18	- 0.22 4	- 0.0 17	- 0.3 44	0.1 87	0.1 43	0.3 34	1.0 00									
(11) COL	0.03 8	0.0 75	- 0.0 04	0.14 4	0.0 03	- 0.0 38	- 0.0 85	- 0.0 40	- 0.0 25	- 0.0 75	1.0 00								
(12) ln_DIA	0.52 7	0.3 15	0.1 19	0.16 4	- 0.0 88	- 0.3 76	0.1 14	0.4 50	0.2 12	0.0 67	0.1 08	1.0 00							
(13) ln_EXR	0.15 4	0.0 10	- 0.0 83	- 0.14 8	0.0 46	- 0.1 47	0.0 90	- 0.0 65	- 0.2 71	- 0.2 50	0.1 08	0.2 33	1.0 00						
(14) INF	- 0.02 4	- 0.0 74	- 0.1 17	0.02 3	0.0 59	- 0.0 81	- 0.0 61	- 0.0 22	- 0.0 16	- 0.0 32	0.0 99	- 0.0 13	- 0.0 28	1.0 00					
(15) CEFTA	0.00 6	- 0.2 46	0.2 22	- 0.16 2	- 0.2 49	- 0.3 86	0.1 18	0.3 22	0.2 59	0.1 30	- 0.0 35	0.1 04	- 0.1 42	- 0.0 12	1.0 00				
(16) SAA_eu	0.13 5	0.0 68	0.4 95	- 0.09 0	- 0.6 83	- 0.0 72	0.1 31	- 0.0 48	- 0.0 75	0.0 01	- 0.0 75	0.0 73	0.1 59	- 0.0 52	- 0.1 05	1.0 00			
(17) FTA_eft a	- 0.02 9	0.0 54	0.1 20	- 0.04 0	- 0.1 75	0.0 33	0.3 68	- 0.0 28	- 0.0 18	0.0 43	- 0.0 18	0.0 35	0.0 33	0.0 16	0.0 24	0.0 43	1.0 00		
(18) FTA_tur	0.05 1	0.0 68	0.0 95	0.08 6	- 0.1 18	- 0.0 22	- 0.0 48	- 0.0 23	- 0.0 14	- 0.0 43	0.5 73	0.0 75	0.0 23	0.0 03	- 0.0 20	- 0.0 43	- 0.0 10	1.0 00	
(19) INST_dist	- 0.12 6	0.1 76	- 0.3 11	- 0.23 4	0.2 80	0.2 76	0.4 31	- 0.2 33	- 0.1 90	0.0 48	0.1 83	0.0 22	0.2 10	- 0.1 05	- 0.2 54	- 0.1 04	0.0 76	0.1 08	1.000

Source: Own elaboration

Comparison of Integrated Variance Forecasts

Vladimír Holý

University of Economics, Prague, Czech Republic, vladimir.holy@vse.cz

Abstract

Daily risk of a financial asset can be measured by integrated variance. This unobserved measure can be estimated from high-frequency prices by many model-free methods in the presence of market microstructure noise. To forecast future integrated variance we need to model dynamics of its estimates (i.e. realized measures). We compare prediction accuracy of several models including popular HAR model and realized GARCH model using one-minute prices of selected stocks.

Keywords: *high-frequency data, integrated variance, market microstructure noise, forecasting.*

Introduction

The term ultra-high-frequency data or simply high-frequency data refers to financial data that consist of observations recorded within seconds or even fractions of seconds (Engle, 2000). A risk of a financial assets such as stocks or currencies can be measured by integrated variance. However, integrated variance is unobserved variable and needs to be estimated. A simple estimator is realized variance, but it can be significantly biased because of so-called market microstructure noise contained in financial high-frequency data. Jacod et al. (2009) propose Pre-averaging estimator which is robust to market microstructure noise under general noise structure. Another problem of integrated variance is its forecasting. For predicting realized variance Corsi (2009) proposes HAR model. For jointly modeling price returns, unobserved volatility and integrated variance estimates Hansen et al. (2012) propose Realized GARCH model.

A wide range of literature is devoted to high-frequency data analysis. Biais et al. (2005) surveys literature about market microstructure, price formation and trading process. McAleer and Medeiros (2008) review so-called realized volatility literature about estimating, modelling and forecasting integrated variance. Holý (2017) surveys a literature about causes, effects and solutions of market microstructure noise contained in high-frequency financial data.

This paper is organized as follows. Section 2 presents general theoretical framework of efficient prices, market microstructure noise, integrated variance and its estimators. Section 3 describes ARIMA, HAR and realized GARCH models of integrated variance estimations used for forecasting. Section 4 studies empirical properties of the models on S&P 500 stock prices. Section 5 concludes with brief discussion about the results.

Integrated Variance Estimation

We describe standard theoretical framework for prices of financial assets as presented for example in Ait-Sahalia and Jacod (2014) or Hautsch (2011). The goal of high-frequency data analysis is to examine logarithmic efficient prices p_t that contain the information about prices of an asset. However, this underlying latent price process is unobservable due to market microstructure noise e_t and we observe only noisy prices

x_t . The market microstructure noise can have various sources such as bid-ask bounce, discreteness of prices and informational effects.

2.1 Efficient Price and Market Microstructure Noise

Observed logarithmic prices x_t are modeled in commonly used additive noise setting

$$x_t = p_t + e_t$$

where p_t is the unobservable efficient price process and e_t is the unobservable market microstructure noise. Furthermore, we denote

$$y_{s,t} := x_t - x_s, \quad r_{s,t} := p_t - p_s, \quad u_{s,t} := e_t - e_s.$$

Under the assumption of no arbitrage, the logarithmic efficient returns $r_{s,t}$ must follow a semi-martingale (Delbaen and Schachermayer, 1994). Let returns $r_{s,t}$ follow a continuous semi-martingale

$$r_{s,t} = \int_s^t \mu(\tau) d\tau + \int_s^t \sigma(\tau) dW_\tau$$

where $\mu(\tau)$ is a finite variation càglàg drift process, $\sigma(\tau)$ is an adapted càdlàg volatility process and W_τ denotes a standard Wiener process. Market microstructure noise e_t can have a very rich structure. Hansen and Lunde (2006) show that the noise present in financial data can be dependent in time as well as dependent on efficient price.

2.2 Integrated Variance, Realized Variance and Pre-Averaging Estimator

For a time interval (a, b) integrated variance is defined as

$$IV_{a,b} := \int_a^b \sigma^2(\tau) d\tau.$$

A natural estimator of integrated variance is realized variance. For prices sampled at times $a = t_0, t_1, \dots, t_n = b$ realized variance is defined as

$$RV_{a,b}^n := \sum_{i=1}^n y_{t_{i-1}, t_i}^2$$

However, in the presence of market microstructure noise realized variance can be significantly biased. When assuming market microstructure noise to be independent white noise the bias is given by

$$E[RV_{a,b}^n] = IV_{a,b} + 2n\omega^2,$$

i.e. realized variance linearly diverges to infinity for $n \rightarrow \infty$. This is caused by the fact that integrated variance is of the same order of magnitude as the time interval while microstructure noise has a roughly constant variability. When the noise is time-

dependent or cross-dependent the bias has more complex structure and can even be negative (Hansen and Lunde, 2006).

To consistently estimate integrated variance we use Pre-averaging estimator proposed by Jacod et al. (2009) with extensions by Hautsch and Podolskij (2013) and Jacod and Mykland (2015). It removes the noise by locally averaging returns before computing realized variance.

However, many other estimators robust to market microstructure noise such as Two-scale realized volatility (Zhang et al., 2005), Realized kernels (Barndorff-Nielsen et al., 2008) or Least squares estimator (Nolte and Voev, 2012) can be used.

Realized Measure Models

Dynamics of integrated variance estimates (i.e. realized measures) can be captured by several models. We estimate integrated variance by Pre-averaging estimator described in Section 2.2. We present three models of realized measures. ARIMA and HAR models are univariate time series models for realized measures without any other variables. However, these models can be extended by adding equation which relates realized measure RM_t to observed returns y_t (Aït-Sahalia and Mancini, 2008). This equation is given by

$$y_t = \mu + \sqrt{RM_t} z_t$$

where z_t are i.i.d. $N(0,1)$. Realized GARCH model has a more complex structure and directly captures relations between observed returns, unobserved volatility and realized measure.

3.1 ARIMA Model

Time series of daily realized measure RM_t can be modeled as ARIMA(p, d, q) process given by

$$\Delta^d RM_t = \sum_{i=1}^p \varphi_i \Delta^d RM_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + v_t$$

where v_t are i.i.d. $N(0, \sigma^2)$ and Δ^d denotes the d -th difference. For the selection of p , d and q parameters we use the framework of Hyndman and Khandakar (2008). Andersen et al. (2003) argued that the distribution of logarithm of realized variance can be closer to normal distribution than the distribution of realized variance. For that reason we also describe logarithmic ARIMA(p, d, q) model given by

$$\ln \Delta^d RM_t = \sum_{i=1}^p \varphi_i \ln \Delta^d RM_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + v_t$$

where v_t are i.i.d. $N(0, \sigma^2)$ and Δ^d denotes d -th difference.

3.2 HAR Model

Corsi (2009) proposes to model daily realized variances by HAR model. It is an autoregressive-type model featuring realized variances over different time horizons motivated by market agents operating at different frequencies (e.g. daily or monthly). Busch et al. (2011) further extended this model to vector HAR model. A standard HAR model explaining daily realized measure by realized measure of last day, week and month as proposed by Corsi (2009) is given by

$$RM_t = \beta_0 + \beta_1 RM_{t-1} + \beta_2 \frac{1}{5} \sum_{i=1}^5 RM_{t-i} + \beta_3 \frac{1}{22} \sum_{i=1}^{22} RM_{t-i} + v_t$$

where v_t are i.i.d. $N(0, \sigma^2)$. Similary as in Section 3.2 we modify this model to feature logarithms of realized measures resulting in logarithm HAR model given by

$$\ln RM_t = \beta_0 + \beta_1 \ln RM_{t-1} + \beta_2 \frac{1}{5} \sum_{i=1}^5 \ln RM_{t-i} + \beta_3 \frac{1}{22} \sum_{i=1}^{22} \ln RM_{t-i} + v_t$$

where v_t are i.i.d. $N(0, \sigma^2)$.

Realized GARCH Model

Hansen et al. (2012) propose Realized GARCH(p, q) to jointly model observed returns y_t , unobserved volatility h_t and realized measure RM_t . It is a modification of classical GARCH model where lagged realized variances are used instead of lagged errors and the measurement equation of realized measure RM_t is added. The whole joint model is given by

$$\begin{aligned} y_t &= \mu_t + \sqrt{h_t} z_t \\ \ln h_t &= \omega + \sum_{i=1}^p \beta_i \ln h_{t-i} + \sum_{i=1}^q \gamma_i \ln RM_{t-i} \\ \ln RM_t &= \xi + \varphi \ln h_t + \tau(z_t) + v_t \end{aligned}$$

where z_t are i.i.d. $N(0,1)$, v_t are i.i.d. $N(0, \sigma^2)$ and $\tau(z_t) = \eta_1 z_t + \eta_2 (z_t^2 - 1)$ which has a property $E[\tau(z_t)] = 0$. We use the $p = 1$ and $q = 1$ specification of Realized GARCH as is common (Hansen et al., 2012).

Results of Model Forecasts

We compare presented models on 245 stocks featured in S&P 500 index. We have 1-minute transaction data available from January 2014 to November 2016. For each day from February 2015 to November 2016 we construct presented models using history of one year, then forecast realized measure (pre-averaging estimate) one day ahead and finally compare it with real realized measure value. As a benchmark, we also include in our analysis a naive model which predict future realized measure to be the same as present realized measure.

Table 1 reports the results of daily forecasts for each model. We measure the deviation of forecasts from real values by mean absolute error (MAE) and root mean squared error (RMSE). Based on MAE we can see that logarithmic versions of ARIMA/HAR models are more precise than regular ARIMA/HAR models. Logarithmic HAR has the lowest MAE while logarithmic ARIMA and Realized GARCH are second and third. RMSE is much more sensitive to outliers and gives a different results. Lowest error is again achieved by HAR model but this time without logarithmic transformation. Realized GARCH has rather high RMSE which may indicate poor robustness to outliers. We also analyze stocks separately and determine which model gives the best forecast (with lowest error). Logarithmic HAR have the lowest MAE for 57.6% symbols while logarithmic ARIMA have 26.5% and Realized GARCH have 15.5%. Similar analysis was performed for each day (daily error is an average of daily errors of all stocks). Realized GARCH have the lowest MAE for 25.8% days while logarithmic HAR have 25.0% and logarithmic ARIMA have 24.9%.

Table 1: Mean absolute errors (MAE) and root mean squared errors (RMSE) for model forecasts in ten-thousandths. Best symbol and best day denote ratio of best prediction among all forecasts for specific symbol/day.

	MAE			RMSE		
	Value	Best Symbol	Best Day	Value	Best Symbol	Best Day
Naive	1.1693	0.0000	0.0576	5.5904	0.1796	0.1643
ARIMA	1.0980	0.0041	0.0744	5.5262	0.2898	0.1671
LogARIMA	1.0416	0.2653	0.2486	5.5688	0.0041	0.1559
HAR	1.0973	0.0000	0.1110	5.5246	0.5224	0.1952
LogHAR	1.0390	0.5755	0.2500	5.5731	0.0041	0.1447
RealGARCH	1.0451	0.1551	0.2584	5.5959	0.0000	0.1728

Discussion

We analyze one-day ahead forecasting performance of ARIMA, HAR and Realized GARCH models using 1-minute prices of 245 stocks from January 2014 to November 2016. We find that HAR model with logarithmic transformation has the lowest MAE. Realized GARCH have slightly higher MAE and is less robust to outliers indicated by high RMSE. Based on MAE we recommend to use logarithmic HAR model for integrated variance forecasting.

Acknowledgements

This paper has been prepared under the support of the project of the University of Economics, Prague - Internal Grant Agency, project No. F4/63/2016.

References

- Aït-Sahalia, Y., & Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press. <http://doi.org/10.1515/9781400850327>
- Aït-Sahalia, Y., & Mancini, L. (2008). Out of Sample Forecasts of Quadratic Variation. *Journal of Econometrics*, 147(1), 17–33. <http://doi.org/10.1016/j.jeconom.2008.09.015>
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2), 579–625. <http://doi.org/10.1111/1468-0262.00418>
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008). Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise. *Econometrica*, 76(6), 1481–1536. <http://doi.org/10.3982/ecta6495>
- Biais, B., Glosten, L., & Spatt, C. (2005). Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications. *Journal of Financial Markets*, 8(2), 217–264. <http://doi.org/10.1016/j.finmar.2004.11.00>
- Busch, T., Christensen, B. J., & Nielsen, M. O. (2011). The Role of Implied Volatility in Forecasting Future Realized Volatility and Jumps in Foreign Exchange, Stock, and Bond Markets. *Journal of Econometrics*, 160(1), 48–57. <http://doi.org/10.1016/j.jeconom.2010.03.014>
- Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2), 174–196. <http://doi.org/10.1093/jjfinec/nbp001>
- Delbaen, F., & Schachermayer, W. (1994). A General Version of the Fundamental Theorem of Asset Pricing. *Mathematische Annalen*, 300(1), 463–520. <http://doi.org/10.1007/bf01450498>
- Engle, R. F. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometrica*, 68(1), 1–22. <http://doi.org/10.1111/1468-0262.00091>
- Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. *Journal of Applied Econometrics*, 27(6), 877–906. <http://doi.org/10.1002/jae.1234>
- Hansen, P. R., & Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics*, 24(2), 127–161. <http://doi.org/10.1198/073500106000000071>
- Hautsch, N. (2011). *Econometrics of Financial High-Frequency Data*. Springer. <http://doi.org/10.1007/978-3-642-21925-2>
- Hautsch, N., & Podolskij, M. (2013). Preaveraging-Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence. *Journal of Business & Economic Statistics*, 31(2), 165–183. <http://doi.org/10.1080/07350015.2012.754313>
- Holý, V. (2017). Market Microstructure Noise. In P. Doucek (Ed.), *Sborník prací účastníků vědeckého semináře doktorandského studia Fakulty informatiky a statistiky VŠE v Praze*, 96–104. Oeconomica. <https://fis.vse.cz/studium/doktorske-studium/den-doktorandu/den-doktorandu-2017>

- Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3), 1–22. <http://doi.org/10.18637/jss.v069.i12>
- Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., & Vetter, M. (2009). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. *Stochastic Processes and Their Applications*, 119(7), 2249–2276. <http://doi.org/10.1016/j.spa.2008.11.004>
- Jacod, J., & Mykland, P. A. (2015). Microstructure Noise in the Continuous Case: Approximate Efficiency of the Adaptive Pre-Averaging Method. *Stochastic Processes and Their Applications*, 125(8), 2910–2936. <http://doi.org/10.1016/j.spa.2015.02.005>
- McAleer, M., & Medeiros, M. C. (2008). Realized Volatility: A Review. *Econometric Reviews*, 27(1–3), 10–45. <http://doi.org/doi.org/10.1080/07474930701853509>
- Nolte, I., & Voev, V. (2012). Least Squares Inference on Integrated Volatility and the Relationship Between Efficient Prices and Noise. *Journal of Business & Economic Statistics*, 30(1), 94–108. <http://doi.org/10.1080/10473289.2011.637876>
- Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *Journal of the American Statistical Association*, 100(472), 1394–1411. <http://doi.org/10.2307/27590680>

Determinants of Trade and the Gravity Model of Trade: the Case of Western Balkan Countries

Visar Malaj

*Department of Economics, Faculty of Economics, University of Tirana Albania
,visarmalaj@feut.edu.al*

Abstract

The magnitude of international trade flows may be affected by social and cultural variables, such as, population structure, common language and colonial links; economic and political variables, such as, economic sizes or incomes, trade costs, trade agreements, exchange rate and relative prices; technical variables, such as, technology advancement, infrastructure condition and geographical distance; and some more complex variables to be anticipated, such as, political conflicts, meteorological conditions and natural catastrophes. The main objective of this work is the empirical analysis of the determinants of bilateral trade between Western Balkan countries and the most important partners, through the gravity theory. We propose a particular gravity equation for trade, including basic and experimental independent variables. We made use of R software for the econometric analysis, considering a panel data estimator for the dependent variable. The statistical effect of the considered explanators is generally confirmed and the resulting adjusted R-squared is relatively high.

JEL Classification: F14, C23, C80

Introduction

The annual increase of world merchandise trade volume in 2014 was 2,5%, less than the last 20-years average of 5,3%. During the time period 2012-2014, world trade volume increased by only 2,4%, suffering the worst performance since the economic and financial crisis of 2008-2009. In the first quarter of 2015, trade continued to grow more slowly; during this period, world trade volume increased by 0,7% compared with the first quarter of 2014, whereas the growth in the fourth quarter of 2014 was 1,8%. This was a consequence of various factors, such as, the decreasing import demand in developed countries, the slow import increase in developing countries, and the insufficient growth of global exports. Total exports of developing countries, including the Western Balkan (WB) region, rose faster than those of developed economies in 2014, 3,1% and 2%, respectively. On the other hand, imports of developing economies increased more slowly than those of developed countries, 1,8% and 2,9%, respectively.³

³ World Trade Report, 2015.

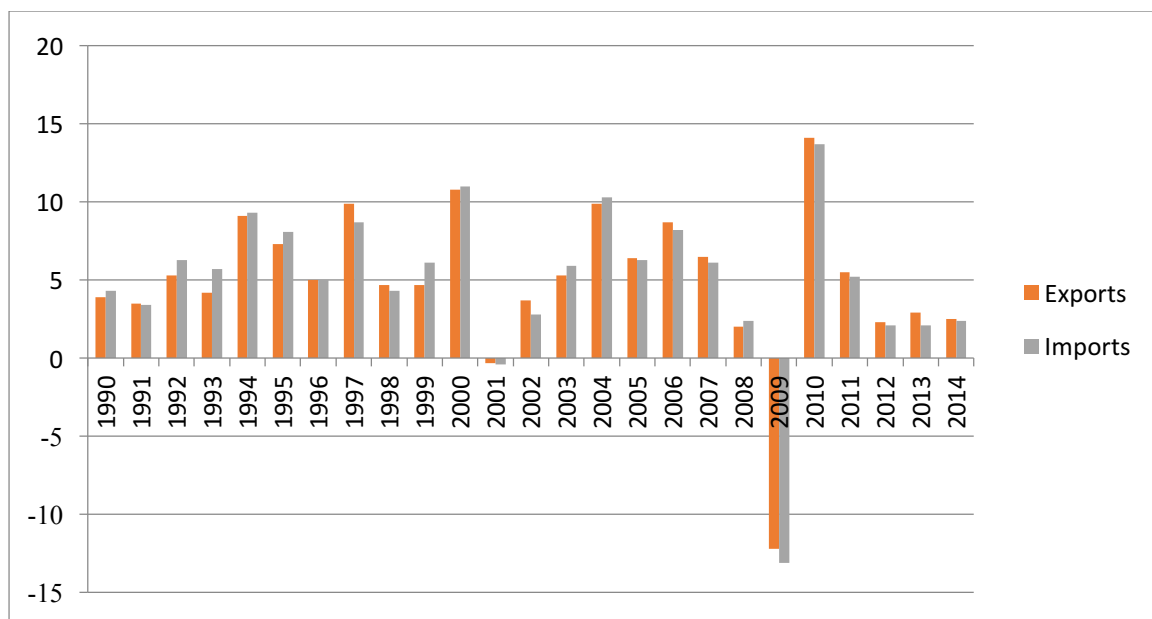


Figure 1. Annual percentage growth rate (base: previous year) of world merchandise trade (imports and exports) volume for the time period 1990-2014.
[Source: Author's elaboration with data from the World Trade Organization (WTO).]

The main objective of this work is the theoretical and empirical analysis of the determinants of bilateral trade between WB countries and the most important partners, through the gravity theory. In the following section, we report some relevant facts and figures regarding international trade and the corresponding determinants in WB countries. Section 3 is dedicated to the theoretical definition of the gravity model of trade. In section 4 we estimate several gravity equations using an appropriate dataset, which includes bilateral trade flows between WB countries and their main partners, and a set of determinants. Concluding remarks are reported in section 5.

Western Balkan countries: trade facts and figures

The WB region is an important part of Eastern Europe. It includes Albania, Macedonia, Bosnia and Herzegovina, Serbia, Montenegro and Kosovo. Croatia is also considered part of the region, but we excluded it from our analysis, mainly because it is a current EU member state and it is characterized by a higher social and economic development compared to the other countries. GDP growth in 2014-2015 was below the expectations in WB countries, also due to the significant decline in commodity prices and to the low level of foreign direct investments. All WB countries ran a current account deficit in 2014. Serbia was the country with the largest deficit (-2,755 billion US\$), while Macedonia recorded the highest current account balance in the WBs (-160 million US\$).

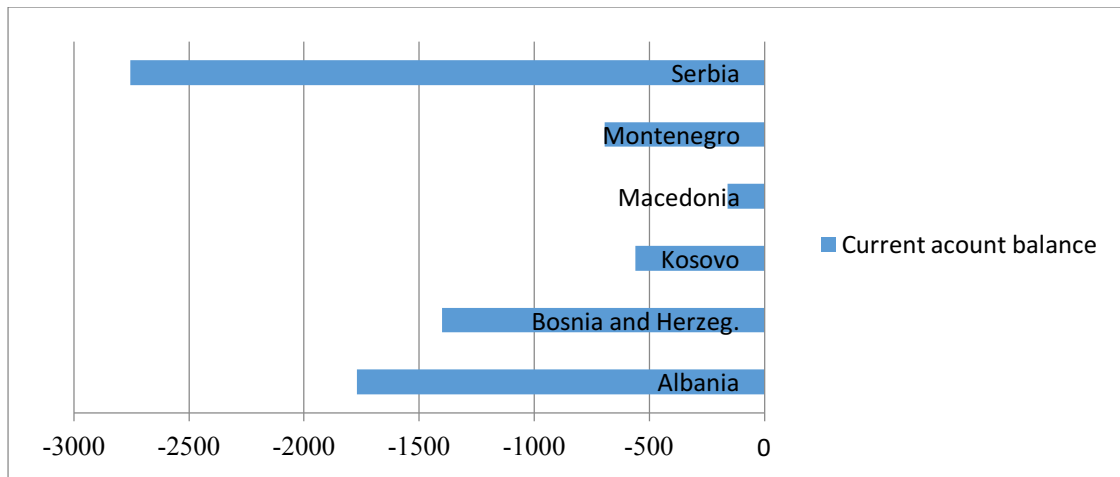


Figure 2. Current account balance (million US\$) in 2014 for Western Balkan countries. [Source: Author's elaboration on World Trade Organization (WTO) and Central Bank of the Republic of Kosovo (BQK) data.]

WB countries have signed several bilateral and multilateral agreements with important EU and extra-EU partners, such as the free trade agreements between Albania, Bosnia and Herzegovina, Macedonia, Serbia and Turkey; Russia, Belarus and Serbia; Albania, Bosnia and Herzegovina, Macedonia, Montenegro, Serbia and the EFTA countries. Croatia, Bosnia and Herzegovina, Albania, Serbia, Moldova, Montenegro, Kosovo and Macedonia signed in December 2006 the Central European Free Trade Agreement (CEFTA), a trade agreement between countries that are not yet EU members. In 2007, CEFTA entered into force for all WB countries. This agreement is considered as an important step before the accession into the EU. CEFTA members commit to reduce trade barriers and to cooperate in order to intensify the bilateral exchanges of goods and services. This agreement is based on WTO rules and procedures and on the EU regulations.

Albania

In June 2006, Albania signed the Interim Agreement on trade and trade-related issues and the SAA with the EU, which entered into force, respectively in December 2006 and in April 2009. The EU countries granted Albania the candidate status in June 2016. Albania has been a World Trade Organization (WTO) member since 2000. Albania's main import partners are the EU and China with respectively 61,1% and 7,3% of total merchandise imports in 2014. The top export destinations of Albania are the EU and Serbia with respectively 77,4% and 8% of total merchandise exports in 2014.⁴

Bosnia and Herzegovina

In July 2008, the Interim Agreement on trade and trade-related issues between Bosnia and Herzegovina and the EU entered into force. In June 2012, EU and Bosnia and Herzegovina started the High-Level Dialogue on the Accession Process. The SAA between Bosnia and Herzegovina and the EU entered into force in June 2015. Bosnia and Herzegovina is continuing the accession negotiations with the WTO, after its application for membership in 1999. The EU and Serbia are the largest import partners of Bosnia and Herzegovina with respectively 58,9% and 10,1% of total merchandized imports in 2014. The EU and Serbia are also the main export destinations for Bosnia

⁴Data from the World Trade Organization (WTO), <http://www.wto.org/>

and Herzegovina with respectively 72,1% and 9,2% of total merchandized exports in 2014.⁵

Kosovo

Kosovo is actually a potential candidate for the accession to the EU. In February 2008, the Assembly of Kosovo declared independence in an extraordinary session. In October 2015, Kosovo signed the SAA with the EU, an important incentive for the implementation of reforms. Kosovo's application and access into the WTO would provide an important stimulus for the international trade. The EU and Serbia are the largest import partners of Kosovo with respectively 42,6% and 14,5% of total merchandized imports in 2014. The main export destinations are the EU and Albania with respectively 30,2% and 13,6% of total merchandized exports in 2014.⁶

Macedonia

The Agreement on trade and trade-related matters and the SAA between the EU and the Former Yugoslav Republic of Macedonia entered in force respectively in 2001 and in 2004. One year ahead, the European Commission accorded the candidate status to Macedonia, fourteen years after the declaration of independence from Ex-Yugoslavia. In 2012, the European Commission initiated a High-Level Accession Dialogue with the Macedonian authorities, which will help to accelerate public administration and electoral reforms, contribute to the protection of minorities' rights, promote market competition and encourage economic growth. Macedonia joined the WTO in 2003. Macedonia's main import partners are the EU and Serbia with respectively 63,5% and 8,7% of total merchandise imports in 2014. The EU and Serbia are also the top export destinations of Macedonia with respectively 76,6% and 9,9% of total merchandise exports in 2014.

Montenegro

Montenegro is an independent state since the dissolution of the Union of Serbia and Montenegro; Montenegro citizens voted for the separation from Serbia in the referendum of May 2006. After this, Montenegro reopened new accession negotiations with the EU; in 2007, the country signed the SAA, which entered into force in May 2010. In January 2008, agreements on trade and trade-related matters, visa facilitation and readmission between Montenegro and the EU entered into force. Montenegro joined the WTO in 2012. Montenegro's main import partners are the EU and Serbia with respectively 45,8% and 26,9% of total merchandise imports in 2014. The EU and Serbia are also the main export destinations of Montenegro with respectively 35,8% and 24% of total merchandise exports in 2014.

Serbia

Serbian government signed the Stabilisation and Association Agreement (SAA) and the Interim agreement on trade and trade-related issues with the EU in April 2008. These agreements are considered as important steps before the accession to the EU. The SAA between Serbia and the EU entered into force in September 2013. In January 2014, the European Council initiated the official accession negotiations with Serbia. Negotiations between WTO and Serbia are still in progress since 2004, when Serbia submitted its membership application. The EU and the Russian Federation are the largest import partners of Serbia with respectively 63,1% and 11,4% of total merchandized imports

⁵Data from the World Trade Organization (WTO), <http://www.wto.org/>.

⁶Data from the Kosovo Agency of Statistics (ASK), available at: <https://ask.rks-gov.net/>

in 2014. The EU and Bosnia and Herzegovina are the main export destinations with respectively 64,6% and 8,9% of total merchandized exports in 2014.

Gravity theory: theoretical foundations

Anderson (1979) laid the foundations of the gravity theory. The author considered two countries (country i and country j) and the corresponding bilateral trade of two differentiated products. Prices are in equilibrium and preferences are assumed to be Cobb-Douglas anywhere. With frictionless trade, the respective equation is as follows.

$$X_{ij} = b_i Y_j \quad (1)$$

where X_{ij} are the exports from i to j , b_i is country i 's share of expenditure on tradable goods and Y_j is country j 's income. The budget constraint is

$$Y_i = b_i \sum_j Y_j \quad (2)$$

Substituting b_i into equation (1), we obtain the following gravity equation.

$$X_{ij} = \frac{Y_i Y_j}{\sum_j Y_j} \quad (3)$$

Anderson extended the analysis, supposing that every country produces two types of goods, tradable and non-tradable, respectively. In this case, the exports from i to j are

$$X_{ij} = \Phi_j \theta_i Y_j \quad (4)$$

where Φ_j is country j 's share of expenditure on tradable goods and θ_i is country i 's share of tradable goods in country j 's total expenditure. The corresponding budget constraint is

$$Y_i \Phi_i = \theta_i \sum_j Y_j \Phi_j \quad (5)$$

Substituting θ_i into equation (4), we obtain

$$X_{ij} = \frac{Y_i \Phi_i Y_j \Phi_j}{\sum_j Y_j \Phi_j} = \frac{Y_i \Phi_i Y_j \Phi_j}{\sum_i \sum_j X_{ij}} \quad (6)$$

Φ_i depends on country i 's income and population and on other variables. We have:

$$\Phi_i = F(Y_i, N_i) \quad (7)$$

We can rewrite equation (6) as

$$X_{ij} = A \frac{Y_i F(Y_i, N_i) Y_j F(Y_j, N_j)}{\sum_j Y_j F(Y_j, N_j)} U_{ij} \quad (8)$$

where A is a constant and U is an error term that satisfies:

$$E(\ln U_{ij}) = 0 \quad (9)$$

Supposing that $F(\cdot)$ is linear, we have that

$$X_{ij} = \left(\frac{A}{K}\right) Y_i^{\alpha_1} N_i^{\alpha_2} Y_j^{\beta_1} N_j^{\beta_2} U_{ij} \quad (10)$$

where

$$K = \sum_j Y_j F(Y_j, N_j) \quad (11)$$

Applying a logarithmic transformation to equation (10), we obtain

$$\ln X_{ij} = \ln \left(\frac{A}{K} \right) + \alpha_1 \ln Y_i + \alpha_2 \ln N_i + \beta_1 \ln Y_j + \beta_2 \ln N_j + \ln U_{ij} \quad (12)$$

Subsequently, Anderson considered trade barriers in the model. τ_{ij} is a transport factor and $X_{ij}\tau_{ij}$ is country j 's export value from country i .

$$X_{ij} = \frac{1}{\tau_{ij}} \Phi_j \theta_i(\tau_j) Y_j \quad (13)$$

The corresponding budget constraint is

$$Y_i \Phi_i = \theta_i(\tau_j) \sum_j \frac{1}{\tau_{ij}} Y_j \Phi_j \quad (14)$$

Substituting $\theta_i(\tau_j)$ into equation (13), we obtain

$$X_{ij} = \frac{1}{\tau_{ij}} \Phi_j \frac{Y_i \Phi_i}{\sum_j \frac{1}{\tau_{ij}} Y_j \Phi_j} Y_j = D_{ij} \frac{Y_i \Phi_i \Phi_j Y_j}{\sum_j D_{ij} Y_j \Phi_j} \quad (15)$$

where

$$D_{ij} \equiv \frac{1}{\tau_{ij}} \quad (16)$$

Anderson also supposed that there exist two types of goods, tradable and non-tradable and every country produce multiple differentiated traded goods classified by the product class k . X_{ijk} is the export volume for good k from country i to country j . We have that

$$X_{ijk} = \frac{1}{\tau_{ijk}} \Phi_j \theta_{ik}(\tau_j) Y_j \quad (17)$$

and

$$X_{ij} = \sum_k X_{ijk} = \Phi_j Y_j \sum_k \frac{1}{\tau_{ijk}} \theta_{ik}(\tau_j) \quad (18)$$

In this case, the budget constraint is

$$Y_i \Phi_i = \left(\sum_k \frac{1}{\tau_{ijk}} \theta_{ik}(\tau_j) \right) \sum_j Y_j \Phi_j \quad (19)$$

Supposing that transport cost depends only on distance, we have that

$$X_{ij} = \left(\sum_k \theta_{ik} \right) \Phi_j Y_j \frac{1}{f(d_{ij})} \quad (20)$$

$$Y_i \Phi_i = \left(\sum_k \theta_{ik} \right) \sum_j Y_j \Phi_j \frac{1}{f(d_{ij})} \quad (21)$$

Substituting $\sum_k \theta_{ik}$ into equation (20), we obtain the gravity equation

$$X_{ij} = \frac{Y_i \Phi_i \Phi_j Y_j}{\sum_j \Phi_j Y_j} \frac{1}{f(d_{ij})} \left[\sum_j \frac{Y_j \Phi_j}{\sum_j \Phi_j Y_j} \frac{1}{f(d_{ij})} \right]^{-1} \quad (22)$$

According to Anderson, the first term in equation (22) represents the economic distance between country i and country j , while the second term is an indicator of the economic distance between country i and the other trading partners with respect to world trade. Anderson also considered constant elasticity of substitution (CES) preferences, including prices into the share of expenditure on tradable goods function. A considerable number of authors followed similar approaches to define the gravity equation (Bergstrand, 1985; Krugman 1979, 1980; Helpman and Krugman, 1985; Helpman, 1987; Bergstrand, 1989). The number of empirical applications of the gravity model to WB trade flows is relatively low. Findings have shown the statistical significance of traditional and new variables, and unexploited trade potentials for most of the WB countries (Christie, 2002; Bussière, Fidrmuc & Schnatz, 2005; Montanari, 2005; Josheski & Apostolov, 2013; Toševska-Trpčevska & Tevdovski, 2014).

Empirical application

We propose a gravity equation for WB international trade flows, including basic and new independent variables. The basic gravity equation includes the economic sizes and the corresponding distance, which approximates the trade costs in our case. GDPs or GDPs per capita are common estimators of the income level and the economic development. So, let us consider the following gravity model of trade, where all variables are expressed in natural logarithm and ε is an error term that follows a known distribution with a mean of zero and constant variance:

$$\begin{aligned} flow_odt = & \alpha_1 + \alpha_2(gdp_ot) + \alpha_3(gdp_dt) + \alpha_4(dist_od) + \alpha_5(pop_sum_odt) \\ & + \alpha_6(BORDER_od) + \alpha_7(FTA_od) + \varepsilon_odt \end{aligned} \quad (45)$$

Table 1 reports the definition and the expected sign, or the corresponding expected effect on trade flows, for each variable. Trade flows and GDPs are expressed in US dollars. We consider three possible dependent variables: import in a WB country from a given trade partner (imp_odt), export of a WB country in a given trade partner (exp_odt), and the sum of imports and exports between a WB country and the corresponding trade partner ($flow_odt$).

Table 1. Variables definition (expressed in natural logarithm) and expected sign.

Variable	Definiton	Expected sign
$flow_odt$	Trade flow (the sum of imports and exports) between a WB country (o) and a partner (d) at year t	Dependent variable
gdp_ot	GDP in country o at year t	+
gdp_dt	GDP in country d at year t	+
$dist_od$	Bilateral distance	-
pop_sum_odt	Sum of populations in country o and in country d at year t	+

<i>BORDER</i> <i>_od</i>	Dummy variable equal to one if countries share a common border, and zero vice versa	+
<i>FTA_od</i>	Dummy variable equal to one if countries have signed a common free trade agreement, and zero vice versa	+

We have built a dataset that includes annual trade flows between WB countries and the most important trade partners. We have considered fifteen partners for each of the six WB countries, for the time period 2010-2014. In table 2 we have listed the considered country pairs for the gravity model estimation.

Table 2. Considered trade partners for each Western Balkan country.

<i>Albania</i>	<i>Bosnia Herz.</i>	<i>Kosovo</i>	<i>Macedonia</i>	<i>Montenegro</i>	<i>Serbia</i>
<i>Austria</i>	<i>Austria</i>	<i>Austria</i>	<i>Austria</i>	<i>Austria</i>	<i>Austria</i>
<i>Belgium</i>	<i>Belgium</i>	<i>Belgium</i>	<i>Belgium</i>	<i>China</i>	<i>Belgium</i>
<i>Bulgaria</i>	<i>Bulgaria</i>	<i>Bulgaria</i>	<i>Bulgaria</i>	<i>Croatia</i>	<i>Bosnia Herz.</i>
<i>China</i>	<i>China</i>	<i>Croatia</i>	<i>Croatia</i>	<i>Denmark</i>	<i>China</i>
<i>Croatia</i>	<i>Croatia</i>	<i>Czech Rep.</i>	<i>Czech Rep.</i>	<i>Germany</i>	<i>Czech Rep.</i>
<i>Czech Rep.</i>	<i>Czech Repub.</i>	<i>Denmark</i>	<i>Denmark</i>	<i>Greece</i>	<i>France</i>
<i>Denmark</i>	<i>Denmark</i>	<i>Germany</i>	<i>France</i>	<i>Hungary</i>	<i>Germany</i>
<i>France</i>	<i>France</i>	<i>Greece</i>	<i>Germany</i>	<i>Italy</i>	<i>Greece</i>
<i>Germany</i>	<i>Germany</i>	<i>Hungary</i>	<i>Greece</i>	<i>Macedonia</i>	<i>Hungary</i>
<i>Greece</i>	<i>Greece</i>	<i>Italy</i>	<i>Italy</i>	<i>Romania</i>	<i>Italy</i>
<i>Italy</i>	<i>Italy</i>	<i>Romania</i>	<i>Romania</i>	<i>Russia</i>	<i>Romania</i>
<i>Macedonia</i>	<i>Russia</i>	<i>Slovenia</i>	<i>Serbia</i>	<i>Serbia</i>	<i>Russia</i>
<i>Serbia</i>	<i>Slovenia</i>	<i>Spain</i>	<i>Spain</i>	<i>Slovenia</i>	<i>Slovenia</i>
<i>Spain</i>	<i>Spain</i>	<i>Turkey</i>	<i>Turkey</i>	<i>Spain</i>	<i>Spain</i>
<i>Turkey</i>	<i>Turkey</i>	<i>UK</i>	<i>UK</i>	<i>Turkey</i>	<i>Turkey</i>

The databases of the United Nations (Comtrade) and CEPII (CHELEM-INT) were our primary sources of bilateral trade flows. GDPs and the corresponding population data were obtained by the World Bank. Bilateral distances between origin and destination countries were collected from CEPII GeoDist database (Mayer, T. and Zignago, S., 2011). Table 3 shows the main descriptive statistics for the gravity model variables.

Table 3. Main descriptive statistics for the gravity model variables.

Variable	Imp	Exp	Flow	gdp_o	gdp_d	Dist	pop_sum	BORDER	FTA
<i>MIN</i>	2184083	8131	3548440	4,09E+09	9,4E+09	156	2623435	0	0
<i>MAX</i>	2,7E+09	2,577E+09	4,89E+09	4,65E+10	1E+13	7686,1	1,37E+09	1	1
<i>Mean</i>	3,5E+08	195524404	5,48E+08	1,59E+10	1,3E+12	1248,2	94717264	0,12	0,73
<i>Median</i>	1,5E+08	47028977	2,09E+08	1,16E+10	4,3E+11	779,1	16333912	0	1
<i>Variance</i>	2,23E+17	1,30E+17	6,15E+17	1,70E+20	3,63E+24	2180952	7,52E+16	0,11	0,2
<i>Std.Deviation</i>	4,7E+08	360772452	7,84E+08	1,3E+10	1,9E+12	1476,8	2,74E+08	0,33	0,44
<i>Asymmetry</i>	2,34	2,97	2,48	1,4	2,68	3,49	4,323881	2,31	-1,1
<i>Kurtosis</i>	5,89	10,72	7,09	0,6	8,22	12,46	17,11251	3,37	-0,9

We report in table 4 the original output of the estimated models, obtained by the R software. We consider three dependent variables: imports, exports and trade flows. We considered a typical estimation method for panel data models, the random effect estimator. The random effect technique is based on the assumption that the variation between country-pairs is unsystematic and it is not correlated with the explanatory variables. In this case, we determine the effect of variables that do not vary with time, as well. The independent variables for the random effect estimations are statistically significant in almost all the cases. We accept the null hypothesis for the free trade agreement parameter in the trade flows and imports equations, and for the border dummy in the export equation. The adjusted R-squared varies from 60,45% (imports equation) to 66,26% (exports equation). Coefficients and R-squared values are comparable to previous similar studies (e.g. Montanari, 2005; Josheski and Apostolov, 2013; Toševska-Trpčevska and Tevdovski, 2014).

Table 4. Commands for estimating the gravity model of trade and the corresponding output from R software. Estimation technique: *Random Effects Estimator (following the 'Swamy and Arora transformation')*. Dependent variables: *'flow' (sum of imports and exports), 'imp' (imports), 'exp' (exports).* [*'Estimate'*=the value of the estimated parameter; *'Std. Error'*=the standard error of the estimated parameter; *'t-value'*=the estimated value of t-test for the estimated parameter; *'Pr(>|t|)'*=p-value for the estimated parameter]

```
> ModRandom <-
plm(flow~gdp_o+gdp_d+dist+pop_sum+BORDER+FTA,data=y,
model="random",index = c("code", "year"))
> summary(ModRandom)
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)
Call:
plm(formula = flow ~ gdp_o + gdp_d + dist + pop_sum + BORDER +
FTA, data = y, model = "random", index = c("code", "year"))
Balanced Panel: n=90, T=5, N=450
Effects:
          var std.dev share
idiosyncratic 0.03508 0.18730 0.041
individual   0.81785 0.90435 0.959
theta: 0.9078
Residuals :
      Min. 1st Qu.  Median 3rd Qu.  Max.
-0.76900 -0.10700  0.00408  0.11500  0.86000
Coefficients :
              Estimate Std. Error t-value Pr(>|t|)
(Intercept) -10.63322    2.62410 -4.0521 5.994e-05 ***
gdp_o         0.86709    0.11013  7.8732 2.686e-14 ***
gdp_d         0.31640    0.10002  3.1635 0.001666 **
dist          -1.34874    0.20836 -6.4731 2.551e-10 ***
pop_sum       0.58262    0.13421  4.3410 1.759e-05 ***
BORDER        0.73344    0.35136  2.0874 0.037419 *
FTA           0.34796    0.24067  1.4458 0.148933
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-Squared:    0.62307
Adj. R-Squared: 0.61805
F-statistic: 35.2382 on 6 and 443 DF, p-value: < 2.22e-16

> ModRandom <-
plm(imp~gdp_o+gdp_d+dist+pop_sum+BORDER+FTA,data=y,
model="random",index = c("code", "year"))
```



```

> summary(ModRandom)
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)
Call:
plm(formula = imp ~ gdp_o + gdp_d + dist + pop_sum + BORDER +
    FTA, data = y, model = "random", index = c("code", "year"))
Balanced Panel: n=90, T=5, N=450
Effects:
      var std.dev share
idiosyncratic 0.03464 0.18611 0.044
individual    0.74462 0.86291 0.956
theta: 0.904
Residuals :
  Min. 1st Qu.  Median 3rd Qu.  Max.
-1.110 -0.109   0.017   0.111   0.821
Coefficients :
      Estimate Std. Error t-value Pr(>|t|)
(Intercept) -9.06601    2.53105 -3.5819 0.0003789 ***
gdp_o         0.81269    0.10634  7.6424 1.333e-13 ***
gdp_d         0.22880    0.09671  2.3658 0.0184198 *
dist         -1.21476    0.19808 -6.1328 1.914e-09 ***
pop_sum       0.63758    0.12836  4.9670 9.727e-07 ***
BORDER        0.66105    0.33340  1.9828 0.0480105 *
FTA           0.18929    0.22896  0.8267 0.4088282
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-Squared:      0.60932
Adj. R-Squared: 0.60451
F-statistic: 33.0664 on 6 and 443 DF, p-value: < 2.22e-16

```

```

> ModRandom <-
plm(exp~gdp_o+gdp_d+dist+pop_sum+BORDER+FTA,data=y,
model="random",index = c("code", "year"))
> summary(ModRandom)
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)
Call:
plm(formula = exp ~ gdp_o + gdp_d + dist + pop_sum + BORDER +
    FTA, data = y, model = "random", index = c("code", "year"))
Balanced Panel: n=90, T=5, N=450
Effects:
      var std.dev share
idiosyncratic 0.4172  0.6459 0.168
individual    2.0650  1.4370 0.832
theta: 0.8029
Residuals :
  Min. 1st Qu.  Median 3rd Qu.  Max.
-3.0600 -0.2080   0.0103   0.2960   2.3300
Coefficients :
      Estimate Std. Error t-value Pr(>|t|)
(Intercept) -26.60909    5.36826 -4.9567 1.023e-06 ***
gdp_o         1.40787    0.22408  6.2829 7.953e-10 ***
gdp_d         0.52883    0.20811  2.5411 0.0113894 *
dist         -1.93459    0.34783 -5.5619 4.622e-08 ***
pop_sum       0.52243    0.24595  2.1241 0.0342125 *
BORDER        0.83722    0.56931  1.4706 0.1421130
FTA           1.41704    0.40856  3.4684 0.0005748 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-Squared:      0.66673

```

Adj. R-Squared: 0.66258 F-statistic: 26.8566 on 6 and 443 DF, p-value: < 2.22e-16
--

Concluding remarks

The main objective of this work was to theoretically and empirically analyze the determinants of bilateral trade between WB countries and the most important partners, through the gravity theory. Our analysis started with a summary of some relevant facts and figures regarding international trade and the corresponding determinants. During the time period 2012-2014, world trade volume suffered the worst performance since the economic and financial crisis of 2008-2009. This negative trend was a result of various factors such as the decreasing import demand in developed countries, the slow import increase in developing countries, and the insufficient growth of global exports. GDP growth trade in 2014-2015 was below the expectations in WB countries, also due to the significant decline in commodity prices and to the low level of foreign direct investments.

The last part of our work was dedicated to the estimation of the gravity equations, using an appropriate dataset, which includes bilateral trade flows between WB countries and their main partners, and a set of determinants. We considered three dependent variables: imports, exports, and the sum of imports and exports. The statistical significant effect of the considered independent variables was generally confirmed. Authorities of WB countries should also orient their policies towards the technological progress, the productivity expansion, and the enhancement of business climate. A competitive environment and the reduction of trade barriers will stimulate foreign direct investments in the region.

References

- Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, 1979, 69(1), 106-116.
- Bergstrand, J. H. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics*, MIT Press, 67(3), 474-81.
- Bergstrand, J. H. (1989). The Generalized Gravity Equation, monopolistic competition, and the factor-proportions theory in international trade. *The Review of Economics and Statistics*, MIT Press, 71(1), 143-53.
- Bussière, M., Fidrmuc, J., & Schnatz, B. (2005). Trade integration of Central and Eastern European countries: lessons from a gravity model. *Working Paper Series 0545, European Central Bank*.
- Christie, E. (2002). Potential trade in Southeast Europe: a gravity model approach. *Working Paper No. 21., The Vienna Institute for International Economic Studies-WIIW*.
- Helpman, E. (1987). Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries. *Journal of the Japanese and International Economies*, 1, 62-81.

- Helpman, E., & Krugman, P. R. (1985). *Market Structure and Foreign Trade. Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge, MA: MIT Press.
- Josheski, D., & Apostolov, M. (2013). Macedonia's exports and the gravity model. *MPRA Paper* No. 48180, July 2013.
- Krugman, P. (1979). Increasing returns, monopolistic competition and international trade. *Journal of International Economics*, 9, 469-479.
- Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70, 950-959.
- Mayer, T., & Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist Database. *CEPII Working Paper* 25.
- Montanari, M. (2005). EU trade with Balkans, large room for growth? *Eastern European Economics*, vol.43, iss.1, pp.59-81.
- Tosevska-Trpcevska, K., & Tevdovski, D. (2014). Measuring the effects of customs and administrative procedures on trade: gravity model for South-Eastern Europe. *Croatian Economic Survey, The Institute of Economics, Zagreb*, vol. 16(1), pages 109-127, April.

Exchange Rate Pass-Through Effect In Albania:A Structural VAR Approach (SVAR)

Aida SALKO¹ Ardit GJECI²

¹University of Ljubljana, Faculty of Economics, Slovenia, MSc. Money and Finance: aida.salko93@gmail.com

²University of Ljubljana, Faculty of Economics, Slovenia, PhD student Money and Finance: ardit_ubt@hotmail.com

Abstract

The aim of this paper is to analyze the exchange rate pass-through into domestic prices in Albania by using Structural Vector Autoregressive (SVAR) model. We used quarterly data, ranging from January 2001 up to December 2015. The source of the data is the Bank of Albania. The variables taken into account are nominal effective exchange rate, consumer price index, gross domestic production and short-term nominal interest rate. By using the SVAR model in our empirical analysis we intent to examine the exchange rate dynamic effects on prices, output and interest rates. Identification of the exchange rate shocks is done by taking into account the recursive ordering as well as checking robustness to alternative orderings. The paper contributes by explaining in detail the procedure of building a SVAR model and interpretation of the impulse response as well as analyzing the impact of exchange rate shocks into the selected variables.

Key Words: Pass-through, Exchange rate, Domestic prices, Monetary policy, Structural Vector Autoregressive (SVAR)

JEL Classification: F41, E23, E3

Introduction

In open economies, the exchange rate fluctuations have received considerable attention in recent theoretical and empirical studies. The main purpose of this research paper is to examine the exchange rate pass through effect in Albania, by using structural vector autoregressive (SVAR) models. The motivation derives from the significant role of exchange rate in a small open economy like Albania, where the exchange rate fluctuations affect the behavior of inflation. This makes the exchange rate pass-through an important consideration with respect to monetary policy by including the exchange rate as one of the main variables in the process of inflation forecasting in the Bank of Albania.

The remainder of this paper is organized as follows. Section 2 reviews the literature on exchange rate pass-through effects. Moreover in this section we pay a special attention to the literature on the role of exchange rate pass-through in candidate countries to European Union (EU) by specifying Albania. Section 3 presents the empirical analyses on exchange rate pass-through in Albania within the context of a SVAR model. Sessions on this part cover data, methodology and the key modeling steps. Furthermore, we perform some robustness tests by using alternative identification schemes. The final section presents main conclusions.

Literature Review

The relation between exchange rate pass through effect and inflation targeting has been empirically studied in detail in the literature. Recent empirical findings indicate that there is an important relationship between the inflation environment and the degree of exchange rate pass through effect. For instance, Choudhri and Hakura (2006), find strong evidence of positive and significant association between pass-through and inflation using a panel of 71 countries for the time period from 1979 to 2000. According to Goldberg and Campa (2002), who offered empirical evidence on exchange rate pass-through in 23 OECD countries over the period 1975 to 1999, scope of exchange rate pass-through to consumer prices tends to be lower in countries with lower inflation and exchange rate volatility. Darvas (2001), treats the possible role of using the exchange rate to control the inflation process in candidate countries for EU, as an empirical issue. His paper is based on common characteristics of these countries which distinguish them from the countries currently inside the European Economic and Monetary Union (EMU). The EU candidate countries place strong emphasis on the objective of low inflation and meantime face the existence of a wide range of different exchange rate regimes. In these countries, domestic prices are likely to change without any movements in the exchange rate and if substantial movements in the exchange rate happen, then the price changes should be carefully decomposed into pass-through and the underlying price convergence process.

The effects of exchange rate pass-through in Albania have been widely studied. Bank of Albania has pursued a monetary targeting regime since 1992, by having price stability as the main objective. Monetary policy is conducted under a flexible exchange rate regime, a choice dictated mostly by weak developments of financial markets, lack of institutional experience and low level of international reserves. Bank of Albania includes the exchange rate as one of the main variables in the process of inflation forecasting, along with other factors (Themeli and Kolasi, 2006). Furthermore, Muco, Sanfey and Taci (2004), found a strong evidence that exchange rate stability has played a crucial role in maintaining low levels of inflation during the transition period in Albania. Peeters (2005), describes the link between inflation and exchange rates as a changeable relationship and extends the doubts whether the exchange rate is still the main channel of monetary transmission mechanism. Meanwhile, Istrefi and Semi (2007), provide empirical evidences on the degree of exchange rate pass-through to consumer prices in Albania and show that there is a fast reaction of consumer prices in the presence of an exchange rate shock.

Empirical Model

Data

The sample period consists of quarterly observations from 2001:Q1 to 2015:Q4. The two most important variables used in the research are: Nominal Effective Exchange Rate (NEER) and the Consumer Price Index (CPI) as a measure of domestic prices. In the mean time, part of the model will be also data on Gross Domestic Production (GDP) as a measure of output and short-term interest rates as a monetary policy instrument. Nominal effective exchange rate is calculated toward two currencies, Euro

and US dollar, taking into account the share that they represent in trade balance. In our model, an increase of NEER means the depreciation of Albanian Currency Lek (ALL). The source of the data is the Bank of Albania.

Methodology

In recent years there is a growing interest in the structural VAR approach (SVAR) and the literature about it shows a common feature: the attempt to "organize", in a "structural" theoretical sense, instantaneous correlations among the relevant variables (Amisano and Giannini, 1997). Several papers used different variants of the VAR models to analyze the impact of exchange rate movements on consumer prices. In addition, McCarthy (2000) used a recursive VAR framework to identify the shocks. In this research paper we will estimate a structural VAR by identifying exchange rate shocks (exogenous change in the exchange rate, originating from the markets, say) in recursive ordering and study their dynamic effects on prices (CPI), output (GDP) and interest rates in Albanian economy. Within this framework, the empirical analysis is conducted by using JMulTi software.

Stationarity Test

In order to estimate the SVAR, all time series are transformed to stationary series. Usually, time series are non-stationary which might lead to misleading results in our estimation. In order to solve this problem we compute Unit root test called Augmented Dickey-Fuller (ADF) test:

$$\Delta y_t = \phi y_{t-1} + \sum_{j=1}^{p-1} \alpha_j^* \Delta y_{t-j} + u_t$$

Where: $H_0: \phi = 0$ versus $H_1: \phi < 0$

If the t-statistic is lower than the critical value, the null hypothesis is rejected; so, the series are non-stationary. Confidence interval is set as 5% (Lütkepohl & Krätzig, 2006). Critical values of the test depend on the deterministic terms which have to be included. Therefore, different critical values are used when a constant or linear trend term is included in the test. All series used in this study have a nonzero mean and also have a linear trend except NEER variable. Results are presented in the following table:

Table 1: Augmented Dickey-Fuller Test Results

Variables	ADF Test Statistic	Deterministic Terms
NEER	-1,4528	Constant
Interest	-1,9657	Constant, Trend
GDP	-0,3785	Constant, Trend
CPI	-1,5915	Constant, Trend

Notes: 5% critical values for ADF test with constant and trend and constant are -3,41 and -2,86, respectively.

Source: Authors calculation

As shown from the results, the values of test statistics are above the critical value with 5% confidence interval, meaning that all series are non-stationary and need to be transformed. First, we use logarithmic values of the variables and then compute first differences to make them stationary. If we compute the augmented Dickey-Fuller test

again, test statistics are below the critical value, enabling us to reject H_0 . Therefore, we achieved stationarity of the time series by using transformations. New results are presented in the following table:

Table 2: Augmented Dickey-Fuller Test Results

Variables	ADF Test Statistic	Deterministic Terms
NEER_log_d1	-3,7579	Constant
Interest_log_d1	-3,7195	Constant, Trend
GDP_log_d1	-4,0635	Constant, Trend
CPI_log_d1	-9,4127	Constant, Trend

Notes: 5% critical values for ADF test with constant and trend and constant are -3,41 and -2,86, respectively.

Source: Authors calculation

Order of the variables

In order to estimate the corresponding VAR and identify the exchange rate shocks, we have to apply the recursive ordering first. In other words, it is important that variables enter the model in a reasonable way. A number of studies on the exchange rate pass-through rely on the ordering by McCarthy (2000). Thus, it is reasonable to order first the most exogenous variables. Based on the exogeneity of the variables and the paper of McCarthy (2000), we decided for the following order :

NEER_log_d1 INTEREST_log_d1 GDP_log_d1
CPI_log_d1

Actually, such ordering is supported even theoretically, as for a small and open economy, like Albania, is suggested that foreign variables (in this case the exchange rate) it should be ranked first. Then, the interest rate being the second means that the central bank decisions are forward looking; therefore it reacts ahead of expected changes in consumer prices.

Lag selection

Next, important step in our analysis is the optimal lag selection. Computing the info criteria in JMulTi, we provide the optimal number of endogenous lags in our VAR model. Results for the most frequently used statistics, namely Akaike info criterion (AIC), Final prediction error (FPE), Hannan-Quinn criterion (HQ) and Schwarz criterion (SC), are showed in the table of output below. Since the AIC criterion tends to overestimate the optimal number of lags we choose a number which is between the results according to AIC and the Schwarz Criterion (conservative). Finally we choose a number of 3 lags as optimal number of lags and take rather a small loss of efficiency than a loss of inconsistency by overestimating the number of lags (Lütkepohl, 2007).

OPTIMAL ENDOGENOUS LAGS FROM INFORMATION CRITERIA

```
endogenous variables:    NEER_log_d1 Interest_log_d1 GDP_log_d1
                        CPI_log_d1
sample range:           [2003 Q4, 2015 Q4], T = 49
```

optimal number of lags (searched up to 10 lags of levels):
Akaike Info Criterion: 10
Final Prediction Error: 3
Hannan-Quinn Criterion: **3**
Schwarz Criterion: 2

Estimation of the model

We use a VAR methodology to investigate exchange rate pass-through to domestic prices in Albanian economy. Cholesky decomposition of variance-covariance matrix of reduced-form VAR residuals is implemented in order to examine responsiveness of domestic price indexes (CPI) to the unexpected exchange rate shock (NEER). The structural form VAR is described as following:

$$Ay_t = A[A_1 \dots A_p]Y_{t-1} + B\varepsilon_t$$

where ε_t is white noise disturbances which represent the unexplained movements in the variables, reflecting the influence of exogenous shocks with $\varepsilon_t \sim N(0, I_K)$. If we define $A = [A_1, \dots, A_p]$, then the corresponding reduced form is:

$$y_t = AY_{t-1} + u_t$$

Moreover, the relationship between reduced-form VAR residuals and structural shocks can be expressed as: $u_t = A^{-1}B\varepsilon_t$. In other words, the A matrix induces a transformation on the disturbance vector ε_t , generating a new vector that can be conceived as being generated by linear combinations (through the B matrix) of n independent disturbances, which we will refer to as e_t (Lütkepohl and Kratzig, 2004). In this model we assume four exogenous shocks that contemporaneously affect endogenous variables - Nominal effective exchange rate shock (NEER), interest rate shock, output shock and internal price shock (CPI). The restrictions on the matrixes A and B, are based on the Cholesky decomposition of the reduced-form VAR residual matrix that define matrix A as a lower triangular matrix which implies a recursive scheme (structural shocks are identified through the reduced-form VAR residuals) among variables (Mirdala, 2014). Estimation is done by maximum likelihood using a scoring algorithm.

The following assumptions are taken into account for the correct identification of exogenous structural shocks:

- Exchange rate (NEER) doesn't contemporaneously respond to the shock from any other endogenous variable in our model.
- Interest rate doesn't contemporaneously respond to output, and internal price shocks, while it is contemporaneously affected only by the exchange rate shock.
- GDP doesn't contemporaneously respond to internal price shocks, while it is contemporaneously affected by exchange rate and interest rate shocks.
- Domestic price index (CPI) is contemporaneously affected by the shocks from all of the endogenous variables of the model.

As mentioned also in section 3.5, the order of variables is crucial for a correct identification of structural shocks and a convenient transmission mechanism of the exchange rate (NEER) shock into the domestic prices. The matrix below describes the relationship between the reduced form VAR residuals and the structural shocks:

$$\begin{bmatrix} u_t^{NEER} \\ u_t^{Interest} \\ u_t^{GDP} \\ u_t^{CPI} \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 \\ b_{31} & b_{32} & b_{33} & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{NEER} \\ \varepsilon_t^{Interest} \\ \varepsilon_t^{GDP} \\ \varepsilon_t^{CPI} \end{bmatrix}$$

Since both A and B have in principle $2K^2$ parameters, the number of restrictions we can impose is given by:

$$2K^2 - \frac{K(K+1)}{2} = 22$$

Model checking

It is important to note that the residuals from the estimated VAR model should be well-behaved, meaning that there should be no problems with autocorrelation and non-normality. Therefore, in order to check for the validity of the model we execute different tests as following:

Breusch-Godfrey LM test for autocorrelation

The Breusch-Godfrey LM test for h -th order residual autocorrelation assumes a model:

$$u_t = B_1^* u_{t-1} + \dots + B_h^* u_{t-h} + error_t$$

and checks : $H_0: B_1^* = \dots = B_h^* = 0$ versus $H_1: B_1^* \neq 0$ or ... or $B_h^* \neq 0$

To perform the test the auxiliary model is estimated:

$$u_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + CD_t + B_1 u_{t-1} + \dots + B_h u_{t-h} + e_t$$

The LM statistic is:

$$LM = T(K - \text{tr}(\sum_R^{-1} \sum_e)) \approx \chi^2(hK^2)$$

Edgerton and Shukur (1999), find that this test may be biased in small samples and therefore another statistic (LMF statistic) which may perform better, is also given for VAR models.

$$LMF = \frac{1 - (1 - R_r^2)^{1/r}}{(1 - R_r^2)^{1/r}} * \frac{Nr - q}{Km}$$

As we see from the output above , the p-value (LMF statistic) = 0.2000, meaning that we cannot reject H_0 , so the residuals are not correlated in our model.

```
LM-TYPE TEST FOR AUTOCORRELATION with 3 lags
Reference: Doornik (1996), LM test and LMF test (with F-
approximation)
LM statistic:          79.1558
p-value:              0.0031
df:                  48.0000
LMF statistic:        1.2251
p-value:              0.2000
```

```
df1: 48.0000
df2: 94.0000
```

Tests for non-normality

Economic data can often fail the assumption of normally distributed residuals due to the outliers and volatile nature of the data. Normality is assumed under a null hypothesis, while alternative hypothesis represents non-normality. Doornik and Hansen proposed an idea to transform the residual vector in such a way that its components are both standardized and independent, and then check the compatibility of skewness and kurtosis with those of a normal distribution. Lütkepohl proposed a different method of computing standardised residuals, therefore the results based on his test statistics are also shown. The normality of the error term can also be tested by using the Jarque-Bera test, which tests for the presence of skewness (non-symmetry) and kurtosis (fat tails) and we will rely on this test. As we see, p-Values(χ^2) show that the first and the third residuals are normally distributed, whereas the second and the fourth are not. This problem, leads to a loss of efficiency, but not consistency of parameters.

TESTS FOR NONNORMALITY

Reference: Doornik & Hansen (1994)

```
joint test statistic: 19.3023
p-value: 0.0133
degrees of freedom: 8.0000
skewness only: 8.3208
p-value: 0.0805
kurtosis only: 10.9815
p-value: 0.0268
```

Reference: Lütkepohl (1993), Introduction to Multiple Time Series Analysis, 2ed, p. 153

```
joint test statistic: 19.9746
p-value: 0.0104
degrees of freedom: 8.0000
skewness only: 9.5539
p-value: 0.0487
kurtosis only: 10.4207
p-value: 0.0339
```

JARQUE-BERA TEST

variable	teststat	p-Value (χ^2)	skewness	kurtosis
u1	13.2178	0.0013	0.86464.6355	
u2	0.3783	0.8277	0.06203.3831	
u3	5.7879	0.0554	0.23094.5057	
u4	2.9407	0.2298	0.50293.4989	

Impulse responses analysis

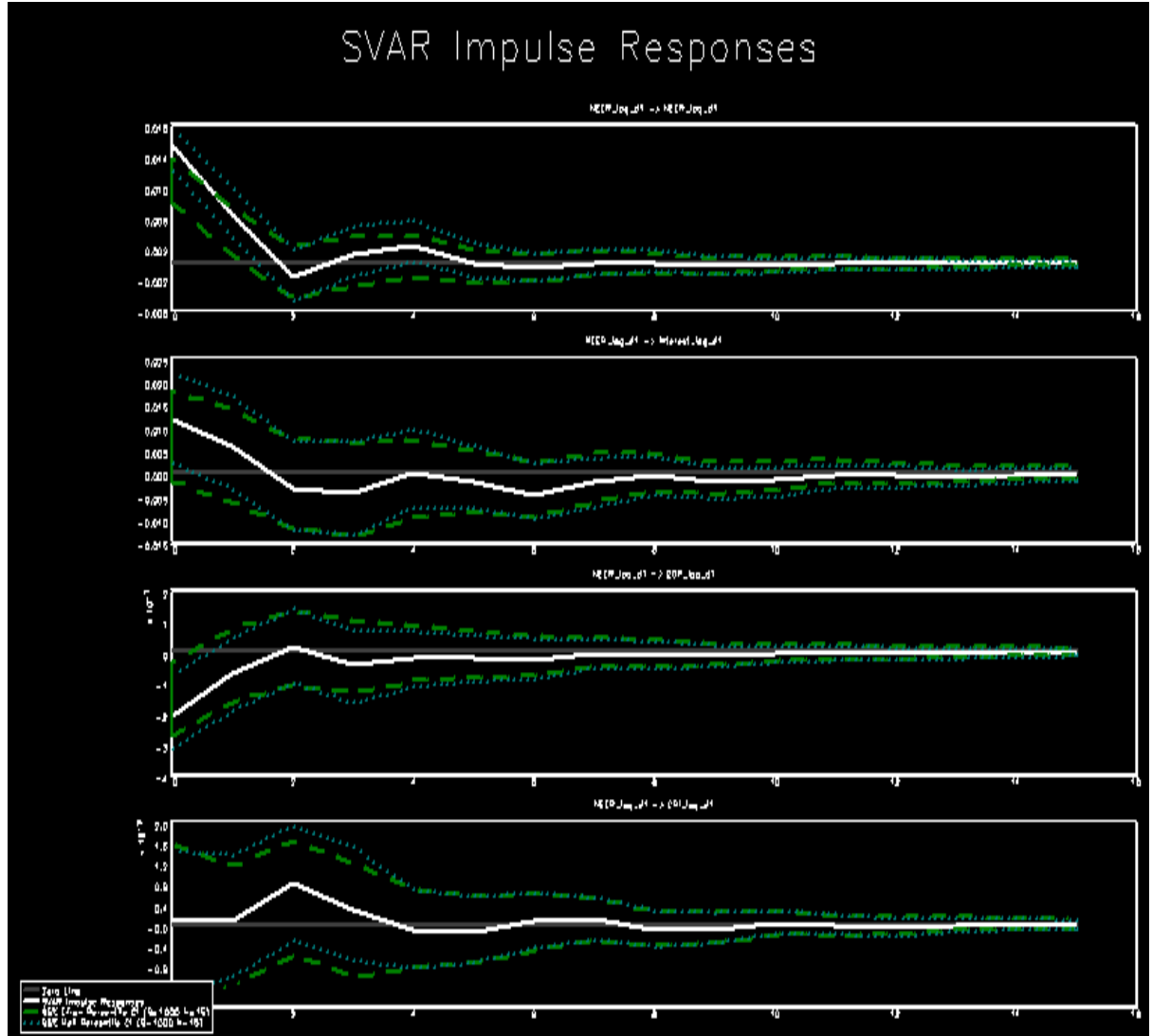
We start now the analysis about Impulse Responses. First of all, we have set number of periods equal to 15. Then we have set 1000 as number of bootstrap replications with 95% confidence interval (which shows us the uncertainty of our estimation). For the impulse response analysis, we have selected Nominal Effective Exchange Rate (NEER), Interest as impulse, and Gross Domestic Production (GDP), Consumer Price Index (CPI) as responses. Before displaying Impulse responses, we select 95% Efron

and Hall Percentile CI for the confidence intervals. The results are presented in Figure 1.

The graph below shows the effect on the interest rates, GDP and domestic prices from the permanent shock of exchange rate (NEER). The horizontal axis shows the time horizon, while the vertical axis shows the percentage change of the price index. All shocks imply a 1% change in the exchange rate while the blue and green lines imply the 95% confidence interval. In this case we are mainly interested in the impact of this shock on consumer prices, said differently, the pass-through of the exchange rate to consumer prices.

In our model, a positive shock of NEER means the depreciation of Albanian currency Lek (ALL). As it is showed in the graph below, the exchange rate has an expected positive effect on CPI that peaks after about two quarters but the effect dies out after eight to ten periods. Figure 1, shows that consumer prices will promptly respond to the shock in the exchange rate by an increase of 0.08% in reaction to the 1% depreciation of the currency. As the economic theory tells us, the depreciation of the national currency (ALL) versus foreign currencies (ex. Euro (EUR) as the main foreign currency in the Albanian economy), makes the imported goods to rise in the domestic market price so it is expected an increase of the Consumer Price Index (CPI) and exported goods are sold cheaper in foreign markets. As a result, this promotes exports at a time when the goods are exported to overseas market competitive by improving the country's trade balance. In other words, an appreciation of the EUR currency, makes imported goods more expensive thus increases the competitiveness of domestic producers (substitution effect dominates), but on the other hand, higher import prices cause an increase on the CPI. In this case we can speak of a kind of "import" of inflation. This is an expected result in a country like Albania, relying heavily on imports in particular. However, this leads to a drop in consumption and a slight increase in exports of Albania, thus GDP does not react a lot in the short run. In this way, Central Bank reacts by increasing interest rates since higher interest rates put downward pressures on inflation. If we consider the overall picture of exchange rate pass-through on prices in Albania, we can say that from 2001 and after, the effect has become weaker. There are a number of possible reasons to explain the decline of exchange rate pass-through such as developments in market structures, the stability of our currency (ALL) as well as the low inflation environment in recent years. Furthermore, the NEER itself has been relatively stable during recent years, showing a decreasing volatility year after year and making it harder to identify significant statistical relationships of exchange rate with other economic variables. Finally, in the Bank of Albania periodicals it is often mentioned the increase of competition in our economy, as a factor that helps in weakening the exchange rate pass-through, even though there is still no empirical study.

Figure 1: SVAR Impulse Responses



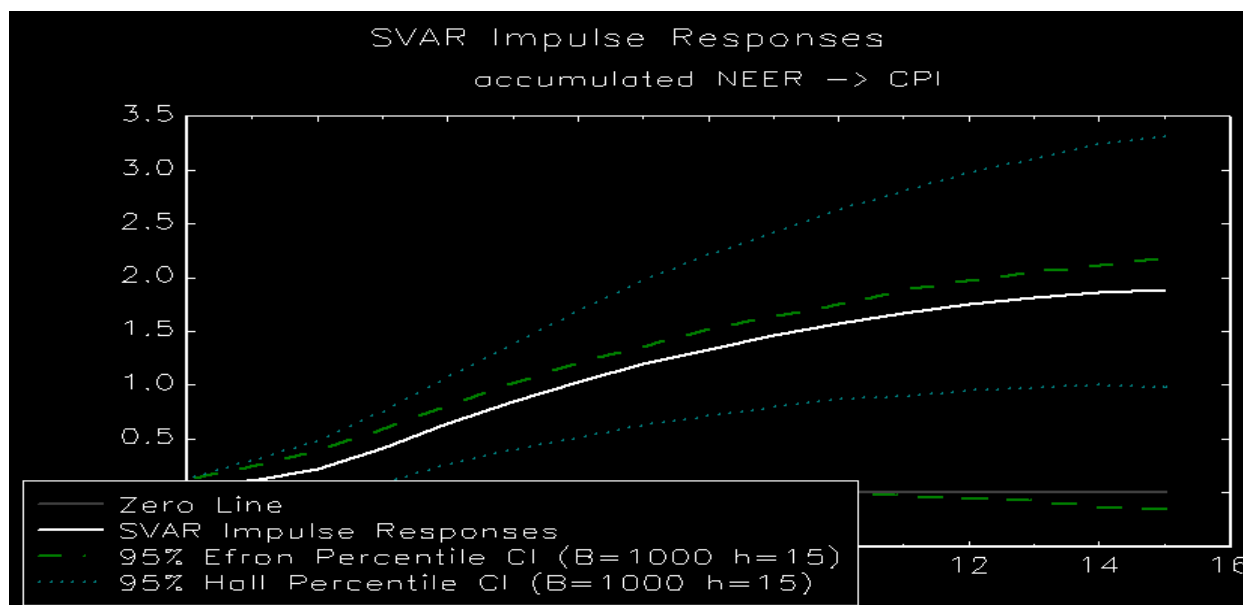
Source: Authors calculation

Figure 1 presents the estimated pass-through of exchange rate to consumer prices when $T=15$ quarters. The dotted lines represent confidence intervals, which indicate that responses are statistically different from zero. Since we are interested mostly in the impact of this shock (NEER) on consumer prices, said differently, the pass-through of the exchange rate to consumer prices, we will analyze it more specifically. We normalize cumulative responses of consumer prices with cumulative responses of exchange rate following Rabanal and Schwartz (2001) and Istrefi and Semi (2007). The pass-through coefficient is thus defined as:

$$PT_{t,t+j} = P_{t,t+j}/E_{t,t+j}$$

where: $P_{t,t+j}$ is the cumulative change in consumer prices and $E_{t,t+j}$ is the cumulative change in the nominal effective exchange rate. Results are showed for both schemes in Figure 6 below.

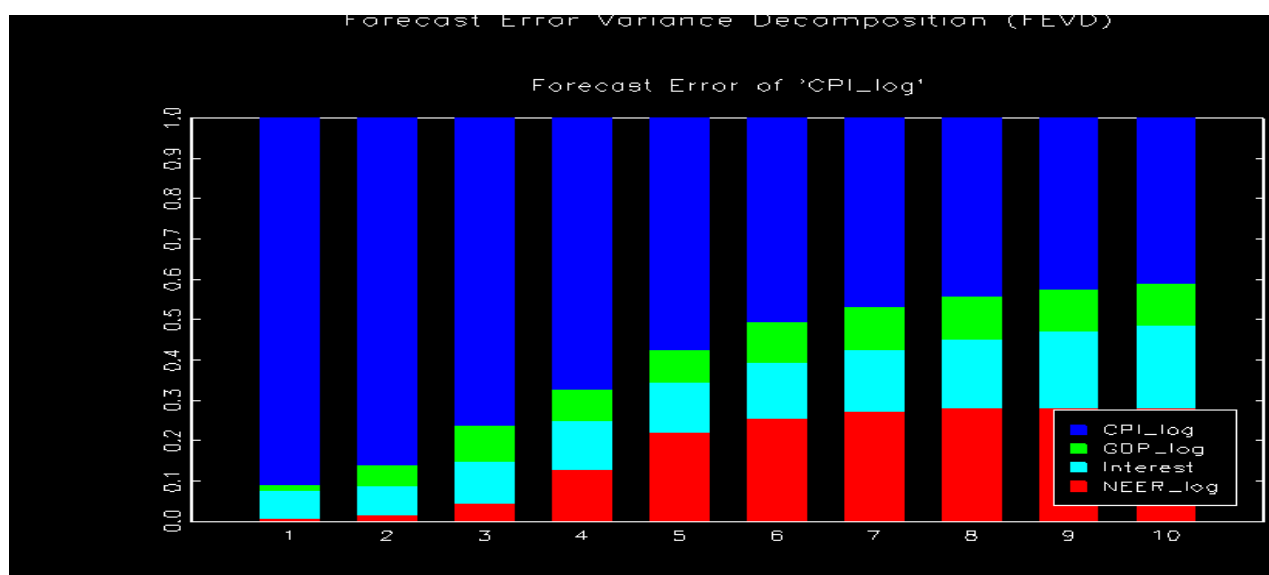
Figure 2: Cumulative responses of consumer prices following a 1% increase in NEER



Source: Authors calculation

Forecast error variance decompositions (FEVDs) are popular tools for interpreting VAR models. In order to see more specifically the effect of an exchange rate shock on consumer prices index, we analyze the variance decomposition of consumer prices which informs us about the relative importance of every random shock on the variance of consumer prices. Variance decompositions measure the percentage of the forecast variance in consumer prices indexes (CPI) that can be attributed to various shocks. As we see from the results, the consumer price variance is explained up to 26 percent by exchange rate shocks but however, the largest part of consumer price variance is explained by its own innovations.

Figure 3: Variance Decomposition CPI



Source: Authors calculation

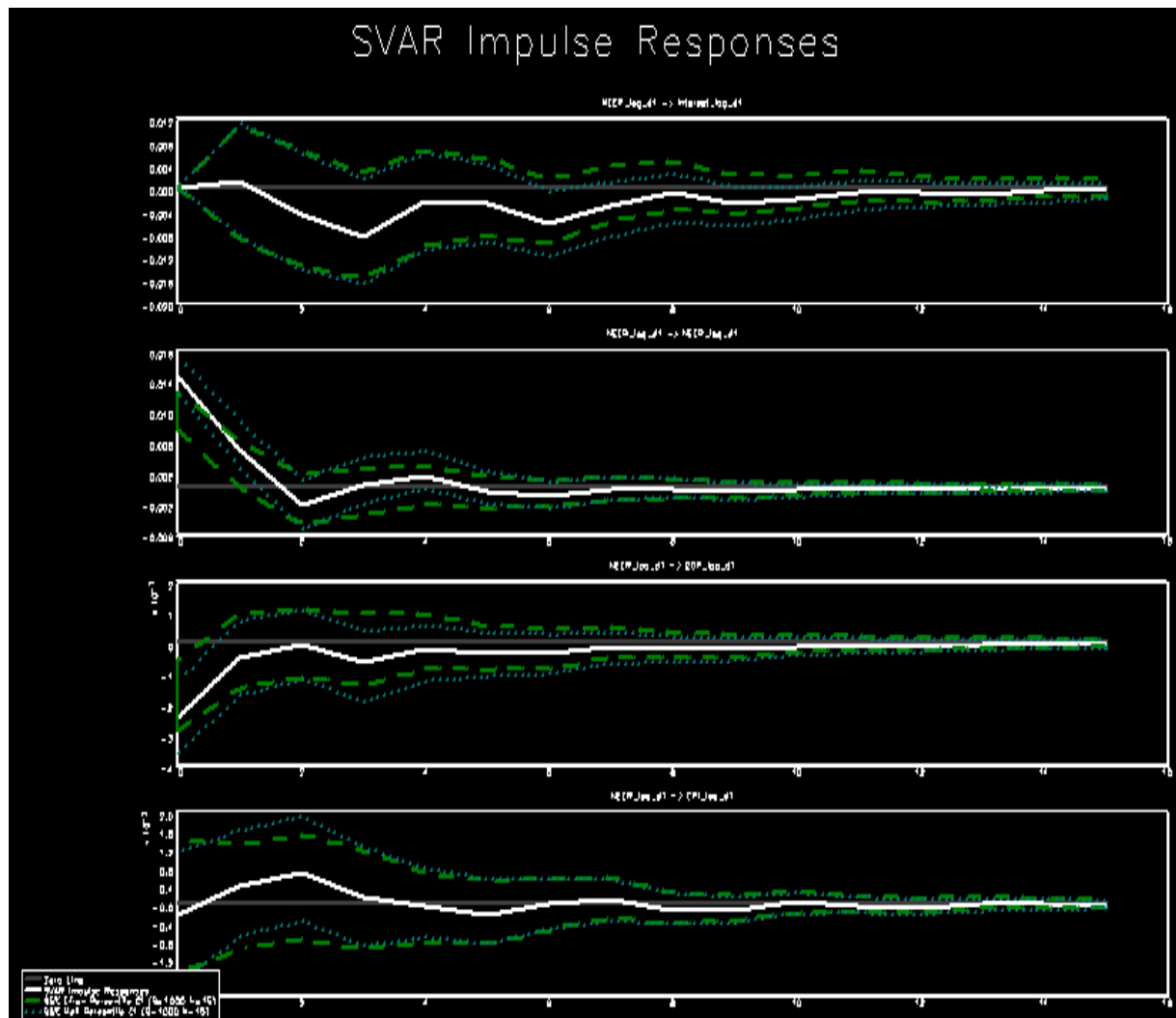
Robustness check

In order to check the robustness of the exchange rate pass-through coefficients, we propose a reordering of the variables and compare the results with the first scheme proposed in section 3.5 above. In this case, the nominal effective exchange rate will be ordered behind the interest rate, allowing it to react to domestic monetary conditions so the new identification scheme is presented as following:

INTEREST_log_d1 $\longrightarrow$ NEER_log_d1 $\longrightarrow$ GDP_log_d1 $\longrightarrow$ CPI_log_d1

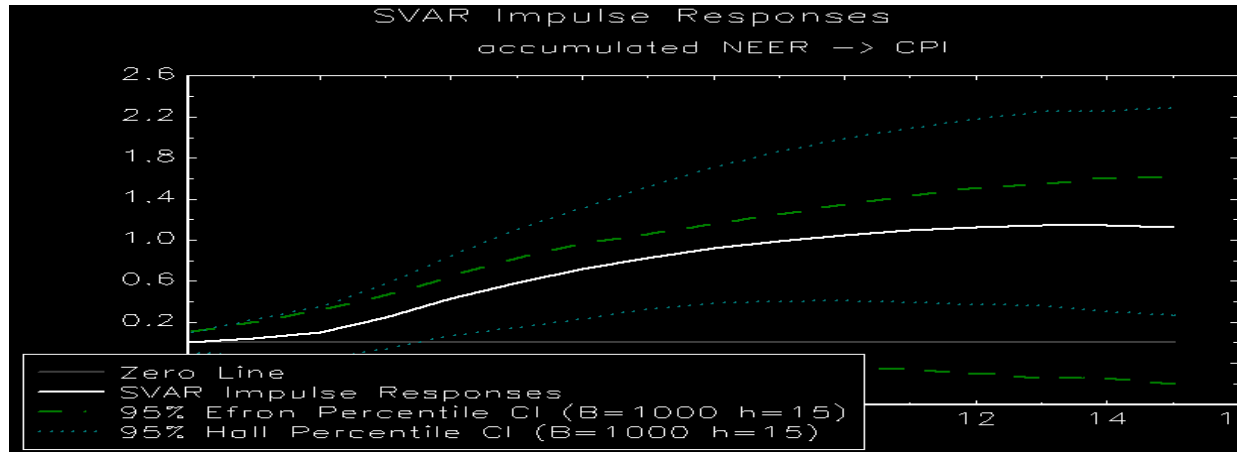
Figure 4 presents the impulse responses based on the new identification scheme. If we compare the first plot of IR (Figure 1) with second plot, where we have changed the order of the NEER and Interest (Figure 4), we can see that results of IR after changing the order of the variables are similar to the results with the previous identification scheme which lead us to the conclusion that our model is robust.

Figure 4: SVAR Impulse Responses



Source: Authors calculation

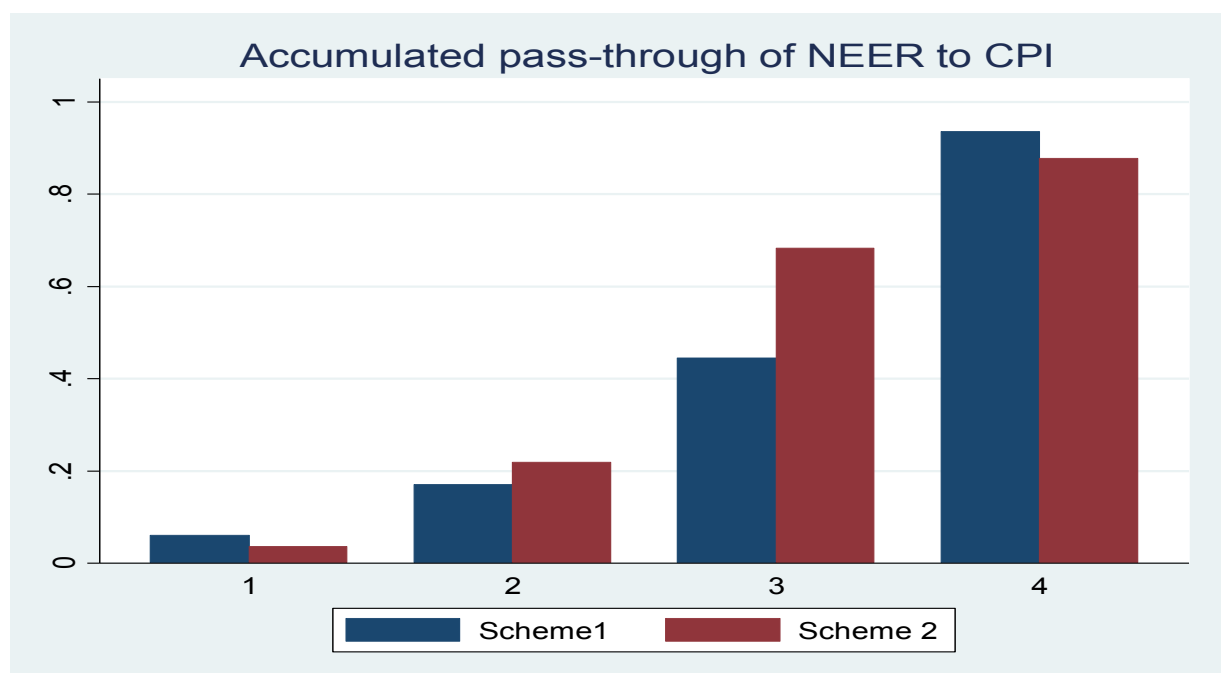
Figure 5: Cumulative responses of consumer prices following a 1% increase in NEER (Scheme 2)



Source: Authors calculation

As explained also above, we normalize cumulative responses of consumer prices with cumulative responses of exchange rate for both identification schemes. From the graph, it is showed that within the year, the exchange rate pass-through reaches the maximum, almost 93 percent (1st scheme, blue column). However, after reordering the variables, the results show that changes between coefficients of pass-through, estimated based on two different specifications are almost insignificant. As a conclusion, we can say that the ordering of the nominal effective exchange rate variable doesnot matter for coefficients of the exchange rate pass-through to be stable.

Figure 6: Accumulated pass through of NEER to consumer prices



Concluding Remarks

In this research, we use SVAR model and impulse response functions in order to shed some light on the transmission of exchange rate movements to consumer prices in Albanian economy. Being in line with the results found also in other countries, the exchange rate pass-through in Albania appears to be in decline. There are a number of possible reasons to explain the decline of exchange rate pass-through such as developments in market structures, the stability of our currency (ALL) as well as the low inflation environment in recent years.

However, there are strong signs that the exchange rate channel is losing its influence at the benefit of the other transmission channels. Nowadays, Albanian economy is still going under major changes. Goods and services markets are developing, there is a gain in business and consumer confidence and the banking sector improves slowly but steadily. As a consequence of all these changes, the transmission channels also change continuously. Furthermore, the exchange rate effect is no more easily captured by direct analyses of data on consumer prices and the exchange rates. It is needed to emphasize that the last years, exchange rate volatility in Albania has been very low, making it hard to identify significant statistical relationships with other economic variables. Despite being in a flexible exchange rate regime, the Bank of Albania has not ignored exchange rate movements in the ALL, taking into account the importance of exchange rate movements in a small and open economy. Monitoring and surveillance of exchange rate can be a tool to keep under control the acceptable level of consumer prices, and maybe should be used especially when fluctuations sensitive to our currency.

References

- Amisano, G., & Giannini, C. (1997). Topics in Structural VAR Econometrics. *Springer*, 181. <https://doi.org/10.1007/978-3-642-60623-6>
- Choudhri, E. U., & Hakura, D. S. (2001). Exchange Rate Pass-Through to Domestic Prices :Does the Inflationary Environment Matter?
- Darvas, Z. (2001). Exchange rate pass-through and real exchange rate in EU candidate countries Zsolt Darvas, (May).
- Edgerton, D. & Shukur, G. (1999). Testing autocorrelation in a system perspective. *Econometric Reviews* 18: 343-386.
- Goldberg, L. S., & Campa, J. M. (2002). Exchange Rate Pass-Through Into Import Prices: a Macro or Micro Phenomenon? *NBER Working Paper*, 53(9), 1689–1699.
- Istrefi, K., & Semi, V. (2007). Exchange rate pass - through in Albania. *Bank of Albania*, (739), 48.

- Lütkepohl, H. (2007). Econometric analysis with vector autoregressive Models. *EUI Working Papers*, 1–56. <https://doi.org/10.1002/9780470748916.ch8>
- Lütkepohl, H., & Kratzig, M. (2006). Initial Analysis in JMulTi Plot Time Series Specification in JMulTi. *Test*, 1–38.
- Lütkepohl, H., & Kratzig, M. (2004). Applied Time Series Econometrics (Themes in Modern Econometrics). Cambridge University Press.
- McCarthy, J. (2000). Pass-Through of Exchange Rates and Import Prices to Domestic Inflation in Some Industrialized Economies. *Eastern Economic Journal*, 33(4), 511–537. <https://doi.org/10.1057/ej.2007.38>
- Mirdala, R. (2014). Exchange Rate Pass-through to Consumer Prices in the European Transition Economies. *Procedia Economics and Finance*, 12(March), 428–436. [https://doi.org/10.1016/S2212-5671\(14\)00364-5](https://doi.org/10.1016/S2212-5671(14)00364-5)
- Muco, M., Sanfey, P., & Taci, A. (2003). Inflation , exchange rates and the role of monetary policy in Albania by, (88), 0–17.
- Peteers, M. (2005) “What about Monetary Transmission mechanism in Albania? Is the exchange rate pass-through (still) the main channel? Fifth Annual Conference of Bank of Albania
- Rabanal, P. & Schwartz, G. (2001). Exchange Rate changes and Consumer Price Inflation: 20 Months after the floating of the Real. IMF Country Report : Selected Issues and Statistical Appendix (Section 5)
- Themeli, E. and Kolasi, G.(2006). ”The IT case in Albania, A tentative road-map for implementing IT in Albania”. *Round Table Inflation Targeting 2. Bank of Albania*, <http://www.gbv.de/dms/zbw/545651395.pdf>
- <https://www.bankofalbania.org/?crd=0,3,0,0,0&ln=Lng1>

Guided Kernel Density Estimator And The Gamma Kernel Estimator

¹Lule Hallaçi, ²Llukan Puka

^{1,2}Department of Applied Mathematics, Faculty of Natural Science, University of Tirana, Tirana, Albania lule.hallaci@fshn.edu.al

Abstract

Parametrically guided nonparametric estimation is a method that allows improving the bias of a nonparametric estimator by using a parametric pilot estimator. Talamakrouni (2016) generalize the parametrically guided nonparametric estimation to randomly right-censored data. The basic idea is to start with any parametric density estimator and then to adjust this first stage parametric approximation using a nonparametric kernel-type estimator of a particular correction factor. However, in many situations, using the classical kernel leads to the well known boundary effect problem, that is, the estimator has a large bias near the endpoints. Bouezmarni (2011) proposed a gamma kernel estimator that corrects for the boundary effects. In this paper we perform a comparison between the guided kernel density estimator, based on Kaplan-Meier (1958) estimator and the gamma kernel estimator, for both the density and the hazard function. Via a Monte Carlo simulation, the finite sample performance of the estimators is investigated under various scenarios.

Keywords: Parametrically guided nonparametric estimation, guided kernel density, gamma kernel estimator.

JEL Classification: C02: Mathematical Methods; C14: Semi parametric and Nonparametric Methods: General; C15: Statistical Simulation Methods: General.

Introduction

Censored data arise in many contexts, for example, in medical follow-up studies in which the occurrence of the event times (called survival) of individuals may be prevented by the previous occurrence of another competing event (called censoring). The estimation of the probability density and hazard function has received considerable attention in such studies, as it allows visualizing and exploring the distribution of data. There is a large variety of approaches to estimate the density and the hazard functions that are parametric, nonparametric, semi parametric and method which use aspects from both the nonparametric and the parametric school. Few of this method have been investigated in the presence of censoring mechanism.

The parametric approach has the advantage of being powerful by its $\sqrt{n}$ rate of convergence and also precise when the chosen family is correctly specified. However, a major complication that is emphasized in parametric modeling is the risk of biased and inconsistent parameter estimation due to misspecification problem. In the fully nonparametric approach, the estimators suffer from the curse of dimensionality and have in general a slower rate of convergence. However, despite its drawbacks, nonparametric approach provides more flexibility since the estimation is not based on any parameterized family of functions and remains more robust and applicable in practice.

Based on the Kaplan-Meier estimator, several nonparametric density estimators have been proposed in the literature. A popular approach for estimating the density function and the hazard rate function is done using a fixed symmetric kernel density with

bounded support and a bandwidth parameter, Blum and Susarla (1980). The kernel determines the shape of the local neighborhood while the bandwidth controls the degree of smoothness. Sabine and Stute (1988) investigated the kernel-type nonparametric estimator in the presence of right-censoring. However when the density function of the data have a bounded support, using the classical kernel leads to an estimator with a large bias near the endpoints. The problem of bias is called also the boundary effect. Boundary effects are well known to be a disturbing nuisance for applications as well as for global measures of performance of kernel estimators. The reason that boundary effects occur for unmodified kernel estimators is that the curve to be estimated has a discontinuity at an endpoint, so that the usual bias expansion which depends on smoothness assumptions cannot be carried out anymore. This is especially the case in survival analysis, since the survival time is assumed to be nonnegative variable. There have been various efforts to modify kernel estimators near boundaries in order to reduce the impact of these boundary effects. Bouezmarni (2011) proposed a gamma kernel (GK) estimator that corrects for the boundary effects.

In the fully nonparametric approach, the estimators have a slower rate of convergence. The parametric approach has the advantage of being powerful by its $\sqrt{n}$ rate of convergence but in parametric modeling is the risk of biased and inconsistent parameter estimation due to misspecification problem. Usually, even when the proposed model is misspecified, parametric estimation can provide valuable information about the phenomenon under study. This motivates the consideration of an approach called parametrically guided nonparametric estimation that contains both a parametric and a nonparametric component. The idea is to multiply an initial parametric density estimate with a kernel type estimate of the necessary correction factor. A guided nonparametric estimator is completely nonparametric in the sense that it does not rely on any assumed global structure. On the other hand, a guided nonparametric estimator takes advantage of both parametric and nonparametric methods: In the complete data case, considerable attention has recently been paid to parametrically guide nonparametric estimation in the literature. The starting point for this method was Hjort and Glad (1995), who introduced the parametric guided kernel (PGK) scheme and proved the bias reduction property of their guided estimator in the context of density estimation. Talamakrouni, Keilegom and Ghouch (2016) adapt and generalize the parametrically guided nonparametric estimation to the censored data case.

The paper is organized as follows. Section 2 introduces the gamma kernel estimators and parametrically guided nonparametric estimation for the density and the hazard rate function for right-censored data. In Section 3 we show the asymptotic properties. Via a Monte Carlo simulation, the finite sample performance of the estimators is investigated under various scenarios in Section 4.

Methodology

Let $T_1, \dots, T_n$ (survival times) be independent and identically distributed (i.i.d) nonnegative random variables with density f and common distribution function F . Let $C_1, \dots, C_n$ be a censoring variable with continuous distribution function G . Under random right censoring, instead of observing T_i , one can only observe (X_i, δ_i) where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. Based on Kaplan-Meier estimator proposed by Kaplan and Meier (1958) several nonparametric density estimators have been proposed.

$$1 - \hat{F}(t) = \prod_{i: X_i \leq t} \left(1 - \frac{1}{\sum_{j=1}^n 1_{\{X_j \geq X_i\}}} \right)^{\delta_i} \quad (2.1)$$

Blum and Susarla (1980) extended the traditional kernel-type nonparametric estimator to censored data

$$\hat{f}(x) = \frac{1}{n} \int_{-\infty}^{+\infty} K\left(\frac{t-s}{h}\right) d\hat{F}(s) \quad (2.2)$$

where K is a kernel function generally chosen to be a symmetric probability density function, $0 < h \equiv h_n$ is a bandwidth sequence and $\hat{F}(\cdot)$ is the Kaplan-Meier estimator. This method is totally nonparametric and admirably impartial to special types of shapes of the underlying density. However classical kernel leads to an estimator with a large bias near the endpoints. Bouezmarni (2011) proposed a gamma kernel estimator that corrects for the boundary effects, defined as follows:

$$\hat{f}_h(x) = \int K(x, h)(t) d\hat{F}(t) = \sum_{i=1}^n K(x, h)(X_{(i)}) W_i \quad (2.3)$$

where the kernel K is given by

$$K(x, h)(t) = \frac{t^{\rho_h(x)-1} \exp(-t/h)}{h^{\rho_h(x)} \Gamma(\rho_h(x))}, \quad \rho_h(x) = \begin{cases} \frac{x}{h} & \text{if } x \geq 2h \\ \frac{1}{4} \left(\frac{x}{h} \right)^2 + 1 & \text{if } x \in [0, 2h) \end{cases} \quad (2.4)$$

The weights W_i are the jumps of $\hat{F}$ at X_i (Suzukawa et al. (2001)).

$$W_i = \frac{\delta_{[i]}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{[j]}}, \quad i = 1, 2, \dots, n \quad (2.5)$$

The gamma kernel estimator for the hazard rate is

$$h_h(x) = \frac{\hat{f}_h(x)}{1 - \hat{F}(x)} \quad (2.6)$$

In the fully nonparametric approach, the estimators suffer from the curse of dimensionality and have in general a slower rate of convergence. However, nonparametric approach provides more flexibility since the estimation is not based on any parametrized family of functions and remains more robust and applicable in practice. The kernel estimator has a rate of convergence of $\sqrt{nh}$, (Lo et al. (1989)) which is slower compared with the $\sqrt{n}$ rate of convergence established in the parametric approach. However, a major complication that is emphasized in parametric

modeling is the risk of biased and inconsistent parameter estimation due to misspecification problem.

Hjort and Glad (1995) proposed a new scheme that contains both a parametric and a nonparametric component, called parametrically guided kernel density estimator. The essential idea behind the guided estimation is to start with a crude parametric estimator which is not necessarily well specified, then to correct this parametric guide using a particular type of correction and a nonparametric estimator. A guided nonparametric estimator takes advantage of both parametric and nonparametric methods. It always converges to the true model no matter if the parametric part is correct or not, and it adapt automatically to the parametric model if the latter is locally or globally close to the true underlying curve.

Talamakrouni, Keilegom and Ghouch (2016) adapt and generalize the parametrically guided kernel estimator to the censored data case defined as follows:

$$\hat{f}_{\hat{\theta}}(x) = \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{x-s}{h}\right) \frac{f_{\hat{\theta}}(x)}{f_{\hat{\theta}}(s)} d\hat{F}(s) = \frac{1}{h} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \frac{f_{\hat{\theta}}(x)}{f_{\hat{\theta}}(X_i)} W_i \quad (2.7)$$

The parametrically guided kernel estimator for the hazard function is

$$\lambda_{\hat{\theta}}(x) = \hat{f}_{\hat{\theta}}(x) / (1 - \hat{F}(x)) \quad (2.8)$$

Asymptotic Properties

In this section the performance of the guided kernel density estimator (2.7) is compared to that of the gamma kernel estimator (2.3). Both estimators have the bias reduction property and allows for a theoretically unbiased estimator. The multiplicative correction used in guided kernel density and hazard function does not affect the variance, the same for gamma kernel estimator. For parametrically guided kernel density estimator, the asymptotic bias and optimal bandwidth are:

$$B_{\theta^*}(x) = \frac{1}{2} h^2 \mu_K^2 \left(\frac{f(x)}{f_{\theta^*}(x)} \right)' f_{\theta^*}(x) \quad \text{and} \quad h_{opt} = \left(\frac{\int \sigma^2(x) dx}{\mu_K^4 \int (f''(x) f_{\theta^*}(x))^2 dx} \right)^{1/5} n^{-1/5}$$

For gamma kernel estimator the asymptotic bias and optimal bandwidth are:

$$B = \begin{cases} \frac{1}{2} x f''(x) h + o(h) + o(n^{-1/2} h^{-1/4}) & \text{if } x \geq 2h \\ \frac{(1-x)(\rho_h(x) - x/h)}{1 + h\rho_h(x) - x} f'(x) h + o(h) + o(n^{-1/2} h^{-1/2}) & \text{if } x \in [0, 2h) \end{cases}$$

$$\text{and } h_{opt} = \left(\frac{\frac{1}{4} \frac{\int \frac{1}{2\sqrt{\pi}} \frac{x^{-1/2} f(x)}{\overline{G}(x)} dx}{\int \frac{1}{2} x f''(x) dx} \right)^{2/5} n^{-2/5}$$

In practice, the choice of the bandwidth is a crucial issue in kernel-based density estimation. To select the bandwidth h in our case, we use unbiased cross validation method (Scott D. W. and Terrell G. R., 1987), adapted to the censoring case.

Simulations Results

In this section is studied the finite sample performance of the guided kernel density estimator and the gamma kernel estimator. Our goal is to compare the performance of the guided kernel density estimator (2.7) with that of the gamma kernel estimator (2.3) and traditional kernel. The comparison is based on Bias, MSE and the optimal bandwidth h .

The model considered is: the survival times follow a Weibull distribution with scale parameter $b=2$ and shape parameter $a=1, 2, 4$. The graphs of the resulting densities are plotted in Figure 1.

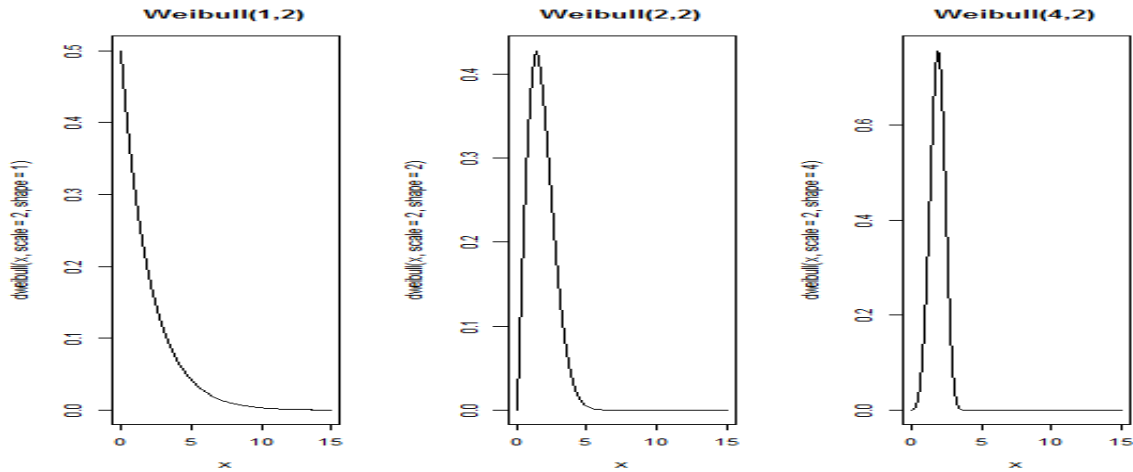


Figure 1: Weibull density with shape parameters $a = 1, 2, 4$ and scale parameter $b = 2$.

The censoring times are also generated from a Weibull distribution with shape parameter a and a scale parameter given by $b((1-p)/p)^{1/a}$, ensuring a degree of censoring equal to p . We consider two censoring rates $p=10\%$ and $p=40\%$ and sample size $n=200$. For parametrically guided kernel density estimator, as a parametric guide we use the exponential density $f_{\theta}(t) = \theta \exp(-\theta t)$ where θ is estimated using the approximated maximum likelihood estimator. The only situation where the guide is correctly specified is the case $a=1$, in the other cases the parametric guide deviates gradually from the true density.

Table 1: Squared bias ($\times 10^5$), MSE ($\times 10^5$), the optimal bandwidth h , for the estimators of several Weibull densities for $a = (1, 2, 4)$, two censoring rates $p = (10\%, 40\%)$ and sample size $n=200$.

	10%					40%		
a	Method	Bias	MSE	h		Bias	MSE	h
1	PGK	0.09	110.7	8		1.67	189.7	7.9
	GK	0.009	112.5	8		0.67	187.6	8.01
	TK	22.98	99.2	9		28.65	260.4	9.01
2	PGK	89	260.7	4.2		89	445.1	4.2
	GK	87.99	244.3	4.5		87.99	446	4.46
	TK	110.14	270.1	4		110.14	589.23	4
4	PGK	87.2	344.2	3		87.2	587	3
	GK	86.51	304.3	2.28		86.51	524.2	3.2
	TK	210.5	689.99	3.9		210.5	890.98	4

We get the best results for the PGK estimator when $a=1$ (a correct parametric guide). The bias of the PGK estimator is significantly reduced compared to that of TK estimator, but GK estimator gives a smaller bias compared to both of them. Regarding the MSE, it is also reduced for the PGK estimator compared to the MSE of the TK estimator. MSE is also reduced for the GK estimator compared to the MSE of the TK and PGK estimator. For $a=2$ and $a=4$, even if the parametric guide is incorrect, the PGK estimator remains significantly better than the TK estimator, while the GK estimator has a significantly smaller bias than both the PGK and TK estimator.

Along the simulations we consider the Gaussian kernel function K (Hansen B, 2009) and, for every estimator, we only show the results corresponding to the optimal tuning parameters, i.e. those which minimize the empirical mean squared error (MSE). The choice of the bandwidth is made by unbiased cross-validation bandwidth selection method, adapted to the censoring case.

Conclusion

In this paper, we investigated a parametrically guided kernel and a gamma kernel estimator for censored data. The PGK estimator is obtained by multiplying an initial parametric estimator by a nonparametric kernel type estimator of a suitable correction function. The simulation results confirm the bias reduction property. We showed that the bias of the PGK estimator and GK estimator can be reduced compared to that of the traditional kernel estimator, while the GK estimator has a significantly smaller bias than both the PGK and TK estimator.

References

- Blum, J. R. and Susarla, V. (1980). Maximal deviation theory of density and failure rate estimates based on censored data. *J. Multiv. Anal.* 5, 213-222.
- Bouezmarni, T., El Ghouch, A. and Mesfioui, M. (2011). Gamma kernel estimators for density and hazard rate of right-censored data. *J. Prob. Stat.* 2011, 16 pages.
- Glad, I. K., Hjort, N. L. and Ushakov, N.G. (2003). Correction of density estimators that are not densities. *Scand. J. Stat.* 30(2), 415-427.
- Hansen B., (2009). Kernel Density Estimation, University of Wisconsin, spring 2009
- Hjort, N.L. (1992). On inference in parametric survival data models. *Int. Stat. Rev.* 60, 355-387.
- Hjort, N. L. and Glad, I. K. (1994). Nonparametric density estimation with a parametric start. Statistical research report, Dept. Mathematics, Univ. Oslo.
- Hjort, N. L. and Glad, I. K. (1995). Nonparametric density estimation with a parametric start. *Ann. Statist.* 23, 882-904
- Kaplan, E. and Meier, P. (1958). "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, vol. 53, pp. 457–481.
- Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis. Techniques for Censored and Truncated Data*. Springer, New York.
- Scott D. W. and Terrell G. R., 1987. Biased and unbiased cross-validation in density estimation. *J. Amer. Statist. Assoc.* 82 1131–1146.
- Suzukawa, A., Imai, H. and Sato, Y. (2001). Kullback-Leibler information consistent estimation for censored data. *Ann. Inst. Statist. Math.* 53, 262-276.
- Talamakrouni, M., Van Keilegom, I. and El Ghouch, A. (2016a). Parametrically guided nonparametric density and hazard estimation with censored data. *Comput. Statist. Data Anal.* 93, 308-323

How do Accounting Professionals Perceive Whistleblowing Reasons and Whistleblowing Preferences?

M. SaitDinc¹, Cemil Kuzey², Ali Haydar Gungormus³, Bedia Atalay⁴

¹Department of Management, International Burch University
Bosnia and Herzegovina, m.sait.dinc@ibu.edu.ba

²Computer Science & Information Systems, Arthur J. Bauernfeind College of
Business

Murray State University, United States of America, cemilkuzey@gmail.com

³a.h.gungormus@gmail.com

⁴bediaatalay@gmail.com

Abstract

Incidences of organizational wrongdoing have been widely spread throughout the business world. Accounting professionals are the key human resources who find evidence of wrongdoing in firms and have the opportunity to report it. The purpose of this study is to examine the relationship between the perception by accounting professionals concerning valid reasons for whistleblowing and their preferences in doing so. Using the survey method, 177 responses were collected from Turkish accounting professionals. A partial least square structural equation model was constructed to test both the reliability and validity of the measurement and the structural model. The results showed that the 'fear of retaliation' dimension has a significant negative influence on 'external whistleblowing' but has a positive influence on 'anonymous whistleblowing'. The accountants' perceptions of 'fear of retaliation' also has a positive relationship for deciding to not blow the whistle. However, thereasons for 'corporate benefit' whistleblowing have a positive effect on both 'anonymous whistleblowing' and 'internal whistleblowing'. They also have a negative impact on the reasons that accounting professionalspreferto not blow the whistle. Finally, the 'ethics and professional benefit' dimension of reasons for whistleblowing only has a significant positive impact on 'internal whistleblowing'.

Keywords: Whistleblowing reasons, Whistleblowing, Accounting professionals, Partial least square structural equation model.

JEL Classification Codes:M40, M41, C83

Introduction

Incidences of wrongdoing in organizations have been spread widely throughout the business world. A recent rash of corporate wrongdoing has made individuals skeptical about many business financial activities. One of the reactions to wrongdoing in organizations is whistleblowing by individuals. Whistleblowing is briefly defined as reporting wrongdoing to an individual or organization who is believed to have the power to correct the problem (Near and Miceli, 1985). Since the benefits of whistleblowing tefor wider society seem to be increasingly well accepted (Park, Blenkinsopp, Öktem, Omurgonulsen, 2008), research interest in the subject has increased (Taylor, and Curtis, 2010; Nayir and Herzig, 2012). According to whistleblowing studies, thegreatest number of wrongdoing reports is made by individuals close to the inner workings of the organization (Mesmer-Magnus, and Viswesvaran, 2005). Among these individuals are accounting professionals.

Accounting professionals are the key human resources who measure and disclose the financial information of organizations to the public (Ferrell, Fraedrich, & Ferrell, 2015). Due to their close relationship with companies, they may discover wrongdoing in firms and blow the whistle about it. The aim of this study is to examine the relationship between the perceptions of accounting professionals' regarding reasons for whistleblowing and their own whistleblowing.

There are several contributions to this study. The focus of whistleblowing research generally identifies the conditions under which whistleblowing intentions are formed and action taken, and retaliation occurs (Ellis and Arieli, 1999; Miceli and Near, 2002). Literature concentrating on the relationship between whistleblowing reasons and whistleblowing is severely limited. Empirical studies which examine this relationship in the context of non-Western countries are also lacking. This study tries to fill this gap in the whistleblowing literature. In this regard, the purpose of the study is to explore the impact of the perception of Turkish accounting professionals regarding whistleblowing reasons on their self-reported reasons to choose specific whistleblowing modes.

Literature Review

Whistleblowing is "the disclosure by organization members (former or current) of illegal, immoral, or illegitimate practices under the control of their employers, to persons or organizations that may be able to effect action." (Near and Miceli, 1985, p. 4). Whistle-blowers can conceivably help organizations to correct wrong and harmful activities such as unsafe products or to prevent fraudulent practices and, thus, to avoid substantial adverse consequences in the long run (Miceli and Near, 1985). On the other hand, they may threaten the organization's authority structure (Weinstein, 1979). Many studies about the ways in which employees might blow the whistle have been made, distinguishing these avenues (Grant, 2002; Park, Rehg and Donggi, 2005). One of these studies is Park et al.'s whistleblowing typology which is based on three dimensions. Each dimension in this typology represents a whistleblowing choice for the employee – formal versus informal, identified versus anonymous, and internal versus external. *Formal whistleblowing* is an institutional form in which a whistleblower reports wrongdoing by following the standard organizational procedures, while the whistleblower reports wrongdoing by personally telling close associates or superiors in the *informal whistleblowing*. Rohde-Liebenau (2006) categorizes this as authorized versus unauthorized whistleblowing. *Identified* versus *anonymous whistleblowing* is based upon whether a whistleblower uses his or her real name in reporting a wrongdoing. The whistleblower sometimes uses an assumed name in the latter case. Finally, internal versus external whistleblowing is related to which inside or outside authority to which the whistleblower provides information. Employees report wrongdoing to a supervisor or another authority within the organization in *internal whistleblowing*, whereas they prefer reporting it to outside agencies in *external whistleblowing*.

Regarding distinguished ways to blow the whistle, research suggests that nearly all whistleblowers attempt to report wrongdoing through internal channels before using external channels (Miceli and Near, 1992, 2002). Despite the less threatening characteristics of internal whistleblowing to organizations, it is often not welcomed and rather, whistleblower reports about wrongdoing are frequently ignored (Miceli et al., 1991). Due to this reason and a belief that no corrective action will be taken, whistleblowers may be reluctant to blow the whistle. Verschoor (2005)'s findings that

almost half (44%) of all employees who are aware of individual or corporate wrongdoing do not disclose their observations to anyone supports this theory of whistleblower² reluctance. However, whistleblowers are likely to prefer external channels rather than attempting to make an internal report when they observe the wrongdoing continuing. This appears to be especially strong when the whistleblower fears retaliation from the organization, its supervisors, or their coworkers (Miceli and Near, 1985).

While retaliation is observed as one of the serious consequences of whistleblowing in the literature, it can serve as an antecedent or predictor as well (Miceli and Near, 1985). Management may take two types of steps when presented with whistleblowing: disregarding or taking appropriate actions and rewarding or retaliating against the whistleblower (Near and Miceli, 1986). The main motivation of organizations to engage in retaliatory acts may be due to their desire to (1) silence the whistleblower completely, (2) prevent the full public knowledge of the problem, (3) discredit the whistleblower, and/or (4) discourage other potential whistleblowers from coming forward (Miceli and Near, 1994; Parmerlee, Near and Jensen, 1982). There are other reasons for whistleblowing. Dasgupta and Kesharwani (2010) categorized the reasons into the altruistic perspective of the whistleblowers, motivational and psychological perspective, and prospective reward. Altruistic concerns briefly refer to concern for the well-being of others (Dinc, and Aydemir, 2014). The altruistic reason for whistleblowing is the desire to correct the wrongdoing which is harmful to the interests of the organization itself, the consumers, the co-workers and/or society at large (Vandekerckhove and Commers, 2004). Whereas the desire for financial recovery can be one motivation for whistleblowing, an expected reward from the organization for reporting the wrongdoings may constitute another reason.

Whistleblowing is perceived as a negative act in Turkey where this study was conducted. According to Nayir and Herzig (2012), openly complaining about ethical misconduct has not been common in Turkey. They proved this observation through evidencing from The Global Corruption Barometer Report published by Transparency International (2009), which demonstrates that many individuals do not lodge formal complaints out of fear of potential harassment and retaliation. There are several studies contributing to the understanding of whistleblowing in Turkey. One of them is Park et al. (2008)'s study which reveals that there is a preference for internal over external reporting in general. Another study has been completed, identifying the relationship between Turkish teachers' perception of whistleblowing and the reasons it occurs (Celep and Konakli, 2012). Their findings demonstrated that teachers have a tendency to blow the whistle because of their positive feelings for their organization and associates. They also found that women have a greater whistleblowing tendency than men. Based on the previous literature discussed above, the following hypotheses are posited:

Hypothesis 1a: The fear of retaliation has a significant negative influence on external whistleblowing.

Hypothesis 1b: The fear of retaliation has a significant positive influence on anonymous whistleblowing.

Hypothesis 1c: The fear of retaliation has a significant negative influence on internal whistleblowing

Hypothesis 1d: The fear of retaliation has a significant positive influence on no whistleblowing

Hypothesis 2a: The corporate benefit has a significant positive influence on external whistleblowing.

Hypothesis 2b: The corporate benefit has a significant positive influence on anonymous whistleblowing.

Hypothesis 2c: The corporate benefit has a significant positive influence on internal whistleblowing.

Hypothesis 2d: The corporate benefit has a significant negative influence on no whistleblowing.

Hypothesis 3a: The ethics and professional benefit has a significant positive influence on internal whistleblowing.

Hypothesis 3b: The ethics and professional benefit has a significant positive influence on external whistleblowing.

Hypothesis 3c: The ethics and professional benefit has a significant positive influence on anonymous whistleblowing.

Hypothesis 3d: The ethics and professional benefit has a significant positive influence on no whistleblowing.

Research Methodology

The study utilized online questionnaire survey to collect data. There has been an increasing trend to utilize online questionnaires recently. The advantages of online questionnaires over traditional questionnaires are perceived as being less costly, faster, and more reliable (Uyar, Kuzey, Güngörmüş, and Alas, 2015). Data was collected during the months of November and December 2016. A total of 177 accountant professionals replied to the survey.

In order to test the hypothetical relationship (Fig. 1), we employed Partial Least Square Structural Equation Modeling (PLS-SEM). Multi-items from prior studies were adapted from English language resources. In addition, 5-point Likert scales were utilized in order to assess the items. Accordingly, 5 represented “strongly agree” while 1 represented “strongly disagree”.

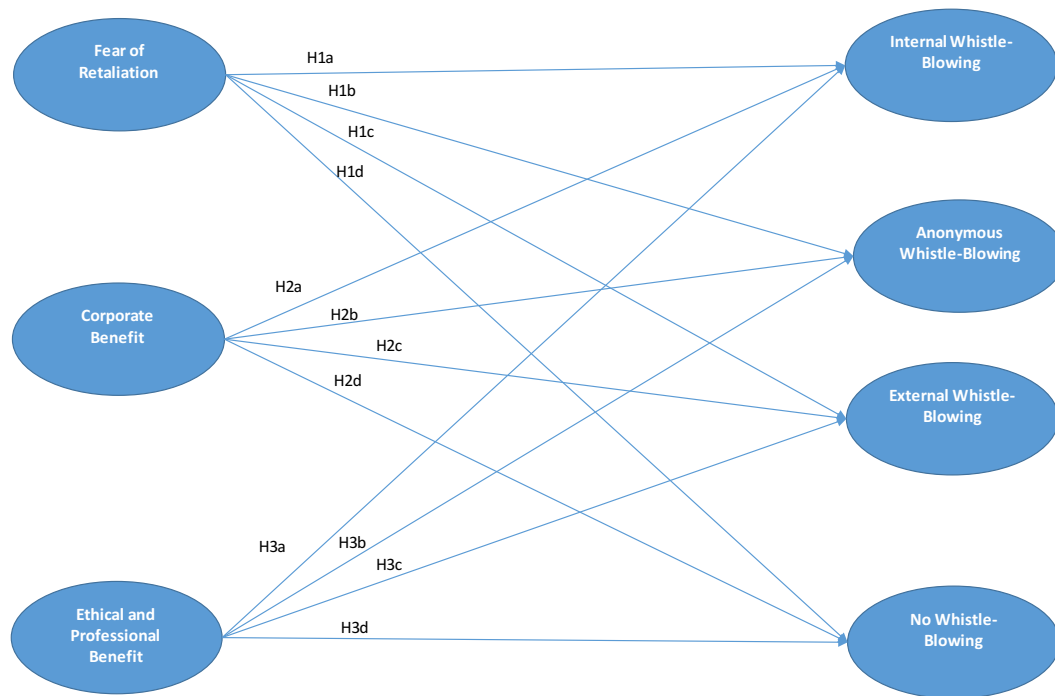


Figure 1: Proposed Model

The Cronbach's Alpha for the constructs ranged between .75 and .94. The whistleblowing reasons scale, including Fears of Retaliation (Cronbach's Alpha=.92), Corporate Benefits (Cronbach's Alpha=.89), Ethical and Professional Benefits (Cronbach's Alpha=.94) dimensions was adapted from Celep and Konakli (2012). The whistleblowing scale which consisted of the dimensions of Internal Whistle-Blowing (Cronbach's Alpha=.75), Anonymous Whistle-Blowing (Cronbach's Alpha=.88), and the External Whistle-Blowing (Cronbach's Alpha=.87) was adapted from Celep and Konakli (2012). They adapted these two constructs from Park et al. (2005) and Park et al. (2008). Finally, the No Whistle-Blowing (Cronbach's Alpha=.76) construct was adapted from Tak, (2010).

Preprocessing of the Data

Preprocessing of the data is important before the testing the hypothesis. In this case, we screened that data by assessing the missing data, imputed, checking the univariate as well as multivariate outlier values, normality, linearity and investigating any unengaged respondents. A missing data analysis was performed. The Little's MCAR test results showed that the missing data were random ($\chi^2=25.46$; df: 33; sig.=.82). The missing data was imputed using linear regression as the model for scale variable. Moreover, the collinearity analysis showed no significantly high correlation among the independent variables since the VIF (Variance Inflation Factor) scores ranged between 1.17 and 1.36 which are significantly lower than the threshold value of 10 (Hair et al. 2010).

Sample

The samples size was 177 after preprocessing steps. The frequency analysis results are shown in Table 1. The results indicated that 79% of the participants were male, that 83.4% were married, that almost 41% were between 41 and 50 years old, that approximately 70% held undergraduate diplomas, that 24.3% had work experience ranging between 11 and 15 years, that almost 70% were chartered public accountants,

and that approximately 25% had an monthly income between 3001 TRY and 4000 TRY.

Table 1: Frequency analysis

Variables		Frequency	Percent
Gender	Female	37	20.90
	Male	140	79.10
	Total	177	100.00
Marital Status	Single	27	15.30
	Married	148	83.60
	Total	177	100.00
Age (in Years)	20-25	2	1.10
	26-60	13	7.30
	31-40	64	36.20
	41- 50	72	40.70
	51 +	26	14.70
	Total	177	100.00
Education	High School	6	3.40
	College	3	1.70
	Undergraduate	123	69.50
	Master's	41	23.20
	Doctorate	4	2.30
	Total	177	100.00
Experience (in Years)	1-5	12	6.80
	6-10	19	10.70
	11-15	43	24.30
	16-20	29	16.40
	21-25	39	22.00
	26 +	35	19.80
	Total	177	100.00
Title	Accounting Internship	13	7.30
	CPA	123	69.50
	CPA, Independent Auditor	29	16.40
	Chartered Accountant	5	2.80
	Independent Auditor	6	3.40
	Academician	1	0.60
	Total	177	100.00
Monthly Income (TRY)	1300-2000	17	9.60
	2001-3000	21	11.90
	3001-4000	44	24.90
	4001-5000	21	11.90
	5001 +	74	41.80
	Total	177	100.00

The average of latent variables ranged between 2.71 and 4.39 (Table 2), also the standard deviation values showed little variation around the mean.

Table 2: Descriptive statistics and Pearson Correlation analysis (N=177)

	Variables	Mean	Std. Deviation	1	2	3	4	5	6	7
1	FR	2.71	1.28	0.87						
2	CB	3.99	1.09	-0.07	0.87					
				-						
3	EPB	4.39	0.94	0.15*	0.61**	0.92				
						0.48*				
4	IW	4.29	0.96	-0.11	0.46**	*	0.82			

				0.23*						
5	EAW	2.10	1.26	*	0.10	0.01	0.12*	0.89		
6	EIW	2.25	1.26	-0.15	0.05	-0.01	0.11	0.40**	0.83	
				0.24*	-	-	-	-	-	0.9
7	NWB	1.78	1.12	*	0.26**	0.20*	0.38**	0.21**	0.29**	0

*IW: Internal Whistle-Blowing; EAW: External Anonymous Whistle-Blowing; NWB: No Whistle-Blowing; EIW: External Identified Whistle-Blowing; FR: Fear of Retaliation; CB: Corporate Benefits; EPB: Ethical Professional Benefits; *p<.05; **p<.01; Values at the diagonals are the square roots of AVE (Average Variance Extracted) scores*

Measurement Model

The factor loading of the constructs is shown in Table 3. Thirty-two items were assessed initially, using PLS-based factor analysis. Following the initial assessment, seven items were eliminated due to low factor loadings. The factor loadings were all above the suggested value of .70, which also demonstrated that the convergent validity was satisfied (Chin.1998). The factor loadings were also used to determine the discriminant validity by comparing the related value of a construct with that of the other construct. It was clear that each factor loading value of a particular latent variable had a higher factor loading than the rest of the values in the row level as well as in the column level. There were no high cross loadings. The Cronbach's Alpha values for each latent variable were above the threshold value of .70

Table 3: Cross factor loadings

Items	FR	CB	EPB	NWB	IW	ECW	EOW
FR1	0.86	-0.06	-0.09	0.20	-0.17	0.26	-0.13
FR2	0.90	-0.05	-0.12	0.19	-0.05	0.22	-0.12
FR3	0.93	-0.08	-0.14	0.21	-0.11	0.20	-0.13
FR4	0.86	-0.06	-0.16	0.23	-0.09	0.16	-0.10
FR5	0.82	-0.07	-0.17	0.22	-0.06	0.15	-0.15
CB2	-0.05	0.74	0.41	-0.17	0.32	-0.02	-0.04
CB3	-0.03	0.91	0.61	-0.27	0.48	0.10	0.07
CB4	-0.06	0.90	0.50	-0.19	0.37	0.13	0.08
CB5	-0.11	0.91	0.57	-0.26	0.42	0.10	0.03
EPB2	-0.16	0.54	0.92	-0.17	0.41	0.01	-0.03
EPB3	-0.19	0.59	0.95	-0.23	0.48	0.03	0.02
EPB4	-0.16	0.58	0.91	-0.22	0.46	0.01	0.04
EPB5	-0.04	0.55	0.91	-0.11	0.43	-0.03	-0.10
NWB1	0.28	-0.19	-0.09	0.90	-0.31	-0.17	-0.28
NWB2	0.28	-0.22	-0.09	0.90	-0.31	-0.17	-0.28
IW3	-0.11	0.25	0.32	-0.24	0.73	0.06	0.04
IW6	-0.01	0.48	0.44	-0.30	0.87	0.09	0.07
IW7	-0.17	0.37	0.40	-0.39	0.84	0.14	0.16
AW1	0.24	0.10	0.00	-0.21	0.17	0.91	0.28
AW2	0.12	0.09	0.02	-0.24	0.07	0.85	0.45
AW3	0.23	0.07	0.00	-0.14	0.07	0.92	0.38
EW1	-0.02	0.15	0.11	-0.32	0.25	0.42	0.73
EW2	-0.12	0.06	-0.02	-0.23	0.10	0.41	0.90
EW3	-0.06	0.14	0.05	-0.23	0.14	0.45	0.83
EW4	-0.19	-0.06	-0.07	-0.25	0.03	0.19	0.87

IW: Internal Whistle-Blowing; AW: Anonymous Whistle-Blowing; NWB: No Whistle-Blowing; EW: External Whistle-Blowing; FR: Fear of Retaliation; CB: Corporate Benefits; EPB: Ethical Professional Benefits

The latent variables were subject to confirmatory factor analysis (CFA) to investigate the construct validity. We employed the maximum likelihood approach for testing the construct validity of the measurement model. Table 5 shows the standardized

Regression weights, as well as the fit measures. The CFA results revealed a good fit since $\chi^2/df = 1.85$; goodness of fit index (GFI) = .84; normed fit index (NFI) = .87; comparative fit index (CFI) = .93; relative fit index (RFI) = .84; incremental fit index (IFI) = .93; Tucker-Lewis index (TLI) = .92; and root mean square error of approximation (RMSEA) = .06. The model fit results showed that the recommended values were satisfied (Hu and Bentler, 1999). The t-statistics of the individual items with respect to their latent variables were statistically significant at the 1% significance level, which in turns indicated that the convergent validity was satisfied.

Table 4: Construct validity and reliability results

Variables	AVE	C.R.	R ²	C.A.
Fear of Retaliation	0.76	0.94		0.92
Corporate Benefits	0.75	0.92		0.89
Ethical Professional Benefits	0.85	0.96		0.94
Internal Whistle-Blowing	0.67	0.86	0.28	0.75
Anonymous Whistle-Blowing	0.80	0.92	0.07	0.88
External Whistle-Blowing	0.70	0.90	0.03	0.87
No Whistle-Blowing	0.81	0.89	0.12	0.76

C.R.: Composite reliability; C.A.: Cronbach's Alpha; AVE: Average Variance Extracted

The association between the latent variables and their corresponding indications were modeled using the reflective approach in order to assess the construct validity. In the reflective model, dropping an item from the current model does not affect the construct. The discriminant validity and construct reliability, as well as the item reliability were investigated using average variance extracted values, correlation coefficients together with the square root of average variance extracted values at the diagonal of the matrix, and composite reliability values. Individual reliability was satisfied since the factor loadings of each item were above .70. The composite reliability values of each latent variable were above the recommended value of .70, which indicated internal consistency (Nunnally, 1987). In addition, the discriminant validity was determined using the average variance extracted values as well as the square root of these values to be compared with the correlation coefficients of the latent variables (Fornell and Larcker, 1981). The average variance extracted values of the latent variables were all above the threshold values of .50 (Table 4). Additionally, the square root of the average variance extracted values of a particular latent variable were located on the diagonal of the matrix (Table 2), which were greater than the correlation coefficients of the other latent variables. These two assessments showed that the discriminant validity was satisfied.

Table 5: Standardized Regression weights

Latent Variables	Items	Estimate	t-statistics
IW →	IW7	0.76	Scaling
IW →	IW6	0.79	8.66
IW →	IW3	0.58	6.89
AW →	AW1	0.80	Scaling
AW →	AW2	0.83	11.85
AW →	AW3	0.89	12.52
NWB →	WBC1	0.76	Scaling
NWB →	WBC2	0.80	6.87
EPB →	EPB5	0.86	Scaling
EPB →	EPB4	0.89	15.89
EPB →	EPB3	0.95	17.84
EPB →	EPB2	0.90	18.14
CB →	CB5	0.84	Scaling

CB	→	CB4	0.82	16.17
CB	→	CB3	0.92	13.90
CB	→	CB2	0.64	9.13
FR	→	FR1	0.78	Scaling
FR	→	FR2	0.88	14.75
FR	→	FR3	0.93	13.51
FR	→	FR4	0.85	12.14
FR	→	FR5	0.78	10.49
EW	→	EW1	0.79	Scaling
EW	→	EW2	0.79	10.71
EW	→	EW3	0.88	11.62
EW	→	EW4	0.63	8.13

IW: Internal Whistle-Blowing; *AW*: Anonymous Whistle-Blowing; *NWB*: No Whistle-Blowing; *EW*: External Whistle-Blowing; *FR*: Fear of Retaliation; *CB*: Corporate Benefits; *EPB*: Ethical Professional Benefits; $\chi^2/df=1.85$; *GFI*= .84; *NFI*= .87; *CFI*= .93; *RFI*= .84; *IFI*= .93; *TLI*= .92; *RMSEA*= .06.

Structural Model

The Structural equation modelling results were based upon the PLS approach. PLSE-SEM was utilized due to its advantage in terms of small sample size and restrictive assumptions. The visual representation of the hypothetical relationships and their corresponding results are illustrated in Figure 2 and Table 6. According to the obtained results, the fear of retaliation had a statistically significant and positive impact on internal whistle-blowing ($\beta=.236$, $p<.01$) and no whistle-blowing ($\beta=.219$, $p<.05$) while it had a negative impact on external whistle-blowing ($\beta=-.153$, $p<.05$). In addition, the corporate benefit was positively associated with internal whistle-blowing ($\beta=.269$, $p<.05$), and with anonymous whistle-blowing ($\beta=.143$, $p<.05$), while it was negatively associated with no whistle-blowing ($\beta=-.231$, $p<.05$). Finally, there was a significantly positive relationship between the ethical and professional benefits and internal whistle-blowing ($\beta=.311$, $p<.05$). Thus, the results indicated that hypotheses H1 and H2 were partially supported, while H3 was marginally supported.

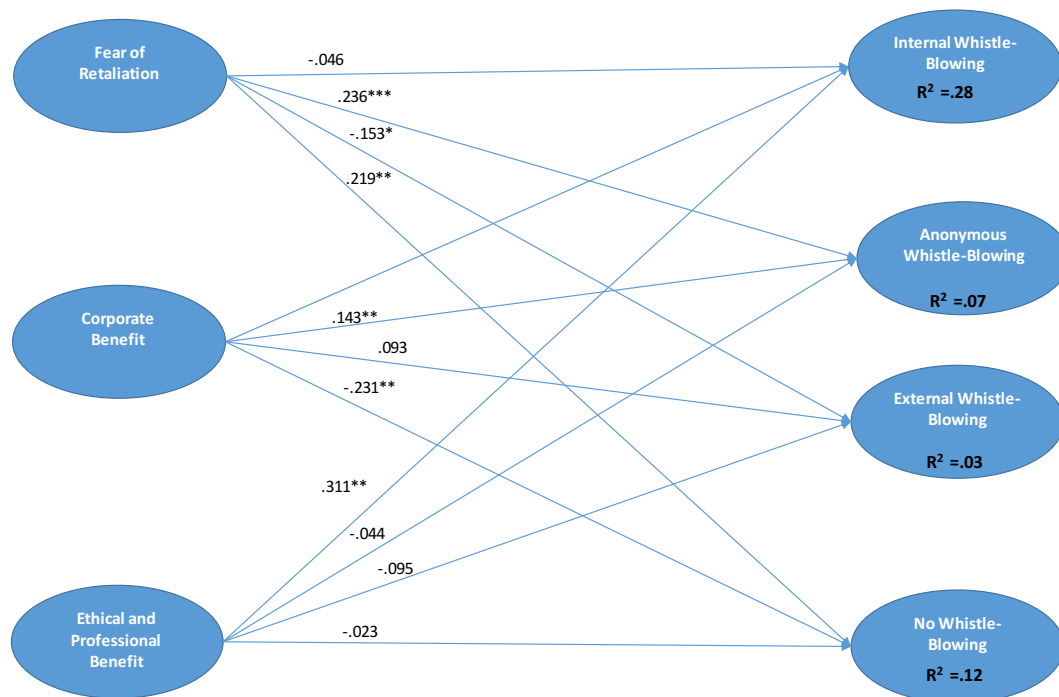


Figure 2: Hypothesis testing(*p<.10; **p<.05; ***p<.01)

The R^2 values were used to show the explained variance by the independent latent variables. The value of R^2 was classified as substantial ($>.26$), moderate (between .13 and .26), and weak ($<.13$). Based on this criteria, Internal Whistle-blowing ($R^2 = .28$) had substantial explanatory power while No Whistle-blowing ($R^2 = .28$) was weakly moderate, and both Anonymous Whistle-blowing ($R^2 = .07$), and External whistle-blowing ($R^2 = .03$) had a weak explanatory power.

Table 6: The structural equation modelling results

Hypothesis	Hypothetical Relationships			Beta	t-statistics	Results
H1a	FR	→	IW	-0.046	0.94	Not Supported
H1b	FR	→	AW	0.236***	2.95	Supported
H1c	FR	→	EW	-0.153*	1.81	Supported
H1d	FR	→	NWB	0.219**	2.87	Supported
H2a	CB	→	IW	0.269**	2.60	Supported
H2b	CB	→	AW	0.143**	2.02	Supported
H2c	CB	→	EW	0.093	0.97	Not Supported
H2d	CB	→	NWB	-0.231**	1.97	Supported
H3a	EPB	→	IW	0.311**	2.46	Supported
H3b	EPB	→	AW	-0.044	0.70	Not Supported
H3c	EPB	→	EW	-0.095	1.15	Not Supported
H3d	EPB	→	NWB	-0.023	0.29	Not Supported

*IW: Internal Whistle-Blowing; AW: Anonymous Whistle-Blowing; NWB: No Whistle-Blowing; EW: External Whistle-Blowing; FR: Fear of Retaliation; CB: Corporate Benefits; EPB: Ethical Professional Benefits; * $p < .10$; ** $p < .05$; *** $p < .01$*

Discussion and Conclusion

The purpose of this study was to examine the relationship between the perception of accounting professionals about reasons for whistleblowing and their preferences for whistleblowing modes. The study revealed expected findings. One of the conclusions of this research is that ‘the ethics and professional benefit’ dimension of whistleblowing reasons has a positive effect on ‘internal’ whistleblowing. Accounting professionals demonstrated that their ethical and professional values direct them to do internal whistleblowing to correct wrongdoing. The finding of the study is consistent with the research about whistleblowing which suggest that nearly all whistleblowers initially attempt to report wrongdoing via internal channels before using external channels (Miceli and Near, 1992, 2002).

Another important conclusion of the study is that the ‘corporate benefit’ dimension of whistleblowing reasons has a positive influence on the ‘external’, and ‘anonymous’, dimensions of whistleblowing preferences of accounting professionals, while it has a negative effect on the ‘no whistleblowing’ construct. Participants showed that accounting professionals with altruistic concerns for corporations do not want their organizations to be harmed by wrongdoing, and thus will engage in external whistleblowing. They cannot keep silent about detrimental misconduct and always blow the whistle anonymously. This conclusion is supported by the literature (Nayir and Herzig, 2012; Vandekerckhove and Commers, 2004; Celep and Konakli, 2012). These studies assume that as long as organizational procedures for internal whistleblowing such as anonymous whistleblowing hotlines are in place people will benefit from this opportunity instead of announcing organizational wrongdoings publicly. But, whistleblowers primarily prefer external whistleblowing ways when internal complaint mechanisms seem to have failed.

Finally, the research implies that ‘the fear of retaliation’ dimension in whistleblowing reasons has a positive effect on ‘anonymous’ whistleblowing, and ‘no whistleblowing’ construct, whereas it has a negative impact on ‘internal whistleblowing’. Accounting professionals stated that they would raise their voice against the wrongdoings they observed by preferring an anonymous form of whistleblowing or even not to blow the whistle due to their perceptions creating fear of retaliation. However, they prefer not to blow the whistle internally. These findings are consistent with whistleblowing literature (Miceli and Near, 1985). They are also related to the study context. Previous research findings show that whistleblowing in Turkey is often viewed as risky for individuals (Nayir and Herzig, 2012). These research findings suggest that whistleblowing is beneficial for organizations in a long run. Because of this reason, whistleblower reports of wrongdoing should not be ignored by organizations and whistleblowers should not have any fear of retaliation from both organization, coworkers and public. To provide it for whistleblowers, officials in Turkey should work on laws and regulations protecting the rights of whistleblowers like the whistleblower protection law in the USA (e.g., Sarbanes–Oxley Act of 2002).

There are some limitations in this research. First, the results of this study come from a limited sample. Surveys with higher sample sizes may give different results. Second, self-reported issue may be another limitation of this sensitive study. Final limitation of this research is the insufficient literature. Future study should add other whistleblowing preferences such as ‘identified’ whistleblowing to whistleblowing reasons and whistleblowing preferences relationship.

References

- Celep, C. and Konakli, T. (2012). Bilgi Uçurma: Eğitim Örgütlerinde Etik ve Kural Dışı Uygulamalara Yönelik Bir Tepki. *E-international journal of education al research*, 4(3), 65-88.
- Chin, W.W. (1998). *The partial least squares approach for structural equation modeling*, in Marcoulides, G.A. (Ed.), *Modern methods for business research*, Lawrence Erlbaum Associates, Mahwah, NJ, 295-336.
- Dasgupta, S. and Kesharwani, A. (2010). Whistleblowing: A Survey of Literature. *The IUP Journal of Corporate Governance*, 9(4), 2-14.
- Dinc, M.S. and Aydemir, M. (2014). Ethical leadership and employee behaviours: an empirical study of mediating factors. *International Journal of Business, Governance and Ethics*, 9(3), 293–312.
- Ellis, S. and Arieli, S. (1999). Predicting Intentions to Report Administrative and Disciplinary Infractions: Applying the Reasoned Action Model. *Human Relations*, 52(7), 947–967.
- Fornell, C. and Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
- Grant, C. (2002). Whistle Blowers: Saints of Secular Culture. *Journal of Business Ethics*, 39(4), 391–399.
- Hair, J.F., Jr., Anderson, R.E., Tatham, R.L. and Black, W.C. (2010). *Multivariate Data Analysis with Readings*. 7th Ed. Englewood Cliffs, Prentice Hall, NJ.
- Hassink, H., Vries, M., and Bollen, L. (2007). A content analysis of whistleblowing policies of leading European companies. *Journal of Business Ethics*, 75, 25-44.

- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.
- Mesmer-Magnus, J. R. and Viswesvaran, C. (2005). Whistleblowing in Organizations: An Examination of Correlates of Whistleblowing Intentions, Actions, and Retaliation. *Journal of Business Ethics*, 62, 277-297.
- Miceli M. P. and Near J. P. (1985). Characteristics of Organizational Climate and Perceived Wrongdoing Associated with Whistle-blowing Decisions. *Personal Psychology*, 38, 525, 544.
- Miceli, M. P. and Near, J. P. (1992). *Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees* (Lexington, New York).
- Miceli, M. P. and Near, J. P. (2002). What Makes Whistle-Blowers Effective? Three Field Studies. *Human Relations*, 55(4), 455-479.
- Miceli, M. P., Near, J. P. and Schwenk, C. R. (1991). Who Blows the Whistle and Why? *Industrial and Labor Relations Review*, 45(1), 113-130.
- Miceli, M. P. and Near, J. P. (1994). Relationships Among Value Congruence, Perceived Victimization, and Retaliation Against Whistle-Blowers. *Journal of Management*, 20(4), 773-794.
- Nayir, Z. D. and Herzig, C. (2012). Value Orientations as Determinants of Preference for External and Anonymous Whistleblowing. *Journal of Business Ethics*, 107, 197-213
- Near, J. P. and Miceli, M. P. (1985). Organizational Dissidence: The Case of Whistleblowing. *Journal of Business Ethics*, 4(1), 1-16.
- Near, J. P. and Miceli, M. P. (1986). Retaliation against Whistle Blowers: Predictors and Effects. *Journal of Applied Psychology*, 71(1), 137-145.
- Nunnally, J. (1978), *Psychometric Theory*, 2nd ed., McGraw-Hill, New York, NY.
- Ferrell, O.C., Fraedrich, J. and Ferrell, L. (2015). *Business Ethics: Ethical Decision Making and Cases*, 10th Ed. Cengage Learning.
- Park, H., Rehg, M. T. and Donggi, L. (2005). The Influence of Confucian Ethics and Collectivism on Whistleblowing Intentions: A Study of South Korean Public Employees. *Journal of Business Ethics*, 58(4), 387-403.
- Park, H., Blenkinsopp, J., Öktem, M.K., Omurgonulsen, U. (2008). Cultural orientation and attitudes toward different forms of whistleblowing: a comparison of South Korea, Turkey and the U.K. *Journal of Business Ethics*, 82, 929-939.
- Parmerlee, M. A., Near, J. P. and Jensen, T. C. (1982). Correlates of Whistle-Blowers' Perceptions of Organizational Retaliation. *Administrative Science Quarterly*, 27, 17-34.
- Rohde-Liebenau, B. (2006). Whistleblowing Rules: Best practice; Assessment and Revision of Rules Existing in EU Institutions. IPOL/D/CONT/ST/2005_58 (European Parliament, Brussels).
- Tak, B. (Nisan 2010). Hastanelerde Hasta Güvenliğini Tehdit Eden Olayların Raporlanması: Türkiye, Azerbaycan, Bosna, Arnavutluk, Lübnan ve Suriye'yi Kapsayan Karşılaştırmalı Bir Araştırma. II. Sağlıkta Kalite ve Performans Kongresi, Antalya
- Taylor, E. Z. and Curtis, M. B. (2010). An Examination of the Layers of Workplace Influences in Ethical Judgments: Whistleblowing Likelihood and Perseverance in Public Accounting. *Journal of Business Ethics*, 93:21-37.
- Uyar, A., Kuzey, C., Güngörmüş, A. H., and Alas, R. (2015). Influence of theory, seniority, and religiosity on the ethical awareness of accountants. *Social Responsibility Journal*, 11(3), 590-604.

- Vandekerckhove W. and Commers M.S.R. (2004). Whistleblowing and Rational Loyalty. *Journal of Business Ethics*, 53(1-2), 225-233.
- Verschoor, C. (2005). Is this the age of whistleblowers? *Strategic Finance*, 86(7), 17–18.
- Weinstein D. (1979). Bureaucratic opposition. New York: Pergamon Press.

Impact of Attention Driven Investments on Agricultural Commodity Prices

Tomas Misecka¹, Miroslava Rajcaniova¹, Jan Pokrivcak¹, Pavel Ciaian²

¹ Slovak Agricultural University in Nitra, Slovakia, Faculty of Economics and Management, Department of Economic Policy

² European Commission, Joint Research Centre, Spain

Abstract

This paper investigates the causality between attention driven investments and the final price of agricultural commodities. We focus on the speculative investments driven by news related to the future commodity price movements. This kind of attention driven investment can be estimated by the number of keywords searched by Google's search engine. We used time series analysis to estimate the impact of speculative investment on commodity prices. Our results show significant impact of attention driven investments on agricultural commodity prices.

Key words: attention driven investments, risk, agricultural commodities

JEL Classification: G23

Introduction

These days financial markets and commodity markets are linked much tighter than ever before. Disturbances in price volatility in one market affect prices in another market and vice versa. Agricultural commodity prices are moving due to factors like economic growth, substitutive investments opportunities, climate changes, market speculations and geopolitical uncertainties. From traders perspective main substitutes of a particular commodity are company stocks, government bonds, investments into currency pairs and other types of commodities.

Investors are trading at spot markets for current prices or at futures markets for future rates. All the wide range of investing possibilities are useless in the period of uncertainty, when the price volatility is high and the movements on the financial markets are difficult to predict. When traders expect that assets of their interests will increase they are entering the buying positions. When they are sure about market decline they can go short and make money on selling the assets or financial instruments. Problem with investing arise when it is difficult to predict an increase or decrease of prices (Fung and Hsieh, 2011).

Usually when value of stocks is decreasing companies do not pay out the dividends. Consequently traders are looking for alternative business opportunities. Trading the government bonds is quite conservative and stable investment, but on the other hand in stable countries interest rates are very, in some cases attacking negative values (Hale and Moore, 2016). Currency pairs are much more dynamic in comparison to government bonds, however much harder to predict.

In times of instability in financial markets, traders are trying to diversify their portfolio and relocate capital from financial assets to more conservative instruments like precious

metals, energy and agricultural commodities. During the financial crisis in 2008-2009, investments to precious metals dramatically increased. In 2007 total physical gold investment was 438 metric tons and silver investment was 1605 metric ton. The next year investments to gold more than doubled to 913 metric tons and investments to silver exploded to 5826 metric tons. In 2009 the prices of gold increased by 24% (Carlson, 2014).

Speculative investments to agricultural commodities are not that known, as investments to precious metals, nevertheless in recent years, they are very common. From 2005 till the mid of 2008 the prices of food commodities doubled in real terms. The main increase was observed in price of soybeans + 86%, wheat +101%, maize + 102%, rice + 110% and palm oil + 140% (Gilbert, 2008). Commodity investments are attractive for traders, repeatedly used as a safe haven in uncertain times. Investors make decisions based on information about future financial derivatives and commodity prices development.

Nowadays, information sources are far more extensive than they were in the past. Historically investors were taking actions based on information from TV, radio or newspapers (Mondria, WU, Zhang, 2009). It was virtually impossible to measure the impact measurement of an information and news on traders behaviour. Nowadays people search for most information on internet using search engines. This enables us to use the searched words in Google search engine to measure the attention driven investments.

The objective of our study is to find out the impact of news, information and attention driven trades on the price of agricultural commodities. More exactly, in this paper we would like to estimate the linkage between number of searched keywords attached to the specific agricultural commodity and its price level movements.

The paper is organized as follows. In section 2 we review the relevant theoretical and empirical literature. Section 3 presents data sources and used methodology. In section 4 we estimate the dependence of agricultural commodity prices from the attention to the news about agricultural commodities. Section 5 summarizes our findings and concludes.

Commodity price formation

Theoretical studies show many triggers, which shift traders from investing at the financial market to the agricultural commodity market. Investors are buying commodities due to their price stability to make their investment portfolio more robust and resistant from financial market volatility. Investments that do not aim to purchase commodities physically are speculative investments. Investors from financial market do not want to own the agricultural commodities, they just want to make profit from price movements. According to Trostle (2008) sharp rise of agricultural commodity prices during financial crisis in 2006 - 2008 had been due to increase in global demand for feedstocks, biofuels and adverse weather in the season 2006 and 2007 in some important grain and oilseed production areas. Additional factors that had put upward pressure on agricultural commodity prices were devaluation of U.S. Dollar, increasing energy prices, increasing costs of production protecting policies adopted by some exporting and importing countries. Gordon and Rouwenhorst (2006) found out that

commodity prices are determined by:

Economic growth: When global economy is expanding, the demand for agricultural commodities follows the growth rate, because consumption rises. When the economy is in recession prices of agricultural commodities do not fall proportionally, while investors are placing their capital into them. This feature makes agricultural commodities to be very stable investment asset.

Prices of substitutive investments: Demand for agricultural commodities is affected by values and riskiness of substitutive investments. Alternative option is investment into government bonds. According to Kat and Oomen (2006) and Gorton and Rouwenhorst (2006) commodities are essentially different from financial assets. As a result they are negatively correlated with stocks and bonds. Instead of longer-term economic expectations, commodity prices are determined by current economic activities. During recession, substitutive investments can be precious metals as well as during expansion investments into energetic commodities.

Weather: Depending on the type of commodity, when commodity price has grown worldwide and its harvest was weak just in some countries, than the global price is not affected dramatically. On the other hand when the yield of corn is poor for example in USA and Ukraine, which are the two of five largest producers than the price level can be affected.

Inflation rate: Higher inflation is having a negative impact on stocks and government bonds returns, but a positive impact on commodities. According to Kat and Oomen (2006), in times of strong economic growth and higher inflation, there will be positive pressure on commodity prices and interest rates. Higher commodity prices and higher interest rates lower the potential growth of companies and reduce the present value of future cash flows. Subsequently returns from stocks and bonds will drop, but commodities, in general will reinforce.

Monetary regime: From the perspective of access to global market we distinguish net commodity exporters, balanced economy and net importers. In an open economy, commodity prices are more less equal to global prices, but in the closed regime the prices of imported commodities can be much more expensive than in the rest of the world.

Market speculations based on financial market volatility: Predictions of price movements are made by analysis of historical prices and information and statements of the politicians, heads of enterprises or central banks (Riley, 2009).

Current market news: Historically it was much harder to obtain relevant, just in time information. Nowadays accesses to the information is much easier and cheaper. Investors use online servers collecting financial news, based on which it is much easier to make the trading decisions.

Investors' attention is an interesting topic for researchers. Investors make decisions influenced by the above mentioned factors. Investing into the agricultural commodities based on news about predictions of commodity price movements or based on information about estimated harvest are speculative investments. Prices of commodities

may be strongly influenced by expected future events. In many cases those factors are changing absolutely unpredictably. For instance grain crop in one country can be damaged by natural disaster and hence grain prices in other countries will increase, another example can be the forecast of dry growing season that raise the prices as well. Those are the reasons why traders are following current news.

Gilbret (2009) explains that companies like agricultural cooperatives, sugar refineries, grain elevator companies or farmers are typical retailers. They operate with small margin between sales and purchase prices, with the consequence that a small decrease of their products prices can eliminate profits on their inventories. They are selling futures contracts (short positions) to offset price exposure. On the other side speculators buy those contracts with the expectations of price increase that will yield them capital gain. According to O'Hara (1995), markets allow traders to trade based on their information. In theory of finance we distinguish informed and uninformed traders. Information about price movements may arise from research or knowledge of the market. In situation, when there are not many informed traders on the market, the informed ones have an opportunity to make profit on their information, but when most of the traders are well informed, price of information become inbounded in the market price. Many investors are not searching the investment opportunities systematically, rather they consider to purchase just the stocks that first catch their attention. This effect causes capital flowing to stocks and other financial derivatives that are more attention grabbing. Selling process is less affected by attention grabbing effect, while individual investors own in general just a small number of stocks or financial derivatives and they are selling just stocks they have (Barber and Odean 2008). In reality investors are heavily influenced by news. This attention-based trading leads investors to buy or sell speculatively and in many cases information have potential to shape the price levels of financial derivatives and commodity prices.

In many cases investors are influenced by information from the entire environment around them, mostly by their vicinity and working place. They tend to invest into the stocks of companies where they are employed or to companies in their surrounding area (Kazantsev 2013). Application of news analytics can be used for the impact evaluation of events that have been previously ignored. They can be used also to create new trading strategies and model behaviour of companies, government bonds, financial derivatives and commodity prices over time.

According to Lamount and Frazzini (2007) in comparison to attention driven trading of company stocks, commodity trading is regardless of size. Frequency of news about current condition of the specific company is dependent on the size or importance of the company. Investors have daily updated news about some companies and on the other hand there is just occasional information about small or less important companies like start-ups traders. For traders it is hard to grab attention in those cases, when the announcement is not reported. In general for merchants it is not important who produced the commodity, while it is always more less the same, the only important is the price. According to the Barber's hypothesis (Barber and Odean, 2011) individual investors have just limited news and when the investing opportunity grab their attention they are more likely to buy the stocks or commodities in comparison to commodities, which haven't been mentioned. By attention driven trading individual investors are more likely to buy the commodities than sell them, regardless the news were good or bad. Second finding is that individual investors are net buyers. Based on the information

and preferences they are more inclined to buy against sell.

According to Kagraoka (2016) we can estimate commodity prices based on the development of four main drivers of commodities prices. Using generalized dynamic factor model Kagraoka found out that US inflation rate, world industrial production, world stock index and the price of crude oil are the main dynamic macroeconomic indicators to determinate commodity prices. Empirical results from examined commodity prices between 1995 and 2015 showed that four dynamic factors explain 68,2% of the total variance in commodity returns.

Chen (2015) used VAR model to investigate the linkage between Chinese commodity sector co-movements and their underlying determinants like global oil prices shocks and domestic fluctuations. He observed strong effect of global oil prices on the common movements across commodity sectors in China at a long horizon. More specifically he illustrated the fact, that following a global oil price shock the common factors of the commodity sector initially increase sharply and then slowly converge to the equilibrium. Chen calculated that the common factor of commodity sectors is significantly and positively correlated with global oil price and with industrial production. The common factor increased by about 2% immediately after the global oil price shock. His results indicate that in the short run the common factor responds much more sensitively to the global oil price shocks than to the industrial production fluctuations. As stated by Lucotte (2016) agricultural commodity prices in the last decade (after commodity boom 2007 -2008) are much more correlated with oil price movement than before commodity boom. Lucotte used VAR model to estimate co-movements between six food price indexes namely cereal price index, dairy price index, meat price index, sugar price index, vegetable price index, food price index and crude oil price index. He concluded that the strong correlation between food price index and oil price index is driven by substitutive effect between biofuels and fossil fuel.

Fowowe (2016) pointed out another view to the interaction between agricultural commodity prices and oil prices. He used cointegration test to determine the long run relationship between prices of maize, sunflower, soybeans in South Africa and international oil price. In the second step he used Gregory and Hansen cointegration technique to endogenously determine the presence of structural breaks. In addition to this test he evidenced nonlinear behaviour between oil prices and agricultural commodity prices. The result indicates that agricultural commodity prices in South Africa are neutral to global oil prices in short and in long run as well. Wang and McPhail (2014) were investigating the impact of energy shocks on US agricultural productivity growth and commodity prices using VAR model. They used annual data between 1948 and 2011. The variables modelled were gasoline prices, agricultural total factor productivity function, real GDP, volume of agricultural export, and real agricultural commodity prices. They focused at the link between energy and agricultural commodity market caused by using corn to produce ethanol as a fuel. They found out that in the short run energy price shocks have negative impact on productivity growth and energy price shocks with agricultural productivity shocks impact US agricultural commodity prices volatility by 10% each. In the long run they find out that energy shocks contribute to about 15% of commodity price's variation.

Bodart et al. (2015) provided empirical evidence about the relationship between real exchange rates and primary commodity prices in developing countries. They estimated

the dependency between structural factors like the degree of trade openness, export diversification, financial openness and exchange rate regime. They used panel cointegration methodology to estimate the impact of structural factors on commodity prices. According to their results, exchange rate, the degree of financial openness and the degree of trade openness are statistically robust and significant determinants of commodity price elasticity in the long run. Huchet and Fam (2016) were investigating how agricultural commodity prices were affected by speculations in international futures market. They estimated causal relationship between future contracts to spot prices of agricultural commodities on a weekly basis between years 1998 and 2013. Using Granger causality test they found out granger causality between speculative investments in future markets and returns of wheat, corn, rice, soy bean, coffee, cocoa and sugar.

Ciaian, Rajcaniova and Kancs (2015) were analysing the influence of investors' attention for BitCoin using VECM. BitCoin is commodity traded just by the Internet so alteration of positive and negative news generate high price cycles. Attention driven investment behaviour can affect either increase or decrease in price, depending on the type of news that dominate in the media (positive/negative). Specifically they identified statistically significant positive correlation between number of Wikipedia views and BitCoin price. They assumed, that Wikipedia views may measure investors' interest in BitCoin and it may reflect improving knowledge about Crypto currencies, which may affect the demand of the BitCoin economy. On the other side Kristoufek (2015) was analysing correlation between online search queries in Google search engine and the Dow Jones Industrial Average components. He found out that there is no universal and global relation between relevant financial variables (traded volume and volatility) and online searches. Public interest is generated by information about stock market. According to Goddard, Kita and Wang (2015) not only retail and small investors are influenced by searched news, but also trading activities of large Forex market participants are influenced by the investors' attention. On the other hand they agree with Kristoufek's outcomes that individual investors are more likely to use Google to acquire relevant information, while large trading companies and dealers use trading platforms such as Bloomberg and Reuters. However small retail investors have contributed significantly to the growth of currency market and may account between 8 to 10 percentage of the total trading volume. Based on a VAR model, they estimate a lead-lag relationship between investors' attention and volatility of foreign exchange rates of major currency pairs that represents more than 69% of the total turnover in Forex markets. They discovered a positive and significant link between investors' attention and volatility. Specifically they claim that based on attention it is possible to forecast the future volatility of currency returns.

According to Bank, Larch and Peter (2011) increase in Google search volume of a company's name is associated with its temporarily higher future returns in the subsequent month. They concluded that the number of searched queries related to the exact company names particularly measures the interest of uninformed traders. Information from Google search leads to the reduction of information asymmetry, improved liquidity and short term buying pressure. Their research was done by panel regression approach between years 2004 – 2010 on the Xetra trading system.

Chan (2003) investigated the impact of public news on the company stock price movements. He examined the differences between monthly returns on investments

from the stocks of the companies that were mentioned in the public news and between company which weren't headlined. He collected information about all stocks in a given month that had at least one headline story and ranked them by monthly raw returns. Then he took the best third of companies – winners and worst third of the companies – losers. He observed a significant difference between those two sets of firms. Companies stocks with negative headlines in public news had a strong drift, stocks that experienced positive news show less drift.

Welagedara, Deb and Singh (2016) used the multivariate regression model to investigate the impact of analysts' recommendations and predictions on the stock price movements. Based on the Google search volume index they measured attention of individual investors. They used a low institutional ownership dummy like a proxy to distinguish between attention of retail (individual) and institutional investors. They observed that positive news caused greater impact on the institutional investors than on individual investors and on the other hand after recommendation downgrade retail investors overreact and show greater price reversal compared to institutional investors.

Methodology and Data

As the first step we need to determine the degree of integration of our time series. Five different unit root tests, Augmented Dickey-Fuller test (ADF), Dickey-Fuller GLS (DF-GLS) test, Phillips-Perron (PP) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and Elliott-Lothman-Stock (ERS) test, were applied, to find out whether the levels of variables are stationary. The number of lags was determined by Akaike Information Criterion.

To estimate the impact of speculative investment on commodity price we use Vector Autoregressive model (VAR). VAR methodology looks like simultaneous-equation modeling in that we consider several endogenous variables together. Each endogenous variable is explained by its lagged, or past, values and the lagged values of all other endogenous variables in the model; usually, there are no exogenous variables in the model (Gujarati, 2004). The VAR model developed by Sims is based on the idea, that if there is true simultaneity among a set of variables, they should all be treated on an equal footing; there should not be any a priori distinction between endogenous and exogenous variables (Sims, 1980).

First we tested the Eigenvalue stability condition, our VAR model satisfies the stability condition as all the eigenvalues lie inside the unit circle.

We used the first difference of each nonstationary variable to construct the VAR model, that captures the inter-relationships among variables. Our estimated VAR model covers the following variables: prices of substitutive commodities e.g. wheat, soybean, prices of precious metals – in our model represented by gold, prices of crude oil, climate changes, inflation rate, monetary regime and number of searches for the commodity under interest.

For a set of n time series variables $y = (y_1, y_2, \dots, y_n)$ the VAR model can be written as:

$$y_t = A_0 + \sum_{l=1}^p A_l y_{t-l} + \varepsilon_t$$

Where p is the lag length, y_t is a vector of endogenous variables, A_0 is a vector of constants, A_l 's are $(n \times n)$ coefficient matrices and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{nt})'$ is an unobservable i.i.d. zero mean error terms, variance-covariance matrix $\Sigma = \{\sigma_{ij}, i, j = 1, 2, \dots, n\}$.

The main uses of the VAR model are the impulse response analysis, variance decomposition, and Granger causality tests.

Granger causality tests highlight the presence of at least unidirectional causality linkages as an indication of some degree of integration. This implies that each market uses information from the other when forming its own price expectations, while unidirectional causality inform about leader-follower relationships in terms of price adjustments (Arshaad, Hameed 2009). Granger causality provides important information about the exogeneity, in other words x_t is defined as an exogenous variable if the current and past values of Y_t do not affect X_t . The Granger causality method is based on the hypothesis that compared series are stationary or $I(0)$. In the absence of cointegration vector, with $I(1)$ series, valid results in Granger causality testing are obtained by simply first differentiating $I(1)$ variables in VAR model.

Since the individual coefficients in the estimated VAR models are often difficult to interpret, the practitioners of this technique often estimate the so-called impulse response function (IRF). The IRF traces out the response of the dependent variable in the VAR system to shocks in the error terms (Gujarati, 2004). We performed the Impulse Response Functions to analyse how a shock in one variable would persist in future periods. The forecast was made considering a ten-week period.

In this paper we consider the news from Internet as a main source of information based on which, investors make their decisions where to place their capital. We have chosen Google search engine, as the most widely used search engine (Sterling, 2015) and as a source of the number of searches we used its Trends database. Google Trends is a Web statistical instrument of Google Inc. Google Trends show how often is a specific expression searched relatively to the total search volume. Our investigation is done based on the global searches and from the filtering options we have chosen attention for news. More specifically we were looking for the name of the commodity (corn, wheat, soy bean, gold and crude oil) in four languages (English, German, French and Spanish) during the period from 2009 to 2015 and come from financial portal Investing.com⁷. Outcome of the search query time series from the Google Trends are normalized and rescaled from 0 to 100 interval that represent the proportion of searched term among the searching period. For the analysis of agricultural commodity price formation, we used weekly data of commodity prices and number of searches for those commodities.

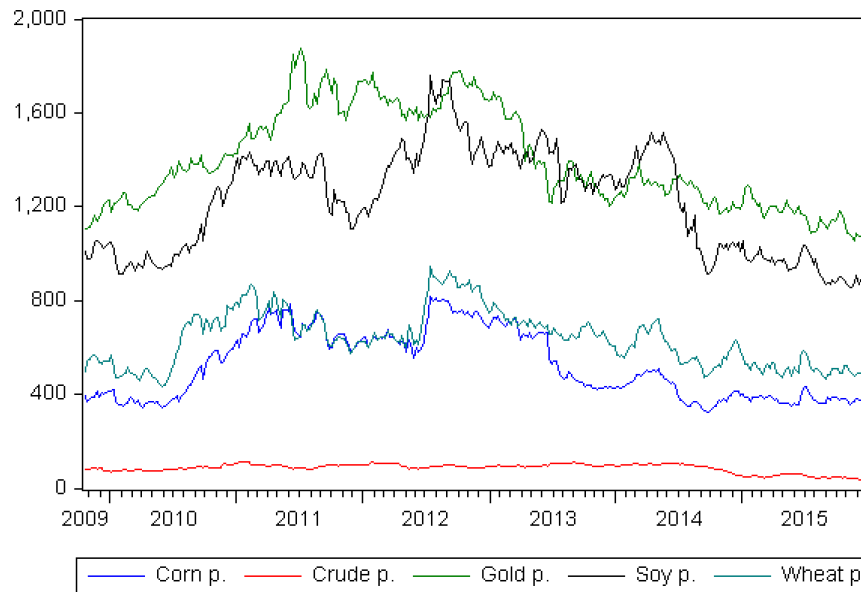
⁷Investing.com provides news, analysis, and streaming quotes about the global financial markets. Advantage of this trading platform is the free access to the historical quotes that are recorded at a weekly frequency.

Table 1: Descriptive Statistics

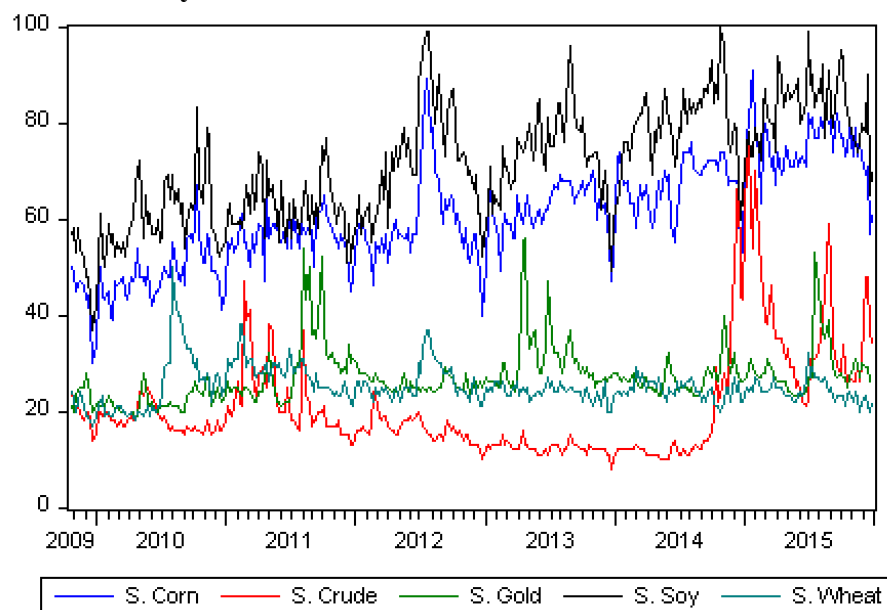
	CORN_P_	WHEAT_P	SOY_P_	GOLD_P_	CRUDE_P_	S_CORN	S_WHEAT	S_SOY	S_GOLD	S_CRUDE
Mean	523.1394	636.4926	1222.575	1399.074	85.21968	60.83951	25.14815	71.20062	27.09259	20.70370
Median	478.8800	633.6850	1269.250	1338.250	91.39000	60.00000	25.00000	72.00000	26.00000	17.00000
Maximum	818.7500	943.8800	1758.380	1873.700	113.9300	91.00000	50.00000	100.0000	56.00000	76.00000
Minimum	323.1200	435.3800	855.6300	1056.200	34.73000	30.00000	17.00000	37.00000	18.00000	8.000000
Std. Dev.	145.6416	118.0100	223.3700	210.8469	19.00977	10.69586	4.132065	12.04275	5.717071	11.40522
Skewness	0.344519	0.410369	0.040302	0.384857	-1.043760	0.129236	1.714166	0.024975	2.299733	2.245406
Kurtosis	1.627023	2.396063	1.892078	1.859886	3.072748	2.627486	8.799021	2.343098	10.56998	8.908119

Source: Own construction based on the data from Investing.com, Google Trends

Figure 1 depicts the development of prices, Figure 2 shows the number of searches.

Figure 1: The Development of Commodity Prices

Source: Own construction based on the data from Investing.com

Figure 2: Commodity Price Search

Source: Own construction based on the data from Google Trends

Volume of speculative investment and value of information affecting the commodity prices are not exact, as in many cases it is difficult to distinguish between investment and speculative trading. Moreover it is difficult to exactly specify based on which information traders and investors made their investment decision.

Results

Time series analysis was used to estimate the link between price development of selected commodities and attention driven investments. As the first step we had to check the stationarity of all variables. We used Augmented Dickey-Fuller test, Dicky-Fuller GLS (DF-GLS) test, Phillip-Perron (PP) test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and Elliot-Rothenberg-Stock Point-Optimal (ERS) test. The results of the tests are summarized in Table 2. Detailed results of the tests are shown in Appendix 1-5.

Table 2: Stationarity Tests Results

Variable	ADF		DF-GLS		PP		KPSS		ERS		Summary
	level	1 st dif.	level	1 st dif.	level	1 st dif.	level	1 st dif.	level	1 st dif.	
Corn price	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	I(1)
Wheat price	Stat	Stat	N-stat	Stat	Stat	Stat	N-stat	Stat	N-stat	Stat	I(1)
Soy price	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	I(1)
Gold price	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	I(1)
Crude oil price	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	N-stat	Stat	I(1)
Search for Corn	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	I(0)
Search for Wheat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	I(0)
Search for Soy	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	I(0)
Search for Gold	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	Stat	I(0)
Search for Crude oil	Stat	Stat	Stat	N-stat	Stat	Stat	N-stat	Stat	Stat	Stat	I(0)

Source: Own construction. Note: Stat means stationary, N-stat means non-stationary time series

Having in mind that some of the variables are stationary in first differences and some in levels, we construct a Vector Autoregressive model in first differences of I(1) variables and levels of I(0) variables. Using VAR we applied Granger causality test to variables modelled to find out, whether there exists at least unidirectional causality linkage between commodity prices and search for information about those commodities. As seen from the results (Appendix 6), there is a unidirectional relationship between wheat price and price of corn and soybean. Our results revealed that the change in the price of wheat does Granger Cause the change in the price of corn and the change in soy bean price. On the other side the changes in wheat price are Granger caused by the shocks in crude oil price. We did not observe any Granger causality running from the number of searched words and commodity prices.

Impulse Response Functions were performed in order to show how a shock in one variable would persist in future periods. The forecast was made considering a ten-week period. As we can see from Appendix 10-12, a shock in the corn price, wheat price and soybean price would result in a mild and temporary response in corn prices. However, there is only minor reaction of corn price to search of wheat, corn or soybean. As seen from Appendix 11, the prices of wheat are more responsive, compared to corn prices.

Wheat prices are influenced by the shocks in corn prices, crude oil prices, their own past values and also by soybean prices and gold prices. Wheat prices react also to shocks in searching for gold, wheat and corn. Soybean prices are influenced the most by the shocks in crude oil prices, gold price and wheat prices. There is also an impact of search for gold, search for corn and minor impact of search for soybean and wheat to prices of soybean.

Conclusions

The main purpose of this paper is to analyze the statistical relationship between agricultural commodity prices, namely wheat, corn and soybean prices and attention driven by news and information. This kind of attention driven investment can be estimated by the number of keywords searched by Google's search engine. We used time series analysis to estimate the impact of speculative investment on commodity prices. Our analysis revealed that search for soybean and wheat does have impact on soybean prices. Wheat prices react to shocks in searching for gold, wheat and corn. However, there is only minor reaction of corn price to search of wheat, corn or soybean.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0552 and VEGA 1/0797/16.

References

- Ankrim, E.M., Hensel, Ch.R. 1993. Commodities in Asset Allocation: A Real-Asset Alternative to Real Estate?. In: Financial Analysts Journal, vol. 49, 20-29.1993.
- Arshad, F.M., Hameed, A.A.A. 2009. The Long Run Relationship Between Petroleum and Cereals Prices. Global Economy & Finance Journal Vol.2 No.2 March 2009 Pp. 91-100
- Auerbach, A.J., Gorodnichenko, Y. 2013. Fiscal Multipliers in Recession and Expansion. Published by the National Bureau of Economic Research. Chicago Press. 2013 [accessed. January 2016] viewed at: <http://www.nber.org/chapters/c12634.pdf>.
- Ciaian, P., Rajcaniova, M., Kancs, A. 2015. The economics of BitCoin price formation. In: Applied Economics, vol. 48, 1799 – 1815. 2015.
- Bank, M., Larch, M., Peter, G. 2011. Google Search Volume and its Influence on Liquidity and Returns of German Stocks. In: Financial Markets and Portfolio Management , Vol. 25, No. 3, 239 – 264. 2011.
- Barber, B.M., Odean, T. 2006. All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. [online] Berkley: Berkley, CA 94720(510)642-6767, 2008. [accessed. December 2016] viewed at: <http://faculty.haas.berkeley.edu/odean/papers/attention/all%20that%20glitters.pdf>
- Barber, B.M., Odean, T. 2011. The Behaviour of Individual Investors. [online] Berkley: Berkley, CA 94720(510)642-6767, 2008. [accessed. December 2016] viewed at: <http://www.umass.edu/preferen/You%20Must%20Read%20This/Barber-Odean%202011.pdf>.
- Boart, V., Candelon, B., Carpentier, J. 2015. Real exchanges rates, commodity prices

- and structural factors in developing countries In: *Journal of International Money and Finance*. 2015. Vol. 51, 264 – 284. 2015.
- Büyüks, B., Robe, M.A. 2014. Speculators, commodities and cross-market linkages .In: *Journal of International Money and Finance*. Vol.42, 38–70, 2014.
- Carlson, D. 2014. 2008 Financial Crisis Set Stage For Gold Rally. In: *Kitco News*. [online] Montreal: Kitco Metals Inc. [accessed. December 2015] viewed at: <http://www.kitco.com/news/2014-10-13/2008-Financial-Crisis-Set-Stage-For-Gold-Rally.html>
- Erb, C., Campbell, H. 2006. The Tactical and Strategic Value of Commodity Futures. In: *Financial Analysts Journal*, March/April, 69–97. 2006.
- Fowowe, B. 2016. Do oil prices drive agricultural commodity prices? Evidence from South Africa In: *Energy*. 2016. vol. 104, 149 – 157. 2016
- Fung, W., Hsieh, D.A. 2011. The risk in hedge fund strategies: Theory and evidence from long/short equity hedge funds. [online] London: Hedge Fund Research Centre of the London Business School, United Kingdom, 2013. [accessed. January 2016] viewed at: <http://www.sciencedirect.com/science/article/pii/S0927539811000211>.
- Gilbert, Christopher L. 2008. How to Understand High Food Prices. [online] Trento: Department of Economics, University of Trento, 2008. [accessed. September 2015] viewed at: <http://core.ac.uk/download/files/153/6262849.pdf>.
- Godard, J. - Kita, A. - Wang, Q. 2015. Investor attention and FX market volatility. In: *Journal of International Financial Markets, Institutions & Money*. 2015. vol. 38, 79 – 96. 2015.
- Gordon, G., Rouwenhorst, G.K. 2006. Facts and Fantasies about Commodity Futures. In: *Financial Analysts Journal*. 2006. Volume 62. March /April 2006.
- Gujarati, D.N., 2004: *Basic Econometrics*, Fourth Edition, The McGraw–Hill Companies, 2004
- Hale, T., Moore, E. 2016. Germany reaches negative rate milestone . [online] London: The Financial Times Limited. 2016. [accessed. July 2016] viewed at: <http://www.ft.com/cms/s/0/3ef5136c-2bf7-11e6-bf8d-26294ad519fc.html#axzz4DQZ02ZHF>
- Huckett, N., Fam, P.G. 2016. The role of speculation in international futures markets on commodity prices In *Research in International Business and Finance*. 2016. Vol. 37, 49 – 65. 2016.
- Chan, S.W. 2003. Stock price reaction to news and no-news: drift and reversal after headlines. In: *Journal of Financial Economics*. 2003. vol70, 223–260, 2003.
- Chen, P. 2015. Global oil prices, macroeconomic fundamentals and China's commodity sector comovements. In: *Energy Policy* 2015. vol. 87, 284–294, 2015.
- Kagraoka, Y. 2016. Common dynamic factors in driving commodity prices: Implications of a generalized dynamic factor model . In: *Economic Modelling*. 2016. vol. 52, 609 – 617. 2016.
- Kat, H.M., Oomen, R.C.A. 2006. What Every Investor Should Know About Commodities Part II: Multivariate Return Analysis. [online] London: Alternative Investment Research Centre Cass Business School, City University 106 Bunhill Row. 2016. [accessed. February 2016] viewed at: <http://poseidon01.ssrn.com/delivery.php?ID=801118100100122086092117121090016118097047072050071009098005066067072122011025030106118049038127005002035077116070031078107098070035043082064127064081064074072032077000113103020068089115012113125116001004013071066024123077008107105070070000124066&EXT=pdf>

- Kazantsev, G. 2013. News Analytics in Finance. [online] New York: Bloomberg, 2013. [accessed. December 2015] viewed at: <http://www.newsanalytics.net>.
- Knittel, Ch.R., Pindick, R.S. 2013. The Simple Economics of Commodity Price Speculation. [online] Cambridge: A Joint Centre of the Department of Economics, MIT Energy Initiative and MIT Sloan School of Management. 2013. [accessed. February 2016] viewed at: <http://web.mit.edu/ceepr/www/publications/workingpapers/2013-006.pdf>
- Kristoufek, L. 2015. Power-law correlations in finance-related Google searches, and their cross-correlations with volatility and traded volume: Evidence from the Dow Jones Industrial components In: *Physica* 2015. Vol 428, 194 – 205. 2015.
- Krugman, P. 2008. More on Oil and Speculation. In: *New York Times*, May 13. 2008. [online] New York: [krugma.blogs](http://krugman.blogs.nytimes.com/2008/05/13/more-on-oil-and-speculation/?_r=0). 2008. [accessed. November 2015] viewed at: http://krugman.blogs.nytimes.com/2008/05/13/more-on-oil-and-speculation/?_r=0.
- Lamount, O., Frazzini, A. 2007. The earning announcement and trading volume. [online] Cambridge: National Bureau of Economic Research, 2007. [accessed. March 2015] viewed at: <http://poseidon01.ssrn.com/delivery.php?ID=965089116025102064021023113110009070037032032002040049023055006036055029013000010002121026098010029103094006077082039101118114096124001008067107125094023019093116069082009004089079072104068021068&EXT=pdf>
- Lucotte, Y. 2016. Co-movements between crude oil and food prices: A post-commodity boom perspective. In: *Economics Letters*. 2016. Vol. 147, 142 – 147. 2016.
- Mondria, J., Wu, T., Zhang, Y. 2009. The determinants of international investment and attention allocation: Using internet search query data. In: *Journal of international Economics*. [online] Toronto: Economics Department, University of Toronto [accessed. December 2015] viewed at: <http://www.sciencedirect.com/science/article/pii/S0022199610000449>.
- O'Hara, M. 1995. *Market Microstructure Theory*. 6. pub. Malden: Library of Congress Cataloguing-in-Publication Data, 2004. ISBN: 978- 0631207610.
- Riley, J. 2009. What are commodities and how are their prices determined? In: *Tutor2u*, May 2009 [online] Boston Spa 2009, [accessed. Januar 2016] viewed at: <http://www.tutor2u.net/business/blog/qa-what-are-commodities-and-how-are-their-prices-determined>
- Spurgin, R. 2001. A Benchmark for Commodity Trading Advisors. In: *Journal of Alternative Investments*. 1999. vol. Summer 1999.
- Sterlin, G. 2015. Google Controls 65 Percent Of Search, Bing 33 Percent [comScore]. [online] Search Engine Land. [accessed. December 2015] viewed at: <http://searchengineland.com/google-controls-65-percent-of-search-bing-33-percent-comscore-228765>.
- The Economist. 2014. Fixing the fix In: *The Economist*. [online] Feb 8th 2014, Brussels. [accessed April 2016] viewed at: <http://www.economist.com/news/finance-and-economics/21595943-european-union-wants-change-how-commodity-benchmarks-are-set-fixing-fix>
- Tomek, W.G., Robinson, K.L. 1990. *Agricultural Product Prices*. 3rd edition. New York: Cornell University Press, 124 Roberts Place. 1990. 360 p. ISBN 0-8014-2451-8.
- Trostle, R. 2008. Global Agricultural Supply and Demand: Factors Contributing to the Recent Increase in Food Commodity Prices. [online] Washington, D.C: U.S. Department of Agriculture, 1400 Independence Ave. 2008 [accessed. April

- 2016] viewed at:
http://www.growthforce.orgwww.growthenergy.org/images/reports/USDA_Global_Agricultural_Supply_and_Demand.pdf
- U.S. Commodity Futures Trading Commission (2006). A Guide to the Language of the Futures Industry. 2006. [online] The Commodities Futures Trading Commission. 2006. [accessed September 2015] viewed at: <http://www.cftc.gov>
- Wang, S.L., McPhail, L. 2014. Impacts of energy shocks on US agricultural productivity growth and commodity prices—A structural VAR analysis. In: Energy Economics. 2014. vol 46, 435–444. 2014.
- Welagedara, V., Deb, S., Singh, H. 2016. Investor attention, analyst recommendation revisions, and stock prices. In: Pacific-Basin Finance Journal. [online] Deakin Business School, Department of Finance, Deakin University, Geelong, Australia. 2016 [accessed. October 2016] viewed at: http://ac.els-cdn.com/S0927538X1630066X/1-s2.0-S0927538X1630066X-main.pdf?_tid=4624a98e-9bb1-11e6-bc52-00000aacb360&acdnat=1477509817_d43578869e68de0cde151111d0dc1ff8

Appendix 1: Augmented Dickey-Fuller test results

Levels	Constant	Constant & Trend
Corn price	-0.011	-0.016
Wheat price	-0.026**	-0.032**
Soy price	-0.013	-0.015
Gold price	-0.011	-0.020**
Crude price	-0.002	-0.010
Search for Corn	-0.110***	-0.307***
Search for Wheat	-0.139***	-0.144***
Search for Soy	-0.087***	-0.247***
Search for Gold	-0.211***	-0.232***
Search for Crude	-0.056**	-0.066***
1st differences		
Corn price	-0.955***	-0.960***
Soy price	-0.943***	-0.949***
Gold price	-0.938***	-0.948***
Crude price	-0.954***	-0.963***

Source: Own calculation

Note: The null hypothesis of this test is that the time series has a unit root, we reject the H_0 , when the p-value is less than 0.05 marked with ** or p-value is less than 0.01 marked with ***

Appendix 2: DF-GLS results

Levels	Constant	Constant & trend
Corn price	-0.007	-0.008
Wheat price	-0.013	-0.009
Soy price	-0.009	-1.080
Gold price	-0.005	0.005
Crude price	-0.004	-0.007
Search for Corn	-0.061**	-0.270**
Search for Wheat	-0.130**	-0.134**
Search for Soy	-0.036**	-0.236**
Search for Gold	-0.114**	-0.218**
Search for Crude	-0.053**	-0.055**
1st differences		
Corn price	-0.076	-0.226**
Wheat price	-0.062	-0.245**
Soy price	-0.190**	-0.797**
Gold price	-0.930**	-0.933**
Crude price	-0.930**	-0.941**

Source: Own calculation

Note: The null hypothesis of this test is that the time series has a unit root. Time series with asterisks are stationary at the 5% level by t-Statistics.

Appendix 3: Philips – Perron test results

Levels	Constant	Constant & Trend
Corn price	-0.011	-0.032
Wheat price	-0.026**	-0.032**
Soy price	-0.013	-0.015
Gold price	-0.011	-0.021
Crude price	-0.002	-0.010
Search for Corn	-0.110**	-0.307**
Search for Wheat	-0.165**	-0.169**
Search for Soy	-0.151**	-0.309**
Search for Gold	-0.211**	-0.232**
Search for Crude	-0.084**	-1.155**
1 st differences		
Corn price	-0.955**	-0.960**
Wheat price	-0.942**	-0.945**
Soy price	-0.943**	-0.949**
Gold price	-0.938**	-0.948**
Crude price	-0.954**	-0.963**
Search for Corn	-1.134**	-1.134**
Search for Wheat	1.223**	-1.223**
Search for Soy	-1.318**	-1.318**
Search for Gold	-1.068**	-1.068**
Search for Crude	-1.027**	-1.027**

Source: Own calculation

Note: The null hypothesis of this test is that the time series has a unit root. Time series with asterisks are stationary at the 5% level by t-Statistics.

Appendix 4: Kwiatkowski-Phillips-Schmidt-Shin test results

Levels	Constant (LM-stat)	Constant & trend (LM-stat)
Corn price	0.649	0.398
Wheat price	0.506	0.344
Soy price	0.471	0.433
Gold price	0.784	0.430
Crude price	0.678	0.371
Search for Corn	1.868	0.068**
Search for Wheat	0.241**	0.106**
Search for Soy	1.625**	0.048**
Search for Gold	0.524	0.134**
Search for Crude	0.468	0.274
1st differences		
Corn price	0.259**	0.073**
Wheat price	0.159**	0.033**
Soy price	0.250**	0.035
Gold price	0.401**	0.060**
Crude price	0.318**	0.044**
Search for Corn	0.323**	0.287
Search for Wheat	0.018**	0.014**
Search for Soy	0.057**	0.037**
Search for Gold	0.037**	0.026**
Search for Crude	0.050**	0.025**

Source: Own calculation

Note: The null hypothesis of this test is that the time series are stationary. Time series with asterisks are stationary at the 5% level by LM-Statistics.

Appendix 5: Elliott–Rothenberg-Stock test results

Levels	Constant	Constant & trend
Corn price	11.073	35.562
Wheat price	6.522	18.322
Soy price	9.217	27.910
Gold price	17.543	49.229
Crude price	11.172	31.139
Search for Corn	1.481**	1.439**
Search for Wheat	0.745**	2.696**
Search for Soy	3.980	1.910**
Search for Gold	0.847**	1.516**
Search for Crude	2.078**	7.210
1 st differences		
Corn price	0.419**	0.870**
Wheat price	0.571**	0.905**
Soy price	0.252**	0.701**
Gold price	0.157**	0.581**
Crude price	0.157**	0.582**
Search for Corn	0.094**	0.314**
Search for Wheat	0.158**	0.466**
Search for Soy	0.020**	0.073**
Search for Gold	0.030**	0.105**
Search for Crude	0.087**	0.299**

Source: Own calculation

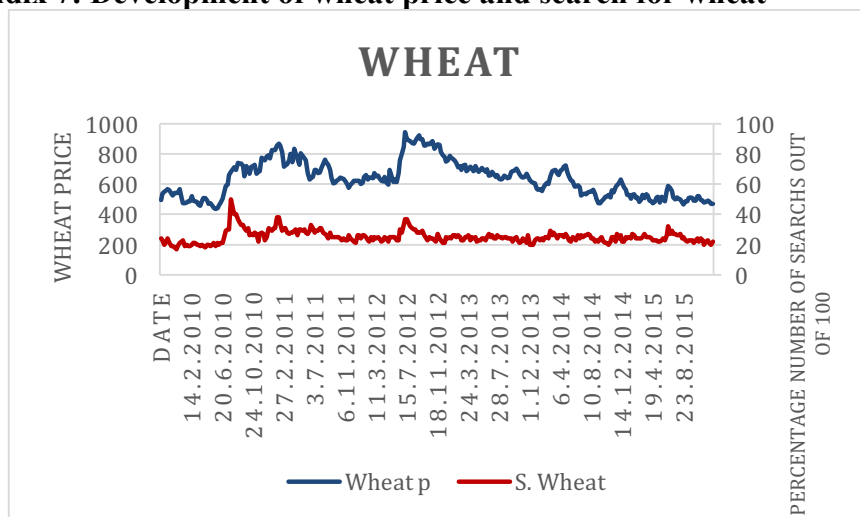
Note: The null hypothesis of this test is that the time series has a unit root. Time series with the asterisk are stationary at the 5% level by P-Statistics.

Appendix 6: Pairwise Granger Causality Tests

Null Hypothesis:	F-Statistic	Prob.
Wheat price does not Granger Cause Corn price	3.09319	0.0272
Soy price does not Granger Cause Corn price	1.40400	0.2416
Gold price does not Granger Cause Corn price	0.81767	0.4849
Crude oil price does not Granger Cause Corn price	0.57099	0.6345
Wheat search does not Granger Cause Corn price	1.02899	0.3799
Corn search does not Granger Cause Corn price	0.24796	0.8628
Soy search does not Granger Cause Corn price	0.66178	0.5761
Gold search does not Granger Cause Corn price	1.66917	0.1736
Crude oil search does not Granger Cause Corn price	0.34216	0.7949
Corn price does not Granger Cause Wheat price	1.03782	0.3760
Soy price does not Granger Cause Wheat price	0.48284	0.6944
Gold price does not Granger Cause Wheat price	0.35585	0.7849
Crude oil price does not Granger Cause Wheat price	2.66772	0.0478
Corn search does not Granger Cause Wheat price	0.52599	0.6647
Wheat search does not Granger Cause Wheat price	0.81235	0.4878
Soy search does not Granger Cause Wheat price	0.74660	0.5250
Gold search does not Granger Cause Wheat price	0.91507	0.4339
Crude oil search does not Granger Cause Wheat price	0.50806	0.6770
Corn price does not Granger Cause Soy price	1.36556	0.2533
Wheat price does not Granger Cause Soy price	2.26313	0.0811
Gold price does not Granger Cause Soy price	0.57447	0.6322
Crude oil price does not Granger Cause Soy price	1.49290	0.2164
Corn search does not Granger Cause Soy price	0.82511	0.4808
Wheat search does not Granger Cause Soy price	0.58307	0.6265
Soy search does not Granger Cause Soy price	0.34305	0.7942
Gold search does not Granger Cause Soy price	1.96118	0.1198
Crude oil does not Granger Cause Soy price	0.18782	0.9047

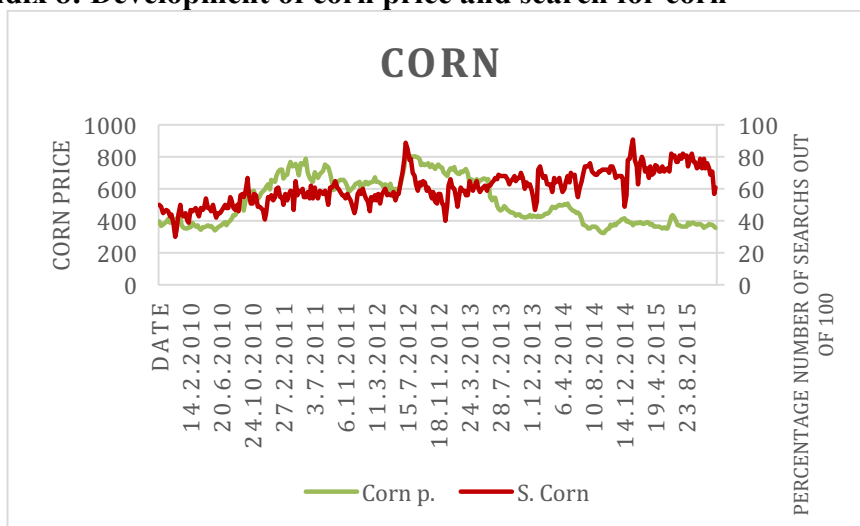
Source: Own calculation

Appendix 7: Development of wheat price and search for wheat



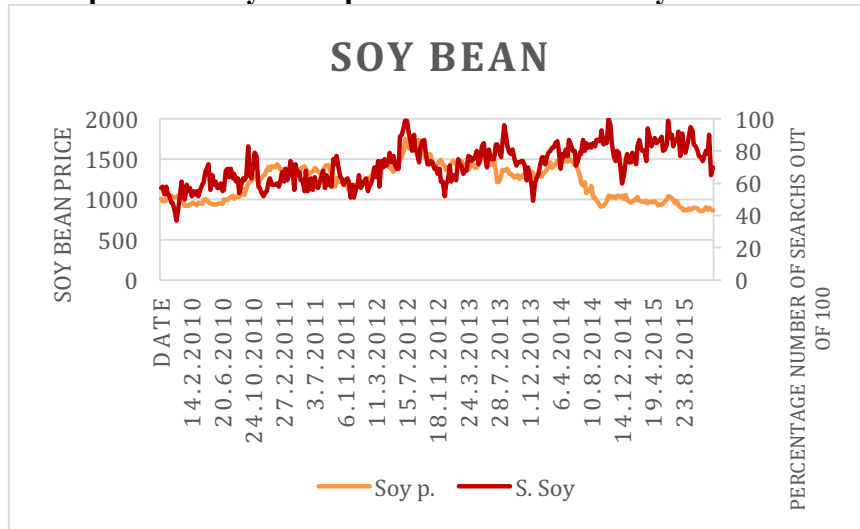
Source: Own elaboration

Appendix 8: Development of corn price and search for corn



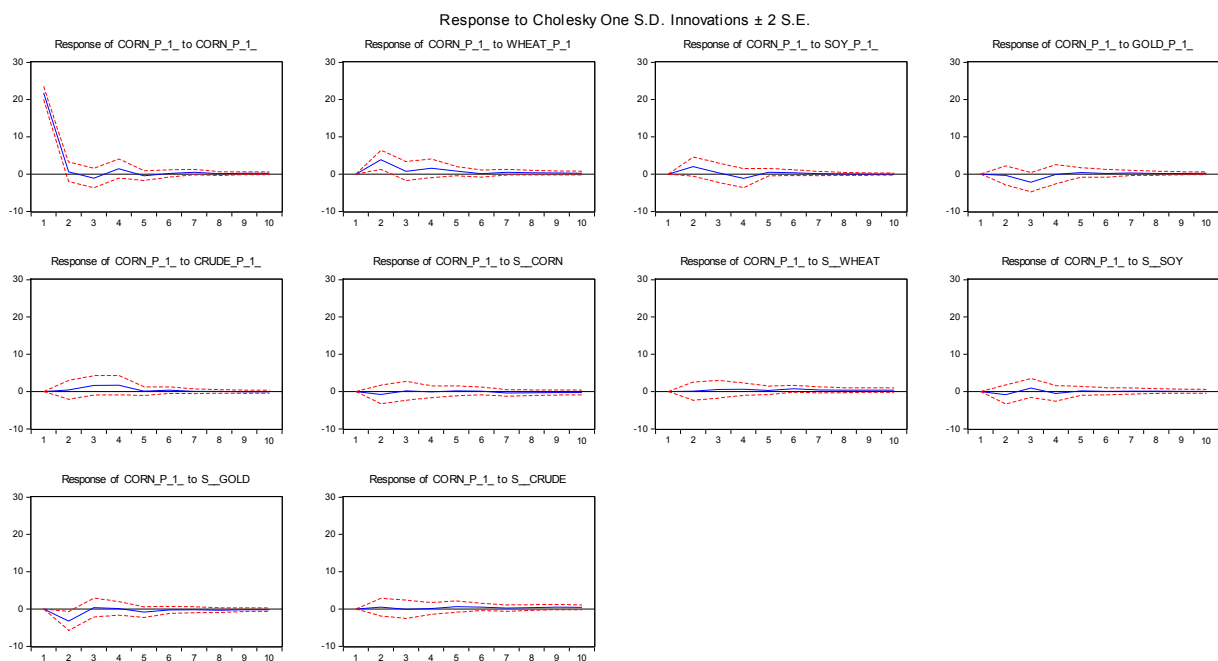
Source: Own elaboration

Appendix 9: Development of soy bean price and search for soy bean



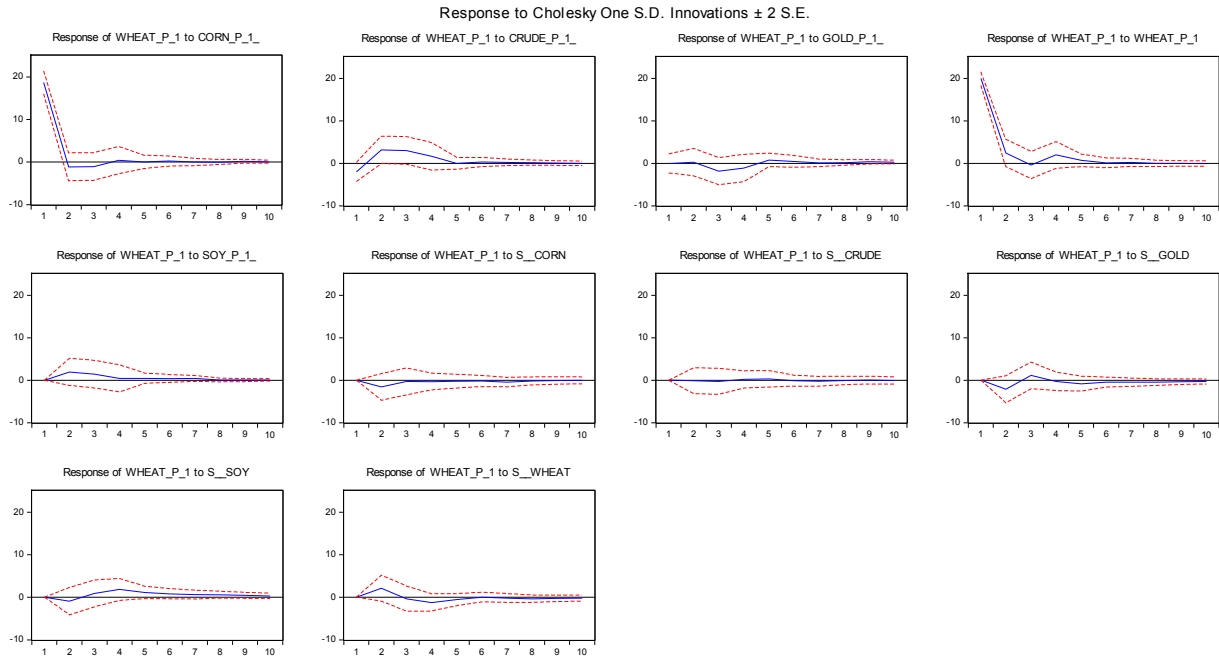
Source: Own elaboration

Appendix 10: Impulse response to differences in CORN prices



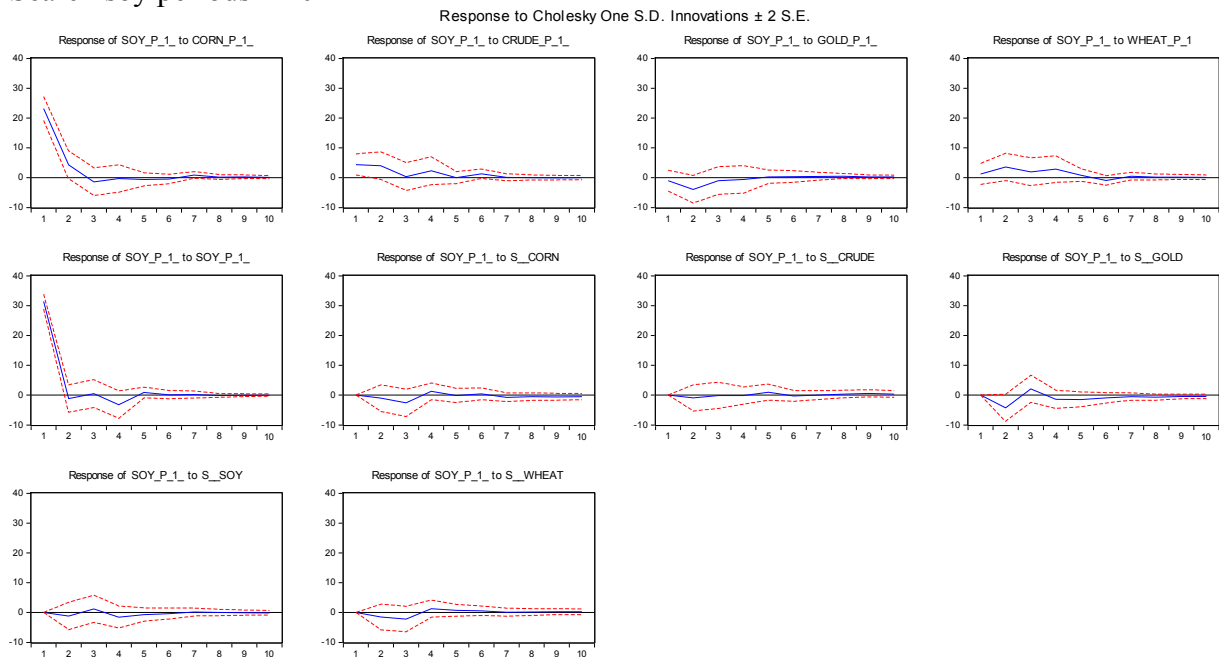
Source: Own elaboration

Appendix 11: Impulse response to differences in WHEAT prices



Appendix 12: Impulse response to differences in SOY prices

Search soy periods – 10



Inflation Dynamics in Albania: A Markov Regime-Switching Approach

Anisa Plepi

University of Tirana, Faculty of Economics ,anisaplepi@feut.edu.al

Abstract

Over the years, inflation has exhibited distinct dynamic patterns, ranging from near hyperinflation levels throughout the process of transition towards a market-based economy, disinflation following the post-crisis of 1997 to moderate values after 2000. In this context, when modeling inflation it is crucial to capture the alteration in inflation behavior that is due to the regime switch. In this paper, inflation is characterized by a Markov Switching model in which the Albanian economy can potentially switch between regimes of low and high inflation and vice versa, which enables us to seize the distinct dynamic behavior inflation exhibits when in different regimes. The empirical results reveal that the Albanian economy has been in the regime of low/stable inflation for the last 19 years with the probability of departing from price stability only 0.02. Furthermore, the estimates suggest that the economy stays longer (twice as long) in the regime of price stability than in that of high inflation.

Keywords: *Inflation dynamics, Markov switching model, transition probabilities, price stability*

JEL codes: *C24, C53, E31*

Introduction

Over the years, inflation has undergone various phases following the transformation of the Albanian economy, as reflected by significant changes in the policy-making environment and institutional structures. The transition from a centrally planned to a market-oriented economy was characterized by near hyperinflation levels that appeared to slow down by 1995 through a prudent monetary management. Nevertheless, as stated by Kalra (1998a), due to a combination of demand pressures following the lax fiscal stance ahead of the 1996 parliamentary elections and a sharp depreciation of the exchange rate as a result of the crisis of 1997 which culminated with the civil disorder that followed the collapse of the pyramid schemes, by 1997 strong inflationary pressures reappeared. The post-crisis period on the other hand, was characterized by substantial disinflation only to be drawn close to the 3 ± 1 percent target after 2000 Kota (2011a). In this context, when addressing inflation, a model that allows for distinct dynamic behavior during different periods where each is characterized by its own time series properties, would prove to be useful.

The existing literature concerning inflation in Albania focuses mainly on determining the driving forces behind inflation by relying on a single model for the conditional mean to represent the patterns inflation exhibits over time, thus it fails to take into consideration the potential switches between regimes of low and high inflation and the alteration in inflation behavior that is due to the regime switch. More specifically, Kalra (1998b), Domac & Elbirt (1998) are among the first attempts to identify the determinants of inflation in Albania during the period 1993-1997 through a linear Error Correction Model (ECM). A similar methodology was employed by Rother (2000) in order to investigate the short and long run dynamics of inflation and relative price adjustments during the transition period 1993-2000. Furthermore, Kunst & Luniku (1998) attempt to identify the potential causes of inflation focusing particularly on monetary and fiscal influences for a time period from 1993-1997 through a simple linear regression model that allows for a single structural

break. Kota (2011b) on the other hand, focuses on estimating persistence for headline and core inflation in Albania while checking for the presence of structural breaks for a time period from 1993-2008 through a univariate approach where inflation is modeled as a simple autoregressive (AR) process.

This paper provides the first attempt to fill the existing gap in literature concerning inflation in Albania by relying on a Markov Switching framework in modeling inflation, as proposed by Hamilton (2010). This methodology is different from the standard linear approach widely employed in literature. In addition, instead of treating inflation as a unit root process, inflation is characterized by a regime switching model in which the economy can potentially shift between regimes of low and high inflation and vice versa, which enables us to capture the distinct dynamic behavior inflation exhibits when in different regimes. In this context, the purpose of this paper is two-fold: to specify a model that seizes inflation dynamics throughout the different stages of transformation that the Albanian economy has undergone from 1996-2017 and from here describe the characteristics of inflation behavior that are specific to each regime.

The paper is structured as follows. Section 2 gives a brief review of the literature concerning previous applications of the Markov-Switching framework in order to characterize inflation behavior in a vast amount of countries. Section 3 presents the features of the econometric model employed, the dataset and the estimation methodology while Section 4 provides the empirical results. Section 5 finally concludes.

Literature Review

This paper builds on a vast amount of literature in which inflation is modeled as a Markov-Switching process. More specifically, Ricketts & Rose (1995) report that two state Markov-Switching models are favored over one state representations of inflation data for the G-7 countries whereas three-state models prove useful in explaining specific episodes in history for some countries.

Simon (1996) provides evidence that inflation in Australia since the early 1960-s, is well characterized by a two state Markov Switching autoregressive model with a single output gap term that can provide additional information regarding inflation behavior.

Ayuso et al. (2003) rely on a similar approach for Spain, where the stochastic process followed by inflation from 1962-2001 is best described by a Markov-Switching model with three states: a low and stable inflation regime, a medium and more volatile inflation regime and a high and volatile inflation regime.

Bredin & Fountas (2006) on the other hand, rely on a Markov-Switching heteroskedasticity model which allows for shifts in both the mean and variance of inflation to investigate the nature of the relationship between inflation and its uncertainty for four European countries: Italy, Holland, Germany and United Kingdom, at short and long time horizons.

Pagliacci & Barraez (2010) further sustain that a two state Markov –Switching estimation of the Philips curve provides a good characterization of inflation dynamics in Venezuela which in terms

of expectations distinguishes between a “normal or backward looking” regime and “a rational expectation” regime.

Amissano & Fagan (2010a) on the other hand, when modeling the inflation process of Euro Area, US, UK and Canada extend the simple Markov-Switching model by allowing the transition probabilities to vary over time conditional on a smoothed measure of broad money growth corrected for trend velocity and output growth, as a potential leading indicator that contains important information regarding potential switches between regimes of low and high inflation.

Having said that, Markov Switching models provide an alternative approach that is able to capture not only inflation dynamics but also the alteration in behavior that is due to the regime, following the distinct stages of development each country has undergone.

Empirical Analysis

III.1 Model Specification and Estimation Procedure

In this paper, following Amisano and Fagan (2010b), inflation is modeled as a stationary first-order autoregressive process, governed by two distributions with distinct means, conditional on an unobservable discrete state variable s_t that follows a first order Markov chain and determines the switching between distributions (regimes):

$$y_t = \alpha_{s_t} + \Phi y_{t-1} + \varepsilon_{st} \quad \varepsilon_{st} \sim N(0, \sigma_{st}) \quad (1)$$

where ε_{st} is a Gaussian disturbance with a state-dependent standard deviation.

More specifically, there are two regimes in which inflation can potentially be in: $s_t=1$ (low inflation) and $s_t=2$ (high inflation) with the probability of a change in regime depending only on the value of the previous state (regime):

$$P(s_t=j \mid s_{t-1}=i) = p_{ij} \quad i,j=1,2 \quad (2)$$

with the transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad (3)$$

Where $p_{12} = 1 - p_{11}$; $p_{21} = 1 - p_{22}$

Furthermore, considering the model specification, only y_t (inflation) can be observed directly, whereas s_t which dictates the regime switch is a latent (unobservable) variable. In addition, in order to make an inference about the value of s_t that determines the regime inflation currently is in (low, high), we rely on the information derived by y_t . That being said, to assess the likelihood of the discrete state variable s_t , it is necessary to estimate the conditional expectations of $s_t=j$, $j=1,2$ given different information sets.

Additionally, given the collection of observed variables up to time t (in our case inflation) $Y^t = (y_t, y_{t-1}, \dots, y_1)$ which represents the information set available at time t , the information set that

covers the full sample Y^T and the vector of parameters $\ddot{Y} = (\alpha_1, \alpha_2, \Phi, \sigma_1, \sigma_2, p_{11}, p_{22})$, the prediction $P(s_t = j | Y^{t-1}; \ddot{Y})$, filtered $P(s_t = j | Y^t; \ddot{Y})$ and smoothed probabilities $P(s_t = j | Y^T; \ddot{Y})$ can be evaluated.

The inference is implemented iteratively for $t=1, 2, \dots, T$ with the prediction probability

$$\eta_{i,t-1} = P(s_{t-1} = i | Y^{t-1}; \ddot{Y}) \quad (4)$$

acting as an input for $i=1, 2$ and in exchange producing as an output:

$$\eta_{j,t} = P(s_t = j | Y^t; \ddot{Y}) \quad (5)$$

Having said that, the density of y_t conditional on the information set available at time $t-1$, Y^{t-1} and the current state of inflation $s_t = j$, $j=1, 2$ is:

$$V_{jt} = f(y_t | s_t = j, Y^{t-1}; \ddot{Y}) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left[-\frac{(y_t - \alpha_j - \Phi y_{t-1})^2}{2\sigma_j^2} \right] \quad (6)$$

Furthermore, given the prediction probability $\eta_{i,t-1}$ specified in (4) as an input, the density of y_t conditional only on the information set available at time $t-1$, Y_{t-1} can be obtained from (6):

$$f(y_t | Y^{t-1}; \ddot{Y}) = \sum_{i=1}^2 \sum_{j=1}^2 p_{ij} \eta_{i,t-1} V_{jt} \quad (7)$$

From here, given the Bayes theorem, the filtered probabilities of being in regime j in time t are derived:

$$\eta_{jt} = P(s_t = j | Y^t; \ddot{Y}) = \frac{P(s_t = j | Y^{t-1}; \ddot{Y}) * f(y_t | s_t = j, Y^{t-1}; \ddot{Y})}{f(y_t | Y^{t-1}; \ddot{Y})} \quad (8)$$

$$= \frac{P(s_t = 1 | Y^{t-1}; \ddot{Y}) * V_{1t}}{f(y_t | Y^{t-1}; \ddot{Y})}$$

The relationship between the filtered and prediction probabilities can be expressed as:

$$P(s_{t+1} = i | Y^t; \ddot{Y}) = p_{1j} P(s_t = 1 | Y^t; \ddot{Y}) + p_{2j} P(s_t = 2 | Y^t; \ddot{Y}) \quad (9)$$

Where $p_{1j} = P(s_{t+1} = i | s_t = 1)$ and $p_{2j} = P(s_{t+1} = i | s_t = 2)$ are the transition probabilities.

By iterating equations 6-9, assuming the Markov chain to be ergodic (the starting value of $\eta_{i0} = P(s_0 = i) = \frac{1-p_{jj}}{2-p_{ii}-p_{jj}}$), the quasi-log-likelihood function can be derived:

$$L_T(\ddot{Y}) = \log f(y_1, y_2, \dots, y_T | y_0; \ddot{Y}) = \sum_{t=1}^T \log f(y_t | Y^{t-1}; \ddot{Y}) \quad (10)$$

The QMLE $\check{Y}_T$ can be evaluated through the maximization of (10) through a numerical optimization algorithm.

Furthermore, by relying on Kim (1994) smoothing algorithm, the smoothed probabilities can be derived:

$$\begin{aligned} P(s_t=i | Y^T; \check{Y}) &= P(s_{t+1}=1 | Y^T; \check{Y}) * P(s_t=i | s_{t+1}=1, Y^T; \check{Y}) + \\ &P(s_{t+1}=2 | Y^T; \check{Y}) * P(s_t=i | s_{t+1}=2, Y^T; \check{Y}) \\ &= \left(\frac{p_{i1} P(s_{t+1}=1 | Y^t; \check{Y})}{P(s_{t+1}=1 | Y^t; \check{Y})} + \frac{p_{i2} P(s_{t+1}=2 | Y^t; \check{Y})}{P(s_{t+1}=2 | Y^t; \check{Y})} \right) * P(s_t=i | Y^T; \check{Y}) \end{aligned} \quad (11)$$

III.2 Statistical tests: Stationarity and inflation regimes

In this subsection, we examine whether the specified model of inflation behavior in Albania is consistent with the pattern conveyed by inflation data as measured by the Consumer Price Index (CPI)⁸.

In addition, the focus is put on two key issues: 1) Can inflation be treated as a stationary process? 2) Is the Markovian structure supported in the case of Albania or the state variables are independent?

In order to judge on the first issue, thus determine whether the presence of a unit root in inflation data can be rejected in the case of Albania, we rely on several tests as reported by Table 1 (Appendix).

More specifically, according to augmented Dickey-Fuller (1979) ADF test, Phillips-Perron (1988) PP test, the null hypothesis of a unit root can be rejected in both cases with and without a time trend at the 0.05 level of significance. The Kwiatkowski, Phillips & Schmidt (1991) KPS test on the other hand, fails to reject the null hypothesis of inflation as a stationary process against the alternative of the presence of a unit root at the 5 percent level of significance without a time trend, whereas when a time trend is included in the test equation, the null can not be rejected at the 1 percent level.

Regarding the second issue, the independence of the state variables would imply that the current state is not affected by the previous one. In this context, inflation would have the same probability of being in the low/high regime in spite of the previous one: $p_{11}=p_{21}$ and $p_{12}=p_{22}$ therefore a simple switching model could be employed to characterize inflation dynamics. Considering that $p_{11}+p_{12}=1$ and $p_{21}+p_{22}=1$ the null hypothesis of independent states can be represented as: $H_0: p_{11}+p_{22}=1$.

⁸ The quarterly data on Consumer Price Index (CPI) are taken from the database of the Bank of Albania (BoA) and cover a time period from 1996:Q1-2017:Q1, the longest period available; a total number of 84 observations. In addition, inflation is represented by $\log(CPI)$.

In addition, as shown by the Wald test results in Table 2 (Appendix), it can be seen that the null hypothesis of independent states is not supported therefore there is no evidence against the Markovian structure specified in the previous subsection in order to characterize inflation dynamics in Albania.

Results/Findings

This section presents the estimation results and their implication for the regimes inflation can potentially be in (low or high). Given the Quasi-Maximum Likelihood estimates shown in Table 3 (Appendix), it can be determined that $s_t=1$ is the state of low inflation with a mean of 0.59% whereas $s_t=2$ the state of high inflation with a mean of 6.81%.

In addition, by relying on the smoothed probabilities $P(s_t=1|Y^T)$, plotted in Figure 1, it can be seen that after 1998 these probabilities are almost 1, thus it is highly likely that in the last 19 years, the Albanian economy has been in the regime of low inflation.

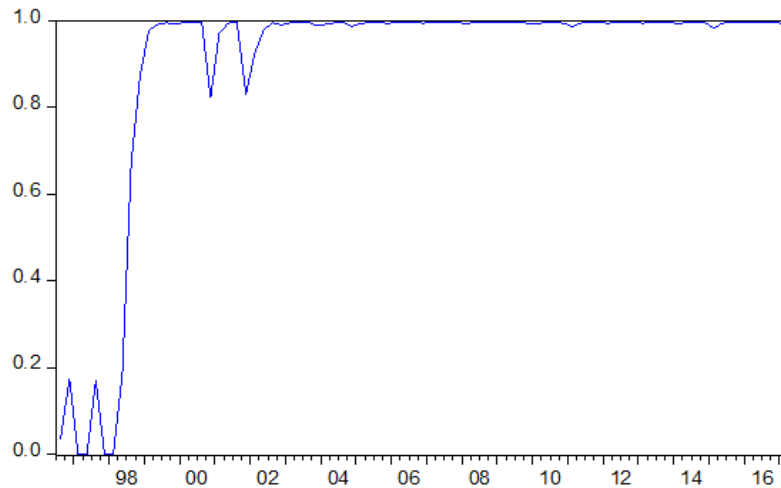


Figure 1: The smoothed probabilities of $s_t=1$, 1996:Q1-2017:Q1

Given the substantial time period throughout which the economy has been under the low inflation regime, it is necessary to provide an explanation as to what is intended with low inflation regime. In this paper, it refers to the time period in which the economy has been characterized by low inflation rates, slightly below or close to the 3% target, thus $s_t=1$ includes periods of price stability and slight deviations from it. Alternatively following the approach of Amisano and Fagan (2010c), $s_t=1$ can be interpreted as the regime of price stability.

In the same way, the smoothed probabilities $P(s_t=2|Y^T)$ as shown in Figure 2, indicate that

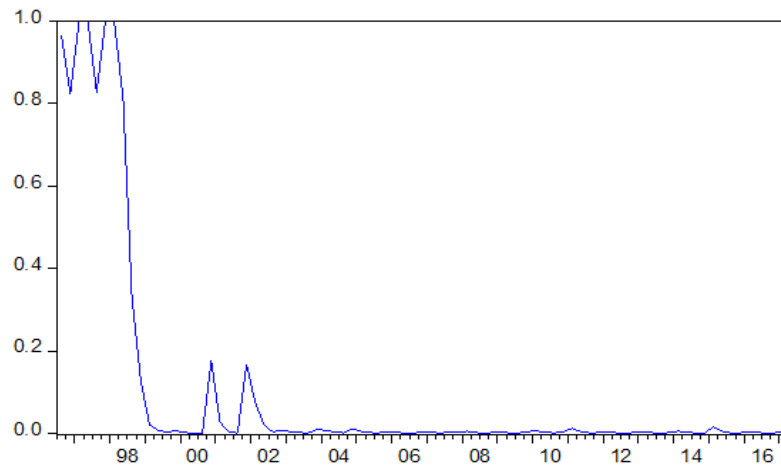


Figure 2: The smoothed probabilities of $s_t=2$, 1996:Q1-2017:Q1

after 1998, the probability of the economy being in the high inflation regime is almost 0 whereas in the years before (1996,1997 given the sample that is used in estimating the model) these probabilities are very close to 1, which indicate that during the time period 1996-1997 the economy has departed from price stability and has been in the high inflation regime.

This is further sustained by a closer look at the inflation data as illustrated by Table 4, where deviations less than 1% from the set target are thought to reflect periods of price stability, deviations more than 1% below the set target, periods of low inflation whereas deviations more than 1% above, periods of high inflation.

Given the classification of regimes in Table 4, it can be seen that in the case of 2002 the deviation is almost negligible compared to 1996,1997 that are periods with the economy in the high inflation regime according to both observed data and the smoothed probabilities. This explains the low value of the estimated smoothed probability $P(s_t=2)$ in 2002 (lower than 0.2) although it is classified as a high inflation regime by relying on inflation data.

In this context, the estimated MSAR(1) model is able to capture the strong inflationary pressures during 1996 as a consequence of the lax fiscal policy ahead of the parliamentary elections in May 2016, that of 1997 when the civil disorder that resulted from the collapse of pyramid schemes combined with enormous supply disruptions, the slump in remittances led to a depreciation of the exchange rate that were reflected in higher inflation, the disinflation that followed the post-crisis of 1997 and the period of low and stable inflation after 2000 when the switching to indirect instruments for conducting monetary policy took place (second half of 2000).

Table 4: Inflation Regimes in Albania 1992-2016

Year	Inflation ⁹	Deviation from Target ¹⁰	Inflation Regime
1992	226	223	High
1993	85	82	High
1994	22.6	19.6	High
1995	7.8	4.8	High
1996	12.7	9.7	High
1997	33.2	30.2	High
1998	20.6	17.6	High
1999	0.4	-2.6	Low
2000	0.1	-2.9	Low
2001	3.1	0.1	Stable
2002	7.8	4.8	High
2003	0.5	-2.5	Low
2004	2.3	-0.7	Stable
2005	2.36	-0.64	Stable
2006	2.37	-0.63	Stable
2007	2.9	-0.1	Stable
2008	3.4	0.4	Stable
2009	2.3	-0.7	Stable
2010	3.6	0.6	Stable
2011	3.5	0.5	Stable
2012	2	-1	Low
2013	1.9	-1.1	Low
2014	1.6	-1.4	Low
2015	1.88	-1.12	Low
2016	2.2	-0.8	Stable

An issue of great importance for policymakers is the duration of each inflation state and the possibility of the economy departing from the regime of price stability.

In addition, given the estimates of the transition probabilities as shown by Table 3, it can be determined that the probability of remaining in the low/stable inflation regime p_{11} is 0.98 whereas the probability of the Albanian economy departing from it is $p_{12}=1-p_{11}=0.02$.

On the other hand, the probability of staying in the high inflation regime in the next period when the current state is that of high inflation is $p_{22}=0.96$ whereas the probability from transitioning into a regime of low inflation/price stability is $p_{21}=1-p_{22}=0.04$.

These estimates indicate significant state dependence in the transition probabilities with a higher probability of remaining in the origin state (0.98 for the stable inflation and 0.96 for the high inflation regime). Moreover given the estimates, it can be seen that the low/stable inflation regime is more persistent than the high one.

The expected duration of the regime of low/stable inflation is approximately $1/1-p_{11} \approx 50$ quarters whereas that of the regime of high inflation approximately $1/1-p_{22} \approx 25$ quarters.

These estimates suggest that the Albanian economy stays longer (twice as long) in the regime of price stability than in that of high inflation.

Conclusions

⁹Consumer Prices (Annual %) derived from the World Bank Database

¹⁰ As a reference point was taken the 3% target set by the Bank of Albania in 2015

This paper, in contrast to the standard linear approach that dominates the literature concerning inflation in Albania, presents an alternative view for modeling inflation dynamics. In addition, by relying on a Markov-Switching framework in which the Albanian economy can potentially shift between regimes of low and high inflation and vice versa, the model is able to capture not only the dynamics of inflation behavior but also the alteration in behavior that is due to the regime switch.

The estimated MSAR(1) model provides a good characterization of inflation behavior in Albania by capturing inflation dynamics throughout the different stages of transformation that the Albanian economy has undergone from 1996-2017, starting with the strong inflationary pressures of 1996-1997 as a combination of the loose fiscal policy ahead of the parliamentary elections and the civil disorder that resulted from the collapse of pyramid schemes, the disinflation that followed the post-crisis of 1997 and the period of low and stable inflation after 2000 when the switching to indirect instruments for conducting monetary policy took place.

The empirical results reveal that the Albanian economy has been in the regime of low/stable inflation for the last 19 years with the probability of departing from price stability only 0.02 whereas the probability of remaining in the current regime in the next period 0.98 which shows the low/stable inflation regime to be quite persistent.

Furthermore, the estimates indicate significant state dependence in the transition probabilities with a higher probability of remaining in the origin state and suggest that the Albanian economy stays longer (twice as long) in the regime of price stability than in that of high inflation.

Having said that, in spite of the restrictions posed by the limited availability of data in the case of Albania, Markov-Switching models provide a useful tool in capturing certain nonlinear patterns in economic time series that have the potential of conveying valuable information for policymaking.

References

- Amisano, G., & Fagan, G. (2013), "Money growth and inflation: A regime switching approach", *Journal of International Money and Finance*, 33, 118-145.
- Ayuso, J., Kaminsky, G. L., & Salido, D. L. (1998), "Inflation regimes and stabilization policies, Spain 1962-1997", (No. 10), FEDEA.
- Bredin, D., & Fountas, S. (2006), "Inflation, inflation uncertainty, and Markov regime switching heteroskedasticity: evidence from European countries", *Economic Modeling*, 36(9), 112-230.
- Domaç, I., & Elbirt, C. (1998), "The main determinants of inflation in Albania", (No.1930), World Bank Publications.
- Hamilton, J. D. (2010), "Regime switching models", *Macroeconometrics and Time Series Analysis* (pp. 202-209). Palgrave Macmillan UK.
- Kim, C. J. (1994), "Dynamic linear models with Markov-switching", *Journal of Econometrics*, 60(1-2), 1-22.
- Kota, V. (2011). "The Persistence of Inflation in Albania," Special Conference Paper No. 3, Bank of Greece

- Kalra, M. S. (1998), "Inflation and money demand in Albania", (No. 98-101), International Monetary Fund.
- Kunst, R. M., & Luniku, R. (1998), "Inflation, its dynamics, and its possible causes in Albania", (No. 57). Institute for Advanced Studies.
- Pagliacci, C., & Barráez, D. (2010), "A Markov-switching model of inflation: looking at the future during uncertain times", *Análisis Económico*, 25(59), 25.
- Ricketts, N., & Rose, D (1995), "Inflation, learning and monetary policy regimes in the G-7 economies", (Vol. 95, No. 6), Bank of Canada.
- Simon, J. (1996), "A Markov-switching model of inflation in Australia ", (No. rdp9611), Reserve Bank of Australia.

Appendix

Table 1: Testing for the Presence of Unit Roots in Inflation 1996:Q1-2017:Q1

Test	ADF t-statistic	
ADF	-5.34	-5.02*
PP t-statistic		
PP	-8.34	-8.84*
LM-statistic		
KPS	0.43 (0.46)	0.18* (0.21)

Note: Each test is conducted with and without a time trend, the values in * report each test's t-statistic when a time trend is included. Furthermore, the values in brackets () report the asymptotic critical values in the case of KPS test at the 0.05 significance level and the 0.01 significance level when a time trend is included.

Table 2: The results of the Wald Test on the independence of the state Variables

Null Hypothesis	t-Statistic	F-statistic
$H_0: p_{11} + p_{22} = 1$	-9.2	84.7

Table 3: The estimation results of the MSAR(1) on inflation 1996-2017

Parameters	s _t =1	s _t =2		
α	0.59 (0.002) [0.003]	6.81 (0.018) [0.000]	Log likelihood=190.58	
Φ	-0.22 (0.112) [0.045]		AIC=-4.42	
σ	-3.86 (0.085) [0.000]	-2.88 (0.27) [0.000]	SC=-4.21	
Transition Probabilities				
	low	high	Expected Durations	
low	0.98	0.02	1 (low inflation)	2 (high inflation)
high	0.04	0.96	50	25

Note: The reported values in () are the estimated Std. errors whereas the values in [] the estimated p-values

Land Cover Statistics as a Measure of Natural Capital Distribution Fairness among altered Administrative Territorial Divisions

Artan Hysa^{1,2}

Epoka University¹, Istanbul Technical University² Albania¹, Turkey² ahysa@epoka.edu.al

Abstract

This paper makes a case on a comparative analysis of two sets of statistical data derived from a single land cover evidence being spatially subdivided according to the spatial pattern of two different versions of local administrative map of Albanian territory. Namely, the first pattern refers to the spatial distribution of local administrative division dating before 2014 and consisting of 36 districts. The second configuration is derived based on the territorial partition as approved in the new administrative reform of 2014, by reorganizing the 36 existing districts into 61 local administrative units.

According to the officially published criteria of the reform, the economic and cultural properties of the territory are highly considered during the decision making process. Whereas, there is no evidence of such an attention on the environmental assets of the land, which in fact is strongly considered as the third pillar of sustainable development. Bearing in mind that the Land Cover is a dimension of environmental properties of a geography, by utilizing it, this paper aims to evaluate the recent Albanian territorial restructuring.

In this research, the land cover information belongs to 2012 CORINE land cover (CLC) spatial data. Two collections of spatial statistical data have been produced utilizing the GIS applications. A comparative analysis between sets is performed by interpreting mainly a single inferential statistical measure, the coefficient variation (CV) or the relative standard deviation (RSD), aiming to express the difference of the relative distribution behaviors of each land cover type separately.

Furthermore, the discussion on distribution fairness of land cover types among municipalities being responsible for managing/ benefiting from them, goes beyond the numerical facts of a coverage map. The ecological service function of each land type and their distribution is an important measure in calculating the natural capital of a territory. The statistical outcomes aim to serve as a judgment of the recent administrative reform. Moreover, the research gives clues on how to introduce the environmental factors of a territory in the decision making process of an administrative territorial reform.

Keywords: Coefficient of Variation (CV), Environmental Economics, Land Cover statistics, Administrative Territorial Reform, Albania

JEL Classification: Q15

Introduction

The problem of natural capital exclusion from the processes of policy making is a well debated issue in global scale (Costanza, et al., 2016). Beyond the reasonable criticism on the assignment of real market values to the natural capital, there are successful efforts on considering the natural capital as ecological services and introducing it to agendas of policy making in developed

countries. On the other hand, the case is not the same in developing geographies, where the short term development pressures cast a total shadow on the values of ecological services in the territory.

In this context, this study makes a case of evaluating the recent (2014) territorial/ administrative reform (as a policy making sample) in Albania (as a developing country). The research goes beyond a critical reading of the objectives and goals of the reform to highlight the lack of the abovementioned concern during the process of the reform. Furthermore, the study relies on a statistical analysis methodology, in order to stress the importance of considering the natural assets during the decision making process. This is done by statistically analyzing the distribution fairness of the natural capital among administrative units in two different versions of spatial division of the territory. Land cover statistics are analyzed under two spatial distribution schemes of pre-reform and post-reform versions of Albanian territory, respectively consisting of 36 and 61 administrative units.

Land Cover Statistics as a Measure of Natural Capital

Ecosystem services flows and natural capital stocks as concepts make ground for very useful methods in order to highlight, measure, and assess the degree of interdependence among humans and nature (Costanza, et al., 2014). The awareness about this interconnection, urges for further attention on the natural values of the territory.

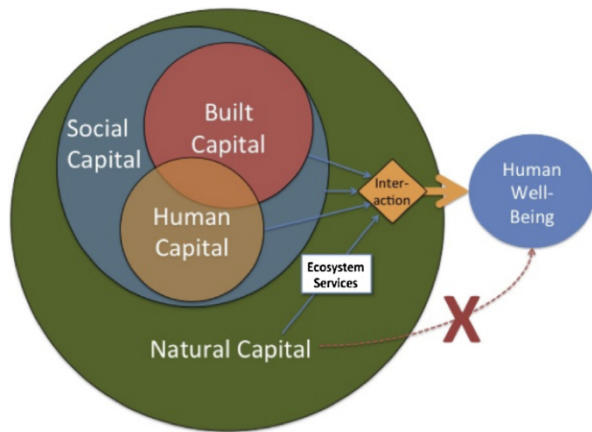


Figure 1. Interaction between built, social, human and natural capital required to produce human well-being. Built and human capital (the economy) are embedded in society which is embedded in the rest of nature. (Costanza, et al., 2014).

Land cover properties of a geography may give reliable facts on the state of its natural assets. Even though, majorly the land cover statistics are utilized to compare and contrast the changes of the natural capital through certain temporal sequences due to land use alterations (Zhang, Zhao, Liu, Liu, & Li, 2015), in this study they are employed to compare and contrast two different distribution versions of the same land cover dataset. Consequently, via this method it can be derived a critique on the natural capital distribution fairness according to dissimilar spatial division of a region.

Albanian Territorial/ Administrative Reform in the Scope of Ecological Services

In 2014, Albania has ratified the national administrative and territorial reform. Basically, the territory is re-organized from 36 administrative units to 61 municipal districts. According to the report on the technical criteria of the division process, the proposal highly consider the economic and social properties of the territory by bringing the concept of *Functional Zones*. The notion infer to a territorial area that is characterized by dense and substantial interaction between citizens and

institutions with economic, social, development and cultural interest (MLA, 2014). Basically, any area that have a relatively considerable level of economic and cultural state, deserves to become a local administrative unit under the new division. On the contrary, the environmental concern which is principally accepted as the third pillar of an integrated sustainable development model (Giovannoni & Fabietti, 2013) is not introduced as part of the decision making process. The anthropocentric approach of focusing on the economic and socio-cultural assets, may successfully guide the definition of the centers of the new administrative units, but it is far away from assisting in defining the borders of rural and natural lands surrounding the centers.

Even though, the environmental concerns seems to be neglected during the recent administrative reform in Albania, some of the physical assets of the territory are increasingly becoming more valuable in the new local administration law. For example, according to the new reform the management and benefiting from Forested Lands is exclusively given to the municipality governing those areas. According to the Directive by the Ministry of Environment dating June 2016, the local government can rent areas from the Forest Fund contributing directly to their institutional budgetary income (MEA, 2016). Besides the forest fund, municipalities may rent other natural lands, like grasslands, marshes, and heathlands for mainly supporting the livestock industry. All these bring new financial opportunities for local government based on the natural properties of the lands they own and are responsible of, which is something worth to be assessed.

Research Questions

Considering all above-mentioned information there emerges a list of researchable queries. The objective of this paper is based on discussing and searching for answers to the following questions;

- Can the territorial reform be criticized/ evaluated based on the Land Cover territorial data?
- Which administrative territorial division is providing a fairer distribution of land cover?
- Can this method tried via Land Cover be comprehensively expended by including other parameters of economic, cultural and infrastructural character?
- Can this method become significant during the process of decision making in the scope of a Territorial Administrative Reform?

Methodology

CORINE Land Cover

In this study, the land cover statistics are used as the specimen mean in measuring the Natural Capital of the territory. More precisely, the experiment will rely on the CORINE Land Cover (CLC) data. CLC provides structured spatial data on the land cover under certain typologies (JRC-EEA, 2005). The CLC nomenclature is structured in three hierarchical levels of surface cover types. The first divides the land surfaces into five main categories; artificial surfaces, agricultural areas, forests and semi-natural areas, wetlands, and waterbodies. The subcategories follows detaining into two other typological sets. In figure 2, it is shown the hierarchical typological division of forests and semi-natural areas. The analytical part of this study will focus on the statistical data of the 3rd category (CLC 300, Forests and semi-natural areas) including all its subclasses. The main selection criteria is based on the high natural capital value this land surface has compared with the other types (Costanza, et al., 2011).

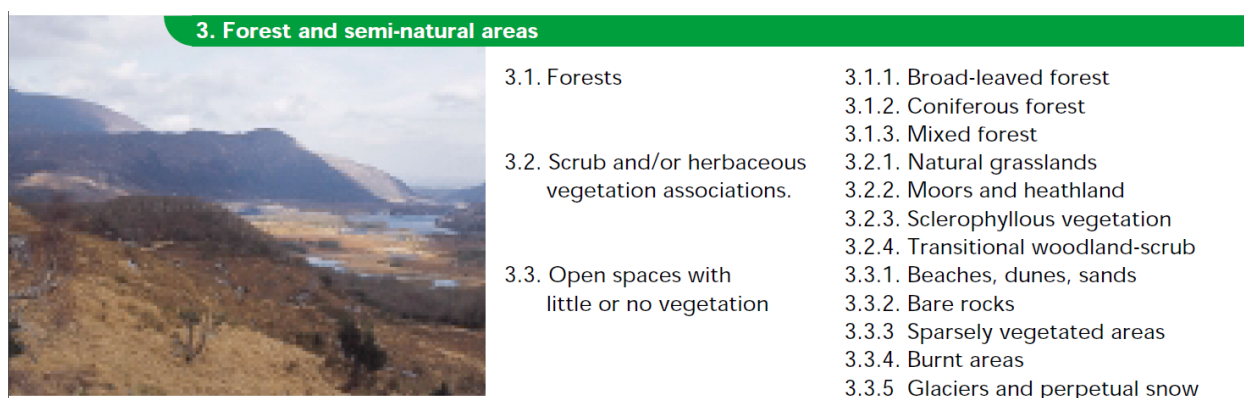


Figure 2. The hierarchical typological division of Forests and semi-natural areas under CLC nomenclature (European Topic Center, 1999).

Spatial Statistics via ArcGIS as Measuring Medium

CLC data of Albanian territory is serving as the main input of this experiment. Open source spatial data derived from EIONET database, provide enough spatial distribution information about the land cover types of several year intervals. In this study it is decided to proceed with the data of 2012, since they coincide with the timeline of the decision making process of the new territorial reform happening between 2013 and 2014.

The original shapefile is spatially split according to two versions of administrative units utilizing the ArcGIS package, version 10.2.2. As an output at this stage a set of statistical tables is produced to be further numerically analyzed via certain statistical measure.

Coefficient of Variation as a Measure of Fair Distribution

Since the same surface area of land cover data is subdivided into two different distribution schemes, the results have to be analyzed in a relative rather than absolute method. This becomes even more crucial while the subdivisions increases from 36 to 61 districts, decreasing the mean surface area of districts by approximately 30%, respectively from 78488 ha to 46569 ha. Thus, absolute statistical values on natural lands surface areas cannot provide ground for correct comparative analysis.

Instead of that the Coefficient of Variation (CV) or the Relative Standard Deviation (RDS) is used to measure the relative distribution of natural surfaces among different local administrative units. Furthermore, CV is advocated to fulfill the requirements for a measure of economic and distribution inequalities (Champernowne & Cowell, 1999) which adds a further dimension to the discussion on equal territorial distribution of natural resources.

Results and Discussion

As the first stage, the Albanian CLC data of 2012 derived via EIONET (figure 2.c), is subdivided according to two different distribution schemes. Firstly, it is split into 36 subdivisions according to the local administrative map of pre-2014 (figure 2.a), and later according to the post reform version it is divided into 61 subsets (figure 2.b). The operation is utilizing the *split* analysis tool of ArcGIS package.

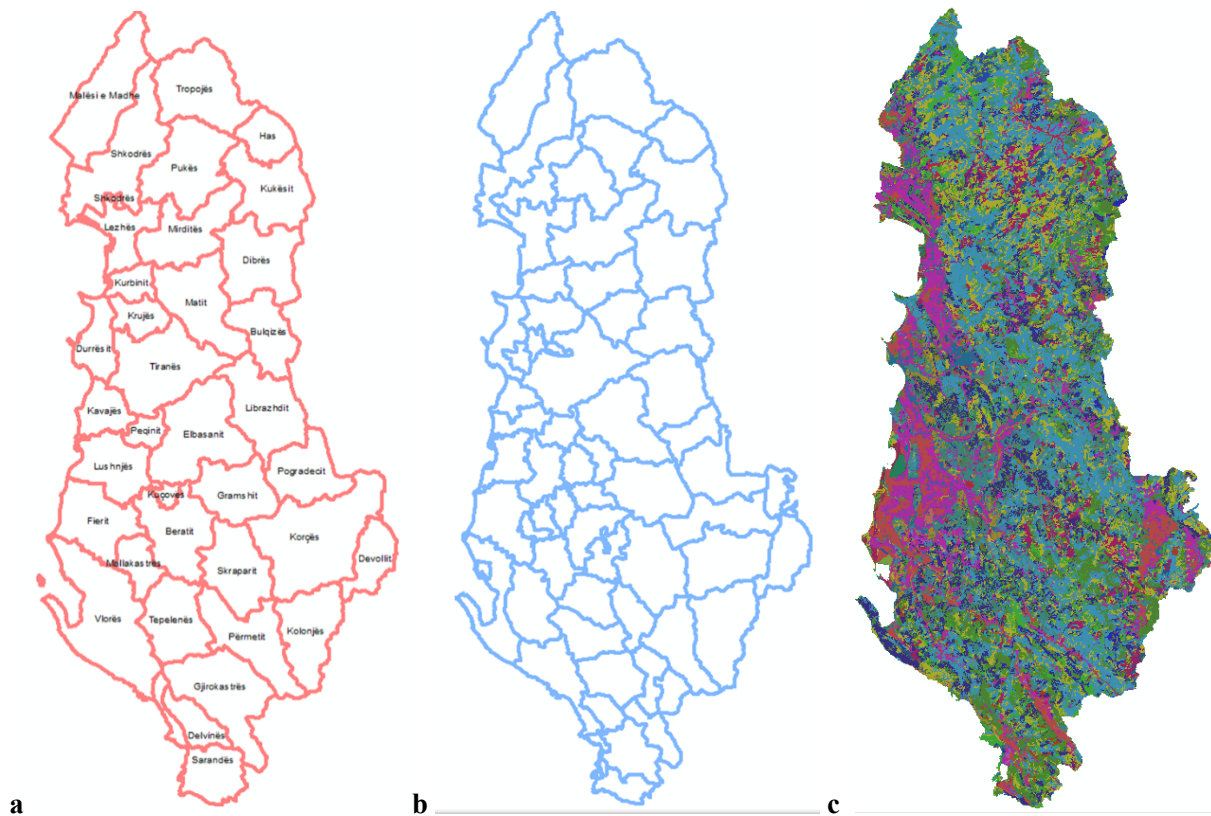


Figure 3. Local Administrative Units in Albania; [a] before 2014, [b] after 2014 and [c] CORINE Land Cover map of 2012

The spatial distribution of CLC data into two different versions leads to statistical data about the surface areas (ha) of each CLC class under each local administrative unit. Consequently, two separate tables of numerical data is produced. For example, in table 1 it is shown the numerical data about the distribution of each 3rd level CLC surface areas among each municipality in both spatial division versions. Besides, a cumulative column represent the total surface area of natural lands (CLC-300) inside the borders of each administrative unit. Furthermore, based on the numerical data of the same table, there are derived other statistical information about the mean, standard deviation and the coefficient of variation of surface areas distribution of each CLC subclass. All these measures makes ground for further discussion on the natural surfaces distribution behaviors under different territorial divisions.

First of all, referring to the absolute values of surface areas of each 3rd level CLC class, it can be stated that the municipality that leads the possession of broad-leaved forested surfaces (clc-311) remains the same as Tropoje, but municipalities like Librazhd, descends from the 2nd place to the 4th after the new administrative reform, with a total broad leaved forest area reduction of 5000 ha. Similarly, considering the numbers of coniferous forested surfaces (clc-312), after the reform there is a decrease by 50 % in the case of the former leading municipality, Puke. Furthermore, referring the summary (SUM) column of both tables, it can be stated that Korca as the former leader district has dropped to the 14th place among municipalities possessing the largest amount of natural lands under their administration. The comparative analysis of the absolute values may lead to further

comparison between different municipalities and the municipality in itself. Besides that, additional important deductions can be drawn via inferential statistics derived from the tables.

Table 1. Numerical data about surface areas (ha) of CLC-300 classes in two versions of 36 (a) and 61 (b) districts.

		Broad-leaved forest	Coniferous forest	Mixed forest	Natural grasslands	Moors and heathland	Scierophyllous vegetation	Transitional woodland-shrub	Beaches, dunes, sands	Bare rocks	Sparsely vegetated areas	Burnt areas	SUM
		311	312	313	321	322	323	324	331	332	333	334	
Berat	1	13079	1899	418	10117		12265	7072	603		1451	84	46988
Bulqizë	2	22801	474	42	16104	140	6332	9809	198		1371		57270
Delvine	3	4923	319		10768		2187	4257	179		2243	458	25333
Devoll	4	9023	1529	1194	3658	48	1842	7699			214	820	26026
Diber	5	16094	6175	4687	18996	782	6027	18520	608		5897	68	77854
Durrës	6	6565	554		1787		1312	863	138		29		11248
Elbasan	7	25912	1471	253	12540		16157	10420	2936		4055		73742
Fier	8	1514	1769	83	609		2135	2107	314		448		8979
Gjirokastër	9	22240	1470	188	18897	2403	12241	13921	846	54	9263	2541	84064
Gramsh	10	19856	5438	1855	8883		12329	7550	2687		1437		60036
Has	11	12209			5009		2696	7513			947		28375
Kavajë	12	867	589		3375		5694	1001	99		168		11793
Kolonjë	13	18333	6812	3200	14583	4654	3832	8954	152		4201		64721
Korçë	14	29995	10521	4254	25588		14179	23790	325	116	4830	110	113708
Krujë	15	6972	1467	59	1012		1690	2853	169		687		14909
Kuçovë	16	217			264		731	59	27				1298
Kukës	17	16436	3770	4449	17047	1373	6596	19421	206		6893		76190
Kurbin	18	9202	226	52	340		1326	1407	35		193		12782
Lezhë	19	7848	1205	855	2337		3207	7209	223		4780		27664
Librazhd	20	35840	331	247	13600	1840	10235	11878	710		4294	307	79282
Lushnjë	21	2338	873	1	989		571	1235	106		56		6169
Malesi e Madhe	22	22776	1438	2602	7124	271	7475	14835	334	1852	14217		72925
Mallakastër	23	3720	123	209	1725		7943	1721	650		1120		17210
Mat	24	35069	3272	905	11399		14807	12652	392		2407		80904
Mirditë	25	30725	7866	3514	3194		4839	22135	199		5627		78100
Peqin	26	245	910		2723		1764	1450	9		162		7262
Permet	27	29943	2427	267	6842	3654	7163	10669	10		9658		70634
Pogradec	28	23042	480	153	5766		3512	9384			66		42403
Pukë	29	31204	14336	6152	2376		6614	28181	158		6193		95215
Sarandë	30	6780	93	105	17580		6220	7869	32		6717	1418	46815
Shkodër	31	29281	818	653	5204	65	10263	20216	1539	642	17161	127	85968
Skopar	32	21573	5266	417	10977		8356	10230	1059		6796		64674
Tepelenë	33	14226	709	447	14444	350	13447	11078	770		4946	1288	61704
Tiranë	34	31689	553	477	10814		18360	9294	1601		4209		76998
Tropojë	35	35931	981	2570	16143	50	8682	15017	158	3731	8273		91536
Vlorë	36	22934	3406	1235	16962	172	32788	14341	2148	176	5814	1225	101200
Grand Total		621403	89570	41542	319776	15801	275816	356610	19621	6571	146824	8446	1901979
Mean	a	17261	2634	1385	8883	1215	7662	9906	595	1095	4195	768	52833
Standard Dev.	b	11434	3261	1718	6788	1520	6469	7113	763	1457	4081	785	32217
CV [b/a]	a	0.66	1.24	1.24	0.76	1.25	0.84	0.72	1.28	1.33	0.97	1.02	0.61

		311	312	313	321	322	323	324	331	332	333	334	SUM
Belsh	1	1103			1843		442	235	136		47		3807
Berat	2	5740	540	117	4048		5568	3033	424		326	84	19879
Bulqizë	3	21511	1118	190	16139		7854	11513	276		1007		59608
Cerrik	4	816	204		2391		3973	308	784		119		8595
Delvine	5	2883	389		5139		1214	2292	178		1259	303	13656
Devoll	6	9188	1375	1128	3703	297	1957	8377			211	819	27056
Diber	7	15734	5839	4264	18648	714	6098	17871	547		5902	68	75683
Divjake	8	1752	916		174		44	1097	402		34		4419
Dropull	9	9188	793	55	6092	1629	3658	6287	374	54	3803	1395	33327
Durres	10	6002	558		1417		944	647	349		46		9961
Elbasan	11	24243	1201	277	8234		12001	9438	2024		3654		61071
Fier	12	1316	1720	84	315		1732	1935	558		549		8209
Finiq	13	5691	93	105	12648		3078	5155			2769	1323	30862
Fushëarrez	14	19914	7051	3575	592		1803	13103			2207		48245
Gjirokaster	15	9154	472	175	7881	653	6145	3756	362		3544	1152	33293
Gramsh	16	19315	5588	1831	9406		12525	7557	2650		1477		60350
Has	17	12462			5263		2936	7508			970		29138
Himare	18	14807	183	179	11026	172	8432	8153	490	137	4148	471	48200
Kamez	19	74					69		30				173
Kavajë	20	223	194		1926		3126	485			47		6000
Kelcyre	21	5368	481		2989	479	4162	2509			3700		19689
Klos	22	9867	2416	465	4099		7363	4150	93		1084		29537
Kolonjë	23	18835	7287	3675	15433	4777	3913	9303	152		4205		67579
Konispol	24	581			6460		3202	3414			1605	60	15322
Korce	25	11903	8609	3357	11847		3706	11598	69	116	3693	55	54951
Krujë	26	7549	1570	59	1005		1438	2988	154		685		15448
Kucovë	27	2419			1631		2492	313	94				6949
Kukës	28	16734	3794	4778	16534	1336	6497	19350	227		6890		76140
Kurbin	29	9385	316	65	391		1399	1727	83		234		13599
Lezhë	30	7006	1186	916	2441		3284	7473	379		4946		27632
Libohovë	31	3662	35		4383	1226	2122	3989	100		2361	35	17913
Librazhd	32	30792	274	201	11711	1603	8644	8024	720		4203	72	66242
Lushnjë	33	672	2		773		538	578			38		2601
Malesi e Madhe	34	22438	1438	2601	7076	271	6989	14774	368	1830	12840		70625
Maliq	35	11774	1547	945	9064		7684	6242	252		1055	57	38620
Mallakastër	36	3805	123	313	1924		7960	1492	473		838		16925
Mat	37	20108	507	741	4197		4957	6716	299		1301		38827
Mirditë	38	30993	7789	3534	3223		4847	21900	199		5610		78094
Mëmëaliq	39	5061	300	447	4416		8702	3932	484		2426	233	26001
Patos	40	462			216		166	29			82		953
Peqin	41	264	926		2681		1640	1379	2		159		7052
Permet	42	24874	2057	284	3095	1932	3610	8042	15		5036		48946
Perrenjas	43	11967	122	40	4989	301	2757	5107	36		666	235	26221
Pogradec	44	22697	464	150	5542		3654	9392			70		41968
Polican	45	4922	1013	250	2552		3483	3261	145		1094		16720
Pukë	46	10973	7331	2127	1593		4706	14827	111		3477		45146
Pustec	47	6655			3183		2580	6030					18447
Roskovec	48	238			97		769	26			81		1211
Rrogozhinë	49	573	396		1644		2811	826	169		149		6568
Sarandë	50	146			42		237	588			2529		3542
Selenice	51	7209	50	72	7634		10837	3874	1327		2558	382	33941
Shijak	52				225		216	107					548
Shkodër	53	21110	213	653	2794	65	5495	11856	1148	489	12070	127	56019
Skrapar	54	21680	5595	319	12146		7759	10230	1012		6755		65497
Tepelenë	55	8652	421		10310	367	4237	7329	334		2774	1047	35470
Tiranë	56	30948	513	477	10425		17055	8947	1584		4163		74112
Tropojë	57	35258	1013	2575	15960	50	8462	14913	158	3903	8204		90495
Ura Vajgurore	58	595	150	1	1939		707	324			5		3721
Vau i Dejës	59	9451	571		2500		5035	8306	505		6928		33297
Vlorë	60	4197	3142	985	3988		14936	3665	550	39	952	622	33077
Vorë	61	667			590		1397	212			27		2893
Grand Total		623603	89883	42009	320627	15871	276045	358491	20823	6568	147610	8540	1910071
Mean	a	10393	1798	1105	5344	992	4525	5975	463	938	2590	449	31313
Standard Dev.	b	9442	2408	1396	4880	1181	3729	5285	539	1455	2878	468	24621
CV [b/a]		0.91	1.34	1.26	0.91	1.19	0.82	0.88	1.16	1.55	1.11	1.04	0.79

b

Referring to table 2, with the increase of the number of administrative units, the mean of surface areas of each land cover class in focus per each district, is reduced by 32 %. Similarly, standard deviation values for the distribution of each CLC-300 class, are decreased with an average of 25 %. Thus, relying on the absolute values of mean and standard deviation while trying to compare the distribution behaviors is not useful. Instead, the relative inferential values may imply reliable information about the distribution fairness of the same resource into two different dispersal schemes.

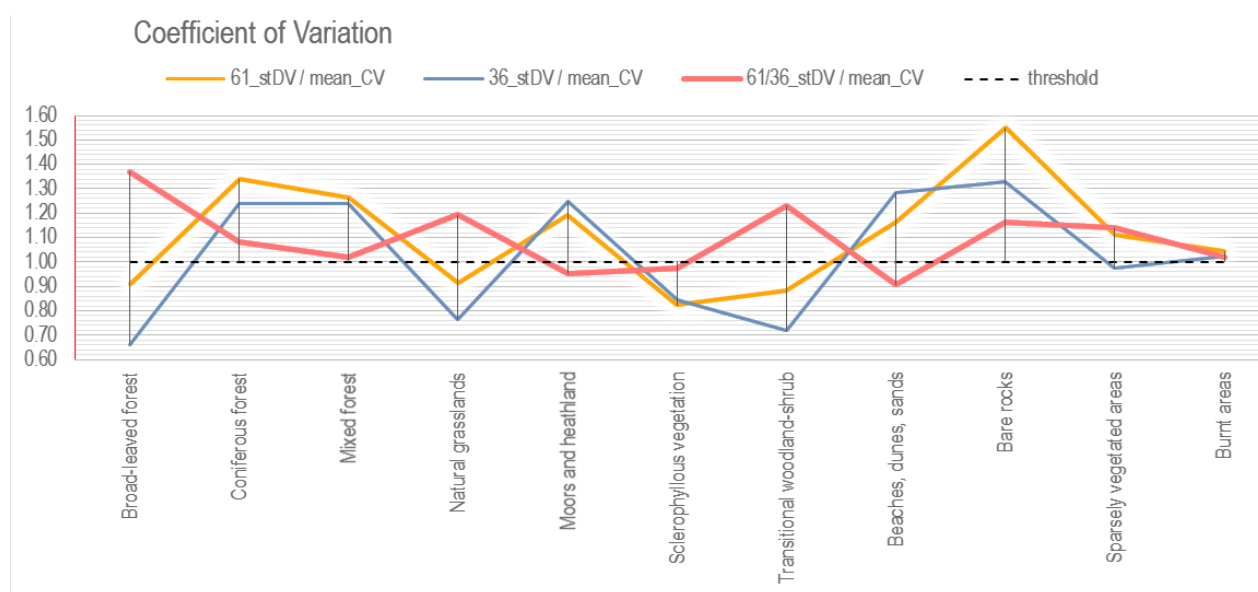
Table 2. Joint inferential statistics of CLC-300 classes

CLC-Classes	CLC-code	61_Grand Total	61_mean	61_stdV	61_stdV / mean of stdV	61_stdV / mean_CV	36_Grand Total	36_mean	36_stdV	36_stdV / mean of stdV	36_stdV / mean_CV	61/36_mean	61/36_stdV	61/36_stdV / mean of stdV	61/36_stdV / mean_CV
Broad-leaved forest	311	623603	10393	9442	3.09	0.91	621403	17261	11434	2.77	0.66	0.60	0.83	1.11	1.37
Coniferous forest	312	89883	1798	2408	0.79	1.34	89570	2634	3261	0.79	1.24	0.68	0.74	1.00	1.08
Mixed forest	313	42009	1105	1396	0.46	1.26	41542	1385	1718	0.42	1.24	0.80	0.81	1.10	1.02
Natural grasslands	321	320627	5344	4880	1.59	0.91	319776	8883	6788	1.65	0.76	0.60	0.72	0.97	1.19
Moors and heathland	322	15871	992	1181	0.39	1.19	15801	1215	1520	0.37	1.25	0.82	0.78	1.05	0.95
Sclerophyllous vegetation	323	276045	4525	3729	1.22	0.82	275816	7662	6469	1.57	0.84	0.59	0.58	0.78	0.98
Transitional woodland-shrub	324	358491	5975	5285	1.73	0.88	356610	9906	7113	1.72	0.72	0.60	0.74	1.00	1.23
Beaches, dunes, sands	331	20823	463	539	0.18	1.16	19621	595	763	0.18	1.28	0.78	0.71	0.95	0.91
Bare rocks	332	6568	938	1455	0.48	1.55	6571	1095	1457	0.35	1.33	0.86	1.00	1.35	1.17
Sparsely vegetated areas	333	147610	2590	2878	0.94	1.11	146824	4195	4081	0.99	0.97	0.62	0.71	0.95	1.14
Burnt areas	334	8540	449	468	0.15	1.04	8446	768	785	0.19	1.02	0.59	0.60	0.81	1.02
cumulative mean		173643	3143	3060	1.00	1.11	172907	5054	4126	1.00	1.03	0.68	0.75	1.00	1.10

The coefficient of variation (CV) is selected as the proper inferential value to be used in this study. CV consists of the ratio between the standard deviation and the mean of surface area values [ha] of each natural land cover class. As it is represented in table 2, CV is calculated for both territorial division versions; 61 and 36 administrative units. In principle it can be stated that the lower the CV value, the fairer distributed natural lands we have. For example, from the same board it can be inferred that the fairest dispersed natural land under the new reform is CLC-323 (Sclerophyllous vegetation) having a CV value of 0.82. On the other hand, the least fair dispersed natural land under the same scheme is CLC-332 (Bare rocks) with a value of 1.55. Whereas, considering the scheme of 36 districts, CLC-311 (Broad leaved forests) is the fairest distributed land cover class with a CV of 0.66. Whereas, the least fairly spread areas in the pre-reform distribution scheme again is CLC-332 (Bare rocks). Meanwhile, land cover classes such as *mixed forests and burned areas*, face almost no change in their CV values, implying an unchanged distribution behavior under two different schemes, consisting of difference of 0.02.

Another inferential statistical instrument is the ratio of CV_{61} to CV_{36} . The result is a coefficient that implies direct comparison between two spatial division conditions being relatively contrasted. Considering the previous assumption that the lower the CV, the fairer the distribution, it can be further assumed that the lower ratio CV_{61}/CV_{36} indicates for an improvement in distribution fairness of natural lands under the new reform. More precisely, the ratio value above 1.00 imply for a worsening of the distribution fairness. On the contrary, the values below 1.00 indicate increases of dispersal equality of land cover surfaces. Referring to the final column of table 2, it can be identified that the most negatively affected distribution with the new reform resulted in *broad leaved forested* surfaces (CLC-311) which stands with the highest ratio of CV_{61}/CV_{36} as 1.37. Whereas the most positively affected one is *beaches, dunes and sand* surfaces (CLC-331) holding a ratio of 0.91.

Chart 1. The Coefficient of Variation [stDV/mean] of two versions [36 and 61], and the ratio among them [61/36]



In addition, referring to chart 1, the land cover classes being above the threshold line of 1.00, thus being negatively affected by the new territorial reform are 8 out of 11 land surface types. Only, *moors and heathlands*, *sclerophyllous vegetation* and *beaches, dunes and sands* surfaces are positively affected by the new reform. Moreover, according to table 2, the cumulative mean value of separate CV_{61}/CV_{36} ratios is 1.10. In other words, it indicates for an overall worsening of the natural lands distribution under the new reform, compared with the pre-reform territorial subdivision scheme of 36 districts.

Conclusions and Final Remarks

This paper presents a methodology of evaluating a policy making process such as territorial reform, relying on the concept of natural capital distribution fairness. The method is applied to the case of the Albanian administrative/ territorial reform approved in 2014. On the other hand, the land cover properties of the territory have been used as tangible measures of the natural capital. The specimen of the experiment is the 3rd category of the first level of CORINE Land Cover classification (CLC-

300). Under this main group, there are 11 subclasses of natural land surfaces such as broad leaved and coniferous forests, which are considered to have the highest natural capital values among other land cover types.

The open source CLC data via EIONET portal, split into two different spatial distribution schemes of pre-reform and post-reform versions, resulted in referable numerical data on surface areas of each land cover class. The need for a relative comparative statistical evaluation tool proceeded to the introduction of the *coefficient of variation* concept. Absolute values of CV_{61} and CV_{36} , as well as the ratio between them, guided the evaluation process to remarkable findings.

First of all, it can be concluded that in overall the natural surfaces- as part of the natural capital- of Albanian territory are less fairly distributed under the new reform than it was under the case of the pre-reform local administrative division map. Meanwhile, among 11 subclasses of natural surfaces, only 3 of them seems to be better fairly distributed than before; being *moors and heathlands*, *sclerophyllous vegetation* and *beaches, dunes and sands* surfaces.

Furthermore, it can be stated that with the new reform the most adversely affected distribution fairness, happened to the broad-leaved surfaces. This fact becomes bolder, while according to literature, this land cover type holds the highest natural capital value among others. On the other hand, the most positively affected land cover type by the reform is *beaches, dunes and sand*. This can be considered positive while thinking of the touristic potential of beaches along Adriatic and Ionian seas.

Also, according to the results of the study, the fairest distributed natural land cover type in the map of 36 districts was *broad-leaved forests*, which drops to the 4th place after the redistribution into 61 units. Whereas, the fairest dispersed natural surface after the reform is *sclerophyllous vegetation*, climbing from the 3rd place of the pre-reform distribution.

In overall the work presented in this paper can be considered as an original utilization of a statistical concept such as the *coefficient of variation* in assessing the natural capital distribution fairness among altered administrative territorial divisions. Additionally, the technique experimented in this work can be considered as rational critique to a decision making process such as territorial/ administrative reform in terms of fair distribution of resources.

Finally, the method proposed in this study can be accepted as a contribution to the comprehensiveness of a policy making process such as territorial reform. Even though, relying just on the land cover statistics is not strong enough, the strategy holds the potential to be expanded by including a variety of environmental properties of a territory facing a local administrative reform.

Acknowledgment

The author is grateful to the European Environment Information and Observation Network (EIONET), for providing CORINE Land Cover data as an open source.

References

- Champernowne, D. G., & Cowell, F. A. (1999). *Economic Inequality and Income Distribution*. Cambridge: Cambridge University Press.
- Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., . . . van den Belt, M. (2016). The Value of the World's Ecosystem Services and Natural Capital . In P. Newell, *The Globalization and Environment Reader* (pp. 117-134). West Sussex: John Wiley & Sons.
- Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I., . . . Turner, R. K. (2014). Changes in the Global Value of Ecosystem Services. *Global Environmental Change*, 152-158.
- Costanza, R., Kubiszewski, I., Ervin, D., Bluffstone, R., Boyd, J., Brown, D., . . . Yeakley, A. (2011). Valuing Ecological Systems and Services. *F 1000 Biology Reports*. doi:10.3410/B3-14
- European Topic Center. (1999). *Corine Presentation*. Retrieved 02 20, 2017, from <https://faculty.washington.edu/eliezg/research/FinlandWolves/corinerepresentation.pdf>
- Giovannoni, E., & Fabietti, G. (2013). What Is Sustainability? A Review of the Concept and Its Applications. In C. Busco, M. L. Frigo, A. Riccaboni, & P. Quattrone (Eds.), *Integrated Reporting; Concepts and Cases that Redefine Corporate Accountability* (pp. 21-40). Switzerland: Springer International Publishing.
- JRC-EEA. (2005). *CORINE land cover updating for the year 2000: image 2000 and CLC2000*. Ispra: JRC-Ispra. Retrieved 02 15, 2017, from http://image2000.jrc.ec.europa.eu/reports/image2000_products_and_methods.pdf
- MEA. (2016, 06 09). *Forest Fund Legislation*. Retrieved 02 25, 2017, from Ministry of Environment of Albania: http://www.mjedisi.gov.al/files/userfiles/Pyjet/Udhezimi_Nr._1_Date_09.06.2016_merge_d.pdf
- MLA. (2014, 03 06). *Albanian Administrative- Territorial Reform; Technical Criteria*. Ministry of Local Affairs. Tirana: MLA. Retrieved 03 06, 2017, from http://www.reformaterritoriale.al/images/presentations/KRITERET%20TEKNIKE%20PROPOZIMI%20PER%20KOMISIONIN_28Prill2014.pdf
- Zhang, Y., Zhao, L., Liu, J., Liu, Y., & Li, C. (2015). The Impact of Land Cover Change on Ecosystem Service Values in Urban Agglomerations along the Coast of the Bohai Rim, China. *sustainability*(7), 10365-10387. Retrieved 2017, from www.mdpi.com/journal/sustainability

Macro-Economic Factors Affecting Ease of Business in Balkan Countries

Bora Kokalari, Emanuela Buci

Epoka University

Abstract

The economy is like the universe, it keeps expanding. The businesses now days are getting more and more interrelated with each other because of market globalization. The ease of doing business in a country is affected by various factors such are: the macro and microeconomic indicators, economic and political stability as well as the policies of the government. This paper aims to examine the macroeconomic factors affecting the ease of business in Balkan countries. Multiple regression analysis was used to examine the relationship between macroeconomic indicators and the ease of business index. The macroeconomic data include variables such are: GDP per capita, unemployment rate, inflation, lending rate and new business density. The data were retrieved from the World Bank Database. In order to test the hypothesis a VAR model is used.

Keywords: *ease of business, macroeconomic indicators, multiple regression, globalization*

Introduction

Economic growth always remains one of the top priority agendas for the governments in the global economic and political environment. Regulators try to provide better regulatory environment for various stakeholders of the economy. It is true that every country may have different problems and challenges, and due to that initiatives, which are taken up by regulatory bodies will be different. Because of this it is difficult to compare the regulatory mechanism by all the stakeholders and even by the regulators themselves. Without comparing with other country, the government may not decide whether the practice followed by them is best or yet there is a scope to improve.

Each country has its own environment created by unique types of policies, demographics, and economic conditions. Knowing which factors and policies are conducive to the growth of new business could be extremely beneficial for several reasons. Countries wishing to increase the amount of innovation and expand their markets could use such information to implement new policies to move their country in that specific direction. The lack of research done in this particular subject is detrimental to developing countries that wish to emulate the previous successes of more developed countries. Therefore, awareness of the industrial demographic is essential.

Research objectives:

1. To assess empirically the influence of macroeconomic factors on the ease of doing business in Balkan countries.
2. To analyse or interpret links between macroeconomic factors and ease of doing business in Balkan countries.
3. To observe the speed of the effect that the macroeconomic factors have on the ease of doing business (if any) in Balkan countries.

Literature Review

According to the research conducted for this topic, there is no previous study for the macroeconomic factors affecting the ease of Business on Balkan Countries. There are many

models showing the gaps of Balkan Countries in the long term productivity competitiveness. Ease of doing business is an index measuring the simplification of regulation for businesses and security of property rights. This index is created by the World Bank. Since there are no previous studies for the region, the literature review will be divided into two sections: the effects of macroeconomic affecting the ease of doing business and the ease of doing business effects on economic growth.

According to the Kosovo Banking Operations (2015), the ease of doing business is affected by the macroeconomic factors such as: inflation according to consumer prices and unemployment. By a comparative analysis conducted by the World Bank, they come up with the following conclusions: Inflation for the period 2008-2013 has been decreasing for all the Western Balkan Countries, thus the cost of doing business has been lower. From 2013-2014 there has been an 37.1% decrease of doing business in Albania, 2.8% increase in Croatia, 7.5% decrease in Kosovo, 3.3% decrease in Macedonia, 14.29% decrease in Montenegro and 18.18% increase in Serbia. These values are also related with the level of unemployment which has been increasing in Albania, increasing in a low values in Croatia, decreasing in Kosovo, decreasing in low values in Macedonia, remaining more or less constant in Montenegro and increasing in Serbia.

There is a convergence challenge in the living standards of Balkan countries in order to enter to the European Union. The main reason is the gap in the competitiveness of the Western Balkan countries in the level of long-term productivity. Peter Sanfey, Jakov Milatovic and Ana Kresic (2016) explain the macroeconomic stability, low unit labor cost and favourable tax regimes combined with the well-educated population are key success indicators for the overall economic growth of these countries.

In a paper by Caliendo and Kritikos (2007), the three most frequently quoted motivations for being self-employed are 1) termination of employment, 2) “being my own boss” and 3) “had first customers”. While the latter two are typically ‘pull’ motives, the first is a ‘push’ motive that creates a job for the self-employed person. At the same time, public authorities provide startup subsidies hoping to create more jobs via these new entrepreneurs

Harris and Hossen (2005) talk about the effects of tax reductions on the growth of businesses. There is an increase in cash flow for even a marginal reduction in taxes and the incentives are two-fold. While the business-owner has a greater rate of return for the effort he or she puts in, there is also an increase in after-tax profits.

It is known that starting a new business requires capital. The expectation is that the easier it gets for potential or nascent entrepreneurs to pay back the ensuing debt, the more they will engage in entrepreneurial activity. Meza and Webb (1996) make an interesting observation about the relationship between entrepreneurial ability and debt. More able entrepreneurs, when financed with debt, make safer choices and, thus, are less likely to default. Less able entrepreneurs, on the other hand, make riskier choices and are more likely to default. Either way, any level of entrepreneurial ability financed with debt results in a great amount of change in the market equilibrium.

Economic growth (especially GDP) has been one of the most important characteristics of the Balkan Countries according to Engell Pere and Albana Hashorva. The economic growth has been

strongly related to the overall administrative facilitations and tax reductions for the sake of doing business. The secondary data show that the economic growth is depended and strongly related to the administrative facilities and rules of doing business. There are also other factors that affect the economic growth such as: income level per capita and improvement of credits facilitation.

Data and Methodology

The focus of our study is to identify factors in the environment of a country that can encourage or hinder business. The variables chosen in this study are GDP per capita, unemployment rate, lending interest rate, Inflation rate and distance to frontier as the main macroeconomic factors that affect new business density. All data are retrieved from World Bank database and they are taken in yearly basis from 2005 until 2015.

GDP per capita is selected to be the first independent variable. Per capita GDP is a measure of the total output of a country that takes gross domestic product (GDP) and divides it by the number of people in the country. The per capita GDP is especially useful when comparing one country to another, because it shows the relative performance of the countries. A rise in per capita GDP signals growth in the economy and tends to reflect an increase in productivity.

Unemployment rate is the share of the labor force that is jobless, expressed as a percentage. It is a lagging indicator, meaning that it generally rises or falls in the wake of changing economic conditions, rather than anticipating them. When the economy is in poor shape and jobs are scarce, the unemployment rate can be expected to rise. When the economy is growing at a healthy rate and jobs are relatively plentiful, it can be expected to fall. The reason why this variable was taken was because that previous studies have shown that when people are left without a job, they have the tendency to become self-employed by opening new businesses. A dearth of jobs could lead entrepreneurs to attempt to create their own jobs as they are willing to take more risks because of a lack of steady income.

Lending interest rates are the interest rates banks provide to the private sector for loans. Interest rates play a key role in economics as government can use them to control monetary policy and slow or speed up an economy. They also play an enormous role in businesses- money is critical to begin a new venture. Higher interest rates deter people from taking out loans, and thus people that need loans to start a business will wait until rates fall to do so.

New business density is defined as new registrations per 1000 people between ages 15-64. The economic reasoning behind choosing this variable is simple- a larger number of businesses opening up will lead to more competition- and thus we expect to find a positive relationship between the two variables (higher business density will lead to higher ease of business numerical values which implies low ease of business).

The distance to frontier score aids in assessing the absolute level of regulatory performance and how it improves over time. This variable shows the distance of the economy from the point that it is in a certain moment to the best performance that it could be. It is measured on a scale from 1 to 100 (from worst to best). We expect to have a positive relationship between distance to frontier and ease of doing business since the better the economy is performing the easier is for businesses to operate.

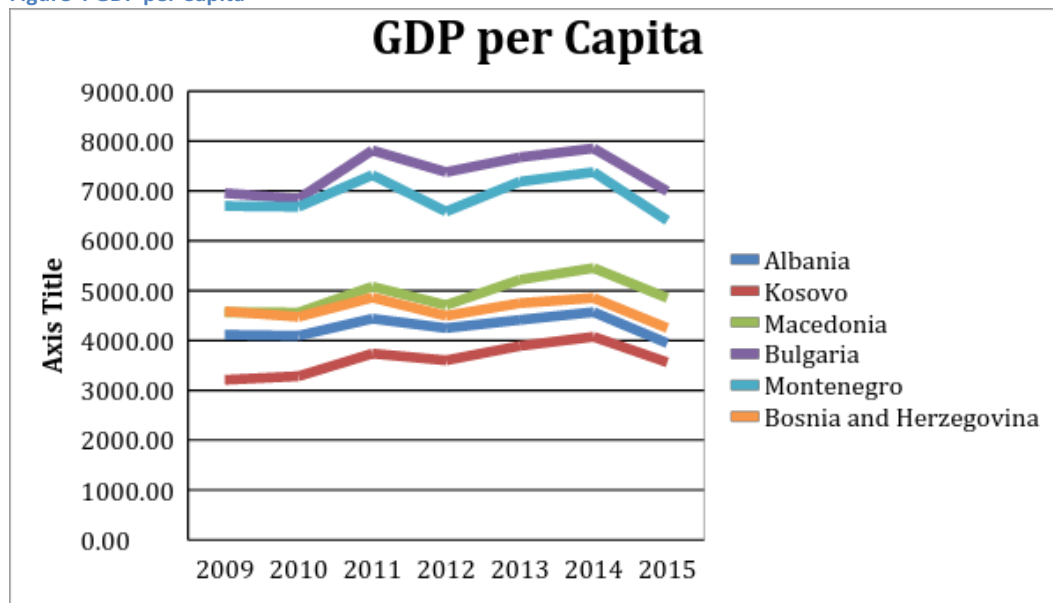
In order to compare the effect that each independent variable has on the ease of doing business for each Balkan country separately a multiple regression was conducted for Albania, Kosova, Bosnia and Herzegovina, Macedonia, Montenegro and Bulgaria. As a second step, a VAR model was conducted to see how these effects have changed over the years for the Balkan region as a whole. Finally, Granger Causality was tested in order to see if there is any causal relationship between the variables.

Empirical Results

Graphical representation

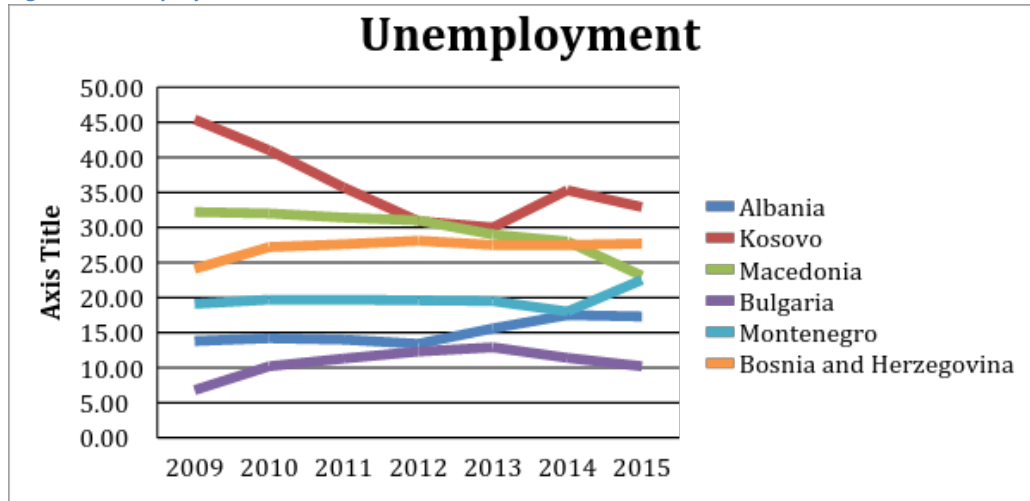
A graphical representation of all the countries chosen in the study was conducted for all the variables separately.

Figure 4 GDP per Capita



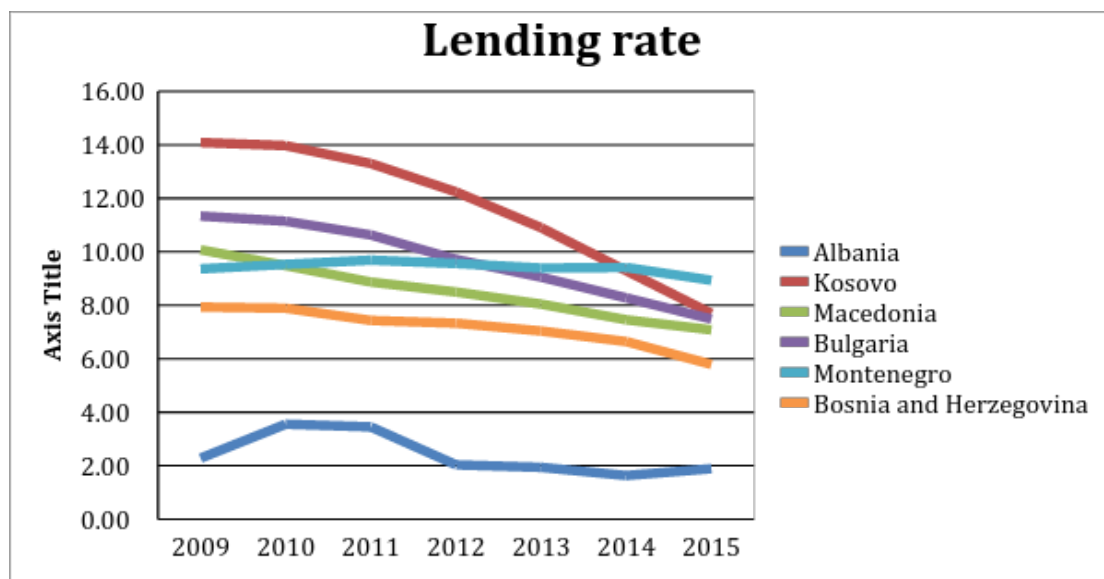
Regarding GDP per capita, Bulgaria has the highest numbers in the Balkan region while Kosova has the lowest one. This can be explained by looking at the economic conditions of both countries. Kosova is a relatively new state and it is still in the transition with unstable political and economical conditions, while Bulgaria is a consolidated state with stable conditions. All the countries tend to fall the same fluctuations over the years but not in the same amount.

Figure 5: Unemployment rate



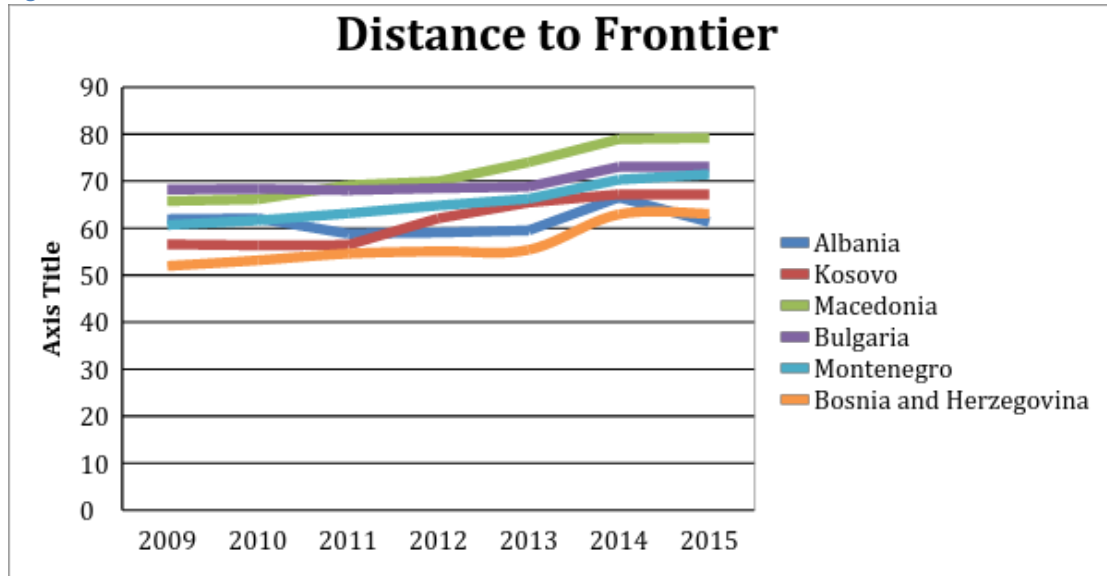
From the graph above we can see that there is no same trend in the unemployment rate between the countries. Kosova has a drastic fall in the unemployment rate from 2009 to 2012, shortly after it became an independent state. However, it still has the highest unemployment rate in the region (accompanied with lowest GDP per capita). On the other hand, Bulgaria (which had the highest GDP per capita) has the lowest unemployment rate in the region. Bosnia and Herzegovina as well as Montenegro seems to have the most stable rate.

Figure 6: Lending interest rate



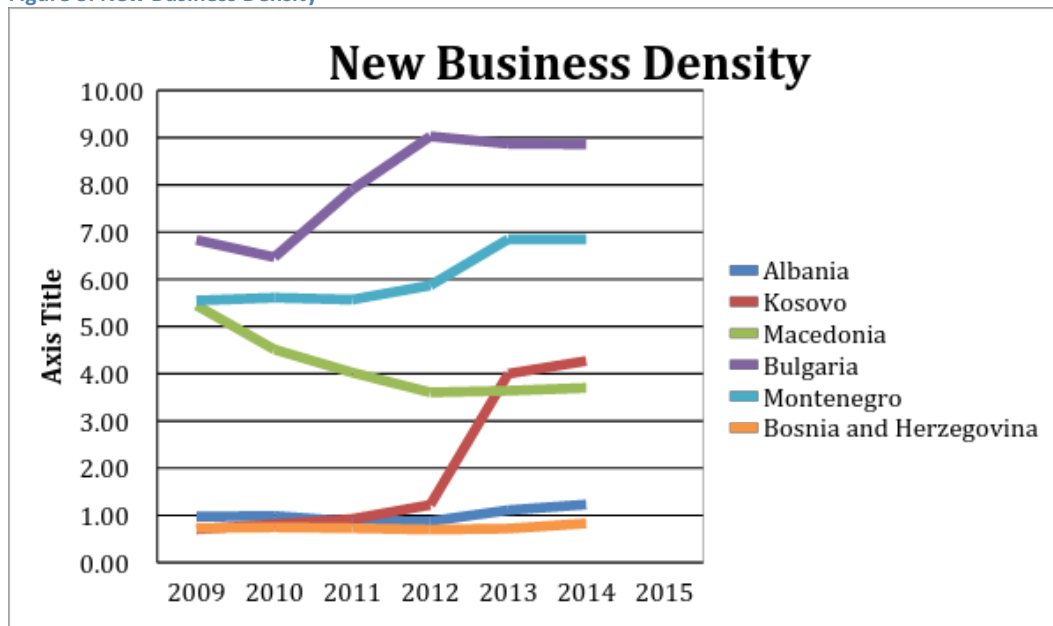
The lending rate in all the countries have a negative trend, apart from Montenegro, which has a stable trend, and Albania, which has fluctuations in lending rate. Kosovo has the highest lending rates for the majority of the years while Albania the lowest.

Figure 7: Distance to Frontier



The distance to frontier variable is almost the same for the countries ranging from 50 to 70. However, Macedonia seems to have made an improvement in the business environment in the past years and it has the highest score in the region as of 2015 while Bosnia and Herzegovina has the lowest throughout the years.

Figure 8: New Business Density



There are significant differences between the countries in the New Business Density variable. Bulgaria has the highest density of new business in the region with a significant difference from Montenegro, which comes second, and of course Bosnia and Herzegovina, which has the lowest density. The countries do not follow any certain pattern in the trend.

Descriptive Statistics

To see how the variables have changed through the years and their distribution descriptive statistics

was examined.

Variable	Mean.	Min.	Max.	Std. Dev.
GDP per Capita	\$ 4669.208	\$ 2135.333	\$ 7553.335	\$ 1569.098
Unemployment	24.08 %	30.2%	5.6%	10.9%
New Business Density	3.86	0.52	9.8	2.8
Lending rate	8.57%	1.63%	14.77%	3.45%
Distance to Frontier	64.74	53.16	79.19	6.39

Results obtained from econometric model

Firstly, a multiple regression model was conducted for all the countries separately. The results are as follows:

Table 3: Albania

Variable	Coefficient	Prob.
Distance to Frontier	0.0253	0.0234
GDP per Capita	1.2001	0.0309
Inflation	-0.6304	0.0223
Lending Rate	-0.3645	0.0135

All the variables chosen in the study have an effect in the ease of doing business in Albania. As expected, inflation and lending rate have a negative effect while distance to frontier and GDP per capita have a positive effect. Also, 82% of variations in new business density are explained by the chosen variables.

Table 4: Kosova

Variable	Coefficient	Prob.
Lending Rate	-0.2079	0.0372
Inflation	-0.0989	0.0074
GDP per Capita	0.0020	0.0316
Distance to Frontier	0.1047	0.8713

Distance to frontier in the case of Kosovo has no impact on the new business density while, as in the case of Albania, lending rate and inflation have a negative effect while GDP per capita a positive but very insignificant one.

Table 5: Macedonia

Variable	Coefficient	Prob.
Lending Rate	-1.5988	0.0151
Inflation	-0.0254	0.0191
GDP per Capita	1.9E-05	0.0687
Distance to Frontier	0.1870	0.0313

Frontier		
----------	--	--

In the case of Macedonia, GDP per capita has no significant impact on the ease of doing business while all the other variables have the same effect as in the previous countries.

Table 6: Bulgaria

Variable	Coefficient	Prob.
Lending Rate	-1.5436	0.0044
Inflation	0.4768	0.1719
GDP per Capita	-0.0004	0.7136
Distance to Frontier	-0.0073	0.9813

When considering Bulgaria the only significant variable seems to be the lending rate, which has a negative effect as expected.

Table 7: Montenegro

Variable	Coefficient	Prob.
Lending Rate	-1.8889	0.0478
Inflation	0.0178	0.0926
GDP per Capita	0.8302	0.0318
Distance to Frontier	0.1307	0.0819

In case of Montenegro lending rate has a strong negative effect in the ease of doing business, while all the other variables have a positive effect. An unexpected result is that in this case inflation has a positive effect while it is expected to have a negative one.

VAR results

	DISTANCE_TO_FRONTIER	GDP_PER_CAPITA	INFLATION	LENDING_RATE	NEW_BUSINESS_DENSITY	UNEMPLOYMENT
DISTANCE_TO_FRONTIER(-1)	0.765075 (0.47420) [1.41341]	-12.38749 (40.7866) [-0.30372]	0.209942 (0.25470) [0.82427]	0.075329 (0.05995) [1.95654]	0.190176 (0.06936) [2.74181]	0.464737 (0.18350) [2.53264]
DISTANCE_TO_FRONTIER(-2)	0.215241 (0.55019) [0.39121]	45.67221 (47.3227) [1.96512]	-0.149215 (0.29551) [-0.50493]	-0.095044 (0.06956) [-1.36643]	-0.132620 (0.08048) [-1.84793]	-0.440154 (0.21291) [-2.06737]
GDP_PER_CAPITA(-1)	0.002511 (0.00178) [1.70979]	0.274443 (0.15321) [1.79132]	-0.001468 (0.00096) [-1.53389]	2.37E-05 (0.00023) [0.10504]	-4.48E-05 (0.00026) [-0.17191]	-0.000309 (0.00069) [-0.44772]
GDP_PER_CAPITA(-2)	-0.001761 (0.00208) [-0.84844]	0.582761 (0.17856) [3.26360]	0.000946 (0.00112) [0.84849]	0.000381 (0.00026) [1.45092]	0.000251 (0.00030) [0.82723]	-7.88E-05 (0.00080) [-0.09810]
INFLATION(-1)	-0.517128 (0.50315) [-1.02777]	-115.3742 (43.2770) [-2.66595]	0.388637 (0.27025) [1.43806]	0.019335 (0.06361) [0.30396]	-0.054272 (0.07360) [-0.73743]	-0.118063 (0.19470) [-0.60637]
INFLATION(-2)	0.480223 (0.49433) [0.97147]	11.61838 (42.5180) [0.27326]	-0.734397 (0.26551) [-2.76598]	-0.065371 (0.06249) [-1.04604]	0.030824 (0.07231) [0.42631]	-0.133655 (0.19129) [-0.69871]
NEW_BUSINESS_DENSITY(-1)	-0.288639 (1.19506) [-0.24153]	-46.62996 (102.789) [-0.45365]	-0.899467 (0.64188) [-1.40130]	-0.128087 (0.15108) [-0.84779]	0.523925 (0.17480) [2.99724]	0.988325 (0.46245) [2.13716]
NEW_BUSINESS_DENSITY(-2)	0.339951 (1.14136) [1.79785]	11.10574 (98.1706) [0.11313]	0.617357 (0.61304) [1.00704]	0.091704 (0.14429) [0.63553]	0.023407 (0.16695) [0.14020]	-1.115180 (0.44167) [-2.52491]
UNEMPLOYMENT(-1)	0.127286 (0.58776) [2.21656]	16.67765 (50.5545) [0.32989]	-0.203681 (0.31570) [-0.64518]	-0.012486 (0.07431) [-0.16803]	-0.136035 (0.08597) [-1.58229]	1.481032 (0.22745) [6.51160]
UNEMPLOYMENT(-2)	0.114523 (0.52188) [1.81944]	-32.47574 (44.8878) [-0.72349]	0.011779 (0.28031) [0.04202]	0.036911 (0.06598) [0.55945]	0.025211 (0.07634) [0.33027]	-0.579269 (0.20195) [-2.86837]

All the significant results of VAR model are highlighted in yellow. In our case, our main focus would be on the new business density. As we can see from the results above the new business density is affected by the first and second lag of distance to frontier, but the first one has a positive effect while the other a negative one.

Causality

A Granger causality test was conducted to see if there is any causal relationship between variables in the Balkan region. From the results we found out that there were some causal linkages between variables. The significant results are as follows:

Table 8: Causality

Hypothesis	Probability
Inflation does not Granger Cause Lending rate	0.0087
GDP per Capita does not Granger Cause Inflation	0.0154
Distance to Frontier does not Granger Cause GDP per Capita	0.0161
GDP per capita does not Granger cause New Business density	0.0185

From the results we can see that a change in inflation would cause a change in the lending rate. This relationship is expected because according to economic theory lending rates change in accordance to inflation changes.

GDP per capita seems to cause inflation. According to economic theory, an increase in inflation means that prices have risen. With an increase in inflation, there is a decline in the purchasing power of money, which reduces consumption and therefore GDP decreases. High inflation can make investments less desirable, since it creates uncertainty for the future and it can also affect the balance of payments because exports become more expensive. As a result, GDP decreases further. So it appears that GDP is negatively related to inflation. However, there are studies indicating that there may also be a positive relationship. The Phillips curve, for example, shows that high inflation is consistent with low rates of unemployment, implying that there is a positive impact on economic growth.

The causality between distance to frontier and GDP per capita is expected. The closer the economy is to the optimum condition, the higher the GDP per capita will be. So, an increase in the points of distance to frontier would cause an increase in GDP per capita.

The last hypothesis tested proves that GDP per capita causes new business density as expected.

Conclusion

This paper aims to study the impact that macroeconomic factors have on the ease of doing business in Balkan countries; Albania, Kosova, Bosnia and Herzegovina, Montenegro, Macedonia and Bulgaria. In order to test the relationship new business density was chosen as the dependent variable while inflation, lending interest rate, unemployment rate, GDP per capita, and distance to frontier were chosen as independent variables. The data were retrieved from the World Bank Database on yearly basis from 2005 to 2015. To prove the research hypothesis a multiple regression was conducted on each country. The results showed that GDP per capita and distance to frontier had a positive effect on the ease of doing business meaning that in times of economic growth, the conditions of starting a new business were facilitated. On the other hand, inflation and lending rates had a negative effect. An increase in prices or in the cost of borrowing would increase the total initial investment. Driven by that increase, the entrepreneurs are not eager in opening new businesses in times of high inflation or lending rates, so the ease of doing business decreases. The new business density has a long run relationship with distance to frontier but the effect changes through the months. Lastly, our study proved that inflation is one of the factors that cause the lending rates to change. GDP per capita causes inflation as well as new business density and it supports the economic theory and the conclusions driven by the Phillips curve. The biggest limitation of the study was the lack of data since the distance to frontier and new business density are two variables that are only available from 2005. Even though this field is little researched, it is very promising for the upcoming researchers.

References

Anderson, J. C., & Narus, J. A. (1991). Partnering as a focused market strategy. *California Management Review*, 33(3), 95-113. Chicago

- Zambujal-Oliveira, J., & Pinheiro-Alves, R. (2010). The Ease of Doing Business Index as a tool for Investment location decisions. *Economic Letters*, 117, 66-70.
- Oliveira, J. Z., & Alves, R. P. (2010). *The Ease of Doing Business Index as a tool for Investment location decisions* (No. 0030). Gabinete de Estratégia e Estudos, Ministério da Economia e da Inovação.
- Fairlie, R. W., & Krashinsky, H. A. (2012). Liquidity constraints, household wealth, and entrepreneurship revisited. *Review of Income and Wealth*, 58(2), 279-306.
- Mazzarol, T., Volery, T., Doss, N., & Thein, V. (1999). Factors influencing small business start-ups: a comparison with previous research. *International Journal of Entrepreneurial Behavior & Research*, 5(2), 48-63.
- Sanfey, P., Milatovic, J., Kresic, A., (January, 2016), How the Western Balkans can catch up?, European Bank for Reconstruction and Development, Working Paper No. 185.
- Pere, E., Hashorva, A., (June, 2013), Business regulation and economic growth in the Western Balkan countries, Eastern Journal of European Studies, Volume 4, Issue 1, pg 10.
- Caliendo, M., & Kritikos, A. S. (2010). Start-ups by the unemployed: characteristics, survival and direct employment effects. *Small Business Economics*, 35(1), 71-92.
- Carroll, R., Holtz-Eakin, D., Rider, M., & Rosen, H. S. (2001). Personal income taxes and the growth of small firms. *Tax policy and the economy*, 15, 121-147.

Mathematical Simulation of an Accident Situation at Intersection

¹MSc. Erjola Cenaj, ²Dr. Raimonda Dervishi, ³Prof. Asoc Shkëlqim Kuka

¹Polytechnic University of Tirana, Faculty of Mathematical Engineering and Physical Engineering, Department of Mathematical Engineering

Tirana, Albania, erjola_cenaj@yahoo.com

²Polytechnic University of Tirana, Faculty of Mathematical Engineering and Physical Engineering, Department of Mathematical Engineering

Tirana, Albania, raimondadervishi@yahoo.com

³Polytechnic University of Tirana, Faculty of Mathematical Engineering and Physical Engineering, Department of Mathematical Engineering

Tirana, Albania, sh.kuka@gmail.com

Abstract

Two basic groups of procedural methods are of practical application for analyzing accident situations, which use mathematical models of system man-vehicle-surroundings and data loggers. In the case of mathematical models, uncertainty in an analysis result mainly from: accuracy of defining the models' parameters and adopted model structure. In the case of the devices recording the motion parameters, adulteration of results may result from measuring errors of the values characterizing the vehicle's motion and the errors resulting from processing of the recorded volumes. This paper focus on the first procedural method: an analysis by means of mathematical models of system man-vehicle-surroundings which are applied on simulation models. We concentrate on the resulting of mathematical simulation of an accident situation at intersection when two vehicles and two pedestrians participate in a crash.

Keywords: Accident investigation, uncertainty analysis, simulation models.

Introduction

In the practice of analyzing accident situations, two basic groups of procedural methods are of practical application, which use mathematical models of system man-vehicle-surroundings and data loggers. In the case of mathematical models, uncertainty in an analysis result mainly from: accuracy of defining the models' parameters and adopted model structure. In the case of the devices recording the motion parameters, adulteration of results may result from measuring errors of the values characterizing the vehicle's motion and the errors resulting from processing of the recorded volumes. We concentrate on the resulting of mathematical simulation of system man-vehicle-surrounding in accident reconstruction. We will examine an accident situation when two vehicles and two pedestrians participate in a crash. The first vehicle moving along the national highway A_1 in the direction of the location A to location B ; A_2 second vehicle moving along the national road in the direction from locality to locality $D-C$ and two pedestrians were crossing the white lines. As a result of the collision of the two vehicles, the collided vehicle hits the two pedestrians as well. Severely damaged in the accident, the driver and passenger of A_1 , also severely damaged remained the two pedestrians who were crossing the white lines.

Statement of the Problem

From the study we will do, the calculation of the speed of movement of vehicles involved in the accident using one of the ways to solve this problem, will be sufficient. The first work to be done is that of the device with a plan as accurate as possible and then draw the locations of the first contact that should be compatible with the tracks left on the road by two means as well as their deformed condition. By these properties it is then possible to evaluate the spins and linear movements under which the two automobiles have gone under in the movement from the first contact location on to the final location.

Spins and displacement (centre of importance) provided as follows:

$$\alpha_1 = 137^\circ = 2,4 \text{ rad} - \text{rotation } A_2 \text{ (running clockwise)}$$

$$\alpha_2 = 107^\circ = 1,9 \text{ rad} - \text{rotation } A_2 \text{ (running anti-clockwise)}$$

$$S_1 = 7.90 \text{ m} - \text{linear displacement of the center of importance of } A_2$$

$$S_2 = 7.40 \text{ m} - \text{linear displacement of the center of importance of } A_1$$

To calculate the velocity, we apply the principle of sustainability of the amount of movement principle that physically adjusts the phenomenon of collision mechanics. Note: m_1 - measures of A_2 with two people on board, m_2 - measures of A_1 with two people on board, v_1 - the velocity of A_2 before the collision, v_2 - the velocity of A_1 before the collision, v'_1 - the velocity of A_2 after the collision, v'_2 - the velocity of A_1 after the collision, θ_1 - the angle formed by the vector v_1 , θ_2 - the angle formed by the vector v_2 , θ'_1 - the angle formed by the vector v'_1 , θ'_2 = the angle formed by the vector v'_2

Analysis of the accident situation at intersection

The principle of sustainability vector form of the quantity of movement is

$$m_1 v_1 + m_2 v_2 = m_1 v'_1 + m_2 v'_2$$

$$v_1 + \frac{m_2}{m_1} v_2 = v'_1 + \frac{m_2}{m_1} v'_2$$

$$v_1 \cos \theta_1 + \frac{m_2}{m_1} v_2 \cos \theta_2 = v'_1 \cos \theta'_1 + \frac{m_2}{m_1} v'_2 \cos \theta'_2 \quad (1)$$

then

$$m_1 = \frac{2.080 + 1.40}{9.81} = \frac{2.220}{9.81} = 226 \text{ kg m}^{-1} \text{ s}^2$$

$$m_2 = \frac{890 + 140}{9.81} = \frac{1.030}{9.81} = 105 \text{ kg m}^{-1} \text{ s}^2$$

$$\frac{m_2}{m_1} = \frac{105}{226} = 0.46$$

$$\theta_1 = 0 \text{ and } \cos 0 = +1,00, \theta_2 = 111^\circ \text{ and } \cos 111^\circ = -0,36$$

$$\theta_1' = 4^\circ \text{ and } \cos 4^\circ = +0,99, \theta_2' = 16^\circ \text{ and } \cos 16^\circ = +0,96$$

$$\text{Now, equation (1) is the same with: } v_1 - 0.17 v_2 = v_1' + 0.46 v_2' \quad (2)$$

We see that after the collision velocities— listed with v_1' and v_2' - can be determined primarily on the theoretical route by considering only linear displacements S_I and S_2 by the formulas:

$$\text{For the second vehicle } A_2, v_{1_0}' = \sqrt{2g f S_1} = \sqrt{2 \times 9.81 \times 0.40 \times 7.90} = 28 \text{ km/h}$$

$$\text{For the first vehicle } A_I, v_{2_0}' = \sqrt{2 g f S_2} = \sqrt{2 \times 9.81 \times 0.40 \times 7.40} = 27 \text{ km/h}$$

Also, for the first and second vehicle (A_2 and A_I) we calculate:

$$v_{*1}' = \sqrt{g f p \alpha_1} = \sqrt{9.81 \times 0.45 \times 2.54 \times 2.4} = \sqrt{27} = 5.2 \text{ m/s} = 5.2 \times 3.6 = 19 \text{ km/h}$$

$$v_{*2}' = \sqrt{g f p \alpha_2} = \sqrt{9.81 \times 0.45 \times 2.39 \times 1.9} = \sqrt{20} = 4.5 \text{ m/s} = 4.5 \times 3.6 = 16 \text{ km/h}$$

Now we obtain the velocity after collision:

For vehicle A_2

$$v_1' = v_{1_0}' + v_{*1}' = 28 + 19 = 47 \text{ km/h} = 47/3.6 = 13.1 \text{ m/s}$$

For vehicle A_I

$$v_2' = v_{2_0}' + v_{*2}' = 27 + 16 = 43 \text{ km/h} = 43/3.6 = 11.9 \text{ m/s}$$

But in fact due to the collision with the small wall of the A_I (to the point that a thicker part collapsed) obtained value of 11.9 m/s must be increased to at least 40% so the velocity will handle the final value of $v_2' = 1.40 \times 11.9 = 16.7 \text{ m/s}$. After the end velocities v_1' dhe v_2' are determined and their values in formula (2) can be substituted

$$v_1 - 0.17 v_2 = v_1' + 0.46 v_2'$$

$$v_1 - 0.17 v_2 = 20.8 \quad (3)$$

Since (3) contains two unknowns and should be placed beside a second equation which is one that is obtained by examining the collision coefficient ε determined by:

$$\varepsilon = \frac{v_2' - v_1'}{v_1 - v_2} = \frac{v_2' \cos \theta_2' - v_1' \cos \theta_1'}{v_1 \cos \theta_1 - v_2 \cos \theta_2}$$

For this case we have:

$$\varepsilon = \frac{v_2' (+0,96) - v_1' (+1,00)}{v_1 (+1,00) - v_2 (-0,36)} = \frac{0.96 v_2' - v_1'}{v_1 + 0.36 v_2}$$

$$\text{Or } v_1 + 0.36 v_2 = \frac{0.96 v_2' - v_1'}{\varepsilon}$$

after substituting the formula of v_1' dhe v_2' we have

$$v_1 + 0,36 v_2 = \frac{2.9}{\varepsilon} \quad (4)$$

$$\text{Combining (3) and (4) we obtain } v_2 = - \frac{20.8 \varepsilon - 2.9}{0.53 \varepsilon}$$

Giving ε a value acceptable to the characteristics of the inelastic collision tends due to large deformations consistent findings on two vehicles, ($\varepsilon \in (0.20, 0.30)$) and have averaged $\varepsilon = 0.25$, then

$$v_2 = |17.4| \text{ m/s} = 17.4 \times 3.6 = 63 \text{ km/h}$$

And the absolute value $v_2 = 17,4 \text{ m/s}$.

By substituting the value found of v_2 in the formula (3) we have:

$$v_1 = 17.8 \text{ m/s} = 17.8 \times 3.6 = 64 \text{ km/h}$$

So basically the calculation brings subsequent valuations of the speeds of two vehicles at the time of the crash:

$$v_1 = 64 \text{ km/h} = 17.8 \text{ m/s} - \text{velocity of } A_2$$

$$v_2 = 63 \text{ km/h} = 17.4 \text{ m/s} - \text{velocity of } A_1$$

V_i movement velocity is achieved by examining the braking distance before contact was:

- vehicle A_2 : $S_1 = 10.80 \text{ metra}$

$$V_1 = \sqrt{v_1^2 + 2 g f S_1} = \sqrt{17.8^2 + 2 \times 9.81 \times 0.80 \times 10.80} = 80 \text{ km/h}$$

$$\text{- vehicle } A_1: S_2 = \frac{3.50 + 6.00}{2} = \frac{9.50}{2} = 4.75 \text{ m}$$

$$V_2 = \sqrt{v_2^2 + 2 g f S_2} = \sqrt{17.4^2 + 2 \times 9.81 \times 0.90 \times 4.75} = 70 \text{ km/h}$$

References

- Brach, R., Guzek, M., & Lozia, Z., Uncertainty of road accident reconstruction computations, 2007.
- Brach R., Brach M. Vehicle accident analysis and reconstruction Methods, SAE International, Warrendale, Pennsylvania USA, 2005
- Cohen, J., Preston, B: Causes and Prevention of Road Accidents, London, 1968
- Highway Code of Albania and articles for its implementation, Tiranë, 1998
- Dervishi. R, Modeling Roadway Crashes Using Statistical Methods And GIS, PhD. Dissertation, Polytechnic University of Tirana, 2016

Metadata, the DNA of Statistical Data

Ertugrel Curumi¹, Ilda Shabani², Olta Kodra³

Institute of Statistics, INSTAT, Albania

¹ecurumi@instat.gov.al, ²ishabani@instat.gov.al, ³okodra@instat.gov.al

Abstract

Do you feel lost in statistical data? So many definitions, variables, populations and lots of grouping code lists make impossible to fully understand and compare data. Metadata, like DNA molecules, are complex dimensional data which contain the assembly instructions for statistical data. This is the reason why metadata's role in statistics is vital. This paper analyses with practical examples how metadata improve the relevance of statistical data and why harmonized metadata make statistics coherent and comparable in cross domain, national or European level. Metadata upgrade the clarity of data in another level according to user oriented perspective. Metadata also increase the awareness of accuracy and quality of data for statistical producers. Metadata mean communication and the production of qualitative metadata requires considerable effort and expertise. Nevertheless, it is a crucial process for the evolution of statistical data from simply data to information and knowledge.

Keywords: Metadata, Statistical Data, Management, Quality, Statistics

Introduction

Nowadays it is becoming more and more necessary to exchange information obtained from statistical data between users and producers. This has pointed out the importance of statistical metadata. Statistical metadata are actually data, which are essential for proper production and usage of statistical data because they describe not only the data itself but also the processes and tools used to produce statistical information. Metadata make data significant. Without metadata, data would not be comprehensive. For example, the number 14.2 is only a number until it is associated with the information that it is the official estimate for unemployment rate in the fourth quarter of 2016 in Albania, for the population aged 15 years and over. But this given information, known as metadata may or may not be enough depending on the intended use of this number. If you are a statistician or researcher and you need in depth information, you may want to know an estimated coefficient of variation or a confidence interval. If you are a politician or policy analyst, definitions used to classify people as employed or unemployed are required to proceed with further political analyses. If you work with surveys and survey methods, you may want to know the questions used and maybe the response rate. And these are just a small part of all metadata that could be available for this number.

Data not accompanied with appropriate metadata are in most cases more harmful than beneficial. There is no producer of statistical information, who would like to take the risk that potential users misinterpret data to fit their purposes in the absence of appropriate metadata. Such misuse cannot be completely avoided, even if there are detailed qualitative metadata, but at least there is an objective information basis to argue from. Metadata compensate distance in time and space between production and usage of data. For example, a user of historical data may want to use data collected when he/she were not even born.

Metadata not only explain form and content of data, but also describe managing facts about data like when were they created and who created them. This way, it is easier to search and locate them. This kind of metadata include also processes before storing data in a database like questions asked, collection methods, processing software, etc.

Considering what was mentioned above, the conclusion is that when producing information, data and metadata should be considered as complementary to one another and they should not be divided and regarded individually.

The goal of this paper is to highlight the importance of metadata in the context of official statistics and the agencies that produce them. Firstly some literature review is done to analyse the concept of metadata. In the next part, the relationship between metadata and quality of statistical data is explained and how qualitative metadata affect in reliability of statistical data. Finally the current situation and future plans regarding the work with metadata in Albanian Institute of Statistics (INSTAT) is presented. Understanding how important and relevant metadata are for both users and producers, will affect in first of all creating metadata and then making them more and more useful by raising their quality.

Literature review

First definition for the concept metadata and other synonyms of it like metadata information or metadatabase were given in Sundgren (1973). The most common definition for metadata is “data about data”, which means that the discussion is about secondary data. Computer and information scientists think of metadata as how data are stored and formatted. They relate the concept metadata with description and type of variables and with the meaning of data which may be formal or not, structured or not. Metadata are often free text descriptions.

Statistical institutions were the first to give the deserved importance to metadata, but that took about two decades and some not successful projects. In the 1980's and the 1990's UN/ECE (the statistical division) arranged some meeting about metainformation systems (METIS). An effective result was a Guideline; Sundgren (1993). EUROSTAT in 1993 organized a workshop on statistical metadata which was attended from a large number of participants. The Compstat conference in 1994 had a session on statistical metadata; Sundgren (1994).

Some years ago, the Open Forum on Metadata members concluded in the following definition about metadata: “Statistical metadata describes or documents statistical data, i.e. microdata, macrodata, or other metadata. Statistical metadata facilitates sharing, querying, and understanding of statistical data over the lifetime of the data.” As it is obvious this is an accurate and clear definition. Nevertheless it should be considered that maybe this definition is not sufficient to suit all users due to different comprehension ability. To have a more precise definition, one must examine the different needs that metadata fulfils in official statistics.

Currently EUROSTAT includes in what is called metadata the following: Euro-SDMX Metadata Structure (ESMS) that is a set of international standards for exchange of statistical information between organisations; International statistical classifications and nomenclatures; EU legal acts and methodological manuals relating to statistics; CODED (EUROSTAT's Concepts and Definitions Database) and other online glossaries relating to survey statistics; Survey

methodologies used at national level to produce EU statistics, quality reports, etc.; Standard cross-domain code lists used in the reference database, recommended for production databases and data transmission. INSTAT is always trying to be more and more coherent with EUROSTAT and that is why the different metadata components mentioned above are being implemented through different IT systems and standards which will be explained further.

Metadata and statistical data quality

What defines statistics with good quality? Quality is a difficult term to be defined and measured in many business processes, but when it comes to statistical data it becomes more challenging. Quality in Official Statistics is a multidimensional concept and plays an important role in statistical production process. ISO standard 9000:2005 defines quality as the "degree to which a set of inherent characteristics fulfils requirements". Therefore, based on this definition, the quality of data should have in focus user needs and how much relevant are statistical data for user needs.

Users seek data of good quality, but they do not have the necessary information to evaluate how "good" the data are. Therefore, it is very important that the metadata accompanying statistical data provide users with necessary material or knowledge to evaluate themselves the quality of those data. From the statistical perspective the main quality dimensions addressed to statistical data are presented below.

First of all what characterizes qualitative statistics is the fact that the data should be relevant for user's needs. *Relevance* in statistics means to identify users and their expectations. To achieve this, a lot of metadata should be used in order to understand the meaning of the data and how they are treated. For example, users of labour market statistics some of which are the Government (e.g. ministries), National Accounts, researchers, enterprises, etc. at most cases need statistics on people in employment or not, people searching for job, hours worked, income from work and benefit, etc. Since labour market covers almost all aspects of people's work and these statistics are available, it can be said that labour market statistics are relevant to user needs.

Another quality dimension is *accuracy*, and that is the degree of closeness of computations or estimates to the exact or true values that the statistics were intended to measure. Accuracy is composed of several statistical indicators such as sampling and non-sampling error, coverage, measurement and processing error, non-response rates. The improvement of metadata on accuracy and precision should be an integral part of the statistics producer's work program. Considering the same example of Labour Market statistics, the unemployment rate in fourth quarter of 2016 is 14.5% (for population aged 15-64 years) and for a 95 % confidence level, it lies within the range 14.4 % to 14.6 %, with a relative standard deviation of 0.5 %. In general this means that the lower the relative standard deviation of an estimate, the higher the accuracy level of that estimate.

Timeliness is the length of time between data being made available and the event they describe. Users need that the statistical information to be up to date and published frequently and on time at pre-established dates. If the metadata system it is managed in a good way, it is easier reducing the time lag between design and implementation by decreasing development time through reuse.

Furthermore, by considering metadata management as an integral part of production process the timeliness and quality of dissemination products can be improved. INSTAT produces monthly, quarterly and yearly statistics. There are also some event publications that are produced once a four year or a ten year period like censuses.

Accessibility refers to the conditions and modalities, by which users can obtain, use and interpret data. As it is mentioned in the introduction, metadata about user access helps searching and locating data. Accessibility includes press releases regarding the data, publications in which they are made available for the public, information if micro-data are disseminated. In INSTAT published data are available through news releases, dedicated publications and online statistical database. Users can submit specific requests for data through the INSTAT website.

Clarity refers to the extent in which statistical data are understandable by users. In order to provide enough clarity, data should be accompanied by sufficient and appropriate metadata and documentation on quality and methodology. Currently INSTAT includes in the end of press releases and publications short explanations related to the definitions of the main concepts and methodological explanations. Additional support information is given to internal users when needed or required.

Coherence reflects the degree to which the data and information from a single statistical programme are brought together with other data and information, and how they are logically connected and completed. Again metadata plays a key role to define and unify the concepts and target population.

Comparability is the extent to which statistics for a given characteristic enable reliable comparisons of values across geography and over time. Comparability can be assured through accurate metadata. In this point it is very important to manage metadata related to definitions, unit of measure, the classification or geographical changes. For instance CPI (Consumer Price Index) data in Albania are comparable between prefectures because the method used for collecting, processing and calculating is the same in all the territory of Albania. All prefectures are covered to the same extent. Also CPI data are comparable with EU CPI statistics considering that the methodology is based on EU regulations. Since CPI data are calculated on monthly basis with the same methodology, they are fully comparable over time.

Metadata and quality are both important and they need to be connected very well with each-other. Metadata should be seen as a tool for enhancing data quality, which for sure is affected by the quality of metadata itself. The quality and completeness of metadata for a dataset, directly affects their search-ability and reuse. In statistics, metadata quality refers to the following components:

The accuracy of metadata - are the characteristics of the resource correctly reflected? Indicating the right title, the right license, the right publisher enables users to discover resources that they need.

The availability of metadata – are metadata accessed now and over time into the future? Making it available for indexing and downloading, and include it in a regular back-up process.

The completeness of metadata – are all relevant characteristics of the resource captured. Identifying necessary fields to be present in a metadata record or the format of the distribution enables filters on those aspects.

The conformance of metadata to accepted standards – is the metadata conforming to a specific metadata standard? The definitions of variables, description of a dataset etc. should be based on a metadata standard.

The consistency of metadata – does the data not contain contradictions? The data should not have multiple and contradictory permission statements for the same piece of data.

The credibility and source of metadata – is the metadata based on reliable sources? It is important to link the reference of the data publisher, in this case the NSI.

The process-ability of the metadata – is the metadata properly machine-readable? It is very important making the metadata of a dataset available in RDF and/or XML, and not as free text.

The timeliness of metadata – is the metadata corresponding to the actual characteristics of the resource and is it published presently enough? Metadata should indicate the last modification date of the resource, thus making sure the metadata is new so that users will see the latest information. Moreover metadata needs to be kept up-to date to the extent possible taking in consideration the available time and budget.

By analyzing all the above components, the conclusion is that the same quality considerations should be applied to both data and metadata. Since metadata provides information on data and resources, the quality of the metadata directly affects the discoverability and reuse of the resources. Moreover, the lifecycle of metadata is longer than the lifecycle of data, metadata are created before the publications of data and they remain even if the data are deleted.

Managing metadata in INSTAT

Albanian Institute of Statistics (INSTAT) has the mission to provide transparent, neutral and timely statistics that help the user to judge on the developments of the transformation processes within the country. Metadata plays an enormous role in achieving this mission. There have been and probably there will be many times in Albania when users do not understand correctly the figures presented to them and therefore they misinterpret statistics converting them sometimes to false information. This is a huge problem especially when these users are the politicians or the media. To minimize the number of times when data are misinterpreted, metadata comes in handy and gains the proper importance.

Users of statistical data have different needs and purposes, which means that their needs (where to look for the intended data, which period does the data correspond to, how they are collected, main definitions of variables and indicators calculated, etc.) for metadata differ. Some of the users of statistical data and therefore potential users of statistical metadata in Albania are:

- The Government and politicians who could use metadata related especially to definitions of indicators used to plan, monitor, and evaluate actions;
- Companies can use metadata to find the appropriate statistics to make business decisions (statistics on wages, prices, etc.) and stay in current with these statistics;
- Organizations could use metadata to make negotiations;
- Researchers, journalists and students may need metadata related to quality of data, metadata on time period that data covers and where to find them, to make their own analyses, to understand, and explain real-world phenomena;
- General public could use metadata that simply and clearly explain statistical indicators, etc.

Another very significant user of statistical metadata in Albania is INSTAT itself which can benefit from them in the following aspects.

Statistical units within the institution should speak the same language to provide consistent statistical information. Yet there are problems related to concepts and terminology, for example the discrepancy among concepts and technical terms that have multiple, conflicting definitions. Metadata can help to solve these types of problems. Currently a system that manages them (MetaPlus) is being used in INSTAT, which will be discussed in more details later on.

Statistical metadata within INSTAT can be used in the production of statistical data as knowledge bases. *Designers of statistical surveys* such as subject matter statisticians, statistical methodologists will need to know about user needs, how similar surveys have been designed in the past and/or by other statistical services in order to conduct new surveys or improve the existing ones. *Input data providers* such as respondents, contact persons, etc. to a statistical survey will be interested to know about the purposes of the survey and about the costs and benefits of his/her participation. In INSTAT, several surveys attach to the questionnaire also a leaflet for the respondents to read so they can understand that their input is very important and it does not harm them.

One of the main priorities in five years strategic plan named Official Statistical Program (PSZ 2017-2021) is creating the data warehouse (DW) in order to integrate administrative data and survey data. Within the context of a statistical office, metadata is known as a necessary input for a successful data warehouse. INSTAT is currently working with MetaPlus system which will be one of the most important sources for the data warehouse, regarding metadata.

To accomplish these statistical metadata needs, INSTAT is working with two different types of metadata:

Structural metadata are used to define the data structures. Variable names, classifications, standard code lists, variable types, data set definitions are part of structural metadata (Harmonizing concepts, classifications and variables, promoting comparability of information). INSTAT is currently working on the implementation of MetaPlus, a standardized system for documentation. MetaPlus it's a metadata driven application developed by Statistic Sweden and has been implemented to INSTAT needs. MetaPlus was implemented in October 2013 and was fully operational in May 2015.

Reference metadata describe the content and quality of statistical data, data collection and processing methods. In order to judge the usefulness of some specific statistical data (macrodata and/or microdata) for his/her purpose, a potential user needs to have information on all the quality dimensions mentioned above like: relevance, accuracy, accessibility, etc. and these are known as

referential metadata. In INSTAT the standard implemented for documenting and disseminating referential metadata is SIMS (Single Integrated Metadata Structure). In this structure, all statistical concepts of the two existing ESS (European Statistical System) report structures (ESMS and ESQRS) have been included and streamlined, by assuring that all concepts appear and are consequently reported upon only once. From the beginning of 2017, INSTAT is publishing referential metadata in ESMS (Euro SDMX Metadata Structure) standard. The decision of working in this standard was made because of two reasons: first because it is very detailed in the information it provides and can respond to a considerable amount of requests for statistical metadata and second because it is a standard which will be used to report to EUROSTAT and will help in making our statistics comparable within EU countries. A desktop application is being developed in INSTAT to manage these metadata. It will be functional by the end of 2017 and it will be used to store, manage, publish and send metadata to EUROSTAT.

These two types of metadata are incorporated in different phases of business processes needed to produce official statistics. The model implemented in INSTAT as a template for process documentation, harmonization of statistical computing infrastructures, and assessment of quality is Generic Statistical Business Process Model (GSBPM). The GSBPM is intended to apply to all activities undertaken in the production of official statistics. It is designed to be independent of the data source, so it can be used for processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources. The GSBPM can also be used for integrating data and metadata standards. In order to have an efficient statistical system, metadata should be collected along all the phases of GSBPM.

Since Albania is a candidate country of European Union, data and metadata in INSTAT are not collected and documented only for internal use but also to be reported to EUROSTAT for certain domains. The standard used to exchange data and metadata in this case is Statistical Data and Metadata Exchange (SDMX). As EUROSTAT has an aim of exchanging all data and metadata in SDMX format with National Statistical Institutes, our challenge is to send all data and metadata to EUROSTAT in SDMX format.

Figure 1 shows the phases of GSBPM and when structural and referential metadata are used, which are managed through MetaPlus system and SIMS (ESMS/ESQRS) standard respectively. However these standards and systems deal with different types of metadata so they should be used with harmony and in combination with each other by reflecting an ideal metadata system.

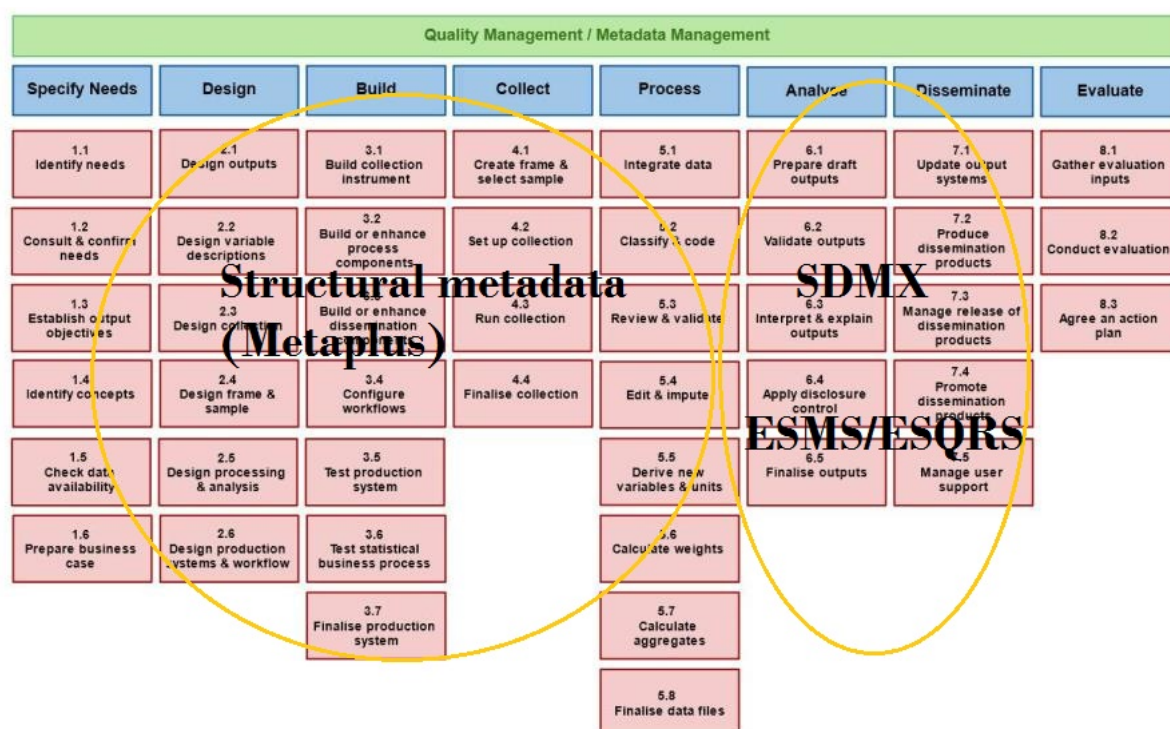


Figure 1: Structural and referential metadata along the processes of GSBPM

Results/Findings

The invocation for metadata grows with an increasing demand on statistics and analyses based on different sets of statistics. Metadata has a core role in the management of data quality and it is an important component of overall management of a statistical institution.

Data quality and metadata quality should have the same degree of importance in a statistical institution, since the dimensions that characterize good quality for both of them are alike. The best approach in managing metadata is to capture them in each phase of the statistical production chain, at the moment when they occur.

The main challenge in a national statistical institution is how to better produce metadata and provide them to appropriate users and how to implement various metadata systems to promote efficiency in production processes and dissemination. INSTAT is dealing with this challenge and is now publishing adequate referential metadata needed to understand and to make use of every aspect of a statistical product. There are still some issues to be solved, and they relate to the prioritization of metadata in our institution. But INSTAT is in the first steps of building structural and referential metadata and in time will learn more and more effective ways to manage and make the most out of metadata.

Conclusions

Everyone who uses and produces statistical information needs statistical metadata to be able to interpret, understand, and analyse them, even if they have not themselves participated in the production processes of these data. Statistical metadata are needed to help a human user to

transform statistical data into information. However metadata studies still remain as a difficult and important issue for national statistical institutions. INSTAT is making a lot of effort on harmonisation of metadata in various processes and standards and connect these standards with each-other. Developing and managing a statistical metadata system is not just an issue for information technologies but a concern that should involve the whole statistical institution.

References

- Bergamasco. S., & Cardacino. A., & Rizzo. F., & Scanu. M., & Vignola. L. (2013). A Strategy on Structural Metadata Management based on SDMX and the GSIM Models. Retrieved from <https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2013/WP4.pdf>
- Hustoft, A. G., & Sæbø, H. V. (2005). Some key concepts related to metadata – to be used in Statistics Norway’s metadata systems. Retrieved from <https://www.ssb.no/a/english/metadata/definitions/concepts.pdf>
- INSTAT. (2017). Quality Reports. Retrieved from <http://www.instat.gov.al/en/about-us/quality-in-statistics/qualityreports.aspx>
- Macedonian State Statistical Office. (2013). Metadata Strategy 2013-2015. Retrieved from http://www.stat.gov.mk/Dokumenti/strategii/SSOMetadata%20strategy_20140516.pdf
- Sundgren. B. (1995). Statistical Metadata. Retrieved from <http://www.scb.se/H/Teori%20och%20metod/R%20and%20D%20Report%201988-2004/RnD-Report-1995-05-yellow.pdf>
- Sundgren. B. (2006). Reality as a Statistical Construction – Helping Users Find Statistics Relevant for Them. Retrieved from <http://ec.europa.eu/eurostat/documents/64157/4374310/08-REALITY-AS-A-STATISTICAL-CONSTRUCTION-SE-2006-EN-Q2006-1.pdf/e63ba6fb-e541-4573-8a4a-ade3f6e3302>
- United Nations Statistical Commission and Economic Commission for Europe of the United Nations (UNECE). (1995). Guidelines for the Modelling of Statistical Data and Metadata. Retrieved from <https://www.unece.org/fileadmin/DAM/stats/publications/metadatamodeling.pdf>

Statistical Indicators as Potential Early Signals of Transitions in Time Series Obtained by a Statistical Model: Geomagnetic Field Case

Klaudio Peqini^{1,2*}, Bejo Duka¹

¹ Faculty of Natural Sciences, University of Tirana, Tirana, Albania, *

klaudio.peqini@fshn.edu.al

² Epoka University, Tirana, Albania

Abstract

Applications of statistical methods in studies of irregular time series of different physical or non – physical quantities are rapidly and constantly increasing. These quantities characterise different phenomena of quite different fields like economy, ecology, climate science, geophysics, etc. The statistical methods provide tools to understand the behaviour of these series and its underneath as well as some limited future prediction. These time series often show transitions from one stable state to another. Statistical tools can provide early signals of such potential transitions. These tools are generally applied to time series from ecological systems, or economy though having specific implications for each system. The time series usually are recovered by the real systems, but they could be obtained from simulating various kinds of numerical models that mimic these systems.

We have used a statistical model known as the “domino model” to simulate long time series of dipolar geomagnetic field. The simulated dipolar field like as the observed geomagnetic field has two stable states and the magnitude of the dipolar field irregularly oscillates between these states. We applied the statistical tools to these time series with the aim to find some early indicators of such transitions, known in geophysics as reversals of the dipolar geomagnetic field. We find some clues and positive results in identifying some prior indicators of future potential transitions.

Key Words: statistical model, time series, geomagnetic field, dipolar geomagnetic reversal, numerical simulation

Introduction

Many complex systems exhibit sharp transitions that occur without any noticeable warning sign (Dakos et al., 2012). These transitions may occur in climate (Lenton et al., 2008), lakes and coral reefs (Scheffer et al., 2001), financial markets (Johannes, 2004) or ecological systems (Dakos et al., 2012). Often in these systems exist one or more bifurcation point where depending on the dynamics of the system may occur transitions to different states or regimes (Strogatz, 1994). Of special interest are generally the critical transition where a given systems jumps from one state to another often distant state. This transition is reflected in drastic changes in real world systems. Often these transitions are not predictable but there seems to be some indicators that show when the system is approaching the transition (Dakos et al., 2012).

There have been extensive studies to find such indicators and many are applied to different real world systems. In table 1 of Dakos et al., 2012 are given the results (together with the respective references) for many indicators starting from statistical parameters like standard deviation, variance, skewness, kurtosis, autocorrelation etc., to more sophisticated models like potential well estimator or autoregressive model of order p . Normally there are used real world data, but when not available, there are used synthetic data produced by numerical models that somehow mimic the real system.

Although there are many studies for many systems, there seems to be no studies done on geomagnetic time series. These time series are constructed from palaeomagnetic measurements or generated by diverse numerical dynamo models. They have an expressed random nature (Shmitt et al., 2001) and this randomness is related to the complicated processes that occur in the liquid outer core of Earth. The geomagnetic field has one main dipolar field and minor more complicated contributions (Jacobs, 1994). In the long history of the geomagnetic field there have occurred drastic transitions of the dipolar field from one polarity to the other. These phenomena are known as reversals happen in an irregular fashion. We will call them in this paper as phase transition where the two phases are related to the normal polarity state and reversed polarity state. By normal polarity we mean the current polarity of the geomagnetic field. A way of forecasting these events would be of much interest in the understanding of the underlying phenomena that govern the time evolution of the geomagnetic field.

Lacking of enough real world geomagnetic data we can make use of the time series produced by numerical models. The time series produced by complex dynamo models are very expensive to compute due to the complexity of the dynamical equations. The use of simpler models, often referred to as statistical models or toy models, generally eliminates the computational obstacles, although there is a loss in the reproducibility of the details of the real geomagnetic field. However the main features are preserved. The “domino” model is one of these toy models and is studied extensively elsewhere (Mori et al., 2013; Duka et al., 2015; Peqini et al., 2015). The results show that this model is adequate to describe not only reversals but also other phenomena like the secular variation (SV) or rate of change of the geomagnetic field. In this paper we analyse the time series produced by the “domino” model from the perspective of statistical analysis of time series in general and are interested in identifying possible early warning signs for the phase transitions that in the case of the geomagnetic field are the reversals.

The structure of the paper is as follows: in section 2 is shortly described the “domino” model. In section 3 are introduced the statistical tools we use in this study and is explained the method that is applied for each of them. The results are reported in section 4 followed by the discussions and conclusions and some recommendations for further studies in section 5.

Model

The model we study in this paper is known as the “domino” model. It inspired from several physical assumptions that are embodied in weakly driven dynamos (Davidson, 2013). The dynamo mechanism takes place in the liquid outer core of the Earth where in the weakly driven dynamo regime is organised in so called columnar convection cells. These structures are identified in numerical simulations (Kageyama and Sato, 1997). Further details on the subject and in the physical assumptions can be found in the papers Duka et al., 2015, Peqini et al., 2015 and the literature referenced there. This paper is focused on analysis of the time series generated by the “domino” model hence we will not mention any further the physical arguments underlying this model.

The “domino” model consists of a circular alignment of N macro-spins that interact pair-wise. These macro-spins are embedded in a uniformly rotating with unit angular velocity $\mathbf{\Omega} = (0,1)$ along the rotational axis (fig. 1). Each of the macro-spins has unit length and each of them can be described dynamically through the angle each of them forms with the rotational axis. Consequently an individual macro-spin is $\mathbf{S}_i = (\sin \theta_i, \cos \theta_i)$.

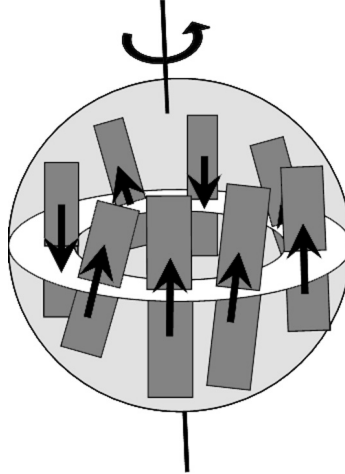


Fig. 1 Sketch of the “domino” model.

Two essential ingredients of the model are: the tendency of each macro-spin to align with the rotational axis (dynamics in rotating systems) and the spin-spin interactions of the macro-spins. In the latter case is adopted an Ising-like interaction. Mathematically these interactions compose the potential energy which reads

$$P(t) = \gamma \sum_{i=1}^N (\mathbf{\Omega} \cdot \mathbf{S}_i)^2 + \lambda \sum_{i=1}^N (\mathbf{S}_i \cdot \mathbf{S}_{i+1}), \quad (1)$$

where γ and λ characterise numerically the orientation tendency and the spin-spin interaction respectively. Also when $i = N$, $\mathbf{S}_{i+1} = \mathbf{S}_i$. The second term in equation 1 actually considers an interaction among neighbouring macro-spins. Other fashions may be used like mean field

“domino” model (Duka et al., 2015; Pegini et al., 2015) but in this paper we focus in this type of interaction only.

The kinetic energy of the system of macro-spins is

$$K(t) = \frac{1}{2} \sum_{i=1}^N \dot{\theta}_i^2. \quad (2)$$

A Lagrangian $L = K(t) - P(t)$ is written, where more explicitly we write

$$L = \frac{1}{2} \sum_{i=1}^N \dot{\theta}_i^2 - \gamma \sum_{i=1}^N (\mathbf{\Omega} \cdot \mathbf{S}_i)^2 - \lambda \sum_{i=1}^N (\mathbf{S}_i \cdot \mathbf{S}_{i+1}). \quad (3)$$

Then a Langevin – type equation is set up as follows:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}_i} \right) = \frac{\partial L}{\partial \theta_i} - \kappa \dot{\theta}_i + \frac{\varepsilon \chi_i}{\sqrt{\tau}}, \quad (4)$$

where the term $-\kappa \dot{\theta}_i$ describes energy dissipation and κ is the parameter of dissipation; $\varepsilon \chi_i / \sqrt{\tau}$ is the random force acting on each spin whose strength is characterised by the parameter ε . χ_i is a Gaussian-distributed random number with zero mean and unit variance and is associated to each spin. The value of this random variable is updated each correlation time τ . Substituting 3 into 4 yields the system of differential equations of the model

$$\ddot{\theta}_i - 2\gamma \cos \theta_i \sin \theta_i + \lambda [\cos \theta_i (\sin \theta_{i-1} + \sin \theta_{i+1}) - \sin \theta_i (\cos \theta_{i-1} + \cos \theta_{i+1})] + \kappa \dot{\theta}_i - \frac{\varepsilon \chi_i}{\sqrt{\tau}} = 0. \quad (5)$$

The model has a system of second order ordinary differential equations (ODE) and periodic boundary conditions are considered, where $i = 1, 2, \dots, N$, $\theta_0 = \theta_N, \theta_{N+1} = \theta_1$.

The system of N second order ODEs is transformed into a system of $2N$ first order ODEs. Then we integrated them with a 4th order Runge – Kutta algorithm in the MATLAB platform using an internal function (ode45). The initial values of the angles θ_i are random and uniformly distributed in $(0, 2\pi)$, whilst the macro-spins are considered to be initially at rest. In summary, the “domino” model has 6 independent parameters: $N, \gamma, \lambda, \kappa, \varepsilon$ and τ . With except of the last parameter which is a technical detail of the numerical simulations, the remaining parameters characterise numerically the physical processes modelled with the respective terms.

The output of each simulation is the cumulative orientation of all macro – spins or also known as axial orientation. We will refer to it as magnetisation and actually it is the scaled non – dimensional quantity corresponding to the dipolar geomagnetic moment. It is calculated by

$$M = \frac{1}{N} \sum_{i=1}^N (\mathbf{\Omega} \cdot \mathbf{S}_i) = \frac{1}{N} \sum_{i=1}^N \cos \theta_i(t). \quad (6)$$

Statistical parameters

The standard deviation and variance are two of the statistical parameters we analyse in the present study and are shown to give positive results in forecasting transitions in time series (Carpenter and Brock, 2006). The former parameter gives an estimate of the degree by which the members of a group differ from the mean values of the group. The latter parameter gives a measure of how far apart are a set of (random) numbers from their mean. From the definitions one can conclude that these parameters are very similar in what they estimate. Hence there are expected similar results when applied to time series.

Each of the statistical parameters is calculated for moving windows of a given width. There are chosen several values that span from 17 to 1000 time units. The smallest window width corresponds to the smallest recorded interval between two consecutive reversals in the history of the Earth that is approximately 20,000 years (Gubbins, 1999). The standard deviation and variance are calculated for each window and there are obtained time series of these quantities. These time series are plotted in the same frame as the magnetisation time series for comparison.

The last statistical analysis is performed by calculating the power spectral density (PSD) of portions of the time series. In principle a time series can be decomposed in several Fourier terms with specific frequencies. The PSD shows the amount of energy per unit of time that is comprised in individual frequencies. Rare or low frequency phenomena have small PSD values. Phase transitions are very rare compared to other phenomena and their presence should be accompanied with an increase in low frequency edge of PSD plot. The PSD results to be successful in forecasting transitions in time series (Kleinen et al., 2003). The PSD is calculated for a standard window with width of 1000 time units. We made this choice after several trials with the aim to obtain better plots. The window is then moved by a step of 500 time units and several PSDs are obtained.

Results

In fig. 2 is shown a typical time series generated by the “domino model”. The full run comprises 300,000 time units, while in the figure are shown the first 30,000 time units. The random nature of the results of the model is evident. Also it is clear the existence of two symmetric states and that the magnitude of the magnetisation oscillates irregularly between them. The symmetry arises from the symmetry in the dynamical equations written for the liquid outer core (Rüdiger and Hollerbach, 2004). In support of this picture is also a statistical model of the amplitude of the dipolar moment known or bistable geodynamo model (Schmitt et al., 2001).

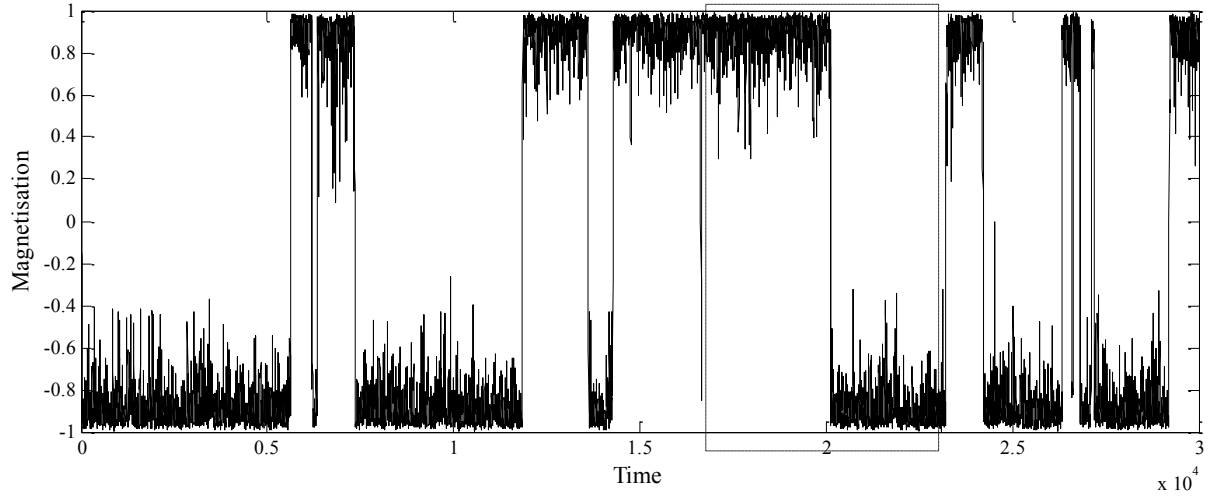


Fig. 2 Magnetisation for a typical run where are shown the first 30,000 time units. Inside the dashed box is the portion studied in this paper.

The effect of the values of the coefficients is substantial to the dynamics of the magnetisation and consequently to the time series produced by the model. The parameters space of the “domino” model is studied by several authors (Mori et al., 2013; Duka et al., 2015; Peqini et al., 2015) although it requires additional work to have a more complete view of different regimes. However we will study the time series produce by the model for parameters values: $N = 8$, $\gamma = -1$, $\lambda = -2$, $\kappa = 0.1$, $\varepsilon = 0.4$ and $\tau = 0.01$ (equal to the time step). Actually we show in fig. 1 a hundredth of the full time steps such that the time series would be easier to study. The fact that the correlation time is equal to the time step, this means that the Gaussian distributed number is updated after each time step.

In our study we focus on one piece of the full time series. The fact that the statistical nature of the time series does not change in all its extension guaranties that the analysis with the statistical parameters is independent on where do we apply these parameters. From this point of view there is no particular reason why we chose the specific piece of the time series. This specific piece of the time series is inside the dashed box and is shown in fig. 2. In this portion is only one reversal, i.e. one phase transition from the normal polarity to the reversed polarity. Also there are present some minor jumps in the time series that indicate of failed attempts to reverse polarity.

The first statistical parameters we study are standard deviation and variance. We study them together because they are similar parameters and in fig. 3 – 6 are plotted on the same frame as the magnetisation time series. In the different panels, the window width is different. The minimal width is 17 time units. This corresponds to the minimal length of a chron (the time interval between two consecutive reversals). Then we enlarge the window width to 50, 100 and 500 time units. In every panel is clear that the when approaching the reversal the standard deviation and the variance increase. However in fig. 3 we see that these parameters are not very effective to predict in advance the eminent reversal, i.e. phase transition. Enlargement of the window width should make this parameters better in forecasting the future transitions. This can be seen in figures 4 – 6 where the

standard deviation and variance become better in forecasting the eminent reversal when the window width increases.

When the window width is small normally the higher frequency variations like secular variation (SV) (Pegini et al., 2015) are captured. The sharp changes in magnetisation are accompanied with sharp changes in standard deviation and variance. When the window width is increased then the smaller variations are somehow obscured. This becomes evident in fig. 6 where these parameters are almost constant except for the middle section of the time series where is located the reversal.

The last statistical analysis focuses on the power spectrum density (PSD) estimation. This analysis is very helpful to determine the weight of each of the processes with different proper characteristic frequencies. The reversals represent the phenomenon with the largest period, i.e. smallest frequency, in the spectrum of temporal variations of the geomagnetic field. As such, the inclusion of a reversal in the time series is reflected in the increase of the weight of low frequencies. In fig. 7 are shown 11 panels where for each panel is calculated the PSD for a time series of 1000 time units. Then the window slides to cover the whole time series. The increase in low frequencies can be seen in panel f where the reversal is included.

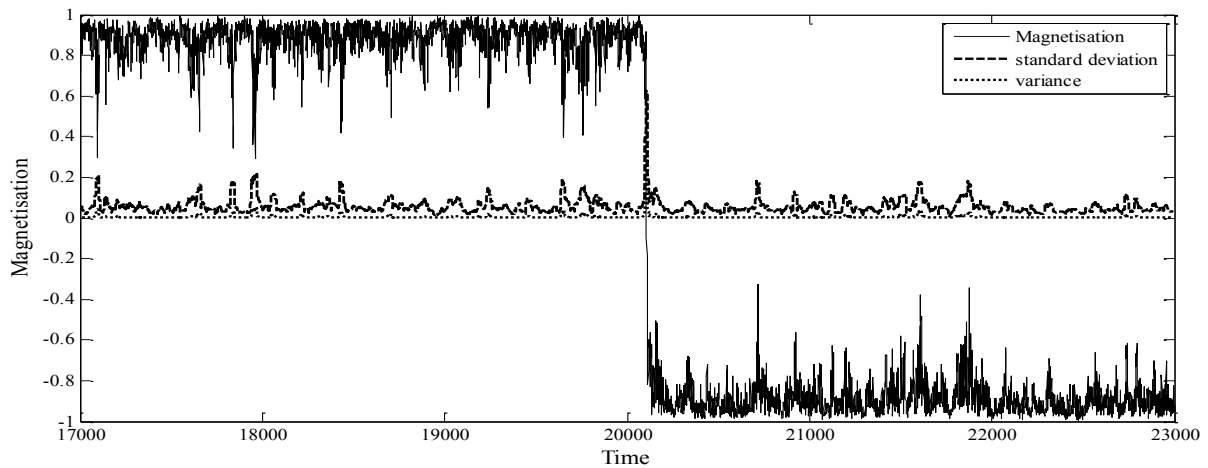


Fig. 3 Standard deviation (dashed) and variance (dotted) for the interval with width 17 time units.

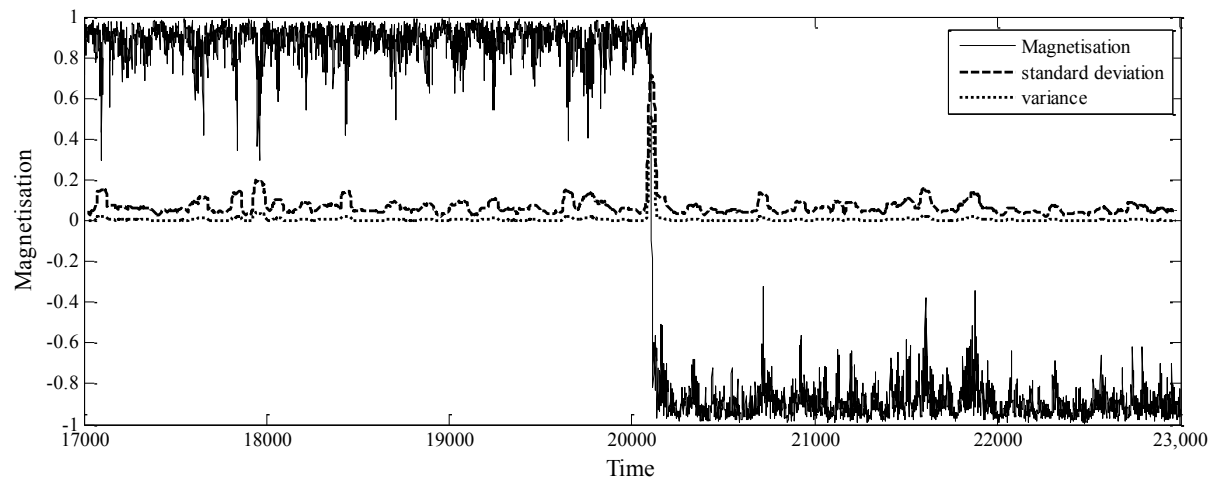


Fig. 4 Standard deviation (dashed) and variance (dotted) for the interval with width 50 time units.

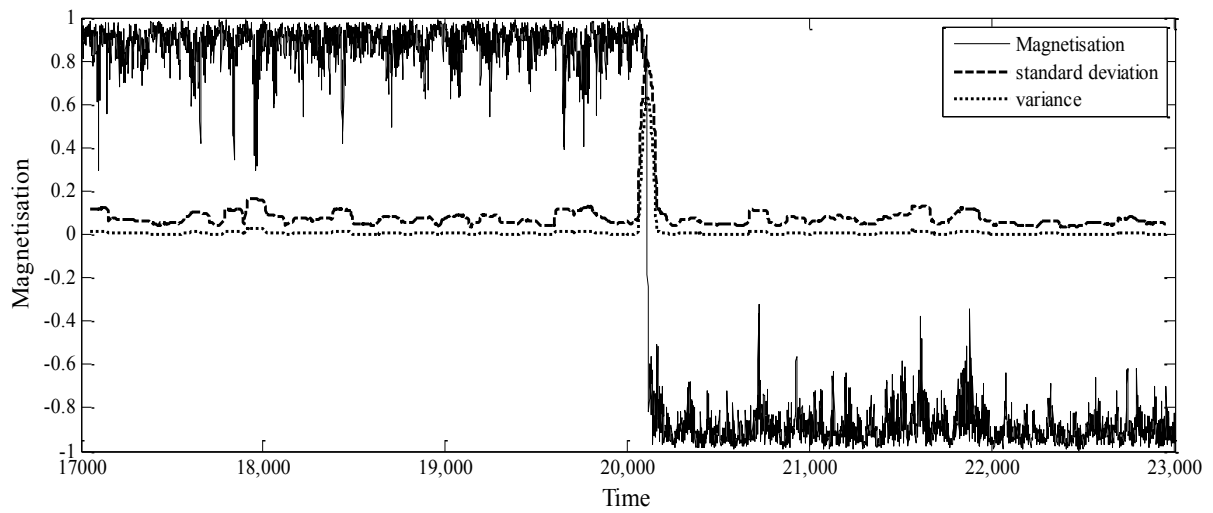


Fig. 5 Standard deviation (dashed) and variance (dotted) for the interval with width 100 time units.

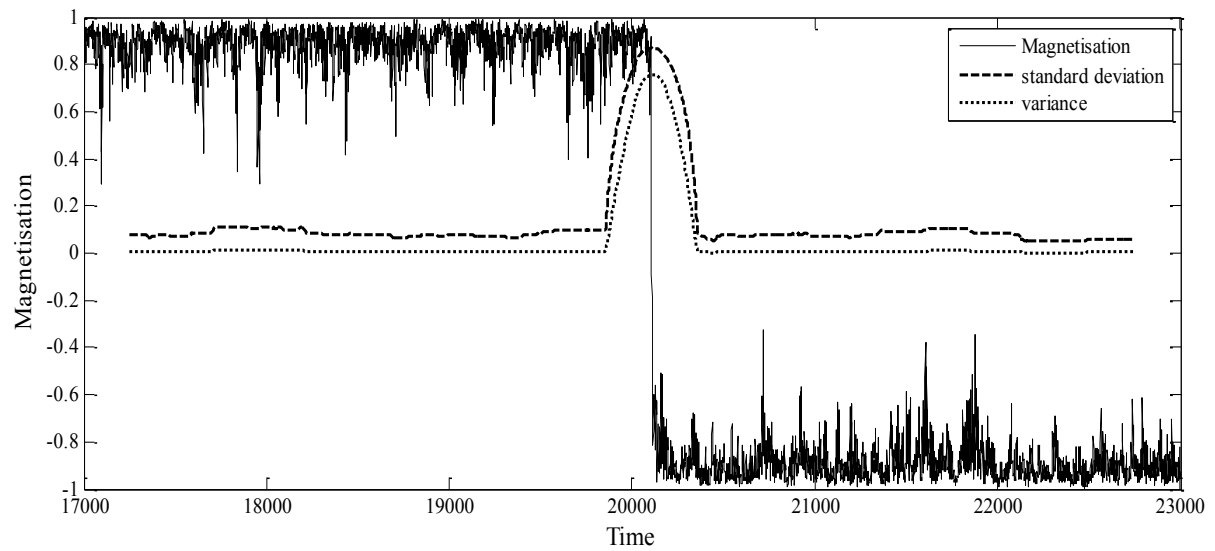


Fig. 6 Standard deviation (dashed) and variance (dotted) for the interval with width 500 time units.

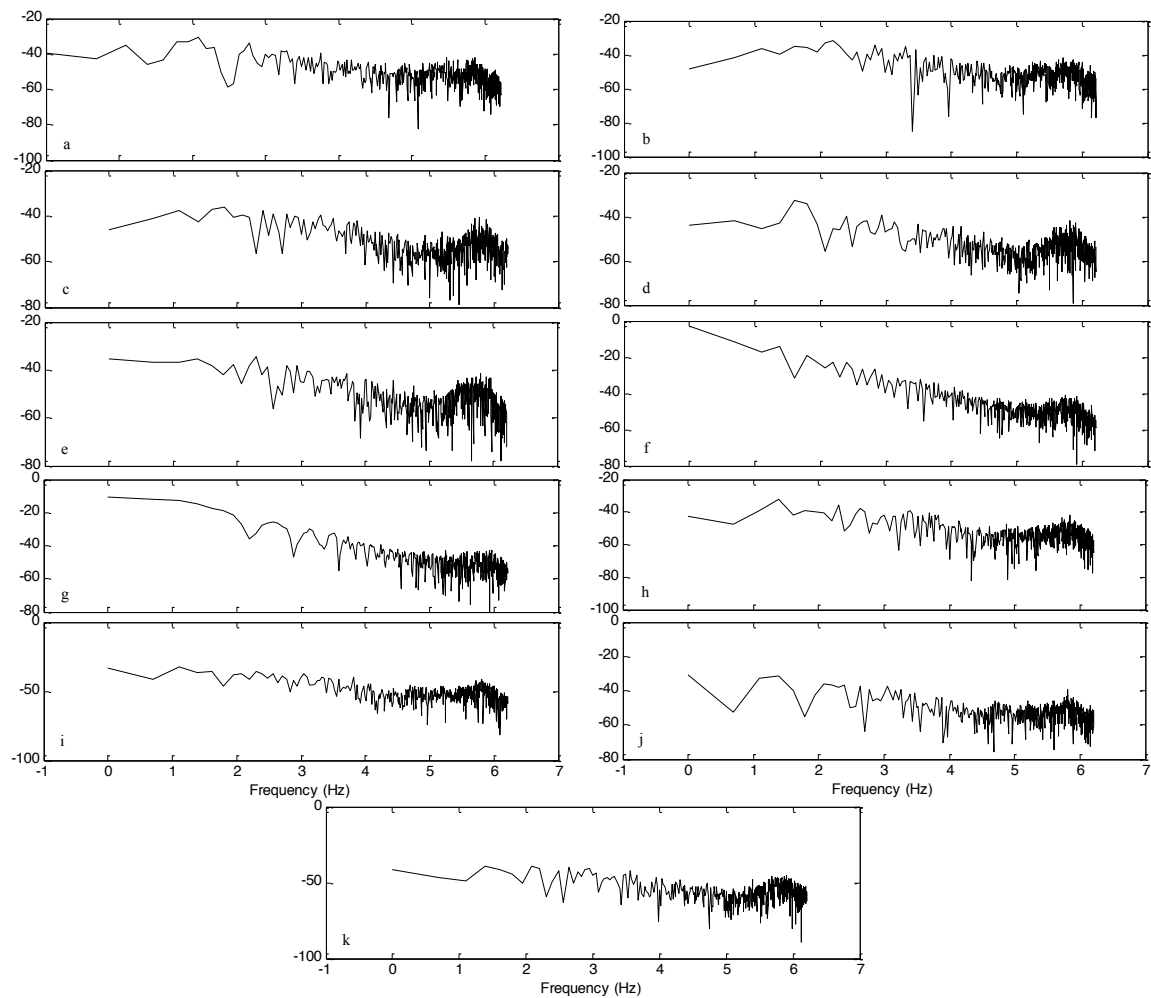


Fig. 7 Power Spectrum Density (PSD) with a window width of 1000 time units. Panels: a) 1 – 1000, b) 500 – 1500, c) 1000 – 2000, d) 1500 – 2500, e) 2000 – 3000, f) 2500 – 3500 (reversal included), g) 3000 – 4000, h) 3500 – 4500, i) 4000 – 5000, j) 4500 – 5500, k) 5000 – 6000. Frequency is in the linear scale, while the vertical axis is in log scale.

There can be seen that the weight of low frequencies diminishes when going further away from the reversal in both directions. However in panels e and g one can see that the weight of low frequencies in the PSDs is significantly larger than the in the cases of the windows at the edges of the time series. In these last cases the low frequency phenomenon of the reversal is not included. This sequence of PSDs shows that when approaching a reversal there is a reflection in frequency contribution which signs the future phase transition.

Discussions and Conclusions

In this paper we analysed a time series produced by a stochastic model of the geomagnetic field from the perspective statistical analysis to find possible early warning signs that could serve as possible methods to forecast future phase transitions also known as reversal of polarity of the dipolar field.

As approaching the reversal, the values of magnetisation tend to become more distant to each other and the difference arises closer to the transition. This yields in increased values of standard deviation and variance. The apogee is achieved when the midpoint of the reversal coincides with the midpoint of the window; hence the maximum is located in the middle of the reversal interval. When leaving the reversal interval, the subsequent values of magnetisation become more like to each other and consequently the values of standard deviation and variance decrease considerably. However the window width substantially affects the ability of a given statistical parameter to potentially predict the phase transition. The optimal window width has to be chosen by trial – and – error taking into consideration that an exceedingly small width does not allow an early forecast, while on the other end seems that many details are obscured.

The approach to an eminent transition is reflected to the PSD of the time series. The reversal lies in the low frequency end of the spectrum of time variations of the geomagnetic field. The inclusion of the reversal in a time series increases the weight of low frequencies. This increase is reflected in the PSD with a substantial negative increase in the slope for the section low – to – middle frequencies. When getting away from the reversal interval the slope in that specific part of the PSD decreases considerably as can be seen in panels a – d and h – k in fig. 7.

In this study we have confined ourselves in reversals only. It would be very interesting to perform the analysis of standard deviation, variance and PSD when there are included excursions. This phenomenon is considered to be a set of two consecutive reversals that are bracketed by an aborted polarity interval (Valet et al., 2008). Also excursions have a higher frequency compared to reversals and may be reflected in principle by an increase of the weight in the low frequency – middle frequency band border. However this hypothesis remains to be studied in further studies.

There can be identified a problem with the methods described in this paper. In fig. 3, 4 can be seen that not always an increase in standard deviation of variance leads to a transition. This fact complicates the analysis in real world time series where temporary increases of the statistical parameters in use may erroneously lead to the idea that a transition will soon occur. The application of the methods described here in real world time series that do not contain phase transitions is extremely crucial in validating or discarding the methods themselves.

It would be of much interest to test other statistical parameters for geomagnetic time series to eventually find possible early warning signs. Possible candidates are: autocorrelation at $-lag - 1$, detrended fluctuation analysis indicator, skewness or kurtosis. Very useful would be the use of models like potential analysis (potential well estimator) to study the time series from early warning signs perspective. These tests remain to be done in forthcoming studies.

The “domino” model is considered to be a toy model and actually mimics the geomagnetic dipolar field. However the analysis of real geomagnetic time series constructed from palaeomagnetic measurements would naturally be the next study to be done. The statistical parameters give encouraging results in time series that have a phase transition like the reversal. Very fruitful is the analysis of time series generated by full dynamo models that are considered to be very close to real geomagnetic field time series. These studies may lead to a refinement of the present methods or to the invention of new methods.

References

- Carpenter, S. R., Brock, W. A. (2006). Rising variance: a leading indicator of ecological transition. *Ecol Lett* 9: 311–318.
- Dakos, V., Carpenter, S. R., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H., Scheffer, M. (2012). Methods for Detecting Early Warnings of Critical Transitions in Time Series Illustrated Using Simulated Ecological Data. *PLoS ONE* 7(7): e41010. doi:10.1371/journal.pone.0041010
- Davidson, P. A. (2013). *Turbulence in Rotating, Stratified and Electrically Conducting Fluids*. New York, NY: Cambridge University Press.
- Duka, B., Pegini, K., De Santis, A., and Pavon-Carrasco, F. J. (2015). Using “domino” model to study the secular variation of the geomagnetic dipolar moment. *Phys. Earth. Planet. Inter.* 242, 9–23. doi:10.1016/j.pepi.2015. 03.001
- Gubbins, D. (1999). The distinction between geomagnetic excursions and reversals. *Geophys. J. Int.* 137, F1–F3. doi:10.1046/j.1365-246x.1999. 00810.x
- Jacobs, J. A. (1994). *Reversals of the Earth’s Magnetic Field*, 2nd Edn. New York, NY: Cambridge University Press.
- Johannes, M. (2004). The statistical and economic role of jumps in continuous time interest rate models. *J. Finance* 59: 227–260.
- Kageyama, A., Sato, T. (1997). Velocity and magnetic field structures in a magnetohydrodynamic dynamo. *Phys. Plasmas* 4, 1569–1575.

- Kleinen, T., Held, H., Petschel-Held, G. (2003). The potential role of spectral properties in detecting thresholds in the Earth system: application to the thermohaline circulation. *Ocean Dynamics* 53: 53–63.
- Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., et al. (2008). Tipping elements in the Earth's climate system. *Proc Nat Acad Sci USA* 105: 1786–1793.
- Mori, N., Schmitt, D., Wicht, J., Ferriz-Mas, A., Mouri, H., Nakamichi, A., Morikawa, M. (2013). Domino model for geomagnetic field reversals. *Phys. Rev. E* 87:012108. doi:10.1103/physreve.87.012108
- Peqini K., Duka B., De Santis A. (2015). Insights into pre-reversal paleosecular variation from stochastic models. *Front. Earth Sci.* 3:52. doi: 10.3389/feart.2015.00052
- Rüdiger, G., and Hollerbach, R. (2004). *The Magnetic Universe: Geophysical and Astrophysical Dynamo Theory*. New York, NY: Wiley-VCH.
- Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature* 413: 591–596.
- Schmitt, D., Ossendrijver, M. A. J. H., and Hoyng, P. (2001). Magnetic field reversals and secular variation in a bistable geodynamo model. *Phys. Earth Planet. Inter.* 125, 119–124. doi:10.1016/S0031-9201(01)00237-0
- Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering (Studies in Nonlinearity)*. Reading, MA: Perseus Books.
- Valet, J. P., Plenier, G., and Herrero-Bervera, E. (2008). Geomagnetic excursions reflect an aborted polarity state. *Earth Planet. Sci. Lett.* 274,472–478. doi: 10.1016/j.epsl.2008.07.056

The Importance of Inference Bayesian in Telecommunication Industry-Albanian Market-

Eralda Caushi

Università Cattolica "Nostra Signora del Buon Consiglio" , Albania
eralda_caushi@hotmail.it

Abstract

One of the most important steps in Telco industry is to understand the behavior of the customers, encourage them in spending more and then predicting their future by preventing their attrition. The churn might be voluntary in cases they want to leave the network they actually are using, or involuntary churn in case of unpaid bills. This paper intends to describe the role of Bayesian Statistics in the area of Telco, in order to identify an appropriate model that have at least a minimum of pre-defined propensity of churning the network they are actually using. The data used for this work are downloaded from the site GSMA Intelligence Markets. The market of reference is the telecommunication market in Albania. Main variables have been identified conducting a survey that identifies the reasons why a customer leaves the network. The findings from this study are helpful for telecommunications companies in order to minimize the customer attrition, and therefor to optimize their customer retention.

Keywords-*Churn, Bayesian inference, Bayesian Regression, Maximum Likelihood Estimation, Probability distribution, Prediction, GSMA Intelligence, Albanian Market*

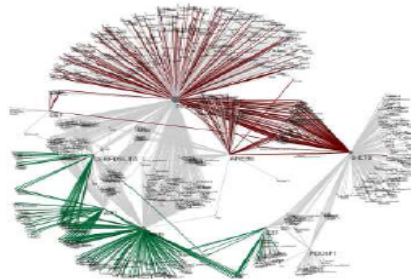
JEL Classification-*L96, C11, F63*

Introduction

The Bayesian approach might have a very important role in Telco industry due to computational reasons. Epistemological reasons and pragmatic reasons are also considerable reasons that guide us in our study. As explained by Brunero Liseo in “Introduzione alla statistica Bayesiana”¹¹, from an epistemological point of view the reasons for using this method are based on a simple and direct inductive reasoning method, according to the information available on a certain set of phenomena, in a certain moment of life that wants to calculate the probability of future events or, more generally of events for which it is not known whether they are verified or not. Bayesian logic is consistent with very logical basis and free of risk counterexamples, always waiting for innovation when it is used the method of induction, and it is necessary to produce statements of probabilistic nature of events that we do not know if it will happen or less. Pragmatic reasons are related to the need of taken in consideration the extra-experimental information of the problem that need to be solve. That refers to the Bayesian setting. In telecommunication, for example, when assessing the probability that a customer might leave the network due to the reduction of some particular offer those that are the a-priori probabilities (extra-experimental info) and are nothing else but the information on the specific offer we need to include in our problem solving. Also very useful in this sector is to have the information at a level of disaggregation sufficiently high. This need goes under the name of “small area estimation” that refers to the difficulty of producing information for

¹¹Introduzione alla statistica Bayesiana”¹¹ –Dispensa di Brunero Liseo Settembre 2008

areas of which we do not have access to the sample. So, estimating the possibility to churn for a single customer that belongs to a sample of a company for which we do not have data, might be possible using the Bayesian method. So, an intrinsic characteristic of the Bayesian method is precisely that of being able to assume, in a simple and natural, different levels of association between the units of the sample, allowing the phenomenon of “borrowing strength” which allows the production of estimates sufficiently stable for those areas with no sample data. The Bayesian Method gives the possibility to integrate, using Bayes theorem, all the information generated by the statistical experiment with the “a priori” data. Monte Carlo methods, based or not on the properties of Markov chains, gives the possibility to generate a sample of whatever dimensions, independent identically distributed by the distribution a posteriori of the parameters we are interested in. That’s why in a very large contest the Bayesian approach permits the flexibility of a model which is very difficult to be achieved through classical methods. The following figure shows



clearly the concept of borrowing concept¹².

In high dimensions, as we are considering telco data, the potential gain is large. A-priori knowledge should make this gain even bigger.

Bayes Theorem

In probability theory and statistics, Bayes' theorem describes the probability of an event, based on conditions that might be related to the event. Bayes' theorem is named after Rev. Thomas Bayes (1701–1761), who first showed how to use new evidence to update beliefs. It was further developed by Pierre-Simon Laplace, who first published the modern formulation in his publication: “Théorie analytique des probabilités” 1812.

Bayes' theorem is stated mathematically as the following equation:

$$P(F|E) = \frac{P(F)P(E|F)}{P(E)} \quad (1.1)$$

Where F and E are [events](#)

P (F) and P (E) are the [probabilities](#) of F and E without regard to each other.

P (F | E), a [conditional probability](#), is the probability of F given that E is true.

P (E | F), is the probability of E given that F is true.

When applied, the probabilities involved in Bayes' theorem may have different interpretations.

¹² Geert Geeven, 2010

Frequentist interpretation

Let's do a simple example in order to explain what Bayes Theorem means using [frequentist interpretation](#). We want to study the probability that measures a proportion of Telco customers. For example, suppose an experiment is performed many times. $P(F)$ is the proportion of customers which have a Vodafone SIM card, and $P(E)$ is the proportion of customers which have another SIM card which is not registered in the Vodafone network. $P(E|F)$ is the proportion of customers which do not have a Vodafone SIM card out of those with Vodafone SIM card, and $P(F|E)$ the proportion of customers with Vodafone SIM card out of those with SIM cards that are not part of Vodafone Network.

Bayesian interpretation

In the [Bayesian interpretation](#), probability measures a degree of belief. Bayes' theorem then links the degree of belief in a proposition before and after accounting for evidence. For example, suppose in the Albanian market there are only two Telco companies. It is believed with 50% certainly that a customer is twice more likely to have a Vodafone SIM card rather than a TELEKOM SIM card. If the number of customer we are considering in our observation is higher than the degree of belief may rise, fall or remain the same depending on the results that have been generated.

For proposition A and evidence B,

$P(F)$, the prior, is the initial degree of belief in F.

$P(F|E)$, the posterior, is the degree of belief having accounted for E.

The quotient $P(E|F)/P(E)$ represents the support E provides for F.

A-Priori Probabilities and Likelihood

Bayes theorem: Let $F_1, F_2, \dots, F_k$ be a set of mutually exclusive events that together form the sample space θ . Let E be any event from the same sample space, such that $P(E) > 0$. Then,

$$P(F_j | E) = \frac{P(F_j \cap E)}{P(F_1 \cap E) + P(F_2 \cap E) + \dots + P(F_k \cap E)}$$

$$P(F_1 \cap E) + P(F_2 \cap E) + P(F_k \cap E)$$

Note: Invoking the fact that $P(F_j \cap E) = P(F_j) P(E|F_j)$, Bayes' theorem can also be expressed as:

$$P(F_j | E) = \frac{P(F_j) P(E|F_j)}{P(F_1 \cap E) + P(F_2 \cap E) + \dots + P(F_k \cap E)}$$

The formula 1.1 refers to the posteriori probabilities of the F event, once it's been realized the E event. The denominator refers to a simple normalization factor. Meanwhile as the nominator, we have two quantities: $P(F_j)$ is the a-priori probability of the F_j event, and the second one $P(E|F_j)$ is the likelihood of the F_j event, which also refers to the probability of the event E, when it's known

the F_j event. Let's consider now the ratio between the two posteriori probabilities $P(F_h | E)$ and $P(F_j | E)$:

$$\frac{P(F_h | E)}{P(F_j | E)} = \frac{P(F_h) P(E | F_h)}{P(F_j) P(E | F_j)}$$

Posterior Probability $\propto$ Likelihood * A – Priori Probability

The normalization factor is canceled, by having so the simplified ratio between the final probabilities. The ratio between the two a – priori probabilities might be considered as a new one a-priori probability called a-priori ratio. The ratio between the two likelihood portions is called Bayes Factor and it is usually indicated with the letter B. In case of testing two hypotheses F_h with regard of any other F_j hypothesis, there will be taken in consideration the facts of events E and not bases in the other events such as F_j . In case of Bayes Factor ($B=1$), then both hypothesis have same evidences.

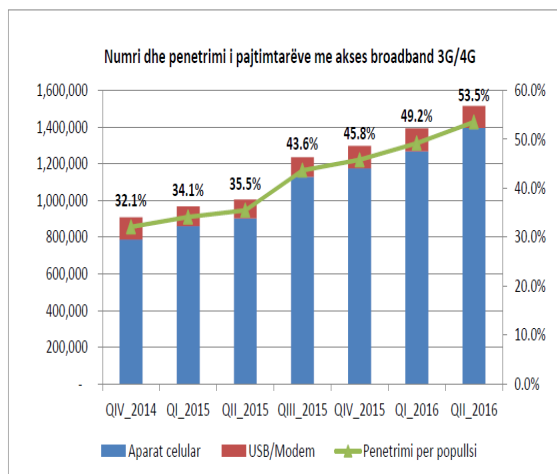
Albanian Market

Albania is home to four wireless operators: UK-backed Vodafone Albania, ex Albania Mobile Communications (AMC) and now Telekom, Eagle Mobile – a former subsidiary of fixed line incumbent Albtelecom, now merged with the telco – and the newest entrant to the market PLUS Communications which has 100% Albanian shareholders. The segment has long been dominated by TELEKOM and Vodafone, though the arrivals of Eagle and PLUS in the market in 2008 and 2010 respectively paved the way for an intensification of competition in an already near-saturated market. Population penetration rates have been excess of 100% since 2009, passing the 150% milestone two years later and around 190% by end-September 2013, with more than six million registered subscribers at that date. Sector regulator the Electronic and Postal Communications Authority (AKEP) noted at the end of 2012, however that the number of active subscribers was significantly lower and that only around 3.5 million of the 5.62 million registered users at end-2012 had been active in the preceding three months, equating to a population penetration of 125%. Pressure from the two smaller operators has slowed Vodafone's subscriber growth whilst eroding TELEKOM's user base, allowing Vodafone to overtake the former frontrunner in the first three months of 2012. Finally, despite offering cheaper tariffs than its larger rivals, PLUS has struggled to compete against its three larger rivals even due to late acquisition of 3G license. In the mobile data market, first mover Vodafone leads the segment in terms of subscribers, but has been surpassed technologically by the newcomers due to exclusivity of 3G license for about one year. By the end of the first year, Vodafone had signed up 228,249 3G users. TELEKOM was awarded the second concession in October 2011, paying EUR15.1 million for the rights, and launched its own service in January 2012. At the end of the first year of full competition, TELEKOM had signed up 250,715 3G subscribers compared to the 408,998 served by Vodafone, according to AKEP's figures. AKEP has faced harsh criticism from wireless operators for its process of allocating licenses. The allocation of a single concession would allow only the operator with the deepest pockets the rights to offer 3G, rather than the operator that would provide the best service. Despite their complaints, AKEP forged ahead, offering the first two licenses as individual concessions. As predicted, the three firms backed by larger international telecoms groups (Vodafone, Cosmote/OTE and Calik Holdings, German Telekom) secured the first

concessions. Vodafone Albania and Telekom have owned the contract and now 4G is in use by the customers. Eagle Mobile and Plus had access to the frequencies in a later moment.

Telco Overview

Telecommunication is the transmission of signs, signals, messages, writings, images and sounds of any nature by wire, radio, optical or other electromagnetic systems. Telecommunication occurs when the exchange of information between communication participants includes the use of technology. It is transmitted either electrically over physical media, such as cables, or via electromagnetic radiation. The term is often used in its plural form, telecommunications, because it involves many different technologies. Mobile telephony is the provision of telephone services to phones which may move around freely rather than stay fixed in one location. Mobile phones connect to a terrestrial cellular network of base stations (cell sites), whereas satellite phones connect to orbiting satellites. Mobile telephone services in Albania are offered by four companies: Telecom (ex-Albanian Mobile Communications), Vodafone Albania, Eagle Mobile and Plus Communications. The focus of this paper is on voice and GPRS data. The growing demand and consumption of data services is leading the scientist community in analyzing and finding new models that better fit to those data. Mobile telephony is highly used by consumers and enterprises as well. As traffic increases, customer and network data increases as well. Furthermore, traffic carried on telco networks is becoming increasingly rich, with voice being supplanted by data services such as streaming audio and video. In Albania telco networks, GPRS now accounts for as much as 53.5 % of total customers¹³.



LTE deployments and associated marketing campaigns have boosted the consumption of data services. This increased data traffic comes from a variety of sources (smartphones, TVs, tablets, and laptops) and multiple channels (social media, web chat, email, and voice calls). Machine-to-machine (M2M) communications have also increased the data service usage as customers use the network to control devices in places like the home or their car. Meanwhile, the telco's companies also have operational data such as billing, network, location data, and call detail records, presented in a structured format, which is typically contained in SQL databases. Semi structured and unstructured data such as call logs, social media messages, text messages, emails, customer

¹³ http://www.akep.al/images/stories/AKEP/statistika/2016/raport_QII_2016.pdf

feedback documents, system logs, and sensor data is also present. The rate of data growth will exceed the capacity of Telco's existing data warehouses. To support internal decision-making processes, these new and varying datasets need to be ingested by storage platforms and processed using analytic tools to gain insights. The insights will inform how to drive increased ARPU, predict and reduce customer churn, deliver improved customer experience through personalized services, and limit the operational costs associated with network management (through network design, planning, and optimization).

Cost implications

Borrowing strength phenomena and is largely applied to the clinical data Cost has had an impact on the slow adoption of new big data technologies by Telco industries. The telecomssector has invested heavily in enterprise data warehouse (EDW) infrastructure and was an earlyadopter of analytics. Such legacy investment decisions will continue to have an impact, as are the operators in fast growing and highly-competitive market like Albanian market.

Lack of relevant skills required to exploit big data analytics

There is a significant skills shortage to support the maximum utilization of big data across all industries. Data scientists are one example; these scientists use advanced statistics and machine learning to create programs that analyze large volumes of data in big data platforms such as OBIEE. Although the development of big data analytics platforms could reduce the requirement for these highly specialized skillsets, these platforms need to merge big data analytics expertise with the telco domain knowledge. Vendors providing big data analytics platforms should have the telco's business context in mind when developing these programs to ensure that the insights derived from analytic activities such as data correlation are meaningful to the telco business. Furthermore, big data analytics solutions need to be developed for ease of use by business users. They should be designed to simplify the daily activities of non-technical business users who are dependent on these platforms to solve their business challenges. Data scientists and telco IT practitioners will need to have a clear understanding of the telco's business objectives and identify how big data as a tool can support the realization of these goals.

References

S.M. Ross, Calcolo delle Probabilità. Apogeo, 2013

S.I. Resnick, A Probability Path. Birkhauser, 1999

Luigi Pace , Alessandra Salvan, Introduzione alla statistica - II - Inferenza, verosimiglianza, modelli. Padova: Cedam, 2001

Adelchi Azzalini, Inferenza statistica: una presentazione basata sul concetto di verosimiglianza Milano: Springer-Verlag Italia, 2001

Brunero LiseoIntroduzione alla statistica Bayesiana –Dispensa di Settembre 2008

<http://www.instat.gov.al>

<http://www.akep.al/>

Time scale regression analysis of oil and interest rate on the exchange rate: A case study for the Czech Republic

Lukáš Frýd

University of Economics, Prague

Czech Republic

Lukas.fryd@gmail.com

Abstract

This paper studies the impact of the Pribor and Oil return rate on the CZK/USD exchange rate return in the different time scale. The time scales were got from the Maximum Overlap Discrete Wavelet Transformation. Thanks to this method we were able to analyse the data generation process in different time horizons. We applied the regression on the wavelet series coefficients which represents the different time scale. The most important results were found for the series from the low-pass filter, which represents the slow movement in the time series. Both regression parameters were significant with a negative sign on this time scale, in comparison with non-transformed data when Pribor did not have a significant impact on CZK/EUR. This conclusion confirms the usefulness of the wavelet transformation for the macroeconomics analysis.

Keywords: wavelet, oil price, time scale regression, multiresolution analysis.

JEL Classification[1]:C22, C46, C51

Introduction

The main object of Czech national bank is price stability. For this purpose, CNB uses inflation targeting. One of the most important factors in price determination and macroeconomics condition is oil price. Because the settlement currency in the oil market is US dollar, the main channel of oil shock transmission is through the Exchange rates. So we can decompose the oil price in Czech crown into the two component:

$$p_{CZK} = e_{USD} + p_{USD} \quad (1)$$

where p_{CZK} is the logarithmic oil price in foreign currency, e_{USD} is logarithmic foreign currency per unit of USD and p_{USD} is the logarithmic price of oil in USD. Following Krugman (1983) the rise in the oil price has negative impact on the balance of payment for oil importing countries such as Czech republic. Zhou in 1995 found that oil price fluctuation is one of the most important shock in exchange rate movement. The similar conclusion bring Camarero and Tamarit (2002) Huang and Guo (2007) and Lizardo and Mollick (2010).

The correlation between exchange rate returns and oil return analysed Cifarelli and Paladino (2010) and Reboredo (2012). Cifarelli and Paladino (2010) utilized multivariate generalized autoregressive conditional heteroskedasticity model when they found along-term negative correlation between these time series. On the other hand, Reboredo (2012) utilize copula function and found a weak correlation with some substantially rises. The possible puzzle in these conclusions could rise from the strong assumption about homogenous market participants. On the

markets operate different agents with different utility functions. This heterogeneity is connected with different investment time horizons. For example, Reboredo and Rivera-Castro (2013) analyzed the correlation between oil return and seven currency returns through wavelet correlation. They found that for the pre-crisis period doesn't exist significant correlation on any scales. However, after global crisis period, they found negative correlation value for all scales. The similar pattern found Fryd (2017) when utilize wavelet coherency methodology on the Czech crown and oil return series. So it's obvious that it is crucial to distinguished the investment horizons. However, the question is if it is sufficient to estimate only the correlation between exchange rate returns and oil return. For example, MacDonald (1998) divide the real exchange rate determinants into the two groups. One of them is created by real interest rate and the second one contains fundamental factors when one of them is oil price. From this purpose, we will analyze the impact of interest rate return (Pribor) and oil return (Oil) on the CZE/USD return on the different time scales. Ramsey (2011) used the same methodology for the estimation of Phillips curve in the US for different time scales.

In this article, we utilize the wavelet transformation for analyzing the following data generation process:

$$\Delta CZK/EUR_t = \alpha_0 + \alpha_1 \Delta Pribor_t + \alpha_2 \Delta WTI_t + \epsilon_t \quad (2)$$

We show that the significance of $\alpha_{0,1,2}$ is changing with time scales. Following Ramsey (2011) we use only basic equation with contemporaneous variables, in order to focus on the usefulness of the wavelet transformation for the analysis of economic relationships.

The paper is organized as follows. We begin with methodology description presenting wavelet analysis. A data description and an empirical analysis of oil return, CZK/USD dependency and Czech money market rate Pribor follow. The last part is devoted to the discussion and conclusion.

Methodology

The principle of wavelet transformation is in the filtration of the original time series with a wavelet function. The wavelet transform uses a basis function that is dilated or compressed and shifted along the time series. The transformation's output provides a time-frequency representation where the information is associated with specific time scales and locations in time. There are two basic wavelet functions, mother wavelet ψ and father wavelet ϕ . The mother wavelet is defined as:

$$\psi_{j,s} = 2^{-\frac{j}{2}} \psi\left(\frac{t - 2^j k}{2^j}\right) \quad (3)$$

Moreover, father wavelet is defined as:

$$\phi_{J,k} = 2^{-\frac{J}{2}} \phi\left(\frac{t - 2^J k}{2^J}\right) \quad (4)$$

Where $j = 1, \dots, J$ is the scaling parameter and k is the translation (or shift) parameter.

The coefficients from wavelet transformation are obtained by projecting the wavelet $\psi(\cdot)\phi(\cdot)$ onto time series $x(t)$. We distinguished three wavelet transformation, the Discrete Wavelet Transformation, Continuous Wavelet Transformation and Maximum Overlap Discrete Wavelet Transform (MODWT). Following Ramsey (2011) we use MODWT which allow us to compute the wavelet series coefficients at all scales given by:

$$W_{j,k} \equiv \sum_k^N \psi_{j,k} x(t) \quad (5)$$

$$V_{J,k} \equiv \sum_k^N \phi_{J,k} x(t) \quad (6)$$

Where $W_{j,k}$ represent the wavelet coefficient at level j . and $V_{J,k}$ represent scaling coefficient at level J . Sometimes we call the mother wavelet as high-pass filter because capture the high movement in the time series.

For example, $j=1$ is connected with the passband $\frac{1}{4} < f < \frac{1}{2}$, for $j=2$ $\frac{1}{8} < f < \frac{1}{4}$ etc. On the other hand the father wavelet is low-pass filter and captures the slow movement in the time series such a long trend. The passband for father wavelet is $0 < f < \frac{1}{2^{J+1}}$.

Thanks to the wavelet transformation we can decompose the original time series $x(t)$ to the component which captures the different time scales:

$$x(t) \approx V_J + W_J + W_{J-1} + \dots + W_j + \dots + W_1 \quad (7)$$

We use this multiresolution decomposition and separate the information from each time series at each scale. Then we apply the following regressions:

$$\Delta CZK/EUR[W_j]_t = \beta_{j0} + \beta_{j1} \Delta Pribor[W_j]_t + \beta_{j2} \Delta WTI[W_j]_t + \epsilon_t \quad (8)$$

$$\Delta CZK/EUR[V_J]_t = \alpha_{j0} + \alpha_{j1} \Delta Pribor[V_J]_t + \alpha_{j2} \Delta WTI[V_J]_t + \epsilon_t \quad (9)$$

Data

For our analysis, we use WTI daily price, CZK/USD daily price and Pribor rate. The sample period spans from 06/01/1993 until 30/09/2009. From the reason of data nonstationary we computed crude oil price exchange rate returns and Pribor (Prague InterBank Offered Rate) change on a continuous compounding basis as the difference between the log of the current price and that of the one-period lagged price. The return series are stationary processes. In the following analysis,

the variables will be in the log-difference. Following Ramsey (2011) we use the Daubechies least asymmetric (LA8) wavelet of length $L = 8$ (Daubechies, 1992) with $J=4$.

Empirical part

In the first, we estimate the equation (2). The results are in the table (1). From the reason of significant autocorrelation, we used HAC estimator for standard errors. We can see that Pribor variable doesn't have as significant influence on the exchange rate. On the other hand, the WTI return has as significant impact on the Exchange rate return. This estimation will be the benchmark for the next outputs.

Table 1. Results from estimation model from equation (2).

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.46E-05	3.44E-05	-1.874719	0.0609
PRIBOR	0.002321	0.004518	0.513661	0.6075
WTI	-0.015010	0.005677	-2.643798	0.0082
R-squared	0.004999	Mean dependent var		-6.95E-05
Adjusted R-squared	0.004306	S.D. dependent var		0.002242
S.E. of regression	0.002237	Akaike info criterion		-9.366222
Sum squared resid	0.014368	Schwarz criterion		-9.359997
Log likelihood	13462.26	Hannan-Quinn criter.		-9.363978
F-statistic	7.212237	Durbin-Watson stat		2.324270
Prob(F-statistic)	0.000751	Wald F-statistic		3.694493
Prob(Wald F-statistic)	0.024978			

The next estimation is from the equation (8) where $j=1$. The W_1 represents the frequency range bandpass $\frac{1}{4} < f < \frac{1}{2}$. This range corresponds to the most fast movement in the time series.

Again we used HAC with results in Table(2). For this model, we cannot see the statistically significant impact on both variables. The p-value for a robust version of Wald test is 0.41.

Table 2. Results from estimation model from equation 8 for scale $j=1$.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.47E-21	7.82E-06	-1.88E-16	1.0000
PRIBOR	-0.003726	0.007669	-0.485830	0.6271
WTI	-0.010315	0.008277	-1.246204	0.2128
R-squared	0.002208	Mean dependent var		4.38E-21
Adjusted R-squared	0.001513	S.D. dependent var		0.001628
S.E. of regression	0.001626	Akaike info criterion		-10.00382
Sum squared resid	0.007595	Schwarz criterion		-9.997593
Log likelihood	14378.49	Hannan-Quinn criter.		-10.00157
F-statistic	3.176414	Durbin-Watson stat		3.430482
Prob(F-statistic)	0.041882	Wald F-statistic		0.881667
Prob(Wald F-statistic)	0.414204			

The next model represents the frequency bandpass $\frac{1}{8} < f < \frac{1}{4}$. The estimation output with HAC standart errors is in the table (3). We can see similar conclusions as in the table (1). WTI return has significant impact on the exchange rate return. The $\beta_{2,2}$ from equation (8) is significant for $\alpha = 0.05$.

Table 3. Results from estimation model from equation 8 for scale j=2.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.61E-20	7.70E-06	-2.10E-15	1.0000
PRIBOR	0.014040	0.015192	0.924223	0.3554
WTI	-0.019441	0.007816	-2.487336	0.0129
R-squared	0.011411	Mean dependent var	-3.26E-20	
Adjusted R-squared	0.010722	S.D. dependent var	0.000867	
S.E. of regression	0.000862	Akaike info criterion	-11.27334	
Sum squared resid	0.002134	Schwarz criterion	-11.26711	
Log likelihood	16202.79	Hannan-Quinn criter.	-11.27109	
F-statistic	16.56933	Durbin-Watson stat	1.295281	
Prob(F-statistic)	0.000000	Wald F-statistic	3.513159	
Prob(Wald F-statistic)	0.029931			

For frequency bandpass $\frac{1}{16} < f < \frac{1}{8}$ or equivalently $j=3$ we have estimation output in table (4). For this time scale we can see significant parameter $\beta_{3,1}$ for $\alpha = 0.01$. On the other hand the WTI return does not have significant impact on the exchange rate return.

Table 4. Results from estimation model from equation 8 for scale j=3.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.57E-21	8.56E-06	3.01E-16	1.0000
PRIBOR	0.028873	0.010916	2.644997	0.0082
WTI	-0.007419	0.007798	-0.951402	0.3415
R-squared	0.019080	Mean dependent var	1.89E-21	
Adjusted R-squared	0.018397	S.D. dependent var	0.000529	
S.E. of regression	0.000524	Akaike info criterion	-12.27008	
Sum squared resid	0.000788	Schwarz criterion	-12.26386	
Log likelihood	17635.10	Hannan-Quinn criter.	-12.26784	
F-statistic	27.92212	Durbin-Watson stat	0.355361	
Prob(F-statistic)	0.000000	Wald F-statistic	4.062918	
Prob(Wald F-statistic)	0.017298			

The last coefficient series from the mother wavelet filter represents the bandpass $\frac{1}{32} < f < \frac{1}{16}$. The estimation output is displayed in table (5). In this situation we can not reject the null hypothesis for t-test. The p-value from robust Wald test is 0.098. This value is too close to the 0.1 and so we do not reject the joint hypothesis.

Table 5. Results from estimation model from equation 8 for scale j=4.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.85E-21	1.47E-05	1.94E-16	1.0000
PRIBOR	-0.010966	0.010202	-1.074911	0.2825
WTI	-0.011528	0.007038	-1.637954	0.1015
R-squared	0.007399	Mean dependent var		5.70E-21
Adjusted R-squared	0.006708	S.D. dependent var		0.000364
S.E. of regression	0.000363	Akaike info criterion		-13.00539
Sum squared resid	0.000378	Schwarz criterion		-12.99916
Log likelihood	18691.74	Hannan-Quinn criter.		-13.00314
F-statistic	10.70083	Durbin-Watson stat		0.094478
Prob(F-statistic)	0.000023	Wald F-statistic		2.321921
Prob(Wald F-statistic)	0.098269			

The last transformation with the father filter represents the long-term behaviour of the time series for $\text{bandpass} 0 < f < \frac{1}{32}$. The estimation results are shown in the table (6). This result is very interesting because we can see significant $\alpha_{4,1}$ from the equation (9) for $\alpha = 0.05$ and $\alpha_{4,2}$ for $\alpha = 0.01$.

Table 6. Results from estimation model from equation 9 for scale J=4.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.15E-05	2.18E-05	-2.824740	0.0048
PRIBOR	-0.021582	0.010894	-1.981050	0.0477
WTI	-0.039664	0.011805	-3.360004	0.0008
R-squared	0.056811	Mean dependent var		-6.95E-05
Adjusted R-squared	0.056154	S.D. dependent var		0.000411
S.E. of regression	0.000399	Akaike info criterion		-12.81314
Sum squared resid	0.000458	Schwarz criterion		-12.80691
Log likelihood	18415.48	Hannan-Quinn criter.		-12.81089
F-statistic	86.46463	Durbin-Watson stat		0.012560
Prob(F-statistic)	0.000000	Wald F-statistic		6.891651
Prob(Wald F-statistic)	0.001033			

Conclusion

In this article, we analysed the dependency of the CZK/USD exchange rate on the Pribor and oil price respectively returns. We used wavelet methodology for the time series decomposition to the different time scale. Concretely the Daubechies least asymmetric (LA8) wavelet of length $L = 8$ was used with multi-resolution level $J=4$. We found a significant impact of WTI on the CZK/USD for bandpass $\frac{1}{8} < f < \frac{1}{4}$ and $0 < f < \frac{1}{32}$. The Pribor has a significant impact on the exchange rate for bandpass $\frac{1}{16} < f < \frac{1}{8}$ and $0 < f < \frac{1}{32}$. The both variables were significant only for bandpass $0 < f < \frac{1}{32}$ with negative signs. This conclusion is very interesting because it suggests that the Pribor and Oil influence the CZK/USD in the longer time horizon. If we compare this result with classical estimation for non-transformed data, we get a better view on the data generation process. This result could be important for monetary policy.

Acknowledgements

The work on this paper was supported by the grant IGS F4/73/2016 of the University of Economics, Prague.

Bibliography

- Camarero, M. and Tamarit, C. (2002). Oil prices and Spanish competitiveness. *Journal of Policy Modeling*, 24(6), pp. 591–605. doi: 10.1016/s0161-8938(02)00128-x.
- Cifarelli, G. and Paladino, G. (2010). Oil price dynamics and speculation. *Energy Economics*, 32(2), pp. 363–372. doi: 10.1016/j.eneco.2009.08.014.
- Fryd, L. (2017). A wavelet transformation approach to crude oil price and CZK/USD exchange rate dependence. *International Institute of Social and Economic Sciences*, pp.17-25, Israel, Tel Aviv, March 2017. ISSN 2336-6044
- Gallegati, M. (2012). A wavelet-based approach to test for financial market contagion. *Computational Statistics & Data Analysis*, 56(11), pp. 3491–3497. doi: 10.1016/j.csda.2010.11.003.
- Gençay, R., Selçuk, F., Whitcher, B.J., Gencay, R., Selcuk, F. and Whitcher, on J. (2001). *An introduction to wavelets and other filtering methods in finance and economics*. San Diego, CA: Elsevier Science.
- Huang, Y. and Guo, F. (2007). The role of oil price shocks on china's real exchange rate. *China Economic Review*, 18(4), pp. 403–416. doi: 10.1016/j.chieco.2006.02.003.
- Krugman, P. (1983). Oil and the dollar. In: Bhandari, J.S., Putnam, B.H. (Eds.), *Economic Interdependence and Flexible Exchange Rates*. Cambridge University Press, Cambridge
- Lizardo, R.A. and Mollick, A.V. (2010). Oil price fluctuations and U.S. Dollar exchange rates. *Energy Economics*, 32(2), pp. 399–408. doi: 10.1016/j.eneco.2009.10.005.
- Percival, D.B. and Walden, A.T. (2000). *Wavelet methods for time series analysis*. Cambridge: Cambridge University Press.
- Gallegati, M., Gallegati, M., Ramsey, J. and Semmler, W. (2011). The US Wage Phillips Curve across Frequencies and over Time. *Oxford Bulletin of Economics and Statistics*, 73(4), pp.489-508.
- Reboredo, J.C. (2011). How do crude oil prices co-move? *Energy Economics*, 33(5), pp. 948–955. doi: 10.1016/j.eneco.2011.04.006.
- Reboredo, J.C. (2012). Modelling oil price and exchange rate co-movements. *Journal of Policy Modeling*, 34(3), pp. 419–440. doi: 10.1016/j.jpolmod.2011.10.005.
- Reboredo, J.C. and Rivera-Castro, M.A. (2013). A wavelet decomposition approach to crude oil price and exchange rate dependence. *Economic Modelling*, 32, pp. 42–57. doi: 10.1016/j.econmod.2012.12.028.
- Yousefi, A. and Wirjanto, T.S. (2004). The empirical role of the exchange rate on the crude-oil price formation. *Energy Economics*, 26(5), pp. 783–799. doi: 10.1016/j.eneco.2004.06.001.

Using Bayesian Methods for Categorical Data Analysis

¹MSc. Erjola Cenaj, ²Dr. Raimonda Dervishi

¹Polytechnic University of Tirana, Faculty of Mathematical Engineering and Physical Engineering, Department of Mathematical Engineering

Tirana, Albania, erjola_cenaj@yahoo.com

²Polytechnic University of Tirana, Faculty of Mathematical Engineering and Physical Engineering, Department of Mathematical Engineering

Tirana, Albania, raimondadervishi@yahoo.com

Abstract

There are a variety of methods available to analyse categorical data. This paper surveys Bayesian methods for categorical data analysis as in many other areas of statistics, with primary emphasis on contingency table analysis. We focus on how the Bayesian approach can be used to estimate cell probabilities and an overview of the theoretical principles underlying the Bayesian approach. This will be followed by a literature review to show how the Bayesian approach has been applied in practice.

Keywords: Bayes' Theorem, Contingency table, Binomial distribution, Dirichlet distribution.

Introduction

The Bayesian paradigm is increasingly as computations become easier to implement and methodology using the Bayesian paradigm has advanced tremendously, also is used for categorical analysis as in many other areas of statistics. For multiway contingency table analysis, partly this is because of the plethora of parameters for multinomial models, often necessitating substantial prior specification. We present only problems in which the Bayesian approach applies quite naturally and is more appealing than Maximum Likelihood approach. Then we summarize more complex developments.

Bayes' Theorem

Bayesian statistical methods originate from an alternative philosophical viewpoint, requiring the analyst to reframe problems in terms of Bayesian logic. This logic combines “subjective” or prior knowledge, typically in the form of statistical distributions, with “objective” current information (data), to derive meaningful “posterior” distributions.

With parameter vector β , in Bayes' theorem current information (or likelihood) provided by data $\mathbf{Y}$, a realization of $\mathbf{Y} \sim f(\mathbf{Y}|\beta)$, is combined with prior information in the form of a prior distribution of unknown parameter values with density $P(\beta)$, which through Bayes' theorem results in the posterior distribution $P(\beta|\mathbf{Y})$. Bayes' theorem is given as

$$\pi(\beta|\mathbf{Y}) = \frac{f(\mathbf{Y}|\beta)P(\beta)}{\int f(\mathbf{Y}|\beta)P(\beta)d\beta}$$

where $m(x) = \int f(\mathbf{Y}|\beta)P(\beta)d\beta$ is the marginal density of $\mathbf{Y}$, or the probability of the data for the model with parameters $\mathbf{b}$. For interesting Bayesian models, the marginal density of $\mathbf{Y}$ presents a difficult integration problem, so a similar expression is often used instead, where

$$\pi(\beta|Y) \approx f(Y|\beta)P(\beta)$$

In this equation the posterior is proportional to the product of the prior and the likelihood. The relative weights of the likelihood and prior are determined by the variances of the distributions, with smaller variance resulting in greater weight in the determination of the posterior. For example, a normal prior of $P(b)_1 \sim N(\mu = \sigma = 10)$ would have less influence on the likelihood than would a normal prior of $P(b)_2 \sim N(\mu = 10, \sigma = 2)$.

The Bayesian theorem asserts that useful information about probabilities is obtained about specific observable events through subjective, expert evaluation or insight. For example, in the simple linear regression model $\hat{y} = b_0 + b_1 x$, an analyst may know something about the expected value of the unknown model parameter b_1 from prior research, from fundamental knowledge, or from prior data. It could be that $b_1 \sim N(\mu, \sigma)$, where μ and σ are known parameters of a normal distribution. Large values of σ might indicate a high degree of uncertainty in the value of b_1 while smaller values may indicate greater confidence. Alternatively, the analyst might believe that $b_1 \sim U(l, u)$, where l and u are the lower and upper bounds of a continuous uniform distribution, respectively.

A common criticism of Bayesian models is the selection of subjective priors to influence parameter values. Often the source of information that guides the selection of priors is the cumulative body of past research. As an example, suppose that ten studies on a phenomenon under investigation show the set of values of b_1 in a linear regression model to be {5.5, 8.1, 9.0, 6.3, 5.3, 6.5, 3.9}. How should this information be expressed in a prior? One could fit these data to a normal distribution and use the parameters obtained as priors. If there was belief that these values should not dominate the model parameters, then the variance could be increased to reflect greater or less belief in these values. However, analysts must defend the choice of priors, and often it is a challenge to argue convincingly for one prior over another. As mentioned previously, a common approach is to assume that nothing informative is known about priors, allowing “ignorance”, “diffuse”, or “noninformative” priors to be used and thus removing the burden to defend subjective priors. As an example, one might assume that a parameter has a mean of zero and standard deviation 1,000—a relatively flat, noninformative prior. However, the assigning of ignorance priors, as in this example, often results in a model with parameters that could have been obtained via maximum likelihood (or other classical) methods. One would question the motive for using Bayes’ theorem under such circumstances.

Bayesian estimation about a binomial parameter

For a test of $H_0: \pi \geq \pi_0$ against $H_a: \pi < \pi_0$, a Bayesian P-value is the posterior probability, $P(\pi \geq \pi_0 | y)$. Routledge (1994) showed that with the Jeffreys prior and $\pi_0 = 1/2$, this approximately equals the one-sided mid P-value for the frequentist binomial test.

Much literature about Bayesian inference for a binomial parameter deals with decision-theoretic results. For estimating a parameter θ using estimator T with loss function $\omega(\theta)(T-\theta)^2$ the Bayesian estimator is $E(\theta\omega(\theta)|y)/E(\omega(\theta)|y)$. With loss function $(T-\pi)^2/(\pi(1-\pi))$ and uniform prior distribution, the Bayes estimator of π is the ML estimator $p = y/n$. Johnson (1971) showed that this is an admissible estimator, for standard loss functions. Rukhin (1988) introduced a loss function

that combines the estimation error of a statistical procedure with a measure of its accuracy, an approach that motivates a beta prior with parameter settings between those for the uniform and Jeffreys priors, converging to the uniform as n increases and to the Jeffreys as n decreases.

Diaconis and Freedman (1990) investigated the degree to which posterior distributions put relatively greater mass close to the sample proportion p as n increases. They showed that the posterior odds for an interval of fixed length centered at p is bounded below by a term of form ab^n with computable constants $a > 0$ and $b > 1$. They noted that Laplace considered this problem with a uniform prior in 1774. Related work deals with the consistency of Bayesian estimators. Freedman (1963) showed consistency under general conditions for sampling from discrete distributions such as the multinomial. He also showed asymptotic normality of the posterior assuming a local smoothness assumption about the prior. For early work about the asymptotic normality of the posterior distribution for a binomial parameter.

Draper and Guttman (1971) explored Bayesian estimation of the binomial sample size n based on r independent binomial observations, each with parameters n and π . They considered both π known and unknown. The π unknown case arises in capture-recapture experiments for estimating population size n . One difficulty there is that different models can fit the data well yet yield quite different projections. A later extensive Bayesian literature on the capture-recapture problem includes Smith (1991), George and Robert (1992), Madigan and York (1997). Madigan and York (1997) explicitly accounted for model uncertainty by placing a prior distribution over a discrete set of models as well as over n and the cell probabilities for the table of the capture-recapture observations for the repeated sampling. Fienberg, Johnson and Junker (1999) surveyed other Bayesian and classical approaches to this problem, focusing on ways to permit heterogeneity in catchability among the subjects. Dobra and Fienberg (2001) used a fully Bayesian specification of the Rasch model to estimate the size of the World Wide Web.

Joseph, Wolfson, and Berger (1995) addressed sample size calculations for binomial experiments, using criteria such as attaining a certain expected width of a confidence interval. DasGupta and Zhang (2005) reviewed inference for binomial and multinomial parameters, with emphasis on decision-theoretic results.

Dirichlet prior and posterior for multinomial parameters

Results for the binomial with beta prior distribution generalize to the multinomial with a Dirichlet prior (Lindley 1964, Good 1965). With c categories, suppose cell counts $(n_1, \dots, n_c)$ have a multinomial distribution with $n = \sum n_i$ and parameters $\pi = (\pi_1, \dots, \pi_c)$. Let $\{p_i = n_i/n\}$ be the sample proportions. The likelihood is proportional to $\prod_{i=1}^c \pi_i^{n_i}$.

The conjugate density is the Dirichlet, expressed in terms of gamma functions as

$$g(\pi) = \frac{\Gamma(\sum \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_{i=1}^c \pi_i^{\alpha_i-1} \text{ for } 0 < \pi_i < 1 \text{ all } i, \sum \pi_i = 1 \text{ where } \alpha_i > 0.$$

Let $K = \sum \alpha_i$. The Dirichlet has $E(\pi_i) = \alpha_i/K$ and $\text{Var}(\pi_i) = \alpha_i(K - \alpha_i)(K^2(K+1))$. The posterior density is also Dirichlet, with parameters $\{n_i + \alpha_i\}$, so the posterior mean is

$$E(\pi_i/n_1, \dots, n_c) = (n_i + \alpha_i)/(n + K)$$

Let $\gamma_i = E(\pi_i) = \alpha_i/K$. This Bayesian estimator equals the weighted average

$$n = (n + K)p_i + [K/(n + K)]\gamma_i$$

which is the sample proportion when the prior information corresponds to K trials with α_i outcomes of type i , $i = 1, \dots, c$.

Good (1965) referred to K as an attenuating constant, since with identical $\{\alpha_i\}$, this estimate shrinks each sample proportion toward the equi-probability value $\gamma_i = 1/c$. Greater attenuating occurs as K increases, for fixed n . Good (1980) attributed $\{\alpha_i = 1\}$, to De Morgan (1847), whose use of $(n_i + 1)/(n + c)$ to estimate π_i extended Laplace's estimate to the multinomial case. Perks (1947) suggested $\{\alpha_i = 1/c\}$, noting the coherence with the Jeffreys prior for the binomial. The Jeffreys prior sets all $\alpha_i = 0.5$.

Hoadley (1969) examined Bayesian estimation of multinomial probabilities when the population of interest is finite, of known size N . He argued that a finite-population analogue of the Dirichlet prior is a compound multinomial prior, which leads to a translated compound multinomial posterior. Let N denote a vector of nonnegative integers such that its i -th component N_i is the number of objects (out of N total) that are in category i ; $i = 1, \dots, c$. If conditional on the probabilities and N , the cell counts have a multinomial distribution, and if the multinomial probabilities themselves have a Dirichlet distribution indexed by parameter α such that $\alpha_i > 0$ for all i with $K = \sum \alpha_i$, then unconditionally N has the compound multinomial mass function,

$$f(N/N, \alpha) = \frac{N! \Gamma(K)}{\Gamma(N + K)} \prod_{i=1}^c \frac{\Gamma(N_i + \alpha_i)}{N_i! \Gamma(\alpha_i)}$$

This serves as a prior distribution for N . Given cell count data $\{n_i\}$ in a sample of size n , the posterior distribution of $N - n$ is compound multinomial with N replaced by $N - n$ and α replaced by $\alpha + n$. Ericson (1969) gave a general Bayesian treatment of the finite-population problem, including theoretical investigation of the compound multinomial.

Here is a summary of other Bayesian literature about the multinomial: Good and Crook (1974) suggested a Bayes / non-Bayes compromise by using Bayesian methods to generate criteria for frequentist significance testing, illustrating for the test of multinomial equiprobability.

An example of such a criterion is the Bayes factor given by the prior odds of the null hypothesis divided by the posterior odds. Dickey (1983) discussed nested families of distributions that generalize the Dirichlet distribution, and argued that they were appropriate for contingency tables. Sedransk, Monahan, and Chiu (1985) considered estimation of multinomial probabilities under the constraint $\pi_1 \leq \dots \leq \pi_k \geq \pi_{k+1} \geq \dots \geq \pi_c$, using a truncated Dirichlet prior and possibly a prior on k if it is unknown. Delampady and Berger (1990) derived lower bounds on Bayes factors in favor of the null hypothesis of a point multinomial probability, and related them to P -values in chi-squared tests. Bernardo and Ramo (1998) illustrated Bernardo's reference analysis approach by applying it to the problem of estimating the ratio π_i/π_j of two multinomial parameters. The posterior

distribution of the ratio depends on the counts in those two categories but not on the overall sample size or the counts in other categories. This need not be true with conventional prior distributions. The posterior distribution of $\pi_i/(\pi_i+\pi_j)$ is the beta with parameters $n_i + 1/2$ and $n_j + 1/2$, the Jeffreys posterior for the binomial parameter.

References

- Agresti, A and Hitchcock, D. Statistical Methods & Applications(2005) 14: 297–330DOI: 10.1007/s10260-005-0121-y,”Bayesian inference for categorical data analysis”.
- Bernardo, J. M. and Ramon, J. M. (1998) An introduction to Bayesian reference analysis: Inference on the ratio of multinomial parameters. *The Statistician*, 47, 101-135.
- Delampady, M. and Berger, J. O. (1990) Lower bounds on Bayes factors for multinomial distributions, with application to chi-squared tests of χ^2 . *The Annals of Statistics*, 18, 1295-1316.
- Diaconis, P. and Freedman, D. (1990) On the uniform consistency of Bayes estimates for multinomial probabilities. *The Annals of Statistics*, 18, 1317-1327.
- Ericson, W. A. (1969) Subjective Bayesian models in sampling finite populations. *Journal of the Royal Statistical Society, Series B, Methodological*, 31, 195-233.
- Good, (1965) *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Cambridge, MA: MIT Press.
- Good, I. J. and Crook, J. F. (1974) The Bayes/non-Bayes compromise and the multinomial distribution. *Journal of the American Statistical Association*, 69, 711-720.
- Hoadley, B. (1969) The compound multinomial distribution and Bayesian analysis of categorical data from finite populations. *Journal of the American Statistical Association*, 64, 216-229.
- Johnson, B. M. (1971) On the admissible estimators for certain fixed sample binomial problems. *The Annals of Mathematical Statistics*, 42, 1579-1587.
- Lindley, D. V. (1964) The Bayesian analysis of contingency tables. *The Annals of Mathematical Statistics*, 35, 1622-1643.
- Madigan, D. and York, J. C. (1997) Bayesian methods for estimation of the size of a closed population. *Biometrika*, 84, 19-31.
- Perks, W. (1947) Some observations on inverse probability including a new indifference rule. *Journal of the Institute of Actuaries*, 73, 285-334.
- Routledge, R. D. (1994) Practicing safe statistics with the mid-p*. *The Canadian Journal of Statistics*, 22, 103-110.
- Rukhin, A. L. (1988) Estimating the loss of estimators of a binomial parameter. *Biometrika*, 75, 153-155.
- Smith, P. J. (1991) Bayesian analyses for a multiple capture-recapture model. *Biometrika*, 78, 399-407

ICASE 2017
International Conference on
Applied Statistics and Econometrics

ISBN 978-9928-135-20-9



9 789928 135209 >