Mansor H. Ibrahim[1] **UDK 336.76:061.1 ASEAN**

ASYMMETRY IN MEAN-REVERTING BEHAVIOR OF ASEAN STOCK MARKET RETURNS

ABSTRACT

The present paper characterizes the mean-reverting behavior of six ASEAN markets – Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam – using an autoregressive exponential GARCH-in mean model and daily data from August 2000 to May 2010. The results indicate fast speed of mean-reversion in the returns of these markets but with quite distinct patterns of return dynamics. The evidence seems strong to suggest asymmetric mean reversion and overreaction during market downturns in the Indonesian market. The Vietnamese market exhibits most persistent return autocorrelation with some evidence pointing to higher persistence during market downturns. However, there is no evidence indicating significant serial correlation in the markets of Singapore and Thailand. Finally, the leverage effect is documented in all markets except Vietnam. We tentatively attribute these differences to stages of market development and, accordingly their levels of efficiency, and to the degree of market volatility.

Keywords: *Asymmetry, Mean Reversion, Volatility, AR-EGARCH(1, 1), ASEAN Markets*

JEL Classification: *G10, G12*

I. INTRODUCTION

The presence of mean-reverting behavior in stock index returns is a subject that has captured great interest. Arguably, the autocorrelation in market returns can be either positive or negative. According to Koutmos (1998, 1999), the presence of positive autocorrelation suggests partial adjustment of stock prices to their intrinsic values, which can be attributed to non-synchronous trading, time-varying short-term expected returns or risk premia, and market frictions. Meanwhile, the negative autocorrelation indicates that stock price changes tend to be followed by predictable changes in the opposite direction, which is consistent with the view that market agents overreact irrationally to shocks or deviations of stock returns from their long run values (De Bondt and Thaler, 1985, 1987). Pinpointing the presence of stock return autocorrelation and whether it is positive and negative is essential to investors for carving appropriate trading strategies. For instance, empirical evidence for significant autocorrelation indicates predictable return behavior and, if it is further found that the autocorrelation is negative, a contrarian trading strategy can be profitable.

[1] Department of Economics, Faculty of Economics and Management, Universiti Putra Malaysia, 43400 Serdang, Selangor MALAYSIA, Email: mansorhi@econ.upm.edu.my. I would like to express my thanks to the anonymous referees of the journal for helpful comments. The usual disclaimer, however, applies.

Recent literature has also acknowledged the possibility of asymmetric mean-reverting behavior of stock returns. This is in cognizant with the widely established results of asymmetric volatility of stock return that stock market declines tend to generate proportionately higher volatility than stock market upturns, the so-called leverage effect. As noted by Koutmos (1998, 1999), if the volatility responds asymmetrically to bad and good news, there is also the possibility that mean return can behave asymmetrically. This possibility stems from investors' higher risk aversion during market downturns and market specialists' easier task to maintain price continuity during market upturns, which result in faster return adjustments during market declines. The presence of asymmetry can also be due to asymmetry in business cycles where, as suggested by Sichel (1993), contractions tend to be steeper than expansion (i.e. sharpness) and troughs tend to be more pronounced than peaks (i.e. deepness). Since stock markets are fundamentally tied to business conditions, it is highly potential that the stock market will exhibit asymmetric patterns as well.

The issue of asymmetric adjustment of stock returns has been increasingly addressed and established in various empirical studies on particularly the developed markets. These include Koutmos (1998, 1999), Nam (2001, 2003), Nam, Pyun and Arize (2002), Nam, Pyun and Avard (2001), Nam, Pyun and Chu (2005), and Kulp-Tag (2007). Koutmos (1998) examines asymmetries in conditional means and conditional variances of nine developed markets and find clear evidence for asymmetric mean reversion for all markets except the UK and US. The serial correlations of stock returns in these markets are positive and significant following good news and, in most cases, are insignificant following bad news. These patterns indicate complete adjustments of the stock prices to their intrinsic values when they are undervalued. In addition, the leverage effect is documented in all markets. In a subsequent study, Koutmos (1999) examines the issue for G-7 markets and arrive at similar findings. A series of studies by Nam (2001, 2003), Nam, Pyun and Avard (2001), Nam, Pyun and Arize (2002) and Nam, Pyun and Chu (2005) further reaffirm faster adjustments of negative market returns. Nam, Pyun and Chu (2005) provide further evidence of negative serial correlation of stock returns after market downturns. Kulp-Tag (2007) applies Nam's (2001) framework to Nordic stock markets and document similar results for market overreaction during market downturns.

More recently, few studies have also extended the analysis of stock market mean-reverting behavior to emerging markets. Zhang and Li (2008) examine asymmetric dynamics of Chinese stock market. Their empirical evidence indicates that stock returns in this market do exhibit asymmetric adjustment and negative returns tend to lead to overreaction. They further find that, as the market progresses, the leverage effect in volatility pattern becomes more apparent. Meanhile, Liau and Yang (2008) examine the mean and volatility asymmetry of 7 Asian markets using daily data from 3 January 1994 to 31 March 2005 and find evidence for asymmetric mean reversion in these markets.

Following this line of inquiry, the present paper examines the asymmetric mean-reverting behavior in market returns of six ASEAN stock markets – Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam – over recent years. The relatively little empirical attention given to these ASEAN markets does not commensurate wide interest in these markets in recent years and their increasing importance in the global financial scene. As these markets are less developed and are thus less efficient, we should expect more evidence for return serial correlation. However, whether their autocorrelations are positive or negative remain vague as they are characterized by a larger extent of inefficiency and frictions, a larger proportion of less-

informed and thus probably irrational traders and a relatively high volatility. Given these differences, the early noted findings from the developed markets may not necessarily apply to these markets. Indeed, by including more advanced market of Singapore and the least developed but recently fast-growing market of Vietnam in the region, we even expect differences in their return dynamics. The present analysis attempts to cater this interest and, at the same time, enriches the empirical literature on the subject for the emerging markets.

To anticipate the results, we observe different patterns of their conditional means. The markets of Singapore and Thailand exhibit no autocorrelation in their returns suggesting complete mean reversion or random-walk behavior of stock prices. The positive autocorrelation is generally documented for the remaining markets. While there seems to be no asymmetric mean reversion in the markets of Malaysia and the Philippines, there is evidence of overreaction to negative news for the Indonesian market. The Vietnamese market demonstrates most persistence in its return dynamics and, interestingly, its persistence is enhanced during market downturns. Finally, except Vietnam, we document the presence of leverage effect in all markets.

The paper is organized as follows. In the next section, we describe our empirical framework. Section 3 presents data preliminaries while section 4 discusses estimation results. Finally, section 5 concludes with the main findings and some concluding remarks.

II. EMPIRICAL APPROACH

The present paper provides the empirical framework for the analysis. To motivate our final specification and clarify model interpretation, we follow Koutmos (1998) by characterizing the mean equation of stock return to follow an autoregressive process of order 1 as:

$$R_t = \mu + \beta R_{t-1} + \varepsilon_t \qquad (1)$$

where R is stock return computed as the logarithmic difference in the daily index under study. The focal parameter in (1) is β, which measures the degree of frictions in the market and serial correlation in market return. If the coefficient β is 0, then the market price completely adjusts to its mean or intrinsic value. The larger the value of β, the larger the degree of market frictions is or, alternatively, the more persistent the market return is. In Koutmos' (1998) partial adjustment framework, the coefficient β is expected to be positive. However, the market can exhibit negative serial correlation arising from overreaction of market agents to past information. It needs mentioning that the market return exhibits a random walk behavior or contains a unit root if β is indistinguishable from unity. Thus, the formulation of the mean equation obviates the need to conduct formal unit root tests to the return series. In our context, it is expected that $|\beta| < 1$ for the return to be stationary or to exhibit mean-reverting behavior.

The above equation assumes that the autocorrelation in stock return or the adjustment process is symmetric. If the adjustment process to good and bad news is asymmetric, the above specification is mis-specified (Koutmos, 1998). Existing empirical literature tends to agree that the mean-reverting process of stock market following a negative return is faster than following a positive return. In other words, the market tends to be more persistent in up direction than in

down direction. As noted by Kulp-Tag (2007), allowance for this asymmetry paves the way for the possibility of market overreaction following market downturns and thus evidence in favor of a technical trading strategy. Following Nam (2001) and Kulp-Tag (2007), we re-specify equation (1) to incorporate asymmetric mean-reversion in ASEAN markets as:

$$R_t = \mu + [\beta_1 + \beta_2 D_1(R_{t-1} < 0)]R_{t-1} + \varepsilon_t \quad (2)$$

where D_1 is a dummy variable taking the value of 1 if $R_{t-1} < 0$ and of 0 otherwise. Based on (2), the degree of market persistence or serial correlation following positive returns is given by β_1 while that following negative returns is measured by $\beta_1 + \beta_2$. Thus, the significance of β_2 serves as an evidence of asymmetric mean reverting behavior of stock returns. With $\beta_1 > 0$, $\beta_2 < 0$ indicates faster mean-reversion or adjustment speed and $\beta_2 > 0$ more persistence in returns following market downturns. Then, it is possible that $\beta_1 + \beta_2 < 0$, which is taken to indicate agents' overreaction following negative returns.

In addition to (2), we also examine the return dynamics of stock returns following consecutive negative market returns for up to 4 days as in Kulp-Tag (2007). This is done by estimating the following mean equations:

$$R_t = \mu + [\beta_1 + \beta_2 D_2(R_{t-1} < 0, R_{t-2} < 0)]R_{t-1} + \varepsilon_t \quad (3)$$

$$R_t = \mu + [\beta_1 + \beta_2 D_3(R_{t-1} < 0, R_{t-2} < 0, R_{t-3} < 0)]R_{t-1} + \varepsilon_t \quad (4)$$

$$R_t = \mu + [\beta_1 + \beta_2 D_4(R_{t-1} < 0, R_{t-2} < 0, R_{t-3} < 0, R_{t-4} < 4)]R_{t-1} + \varepsilon_t \quad (5)$$

D_i for i = 2, 3, and 4 takes the value of 1 if the past returns for consecutive i days are negative and of 0 otherwise. As will be presented later, the occurrences of consecutive negative returns over three or four days in these ASEAN markets are not uncommon. Accordingly, this extension is worth pursuing since the market psychology can be different following a string of negative market returns. These models are referred as Model 1, Model 2, Model 3 and Model 4 for respectively one-period, two-period, three-period and four-period negative returns.

To complete our specification, we specify the error variances of equations (2) to (5) to be time-varying using GARCH-type models and consider incorporation of the time-varying volatility in the mean equations to capture mean-variance tradeoff in the return dynamics. It is well acknowledged that the financial time series tend to exhibit a leptokurtic property, i.e. a high-peaked and fat-tailed distribution, implying autocorrelation in their variance process. Engle (1982) develops the autoregressive conditional heteroskedasticity (ARCH) model to capture this leptokutic behavior of time series, which is later generalized by Bollerslev (1986) and hence the generalized ARCH model or GARCH. Nelson (1991) further extends the GARCH model by allowing the presence of asymmetric volatility in market returns by using the exponential GARCH model or EGARCH. With recent recognition of the asymmetric volatility, we employ an EGARCH specification in our modeling.

More specifically, our specification is the Asymmetric Autoregressive – Exponential GARCH – in mean model or, in short, asymmetric AR-EGARCH(1, 1)-M model. This model is expressed as:

$$R_t = \mu + [\beta_1 + \beta_2 D_i]R_{t-1} + \lambda \log(h_t) + \varepsilon_t \quad (6)$$

$$\varepsilon_t \mid I_{t-1} \sim GED(0, h_t, v) \quad (7)$$

$$\log h_t = \theta_0 + \theta_1 \log h_{t-1} + \theta_2 \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + \theta_3 \frac{\varepsilon_{t-1}}{h_{t-1}} \quad (8)$$

Equation (6) is the mean equation as elaborated above but extended to include time-varying variances in natural log form. Then, given the information set up to time t -1, the model error term is assumed to have the generalized error distribution with mean 0, time-varying variance h_t and measure of thickness v. Finally, we use the exponential EGARCH to characterize the asymmetric time-varying variances. In the empirical implementation, we estimate the specified asymmetric AR-EGARCH(1, 1)-M first. If we find no evidence of asymmetric volatility, the asymmetric AR-GARCH(1, 1)-M is used instead. The "in-mean" term is further dropped from the model if it is found to be insignificant.

The choices of GED and EGARCH specification are in line with existing studies, the brief explanation of which is in order. As shown by Lee, Chen and Rui (2001) and De Santis and Imrohoroglu (1997), the GED has the ability to well capture the leptokurtic properties of financial series. The GED distribution also nests other distributions as special cases. These include normal distribution ($v = 2$), double Exponential or Laplace distribution ($v = 1$) and uniform distribution ($v \rightarrow \infty$). The value of parameter v below 2 suggests thicker tails of the distribution than the standard normal while the value above 2 indicates thinner tails. The GED distribution also has an advantage in that extreme observations exert no excessive influence on the parameter estimates (Koutmos, 1998). As noted, the EGARCH variance specification is based on widely documented evidence that positive and negative shocks have asymmetric influences on stock market volatility (see the aforementioned studies on developed markets). Based on (8), negative news (i.e. $\varepsilon_t < 0$) leads to higher volatility if $\theta_3 < 0$. This is termed as the leverage effect in the finance literature. Note also that the EGARCH requires no non-negativity restrictions of the parameters in the variance equation.

III. DATA PRELIMINARIES

We employ aggregate stock market indexes of the six ASEAN markets, namely, the Jakarta Stock Exchange composite index for Indonesia (JSE), the Kuala Lumpur composite index for Malaysia (KLCI), the Philippines Stock Exchange composite index for the Philippines (PSE), the Straight Times Index for Singapore (STI), the Stock Exchange of Thailand composite index (SET) for Thailand, and the Ho Chi Min VSE price index (VSE) for Vietnam. All indexes are expressed in local currency. The data are daily from August 1, 2000 to May 31, 2010, a total of 2565 observations. The data sample starts beyond the 1997/1998 Asian crisis so there should be no contamination effect of the crisis experienced by the markets in this region on their return dynamics. Moreover, the price index of Vietnam is only available from August 2000 onwards. These data are retrieved from *Datastream International*. We compute daily returns for each market as the logarithmic difference in its corresponding market index. Table 1 provides

descriptive statistics of the return series. Meanwhile, Figure 1 graphs the stock market indexes (in natural logarithm) and their returns for the six ASEAN markets.

Table 1

Descriptive Statistics of Market Returns

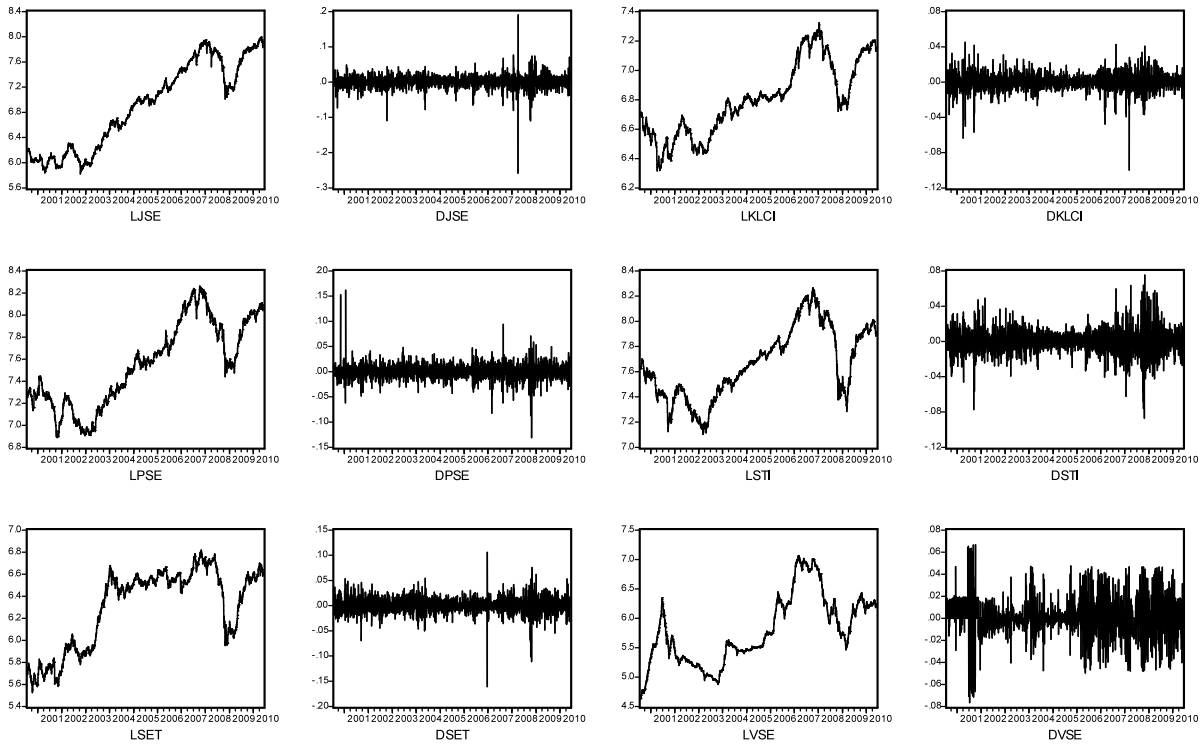
	DJSE	DKLCI	DPSE	DSTI	DSET	DVSE
Mean	0.000680	0.000183	0.000328	0.000110	0.000363	0.000627
Maximum	0.190719	0.045027	0.161776	0.075305	0.105770	0.066561
Minimum	-0.257802	-0.099785	-0.130887	-0.086960	-0.160633	-0.076557
Std. Dev.	0.016168	0.008929	0.014066	0.012970	0.014530	0.017062
Skewness	-1.471043	-0.966266	0.595114	-0.156424	-0.764576	-0.234291
Kurtosis	39.86122	13.56854	20.61313	7.909412	13.82205	5.626976
Jarque-Bera	146084.4	12331.63	33293.42	2585.388	12761.79	760.7143
P-values	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Notes: the prefix "D" denotes first difference. JSE = Jakarta Stock Exchange composite index; KLCI = Kuala Lumpur composite index; PSE = Philippines Stock Exchange composite index; STI = Singapore Straight Times index; SET = Stock Exchange of Thailand composite index; and VSE = Ho Chi Min VSE price index. The source of these data is *Datastream*.

From Table 1, we may note that all markets exhibit positive returns over the sample period with the Indonesian market recording the highest daily return (0.068%) followed by the Vietnamese market (0.063%). The market of Singapore experiences the lowest return (0.011%) followed by the Malaysian market (0.018%). As reflected by the standard deviation, the markets of Indonesia and Vietnam also have the highest unconditional volatility while the Malaysian market records the least volatility. It is interesting to note that the market of Indonesia has the most extreme positive and negative returns and yet is less volatile than the Vietnamese market. This shows that these extreme values are rare occurrences in the Indonesian market while the Vietnamese market returns fluctuate widely within the minimum and maximum returns. All market returns except the Philippines market are negatively skewed. The kurtosis statistics, which are substantially higher than 3, indicate excess peakness of the return distribution in all markets. The Jarque-Bera test statistics for normality soundly indicate deviations from the normal distribution in all ASEAN markets under study, which justify the use of GARCH-type models. These can be clearly seen in the return plots presented in Figure 1.

Figure 1:

Time-Series Plots of Stock Market Indexes and Returns



Notes: The prefix "L" denotes natural logarithms. See also notes from Table 1.

As a precursor to our formal analysis in the next section, we present the number of daily positive and negative returns and the number of consecutive positive and negative returns over 2, 3, and 4 days in Table 2. From Table 2, the six ASEAN markets can be divided into two groups. The first group records higher counts of consecutive positive returns than consecutive negative returns. These are the markets of Indonesia, Malaysia, Singapore and Thailand. Among these markets, the ratio of consecutive positive to negative return counts is highest for Indonesia and lowest for Thailand. Based on this, the Indonesian market seems to have the most potential in exhibiting asymmetric return dynamics. The markets of the Philippines and Vietnam make up the second group with the number of consecutive positive returns to be lower than that of the consecutive negative returns. While this pattern for the Philippines seems to accord well with its positively skewed distribution, the noted observation for the Vietnamese market is interesting. Particularly, for the Vietnamese market, the return dynamics may be more persistent following consecutive market downturns. From this preliminary analysis of the return series, we may expect different return dynamics among these markets. We turn to our formal analysis next to concretely uncover their potential asymmetric mean reversions.

Table 2:

Number of Consecutive Positive and Negative Returns

	DJSE	DKLCI	DPSE	DSTI	DSET	DVSE

Total Observations	2564	2564	2564	2564	2564	2564
Positive Observations	1398	1425	1260	1374	1391	1437
Negative Observations	1066	1139	1204	1190	1173	1127
Two Positive Observations	732	696	618	606	593	642
Two Negative Observations	505	563	637	542	558	667
Three Positive Observations	384	376	315	270	294	360
Three Negative Observations	243	280	345	243	266	394
Four Positive Observations	207	215	144	134	131	216
Four Negative Observations	110	135	173	106	123	247

Notes: the prefix "D" denotes first difference. JSE = Jakarta Stock Exchange composite index; KLCI = Kuala Lumpur composite index; PSE = Philippines Stock Exchange composite index; STI = Singapore Straight Times index; SET = Stock Exchange of Thailand composite index; and VSE = Ho Chi Min VSE price index. The source of these data is *Datastream*.

4. ESTIMATION RESULTS

This section discusses the estimation results of the four models examined, which are implemented using EVIEWS. Note that the sample runs through the recent 2007/2008 global financial crisis. To ascertain whether the crisis has any influence on the ASEAN market return dynamics, we also perform the analysis using the data up to 2006. The results are largely similar to the whole sample and, accordingly, not reported to conserve space. Take note that, for the Vietnamese market, we find no evidence of asymmetric volatility. Accordingly, the time-varying variance is modeled using the standard GARCH(1, 1). Moreover, the "in-mean" term is found to be significant on in the markets of Singapore and Vietnam. Estimation results for one-period negative return (Model 1) to four-period consecutive negative returns (Model 4) are given respectively in Table 3 to Table 6.

The coefficients of variance equations are statistically significant in all markets and all models. A common feature of the results is that the market volatilities of the ASEAN markets examined tend to be persistent depending mostly on past volatilities. Meanwhile, past shocks only assume a secondary role. Except Vietnam, we also observe the presence of leverage effect in all markets. The coefficient of the standardized error in the variance equation is negative and significant at 1% significance level. This means that past negative shocks or bad news is likely to generate proportionately higher volatility than past positive shocks of the same magnitude. The impact of negative shocks tends to be strongest for the Indonesian market, as reflected by the magnitude of its once-lagged standardized error coefficient. The heightened market volatility of the Indonesian market following market downturns in magnitude substantially higher than any other markets seems to suggest possible overreaction of market agents in Indonesia to negative shocks.

As expected, we observe different return dynamics of these six ASEAN stock markets. The common finding among these markets is the mean-reversion pattern of their returns with fast adjustment towards the mean value following either positive or negative returns. This is reflected by low estimated coefficients of the autoregressive term in the mean equation in all models and in all markets. Still, the results for the mean equation roughly divide these markets into four distinct return dynamics. The market of Indonesia stands alone as the only market that

exhibits asymmetric mean reversion and overreaction following market downturns. The autocorrelation of the Indonesian market during market downturns is significantly lower than that of market upturns and, reflecting overreaction, the autocorrelation during market downturns (i.e. $\beta_1 + \beta_2$) is negative and is significant at better than 10% significance level following three and four consecutive negative returns. This result conforms well to the noted relatively high volatility pattern in the Indonesian market. This means that, in Indonesia, a technical trading strategy based on buying the losers during market downturns, or the contrarian strategy, can be profitable.

Table 3:

Estimation Results - One Period

	DJSE	DKLCI	DPSE	DSTI	DSET	DVSE
(a) Mean Equation						
μ	0.0000	0.0001	0.0000	-0.0042	0.0000	0.0043**
β_1	0.1055***	0.0995***	0.0717***	-0.0147	0.0010	0.2166***
β_2	-0.1053**	-0.0149	0.0025	-0.0385	-0.0011	0.0165
λ	---	---	---	-0.0005*	---	0.0004**
(b) Variance Equation						
θ_0	-1.3613***	-0.3991***	-0.8931***	-0.2366***	-0.6304***	0.0000***
θ_1	0.8607***	0.9733***	0.9155***	0.9872***	0.9430***	0.7699***
θ_2	0.2429***	0.1923***	0.2195***	0.1609***	0.1947***	0.2686***
θ_3	-0.1549***	-0.0585***	-0.0710***	-0.0659***	-0.0594***	---
v	1.0016***	1.1306***	1.0662***	1.3282***	1.0753***	1.2812***
(c) Model Statistics						
LL	7524.82	8942.53	7693.55	7794.60	7574.31	7705.07
F-Stat	3.2886***	7.6251***	4.4064***	0.0000	0.0000	21.892***

Notes: the model estimated Model (1). *LL* = Log Likelihood and *F-Stat* = F statistics for the model's overall significance. ***, ** and * denote significance at 1%, 5% and 10% respectively.

Source: author's estimation using EVIEWS.

Table 4:

Estimation Results - Two Periods

	DJSE	DKLCI	DPSE	DSTI	DSET	DVSE
(a) Mean Equation						
μ	0.0003	0.0001	0.0000	-0.0038	0.0000	0.0038**
β_1	0.0779***	0.1015***	0.0750***	-0.0241	0.0005	0.2344***
β_2	-0.0632	-0.0316	-0.0061	-0.0210	-0.0331	-0.0304
λ	---	---	---	-0.0004*	---	0.0004**
(b) Variance Equation						
θ_0	-1.3701***	-0.4064***	-0.8788***	-0.2394***	-0.6357***	0.0000***
θ_1	0.8601***	0.9727***	0.9168***	0.9871***	0.9424***	0.7706***
θ_2	0.2460***	0.1943***	0.2162***	0.1633***	0.1940***	0.2676***
θ_3	-0.1557***	-0.0591***	-0.0698***	-0.0655***	-0.0594***	---
v	1.0142***	1.1297***	1.0661***	1.3283***	1.0710***	1.2848***
(c) Model Statistics						
LL	7518.77	8939.72	7691.59	7991.47	7572.48	7701.92
F-Stat	2.8565***	8.1279***	4.4711***	0.0000	0.3930	22.153***

Notes: the model estimated Model (2). *LL* = Log Likelihood and *F-Stat* = F statistics for the model's overall significance. ***, ** and * denote significance at 1%, 5% and 10% respectively.

Source: author's estimation using EVIEWS.

The Vietnamese market is distinct from the rest in several aspects. Apart from the absence of the leverage effect, the Vietnamese market is the only market that exhibits positive and significant mean-variance relationship. Second, while we note no evidence of asymmetric mean reversion following a negative return, two-period consecutive returns, and three-period consecutive returns, the market tends to exhibit most persistence among the market considered. The serial correlation coefficient is found to be higher than 0.2 in all models estimated, which is more than double to the ones estimated for other markets. Moreover, we do find asymmetric mean reversion in the Vietnamese stock returns following four consecutive negative returns with the pattern to be opposite to the one observed for Indonesia. Namely, following market downturns over 4 consecutive days, the Vietnamese market tends to exhibit more persistence having serial correlation of 0.36 during downturns as opposed to 0.20 during upturns. This finding is interesting since it counters the widely-noted finding of faster adjustment of stock markets during the downturns in other studies on developed markets. In short, what is observed for the developed markets or even other emerging markets as in our study may not necessarily be true for a specific emerging market at its early stage of development. Zhang and Li (2008) also document similar results during the initial sample periods in their analysis of the Chinese markets (see Zhang an d Li (2008, 961)).

Table 5:

Estimation Results - Three Periods

	DJSE	DKLCI	DPSE	DSTI	DSET	DVSE
(a) Mean Equation						
μ	0.0004*	0.0001	-0.00002	-0.0039	0.0000	0.0041***
β_1	0.0780***	0.1040***	0.0847***	-0.0200	0.0004	0.2211***
β_2	-0.1131**	-0.0746	-0.0853*	-0.0907	-0.0472	0.0235
λ	---	---	---	-0.0004*	---	0.0004***
(b) Variance Equation						
θ_0	-1.3554***	-0.3967***	-0.8715***	-0.2346***	-0.6378***	0.0000***
θ_1	0.8617***	0.9734***	0.9175***	0.9874***	0.9421***	0.7699***
θ_2	0.2433***	0.1902***	0.2135***	0.1600***	0.1939***	0.2687***
θ_3	-0.1545***	-0.0587***	-0.0671***	-0.0655***	-0.0597***	---
v	1.0157***	1.1266	1.0578***	1.3260***	1.0689***	1.2780***
(c) Model Statistics						
LL	7516.23	8937.18	7690.17	7988.60	7570.07	7699.37
F-Stat	7.1837***	8.5150***	4.3692***	0.0000	0.1352	21.721***

Notes: the model estimated Model (3). *LL* = Log Likelihood and *F-Stat* = F statistics for the model's overall significance. ***, ** and * denote significance at 1%, 5% and 10% respectively.

Source: author's estimation using EVIEWS.

The markets of Malaysia and the Philippines exhibit no asymmetric mean reversion except for one case for the Philippines. Given low serial correlation in their returns, the aggregate stock prices of Malaysia and the Philippines tend to adjust fast to their intrinsic values. In the case of the Philippines, we note complete adjustment following three consecutive returns whereby the sum $\beta_1 + \beta_2$ is insignificantly different from zero. Finally, the return autocorrelation of the Singapore and Thai markets is indistinguishable from zero regardless of the directions of the market movements. Thus, only in these two markets, stock prices tend to exhibit a clear random walk behavior and adjust completely to their intrinsic values.

The results we obtain are intriguing when viewed in lights of the general uniformity of return dynamics documented for the developed markets by Koutmos (1998, 1999). The return dynamics in the ASEAN markets are far from being uniform, which begs for explanation. We believe that stages of stock market development and degree of volatility may account for the results. Among the markets examined, the Vietnamese stock exchange is the recent upcoming market but still lags behind other markets in region. Interestingly, the behavior of the Vietnamese market tends to mimic the Chinese stock market during 1990 and 1991 as documented by Zhang and Li (2008). The Indonesian market is relatively more volatile. This is consistent with the finding of LeBaron (1992) that the return autocorrelation and volatility are negatively related. Thus, the heightened volatility during market downturns as captured by the leverage effect may account for lower or even negative return autocorrelation during market downturns. Finally, the market of Singapore is the most advanced in the region and,

accordingly, the finding of complete mean-reverting behavior comes at no surprise. We admit that this explanation is at best tentative and it remains intriguing why the Thai market tends to behave in the same way as the market of Singapore. In short, further research is needed to account for different dynamic patterns of stock market returns in these countries.

Table 6:

Estimation Results - Four Periods

	DJSE	DKLCI	DPSE	DSTI	DSET	DVSE
(a) Mean Equation						
μ	0.0004**	0.0002	0.0000	-0.0035	0.0000	0.0041***
β_1	0.0718***	0.0946***	0.0785***	-0.0269	0.0001	0.2032***
β_2	-0.1951**	-0.0328	-0.0785	-0.0318	-0.0029	0.1600***
λ	---	---	---	-0.0004*	---	0.0004***
(b) Variance Equation						
θ_0	-1.3694***	-0.3947***	-0.8720***	-0.2346***	-0.6521***	0.0000***
θ_1	0.8605***	0.9737***	0.9173***	0.9874***	0.9406***	0.7684***
θ_2	0.2482***	0.1909***	0.2135***	0.1605***	0.1964***	0.2718***
θ_3	-0.1555***	-0.0582***	-0.0680***	-0.0663***	-0.0609***	---
v	1.0192	1.1294***	1.0545***	1.3290***	1.0659***	1.2510***
(c) Model Statistics						
LL	7513.24	8932.53	7685.40	7984.27	7567.39	7698.89
F-Stat	0.3896	7.7610***	4.3351***	0.0000	0.0000	20.807***

Notes: the model estimated Model (4). *LL* = Log Likelihood and *F-Stat* = F statistics for the model's overall significance. ***, ** and * denote significance at 1%, 5% and 10% respectively.

Source: author's estimation using EVIEWS.

5. CONCLUSION

With the interest in dynamic patterns of stock returns in ASEAN markets, this paper empirically analyzes the presence asymmetric mean reversion and asymmetric volatility of six ASEAN markets – Indonesia, Malaysia, the Philippines, Singapore, Thailand, and Vietnam. Among these markets, the Singapore market is the most advanced and the Vietnamese market is least developed. Moreover, the Indonesian and Vietnamese markets are relatively more volatile while the Malaysian and Singapore markets are at the other end of the volatility spectrum. In the analysis, we make use of the asymmetric autoregressive specification of the mean equation and the exponential-GARCH specification of the variance equation to characterize stock return dynamics. In general, the adjustment process of their market returns is fast but with evidence of quite distinct dynamic behavior of these markets and over August 2000 to May 2010.

The Indonesian markets stand out as the only market that exhibits overreaction during market downturns while the Vietnamese market demonstrates most persistence in its return behavior with the degree of persistence to be higher during the market downturns. No evidence is

uncovered for the markets of Malaysia and the Philippines and no evidence is found for significant serial correlation in Singapore and Thai stock returns. Finally, we find the presence of asymmetric volatility in all markets except the Vietnamese market. Consistent with the leverage effect, negative shocks tend to result in proportionately higher volatility than positive shocks of the same magnitude do. We tentatively attribute the differences across these markets to the stages of market development and the degree of volatility. However, to be concrete as to the reasons underlying these differences, further research is needed.

REFERENCES

Bollerslev, T., (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

DeBondt, W.F.M., and Thaler, R. H., (1985), "Does the Stock Market Overreact?' *Journal of Finance*, 40, 793-805.

DeBondt, W.F.M., and Thaler, R. H., (1987), "Further Evidence on Investor Overreaction and Stock Market Seasonality," *Journal of Finance*, 42, 557-581.

De Santis, G., and Imrohoroglu, S., (1997), "Stock Returns and Volatility in Emerging Financial Markets," *Journal of International Money and Finance*, 16, 561-579.

Engle, R. F., (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987-1008.

Koutmos, G., (1999), "Asymmetries in the Conditional Mean and the Conditional Variance: Evidence from Nine Stock Markets," *Journal of Economics and Business*, 50, 277-290.

Koutmos, G., (1999), "Asymmetric Index Stock Returns: Evidence from the G-7," *Applied Economic Letters*, 6, 817-820.

Kulp-Tag, S., (2007), "Short-Horizon Asymmetric Mean Reversion and Overreactions: Evidence From the Nordic Stock Markets," *Meddelanden Working Paper*, no. 524.

LeBaron, B., (1992), "Some Relations between Volatility and Serial Correlations in Stock Market Returns," *The Journal of Business*, 199-219.

Lee, C. F., Chen, G. M., and Rui, O., (2001), "Stock Returns and Volatility in China's Stock Markets," *Journal of Financial Research*, 24, 523-543.

Liau, Y.-S., and Yang, J.J.W., (2008), "The Mean/Volatility Asymmetry in Asian Stock Markets," *Applied Financial Economics*, 18, 411-419.

Nam, K., (2001), "Stock Overreaction and Contrarian Profits: Findings from Asymmetric Reversal Intrinsic in Time Series," Paper presented at the 2001 FMA Annual Meeting in Toronto, Canada, October.

Nam, K., (2003), "The Asymmetric Reverting Property of Stock Returns," *Studies in Nonlinear Dynamics and Econometrics*, 6(4), Article 2.

Nam, K., Pyun, C.S., and Arize, A., (2002), "Asymmetric Mean-Reversion and Contrarian Profits: ANST-GARCH Approach," *Journal of Empirical Finance*, 9, 563-588.

Nam, K., Pyun, C. S., and Avard, S., (2001), "Asymmetric Reverting Behavior of Short-Horizon Stock Returns: An Evidence of Stock Market Overreaction," *Journal of Banking and Finance,* 25, 807-824.

Nam, K., Washer, K. M., and Chu, Q. C., (2005), "Asymmetric Return Dynamics and Technical Trading Strategies," *Journal of Banking and Finance*, 29, 391-418.

Nelson, D. B., (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica,* 59, 347-370.

Sichel, D. E., (1993), "Business Cycle Asymmetry: A Deeper Look," *Economic Inquiry*, XXXI, 224-236.

Zhang, B., and Li. X., (2008), "The Asymmetric Behavior of Stock Returns and Volatilities: Evidence from Chinese Stock Market," *Applied Economics Letters*, 15, 959-962.

ASIMETRIJA U PONAŠANJU KRETANJA PROSJEČNIH VRIJEDNOSTI ZARADA BURZE ASEAN-a

SAŽETAK

Ovaj rad obrađuje ponašanje kretanja prosječnih vrijednosti šest tržišta ASEAN-a – indonezijskog, malezijskog, filipinskog, singapurskog, tajlandskog i vijetnamskog – koristeći eksponencijalni autoregresijski GARCH-M model i dnevne podatke od kolovoza 2000. do svibnja 2010. Rezultati ukazuju na veliku brzinu povratka na prosječne vrijednosti u zaradama ovih tržišta no s jasno vidljivim uzorkom dinamike zarade. Postoje čvrsti dokazi koji ukazuju na asimetrično kretanje prosječnih vrijednosti i pretjerane reakcije za vrijeme pada tržišta na indonezijskom tržištu. Vijetnamsko tržište pokazuje najustrajniju autokorelaciju zarade s nekim dokazima koji upućuju na veću ustrajnost za vrijeme pada tržišta. Ipak, ne postoje dokazi koji bi upućivali na značajniju serijsku korelaciju na tržištima Singapura i Tajlanda. Na kraju, efekt poluge je zabilježen na svim tržištima osim vijetnamskog. Privremeno te razlike pripisujemo fazama razvoja tržišta i, u skladu s njihovim razinama efikasnosti, stupnju volatilnosti tržišta.

Ključne riječi*: asimetrija, lretanje prosječnih vrijednosti, volatilnost, AR-EGARCH(1,1), tržišta ASEAN-a*

JEL klasifikacija*: G10, G12*

Jiancai Pi, Ph.D.[1] **UDK 061.232-055.5/.7(510)**

FAIRNESS AND INCENTIVES IN RELATION-BASED SOCIETIES

ABSTRACT

This paper mainly discusses the effects of fairness on incentives in relation-based societies (e.g., China) through the principal-agent framework. Our analyses give the conditions under which the consideration of fairness will decrease or increase the agent's efficiency wage. At the same time, our analyses give the conditions under which taking into account fairness will make the principal's constraint to incentivize the agent easier or harder to be satisfied. In a word, this paper finds that the incentive effects of fairness are condition-dependent, and that moral hazard problems are more subtle and difficult to be tackled when fairness is taken into account.

Keywords: *Fairness, Incentive, Principal-agent, Efficiency Wage*

JEL Classification: *A12, D63, J33*

1. INTRODUCTION

More and more economists pay attention to the effects of fairness on economic relations. However, there are different definitions of fairness. In his original paper, Rabin (1993) develops a formal benchmark which incorporates fairness into game theory and economics. Rabin's fairness means that people like to help those who are helping them and hurt those who are hurting them. Zajac (1995) holds that "strategic uses of fairness to advance self-interest is commonplace," and that to develop a theory of fairness strategizing needs to incorporate two feathers of human behavior. The first feather is that people are not ususally energized by a sense of fairness but by a feeling of unfairness. The second feather is that "people have a marvelous ability to deny." Konow (1996) generalizes a conclusion from the empiricial research that "fairness is a highly differentiated phenomenon which varies widely with context, e.g. social, institutional or cultural context." Kaplow and Shavell (2002) are of the opinion that "notions of fairness typically are used to reach conclusions based upon situational characteristics of events." Alvi (1998) holds that there are three sources of fairness: moral precept, convention, and reciprocity. The first source means that "fairness is driven by moral prerogatives." The second source implies that "fairness is a stable convention brought

1 Associate Professor

Department of Economics, School of Business, Nanjing University, 22 Hankou Road, Nanjing 210093, P. R. China

Phone: 86-25-83621121 (O)

E-mail: jiancaipi@hotmail.com; pi2008@nju.edu.cn

Academic Degree: Ph.D. in Economics

Current Position: Associate Professor with Department of Economics, School of Business, Nanjing University

Homepage: http://nubs.nju.edu.cn/en/faculty.php/160

http://nubs.nju.edu.cn/faculty.php/160

about by an evolutionary process." The third source signifies that fairness is the result of reciprocity. In a word, different people hold dissimilar and even conflicting ideas about fairness. In order to unify some concepts in the existent theories of fairness, Corchón and Iturbe-Ormaetxe (2001) try to propose a procedure based on the aspiration functions, which represents an important direction and a potential mathematical framework for the future research.

Although many mathematical tools are used to handle fairness, such as game theory (Rabin, 1993; Morelli and Sacco, 1997; Cox et al., 2007), bargaining theory (Bereby-Meyer and Niederle, 2005; von Siemens, 2009), and incentive theory (Fehr and Schmidt, 2000; Fehr and Schmidt, 2004; Fehr et al., 2007), there are three very distinct approaches among them. The first approach is to regard fairness as a variable of the utility function (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Trautmann, 2009). These seminal works have made a great progress in incorporating fairness into formal analyses, although they apply different processing methods, which makes it easy for them to make some specific predictions about many phenomena. The second approach is to treat fairness as a an extrinsic constraint (e.g., Pi, 2007, 2008). This approach argues that fairness is a psychological factor which is determined by the cultural background and special situation to a great extent (Kahneman et al., 1986; Akerlof and Yellen, 1988; Konow, 1996; Shavell, 2002), and can conduct formal mathematical analyses easily by introducing some constraint conditions. The third approach is to introduce the psychological cost or painful guilt resulting from unfairness (e.g., Gill and Stone, 2010). People can easily understand this approach by intuition. As a whole, our paper will adopt the third approach.

There are two papers which are closely related to our paper. Fehr and Schmidt (2000) and Fehr et al. (2007) also examine fairness in strategic interactions, and find that fairness concerns can affect the provision of incentives through different contracts. There are three important distinctions between our paper and theirs. Firstly, their models adopt the first approach to some extent, while our model adopts the third approach in the mass. Thus, our main variables are fundamentally different from theirs. Secondly, their models are more simplified and focuses mainly on the choice of contracts, while our model is more complex and focuses mainly on the different equilibrium outcomes without and with fairness. Thirdly, our analyses are on the basis of relation-based societies (e.g., China). In relation-based societies, relation (or guanxi in Chinese language) is an important factor that should be considered. In these kinds of societies, the principal and the agent are thoroughly familiar with each other. In other words, the principal has intimate knowledge of the agent, and the agent does so, too. This relation or guanxi can ensure that they have common knowledge about cost parameters related to unfairness through repeated social interactions. Furthermore, the existent literature shows that small group sizes more favor fairness (Höffler, 1999).

The rest of the paper is organized as follows. Section 2 is the basic setup. Section 3 provides the model without fairness. Section 4 offers the model with fairness. Section 5 conducts a comparative analysis of outcomes derived from the two different models. Some concluding remarks are made in Section 6.

2. THE BASIC SETUP

In this section, we follow Laffont and Martimort's (2002) analytical framework. There is a principal-agent relationship which is subject to moral hazard. It is assumed that the agent is

risk-neutral. If he exerts effort level $e \in \{0,1\}$, the principal's added-value will be $\overline{V}$ with probability $\pi(e)$, and $\underline{V}$ with probability $1-\pi(e)$, where $0<\pi(e)<1$. $\pi(e)$ and $1-\pi(e)$ can be seen as success and failure probabilities, respectively. When the agent's performance is good, he can get a bonus; however, when his performance is bad, he will not be punished. That is to say, he is protected by limited liability. When he exerts no effort, his effort cost is $\Psi_0 = 0$. When he exerts effort, his effort cost is $\Psi_1 = \Psi > 0$. The subscripts 0 and 1 represent $e=0$ and $e=1$, respectively. For the sake of narrative simplicity, we call Ψ the exertion cost. The following mathematical definitions should be noted, $\pi(1)=\pi_1$, $\pi(0)=\pi_0$, $\Delta\pi = \pi_1 - \pi_0 > 0$, $\Delta V = \overline{V} - \underline{V} > 0$.

In order to overcome the moral hazard problem, the principal has to adopt an appropriate scheme to incentivize the agent. When the added-value is $\overline{V}$, the principal offers an efficiency wage $\overline{w} \geq 0$ to the agent. However, when the added-value is $\underline{V}$, the principal offers an efficiency wage $\underline{w} \geq 0$ to the agent. It is assumed that $\overline{w} \geq \underline{w}$.

We consider two distinct cases in this paper. The first case is that the principal and the agent do not care for fairness. This case is the same as that in the standard incentive theory. The second case is that the principal and the agent take fairness into account. This case is greatly different from that of the traditional literature. If the agent exerts effort and succeeds, both the principal and the agent will take it as fair. However, if the agent exerts effort and fails, both the principal and the agent will take it as unfair, which will bring unfairness costs $F_h \geq 0$ and $f_h \geq 0$ to the principal and the agent, respectively. Similarly, if the agent exerts no effort and succeeds, both the principal and the agent will take it as unfair, which will bring unfairness costs $F_b \geq 0$ and $f_b \geq 0$ to the principal and the agent, respectively. Unfairness related to F_h and f_h is called the first-type unfairness, and hence F_h and f_h are called the first-type unfairness cost to the principal and the agent, respectively. Similarly, unfairness related to F_b and f_b is called the second-type unfairness, and hence F_b and f_b are called the second-type unfairness cost to the principal and the agent, respectively. Unfairness leads to some kind of disutility cost to the principal and the agent who receive this kind of feeling.

The introduction of F_h, f_h, F_b and f_b is inspired by Gill and Stone (2010), who treat the unfairness in an indirect way by introducing some ad hoc cost parameters which are different from ours because they deal with a dissimilar issue. Here, the relationship between f_h and f_b should be noted. f_h and f_b reflect the agent's degree of psychological hatred

to "reaping where one has not sown" (laoerbuhuo in Chinese language) and "getting something for nothing" (bulaoerhuo in Chinese language) in relation-based societies (e.g., China), respectively. So, the relation may be that $f_h > f_b$, $f_h < f_b$ and $f_h = f_b$, which is obviously ambiguous. Similarly, the relation between F_h and F_b is uncertain, too.

The timing of the principal-agent game is as follows.

(1) At *t*=0, the principal offers an incentive contract $\{(\underline{w}, \overline{w})\}$ to the agent.

(2) At *t*=1, the agent rejects or accepts the offer.
(3) At *t*=2, the agent chooses an effort, which is equal to 1 or 0.
(4) At *t*=3, the principal's added-value is realized.
(5) At *t*=4, the signed contract is enforced.

3. THE MODEL WITHOUT FAIRNESS

When the principal and the agent do not consider fairness, there is no unfairness cost at all. In order to make the agent exert effort, the principal must find an optimal compensation plan $\{(\underline{w}, \overline{w})\}$. The principal's programming problem will be:

$$\max_{\{(\underline{w},\overline{w})\}} \pi_1(\overline{V} - \overline{w}) + (1-\pi_1)(\underline{V} - \underline{w})$$

$$s.t. \quad \pi_1 \overline{w} + (1-\pi_1)\underline{w} - \Psi \geq \pi_0 \overline{w} + (1-\pi_0)\underline{w} \qquad (1)$$

$$\pi_1 \overline{w} + (1-\pi_1)\underline{w} - \Psi \geq 0 \qquad (2)$$

$$\underline{w} \geq 0 \qquad (3)$$

(1), (2), and (3) are the agent's incentive compatibility, participation, and limited liability constraints when fairness is not considered, respectively.

According to the standard incentive theory, it is easy for us to find that constraints (1) and (3) are binding.

Solving this programming problem, we obtain:

$$\underline{w}^{N*} = 0 \qquad (4)$$

$$\overline{w}^{N*} = \frac{\Psi}{\Delta\pi} \qquad (5)$$

The superscript $N*$ stands for second-best state with no consideration of fairness. $\overline{w}^{N*}$ is the agent's efficiency wage with no consideration of fairness.

In addition, we need to consider the principal's constraint to incentivize the agent:

$$\pi_1(\overline{V} - \frac{\Psi}{\Delta\pi}) + (1-\pi_1)\underline{V} \geq \pi_0 \overline{V} + (1-\pi_0)\underline{V} \qquad (6)$$

It is quite obvious that (6) can be simplified into:

$$\Delta\pi\Delta V \geq \frac{\pi_1 \Psi}{\Delta\pi} \qquad (7)$$

Only when (7) is satisfied does the principal have the incentive to make the agent exert

effort, otherwise he will prefer not to do it. Therefore, this constraint should be taken seriously during the course of our analyses.

4. THE MODEL WITH FAIRNESS

When the principal and the agent do consider fairness, the unfairness costs should not be neglected. In this case, the principal's new programming problem will become:

$$\max_{\{(\underline{w},\overline{w})\}} \pi_1(\overline{V}-\overline{w})+(1-\pi_1)(\underline{V}-\underline{w}-F_h)$$

$$s.t. \quad \pi_1\overline{w}+(1-\pi_1)(\underline{w}-f_h)-\Psi \geq \pi_0(\overline{w}-f_b)+(1-\pi_0)\underline{w} \qquad (8)$$

$$\pi_1\overline{w}+(1-\pi_1)(\underline{w}-f_h)-\Psi \geq 0 \qquad (9)$$

$$\underline{w} \geq 0 \qquad (10)$$

(8), (9), and (10) are the agent's incentive compatibility, participation, and limited liability constraints when fairness is considered, respectively.

Two irrelevant constraints have been neglected and omitted. For the sake of analytical simplicity, we assume that $f_b \leq \min\{\frac{\Psi+(1-\pi_1)f_h-\pi_0 f_b}{\Delta\pi}, \frac{\Psi+(1-\pi_1)f_h}{\pi_1}\}$ or $\min\{\frac{\Psi+(1-\pi_1)f_h-\pi_0 f_b}{\Delta\pi}, \frac{\Psi+(1-\pi_1)f_h}{\pi_1}\} < f_b \leq \max\{\frac{\Psi+(1-\pi_1)f_h-\pi_0 f_b}{\Delta\pi}, \frac{\Psi+(1-\pi_1)f_h}{\pi_1}\}$, which can ensure that the agent's participation constraint when he exerts no effort (namely, $\pi_0(\overline{w}-f_b)+(1-\pi_0)\underline{w} \geq 0$) will always be satisfied. That is to say, this constraint is irrelevant.

According to the standard incentive theory, it is easy for us to find that constraint (10) is binding and that constraint (8) is binding when $\Psi \geq \pi_1 f_b-(1-\pi_1)f_h$, and that constraint (9) is binding when $\Psi < \pi_1 f_b-(1-\pi_1)f_h$.

Solving this new programming problem, we obtain:

If $\Psi \geq \pi_1 f_b-(1-\pi_1)f_h$, then

$$\underline{w}^{F*}=0 \qquad (11)$$

$$\overline{w}^{F*}=\frac{\Psi+(1-\pi_1)f_h-\pi_0 f_b}{\Delta\pi} \qquad (12)$$

If $\Psi < \pi_1 f_b-(1-\pi_1)f_h$, then

$$\underline{w}^{F*}=0 \qquad (13)$$

$$\overline{w}^{F*}=\frac{\Psi+(1-\pi_1)f_h}{\pi_1} \qquad (14)$$

The superscript $F*$ stands for second-best state with the consideration of fairness. $\overline{w}^{F*}$ is the agent's efficiency wage with the consideration of fairness.

If $\Psi \geq \pi_1 f_b - (1-\pi_1) f_h$, then the principal's constraint to incentivize the agent will be:

$$\pi_1(\overline{V} - \frac{\Psi + (1-\pi_1) f_h - \pi_0 f_b}{\Delta\pi}) + (1-\pi_1)(\underline{V} - F_h)$$

$$\geq \pi_0(\overline{V} - F_b) + (1-\pi_0)\underline{V} \qquad (15)$$

It is quite obvious that (15) can be simplified into:

$$\Delta\pi\Delta V \geq \frac{\pi_1[\Psi + (1-\pi_1) f_h - \pi_0 f_b]}{\Delta\pi} + (1-\pi_1)F_h - \pi_0 F_b \qquad (16)$$

If $\Psi < \pi_1 f_b - (1-\pi_1) f_h$, then the principal's constraint to incentivize the agent will be:

$$\pi_1[\overline{V} - \frac{\Psi + (1-\pi_1) f_h}{\pi_1}] + (1-\pi_1)(\underline{V} - F_h)$$

$$\geq \pi_0(\overline{V} - F_b) + (1-\pi_0)\underline{V} \qquad (17)$$

It is quite obvious that (17) can be simplified into:

$$\Delta\pi\Delta V \geq \Psi + (1-\pi_1)(f_h + F_h) - \pi_0 F_b \qquad (18)$$

The implications of (16) and (18) are similar to that of (7) which has been stressed in Section 3.

5. A COMPARATIVE ANALYSIS

In this section, we will conduct a comparative analysis between the outcome without the consideration of fairness and that with the consideration of fairness. By comparison, it is easy for us to obtain the following four propositions.

Proposition 1: When $\Psi \geq \pi_1 f_b - (1-\pi_1) f_h$, if $f_h < \frac{\pi_0 f_b}{1-\pi_1}$, then taking into account fairness will decrease the agent's efficiency wage; if $f_h \geq \frac{\pi_0 f_b}{1-\pi_1}$, then taking into account fairness will increase the agent's efficiency wage.

Proof: If $\Psi \geq \pi_1 f_b - (1-\pi_1) f_h$ and $f_h < \frac{\pi_0 f_b}{1-\pi_1}$, then from (5) and (12), we obtain:

$$\overline{w}^{N*} - \overline{w}^{F*} = \frac{\Psi}{\Delta\pi} - \frac{\Psi + (1-\pi_1) f_h - \pi_0 f_b}{\Delta\pi} = -\frac{1-\pi_1}{\Delta\pi}(f_h - \frac{\pi_0 f_b}{1-\pi_1}) > 0.$$

If $\Psi \geq \pi_1 f_b - (1-\pi_1) f_h$ and $f_h \geq \frac{\pi_0 f_b}{1-\pi_1}$, then from (5) and (12), we obtain:

$$\overline{w}^{N*} - \overline{w}^{F*} = \frac{\Psi}{\Delta\pi} - \frac{\Psi + (1-\pi_1)f_h - \pi_0 f_b}{\Delta\pi} = -\frac{1-\pi_1}{\Delta\pi}(f_h - \frac{\pi_0 f_b}{1-\pi_1}) \le 0 . \quad \square$$

Proposition 1 implies that when the exertion cost is large enough and the the first-type unfairness cost to the agent is small enough, the consideration of fairness will decrease the agent's efficiency wage, and that when the exertion cost and the the first-type unfairness cost to the agent are both large enough, the consideration of fairness will increase the agent's efficiency wage.

Proposition 2: When $\Psi < \pi_1 f_b - (1-\pi_1) f_h$, if $\Psi \ge \frac{\Delta\pi(1-\pi_1) f_h}{\pi_0}$, then taking into account fairness will decrease the agent's efficiency wage; if $\Psi < \frac{\Delta\pi(1-\pi_1) f_h}{\pi_0}$, then taking into account fairness will increase the agent's efficiency wage.

Proof: If $\Psi < \pi_1 f_b - (1-\pi_1) f_h$ and $\Psi \ge \frac{\Delta\pi(1-\pi_1) f_h}{\pi_0}$, then from (5) and (14), we obtain:

$$\overline{w}^{N*} - \overline{w}^{F*} = \frac{\Psi}{\Delta\pi} - \frac{\Psi + (1-\pi_1) f_h}{\pi_1} = \frac{\pi_0}{\Delta\pi\pi_1}[\Psi - \frac{\Delta\pi(1-\pi_1) f_h}{\pi_0}] \ge 0 .$$

If $\Psi < \pi_1 f_b - (1-\pi_1) f_h$ and $\Psi < \frac{\Delta\pi(1-\pi_1) f_h}{\pi_0}$, then from (5) and (14), we obtain:

$$\overline{w}^{N*} - \overline{w}^{F*} = \frac{\Psi}{\Delta\pi} - \frac{\Psi + (1-\pi_1) f_h}{\pi_1} = \frac{\pi_0}{\Delta\pi\pi_1}[\Psi - \frac{\Delta\pi(1-\pi_1) f_h}{\pi_0}] < 0 . \quad \square$$

Proposition 2 implies that when the exertion cost is moderate, the consideration of fairness will decrease the agent's efficiency wage, and that when the exertion cost is small enough, the consideration of fairness will increase the agent's efficiency wage.

Proposition 3: Taking into account fairness will make the principal's constraint to incentivize the agent easier to be satisfied when $\Psi \ge \pi_1 f_b - (1-\pi_1) f_h$ and $f_h \ge \frac{\pi_0 \Delta\pi F_b - (1-\pi_1)\Delta\pi F_h + \pi_0\pi_1 f_b}{\pi_1(1-\pi_1)}$ or when $\Psi < \pi_1 f_b - (1-\pi_1) f_h$ and $\Psi \le \frac{\Delta\pi(1-\pi_1)(f_h + F_h)}{\pi_0} - \Delta\pi F_b$.

Proof: When $\Psi \ge \pi_1 f_b - (1-\pi_1) f_h$ and $f_h \ge \frac{\pi_0 \Delta\pi F_b - (1-\pi_1)\Delta\pi F_h + \pi_0\pi_1 f_b}{\pi_1(1-\pi_1)}$, then from (7) and (16), we obtain:

$$\frac{\pi_1 \Psi}{\Delta\pi} - \{\frac{\pi_1[\Psi + (1-\pi_1) f_h - \pi_0 f_b]}{\Delta\pi} + (1-\pi_1) F_h - \pi_0 F_b\}$$
$$= -\frac{\pi_1[(1-\pi_1) f_h - \pi_0 f_b]}{\Delta\pi} - (1-\pi_1) F_h + \pi_0 F_b$$

$$= -\frac{\pi_1(1-\pi_1)}{\Delta\pi}[f_h - \frac{\pi_0\Delta\pi F_b - (1-\pi_1)\Delta\pi F_h + \pi_0\pi_1 f_b}{\pi_1(1-\pi_1)}] \le 0.$$

When $\Psi < \pi_1 f_b - (1-\pi_1) f_h$ and $\Psi \le \frac{\Delta\pi(1-\pi_1)(f_h+F_h)}{\pi_0} - \Delta\pi F_b$, then from (7) and (18), we obtain:

$$\frac{\pi_1\Psi}{\Delta\pi} - [\Psi + (1-\pi_1)(f_h+F_h) - \pi_0 F_b]$$

$$= \frac{\pi_0}{\Delta\pi}\{\Psi - [\frac{\Delta\pi(1-\pi_1)(f_h+F_h)}{\pi_0} - \Delta\pi F_b]\} \le 0. \quad \square$$

Proposition 3 implies that when the exertion cost and and the first-type unfairness cost to the agent are both large enough, the principal is more likely to incentivize the agent, and that when the exertion cost is small enough, the principal is more likely to incentivize the agent.

Proposition 4: Taking into account fairness will make the principal's constraint to incentivize the agent harder to be satisfied when $\Psi \ge \pi_1 f_b - (1-\pi_1) f_h$ and $f_h < \frac{\pi_0\Delta\pi F_b - (1-\pi_1)\Delta\pi F_h + \pi_0\pi_1 f_b}{\pi_1(1-\pi_1)}$ or when $\Psi < \pi_1 f_b - (1-\pi_1) f_h$ and $\Psi > \frac{\Delta\pi(1-\pi_1)(f_h+F_h)}{\pi_0} - \Delta\pi F_b$.

Proof: When $\Psi \ge \pi_1 f_b - (1-\pi_1) f_h$ and $f_h < \frac{\pi_0\Delta\pi F_b - (1-\pi_1)\Delta\pi F_h + \pi_0\pi_1 f_b}{\pi_1(1-\pi_1)}$, then from (7) and (16), we obtain:

$$\frac{\pi_1\Psi}{\Delta\pi} - \{\frac{\pi_1[\Psi + (1-\pi_1)f_h - \pi_0 f_b]}{\Delta\pi} + (1-\pi_1)F_h - \pi_0 F_b\}$$

$$= -\frac{\pi_1[(1-\pi_1)f_h - \pi_0 f_b]}{\Delta\pi} - (1-\pi_1)F_h + \pi_0 F_b$$

$$= -\frac{\pi_1(1-\pi_1)}{\Delta\pi}[f_h - \frac{\pi_0\Delta\pi F_b - (1-\pi_1)\Delta\pi F_h + \pi_0\pi_1 f_b}{\pi_1(1-\pi_1)}] > 0.$$

When $\Psi < \pi_1 f_b - (1-\pi_1) f_h$ and $\Psi > \frac{\Delta\pi(1-\pi_1)(f_h+F_h)}{\pi_0} - \Delta\pi F_b$, then from (7) and (18), we obtain:

$$\frac{\pi_1\Psi}{\Delta\pi} - [\Psi + (1-\pi_1)(f_h+F_h) - \pi_0 F_b]$$

$$= \frac{\pi_0}{\Delta\pi}\{\Psi - [\frac{\Delta\pi(1-\pi_1)(f_h+F_h)}{\pi_0} - \Delta\pi F_b]\} > 0. \quad \square$$

Proposition 4 implies that when the exertion cost is large enough and the the first-type unfairness cost to the agent is small enough, the principal is more unlikely to incentivize the agent, and that when the exertion cost is moderate, the principal is more unlikely to

incentivize the agent.

There is a extreme case that should be stressed and discussed. The case is that there don't exist the second-type unfairness costs at all. That is to say $F_b = f_b = 0$. This special case can act as a benchmark that sets a starting point for our analyses. According to Propositions 1-4, we can find that in this case: (a) $\Psi \geq \pi_1 f_b - (1-\pi_1) f_h$ (namely, $\Psi \geq -(1-\pi_1) f_h$) and $f_h \geq \frac{\pi_0 f_b}{1-\pi_1}$ (namely, $f_h \geq 0$), and (b) $\Psi \geq \pi_1 f_b - (1-\pi_1) f_h$ (namely, $\Psi \geq -(1-\pi_1) f_h$) and $f_h \geq \frac{\pi_0 \Delta\pi F_b - (1-\pi_1)\Delta\pi F_h + \pi_0 \pi_1 f_b}{\pi_1 (1-\pi_1)}$ (namely, $f_h \geq -\frac{\Delta\pi F_h}{\pi_1}$). (a) and (b) imply that taking into account fairness will increase the agent's efficiency wage and make the principal's constraint to incentivize the agent easier to be satisfied.

6. CONCLUDING REMARKS

In this paper, we mainly discuss the effects of fairness on incentives in relation-based societies (e.g., China) through the improved principal-agent framework. Our analyses show that under some conditions the consideration of fairness will decrease the agent's efficiency wage, while under other conditions the consideration of fairness will increase the agent's efficiency wage. At the same time, our analyses show that under some conditions taking into account fairness will make the principal's constraint to incentivize the agent easier to be satisfied, while under other conditions taking into account fairness will make the principal's constraint to incentivize the agent harder to be satisfied. In summary, fairness concerns play an important role in economic relations in relation-based societies, and moral hazard problems are more subtle and difficult to be tackled when fairness is taken into account.

REFERENCES

Akerlof, G. A., Yellen, J. L., (1988), "Fairness and Unemployment", *American Economic Review*, Papers and Proceedings, 78(2): 44-49.

Alvi, E., (1998), "Fairness and Self-Interest: An Assessment", *Journal of Socio-Economics*, 27(2): 245-261.

Bereby-Meyer, Y., Niederle, M., (2005), "Fairness in Bargaining", *Journal of Economic Behavior and Organization*, 56(2): 173-186.

Bolton, G., Ockenfels, A., (2000), "ERC: A Theory of Equity, Reciprocity, and Competition", *American Economic Review*, 90(1): 166-193.

Corchón, L. C., Iturbe-Ormaetxe, I., (2001), "A Proposal to Unify Some Concepts in the Theory of Fairness", *Journal of Economic Theory*, 101(2): 540-571.

Cox, J. C., Friedman, D., Gjerstad, S., (2007), "A Tractable Model of Reciprocity and Fairness", *Games and Economic Behavior*, 59(1): 17-45.

Fehr, E., Klein, A., Schmidt, K. M., (2007), "Fairness and Contract Design", *Econometrica*, 75(1): 121-154.

Fehr, E., Schmidt, K. M., (1999), "A Theory of Fairness, Competition and Cooperation",

Quarterly Journal of Economics, 114(3): 817-868.
Fehr, E., Schmidt, K. M., (2000), "Fairness, Incentives, and Contractual Choices", *European Economic Review*, 44(4-6): 1057-1068.
Fehr, E., Schmidt, K. M., (2004), "Fairness and Incentives in a Multi-Task Principal-Agent Model", *Scandinavian Journal of Economics*, 106(3): 453-474.
Gill, D., Stone, R., (2010), "Fairness and Desert in Tournaments", *Games and Economic Behavior*, 69(2): 346-364.
Höffler, F., (1999), "Some Play Fair, Some Don't: Reciprocal Fairness in a Stylized Principal - Agent Problem", *Journal of Economic Behavior and Organization*, 38(1): 113-131.
Kaplow, L., Shavell, S., (2002), *Fairness and Welfare*, (Cambridge, Massachusetts: Harvard University Press).
Kahneman, D., Knetsch, J., Thaler, R., (1986), "Fairness as a Constraint on Profit-Seeking: Entitlements in the Market", *American Economic Review*, 76(4): 728-741.
Konow, J., (1996), "A Positive Theory of Economic Fairness", *Journal of Economic Behavior and Organization*, 31(1): 13-35.
Laffont, J.-J., Martimort, D., (2002), *The Theory of Incentives*, (New Jersey: Princeton University Press).
Morelli, M., Sacco, P. L., (1997), "Game-Theoretic Definitions of Fairness and the Contractualistic Foundations of Justice", *Research in Economics*, 51(4): 387-400.
Pi, J., (2007), "Fairness Compatibility Constraints and Collective Actions", *Frontiers of Economics in China*, 2(4): 644-652.
Pi, J., (2008), "Fairness as a Constraint", *Humanomics*, 24(3): 239-243.
Rabin, M., (1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, 83(5): 1281-1302.
von Siemens, F. A., (2009), "Bargaining under Incomplete Information, Fairness, and the Hold-Up Problem", *Journal of Economic Behavior and Organization*, 71(2): 486-494.
Trautmann, S. T., (2009), "A Tractable Model of Process Fairness under Risk", *Journal of Economic Psychology*, 30(5): 803-813.
Zajac, E. E., (1995), *Political Economy of Fairness*, (Cambridge, Massachusetts: MIT Press).

PRAVEDNOST I POTICAJI U DRUŠTVIMA ZASNOVANIMA NA RODBINSKIM VEZAMA

SAŽETAK

Rad uglavnom razmatra učinke pravednosti na poticaje u društvima zasnovanima na rodbinskim vezama (npr. Kina) u okviru odnosa principala i agenta. Naše analize prezentiraju uvjete pod kojima će razmatranje pravednosti smanjiti ili povećati agentovu stimulativnu nadnicu. Istovremeno, istraživanje prezentira uvjete pod kojima će uzimanje pravednosti u obzir otežati ili olakšati zadovoljavanje principalovih ograničenja pri poticanju agenta. Jednom rječju, ovaj rad pokazuje da poticajni učinci pravednosti ovise o okolnostima a da moralni problemi postaju istančaniji i teže rješivi kad se pravednost uzme u obzir.

***Ključne riječi**: pravednost, poticaj, principal-agent, stimulativna nadnica*

Robert Maršanić, D. Sc.[1]
Zdenka Zenzerović, D. Sc. [2]
Edna Mrnjavac, D. Sc. [3]

UDK 625.712.63(497.5 Rijeka)

APPLICATION OF THE QUEUING THEORY IN THE PLANNING OF OPTIMAL NUMBER OF SERVERS (RAMPS) IN CLOSED PARKING SYSTEMS

ABSTRACT

The principal objective of this scientific paper is to learn how to efficiently organise traffic areas and especially the size of parking capacities and hence how to ensure a quality parking service to local population and tourists as a component of the overall offer in urban and tourist destinations and how to ensure a return of investments in a reasonable period to parties investing in the parking capacity. What is the optimal capacity and how to calculate it in the best possible way by connecting parking supply and demand? This paper presents the application of the queuing theory to the planning of the optimal number of servers (ramps) in closed parking systems, since parking area can be defined as a queuing system. The illustrated model has been tested on the example of the "Delta" parking area in the City of Rijeka and the particular value of the model is its universal application. This approach has shown that by using the queuing theory, the optimal number of servers (ramps) in closed parking systems can be determined.

Keywords: *optimal number of servers (ramps), queuing theory, queuing system, planning of parking area capacity "Delta" parking area in the City of Rijeka*

1. INTRODUCTION

The demand for parking services is not constant, but it varies from lowest to highest. The span between the lowest and the highest demand and the dynamism of changes are fundamental factors which influence the required size of parking area capacities and the financial effect of parking.

The parking activity, whose objective is to "produce" parking service, is faced with primary difficulties as early as in the stage of planning: how to ensure parking service in the period of increased demand; the dimensioning of the optimal parking area capacity should be performed based on what demand; what to do with the surplus of parking area capacity when the demand is reduced; which percentage of under capacity and over capacity of the parking area is acceptable; how to balance the use of parking area capacity in conditions of fluctuation with the process of long term increase of demand and of capacity?

Besides, unlike the means of transportation which, in a case of variation larger than the planned circulation, can be rented, sold or purchased and therefore, temporarily or permanently, adjust capacity to the demand, for parking areas in the function of parking defined by location and purpose, and whose invested resources and lifetime are long term, this

[1] marsanic@rijekapromet.hr, Rijeka promet d.d., Fiumara 13, 51 000 Rijeka, Republic of Croatia
[2] zenzerov@pfri.hr , Faculty of Maritime Studies Rijeka, Studenska 2, 51 000 Rijeka, Republic of Croatia
[3] ednam@fthm.hr, Faculty of Tourism and Hospitality Management, Ika, Primorska 42, 51 410 Opatija, Republic of Croatia

is not possible. For that reason, when constructing parking locations, it is necessary to also take into account a long term forecast of parking area requirements and demands (particularly important for tourist destinations), a forecast of cyclical oscillations (arrivals of tourists in tourist season requires a larger number of parking areas as against the rest of the year) and the possibility of extensions of parking capacities according to requirements.

Since arrivals of vehicles and length of parking time can be taken as a random (stochastic) variable and the empirical distribution of these variables then approximised with adequate theoretical distributions, for parking areas it is therefore possible to apply the analytical approach, i.e. to use formulas set out by the queuing theory in order to calculate the parking functioning ratio.

The objective of this paper is to demonstrate that by applying the queuing theory the optimal number of servers (ramps) in closed parking areas can be defined and that by applying the laid out model it is possible to make adequate business decisions regarding the planning and development of parking capacities.

In the national and international scientific and professional literature the queuing theory applied to the problematics of parking has not yet been consistently researched and presented to the public, but it has only been partially analysed and treated. In recent time, parking or still traffic, as referred to by some authors, is mentioned by some Croatian authors [5, 6] as a growing problem which will eventually paralise traffic in cities and tourist centres. The issue of the modelling of parking systems is significantly less represented; parking models have been analysed by G. Luburić in 2005 [5] in his PhD dissertation entitled "*Model rješavanja problema parkiranja u gradskim središtima*" (Model for resolving the parking problem in urban centres), whilst in 1991 in the article published in the magazine "*Promet i prostor*" (Traffic and space) Ž Kerkez presents a model of dimensioning of the optimal size of parking area [3], but neither the former nor the latter paper apply the queuing theory.

2. FUNCTION OF MODEL IN THE PLANNING OF PARKING AREA CAPACITIES

Parking area is a complex system consisting of adequate system components and of their inter-dependence; it is therefore necessary prior to analysing and planning of parking capacities to define the model of parking.

The use of model in the analysis of a system has multiple advantages:

1. model enables analysing and experimenting with complex situations, not possible with a real system,
2. model enables to consider in advance most of the problem's relevant elements and to determine the inter-dependence of these elements,
3. experimenting on the model, which is an approximation of the real system, is shorter in time and cheaper than experimenting on the real model, all the more because in practice experimenting with the real system is not always possible,
4. model enables the subsequent insertion of the parameters' changes and the making of adequate business decisions for the system observed over a time.

Depending on the criterion, various types of models can be observed: according to function, structure, degree of randomness, time dependence and degree of quantification.

With relation to the object of research, this paper gives particular importance to stochastic (probabilistic) models having random model variables with known probability distributions.

The methods, i.e. procedures used to resolve the models are: analytical, simulational and heuristic methods.

Analytical methods are based on mathematical analysis procedures which produce solutions for an optimal functioning of the system. The solution of the model is given in the

form of a formula into which various values of input variables and parameters are inserted, depending on the problem content [15].

Simulation is the method which uses trials (experiments) which produce variants and based on the set criteria the variant most likely to be closest to the optimal variant is then determined.

Analytical methods and the simulation method, due to their advantages and disadvantages, complement each other. In researches of traffic and parking systems it is advisable to apply the analytical method in those cases in which it is possible to set a mathematical model for the parking area under observation; or as an addition to the research use the method of simulation to analyse certain sections of more complex models, or when it is not possible to put up a mathematical model, simulational modelling is then applied.

In this paper, since the object of research are parking areas, models of the queuing theory used for the modelling of the queuing process have been selected.

The queuing theory is one of the methods of operational researches studying the problems of waiting lines whose task is to serve randomly arrived units or requests for a service. The queuing theory uses mathematical models to determine inter-dependence among the arrival of units, their waiting to be served, their serving and finally the exit of units from the system.

Basic terms of the queuing theory are the following:

1. customers (service users, clients),
2. servers (service places, locations providing service or performing processing),
3. queue (waiting line, amassing).

Considering the elements used to describe a process of the waiting line, the necessity to denote the type of the queuing problem has arisen. For this purpose, Kendall's notation has been accepted

$$v / w / x / y / z \tag{1}$$

where the letters stand for:

v - distribution of the arrival of units into the system,
w - distribution of time of the service of units,
x - number of servers,
y - capacity of the queuing system,
z - discipline of queue.

A large number of types of queuing problems arise in practice because the six elements used to describe the queuing process can be presented in a large number of variations.

Among all the possible queuing systems, the most common in practice are systems with waiting. These are queuing problems in which customers wait to be served if their number exceeds the number of the servers or wait for service places (servers) if the number of customers is lower than the number of servers. Among queuing systems with waiting, particularly important for parking areas are waiting lines with finite waiting time, i.e. limited number of places in the waiting line.

Basic parameters used for the calculation of adequate indices of the queuing system functioning are the intensity of arrival flow (λ) and intensity of servicing (μ) which in practical examples are determined based on data obtained by statistical observation or evaluation depending on the objective and task of the research.

The application of the queuing theory is shown in section 4 of the present paper.

3. STATISTICAL ANALYSIS OF THE INTENSITY OF VEHICLE ARRIVAL AND OF THE PARKING TIME LENGTH

The arrival of vehicles into parking areas has large oscillations during the year, month, day and hour. It is therefore difficult to pre-determine the number of vehicles arriving to or leaving a certain parking area on a certain day. However, for the planning of a parking area capacity it is useful to determine whether there is a certain regularity in the arrival of vehicles, i.e. in the number of vehicles and in the time of the arrival of these vehicles into the parking area.

If the data about the number of vehicles by days and months are compared, no relation between the day under observation and the following days could be noticed. This conclusion can be verified using the correlation method, i.e. a statistical method verifying the existence and form of relation between two or more observed phenomena. Considering the fact that a high number of value pairs has been set, we have opted for the correlation for grouped elements [10].

For this purpose a correlation table is drawn up in which the denominations of groups of one characteristic are put in the table's row heading (number of vehicles of the following day – phenomenon X), and the denominations of groups of another characteristic in the column heading of the same table (number of vehicles of previous day – phenomenon Y).

In order to determine the intensity of the relation between the number of vehicles of the following day and the number of vehicles of the previous day, it is necessary to calculate the correlation coefficient according to the formula:

$$r = \frac{\sum f_{x_i y_j} \cdot X_i Y_j - N\overline{X}\,\overline{Y}}{N\sigma_x \sigma_y} \quad , \tag{2}$$

where:

N – total number of days
σ_x – standard deviation for phenomenon X (number of vehicles of following day)
σ_y – standard deviation for phenomenon Y (number of vehicles of previous day).

From the obtained value of the correlation coefficient, for which $0 < r < 1$, the dependence between the observed variables can be concluded. If r has a small value, then there is no significant dependence in the order of daily arrivals of vehicles which then means that the arrivals of vehicles can be observed as if they were independent, random in statistical sense and that the number of the vehicles arriving into closed parking area can be taken as a random variable.

Analogously a statistical analysis is carried out of the parking time length and the conclusion is drawn up about significant or random dependence between the parking time length of the following and previous day.

If the arrivals of vehicles into the parking area as well as the length of parking time are random variables, it is necessary to determine the types of intensity of these variables, i.e. to verify if these variables act according to the rules of certain theoretical distributions.

Traffic entities (further on vehicles) arrive to the service place at random time moments. In most cases these time moments are mutually independent. The notion of the arrival of vehicle can be equated with the event which is realised at the entry into the queuing system. We can therefore talk about incoming flow of event.

The arrivals of units into the system can by deterministic and in these cases it is sufficient to know the value of λ (average number of units arriving into the system in a unit of time) or t_{dol} (time period between two consecutive arrivals). However, in practice it is much more

common for arrivals of units to be random (stochastic), in which case besides λ and t_{dol} it is necessary to know their probability functions.

In real terms, an unlimited number of various distributions of arrivals is possible. In theory, they are described in distributions best suited to actual situation, but such distributions are after all approximations of actual processes.

Distributions of arrivals are most commonly classified by type of arrivals:

- regularly distributed arrivals,
- completely random arrivals (exponential distribution of arrivals),
- arrivals distributed according to Erlang distribution of order *k*,
- generally independently distributed arrivals.

Completely random arrivals (marked with *M*) refer to the units arriving into the system at random order with average time interval between two consecutive arrivals $\bar{t}_{dol}$ or with average number of units λ. As these are random arrivals, it is necessary to define the probability function. The distribution of the number of arrived units can be described using the Poisson density of distribution:

$$P_n = \frac{(\lambda t)^n}{n!} e^{-\lambda t};\ t \geq 0, n = 1,2,... \ , \tag{3}$$

and the distribution of time intervals (with *k* consecutive arrivals) with the Erlang density of distribution:

$$E_k(t) = \frac{1}{(k-1)!} \lambda^k t^{k-1} e^{-\lambda t} \quad ; \quad t \geq 0 \quad , \quad k = 1,2,... \tag{4}$$

When for the distribution of time intervals between two consecutive arrivals $k = 1$, the Erlang distribution changes to the exponential density of distribution:

$$Exp(t) = \lambda e^{-\lambda t} \quad ; \quad t \geq 0 \quad . \tag{5}$$

The servicing time is expressed in the number of time units necessary for the servicing of one unit, i.e. for the performing of a certain service. This ratio expresses the throughput of the server.

The servicing time can be analysed in a similar manner as the arrival of units, but there is a substantial difference between these two processes. The arrivals of units are distributed within the entire time interval under observation, whilst the servicing is defined only when there are units inside the system; if there is no unit, the server is unused.

The characteristics of the intensity of vehicles' arrival flow and of the time of servicing of vehicles in parking areas are the following:

▪ Stationarity is the property which shows random hesitation about the mean value. This property can also be accepted for parking systems so the intensity of vehicles' arrival flow does not depend on the time, but it is a constant value and represents the average number of vehicles arrived in a unit of time (λ).

▪ Arrivals of vehicles into the parking area are events which are consecutive one to another in moments randomly distributed in the interval under observation and they represent the incoming flow. Analogously the outgoing flow can be defined as well, i.e. the exiting of vehicles from the parking.

▪ Flows of events on parking areas are non-uniform (non-homogeneous) event flows because the demands for servicing vary according the type and the internal structure. It has been expected considering the task of the parking areas, their functioning within the traffic chain and technology of work in parking areas. However, in this paper we accept the

hypothesis about the homogeneous flow considering the fact that passenger vehicles prevail in the total number.

- In parking areas the event flows are mostly non-regular (random) because the demands for servicing do not appear according to pre-determined order.
- In relation to the time of arrival, it can be accepted that event flows in parking areas are ordinary, which means that there is a very small probability that two vehicles will appear at the same time with the same demand, and it is accepted that vehicles enter into the parking area one after the other.
- The arrival of vehicles over a certain period does not depend on the number of vehicles previously arrived into the parking area. That is why we say that the arrivals of vehicles are flows without consequence. This flow property makes sense only if the vehicles arrive from more directions, not only one, which is the commonest case on parking areas.

From the aforementioned properties (stationarity, ordinarity, flow without consequence) it is deducible that flows in parking areas are simple random flows and because of that parking areas can be analysed as a mass service system. However, it can be seen in practice that incoming and outgoing event flows do not have all of these properties; the authors [2, 12] studying the queuing theory recommend not to reject the hypothesis of a simple event flow, but to implement certain generalisations that do not influence substantially the accuracy of the obtained results.

If the number of vehicles and the length of service period are random variables, it is not possible to pre-determine their values, but it is possible to pre-determine their probability distributions.

In order to be able to calculate the probability of the realisation of the random variable representing the number of arrival of vehicles and the number of serviced vehicles, it is necessary to carry out the following: gather the data about the entering, carry out a statystical analysis of these data and to verify the matching of the empirical distribution with the selected theoretical distributions.

The grouping of data results in empirical distributions about the vehicles, the number of entry ramps and all the other information related to the production of the parking service. Basic statistical parameters are calculated for these distributions, like: arithmetic mean, standard deviation, range of variation, measures of asymmetry and roundness. If samples are in question, then it is necessary to apply the sampling method, especially interval evaluations of certain parameters and the verification of the set up hypotheses. The relation between certain variables are tested using the correlation method, linear or non-linear.

Based on the parameters calculated for the assigned empirical distribution, the procedure is continued by calculating theoretical frequencies for the selected theoretical distribution and eventually the coincision between the empirical and the theoretical distribution is tested.

The determination of the type of distribution according to which the arrivals of vehicles and the service time behave, is done by statistical tests, i.e. by testing the hypothesis about the coincision of the assumed theoretical probability distribution with the empirical probability distribution. This paper has shows the χ^2 - test.

When it is determined that the arrival of vehicles and the service time behave according to some theoretical distribution, then it is possible to apply the queuing theory for calculating the index of functioning of the assigned system depending on the distribution of arrival of vehicles and the distribution of the length of service time.

If empirical distributions of the observed variables (arrivals of vehicles and service time) cannot be adapted to any of the theoretical distributions, then, as agreed by most authors [8], it is not possible to use the analytical approach, but it is necessary to apply simulation.

4. APPLICATION OF THE QUEUING THEORY IN THE PLANNING OF PARKING CAPACITIES

The problematics of parking in urban, tourist and other centres requires an interdisciplinary approach which takes into consideration all aspects of transportation needs, preservation of urban area, environment protection and economic plausibility of possible solutions. This paper presents the application of the queuing theory in the function of optimal dimensioning of parking areas using the example of the "Delta" parking area in the City of Rijeka. The analysis of the parking area as a queuing system and the making of business decisions related to the development of parking capacities can be supported by the shown model presenting the parking area.

4.1. PARKING AREA AS A QUEUING SYSTEM

Since the arrival of the vehicles into the parking area is irregular, statistical analysis has confirmed that the arrivals of vehicles and the length of service time can be observed as random variables which can be approximised with adequate theoretical distributions. This further on means that parking areas can be analysed as mass service systems. This paper takes into consideration "closed parking systems"[1] which represent parking locations into which all entering and exiting points are equipped with certain types of ramps (arm barriers), where when entering from the incoming terminal, the driver takes the parking ticket and enters into the parking area and the payment is made in toll booths.

Parking area represents a queuing system with the following structure: customers are vehicles forming (or not) a waiting line (depending on the current situation) in order to be served (parked) in a parking section and after the service has been completed (certain length of parking time), they exit the system.

According to the characteristics of the intensity of vehicles' arrival flow and the service time listed in section 3 of the present paper, ununiformity in use of parking capacity can be deducted; if the number of vehicles arrived into the parking area is greater than the number of vehicles the existing parking capacities can serve in a unit of time, then vehicles are lined up in waiting lines, and in reverse case vehicles do not wait, but however parking capacities are not fully exploited.

Defining the optimal number of parking spaces requires the taking into consideration of all the factors influencing the work of the parking area. The optimal number of parking spaces is the one providing satisfactory level of service to the user, and at the same time has good economic effects, i.e. small number of unserved vehicles and a large number of occupied parking spaces.

For the selected system of incoming terminals into the parking area the arrivals intensity flow λ represents the average number of vehicles arriving into the parking area in a unit of time (hour, day, year etc.) under observation.

For the selected system of incoming terminals into the parking area, the intensity of servicing μ represents also the average number of vehicles which can be served in a unit of time (hour, day, year etc.) under observation. The intensity of servicing is the reciprocal value of the average time necessary for the vehicle to enter the parking and the length of service time ($\mu=1/\bar{t}_{usl}$).

Service time ($\bar{t}_{usl}$) is expressed in the number of units of time necessary for the servicing of one vehicle, i.e. for the entry of vehicles into the parking area. This data is used to express the throughpput of the server. The average time needed for the vehicle to enter into the parking area is calculated as the arithmetic mean of the total time (sum of all waiting times at entry ramps) needed for the vehicles to enter into the parking area and that time consists of: time needed for the driver to stop at the entry ramp, to press the button or to read the bar code

ticket, the time to lift the ramp and the time the driver needs to proceed to the parking space after the ramp has been lifted.

The relation between the intensity of the flow of arrivals and the intensity of the flow of servicing of vehicles is the degree of load of entry ramp $\rho=\lambda/\mu$, i.e. the coefficient of utilisation of the parking area $\rho_s=\lambda/S\mu$, where *S* stands for the number of entry ramps.

If the entry ramps are occupied, the vehicle waits in line until served. Customers are vehicles, and services are performed by servers – entry ramps. Considering the fact that the length of the waiting line is limited by the length of space in which vehicles can halt while waiting to enter the parking area, the parking area is defined as a multi-server queuing system with finite waiting line length.

From the point of view of the queuing theory, the following can be concluded for incoming/outgoing parking area terminals:

- considering that the arrival flow of vehicles is not an integral part of the system, the parking area is an open system,
- considering that more waiting lines are formed at the entry point into the parking area, we can talk about a multi-server queuing system,
- arrivals of vehicles into the parking area are distributed according to certain theoretical distributions (in this paper the distribution is Poisson's),
- service time is also distributed according to certain theoretical distributions (in this paper the distribution is exponential),
- servicing of vehicles is done according to the FIFO method (first-come-first-served),
- considering the fact that outside the entry ramps there is a certain number of space for vehicles waiting to enter the parking area, we can talk about a waiting line with limited length.

Based on the starting parameters and the characteristics of a concrete mass service system, adequate indices are calculated. These are values which express the functioning of a mass service system. According to the obtained dana, adequate conclusions are made for various number of entry ramps.

4.2. DETERMINATING PARKING CAPACITIES

Experience shows that, although the parking area has sufficient number of entry ramps, large traffic congestion creates each day at the entry into the parking. When the parking area fills up, entry ramps automatically prevent new vehicles from entering into the parking area, i.e. the drivers trying to enter are signalled that the parking area is complete and this initiates the creation of a line of vehicles trying to enter into the parking area.

A dilemma arises from the question whether the entry ramp or the parking space is a service place. Considering the fact that there are relatively poor experiences presented in national and international literature, in this paper the authors define entry ramps as service place and they base their analysis upon them; based on the obtained results they have calculated the number of parking spaces and therefore defined the required parking area capacity. It can therefore be deducted that, when the optimal size of the parking area is being defined, it is not sufficient to take into account only the entry ramps, but it is necessary to take also into account the number of parking spaces because the increase of the number of entry ramps does not mean an increase of parking area capacity.

The parking area capacity is expressed in the number of parking spaces, i.e. the number of vehicles which use the parking service. It is necessary to distinguish the statistical from the dynamic parking area capacity. Statistical capacity is expressed in the number of vehicles which can be parked at the same time, and the dynamical in the total number of vehicles counted in a unit of time.

The optimal number of entering points, i.e. ramps, according to experts in garage facilities and closed parking areas construction, amounts to one entering point per 250 parking spaces. This is the statistical parking area capacity.

The dynamical capacity is calculated taking into account the number of vehicles entering into the parking area in a day, then the average parking time length and the total working time of the parking area, so the number of the required parking spaces can be calculated using the formula

$$\Sigma_{ZM} = \lambda \times t_0 / T \tag{6}$$

where:

Σ_{PM} – average number of occupied parking spaces during the day,

λ – average number of vehicles during the day he entered the parking lot,

t – average duration of parking for one car (hours),

T – total daily work time parking (hours).

It is important to underline that once obtained optimal solution for a problem is not optimal forever and that the change in any of the mentioned elements influences a smaller or greater change of the optimal solution. Having obtained the optimal solution does not mean that there will be no waiting line, but it means that we expect this line to be the shortest possible, depending on the set out optimisation criteria: waiting time (W_Q), number of vehicles in the waiting line (L_Q), probability of cancellation (P_{otk}) or costs emerged due to waiting in line and the non-occupance of service spaces, or similar.

Since the waiting time for the vehicles waiting to enter into the parking area is not paid to the driver, the criteria for optimality for the parking area should be the probability of cancellation (P_{otk}) because if this is high, the drivers will opt for another parking area (the service system will not be congested because m spaces have been set in the waiting line, but many will be cancelled). Imagine for example $P_{otk} = 0.33$ which means that the probability is that every third vehicle will be cancelled, which is problematic and disturbing! So, according to experience and practice as well as the peculiarities of each parking location, it is necessary to evaluate the maximum tolerable value of P_{otk} and then to determine the optimal solution which this criterion meets.

Planning the dimensioning of parking area capacity by average value is an important segment in the dimensioning of the optimal parking capacity which cannot be avoided, but however the time unit can be shortened, especially if it is a seasonal phenomenon (oscillations according to months, days, hours). However, even if the planning of service places would be based on the peak load (take one day with the highest number of vehicles in the year) it does not mean that there would be no waiting lines, because the arrivals of vehicles and the parking time length are stochastic variables (discussed in section 3 of the present paper).

All of the above is the consequence of the fact that traffic (parking) service cannot be stocked like, for example a commodity and because of that it is very difficult to obtain the ideal situation. Planning capacity in traffic is one of the most difficult problems with which under-capacity or over-capacity attempts to be avoided, i.e. it is attempted to reach the maximum utilisation. One of the solutions to the problem is the implementation of a varying number of service places (ramps). But with parking areas this is not rational, however a more favourable and acceptable solution is the implementation of a system of guidance and information about free parking spaces on other locations.

4.3. Dimensioning of optimal capacity of parking area illustrated by the example of the "Delta" parking area in the City of Rijeka

The model of defining the optimal size of parking area shown in sections 3, 4.1 and 4.2 of this paper has been applied to the planning of optimal capacity of the "Delta" parking area in the City of Rijeka.

Table 1 shows the number of vehicles arrived into the "Delta" parking area in 2005.

In order to determine the intensity of relation between the number of the vehicle of the following day and the number of vehicles of the previous day, the correlation coefficient has been calculated according to (1) from section 3 of this paper.

The obtained value of the correlation coefficient $r = 0.04$ shows that there is no significant dependence in the sequence of daily arrivals of vehicles and that the hypothesis of the number of vehicles arriving into closed parking areas as a random variable can be accepted.

By doing the statistical analysis of the parking time length, we would arrive to analogous conclusion because the parking time length is also a random variable.

After such a conclusion, it can be examined whether the arrivals of vehicles and the parking time length, being random variables, behave according to some theoretical distributions.

In order to verify the aforementioned hypotheses, figure 1 shows the number of vehicles arrived into the "Delta" parking area in 2004 and 2005.

Table 1. **Number of vehicles arrived into the "Delta" parking area in 2005 by days and months**

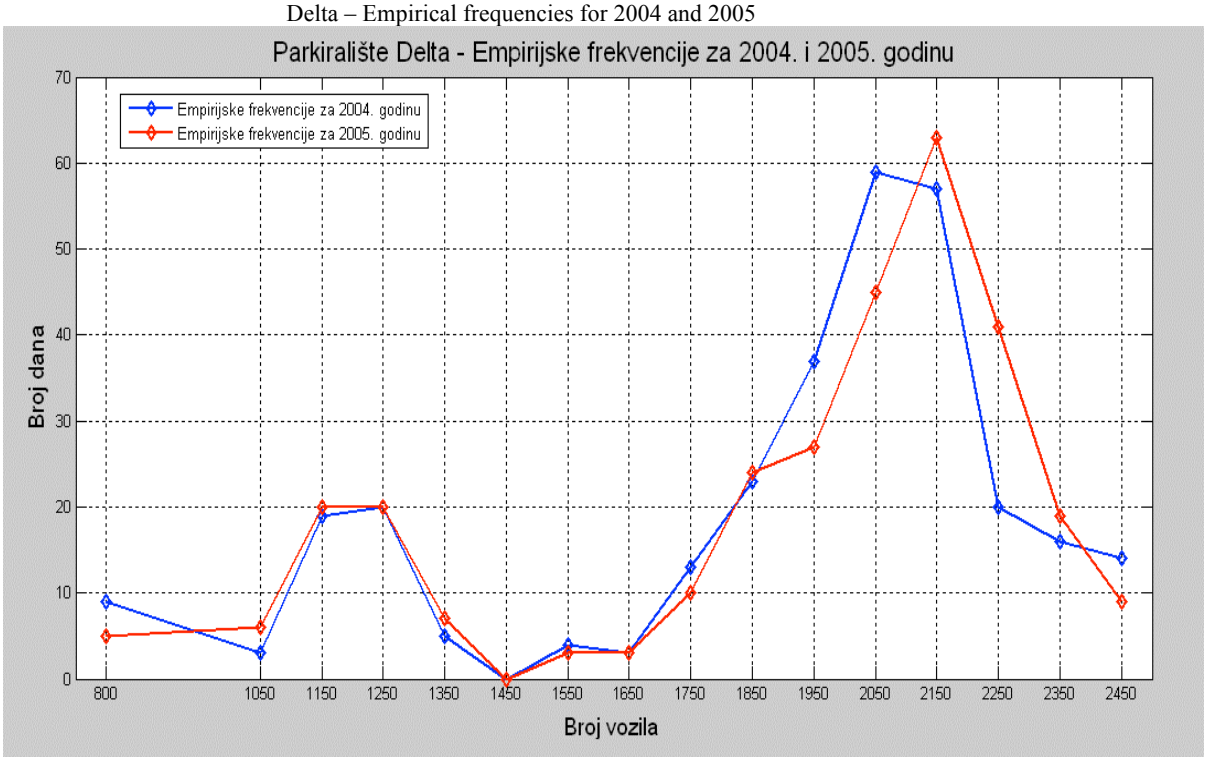
Month / Day	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Total
1.	-	1187	1859	2196	-	2189	2030	1949	2082	1189	-	1930	**16611**
2.	1158	1206	1787	1198	2279	2142	1144	1955	2260	-	2104	2093	**19326**
3.	1589	1681	1794	-	2121	2151	-	2217	1192	2130	2181	1015	**18071**
4.	1753	1824	1909	2070	2122	1225	2022	2407	-	2040	2071	-	**19443**
5.	1205	1159	1176	2012	1995	-	1837	-	2228	1874	1222	2243	**16951**
6.	-	-	-	2039	2198	2229	2116	1302	2143	1862	-	2127	**16016**
7.	1986	2174	2124	2009	1226	1829	2168	-	2040	2202	1871	2078	**21707**
8.	-	1858	1993	2098	-	2112	2220	2434	2181	-	2024	2200	**19120**
9.	1788	1825	2040	1119	2160	2037	1333	2080	2134	-	2105	2170	**20791**
10.	1821	1961	1996	-	1982	2253	-	2290	1290	2210	1998	1071	**18872**
11.	2006	2196	1980	1830	2157	1162	2305	2167	-	2119	2216	-	**20138**
12.	1119	1204	1186	2151	2280	-	2338	2440	2262	2046	1216	2005	**20247**
13.	2156	-	-	2064	2153	2238	2214	1289	2408	2050	-	2132	**18704**
14.	2259	2381	2094	2133	819	2083	1914	-	2255	2167	2201	2049	**22355**
15.	-	1884	2180	2282	-	2087	1984	-	2301	1103	2051	2165	**18037**
16.	1899	1957	2188	1266	2222	2159	1221	2442	2420	-	2099	2029	**21902**
17.	2203	2001	2258	-	2101	2126	-	2388	1399	2057	2115	1205	**19853**
18.	1877	1937	2414	2269	1980	1155	2191	1949	-	2059	2182	-	**20013**
19.	1824	1220	1339	1955	2159	-	2302	2021	2499	1956	1052	2351	**20678**
20.	1920	-	-	1873	2231	2159	1922	1233	2321	1918	-	2361	**17938**
21.	2125	1547	2196	2168	1312	2075	1859	-	2314	1853	2192	2230	**21871**
22.	1187	1608	1889	2387	-	-	2111	2479	2241	1076	2011	2270	**19259**
23.	-	1576	2141	1251	2268	2012	1277	2334	2361	-	1381	2326	**18927**
24.	1927	1846	2212	-	2322	1965	-	2228	1230	2182	2001	1109	**19022**
25.	1000	2082	1730	1945	1990	-	2122	2042	-	2069	1716	-	**16696**
26.	1080	1174	1198	2259	-	-	1785	2371	2259	2098	974	-	**15198**
27.	1223	-	-	2171	2299	1958	1798	1388	2235	2099	-	2257	**17428**
28.	1124	2137	-	2100	1183	1875	1645	-	2179	2214	2390	2053	**18900**
29.	617	-	2234	2329	-	1762	1792	2268	1867	1291	1864	1894	**17918**
30.	-	-	2157	1272	2131	2025	1025	2187	2195	-	2154	2060	**17206**
31.	1287	-	2217	-	2315	-	-	2103	-	2170	-	714	**10806**
Total	**40133**	**41625**	**50291**	**50446**	**50005**	**47008**	**48675**	**51963**	**54296**	**48034**	**47391**	**50137**	**580004**

Note: Sign "—" denotes holidays, i.e. days (Sunday and national holidays) when parking fee was not charged or months with less than 31 days.

Source: Statistical data of the "Delta" database (processed by authors)

Fig 1. **Dynamics of arrivals of vehicles into the "Delta" parking area in 2004 and 2005**

Delta – Empirical frequencies for 2004 and 2005



↑ Number of days → Number of vehicles

From figure 1 it can be noted that the number of days per number of vehicles acts very ununiform, which only corroborates the previous conclusion that the arrival of vehicles is a random variable and that the difference between the two years under observation is not significant.

However, figure 1 shows that regardless of the year, the number of days can be divided into two intervals: the first interval comprises the days with 600 to 1600 vehicles per day and the second interval are days with 1600 to 2500 vehicles per day.

It can therefore be concluded that the average number of vehicles per day will significantly vary in regard to the interval: for the first interval $\overline{X}_1 = 1{,}178$ vehicles per day and for the second 2,089 vehicles per day.

The mentioned fact points to the need for a statistical analysis of each interval distinctly, however, from the practice point of view, this causes particular problems when planning of parking area capacities is concerned. How to plan the required number of parking spaces? If we take the average number per year or the number of vehicles in the first and in the second interval, the parking area will be unutilised or on the other hand the number of parking spaces will be unsufficient in certain days of the year.

Based on the completed statistical analysis, it follows that it is essential to verify whether the assigned distribution follows the rules of a theoretical distribution, i.e. it is essential to compare the coincision of the empirical distribution and the selected theoretical distribution. The comparison has been done by applying the χ^2- test, and among the theoretical distributions, the normal and the Poisson distributions have been selected.

The calculation of theoretical frequencies depends on the selected theoretical distribution. For the Poisson distribution $f_{t_i} = N \bullet P(x)$, however, the number of vehicles is a relatively large number (around 2,000) so the calculation of theoretical frequencies has been rendered very difficult, even with the use of a computer. Statistical literature mentions that the tabulation of values for n > 20 with the Poisson distribution is not practical.

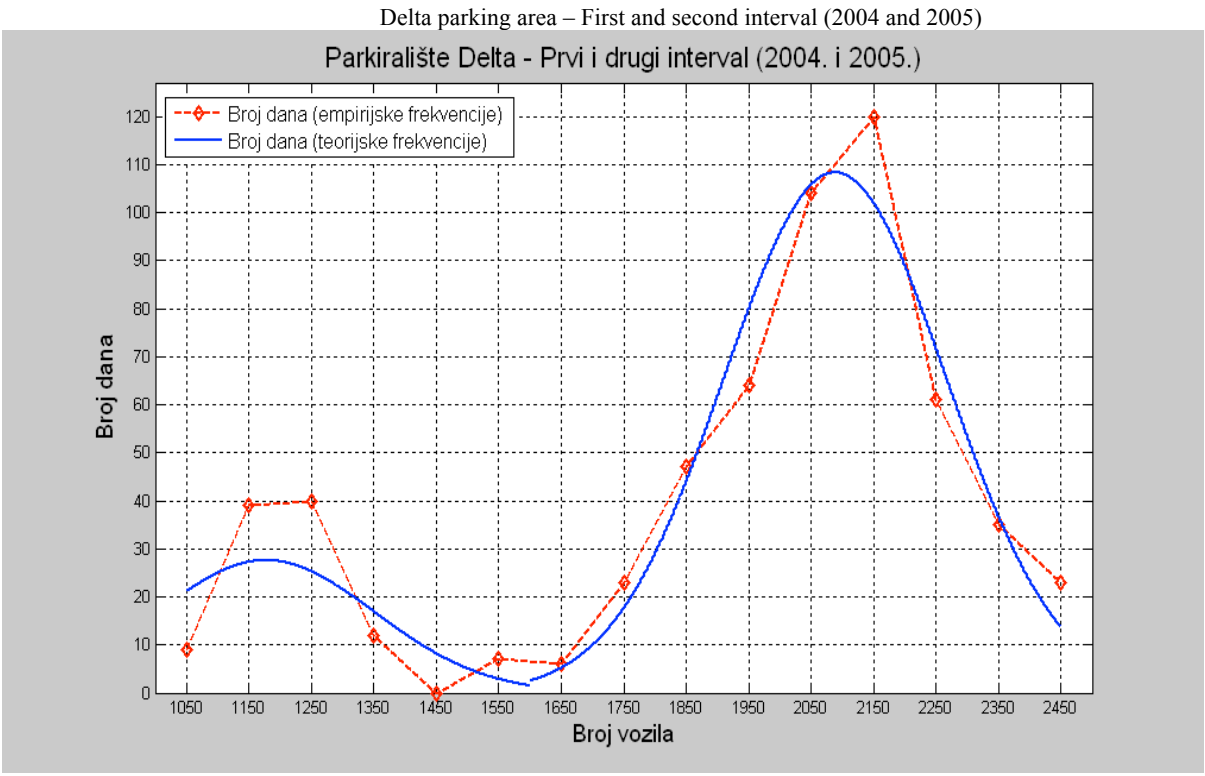
Since the Poisson distribution when $n \to \infty$ becomes less asyimmetrical and aspires towards a normal curve, in these cases $P(\lambda)$ can be approximised with normal distribution, which has been done in figure 2.

The comparison of empirical data with the normal distribution has been done distinctly for the 1st and the 2nd interval. Figure 2 shows evidently that the days in the 2nd interval adjust well to the normal distribution, but in the 1st interval this is not the case. This conclusion is corroborated by the values χ^2:

1st interval: $\chi^2 = 37.116$ $\chi^2_{0..99} = 11.341$; significant difference

2nd interval: $\chi^2 = 16.140$ $\chi^2_{0..99} = 16.812$; statistically not significant difference.

Fig 2. **Comparison of empirical and theoretical frequencies for the "Delta" parking area for 2004 and 2005**



In these cases, when empirical distributions cannot be reduced to some theoretical distributions, in theory it is advisable to apply either simulation or use of hypothesised distribution, at least as approximation of the real problem. The following authors support approximation:

- S. Vukadinović [12] states that the capacity and other characteristics of the service process depend relatively little on the form of distribution, and more on the average value (parameter λ),
- D. Gross [2] indicates that for the statistical queuing models the hypothesis that the intensity of flow of arrivals and the intensity of the serviceing follow the Poisson distribution is most often accepted.

Based on the mentioned facts, we accept the hypothesis that the distribution of days per number of vehicles on the parking behave according to the Poisson distribution.

On the basis of the above results, it follows that when defining the optimal parking area capacity the use of the queuing theory is justified.

The number of spaces in the waiting line: total length of space appointed to the waiting of the vehicles in order to be able to enter into the parking area is 80 m; if the average length of a vehicle in the waiting line is 5 m, it follows that the maximum of 16 automobiles can be present on the reserved space in one moment, i.e. $m = 16$.

Therefore, the observed servicing process is classified as a queuing problem with finite number of vehicles in the waiting line, $M/M/S/16$. Every next vehicle (17^{th} one) in the waiting line will be cancelled from the waiting line because the line of vehicles would otherwise continue on the roads intended for the circulation of motor vehicles.

The intensity of the vehicles' arrival flow: the calculation will use the average number of vehicles arriving daily into the parking area in 2005; $\lambda = 1{,}921$ vehicles per day (with 14-hour working time and 302 days a year because the rest of the days are holidays and the parking fee is not charged) or the average of 137 vehicles/hour, i.e. 302 vehicles/hour in peak hours and maximum load of the parking area.

Intensity of servicing: the intensity of servicing (μ) is obtained in the calculation as a reciprocal value of the average servicing time ($\bar{t}_{usl}$ = arithmetic mean of the servicing time); if the servicing time represents the time necessary for the driver (parking area customer) to stop its vehicle in front of the entry terminal, to take the parking ticket and to enter into the parking area and it amounts to an average of 15 s, then $\bar{t}_{usl} = 15$ s = 0.0041666 hours, and the intensity of servicing

$$\mu = 1/\bar{t}usl = 240 \text{ vehicles/hour.} \quad (7)$$

It is evident that in peak hours more vehicles arrive in a unit of time in relation to the possibility of their servicing with only one entry ramp. Based on the definition of basic parameters on the "Delta" parking area as a servicing system with finite length of the queuing line ($M/M/S/16$), we arrive to $\rho = \lambda / \mu = 302/240 = 1.25833$, and with adequate formulas [14] indices of parking area functioning have been calculated.

Based on the conducted analysis of the functioning of entry ramps on the "Delta" parking area, as a mass service system with finite length of waiting line, it is deductible that the increase of the number of entry ramps influences the increase/decrease of indices values of the parking system. Two variants have been analysed: variant A, when the average number of serviced vehicles in a day with two entry ramps is taken into account, and variant B when the maximum number of serviced vehicles in the peak hour of the occupancy (for example from 8 am to 9 am) with one, i.e. two ramps is taken into account.

All the obtained indices for the average number of vehicles, served during the day, point out that only one ramp is necessary, but the one capable of servicing all the arrived vehicles. Therefore more attention will be paid here to variant B where it is evident that a large problem exists when a high number of vehicles in peak hours want to enter, but also exit the parking system. The question arises whether the reason for the creation of waiting lines in front of the entrance points into the parking system is the insufficient number of parking spaces within the system or lower/insufficient number of service places, in this case entry ramps.

Analysing the indices for variant B and the maximal number of vehicles the following arises:

- in the system with only one entry ramp and the maximum number of vehicles in peak hour, the ramp load degree (ρ) is higher than one which imposes the conclusion that there will be the amassment of vehicles at the entering points into the parking area which will eventually result in the impossibility of a normal functioning of the system and in a "large" probability of cancellation (P_{otk}). The system with 2 entries has ρ_s smaller than

one, and therefore meets the basic requirement of the customer (vehicle) being served eventually,

- probability that there is no vehicle in the queuing system (P_0), i.e. that the capacity of the service place is inused is very small for the the system with 1 entering point, whilst the system with 2 entering points has a significantly higher probability of 22.76%,
- probability that the vehicle entering into the system will not be serviced, i.e. that it will be cancelled, in the system with 1 incoming terminal is 20.86%, whilst for the system with 2 incoming terminals it is minor,
- probability that the vehicle entering into the system will be serviced (servicing probability) in the system with 1 incoming terminal amounts to 79.14%, whilst for the system with 2 entries it aspires to be 100%,
- average number of vehicles in the waiting line for the system with 1 incoming terminal is 12 vehicles, whilst for those with 2 entries in average no vehicles need to wait in line to enter the parking area,
- average number of vehicles currently serviced is 1 – it is the vehicle entering the parking area,
- average waiting time in line for the system with 1 incoming terminal is 148 seconds (2.5 minutes) and 10 seconds for the system with 2 incoming terminals,
- average servicing time for the system with 1 incoming terminal is 11.88 seconds and 15 seconds for the system with 2 incoming terminals,
- average time inside the servicing system for the system with 1 incoming terminal is 160 seconds (2.664 minutes) and 24 seconds for the system with 2 incoming terminals.

Based on the presented it can be easily concluded that having only one entry ramp worsens considerably the quality of the servicing of vehicles with the possibility of cancellation in the system, whilst values for cases when two entry ramps are put on the entering point improves considerably the servicing quality. It can be concluded that two entry ramps is the optimal number considering the intensity of arrival of vehicles into the parking area and their servicing time at the entry of the vehicles.

If we take into consideration the fact that there is an average of 1,915 vehicles arriving daily into the "Delta" parking area and that the average length of parking is 2 hours[2] and the opening hours are 14 a day, the formula (6) shows us that the required number of parking spaces is 274. Since the actual number of parking spaces is around 500[3], the question still arises why there are long lines at the entry. Since 40% of the parking area is occupied by privileged parking ticket holders (residents and companies) we arrive to the real reason of the lack of parking space.[4] This category of users, leaving their vehicles for longer periods, is the cause of the problem.

It is therefore necessary to calculate the required number of parking spaces on the "Delta" parking area in relation to the number of vehicles entering into the parking area during the day, average time of stay in the parking and the total working hours of the parking. If the calculation takes into account the hypothesis about the mentioned percentage of occupancy of the parking area by privileged parking ticket holders (N) and the fact that λ = 1,915 and the average stay time t = 2 hours and the working hours T = 14 hours, by inserting these into the formula for the number of parking spaces (8) we arrive to the required number of parking spaces in the "Delta" parking area of 602 spaces.

$$\sum EM = \sum PM - \frac{N \times t_1}{T} \tag{8}$$

All of the above outlined leads to the conclusion that lines of vehicles forming at the entry into the mentioned parking area are not caused by poor organisation of entry ramps or their insufficient capacity, but by the fact that the City of Rijeka chronically lacks parking space, i.e. there is a disproportion between the supply and the demand for parking space.

5. CONCLUSION

The arrival of vehicles into parking areas has large oscillations during the year, month, day and hour. It is therefore difficult to pre-determine the number of vehicles arriving to or leaving a certain parking area on a certain day. However, for the planning of a parking area capacity it is useful to determine whether there is a certain regularity in the arrival of vehicles, i.e. in the number of vehicles and in the time of the arrival of these vehicles into the parking area.

Since arrivals of vehicles and length of time of their servicing can be taken as random (stochastic) variables and the empirical distribution of these variables then approximised with adequate theoretical distributions, for parking areas it is therefore possible to apply the analytical approach, i.e. to use formulas set out by the queuing theory in order to calculate the parking functioning ratio.

A dilemma arises from the question whether the entry ramp or the parking space is a service place. Considering the fact that there are relatively poor experiences presented in national and international literature, in this paper the authors define entry ramps as service place and they base their analysis upon them; based on the obtained results they have calculated the number of parking spaces and therefore defined the required parking area capacity. It can therefore be deducted that, when the optimal size of the parking area is being defined, it is not sufficient to take into account only the entry ramps, but it is necessary to take also into account the number of parking spaces because the increase of the number of entry ramps does not mean an increase of parking area capacity.

The objective of this paper is to demonstrate that by applying the queuing theory the optimal number of servers (ramps) and the required capacity (number of parking spaces) in closed parking areas can be defined. After all, the verification of the set out model of planning of optimal capacity of parking area capacity upon the actual "Delta" parking area in the City of Rijeka has shown the indisputable applicability of the results of a scientific research to actual parking area capacities. A particular merit of the model is its universal applicability because the presented methodology can be applied to any other closed parking area, i.e. parking area with ramps in current or future, changed conditions.

REFERENTIAL NOTES

[1]"Open parking system" refers to those parking areas in which the service is charged via parking machines and/or mobile phones, through the so-called *m*-parking service.

[2]Average stay on the parking (not including customers with privileged pre-paid tickets) is calculated in the following way: total income of the analysed day/number of charged parking tickets (for example 3.000,00 kuna (income of the analysed day)/475 (number of charged parking tickets) = 6.32 kuna (average income per charged parking ticket). Since parking fee of the "Delta" parking area is 4.00 kuna per hour, i.e. 8.00 kuna for two hours, we calculate that the average time of stay is 2 hours.

[3]Actual number of parking spaces on the "Delta" parking area is 458, but in author's experience, this parking area can easily accommodate around 500 vehicles.

[4]Data based on the experience of one of the authors and documentation of the company Rijeka promet.

REFERENCES

[1] Čerić, V.: **Simulacijsko modeliranje**, Školska knjiga, Zagreb, 1993.

[2] Gross, D., et. all.: **Fundamentals of Queueing Theory**, John Wiley & Sons, 1988.

[3] Kerkez, Ž.: **Model dimenzioniranja optimalne veličine parkirališta**, Promet i prostor, Zagreb, Vol. 3, 1991., 3─4.

[4] Krpan, Lj., Maršanić, R.: **Optimizacija prometnih tokova na ulazu u parkiralište primjenom teorije redova čekanja**, Hrvatsko znanstveno društvo za promet, XIII International Scientific Symposium, Opatija, 2006.
[5] Luburić, G.: **Model rješavanja problema parkiranja u gradskim središtima**, doktorska disertacija, Fakultet prometnih znanosti Zagreb, Zagreb, 2005. (neobjavljena)
[6] Maršanić, R.: **Parkiranje u turističkim destinacijama**, Rijeka, IQ Plus, 2008.
[7] Pašagić, H. et all: Analysis og vehicle arrivals at a selected intersection in the city of Zagreb, Promet – Traffic – Traffico, Vol. 10, No. 5 – 6, 1998.
[8] Petrić, J., et. all.: **Operaciona istraživanja**, Zbirka rešenih zadataka, knjiga druga, Privredno financijski vodič.
[9] Pupovac, D., Maršanić, R.: **Osnovne postavke optimizacije gradskih parkirališnih kapaciteta**, Ekonomski pregled, 61 (7-8), Zareb, 2010.
[10] Serdar, V., et. all.: **Uvod u statistiku**, Školska knjiga, 1986., Zagreb.
[11] Sikavica, P., et.all.: **Poslovno odlučivanje, Teorija i praksa donošenja odluka**, Informator, Zagreb, 1994.
[12] Vukadinović, S.: **Elementi teorije masovnog opsluživanja**, Naučna knjiga, Beograd, 1988.
[13] Zenzerović, Z.: **Optimizacijski modeli planiranja kapaciteta morskih luka**, doktorska disertacija (neobjavljena), Ekonomski fakultet u Rijeci, Rijeka, 1995.
[14] Zenzerović, Z.: **Teorija redova čekanja**, **Stohastički procesi II dio**, autorizirana predavanja, Pomorski fakultet u Rijeci, Rijeka, 2003.
[15] Žiljak, V.: **Simulacija računalom**, Školska knjiga, Zagreb, 1982.

PRIMJENA TEORIJE REDOVA ČEKANJA U PLANIRANJU OPTIMALNOG BROJA USLUŽNIH MJESTA (RAMPI) NA ZATVORENIM PARKIRALIŠNIM SUSTAVIMA

SAŽETAK

Temeljni cilj ove znanstvene rasprave je kako učinkovito organizirati prometne prostore i napose veličinu parkirališnih kapaciteta čime se domicilnom stanovništvu, ali i turistima omogućava kvalitetna parkirališna usluga kao sastavnica cjelokupne ponude gradskih i turističkih destinacija, a investitoru u parkirališni kapacitet povrat investicije u primjerenom roku. Koji je to optimalan kapacitet i kako ga izračunati na najbolji mogući način povezujući parkirališnu ponudu i potražnju ? U ovom je radu prikazana primjena teorije redova čekanja u planiranju optimalnog broja uslužnih mjesta (rampi) na zatvorenim parkirališnim sustavima, budući da se parkiralište može definirati kao sustav opsluživanja. Prikazani je model testiran na primjeru parkirališta „Delta“ u Gradu Rijeci, a posebna je vrijednost modela univerzalna primjenjivost. Takvim pristupom dokazano je da se primjenom teorije redova čekanja može utvrditi optimalan broj uslužnih mjesta (rampi) na zatvorenim parkirališnim sustavima.

Ključne riječi: *optimalan broj uslužnih mjesta (rampi), teorija redova čekanja, sustav opsluživanja, planiranje parkirališnih kapaciteta, parkiralište "Delta" u Gradu Rijeci*

Marinko Škare
Saša Stjepanović

UDK 330.101.52(497.5)

COMPUTABLE GENERAL EQUILIBRIUM MODEL FOR CROATIAN ECONOMY

ABSTRACT

Computable General Equlibrium models or CGE models, are one of the most useful models in a global development planning and macroeconomic analysis. CGE models are discovered in 1960., but there was no major development until 1978. These models have become a standard tool of empirical economic analysis. These models dominate in major part of applied econometric analysis, which is involved on problem solving in economic development and local economic policies. They are inevitable tools for analysis in international trade and government planning, changes in oil markets, and at the same time they are used in the analysis in tax reforms, welfare distributions, and even in the analysis of global warming. From that it concludes that uses of CGE models are very wide-spread. In the last few years, improvements in a specification of the model, availability of data and development in computer technology results in increased efficiency and reduced cost of analysis, which is based on CGE models. CGE models are most commonly used for analysis in countries that are in transition, but basic framework and specification of a model can be used from global to the local level.

Keywords: CGE model, General equilibrium, Croatian Economy.

1. THEORETICAL BACKGROUND OF CGE MODELS

Computable General Equilibrium (CGE) models are simulations that combine the general equilibrium structure formalized with economic data to solve numerically for the levels of supply, demand and price that support equilibrium across a specified set of markets. CGE models are a standard tool of empirical analysis, and are widely used to analyze the aggregate welfare and distributional impacts of policies. They are in use in many different areas like fiscal reform, international trade, development planning, environmental regulation, etc. CGE models are based on circular flow of commodities in a closed economy, showed in Figure 1. In Figure 1. Households own the factors of production and are the final consumers of produced commodities, and Enterprises rent the factor of production from the household for producing goods and services that the household then consumes. Many of CGE models present government that has a direct impact on the economy, but in circular flow government have the more passive role to collect taxes and to distribute revenues in a form of transfer subsidies to the firms and households. There is the two trace in circular flow. One is supply of factor inputs (labor and capital services) that household provides to firms, for exchange in goods and services that households buy from firms. And another trace route in circular flow is the payments that households receives for supply of factor inputs to the firms, which are then spent on goods and services that are produced in firms.

Equilibrium in economic flows in Figure 1. results in protection of both products and value. Protection of products means that quantity of factors that household gives and commodities that are produce by the firms must be absorbed by the firms and household in the rest of the economy. Protection of value means that for the each activity in the economy the value

of expenditure must be in balance with the value of income, and that each unit of expenditure must purchase some amount of some type of commodities. The implication of this is neither product or value can appear out of nowhere, nor can product or value disappears, which means every activity income must be balanced by other expenditure, and any transfer of purchasing power can only be effected through an opposing transfer of some positive amount of some produced good or primary factor service, and vice versa.

These accounting rules represent foundation of Walrasian General Equilibrium. Protection of product implies that firms' outputs are fully consumed by households, and that households primary factors are in turn fully employed by firms. Thus, for a given commodity the quantity produced must be equal to the sum of the quantities of that are demanded by the other firms and households in the economy. For a given factor the quantities demanded by firms must completely exhaust the aggregate supply endowed to the households. This is the familiar condition of market clearance.

Finally, the returns to households' endowments of primary factors, accrue to households as income that the households exhaust on goods purchases. The fact that households' factor endowments are fully employed, so that no amount of any factor is left idle, and that households exhaust their income, purchasing some amount of commodities---even for saving, reflects the principle of balanced budget accounting known as income balance.

The three conditions of market clearance, zero profit and income balance are employed by CGE models to solve simultaneously for the set of prices and the allocation of goods and factors that support general equilibrium. The three conditions define Walrasian General Equilibrium. Consequently, CGE models typically do not explicitly represent money as a commodity. However, in order to account for such trades the quantities of different commodities still need to be made comparable by denominating their values in some common unit of account. The flows are thus expressed in terms of the value of one commodity---the so-called numeraire good---whose price is taken as fixed. For this reason, CGE models only solve for relative prices.

2. MACROECONOMIC TRENDS IN CROATIAN ECONOMY

Croatian Economy, as well as all the transition economies during the period 1994. to 1998. performed under unstable international financial market conditions. Financial crisis from Asian markets spread to almost all emerging markets, causing increased uncertainty and investor confidence, which resulted in turning to the markets of developed countries.

The introduction of the single European currency is definitely an event that marked the end of the 1998th year. European Central Bank fixed euro exchange rate against the currencies of the Member States of the European Monetary Union and thus turned a new leaf in history. In January 1999. the financial crisis in Brazil has caused a sudden drop in the Brazilian currency, real, resulting in a significant increase in price of credits for emerging markets as it was, and the Croatian market. NATO action in Kosovo during the spring 1999th is reflected very negatively on tourism and traffic in Croatia. Croatia in 1999. year had privileged trade status with the European Union, because Croatia was not a member of CEFTA in that period and the export of Croatian products in the Central European transition countries was loaded uncompetitive tariffs.

In 2000 the negotiations were successfully completed, and Croatia joined the World Trade Organization, and also there have been concluded free trade agreements with Slovenia, Bosnia and Herzegovina and Hungary in February 2001. In October 2001 the Croatian candidature has been accepted for membership in CEFTA and signed the Agreement on Stabilization and

Association Agreement, which improved relations between the Croatia and the EU. Strong slowdown in economic activity was marked by the world economy in 2001. year. Therefore, during 2002. the recovery of the world economy was still poor with uncertain prospects. Economic trends were then characterized with the war in Afghanistan and the uncertainty of the war in Iraq and the continued recession in Japan. In this environment of stagnation in the world economy, Croatia enters into the reforms, and in 2003. year comes to a significant slowing economy by all indicators of economic activity. External debt, balance of payments deficit and fiscal deficit marked the Croatian macroeconomic environment in 2003. year.

Economic situation in 2004 years marked the growth of economic activity, reducing the budget account deficit, a slowdown in external debt and maintaining price stability. In 2005 world economic growth reached 4.9%. Despite the slowdown compared with the 2004 when it was 5.3%, such growth is considered quite strong considering that the world is ruled by the trend of rising oil prices and the trend in monetary policy restrictions. The growth of world trade was also very high and amounted to 7.4% in 2005 in relation to the trade growth in 2006 of 8.5%. The growth of the world economy in 2006 years was 5.2%, an increase of 0.3% compared to the previous year. The main generators of growth in 2006 were strong personal consumption, investment and the positive contribution of net exports. In 2006 there was acceleration of real growth in personal consumption, which is the result of positive developments on the labor market.

The growth of the world economy in 2007 amounted to 5.0% slightly slowing down since the previous year. It also slows the growth of world trade that is 6.8% in 2007. Reduction of economic activities that led to deceleration of growth are a consequence of higher prices of raw materials, turbulence in financial markets and the general deterioration in the international environment. The main generators of growth in 2007 were personal consumption and investments.

The Croatian economy is in the period from 1994 to 1998 achieved average annual growth of above 6%. Such high real rates of growth were achieved in conditions of stable prices and exchange rate movements, growth in living standards and the significant growth of savings. Growth of gross domestic product (GDP) is considered one of the most important indicators of macroeconomic trends of each economy. After five years of consecutive growth in 1999 years there were a drop asset value of GDP from -0.9%.

After strong growth in personal and public consumption since 1994 to 1998 in 1999 a drop in personal spending of 3.0% and the fall in the share of consumption in the GDP by 2.3% was registered. Growth of investment has slowed and terminated in 1999 due to the reduction of public investments, reducing liquidity and default problems in the construction sector, but also reduced imports of capital goods in 1999 years of 6.6%.

In 2000 there was a revival of economic activity, which resulted in the growth of gross domestic product of 3.8%. Stronger economic activity in 2000 was supported by a stable macroeconomic environment expressed in the retention of price stability, exchange rate stability and a stable and liquid banking system. In addition, the restrictive fiscal policy has opened the possibility of conducting a more relaxed monetary policy as reflected recovery of monetary aggregates and total sales and intensified credit activity of the banking sector.

Stronger economic activity began in 2000 and continued in 2001 supported by macroeconomic stability. This was the highest growth rate since 1997 year with the lowest inflation since 1994 year. Growth was recorded virtually in all areas of the real economy. Continuing tight fiscal policy was reflected in the reduction of government spending and reduced budget deficit. This movement of fiscal policy and increased investment activity and private

consumption was accompanied by significant growth of monetary aggregates and credit, and a fall in interest rates. Input trend of economic growth that began in 2001 continued in 2002. Subject is still a stable macroeconomic environment that was expressed in the retention price stability, exchange rate stability and a stable banking system.

In the 2004 GDP growth was 3.8%, which is slowing from 0.5% with respect to 2003. The growth of economic activity in 2004 resulted from the growth of net exports, private consumption and investment. Through 2005 and 2006 year GDP growth increases by 0.5% percentage points, while in 2007 year was 5.6%. In addition the largest contribution to growth came from personal consumption and investment.

Unemployment since 1994 to 1998 increased pursuant to the demobilization of military forces, and restructuring of former socially-owned enterprises. The speed of absorption of the surplus labor force was dependent on private sector development, as well as foreign direct investments, which were at 1998 in Croatia at an extremely low level. After the 1998 the unemployment rate is constantly increasing until 2002 when it reaches a record unemployment rate in Croatia about 22.3%. After the 2002 the unemployment rate was in slight decline, and 2007 year reached 14.8% which is also the lowest unemployment rate in Croatia since 1995 year.

Average annual inflation in the period 1994 to 1999 was at 4% annually. The exception was the year 1998 when the inflation rate amounted to 5.7% due to the introduction of value added tax. In 2000 inflation rates reached a highest value and stopped at 6.2%. In the next period rate of inflation was around 3% per annum. Largest increase of inflation rate matches with the rise in the price of crude oil and petroleum products, followed by the increase prices of product tobacco, which have significantly affected the rate of inflation.

The foreign trade of Croatia has a constant growth in balance of payments deficit, which is increasing from year to year with few year exceptions.

3. BUILDING SAM FOR CROATIA

Social accounting matrix (SAM) is a data framework that typically represents the economy of a nation. SAM is a square matrix in which each account is represented by a row and a column. Each cell shows the payment from the account of its column to the account of its row. For each account in the matrix total revenue (row total) equals total expenditure (column total). Social accounting matrix provides data for CGE models.

SAM (Social accounting matrix) is designed like generalized input - output table, and it provides a description of whole economy for the selected country. It has a form of a matrix where the number of rows is equal to the number of columns for each category of commodity, factor production and actors in the economy. Every row represents a source of origin for certain economic resources, which are connected with certain actor in the economy. Every column, opposite of rows, represents the way of use for those resources. Every actor in the economy and his behaviour must respect basic accounting principles of balance and that balance provides balance for social accounting matrix. These actors in the economy are represented by households, companies, government, etc. Exactly, these actors in the economy and their accounts describe in detail relationship between production structures and distributions of welfare, as well as financial transactions in a domestic economy, and later financial transactions with the rest of the world.

Thus, the SAM consists of a set of interrelated subsystems that, on the one hand, give an analytical picture of the studied economy in a particular accounting period and, on the other hand, serve as an instrument for assessing the effects of changes on the particular flows represented by it (injections and leakages in the system), which might be the result of policy measures.

Therefore, the SAM can be seen as a working instrument for quantifying the flows in the economic circuit and for simulating the effects resulting from any changes in such flows. This session will not develop the latter aspect (the SAM modelling).

Table 1.

Aggregate SAM model for Croatia

	Activity	*Commodity*	*Household*	*Government*	*Investment*	*World*	*Total*
Activity		**Domestic sales**				**Exports**	**Total Sales**
Commodity			**Household consumption**	**Government consumption**	**Investment**		**Total Apsorption**
Household	**GDP at factor cost**					**Foreign Remittances**	**Household Income**
Government	**Indirect Taxes**	**Tariffs**	**Income Tax**				**Government Revenue**
Savings			**Household Savings**	**Government Savings**		**Foreign Savings**	**Total Savings**
World		**Imports**					**Total Imports**
Total	**GDP Market Prices**	**Total Supply**	**Total Household Expenditure**	**Government Expenditure**	**Total Investment**	**Total Foreign Exchange**	

Social accounting matrix for Croatia, we have calculated based on data from the Statistical Yearbook, and based on the authors own calculations. It is very important to emphasize that the matrix must be in balance as it is in accounting, matrix must fulfill the principle of balance, as well as in the social accounting matrix the total sum of each column must be equal to the total sum of each row. In order to achieve this so called systemic constraints are in use. Although in practice it is impossible to calculate some parts of the matrix in such a simple way, for the models limits are used such as, for instance: saving is equal to investment. Although in practice this is not absolutely true, for the needs of making the SAM this is taken as a necessary limitation that allows us to bring the matrix into balance, sum of each row must be equal to the sum of each column.

Table 1 presents a matrix of social accounting which will be used as a basis for further work with the CGE model. As it is seen, the matrix is composed of seven rows and seven columns with the same names, which are the following: Activity, Commodity, Household, Government, Investment, World and Total, which allow the equalization of columns with rows.

Table 2.

Aggregate SAM model for Croatia for 2009

Table 2 shows the calculation of the SAM matrix for Croatia in 2009 year. It will be useful as the basis and foundation for the calculation of CGE model for Croatia. Most of the data was used from the Statistical Yearbook of the Central Bureau of Statistics. Data that are not found in the statistical yearbook are calculated by the author based on the system limitations of balance itself. These data are: foreign savings, government savings and foreign remittances.

CGE Model Framework for Croatian Economy

This Computable General Equilibrium model for Croatian economy will be built on Salter Swan framework, which gives excellent groundwork study how macroeconomic imbalance and policy adaptation effects on real sector in a small open economy.

CGE model with Salter Swan framework represents a respectable improvement against neoclassical models of exchange. These results are based on two assumptions. First assumption is that all goods and products in a model are tradable, and second one is that in a selected economy exists perfect substitutions between domestic and foreign goods. These two assumptions implicate method for determine prices for domestic goods and services on the world market.

The most important setting in Salter Swan framework is the differences between goods and services that are tradable and those that are not tradable. Non tradable goods and services are those which price is determined based on demand and supply on the domestic market. Price for goods and services that are tradable are determined on the world market. Fact that goods and services are non tradable may result from nature of goods and services (public goods or construction), or for example, there are very big transport costs for some good that there is no demand for that type of good, makes the good non tradable.

The standard Salter-Swan model is a two-sector, general equilibrium model that consists of three types of products: products that are not traded, export and import products. Country is assumed to have a small foreign trade, and therefore, faces a perfectly elastic supply from the rest of the world. In other words, it cannot affect the terms of trade and exchange of goods and services with the rest of the world. Institutional framework of Salter-Swan model of economy replicates the perfect competition with three principal participants in the market: 1) a producer who maximizes profit based on technical refinement and richness of the primary factors, 2) a consumer commodity that maximizes utility based on the overall budget constraints, and 3) the rest of the world. The balance is supposed to be a balance at the full employment level. Price factor and the product price is flexible enough to maintain the status of full employment.

The standard Salter-Swan model focuses on the effects of external shocks and the real exchange

rate that ultimately affects the direct allocation of resources in the economy. The above assumptions describe the best of both worlds: perfect competition at home and free trade abroad. Given that the country is small, all goods that can be traded are aggregated in one well. Trade balance is exogenous variable in the model, while the homogeneity of factors, combined with price flexibility allows the free market. Finally, domestic and foreign products are perfectly interchangeable in consumption.

Salter-Swan basic model refers to one country with two production sectors and three products. This model is called "1-2-3 model". Two products that country produces are: (1) an export good, E, which are sold to foreigners, and it has no domestic demand, and (2) domestic product, D, which is sold on the domestic market. The third product is import, M, which is not produced in the domestic market. In the model, there is only one consumer who receives all income. Country in the model is very small and has no impact on the world market, that meets with the fixed prices of imports and exports.

4. MATHEMATICAL CHARACTERISTICS OF A GENERAL EQUILIBRIUM MODEL FOR CROATIA

This model has three market participants: producers, households and the rest of the world. First equation defines the limit of domestic production capabilities, which gives the maximum possible combination of export goods and offers that can offer economies. The function is assumed to be concave and will be specified as constant elasticity of transformation (CET) function with elasticity of transformation. Constant in first equation define aggregate production, and it is fixed. Since the model does not have intermedia inputs, this equation is also equal to real GDP. Assume that fixed variable is equal to the assumption of full employment of all inputs of primary factors. Equation four shows the relationship between the effectiveness of domestic exports to output.

(1)

Equation one, defines the relationship between seven different prices. There are fixed world prices of exports and imports, domestic prices of exports and imports, prices of domestic products and prices of two composite products. Equation one and equation two are linearly homogeneous, and they correspond to double-pricing equations, nine and ten.

(2)

Equation two define composite products that create supply and demand for domestic products and imported goods consumed by a consumer. In multi sector models, this equation is used for multiple sectors, assuming that imports and domestic goods in the same sector are imperfect substitutes, and this version is called Armington assumption. Considering this assumption, it is considered that the composite product is expressed with a constant elasticity of a substitution (CES) aggregate function of imported and domestic products, the elasticity of substitution.

(3)

Equation three is in a model instead for many complicated systems of equations of consumption, which is located in multi sector models and showing the important feature of all complete systems of consumption. The value of goods for which there is a demand must be equal to aggregate consumption. Equation three and five use zero level in prices, duplicated all prices, for example, offers a realistic level and the desired ratio of exports and imports remains unchanged. Since only relative prices are important, it is necessary to define a value (numeraire) price, which is determined in equation eleven, and it is the exchange rate.

(4)

(5)

Consumers maximize utility, which is equivalent to maximizing the composite product in this model, and then the equation five shows the desired ratio between the imported products and domestic products as a function of relative prices.

(6)

Equation six determines income households. Equation three defines household demand for composite product. Here it should be noted that all income is spent on a composite good.
Prices:

(7)

(8)

(9)

Equation nine defines the composite price of the product, and it is a function of the cost taking into account the previous equation. Composite price of the product is equal to the GDP deflator.

(10)

The equation ten defines the cost of composite products. It is a function of costs with equation five. Price match aggregate consumer price index basket.

(11)
Equilibrium conditions:

(12)

(13)

(14)

Equations twelve, thirteen and fourteen define the market equilibrium conditions. Supply must be equal to the demand for goods and limitation in the balance of trade must be satisfied. The entire model has fourteen equations and thirteen endogenous variables. There are three conditions of equilibrium, but not the entire independent. Any of them can be dropped from the model, and pursuant to the model will be fully determined.
Identities:

(15)

(16)

(17)

To prove that the three equilibrium conditions are not independent, it is enough to show that the model satisfies Walra's law. Such a model is closed in the way that no outside or outflows of funds exists in the economy. Identities that the model satisfies are shown in equations 15, 16 and 17. The first two derive from the assumption of homogeneity, while the third stems from the fact that, in each system of equations of consumption, the value of purchase must be equal to the total consumption. When we multiply 12th equation and 13th with their respectable prices, the sum of equations 12th 13 and 14 must be zero, as identity.

1-2-3 model is different from the standard neoclassical model of exchange in which all goods, tradable or not tradable, are perfectly replaced with domestic products. Empirical model that consists of these assumptions contains "one price law", which states that tradable domestic relative prices of goods, are determined based on world prices. This model tends to extreme specialization in production and unrealistic changes in domestic prices in response to changes in trade policies and world prices. Empirically is proven that the changes in the prices of imports and exports only partially affect the price level of domestic products. In addition, this (neoclassical) model can display two-way exchange in any sector, which is usually observed at higher level aggregation. This is why the 1-2-3 model is used, in which is stated imperfect substitution and transformation. In this case, all domestic manufactured products that are exported, are effectively treated like products, which are not traded. The share of these products in the GDP is now minus one share of exports, which is extremely large, and so all sectors are treated symmetric.

5. CALIBRATION PROCEDURE, PROBLEMS AND CONSTRAINS TO USE CGE MODELS FOR SMALL OPEN ECONOMIES

The design of a CGE model requires several steps. First, the structure of the general model is determined. Then, a particular functional form has to be chosen for the production and demand functions. Usually Cobb-Douglas, Linear Expenditure System (LES) or Constant Elasticity of Substitution (CES) specifications are selected for this purpose. Finally, the parameter values for the functional forms must be derived. Ideally, all the parameters in the CGE model may be econometrically estimated, using simultaneous equation estimation methods that take into account the overall model structure. However, given the size of CGE models, the required sophistication of techniques, the identification problems and the lack of data, this procedure is considered infeasible (Gunning and Keyzer, 1995). Therefore, the most commonly used procedure to determine the parameter values is calibration (Mansur and Whalley, 1984). The calibration procedure implies that the parameters of the model are identified based on a single observation of the economy. The economy under consideration is assumed to be in equilibrium, a so-called reference equilibrium or benchmark equilibrium. In practice, the benchmark equilibrium or benchmark data set is a Social Accounting Matrix, constructed from national accounts or other governmental data sources. The calibration procedure ensures that the parameters of the model are specified in such a way that the model will reproduce the initial data

set as an equilibrium solution.

A very important characteristic of calibration is that the specification of the calibration model is not statistically tested because of the deterministic procedure by which the values do parameters from the database are derivated. That is why this approach uses a key assumption that the data from the databases represent a balance for economics that is being observed. For that a social accounting matrix is used, which is itself constructed in a way so that the data contained there are in equilibrium. The main advantage of the calibration model is the relative limitations of the data that are used in the model, or data do not need to be complete in order to construct a total model. These data can be used from a single observation, that was not the case with the econometric models. This characteristic is very important for countries in transition where there are very big limitations in data availability.

The main criticism of econometric models came from econometrics side among which are cited: the unreliability of data, the dependence of the quality of the model with the quality of data for the selected year and limited structure of the model. Unreliability of data is a very big problem with CGE models, especially because the value of the parameter is very important for later determination of the results at different simulations. The main problem here is that most data are derived from databases, although there are other data that are not in some databases, so they are being copied from some literature (most are data on the elasticity of substitution). However, the biggest problem is in those data that are not in any databases or literature. Therefore, these data are being affected based on some kind of assumptions. It is, in fact, the problem of unreliability of such data.

Another disadvantage is related to the quality of the data for the selected years, which directly affects on the quality of the whole model. In the econometric models, the stochastic distribution tries to reduce the errors in measurements of endogenous and exogenous variables in the model. However, in the calibration process assumption is the stochastic distribution is equal to zero, which then leads to that the calibration parameters must absorb all the mistakes that have been made in the data for the chosen base year. Furthermore, the social accounting matrix is not always in equilibrium, other words the sum total of the ranks is not equal to the total sum of the column, which also leads to some errors that occur in the process of bringing the matrix into balance.

Limitation of structure of the model as a third deficiency of calibration applies mainly on the functional form of the model. The main disadvantage is that the number of parameters, that are defined through the calibration, cannot be bigger than the number of equations in the model. To solve this problem, there are several solutions and one of these approaches is that parameter estimation must be based on data from more years and not a single base year.

In this section, the model is calibrated and the equilibrium model runs. Mathematical structure that is presented in the specification CGE model depends on several parameters (changes, distribution and elasticity), which must be specified so that the model numerically can be compatible with observations in the base year, which contains a matrix of social accounting. This is done in a calibration process where the values of structural parameters are expressed as a function of relevant variables in the model. When the values of the base year are used in the expression of function, and the model is solved using the values of the resulting parameters, the basic solution is being obtained.

Calibration and model initialization is done because the econometric estimation of parameters for individual equations for the whole model is usually impossible for two reasons. The first reason is the problem of identification of all the variables in the model, because the number of endogenous variables in the model is very large. Another problem of estimating

parameters of the equation is that not all data are available for all equations. For example, for consumption, data on household budget may exist for some years, but certainly there are no aggregate time series' data. That is why in the development of CGE models the two options are in use, that allow solving the above problems. First is used an econometric evaluation components by component, which are observed independently of the production and consumption, and in use are data that are available from a variety of literature. Another solution is to set the other parameters that are not familiar through the calibration model, which is also called benchmarking method.

Calibration of the model is defined as a selection of numerical values of the parameters for the function of demand and supply function. Most often very important parameters, such as in our model the elasticity of export transformation (omega_x) and elasticity of substitution of the exchange of goods and products (sigma_q), are set as exogenous variables whose value is determined based on estimations (method that will be used in the model) or based on data that are calculated in different literatures. In our model, these two parameters will be estimated by the authors of the model based on data from earlier years.

6. EVALUATING POSSIBLE EFFECTS OF THE WORLD CRISIS ON THE CROATIAN ECONOMY

Below we show the possible movement of the value of certain variables in different scenarios. First scenario shows how the export price reductions of 10% have effect on all other indicators in the economy. Second scenario considers how it will affect the increase of price of exports from 10% to value for all other variables in the economy. Third scenario is showing the changes in variables induced by reducing tariffs by 10%, while the fourth scenario shows the changes that will occur in economic variables in the case of any possible changes to increase tariffs by 10%. Based on these selected scenarios we will show the movement and reactions of certain variables in the economic model on the possible external influences.

Table 3.

Household consumption

In the case of Scenario 1, the household consumption, as seen from the table above, total consumption of households falls in the amount of 9 472 million, while in another scenario, the total consumption of households grows to 8 093 million kuna. In the third and fourth scenario, there is a little change in total household expenditure in relation to the previous scenarios. In the case of reduction of tariffs by 10%, leads to increase in total energy consumption of 115 million kuna, while in case of increasing tariffs by 10%, leads to a reduction in total spending by 114 million kuna. From this we can conclude that the movement of household consumption shows a

much greater sensitivity to changes in export rather than to changes in the amounts of custom duties.

Table 4.

Investments

The value of investments in the first scenario, falls for 553 million kuna, while in another scenario, in terms of increasing the value of exports, leads to growth in investments in the amount of 854 million kuna. In the third scenario where tariffs are reduced by 10%, leads to a reduction of investments for 116 million kuna, while the fourth scenario, total investment grew for 114 million kuna. The total value of investments responds to external changes in a similar way as the movement of total household consumption by the individual scenarios.

Table 5.

Import

Trends in the value of imports in scenario 1 are in the direction of reducing for 7 966 million kuna. In the next scenario, there is an increase in total imports for 7 777 million kuna. In the third scenario, there is an increase in demand for imports by 8 million kuna, while in the fourth scenario, leads to a decrease in demand for imports by 9 million kuna. These small values and changes are surprising considering the large and direct dependence of import with the amount of import duties.

Table 6.

Domestic sale

The value of domestic sales in the first scenario, falls in the amount of 1 666 million kuna, while in the second scenario, there is an increase of domestic sales for the amount of 1 531 million kuna. In the third scenario domestic sales is decreased for 9 million kuna, while in the fourth scenario, we have a decrease in domestic sales for 8 million kuna.

Table 7.

Saving

From Table 7 shown above, can be observed the movement of the variable saving through four listed scenarios. In the first scenario, there is a decrease of savings in the amount of 553 million kuna, while in the second scenario leads to increase of savings for 854 million kuna. In the next scenario, we have a case of reduce of saving in the amount of 116 million kuna, while in another scenario, saving is increased for 114 million kuna.

Table 8.

Export

Table 8 shows the movement of total exports considering to the observed scenarios and the changes in them. In the first scenario, is shown that in the case of reduction of price of exports

leads to an increase in total exports for 1 665 million kuna, while in case of the increase of price of exports by 10%, we have a proportional reduction of total exports of 1 531 million kuna. In scenario 3, exports is increased for 9 million kuna, while in the four scenarios exports fall for 9 million kuna. Here, can be emphasized that, although in the third and fourth scenario conditions change in tariffs, is amazing that total exports do not react to these changes and there is a very little change in total export.

Table 9.

Household income

Table 9 observe the movement of household income. Under the first scenario leads to reduced household income in the case of price reductions in the value of exports in the amount of 9 840 million kuna. In the second scenario, there is an increase in household income due to increased export prices, which represents the result of theoretical thinking in the amount of 8 407 million kuna. In the next two scenarios leads to minor changes in overall household income in the amounts of 119 and 120 million kuna. From this can be concluded that the decrease or increase in tariffs has a very little effect on total household income.

Table 10.

Goverment income

In the final table related to scenarios with different changes, we observe the movement of values of the total revenue of the state. In the first scenario, comes a decrease in state revenues in the amount of 5 598 million kuna. In the next scenario leads to the growth of government revenue in the amount of 5 415 million kuna. In scenario 3 comes to a reduction in state revenues in the amount of 166 million kuna. In the fourth scenario, there is an increase in state revenues in the amount of 164 million kuna.

As a conclusion the one fact is imposed, and that is that in the model all the variables range according to certain patterns of behavior that can be identified with theoretical postulates. The only variables that have different reactions to external stimuli are exports and imports who

show little sensitivity to changes in the values of duty, although according to the theory, such as changes should be much higher.

Conclusion

CGE models are one type of tool that can be used for the assessment of external influences on the balance of the economy of a country. One of the advantages of these models is such that they can be successfully developed in those situations where there are no complete data or even no data at all. As Croatia is a young country, and its independence and autonomy is recently acquired, so the data available are not entirely complete with some period or for some period they even don't exist. So it is just a reason why we use CGE model, where with the process of calibration, we can avoid this disadvantage, and that does not violate the integrity of the model nor its quality. For now, there is no generally accepted macroeconomic model, which will give timely decisions as his result, which eased the external influences on the economy. Successful construction of the model of general equilibrium provides quality information and answers the question of how potential external influences can affect the Croatian economy. CGE models provide one of these responses were, based on data that are collected for one calendar year in which we included in the overall economy. We construct a balanced model, which show how external shocks, positive and negative, and changes affect the overall balance. Such models not only show how these influences throw the economy out of balance, but also give an answer how to take economy back into equilibrium, or how to reallocate resources within the economy in order to achieve equilibrium. The main goal of this paper is the very essence of CGE models, and that is that the model can provide answers to questions regarding the behavior of the economy in the future. Model itself is based on data in the present. However, based on these data can be given in response to changes that may occur afterward. The way the model works is to be brought into balance with the current data and then to identify the different scenarios that could happen afterward. The results of these scenarios, showing how the economy might react to changes and show changes in each variable separately. Based on these results later, we can take decisions that will in advance prepare the economy to external shocks.

Application of research results will provide understanding of how potential external shocks may affect the Croatian economy, which will provide high-quality information base for a possible reaction to the world economy crisis. Application of research results can be off in the further development of the CGE model, by spreading it to more sectors and adding more external and internal variables. The model in this range is a simplified representation of the economy, so it is very easy on the results of a simplified model to determine the direction of the development of the model and its expansion into new sectors and new variables that have not been added to the model, partly because of the simplicity of it due to statistical limitations. Using such a model proved to be very important for making decisions just for small transition countries that have not a significant impact on the world economy and so this model is well suited just for Croatia.

BIBLIOGRAPHY

1. James C. Moore, General Equilibrium And Welfare Economics, Springer, 2007.
2. Burkard Heer, Alfred MauBner, Dynamic General Equilibrium Modelling: Computational Metods and Applications, Springer, 2005.
3. Alessandro Citanna, John Donaldson, Herakles Polemarchakis, Paolo Siconolfi, Stephen E. Spear, Essays in Dynamic General Equilibrium Teory, Springer, 2005.

4. H.Chenery, T. N. Srinivasen, Handbook of Develompent Economics vol.2, North-Holland, 1988.
5. Hans Lofgren, Rebecca Lee Harris, A Standard Computable General Equilibrium (CGE) Model in GAMS, International Food Policy Research Institute, 2002.
6. Christoper Dougherty, Introduction to Econometrics third edition, Oxford, 2007.
7. Badi H. Baltagi, Econometrics fourth edition, Springer, 2008.
8. Damodar N. Gujarati, Basic Econometrics fourth edition, The McGraw-Hill Companies, 2004.
9. Fumio Hayashi, Econometrics, Princenton University Press, 2000.
10. A. H. Studenmud, Using Econometrics and Eviews Software Package, Addison Wesley, 2001.
11. William E. Griffits, R. Carter Hill, Guay C. Lim, Using Eviews for Principles of Econometrics, Wiley, 2008.
12. George G. Judge, William E. Griffits, R. Carter Hill, Undergraduate Econometrics, Using Eviews For, Wiley, 2000.
13. Dimitrios Asteriou, Stephen G. Hall, Applied Econometrics, Palgrave Macmillan, 2007.
14. Ross M. Starr, General Equilibrium Theory, Cambridge University Press, 2001.
15. Michael Wickens, Macroeconomic Theory: A Dynamic General Equilibrium Approach, Princeton University Press, 2008.
16. Eurostat, MMF, World Bank, UN, OECD, Sustav Nacionalnih Računa 1993, Državni Zavod za Statistiku, 1993.
17. Donald Brown, Felix Kubler, Computational Aspects of General Equilibrium Theory, Springer, 2008.
18. Doina Maria Radulescu, CGE Models and Capital Income Tax Reforms, Springer, 2007.
19. Shantayanan Devarajan, Delfin S. Go, Jeffrey D. Lewis, Sherman Robinson and Pekka Sinko, Simple General Equilibrium Modeling, Cambridge University Press, 1997.
20. R. R. Johanson, A Guide to Using Eviews with Using Econometrics: A Practical guide, University of San Diego, 2000.
21. Ozge Senay, Disinflation Dynamics in an Open Economy General Equilibrium Model, University of York, 1999.
22. Sergey Paltsev, Moving from Static to Dynamic General Equilibrium Economic Models, MIT Press, 2004.
23. Ian Sue Wing, Computable General Equilibrium Models and Their Use in Economy-Wide Policy Analysis, MIT Press, 2004.
24. Benjamin H. Mitra-Kahn, Debunking the Myths of Computable General Equilibrium Models, City University, 2008.
25. Andrea Moro, Peter Norman, A General Equilibrium Model of Statistical Discrimination, University of Minnesota, 2002.
26. Channing Arndt, Sherman Robinson, Finn Trap, Parameter Estimation for Computable General Equilibrium Model: A Maximum Entropy Approach, International Food Policy Research Institute, 2001.
27. Frank Smets, Raf Wouters, An Estimated Stohastic Dynamic General Equilibrium Model of the Euro Area, European Central Bank, 2002.
28. B. Essama-Nssah, Building an Applied General Equilibrium Model, The World Bank, 2006.
29. B. Essama-Nssah, Intorduction to the Syntax of the Eviews, The World Bank, 2006.
30. Minnesota Implan Group, Elements of the Social Accounting Matrix, IMPLAN group, 2002.
31. Benjamin S. King, What is SAM? A Layman´s guide to social accounting matrices, The World Bank, 1981.

32. Susana Santos, Social Accounting Matrix and the System of National Accounts: An Application, Technical University of Lisbon, 2005.
33. B. Essama-Nessah, Building and Running General Equillibrium Models in Eviews, The World Bank, 2004.
34. http://www.dzs.hr
35. http://epp.eurostat.ec.europa.eu
36. http://www.worldbank.org/
37.http://ec.europa.eu/economy_finance/indicators/annual_macro_economic_database/ameco_en.htm
38. http://unstats.un.org/unsd/sna1993/tocLev8.asp?L1=20&L2=3

MODEL OPĆE RAVNOTEŽE REPUBLIKE HRVATSKE

Sažetak

Computable General Equilibrium modeli ili CGE modeli, odnosno modeli opće ravnoteže su jedni od najkorištenijih modela u globalnom razvojnom planiranju i makroekonomskoj analizi. CGE modeli su otkriveni 1960. godine, međutim do njihovog većeg razvoja došlo je tek nakon 1978. godine. Ti modeli dominiraju u velikom dijelu primjenjivih ekonometrijskih analiza koje se bave problemima ekonomskog razvoja i lokalne ekonomske politike. Oni su neizbježni alati za analizu u međunarodnoj razmjeni i državnom planiranju, previranja na tržištu nafte, a ujedno su i korišteni u analizama porezne reforme, distribuciji blagostanja, te u zadnje vrijeme čak i utjecaja globalnog zagrijavanja. Iz toga vidimo da je vrlo raširena primjena CGE modela. U nekoliko posljednjih godina, napredak u specifikaciji modela, dostupnosti podataka i računalnoj tehnologiji je povećao isplativost i smanjio troškove analiza koje su se temeljile na CGE modelima, što je postavilo preduvjete za sve masovniju i rašireniju upotrebu samih modela. Iako su se CGE modeli najčešće upotrebljavali kod analiza za zemlje u razvoju, osnovni okvir i obilježja modela se mogu primjenjivati i korišteni su od svjetske razine do razine pojedine države, regije, sela ili čak pojedinog kućanstva.

Ključne riječi: CGE model, Općea ravnoteža, Hrvatsko Gospodarstvo.

Yener Altunbas[a]
Ayhan Kapusuzoglu[b]

UDK 620.9:330.35>(410)"1987/2007"

THE CAUSALITY BETWEEN ENERGY CONSUMPTION AND ECONOMIC GROWTH IN UNITED KINGDOM

ABSTRACT

This study aims to examine the relationship between the energy consumption (EC) and economic growth (GDP) in the United Kingdom during the period between 1987 and 2007. Augmented Dickey-Fuller (ADF) and Philips-Perron (PP) unit root tests, the Johansen Cointegration test and standard Granger causality test were applied to examine the relationship between EC and GDP. Since the analysis results indicated no cointegration relationship between the variables of EC and GDP, it was found that there is no long-term relationship between the variables; however, in the short run, there is a unidirectional causality relationship from GDP to EC.

Keywords: *Energy Consumption, Economic Growth, Causality, United Kingdom*
JEL Classification: *C32, Q43*

1.INTRODUCTION

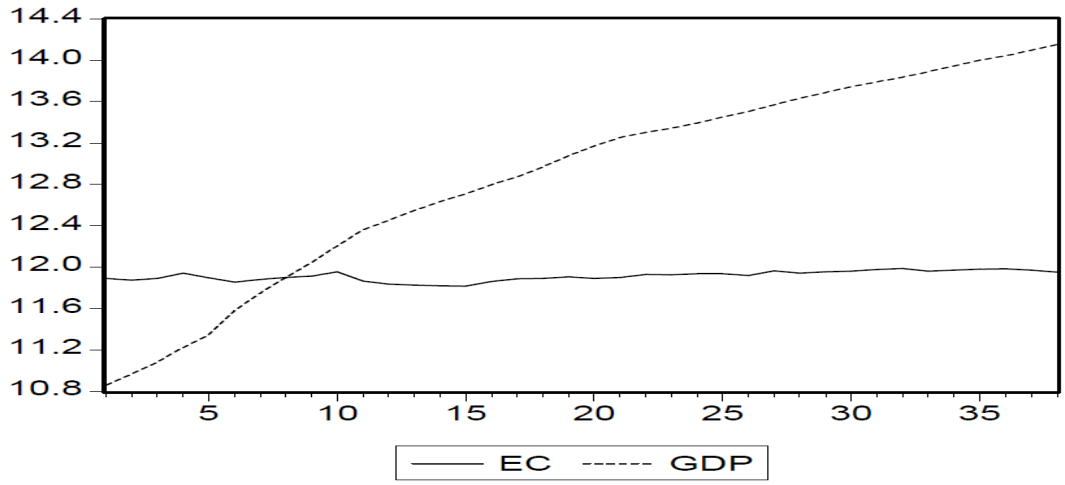
The United Kingdom is one of the countries with a highly developed economic system and thus a high income level in the world. At present, energy constitutes one of the chief factors that operate and guide economic systems. The energy factor could create a lever effect on economic growth, particularly through its contributions within the real sector. The United Kingdom is significant both for its energy use and its resources. Global energy statistics (BP, 2010) indicate that by the end of 2009, the country accounted for 0.2% of oil reserves, 1.8% of oil production, and 1.9% of oil consumption around the world; 0.2% of natural gas reserves, 2% of natural gas production, and 2.9% of natural gas consumption in the world; and 0.3% of coal production and 0.9% of coal consumption although it has no coal reserves. Figure 1 below shows the course followed by its EC and GDP between 1987 and 2007, the analysis period of the study. As an examination of Figure 1 reveals, its GDP followed an increasing trend over years, while its EC followed a stationary trend over the years, expect for some minor fluctuations.

[a]Department of Banking and Finance, Bangor University Bangor Business School, LL57 2DG, Bangor, Gwynedd, UNITED KINGDOM, Tel: +44 124 838 32 31, Fax: +44 124 838 32 28, E-mail: abs011@bangor.ac.uk

[b]Department of Business Administration, Abant Izzet Baysal University, Faculty of Economics and Administrative Sciences, 14280, Golkoy, Bolu, TURKEY, Tel: +90 374 254 10 00, Fax: +90 374 253 45 21, E-mail: ayhkap@gmail.com

Figure 1:

Changing Trends in EC and GDP Logarithmic Data (1987-2007)



The present study investigates the long and short-term relations between gross domestic product (GDP) – one of the main indicators of economic growth – and the amount of energy consumption in the United Kingdom, a country with a developed economic system. In the second part of our study, theoretical backround and findings regarding empirical studies, which can be found in the literature in relation with concerned variables, are presented. In the third part, data and methodology are presented. In the fourth part, our findings from carried out analyses are presented and there will be evaluations made in the light of obtained findings in the last part.

2.THEORETICAL BACKROUND AND LITERATURE REVIEW

The economic growth is based on two major sectors. One of them is the real sector, and the other is the financial sector. As for real sector, one of the most important sources of the real sector is energy factor and it has important impacts on real sector mechanism and thus influences the economic growth process. Interest rates and stock indexes created within the financial sector have an impact on countries' economies in different ways. Although the impacts of energy prices or energy consumption and economic growth on each other are accepted by economists to a large extent, and it is caused by the direction of causality on which this relationship is based. This issue is explained by "Ecological and Neoclassical" approaches which are contrary to each other.

The main idea of Neoclassical approach is to evaluate the economic structure as a closed system. Created products are produced through capital and workforce and the products are exchanged between enterprises and clients (Ockwell, 2008). Neoclassical growth theory takes the energy factor into consideration and this kind of approach is mainly influenced by developed intrinsic growth models, public spending (Barro, 1988), human capital (Lucas, 1988) and the studies of Neoclassical economists, Hamilton (1983) and Burbridge and Harrison (1984) (Aytac, 2010). Ecologist economists criticize the ideas presented by neoclassical approach. Ecological point of view asserts that the closed system adopted by the neoclassical approach isn't realistic and that the economic system should be taken into consideration as an open global system (Ockwell, 2008).

Concerning this significant relationship between energy consumption and economic growth, the relevant literature contains numerous studies examining the relationships between energy consumption and economic growth in the field of energy economy. Yang (2000) investigated the causality relationship between energy consumption and gross domestic product in Taiwan. It is detected that GDP has a bidirectional causality relationship with total

energy consumption, coal consumption and electricity consumption, and a unidirectional causality relationship with natural gas consumption and oil consumption.

Wolde-Rufael (2004) examined in his study the relationship between energy consumption and gross domestic product in Shanghai. A unidirectional causality relationship is detected from coal, coke, electricity and total energy consumption to GDP. However no causality relationship is detected between oil consumption and GDP.

In a study, Lee (2005) investigated the causality relationship between energy consumption and gross domestic product for 18 developing countries during the period between 1975 and 2001 by using panel cointegration and panel error correction models. Research results revealed that long- and short-term energy consumption is a unidirectional causality of GDP and that a high level of energy consumption led to an increase in the GDP level.

Chontanawat, Hunt and Pierse (2006) examined the causality relationship between energy consumption and gross domestic product in 30 OECD and 70 Non-OECD countries. Covering the period between 1970 and 2000 and causality from aggregate energy consumption to GDP and GDP to energy consumption is found to be more prevalent in the developed OECD countries compared to the developing non-OECD countries.

Zou and Chau (2006) investigated the long- and short-term relations between oil consumption and economic growth (gross domestic product) in China. Covering the period between 1953-2002 and using Granger cointegration and causality tests, the study indicated the existence of a long-term relationship between the variables in question, demonstrating that oil consumption is the causality of both short-term and long-term economic growth.

In their study, Mozumder and Marathe (2007) investigated the causality relationship between electricity consumption and gross domestic product in Bangladesh. The study covered the period between 1971 and 1999, and used Johansen-Juselius cointegration test, finding that there exists a unidirectional causality relationship from GDP to electricity consumption.

Hu and Lin (2008) examined in their study the non-linear equilibrium relationship between energy consumption (electricity, gas, coal and oil) and gross domestic product in Taiwan. Applying threshold cointegration analysis to data for the period between 1982 and 2006, the study showed that there is a long-term relationship between energy consumption and GDP, in which energy consumption growth is higher than economic growth.

In a study Belloumi (2009) examined the causality relationship between energy consumption per capita and GDP per capita in Tunisia for the period between 1971 and 2004. Using the vector error correction model and Granger causality test, the study found a long-term bi-directional and a short-term unidirectional causality relationship between the variables.

Odhiambo (2009) studied the causality relationship between energy consumption and economic growth level (gross domestic product per capita) in Tanzania. Covering the period between 1971 and 2006, the study employed the ARDL test, revealing the existence of a long-term relationship between the variables, and a unidirectional causality relationship from energy consumption to economic growth level.

In their study, Yuan, Liu, Fang and Xie (2010) investigated the relationship between economic growth (gross domestic product and added value of primary-secondary-tertiary industries) and energy consumption in China. The study examined four different period (1980-1992, 1993-1996, 1997-2000 and 2001-present day), and employed Grey analysis and Granger causality test, obtaining different results between the variables for different periods.

Arbex and Perobelli (2010) investigated the effects of economic growth (gross domestic product) upon energy consumption in Brazil. Covering the period between 1960

and 2003, the study examined 11 economic sectors in Brazil by using the economic growth model, and found that energy consumption level in each sector was closely correlated with the output growth level in corresponding sectors.

Kapusuzoglu and Karan (2010) studied the causality relationship between electricity consumption and gross domestic product in Turkey. The study covered the period between 1975 and 2006, and used Johansen-Juselius cointegration and Granger causality tests, finding that there exists a unidirectional causality relationship from GDP to electricity consumption.

In a study Quedraogo (2010) investigated the causality relationship between electricity consumption and economic growth (gross domestic product and gross capital formation) in Burkina Faso. Using the ARDL model for the period between 1968 and 2003, the study failed to find any causality relationship between electricity consumption and investments, but detected a long-term bi-directional causality relationship between electricity consumption and GDP.

3.DATA AND METHODOLOGY

The time series data employed for the empirical analysis in the study consist of the data on real gross domestic product (GDP) and aggregate energy consumption (EC) data in the United Kingdom on an annual basis for the period between 1987 and 2007. The EC (thousand tonnes of oil equivalent) data (all final users) were obtained from the Department of Energy and Climate Exchange of the United Kingdom. The GDP (in millions / U.S. Dollars) data were obtained from OECD Stat Extracts. Before starting analysis process, natural logarithm is applied on data. Table 1 presents descriptive statistics for EC and GDP series.

Table 1:

Descriptive Statistics of EC and GDP

Variables	Definition	Usable observations (1987-2007)	Mean	St. Dev.	Min.	Max.
EC[a]	Total Energy Consumption	21	11.945	0.030	11.892	11.988
GDP[b]	Total Gross Domestic Product	21	13.611	0.348	12.968	14.151

[a] thousand tonnes of oil equivalent.
[b] millions, U.S. Dollars.

The study used Granger (1969) causality test to examine the causality relationship between economic growth and energy consumption in the United Kingdom. Causality tests require variables to have the same order of stationarity. Augmented Dickey-Fuller (1979-ADF) and Philips-Perron (1988-PP) unit root tests were performed to examine the stationarity of the EC and GDP series. If these tests do not yield the same order of stationarity for the variables, they need to repeated by using the first differences. The series should have the same order of stationarity so that the cointegration relationship can be investigated between EC and GDP series.

The study investigated the presence of a long-term linear relationship (cointegration) between the series by using the test introduced by Johansen (1988) and Johansen and Juselius (1990) frequently used in investigating cointegration relations. Investigation of the cointegration relationship was based on the results of the trace and maximum eigenvalue likelihood ratio obtained from the test. In the presence of a cointegration relationship, the causality relationship is determined by Granger causality test performed in line with the

VECM model, while the absence of a cointegration relationship requires that causality relationship should be determined by the standard Granger causality test.

4.EMPIRICAL RESULTS

The results in Table 2 concern the findings about the ADF and PP test results for the EC and GDP variables in the United Kingdom. As an examination of Table 2 will clearly show, the EC and GDP variables do not have the same order of stationarity, but only become stationary when their first differences are taken to put it differently, H_0 hypothesis was rejected at the significance level of 1% after taking the first differences for both variables.

Table 2:

Results of ADF and PP Unit Root Tests

	ADF statistics		PP statistics	
	Levels	First differences	Levels	First differences
Variables				
EC	**-1.988**	**-6.112[a] (0.0001)**	**-1.967**	**-6.655[a] (0.000)**
GDP	**-2.420**	**-3.972[a] (0.0075)**	**-2.509**	**-7.319[a] (0.000)**
Critical Values				
1%	**-3.808**	**-3.831**	**-3.808**	**-3.831**
5%	**-3.020**	**-3.029**	**-3.020**	**-3.029**
10%	**-2.650**	**-2.655**	**-2.650**	**-2.655**

Each ADF and PP tests uses an intercept and no trend and lag length has been chosen basen on minimum Schwarz Info Criterion (SIC), p-values are one-sided (MacKinnon 1996).

[a] Implies significance at 1% levels, numbers in paratheses are the corresponding p-values.

Table 3 shows the results of Johansen Cointegration test performed to examine the presence of a cointegration relationship between EC and GDP. The cointegration test uses no intercept and no trend. The optimum lag length for the Johansen cointegration test was determined on the basis of the minimum AIC value obtained as a result of unconstrained VAR analysis. As demonstrated by the results of Johansen cointegration test given in Table 3, it was found that the EC and GDP variables do not have a cointegration relationship both according to the trace and maximum eigen value statistical results (H_0: r=0 not rejected at 10%, 5% and 1% levels).

Table 3:

Johansen Test for the Number of Cointegrating Relationships (EC-GDP)

H_0:H_1	Eigenvalue	Trace statistic	10% critical value	5% critical value	1% critical value
None	**0.373**	**10.905**	**13.428**	**15.494**	**19.937**
At most 1	**0.100**	**2.014**	**2.705**	**3.841**	**6.634**
H_0:H_1	Eigenvalue	Max-Eigen statistic	10% critical value	5% critical value	1% critical value
None	**0.373**	**8.891**	**12.296**	**14.264**	**18.520**
At most 1	**0.100**	**2.014**	**2.705**	**3.841**	**6.634**

r indicates the number of cointegrating relationships. The critical values for trace and max-eigen test satistics are given by Johansen and Juselius (1990). The specification for EC-GDP model includes an intercept and no trend in the cointegrating equations.

Table 4 presents the results of the Pairwise Granger causality test between the EC and GDP variables. The results indicate that there is a unidirectional causality relationship from GDP to EC at 5% significance level, while no causality relationship exists from EC to GDP. This result suggests that energy consumption is affected by economic activities and an increase in the economic growth level results in an increase in energy consumption level.

Table 4:

Pairwise Granger Causality Test

Null hypothesis	F-statistics	Implication
GDP does not Granger cause EC	**5.202 (0.035)[a]**	**GDP causing EC (significant 5%)**
EC does not Granger cause GDP	**0.522 (0.479)**	**EC not causing GDP**

[a] **Implies significance at 0.05 level, numbers in parantheses are the corresponding p-values.**

5.CONCLUSION

The present study investigated the causality relationship between energy consumption (EC) and economic growth (GDP) in the United Kingdom during the period between 1987 and 2007. Granger causality test was used to examine the causality relationship between energy consumption and economic growth, while Johansen cointegration test was performed to examine long-term cointegration relationship. As a result of these analyses, we could not reject the hypothesis that there is no long-term relationship between energy consumption and economic growth, and it was found that there is not long-term relationship between the variables. Nevertheless, we rejected the hypothesis that there is no short-term causality relationship between energy consumption and economic growth. In other words, the study found a unidirectional causality relationship between energy consumption and economic growth from GDP to EC (Mozumder and Marathe, 2007; Belloumi, 2009; Kapusuzoglu and Karan, 2010; Quedraogo, 2010). The results obtained suggest that the economic growth process in the United Kingdom will increase energy consumption in the short run and GDP falls as a result of the unfavorable trends in economic growth process will adversely affect energy consumption; however, in the long run, the changes in economic growth and energy consumption will not have significant influences upon each other. All these results indicate that the economic growth policies implemented in the United Kingdom exert their influence as one of the main factors that affect energy consumption in the short run.

According to the findings, it can be said that in the process of economic growth in United Kingdom, gross domestic product amount was an important variable which affected the energy consumption and therefore it can be said that it is important to provide the primary energy sources used in energy production and in time without any interaction for preventing the failures which may arise during the process of economic growth and for the stability of the economic production and consumption process.

REFERENCES

Arbex, M. and Perobelli, F. S., (2010), "Solow meets Leontief: Economic growth and energy consumption", *Energy Economics*, 32:43-53.

Aytac, D., (2010), "Enerji ve ekonomik büyüme ilişkisinin çok değişkenli VAR yaklaşımı ile tahmini", *Maliye Dergisi*, 158, 482-495.

Barro, R. J., (1988), "Government spending in a simple model of endogenous growth", *NBER Working Paper*, No:2588.

Belloumi, M., (2009), "Energy consumption and GDP in Tunisia: Cointegration and causality analysis", *Energy Policy*, 37:2745-2753.

BP, (2010), *BP Statistical Review of World Energy*, http://www.bp.com/statisticalreview, July 17, 2010.
Burbridge, J. and Harrison, A., (1984), "Testing for the effects of oil prices rises using vector autoregressions", *International Economic Review*, 25:459-484.
Chontanawat, J., Hunt, L. C. and Pierse, R., (2008), "Does energy consumption cause economic growth?:Evidence from a systematic study of over 100 countries", *Journal of Policy Modelling*, 30:209-220.
Dickey, D. and Fuller, W., (1979), "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, 74(366):427–431.
Granger, C. W. J., (1969), "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica*, 37(3), 424–438.
Hamilton, J. D., (1983), "Oil and the macroeconomy since World War II", *Journal of Political Economy*, 91:228-248.
Hu, J. L. and Lin, C. H., (2008), "Disaggreated energy consumption and GDP in Taiwan: A threshold co-integration analysis", *Energy Economics*, 30:2342-2358.
Johansen, S., (1988), "Statistical analysis of cointegration vectors", *Journal of Economic Dynamics and Control*, 12:231–254.
Johansen, S. and Juselius, K., (1990), "Maximum likelihood estimation and inferences on cointegration with applications to the demand for money", *Oxford Bulletin of Economics and Statistics*, 52(2):169–210.
Kapusuzoglu, A. and Karan, M. B., (2010). "An analysis of the co-integration and causality relationship between electricity consumption and gross domestic product (GDP) in the developing countries: An empirical study of Turkey", *Business and Economics Research Journal*, 1(3):57-68.
Lee, C. C., (2005), "Energy consumption and GDP in developing countries: A cointegrated panel analysis", *Energy Economics*, 27:415-427.
Lucas, R. E., (1988), "On the mechanics of economic development", *Journal of Monetary Economics*, 22:3-42.
MacKinnon, J. G., (1996), "Numerical distribution functions for unit root and cointegration tests", *Journal of Applied Economics*, 11(6):601–618.
Mozumder, P. and Marathe, A., (2007), "Causality relationship between electricity consumption and GDP in Bangladesh", *Energy Policy*, 35:395-402.
Ockwell, D. G., (2008), "Energy and economic growth: Grounding our understanding in physical reality", *Energy Policy*, 36:4600-4604.
Odhiambo, N. M., (2009), "Energy consumption and economic growth nexus in Tanzania: An ARDL bounds testing approach", *Energy Policy*, 37:617-622.
Phillips, P. C .B. and Perron, P., (1988), "Testing for unit root in the time series regression", *Biometrika*, 75(2):335–340.
Quedrago, I. M., (2010), "Electricity consumption and economic growth in Burkina Faso: A cointegration analysis", *Energy Economics*, 32:524-531.
Wolde-Rufael, Y., (2004), "Disaggreated industrial energy consumption and GDP: The case of Shanghai, 1952-1999", *Energy Economics*, 26:69-75.
Yang, H. Y., (2000), "A note on the causal relationship between energy and GDP in Taiwan", *Energy Economics*, 22:309-317.
Yuan, C., Liu, S., Fang, Z. and Xie, N., (2010), "The relation between Chinese economic development and energy consumption in the different periods", *Energy Policy*, 38:5189-5198.
Zou, G. and Chau, K. W., (2006), "Short and long run effects between oil consumption and economic growth in China", *Energy Policy*, 34:3644-3655.

KAUZALNOST IZMEĐU POTROŠNJE ENERGIJE I GOSPODARSKOG RASTA U UJEDINJENOM KRALJEVSTVU

SAŽETAK

Rad istražuje odnos između potrošnje energije (EC) i gospodarskog rasta (GDP) u Ujedinjenom Kraljevstvu tijekom perioda od 1987. i 2007. Prošireni Dickey-Fuller (ADF) i Philips-Peron (PP) testovi jediničnog korijena, Johansenov kointegracijski test i standardni Grangerov test kauzalnosti primijenjeni su kako bi se ispitao odnos između EC i GDP. Zaključeno je da ne postoji dugoročan odnos između varijabli; ipak, kratkoročno postoji jednosmjerna kauzalna veza od GDP-a prema EC.

***Ključne riječi**: Potrošnja energije, Gospodarski rast, Kauzalnost, Ujedinjeno Kraljevstvo*

***JEL klasifikacija:** C32, Q43*

Nuno Carlos LEITÃO[1] **UDK 339.5:338.45>(73)**

UNITED STATES' INTRA-INDUSTRY TRADE

ABSTRACT

The main objective of this manuscript is to explain the horizontal and vertical intra-industry trade of United States with trade partner of NAFTA, European Union and ASEAN. Identify the determinants of intra-industry trade, horizontal and vertical. Using a panel data approach, the results show a negative correlation between endowments and intra-industry trade. These results indicate that intra-industry trade occurs more frequently among countries that are similar in terms of factor endowments. The findings support the theory that, in general, there is no positive statistical association between HIIT and HO variables. Our results also confirm the hypothesis that trade increases if the transportation costs decrease.

Keywords: *Horizontal Intra-industry trade, Vertical, United States, Comparative Advantages.*

INTRODUCTION

The intra-industry trade (IIT) or two-way trade is defined as simultaneous exports and imports of a product within country or a particular industry. Nowadays in the developed world, most trade is of the IIT. Intra-industry trade is more intensive within countries and industries with similar income levels. IIT is explained by increasing return-to-scale, monopolistic competition and product differentiation. The traditional IIT models (Krugman, 1979, Lancaster, 1980) predict a negative correlation between comparative advantage and IIT. The literature of IIT emerged with Verdoorn (1960) and Balassa (1966). This phenomenon occurred in the years following the formation of the European Economic Community (EEC). However, it only started to receive increasing attention after Grubel and Lloyd (1975) had introduced an index to measure IIT.

In the 1990s new developments which occurred with a special emphasis to the European Union and central and eastern European countries (Aturupane, Djankov, and Hoekman, 1999, Kandogan, 2003). Clark (2007), Clark and Stanley (2003) Hart and McDonald (1992), and Ekanayake et al (2009) are same examples of studies on the U.S. IIT. The new trade theories (Krugman 1979, Lancaster 1980, Helpman 1981, Eaton and Kierzkowski 1984 and Helpman and Krugman 1985) explained the IIT in imperfect competition and introduce economies of scale and product differentiation.

Greenaway, Hilner and Milner (1994) and Abd-El-Rahaman (1991) introduced new types of differentiation (horizontal and vertical intra-industry). The relative unit values of exports and imports are use for separating HIIT and VIIT. This has been criticized in the literature (Zhang and Clark, 2009). This article uses the methodology of Kandogan (2003) by the fact that solving the problems associated with the dispersion of the unit value.

Horizontal intra-industry trade (HIIT) occurs within similar products. In other words, the products are differentiated by attributes (see Krugman, 1979, Lancaster, 1980, Eaton and Kierzkowski 1984, and Helpman and Krugman 1985).

[1] Ph.D., Professor of Economics, ESGTS, Polytechnic Institute of Santarém, Santarém, Portugal E-mail: nunocarlosleitao@gmail.com; Phone: (+351) 243303200

VIIT intra-industry trade (VIIT) is explained by different quality products (see Falvey, 1981, Falvey and Kierzkowski 1987, and Shaked and Sutton, 1984). This type of trade (VIIT) is also explained by the fragmentation or outsourcing (Lloyd 2004).

This manuscript analyses the determinants of intra-industry trade (IIT) and its components (HIIT and VIIT) between United States and NAFTA, European Union, and ASEAN countries in the period 1995-2008. This study uses country-specific characteristics (per capita income, market size, and geographical distance and factor endowments). The manuscript uses a panel data approach. Section 3 we present the measurement of IIT. Section 4 shows the econometric model. The final section provides conclusions.

Literature Review and Empirical Studies

The models of Krugman (1979) and Lancaster (1980) consider a monopolistic competition with increasing returns to scale. A few years later, Helpman and Krugman (1985) synthesized these type models called Chamberlin – Heckscher -Ohlin. This theoretical construction combines monopolistic competition and the theory of Heckscher - Ohlin, relating differences in factor endowments and horizontal product differentiation. In vertical intra-industry trade we can refer the contributions of Falvey and Kierzkowski (1987) and Shaked and Sutton (1984). It was concluded that the capital-abundant countries have a comparative advantage in goods of high quality and abundant in labour countries exporting low-quality goods. The vertical product differentiation means that different varieties have different types of quality. The demand is made up of consumers with different types of choice, that is, a relationship that emerges from the quality- price. On the supply side is assumed that the products (varieties) are low or high quality. The lower qualities products are labour intensive and higher quality are capital intensive. Falvey and Kierzkowski 1987) followed Linder (1961) theory. The authors consider that vertical differentiation could be explained by differences between per capita incomes. Falvey and Kierzkowski (1987) concluded that countries are capital abundant have higher productivity and higher wages. Symmetrically, the labour abundant country (low-wage country) will have comparative advantages in low-quality varieties that are labour-intensive. Flam and Helpman (1987) contains the differences in technology (labour productivity) that explain VIIT. The country with most productivity has higher wages and exports the higher quality products. In the Shaked and Sutton's article (1984), trade is studied in the context of a natural oligopoly, vertical product differentiation. The IIT is explained by different varieties of quality products (differences in income distribution: lower income country specializing in lower quality products, higher income specializing of quality products).

The study of Ekanayake (2001) examines the determinants of Mexican intra-industry trade. This study concludes that IIT is positively correlated with economic dimension (average of GDP per capita), trade intensity (openness trade), and border. Ekanayake (2001) also shows that IIT is negatively correlated with relative factor endowments and geographical distance.

Mardas and Nikas (2008) analyses the free trade areas (FTA) between the Balkan countries and Greece. The authors conclude that trade liberalization influences the increase of IIT. Recent studies found vertical IIT dominates HIIT in bilateral trade. Yoshida et al. (2009) consider the vertical intra-industry trade (VIIT) between Japan and various European countries. The authors conclude that IIT between European countries and Japan increases with their corresponding Japanese FDI (foreign direct investment), especially for new EU member countries. Ekanayake et al. (2009) analyse the vertical and horizontal IIT between the United States and NAFTA: The authors use the methodology of Greenaway et al[2]. (1994) to calculate the components of IIT, i.e HIIT (horizontal intra-industry trade) and VIIT (vertical intra-industry trade). Ekanayake et al. (2009) find that VIIT predominate in the US-NAFTA.

[2] The unit prices of exports and imports.

Chang (2009) examines the main factors of HIIT and VIIT including investment approaches of a firm in the industry of information technology for Asian, European and U.S. markets. The study uses time series data over the period of 1996-2005. Chang (2009) demonstrates that vertical intra-industry trade is significant between Asia and EU countries. According to Chang (2009) the regional agreements between EU and Association of South East Asian Nations (ASEAN) conduces the vertical specialization.

Kimura et al. (2007) analyzed parts and components trade between East Asia and Europe. This study also shows that VIIT is higher than HIIT.

Wakasugi (2007) constructed an index of vertical intra-industry trade to measure the fragmentation of production. The author used a gravity model and analyzed the impact of VIIT in East Asia, NAFTA, and European Union. Wakasugi (2007) concluded that fragmentation increased with intra-industry trade.

MEASUREMENT OF INTRA - INDUSTRY TRADE

It is usual the empirical studies using the unit prices of exports and imports to determine the horizontal intra-industry trade (HIIT) and vertical intra-industry trade (VIIT). This technique has been criticized by several authors. Most studies show that vertical intra-industry trade is inflated, when using the criterion of Greenaway Hine, and Milner (1994) or Abd-el-Rahman (1991).The present study uses the methodology of Kandogan (2003) for separating IIT into its components (HIIT) and (VIIT) intra-industry trade. Grubel and Lloyd (1975) shows that the products are similar in HIIT and products with different types of quality are VIIT. A large part of total trade (*TT*) in industry is inter-industry trade (*INT*). Kandogan´s methodology is summarized below:

$$TT_i = X_i + M_i \tag{1}$$

$$IIT_i = TT_i - |X_i - M_i| \tag{2}$$

$$INT_i = TT_i - IIT_i \tag{3}$$

$$HIIT_i = \sum \left(X_{ik} + M_{ik} - |X_{ik} - M_{ik}|\right) \tag{4}$$

$$VIIT_i = IIT_i - HIIT_i \tag{5}$$

ECONOMETRICAL MODEL

Following the literature our study applies a gravity equation with panel data. The dependent variable used is U.S. intra - industry trade (IIT_{it}), horizontal IIT ($HIIT_{it}$) and vertical IIT ($VIIT_{it}$). The data for the dependent variable is sourced from OECD at the five-digit level of the Standard International Trade classification (SITC) in US dollars. The explanatory variables are taken form World Development Indicators, the World Bank.

EXPLANATORY VARIABLES

Hypothesis 1: IIT and HIIT predominate between countries that are similar in terms of factor endowments.

Hypothesis 1(a): VIIT predominate between countries that are different in terms of factor endowments.

Economic differences between countries (DGDP): this is difference in GDP (PPP, incurrent international dollars) between U.S. and the partner country:

$$\left|GDP^{U.S.} - GDP^{partner}\right|$$

Following Helpman (1987), and Greenaway Hine, and Milner (1994) we used this proxy to evaluate relative factor endowments. Loertscher and Wolter (1980) and Balassa and Bauwens (1987) suggest a negative sign for the IIT model. Linder (1961) considers that countries with similar demands will trade similar products. Hummels and Levinshon (1995) and Greenaway et al. (1994) found a negative sign. Greenaway Hine, and Milner (1994) also

found a negative relationship between IIT, HIIT, and the difference in income per capita. It is generally agreed that the expected sign for the variable difference of income per capita is negative in the models of IIT and HIIT (Hummels and Levinshon (1995) and positive model VIIT (Greenaway, Hine and Milner1994). It should be noted that the recent study by Zhan and Clark (2009) found a negative relationship to the model VIIT for the case study of North American.

Hypothesis 2: There is a positive relationship between lowest value of GDP per capita and IIT (HIIT, and VIIT).

-MinGDP: this is the lowest value of GDP per capita (PPP, in current international dollars) between U.S. and the partner country:

$Min(GDP^{U.S.}, GDP^{partner})$

This variable is included to control for relative size effects. According to Helpman (1987) and Hummels and Levinshon (1995), a positive sign is expected, which is consistent with the hypothesis of a positive correlation between the share of IIT(HIIT, VIIT) and dissimilarity in per-capita GDP.

Hypothesis 3: There is a negative relationship between highest value of GDP per capita and IIT (HIIT, and VIIT).

MaxGDP: this is the higher/highest value of GDP per capita (PPP, in current international dollars) between U.S. and the partner country.

$Max(GDP^{U.S.}, GDP^{partner})$

This variable is also included to control for relative size effects. A negative sign is expected (Helpman 1987, Hummels and Levinshon 1995, and Greenaway Hine, and Milner (1994). A negative sign is consistent with the hypothesis that the more similar countries are in economic dimension, the greater the IIT between them.

Hypothesis 4: There is a positive relationship between comparative advantages and VIIT.

INT: this is inter-industry trade. Following the literature (Grubel and Lloyd, 1975, and Kandogan, 2003) we expected a negative sign to IIT and HIIT models, and positive to VIIT model.

Hypothesis 5: Trade increases when partners are geographically close.

DISTxDGDP: this is geographical distance multiplied by the DGDP between the U.S. and the partner country. Balassa and Bauwens (1987) argue that IIT (HIIT) will be greater when trading partners are geographically close. A longer distance will increase the transaction and transportation costs. Thus, there is a negative relationship between the share of IIT in the industry and geographical distance. Hummels and Levinshon (1995) found a negative sign.

Hypothesis 6: The foreign direct investment influences the volume of trade.

FDI (Foreign Direct Investment inflows): the relationship between IIT and the level of FDI in a particular industry is somewhat ambiguous since FDI may be a substitute for the trade. Gray (1988) considers an ambiguous relationship between FDI and IIT. Greenaway Hine, and Milner (1994) estimated a positive sign for the coefficient of this variable. Markusen (1984) considers that IIT, HIIT and VIIT shares will be positively associated with a trading partner's FDI inflows.

MODEL SPECIFICATION

$$y_{it} = \beta_0 + \beta_1 X_{it} + \delta t + \eta_i + \varepsilon_{it} \qquad (6)$$

Where y_{it} is the intra-industry trade (IIT_{it}) horizontal IIT ($HIIT_{it}$) and vertical IIT ($VIIT_{it}$), X is a set of explanatory variables. All variables are in the logarithm form; ηi is the unobserved time-invariant specific effects; δt captures a common deterministic trend; ε_{it} is a random disturbance assumed to be normal, and identical distributed (IID) with E (ε_{it})=0; Var (ε_{it})$=\sigma^2 \succ 0$.

EMPIRICAL RESULTS

Fixed effects estimates are reported in table 1. The general performance of the three equations is satisfactory.

Table 1: Fixed Effects Estimates

	IIT	HIIT	VIIT
Variables	Coefficient	Coefficient	Coefficient
LogDGDP	-2.515 (-8.925)***	-2.668 (-3.533)***	0.4978 (1.882)*
LogMinGDP	-0.128 (-1.211)	-1.185 (-3.232)***	0.425 (0.988)
LogMaxGDP	0.087 (1.785)*	0.823 (1.726)*	0.107 (4.235)***
LogFDI	0.746 (3.339)***	0.191 (2.675)***	0.654 (0.224)
LINT	-0.074 (-3.198)***	-0.485 (-5.995)***	0.9180 (26.612)***
LogDISTxDGDP	-0.286 (-4.692)***	-0.227 (-1.751)*	0.041 (0.693)
Adj. R^2	0.95	0.86	0.90
Observations	252	252	252

T-statistics (heteroskedasticity corrected) are in round brackets.

***/** - Statistically significant, respectively at the 1%, 5%, and 10% levels.

Our analysis pretends to evaluate the signs of the coefficients and their significances. For the first equation (IIT) all the explanatory variables are significant, which the exception of LogMinGDP.

The estimates obtained for the second model (HIIT) show that all explanatory variables are significant. The results are very similar to the previous model estimated. The third equation (VIIT) presents three significant variables (LogDGDP, LogMaxGDP, and LogINT).

Factor endowment differences (LogDGDP) are consistent with theoretical predictions, i.e, a negative impact on IIT and HIIT, and a positive influence on VIIT. Helpman and Krugman (1985) found a negative sign. Zhang and Clark (2009) also found a negative sign for IIT and HIIT to the North American case. Greenaway, Hine, and Milner (1994) expect a positive relationship between the VIIT and differences in relative factor endowments. Falvey and Kierzkowski (1987) suggest that this relationship will be positive for the VIIT model.

Following Helpman and Krugman (1985) and Hummels and Levinsohn (1995), the study also includes two variables to control for relative size effects. Our results are not according to the theoretical models.

The foreign direct investment (LogFDI) is an important determinant. As in Greenaway, Hine and Milner (1994) we find a positive correlation between FDI and IIT, and HIIT.

The index of inter-industry trade (INT) has a significant effect on IIT and HIIT. This result is according to the literature (Grubel and Lloyd 1975, and Kandogan 2003). Lloyd (2004) refers that vertical IIT is explained by comparative advantages; i.e there is a positive relationship between INT and vertical IIT. This is in accordance with the neo-Heckscher-Ohlin trade theory, which also explains VIIT by comparative advantages.

The geographical distance multiplied by the difference between per-capita incomes (LogDISTxDGDP), has been used a typical gravity model variable. A negative effect of distance on bilateral IIT, HIIT and VIIT was expected and the results confirm this to IIT and HIIT models, underling the importance of neighbour partnerships. Zhang and Clark (2009) also found the same results to U.S. experience. The regression results suggest the influence of U.S. free trade agreements (NAFTA). The predominance of HIIT shows that United States has trade relations with more intensity with Canada and Mexico, which is geographically closer.

CONCLUSIONS

The objective of this manuscript was to analyze the country-specific determinants of intra-industry trade for U.S. The hypotheses put forward in regard to common country characteristics are generally confirmed by the empirical results. Our results are robust with static and dynamic panel data.

The variable (LogDGDP) used to evaluate the relative factor endowments presents a negative impact on IIT, and HIIT, when we used fixed. These results are according to the literature (Loertscher and Wolter, 1980). The study of Zhang and Clark (2009) also found a negative sign to U.S. experience. IIT occurs more frequently among countries that are similar in terms of factor endowment.

We find a positive correlation between LogDGDP and vertical IIT. Our results show that the higher the difference in GDP per capita between U.S. and trade partner, the higher will be VIIT.

The variable foreign direct investment (FDI) is according to the literature, i.e, there is a positive relationship between FDI and IIT. Markusen (1984), Heplaman and Krugman (1995), Greenaway et al. (1994) found a positive sign. The results show that FDI and trade are complementary.

The variable (LogINT) used to analyses the inter-industry (Grubel and Lloyd 1975, and Kandogan 2003) is according to the literature; the inter-industry trade presents a negative correlation with IIT and HIIT.

Vertical IIT is explained by comparative advantages; i.e there is a positive relationship between INT and vertical IIT (Lloyd 2004). This is in accordance with the neo-Heckscher-Ohlin trade theory, which also explains VIIT by comparative advantages.

According to the literature we expected a negative sign to geographical distance. It is usual that the literature attributes a negative sign to geographical distance, i.e. trade increases if the partners are geographically close. The findings support this hypothesis.

In the further, we need to improve research on vertical integration (fragmentation), because this phenomenon is associated to two-trade of different endowments quality products.

REFERENCES

Abel-el-Rahaman, K. (1991), "Firms Competitive and National Comparative Advantages as Joint Determinants of Trade Composition", *Weltwritschaftliches Archiv*, 127, (1) :83-97.

Aturupane, C., Djankov, S. and Hoekman, B. (1999), "Horizontal and vertical intra industry trade between eastern Europe and the European Union", *Weltwirtschaftliches Archiv*, 135: 62-81.

Anderson, J. (1979), "A Theoretical Foundation for the Gravity Equation", *American Economic Review,* 69, (1): 106-116.

Badinger, H. and Breuss, F. (2008), "Trade and Productivity: An industry perspective", *Empirica***,** 35: .213-231.

Balassa, B. (1966),"Tariff Reductions and Trade in manufactures among the industrial countries, *American Economic Review"*, 56: 466-473.

Balassa, B, and Bauwens, L. (1987), "Intra-Industry Trade Specialization in a Multi-Country and Multi- Industry Framework" , *Economic Journal* 97, (388): 923-939.

Clark D. (2006),"Country and industry-level determinants of vertical specialization based trade", *International Economic Journal,* 2: 211–225.

Clark, D. (2007), " Intra-industry Specialization and the United States-Central America-Dominican Republic Free Trade Agreement" , *The Developing Economies* 45 (4): 491–506.

Chang, Shun-Chiao. (2009), "Horizontal and Vertical Intra-industry Trade and Firm's Investment Strategies: Evidence from the IT Industry in the Asian, EU and US", Markets, *Global Economic Review*, 38: 63-76.

Davis, D. R. (1995), "Intra-Industry Trade: A Heckscher-Ohlin-Ricardo Approach", J*ournal of International Economics*, 39 (3, 4): 201-226.

Eaton, J. and Kierzkowski, H. (1984), Oligopolistic competition, product variety, and international trade, in H. Kierzkowski, (eds), *Monopolistic Competition and International Trade*, New York [etc.] : Oxford University Press.

Ekanayake E.M. (2001), "Determinants of Intra-industry trade: The Case of México", *The International Trade Journal,* XV, (1): 89-112.

Ekanayake E.M, Veeramacheneni, B., and Moslares, C. (2009), "Vertical and Horizontal Intra-Industry Trade Between U.S. and NAFTA Partners", *Revista de Análisi Económico*, 24 (1): 21-42.

Falvey, R. E. (1981), "Commercial Policy and Intra-Industry Trade", Journal of *International Economics*, 11, (4), :494-511.

Falvey, R. E., and Kierzkowski, H. (1987), Product Quality, Intra-Industry Trade and (Im)perfect Competition. in H. Kierzkowski (eds), *Protection and Competition in International Trade*, Oxford, UK: Basil Blackwell.

Gray H. (1988), "Intra-Industry Trade: An Untidy Phenomenon", *Weltwirtschaftliches Archiv*, (124), : 17-134.

Greenway, D, Hine, R. and, Milner, C. (1994), "Country-specific factors and the pattern of horizontal and vertical intra-industry trade in UK", *Review of World Economic*, 130, (1), : 77-100.

Grubel, H. and Lloyd, P.(1975), *Intra-industry trade: the theory and measurement of international trade in different products*, London: Macmillan.

Gross M. and E. Helpman (2005)," Outsourcing in a Global Economy", *Review of Economic Studies*, 72:135-159.

Hart, T. and McDonald B. (1992), Intra-Industry Trade Indexes for Canada, Mexico, and the United States 1962-87, *Agricultural and Trade Analysis Division, Economic Research Service, U.S. Department of Agriculture, Staff Report* AGES 9206.

Helpman, E. (1981), " International trade in the presence of product differentiation, economies of scales and monopolistic competition: a Chamberlin-Heckscher-Ohlin Approach", *Journal of Political Economy*, (92) :451-71.

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. (1985), *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition and the International Economy*, 1st ed; London : MIT Press.

Hummels, D. and Levinsohn, J. (1995)," Monopolistic competition and international trade: reconsidering the evidence", *Quarterly Journal of Economics*, 110, (3) : 799-836.

Kandogan, Y. (2003), "Intra-Industry Trade of Transition Countries: Trends and Determinants", *Emerging Markets Review*.4, (3): 272-286.

Kimura, F., Takahashi, Y. and Hayakawa, K. (2007), " Fragmentation and parts and components trade: Comparison between East Asia and Europe", *The North American Journal of Economics and Finance* 18(1): 23–40.

Krugman, P. (1979), "Increasing returns, monopolistic competition and international trade", *Journal of International Economics*,. 9 , (4) :469-479.

Lancaster, K. (1980), "Intra-industry trade under perfect monopolistic competition", *Journal of International Economics*, 10, (2) : 151-175.

Linder, S. B. (1961), *An Eassy on Trade and Transformation*, New York: John Wiley.

Loertscher, R. and Wolter, F.(1980), " Determinants of intra-industry trade among countries and across industries" , *Review of World Economics*, 116,(.2): 280-293.

Lloyd, P., (2004), "How Intra-Industry Trade Changed Our Perception of the World Economy", *The Singapore Economic Review*, 49,(11): 1-17.

Mardas, D. and Nikas, C, (2008), "European Integration, Intra-Industry Trade in Vertically Differentiated Products and the Balkan Countries", *International Advances in Economic Research,* 14: 355-368.

Markusen, J. R. (1994), "Multinational Multiplant Economies and the Gains from Trade", *Journal of International Economics*, 16, (3,4),205-226.

Shaked, A. and Sutton, J (1984). Natural oligopolies and international trade, in H. Kierzkowski,. (eds.). *Monopolistic competition and international trade*, Oxford University Press.

Wakasugi, R. (2007), " Vertical Intra-industry Trade and Economic Integration in East Asia", *Asian Economic Papers,* (6) : 26-45.

Verdoorn, P. J. (1960), The Intra-Bloc Trade of Benelux, in E. A. G. Robinson (eds.), *Economic Consequences of the Size of Nations*, London: Macmillan.

INTRAINDUSTRIJSKA TRGOVINA SJEDINJENIH AMERIČKIH DRŽAVA

SAŽETAK

Cilj ovog rada je objasniti horizontalnu i vertikalnu intraindustrijsku trgovinu SAD-a s trgovinskim partnerima NAFTA-e, Europske Unije i ASEAN-a, te utvrditi determinante intraindustrijske trgovine, horizontalne i vertikalne. Koristeći pristup panelnih podataka, rezultati pokazuju negativnu korelaciju između zaklada i intraindustrijske trgovine. Ti rezultati ukazuju na to da se intraindustrijska trgovina češće odvija među zemljama koje imaju sličnosti po pitanju zaklada. Nalazi potvrđuju teoriju da, generalno gledano, nema pozitivne statističke veze između HIIT i HO varijabli. Naši razultati također potvrđuju hipotezu da trgovine raste ako se smanjuju troškovi prijevoza.

***Ključne riječi**: Horizontalna intraindustrijska trgovina, Vertikalna, SAD, Komparativne prednosti*

Chunchao Wang[1]
Jiliang Hu[2]
Jingwen Yu[3]

UDK <353:330.3>:711.433(512.21)

CAN STRUCTURAL ADJUSTMENT OF GOVERNMENT GOVERNANCE IMPROVE ECONOMIC PERFORMANCE? : THE CASE OF WUHAN METROPOLIS CIRCLE IN CHINA[4]

ABSTRACT

The paper employs M-form and U-form organization theory to analyze the structural innovation of government governance, and tries to study the resources integration and economic performances among different cities in a metropolis circle by using the example of Wuhan metropolis circle in China. Specifically, we focus on analyzing the difference between economic performance before and after the formation of Wuhan metropolis circle. The research result shows that, on the one hand, the formation of Wuhan metropolis circle can make full use of the U-form organization; on the other hand, different cities also benefit from coordinated regional development and rational resources allocation thanks to the formation of metropolis circle. Furthermore, each city has individual characteristics and complementary to other cities. Consequently, the economic performance of theses cities greatly differs from each other.

Keywords: *Wuhan Metropolis Circle in China, Government Governance adjustment, M-form and U-form Organization*

I. INTRODUCTION

The concept of metropolis circle was originally introduced by Gottmann in 1957. As a means for economic and social development within a city, the theory of forming metropolis circle becomes increasingly important in explaining the conformation and development of regional economy as well as China economy as a whole. As one of the largest transforming and developing countries, China has been experiencing the process of rapid urbanization and industrialization since the Opening-up and Reform in 1979, during which process some metropolis circles have been set up one after another. According to the existing literatures, it's found that researches on this field mainly focus on the following two aspects: first, the industrial structure and the theory of agglomeration economy, including effect of agglomeration and effect of radiation. Such theories are used to explain the development of metropolis circle (Boudeville 1966, Mills 1967, Fujita, Krugman and Venables 2001, Xiang 2004, Wei Zhao 2005, Bin Yu, et al. 2007, Wu and Liu 2008). In recent years, relevant researches have been expanded to cover the area external to the metropolis circle and its influence on urban capital accumulation, labor mobility and ecological environment (Pauchard et al. 2006, Fu2007, Saks 2008, Tharakan and Tropeano 2009). Furthermore, some literatures show the significance of government

[1] Associate Professor of Economics, Department of Economics JiNan University, China. No. 601 Huangpudadaoxi Road, Department of Economics, JiNan University, Guangzhou, Guangdong Province, China, e-mail: twangcc@jnu.edu.cn
[2] Candidate for PhD, College of Economics, Huazhong Normal University, China. e-mail: hujiliang_econ@126.com
[3] Correspondent author. Candidate for PhD, National School of Development Peking University, China. e-mail: richard.yo2008@gmail.com
[4] This work is funded by the National Philosophy and Social Sciences Foundation Project (10BJL032), the Humanities and Social Sciences Project, Ministry of Education (09YJC790122), the Industry-School-Research Cooperative Project, Guangdong Science and Technology Department (2010B090400008), the Soft Science Research Project in Guangdong Province (2009B070300050), Social Science Project of the department of Education in Guangdong Province (10WYXM042) and the Innovation Fund of JiNan University. Many thanks for the valuable suggestions and reviews that the anonymous reviewers give to us.

administration and emphasize its important role in economic development of cities, particularly in the overall development of the metropolis circle. A significant issue in this regard lies in the inter-industry coordination and pattern of interests distribution, wherein local governments are required to strengthen coordination and change their coordinating means as well as optionally market intervention (Ostrom, Tiebout and Warren, 1961, Osborne and Gaebler, 1992, Tan, 2000, Feiock, 2004).

It can be understood that previous studies have mainly focused on the interpretation of the metropolis circle in terms of a variety of economic effects, such as agglomeration effects, radiation effects, spatial distribution of such effects and external area. Meanwhile, present research literatures pay little attention to the forming mechanism of metropolis circle and the difference between economic performance before and after its formation. The paper differs from above studies in that it utilizes the M-U theory to analyze regional metropolitan development and presents a new interpretation of the forming mechanism of the metropolis circle from the perspective of inter-organizational structural changes. We try to use exemplary Wuhan Metropolis Circle in China to illustrate the mechanism. The structure of the paper is arranged as follows: the second section is to outline the M-U theory. The third section is to explain the management of organizational performance in Metropolis Circle by applying the M-U theory. The fourth section is a theoretical analysis, wherein an empirical model will be set up to illustrate the improvement of economic performance thanks to the formation of Wuhan Metropolis Circle. The fifth section is a brief conclusion.

II. M-U THEORY AND ITS ANALYSIS FRAMEWORK

M-form organization refers to an organization with quasi-autonomous departments classified in terms of products, trademarks and geographical location. It is an organizational form of corporate governance. U-form organization refers to an organization governed according to the department functions. Qian and Xu (1993) gave a more specific definition. It is said that M-form organization covers many organizations and include all kinds of departmental functions. It emphasizes the combination of all complementary tasks. In such case, it can be regarded as a single complex department, in which every sector is an equivalent to a small U-form organization. On the contrary, U-form organization is formed by similar tasks. For example, as a typical M-form organization, Procter & Gamble organizes shampoo products in terms of the trademark by self-managing. In earlier years, the Ford Company was organized in the form of organizational departments based on production functions. By way of example, the production of car axles and car doors were divided in two different sectors to maintain professional production and economy scale. Hill (1985) further specified the difference between the U-form organizations and the M-form organizations. He believed that the M-form organization has the following five functions not owned by the U-organization: first, the chief manager of the M-form organization can assign full responsibility to various quasi-autonomous departments and each department has its own manager just like a small U-form structure. Second, the decision-maker is only responsible for formulating development strategies, including allocating resources to different sectors within the M-form organization. Third, a group of elites are gathered in the management of the M-form organization to act as leaders. Fourth, the management has resources sufficient to provide incentives and measures so as to make employees work as hard as required by the head. Fifth, the top leader may have a clear understanding of contributions made by various departments within the M-form organization, and thus can facilitate performance evaluation of these departments.

Further research on the U-M organization is aimed at seeking for a form of organization by means of which cities may achieve higher economic performance. Williamson (1975) has proposed an M-form hypothesis. He said that as compared to the U-form organization, people's behavior is closer to the goal of profit maximization within the M-form organizations. In other words, the M-form organization is more efficient than the U-form organization. In fact, as early as in 1962, Chandler conduct studies on the U.S. companies. In the early 20th century, some giant companies, such as General Electronics and the DuPont, have adopted the M-form in response to the increasing complexity of the organization structure. Also, it reflected the diversity of product categories. Burton and Obel (1980) confirmed that the M-form organization is more efficient than the U-form by using the methodology of computational simulation; Qian (1994), Maskin, Qian and Xu (2000) believed that the M-form organization is capable of enhancing the competitions among different sectors. The competition will more effectively force the individual to pursue better economic performance. The researchers set up a theoretical model and used the data of the Chinese state-owned enterprises to validate the theoretical hypothesis. Subsequently, Qian, Roland, and Xu (2006) applied the M-U theory to study the efficiency of a company's governance structure in a transforming economy. Herein we will take a further step to apply the M-U theory in the field of regional economics, especially in the study of metropolis circle. We believe the economic development and cooperation between cities can also be regarded as a huge system. The performance of an economic organization can mostly be determined by the type of organization structure it adopts, that is to say, in practice, different structures of urban governance usually obtain different economic performances. Why the governance structures have a great impact on the economic performance and how it runs? The following analysis is made on both theoretical and empirical bases through a comparative research on urban organizations depending on the framework of the M-U theory.

III. ECONOMIC PERFORMANCE OF METROPOLIS CIRCLE UNDER GOVERNANCE

The study introduces the U-M theory into an analytic framework of regional metropolis circle. It studies administration structure of a government, which leads to different economic performance concerning U-form and M-form metropolis circle. On the one hand, the central government divides the administrative districts according to the economic function, such as the Soviet Union and the Eastern Europe. This is referred to as U-form. In such case, the district runs under the coordination of the central government. The functions of local government should be coordinated among various regions, for instance, a change in one region will inevitably affect the other adjacent ones. As a result, such type of organization facilitates the control of local force by the central government. On the other hand, within the M-form organizations, the central government is capable of reducing the adverse effect of incomplete information by granting local governments more responsibilities and rights in a more decentralized way. It allows the local governments to take actions in the economy based on their own resources, since the local officials know better of local economic characteristics than any others. Moreover, the M-form organization will improve the competitiveness of regional economy, which is conducive to the urban economic development and the prosperity of urban economy. It means that local government has certain economic independence, which may lead to a competition within the local government and improve the operating efficiency. From the comparison, it can be seen that both the M-form organizations and the U-form organizations have their own advantages and disadvantages. Therefore, we need to further discuss the benefits brought by the formation of metropolis circle and an accurate forming mechanism. The preceding issue lies in whether the overall economic performance significantly changes with the local governance restructuring within the metropolis circle. We believe that the formation of metropolis circle has its inherent

characteristics, which will be explained in the summary of improvement in organization efficiency. Furthermore, the radiation effect is also critical to this issue. First, the metropolis circle area may include several cities with different professional functions. Then, the demand for efficiency improvement may lead to structural innovation of both the central and the local government. As a result, a vigorous metropolis circle comprising some adjacent cities can be formed in the end. On the one hand, a provincial government constituted by the local governments becomes an independent M-form organizations so that the issue of incomplete information can be resolved through the metropolis circle. In the M-form organization, each region still runs based on its own comparative advantages, while achieving scale economy without losing their own inherent advantages.

The establishment of Wuhan metropolis circle in China is consistent with above senarios. A "1+8" city's industrial chain has been set up and an integration of industrial structure has been realized since 2005[5]. These cities have different natural resources within the circle, forming different comparative advantages. Meanwhile, cities with different industrial structures can be integrated to to take the advantage of scale economy, despite they have their own particular leading industries. For example, Wuhan City is leading in the industries such as photoelectron information industry, auto manufacturing, steel manufacturing, environmental protection industry, new non-metallic materials, biological engineering; Huangshi City in textile, metallurgy, building materials and machinery; Ezhou City in metallurgy and clothing; Huanggang City in materials and textiles; Xianning City in textiles, foods and pharmaceutical chemicals; Xiantao City in textiles, light industry and food; Qianjiang City in salt chemicals, petrochemical and pharmaceutical chemicals; Tianmen City in foods, produce processing and textiles. In addition, these cities are also adjacent to each other, allowing them to enjoy benefits and advantages in this kind of spatial organization by division of work and cooperation. These nine cities within Wuhan metropolis circle complements with each other. To date, the coordinated development among the cities and appropriate allocation of resources has laid a solid foundation. In such scenario, the metropolis circle may effectively improve economic performance, provided that there is a good organization.

The existing theory about M-form and U-form can be correspondingly extended. In the following, a detailed theoretical illustration of the formation of the metropolis circle will be given and the differences between economic performance before and after the formation of Wuhan metropolis circle will be shown and explained.

IV. THEORETICAL ANALYSIS AND EMPIRICAL STUDY

In Wuhan metropolis circle, all prefecture-level cities have their own unique resource endowments and characteristic industry structure. Consequently, these cities are greatly complementary in function. From a theoretical model, we can recognize key features of the formation of metropolis circle. Under the leadership of the provincial departments, there are various cities, specialized division of labor within some specific industrial cities in a relatively perfect labor market; other cities will have a great impact on this type of industry. Thus, it's assumed that the equivalent of a city has a functional organization. Due to its influence, it can coordinate the functions of other cities. The development of functional departments in each city must be in advantage of the functional status of the city coordinated to the development of functional departments. The provincial government can collect information to coordinate the

[5] The Wuhan Metropolis Circle contains of 8 cities and another 1 central cities, which is as follows: Huangshi, Erzhou, Xiaogan, Huanggan, Xianning, Xiantao, Qianjiang, Tianmen, and the central one is Wuhan City. All the 9 cities constitute of urban communities within 100 km radius, and Wuhan City is the centre within the urban communities.

strategy in this case. However, cities with different functions cannot effectively communicate the business information with each other. It is just like a U-form organization. After the formation of Wuhan metropolis circle, the development of all cities included in Wuhan circle may work well together and the complementary features between different industries will be utilized by the member cities. This makes such formation a self-contained structure, and forms a local government organization of M-form.

The theory is based on the analysis framework by Qian, Roland, and Xu (2006). The overall uncertainty and the distance between regions as well as resources will be integrated into the cost analysis framework. It's assumed that there are three regions, i.e., A, B and C, each region has two functions: function 1 and function 2, and these two functions emerge to the complementary relationship in economic functions. It's further assumed that there is an economic development strategy requiring the two functions to run on a coordinated basis at a certain ratio. If a region contains only 2 functions, then the local government's information can be identified as belonging to local information. It means that the regional governments can acquire all information. As a result, there is no distortion in information communication under the M-form organization. On the contrary, if an area is subject to the U-form organization, then various functional sectors may have to communicate information with each other. An additional information cost will incur. We make further assumptions as follows: A is the center of the metropolis circle. B and C are located at the periphery of the central city and gain R/2 benefits. The process of economic development and the implication of development strategy require a certain cost, indicated as C. Implementation of any development strategy requires a certain kind of cost. Once the development strategy is carried out, the chief executive officer of local government will no longer need to pay additional costs so that C can be regarded as a learning cost.

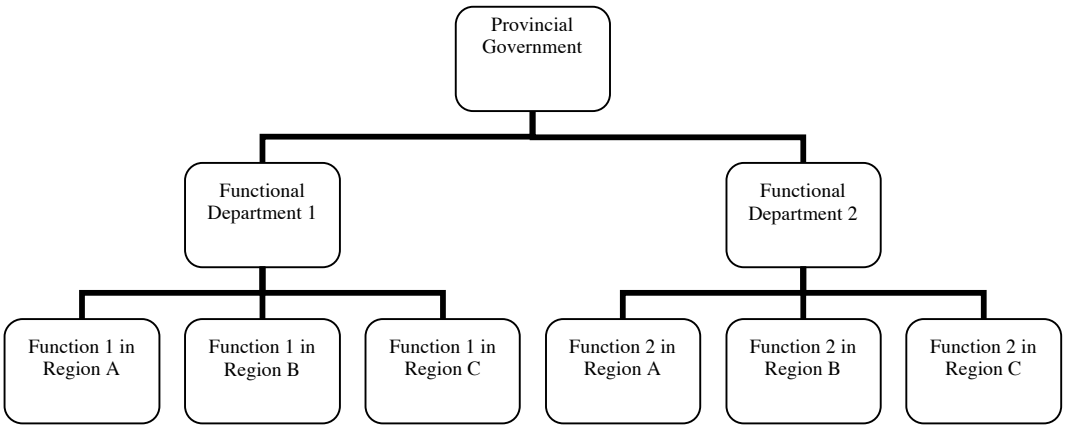
Assumption: $\frac{R}{2(1-\delta)} - C > 0$,i.e., if a regional development strategy can be implemented correctly by probability 1, then net income of the region must be greater than 0.

A. the situation before the formation of metropolitan

In the foregoing analysis, we can regard the provincial and local urban governments as an M-form organization in China. Within the M-form organization, each region can be deemed as a small U-form organization, and the regions are also subject to their internal functions. Before full formation of the metropolis circle, organization of each regional government under this structure is of U-form, as shown in Figure 1.

Figure 1

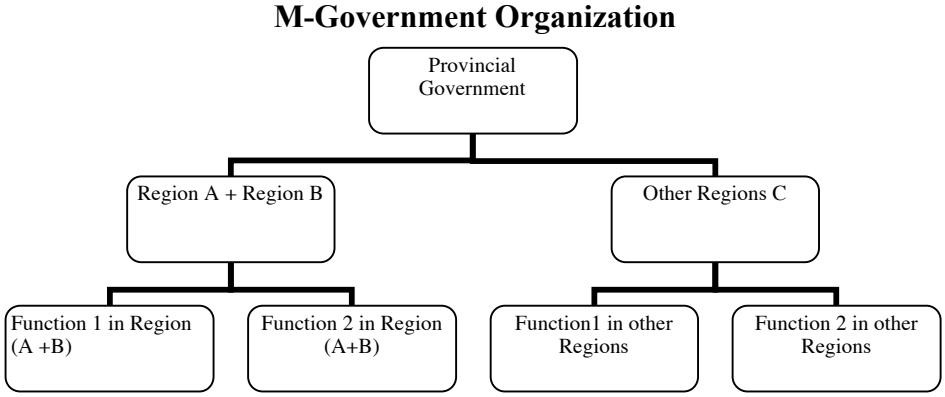
U-Government Organization



B. the situation after the formation of metropolitan

If the plan of establishing the metropolis circles is implemented, then these complementary features of different regions will be organized together into a independent organizational system that can be seen as an M-form structure, as shown in Figure 2.

Figure 2



C. Empirical Analysis

1) Data

Herein we use the panel data of Wuhan metropolis circle from 2002 to 2008. The data from 2002 to 2007 is obtained from Hubei Statistics Yearbook in 2008 and the data of 2008 from the Bulletin Economic and Social Development in China. The data covers the overall individuals rather than sampling data, so we can use the econometric model of fixed effects. We have arranged the panel data for every city included in Wuhan metropolis circle in China.

2) Econometric Model

We calculate the difference between economic performance before and after the formation of Wuhan metropolis circle. The basic econometric model is derived from the Cobb-Douglas production function. Y is output, K represents capital inputs and L represents labor input. In addition, A is the residual, which means the Solow's total factor productivity. The basic equation is as follows:

$$Y = AK^{\alpha}L^{\beta} \quad (1)$$

Equation (1) is the fundamental product function of a firm, which is used to depict the economic development of the whole metropolis circle. According to equation (1), we obtain the following econometric equation (2).

$$LnY_{it} = LnA + \beta_{11} LnLU_{it} + \beta_{12} LnLR_{it} + \beta_2 LnK_{it} + \beta_3 D_{2003} + \beta_4 D_{2004} + \beta_5 D_{2005} + \beta_6 D_{2006} + \beta_7 D_{2007} + \beta_8 D_{2008} + \beta_9 D_{2009} + BX + \varepsilon_{it} \quad (2)$$

In the above equation (2), Y_{it} means gross product of region i in year t; A is the efficiency coefficient; LU_{it} means the total number of employees in the urban areas of region i in year t; LR_{it} means the total number of employees in the rural areas of region i in year t; K_{it} means the total investment in fixed assets of region i in year t; D_t (t =2003, 2004, 2005, 2006, 2007, 2008,2009) is a dummy variable, and $D_t = \begin{cases} 1, & \text{Year t} \\ 0, & others \end{cases}$, which captures two effects: first, the time effect of GDP, i.e., characterized by possible changes of GDP with time; second, the marginal effect of establishment of Wuhan Metropolis Circle, which is of interest herein; LnA is the efficiency in year 2002, which is the benchmark; X represents a group of control variables: FDI_{it}, which stands for foreign direct investment; and $Fiscal_{it}$, which means fiscal expenditure.

In order to eliminate the policy effect of establishment of Wuhan Metropolis Circle, the framework by Wooldridge (2005) is used here to make comparison research on the economic performance within and outside the metropolis circle, including specific analysis steps as: first, the samples composed of all cities within Wuhan Metropolis Circle are used to estimate each parameter in model (2) , with coefficient D_t of interest; second, the samples composed of all cities outside Wuhan Metropolis Circle are used to re-estimate each model (2) , with coefficient D_t of interest; finally, a sequence of disparities between economic performance within and outside Wuhan Metropolis Circle can be obtained after coefficient D_t from estimate of samples within Wuhan Metropolis Circle minus by coefficient D_t from estimate of samples outside Wuhan Metropolis Circle, and then the change in the sequence before and after the formation of Wuhan Metropolis Circle can be studied. If there's remarkable change, then it shows the formation of metropolis circle has marked influence on economic performance, indirectly evidencing the effectiveness of the policy of establishing metropolis circle.

The paper collects sample data from both area within and outside Wuhan Metropolis Circle, and employs fixed effect model to estimate formula (2), with results shown in table 1. The values of Rho and F in Model I and Model II both pass 1% significance test, indicating the model is tenable. Since certain relevance may exist between the variables, the significance and parameter symbols of part of control variables in these two models may differ from those in existing literatures. However, it is neither an interest of this research, nor will it impose any material impact on this research and its conclusions. Thus, we will focus on the analysis of parameter D_t. In Model I, all D_ts pass 5% significance test while D_{2005} fails to pass 10% significance test, indicating the presence of significant time effect or policy effect in GDP of each city within Wuhan Metropolis Circle. In Model II, D_{2004} passes 10% significance test, while other D_ts fails, indicating both the absence of equivalent policy effect in GDP of cities outside Wuhan Metropolis Circle and the lack of time effect which changes as time lapses. Consequently, whether the policy of establishing metropolis circle is effective can be concluded by simply studying Model I. However, for the purpose of robustness, the paper also calculates the marginal effect of the policy of establishing metropolis circle based on the previous methods to validate the reliability of the conclusion. See last column in table for detailed results.

The estimated results are shown in Table 1.

Table1

Estimation Result of Fixed Effect Model

Variables	*Model □（samples collected from within the metropolis circle）*		*Model □（samples collected from outside the metropolis circle）*		*Disparity Coefficient=（1）-（2）*
	（1） Coefficient	Standard Deviation	（2） Coefficient	Standard Deviation	
D_{2003}	0.0758*	0.0430	0.0519	0.0512	0.0239
D_{2004}	0.1872***	0.0661	0.1600*	0.0914	0.0272
D_{2005}	0.1454	0.1001	0.0589	0.1571	0.0865
D_{2006}	0.4753***	0.1477	0.2803	0.2284	0.1950
D_{2007}	0.4193**	0.1944	0.2012	0.3039	0.2181
D_{2008}	0.6108**	0.2504	0.3161	0.3956	0.2947
D_{2009}	0.6716**	0.3241	0.3277	0.5287	0.3439
$LnLU_{it}$	-0.1388**	0.0612	-0.2056**	0.0904	-
$LnLR_{it}$	0.0688	0.1851	-0.1759	0.2556	-
LnK_{it}	0.4347***	0.1334	-0.1003	0.0962	-
$LnFDI_{it}$	-0.0267	0.0193	0.0053	0.0318	-
$LnFiscal_{it}$	-0.2622	0.2183	0.4021	0.3334	-
Constant	4.2524***	0.7364	6.2601***	1.4321	-
Sample	72		48		-
Rho	0.9936		0.9808		-
F	11.81		27.63		-

Notes: * means significant at the level of 10%; ** means significant at the level of 5%; *** means significant at the level of 1%.

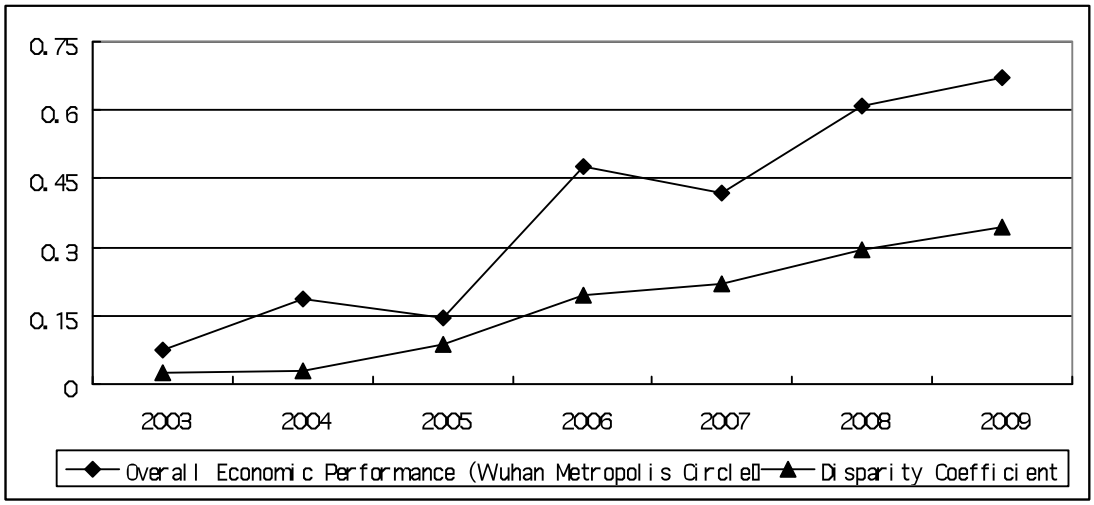
The above econometric results indicate that the overall economic performance of all the cities within Wuhan metropolis circle markedly improves after the formation of Wuhan Metropolis Circle. From 2003 to 2009, the coefficients of dummy variables generally indicate the increasing trend in both Model □ and Model □. The most important conclusion is that the overall economic performance has been gradually increasing since the formation of Wuhan Metropolis Circle as shown in Figure 3. In the analytic framework, the change of the coefficients concerning dummy variables generally shows the change of the economic performance due to the restructure of governance. It's noted that the most important conclusion

is that, 2005 is the 'skip year' for the overall economic performance, when the reform of organizational structure in Wuhan metropolis circle occurred. Upon the occurrence of the reform, the economic performance exhibited an increasing trend in the next few years. In addition, all the dummy variables are significant except for dummy one for 2005. Therefore, it can be seen that there is great difference between before and after the reform of organizational structure in Wuhan metropolis circle. The above empirical conclusion testifies the theoretic conclusion on the economic difference between U-form and M-form organization structure. The M-form shows a better performance for cities in metropolis circle than U-form.

To validate the robustness of above conclusion, the economic performance of samples within Wuhan Metropolis Circle is compared to those outside Wuhan Metropolis Circle (see sixth column in Table 1 and Fig. 3). From the trend reflected in the disparity coefficient in Table 1 and Fig.3, it can be seen there's little disparity between economic performance of cities within and outside the circle before its formation. However, the disparity in economic performance between cities within and outside the circle greatly expands from 0.02 at the early stage to 0.20 in 2006, till 0.34 in 2009, indicating the formation of Wuhan Metropolis Circle indeed contribute to the economic performance, and thus the conclusion is robust and reliable.

Figure 3

Change of Economic Performance



V. CONCLUDING REMARKS

The paper is intended to study the resource integration and economic performance among different regions in metropolis circle. We use the example of Wuhan metropolis circle in China to explain the change caused by the transformation of governance structure. We utilize the M-U organization theory to analyze the development of metropolis circle and the change of economic performance of adjacent cities. We particularly focus on the difference of economic performance between before and after the formation of Wuhan metropolis circle. We believe that the formation of Wuhan metropolis circle makes full use of U-form organization. Furthermore, different regions also make use of M-form organization due to the coordination of regional development and rational resources allocation. In the case of China, each region in metropolis circle has local characteristics and complements to other regions. As a result, the economic performance of each region will show a significant discrepancy. It's shown that M-form organization will probably result in higher economic performance, and reduce the co-operation cost among the cities in metropolis circle. The implication obtained from the theoretic model can

be testified through data from different developing countries or districts around the world. Specifically, we use the Wuhan metropolis circle as an evidence. Indeed, the empirical conclusion tells us that for a developing district, only during the transforming process of urban development can you give an impetus to the economic performance of the whole metropolis circle. Specifically, the government can formulate appropriate policy to promote cooperation among adjacent cities. The simplest way is to construct a metropolis circle. We believe that the development of metropolis circle is probably a key pathway for urban development. Adjacent cities joining in a metropolis circle will reduce the cooperation cost. In conclusion, M-form organization structure is conducive to the management among cities.

The analysis framework and methodology in the paper may be widely used to study economic performance of some other organizations in developing countries and areas. And for some specific organizations, we should especially pay attention to the specific operation conditions among the internal departments of a whole organization.

REFERENCES

Bin Yu, Feng Juan & Zeng Juxin (2007). "Industrial Spatial Organizing based on Cluster Network in Wuhan Urban Cluster". *Economic Geography*, 27 (3), 427-432.

Boudeville, J. R. (1966). *Problems of Regional Economic Planning*. Edinburgh : Edinburgh University Press ,1966

Burton, Richard M. and Obel, B. (1980). "A Computer Simulation Test of the M-Form Hypothesis". *Administrative Science Quarterly*, 25 (3), 457-466.

Chandler, Alfred Jr. (1962). *Strategy and Structure*. New York: Doubleday & Company, Inc.

Feiock, Richard C. (2004). *Metropolitan Governance*. Washington, DC: Georgetown University Press.

Fu, Shihe (2007). "Smart Café Cities: Testing human capital externalities in the Boston metropolitan area". *Journal of Urban Economics*, 61 (1), 86-111.

Fujita, M., Krugman, P. R. & Venables, A. (2001). *The Spatial Economy*. Cambridge, MA: MIT Press.

Gottmann, J. (1957). "Megalopolis or the Urbanization of the Northeastern Seaboard". *Economic Geography*, 33 (3), 189-200.

Hayek, F. (1945). "The Use of Knowledge in Society". *American Economic Review*, 35 (4), 519-530.

Hill, C.W.L. (1985). "Oliver Williamson and the M-Form Firm: A Critical Review". *Journal of Economic Issues*, 19 (3), 731-751.

Maskin, E., Qian, Yingyi & Xu, Chenggang (2000). "Incentives, Information, and Organizational Form". *Review of Economic Studies*, 67(2), 359-378.

Mills, Edwin S. (1967). "An Aggregative Model of Resource Allocation in a Metropolitan Area". *American Economic Review*, 57(2), 197-210.

Osborne, D. & Ted Gaebler (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Reading, MA: Addison-Wesley.

Ostrom, V., Tiebout, Charles M. & Warren, R. (1961). "The Organization of Government in Metropolitan Areas: A Theoretical Inquiry". *American Political Science Review*, 55 (4), 831-842.

Pauchard, A., Mauricio, A., E. Peña & R. Urrutia (2006). "Multiple Effects of Urbanization on the Biodiversity of Developing Countries: The Case of a Fast-growing Metropolitan Area". *Biological Conservation*, 3 (127), 272-281.

Qian, Yingyi (1994). "Incentives and Loss of Control in an Optimal Hierarchy". *Review of Economic Studies*, 61 (3), 1994, 527-544.

Qian, Yingyi, Roland, Gérard & Xu, Chenggang (2006). "Coordination and Experimentation in M-Form and U-Form Organizations". *Journal of Political Economy*, 114 (2), 366-402.
Qian, Yingyi and Chenggang Xu (1993). "Why China's Economic Reform Differ: The M-form Hierarchy and Entry/ Expansion of the Non-state Sector". *Economics of Transition*, 1993, 1 (2), 135-170.
Saks, E.R. (2008). "Job Creation and Housing Construction: Constraints on Metropolitan Area Employment Growth". *Journal of Urban Economics*, 64 (1), 178-195.
Tan, A. H. H. (2000). "Sources of the Asian Currency Crisis: Internal and External" (pp. 306–316). in ASEAN University Network & The Korean Association of Southeast Asian Studies, Economic Crisis in Southeast Asia and Korea.
Tharakan, Joe & Jean-Philippe Tropeano (2009). "On The Impact of Labor Market Matching on Regional Disparities". *Journal of Regional Science*, 49 (1), 57-80.
Williamson, O. (1975). *Markets and Hierarchies*. New York: Free Press.
Jefferey M. Wooldridge (2008). *Introductory Econometrics A Modern Approach.* Cincinnati:South-Western College Publishing.
Wu, Fuxiang and Liu Zhibiao (2008). "Research on the Mechanism of How City Group Drive Economic Growth——Empirical Evidences from 16 Cities of Yangtze River Delta". *Economic Research Journal*, 11, 126-136.
Xiang, Guangqin (19). "World Metroplis Circle Theory and Practice Implications for Urban Development in China". *Forum of World Economy and Politics*, 3, 15-18.
Zhao, Wei (2005). "Functions of a Central City & Development of Wuhan Urban Agglomerations". *Journal of Wuhan University of Hydraulic and Electrical Engineering* (Social Sciences Edition), 58 (3), 300-305.

MOGU LI STRUKTURALNE PRILAGODBE VLADINOG UPRAVLJANJA POBOLJŠATI EKONOMSKE UČINKE? SLUČAJ WUHANSKOG VELEGRADSKOG PODRUČJA U KINI[6]

SAŽETAK

Rad koristi organizacijsku teoriju M-forme i U-forme za analizu strukturalnih inovacija vladinog upravljanja, te pokušava proučiti integraciju resursa i ekonomske učinke vladinog upravljanja među raznim gradovima velegradskog područja koristeći primjer wuhanskog velegradskog područja u Kini. Posebno je usredotočen na analizu razlike između ekonomskih učinaka prije i nakon stvaranja wuhanskog velegradskog područja. Rezultati istraživanja pokazuju da, s jedne strane, stvaranje wuhanskog velegradskog područja može u potpunosti iskoristiti U-formu organizacije, dok s druge strane, različiti gradovi također profitiraju radi koordiniranog područnog razvoja i racionalne raspodjele resursa zahvaljujući stvaranju velegradskog područja. Osim toga, svaki grad ima individualne karakteristike i one komplementarne drugim gradovima. Stoga se ekonomski učinci ovih gradova uvelike razlikuju jedni od drugih.

***Ključne riječi**: wuhansko velegradsko područje, prilagodba vladinog upravljanja, M-forma i U-forma organizacije*

[6] Ovaj rad financira Projekt Nacionalne fondacije za filozofiju i društvene znanosti (10BJL032), Projekt Društvenih znanosti, Ministarstvo obrazovanja (09YJC790122), Projekt istraživačke suradnje industrije i škole, Odjel znanosti I tehnologije u Guangdongu (2010B090400008), Soft Science istraživački projekt provincije Guangdong (2009B070300050), Projekt društvenih znanosti odjela za obrazovanje provincije Guangdong (10WYXM042) i Fond za inovacije Sveučilišta JiNan. Velika zahvala ide vrijednim sugestijama i recenzijama anonimnih recenzenata.

Marc Escrihuela-Villar[1],
Jorge Guillén[2],

UDK 339.13.012.432

ON COLLUSION SUSTAINABILITY WITH STACKED REVERSION

ABSTRACT

We consider a multi-period oligopoly model to analyze cartel sustainability where a subset of collusive firms is exogenously given. We assume that in case of cheating only the cheater is expelled from the cartel and collusion continues without the cheater. We show that, in our model, when firms compete in quantities and the cartel is sufficiently small, a Stackelberg leader cartel can always be sustained if firms are patient enough. Furthermore, in this case collusion is more easily sustained than when firms play grim trigger strategies. The opposite result is obtained in a price-setting supergame with differentiated products.

Keywords: *Collusion; Stacked reversion; Trigger strategies*
JEL Classification: *L13; L40; L41*

1. INTRODUCTION

The analysis of cartel formation in oligopoly markets has a long tradition in the economic literature. It is well-known that although there is a general firms' interest in the existence of a cartel, the benefits of cartel formation are not evenly distributed and often nonmembers obtain higher profits than cartel members.[3]

We address the issue of collusion success by considering a supergame-theoretic approach (see the seminal paper by Friedman (1971)). In this approach, seemingly independent but parallel actions among competing firms in an industry are driven to achieve higher profits and this is termed tacit or implicit collusion. The tacit collusion literature is immense and it has usually focused on the equilibrium that maximizes industry profits[4] (see for example Rothschild (1999)) using subgame perfect Nash equilibria -henceforth, SPNE- as solution concept although, in practice, it can be observed that many collusive agreements do not involve all firms in the industry. In this regard, a significant example is the citric acid industry where three North-American and five European firms were fined for fixing prices and allocating sales in the worldwide market. Their joint market share was around 60 percent (see Levenstein and Suslow (2006)). It is also well known that this repeated game setting exhibits multiple SPNE collusive agreements. Therefore, to select among those equilibria the literature has usually adopted the particular criterion of restricting strategies to grim trigger strategies (see for instance Escrihuela-Villar (2008)) in which defection ruins the relationship forever. These strategies do not consider the possibility that after a defection loyal cartel members might want to continue colluding. In other words, it seems plausible that a large number of firms may not simply fail to coordinate on a punishment strategy for one firm that

[1] Corresponding author mailing address: Universitat de les Illes Balears. Departamento de Economia Aplicada. Edificio Jovellanos. Ctra. Valldemossa km 7.5, 07122 Palma de Mallorca. Balearic Islands - Spain. Tel. +34 971173242. Fax: +34 971172389. Email: marc.escrihuela@uib.es

[2] Mailing address: Escuela de Administración de Negocios para Graduados (ESAN). Alonso de Molina 1652, Monterrico, Surco, Lima - Perú Telf.: (511) 317-7200. Email: jguillen@esan.edu.pe

[3] The intrinsic difficulty in convincing firms to constitute a cartel was pointed out by Stigler (1950, p. 25) in a discussion of mergers: "The major difficulty in forming a merger is that it is more profitable to be outside a merger than to be a participant".

[4] Among the few exceptions are Escrihuela-Villar (2008) that analyzes the price effects of horizontal mergers, Escrihuela-Villar (2009) that considers how the sequence of play between the cartel and the fringe affects cartel stability or Bos and Harrington (2010) that endogenize the composition of a cartel with heterogeneous production capacities.

deviates from the collusive equilibrium. In this sense, there exists empirical evidence that participants in a cartel are not always willing to risk the collapse of collusion by punishing deviators. As an example, the Sugar Institute case explained in Genesove and Mullin (2001) where occasional incidents of cheating were typically not retaliated against or cheating did occur but sparked only limited retaliation.

The main purpose of this paper is to analyze cartel sustainability considering a set of strategies that could eventually be less grim than the trigger strategies. To that extent, we consider that cheating on the cartel agreement results only in the ejection of the defector. Our work is most closely related to Eaton and Eswaran (1998) from where the present paper gets part of the inspiration. To the best of our knowledge, they are the first to propose a set of strategies where non cheating cartel members continue operating as a cartel. They termed this scenario stacked reversion. The present paper differs, however, in some important ways. To begin with, Eaton and Eswaran consider simultaneous decision between the cartel and the fringe. However, in Abreu (1986, 1988) a two-phase output path (with a stick-and-carrot pattern) it is proved to be optimal when firms simultaneously choose their output and, consequently, the optimal set of strategies has already been characterized in a simultaneous decision context.[5] Therefore, and since one of the most widely accepted structures to characterize collusive behavior is that of a leader cartel, we assume the existence of cartel leadership.[6] Second, Eaton and Eswaran provide numerical examples of cartel sustainability with stacked reversion whereas in the present paper, by assuming an inverse demand curve, fairly general results are obtained.

We develop a multi-period oligopoly model with homogeneous firms, an exogenous subset of which are assumed to collude, while the remaining (fringe) firms choose their output levels non-cooperatively.[7] We use SPNE as solution concept assuming that in the event a cartel member cheats there is stacked reversion. Two different scenarios are considered: (i) firms set quantities and therefore, the cartel behaves as a Stackelberg leader with respect to the fringe in each period and (ii) there is cartel leadership with differentiated products and price competition. We obtain that with quantity competition, when the cartel is sufficiently small, a cartel can always be sustained with stacked reversion when firms are patient enough. In this case, stacked reversion also proves to be an easier way to sustain collusion than Nash reversion. Intuitively, the incentives that fringe firms have to free-ride on the cartel's collusive efforts disappear since a firm inside the cartel does not find it desirable to exit because each firm in the cartel earns higher profits than each fringe firm. One possible interpretation of this result is that the literature that deals with tacit collusion and cartel leadership should consider that Nash reversion might not constitute the right way to punish the cheater. On the contrary, these results do not carry over to a model with price competition and differentiated products. With price competition, each firm in the fringe always earns higher profits than each cartel firm. Thus, the free-rider problem that characterizes a cartel-fringe model arises and the cartel cannot generally be sustained with stacked reversion and price competition. In this case, therefore, collusion is always more easily sustained with Nash reversion than with stacked reversion.

The remainder of this paper is structured as follows. In section 2 we present the model and study cartel sustainability with quantity competition in a multi-period oligopoly model with stacked reversion and trigger strategies. In section 3 we also consider differentiated products and price competition. We conclude in section 4. All proofs are grouped together in the appendix.

---

[5] Abreu's notion of optimality implies that the range of time discount factors over which collusion is sustainable is maximized.

[6] There exist several real examples of cartel leadership like the Carbonless paper industry (see for instance Harrington (2006)).

[7] The assumption of a cartel involving a subset of firms is based on the fact that some of the best known examples of cartels involve only a part of the industry. Some significant cases are the citric acid, the carbonless paper or the North Atlantic shipping industries (see Levenstein and Suslow (2006)).

2 THE MODEL AND RESULTS

We consider an industry with $N>2$ firms. Each firm produces a quantity of a homogeneous product and for simplicity it is assumed that the total production cost of the firms is equal to zero. The industry inverse demand is given by the piecewise linear function: $p(Q)=max(0,a-Q)$, where Q is the industry output, p is the output price, and $a>0$. We assume that $K \in [2, N]$ firms, indexed by $k=1,...,K$ -henceforth, cartel firms- behave cooperatively so as to maximize their joint profits. The remaining (N-K) firms constitute the fringe and choose their output in a non-cooperative way. We assume that only one cartel is formed, and we take K as exogenously fixed.[8] The assumption of an exogenously given subset of firms colluding is based on the fact that cartels often involve an agreement between firms which can easily coordinate with each other (e.g. because they are based in the same country or have a common corporate culture). The fringe consists of foreign firms or new entrants that could not coordinate their behavior with the cartel firms even if they wish so.[9]

We assume that firms compete repeatedly over an infinite horizon with complete information (i.e. each of the firms either fringe or cartel observes the whole history of actions) and discount the future using a discount factor $\delta \in [0,1]$. Following the cartel and fringe literature, we assume that in each period the cartel behaves as a Stackelberg leader with respect to the fringe or equivalently, that the strategic advantage of the leadership is only available to cartel firms.[10] Time is discrete and dates are denoted by $t=1,2,....$ In this framework, a pure strategy for firm k is an infinite sequence of functions $\{S_k^t\}_{t=1}^{\infty}$ with $S_k^t : \sum^{t-1} \to \vartheta$ where $\sum^{t-1}$ is the set of all possible histories of actions (output choices) of all cartel firms up to $t-1$, with typical element σ_j^τ, $j=1,...,K$, $\tau=1,...,t-1$, and Q is the set of output choices available to each cartel firm. We restrict our attention to the case where each cartel firm is only allowed to follow the stacked reversion strategies in which in case of cheating, only the cheater is expelled from the cartel and collusion continues without the cheater.[11] Let qK and qF denote the output corresponding respectively to collusion when K firms maximize joint profits and non-cooperative behavior. At $t=1$, $S_k^1 = qK$, while at $t=2,3,...$

$$S_k^t(\sigma_j^\tau) = \begin{cases} q_K \text{ if } \sigma_j^\tau = q_K \text{ for all } j = 1,...,K \text{ and } \tau = 1,...,t-1 \\ q_{K-1} \text{ if } \sigma_j^\tau \neq q_K \text{ for a } j = 1,...,K \; j \neq k \text{ and } \tau = 1,...,t-1 \\ q_F \text{ if } \sigma_K^\tau \neq q_K \text{ for any } \tau = 1,...,t-1 \end{cases} \quad (1)$$

In words, in the event a cartel member cheats instead of dissolving the entire cartel, the loyal members find in their own interest to continue to produce as a cartel but with one firm less. Regarding fringe firms, their optimal response consists of maximizing their current period's payoff in such a way that if each firm in a cartel with K members produces qK, then the output produced by each fringe firm (qF) is

$$q_F = \max\left\{0, \frac{a - Kq_K}{N - K + 1}\right\}$$

[8] The literature on endogenous cartels has frequently used the concept proposed by d' Aspremont et al. (1983) for cartel stability. In this case, cartels containing just over half the firms in the industry are stable (see for instance Donsimoni (1986) or Shaffer (1995)).

[9] As an example, three North-American and five European firms in the citric acid industry were fined for fixing prices and allocating sales in the worldwide market. Their joint market share was around 60 percent. The rest of the producers included a variety of minor companies based in Eastern Europe, Russia and China (see Levenstein and Suslow (2006)).

[10] The seminal papers in this literature are Selten (1973) and d' Aspremont et al. (1983).

[11] Simultaneous deviations are ignored since, as was remarked by Fudenberg and Tirole (1991), "in testing for Nash or subgame-perfect equilibria, we ask only if a player can gain by deviating when his opponents play as originally specified" (p. 281).

We assume also that if a cartel firm deviates fringe firms optimally respond to the deviation. We denote by $\Pi^C(N,K)$ and $\Pi^F(N,K)$ the profit function of a cartel firm and that of a fringe firm when cartel and fringe firms produce respectively qK and qF. Then, cartel firms producing qK in each period can be sustained as a SPNE of the repeated game with the strategy profile (1) if and only if for given values of N,K and δ, the following condition is satisfied

$$\frac{\prod^{C}(N,K)}{1-\delta} \geq \prod^{D}(N,K) + \frac{\delta \prod^{F}(N,K)}{1-\delta}, \qquad (2)$$

where $\Pi^D(N,K)$ denotes the profits attained by an optimal deviation from a collusive output. Then, we denote by δ_K the minimum δ required for the condition (2) to be satisfied. Observe that for the incentive compatibility condition (2) to make sense, it must also be the case that it is satisfied for $K-1$. Otherwise, $\Pi^F(N,K-1)$ would not be the profits that a cheater would obtain since the cartel of $K-1$ would not be itself robust against cheating. Consequently, (2) has to be also satisfied for cartels with less than K firms in such a way that the following set of necessary and sufficient incentive compatibility conditions must all be satisfied

$$\frac{\prod^{C}(N,\mathrm{h})}{1-\delta} \geq \prod^{D}(N,\mathrm{h}) + \frac{\delta \prod^{F}(N,\mathrm{h}-1)}{1-\delta} \quad \forall h = 2,3,...,K \qquad (3)$$

It can be verified that the minimum discount factor needed for the joint profit maximization of cartel firms to be a SPNE of the repeated game with cartel firms playing (1) is

$$\delta_K = \frac{(-1+K)^2(2-K+N)^2}{(-1+K)^2K(4+K) - 2(-1+K)K(3+K)N + (1+K)^2N^2 + 4(1+N)} \qquad (4)$$

This is true because δ_K is the minimum δ required for the condition (2) to be satisfied and since $\frac{\partial \delta_K}{\partial \mathrm{K}} > 0$ this implies that if $\delta \geq \delta_K$ no firm has incentive to deviate from a cartel with K firms. In other words, the set of conditions (3) are also satisfied. Consequently, the following definition applies.

Definition 1 *δ_K is said to be the minimum discount factor required for the cartel of K firms to be sustainable as a SPNE. Then, a cartel of K firms is said to be sustainable if $\delta \geq \delta_K$ and $\delta_K \in (0,1)$.*

Equivalently, a cartel of K firms is more easily sustained than a cartel of $\overline{K}$ firms whenever $\delta_K < \delta_{\overline{K}}$. We are now in the position to characterize cartel sustainability with stacked reversion.

Proposition 1 *For $N \geq 4$ if $K \leq f(N) \equiv \frac{1}{4}(5+3N-\sqrt{N(N-2)-7})$, there exists $\overline{\delta} \in (0,1)$ such that if $\delta \geq \overline{\delta}$ the cartel with K firms is sustainable. For N=3, the cartel K=2 is sustainable if $\delta \geq 0.18$.*

Proposition 1 establishes that whenever the size of the cartel is relatively small and firms discount the future sufficiently little collusion can always be sustained using stacked reversion strategies. The intuition behind this Proposition is as follows. The major difficulty in forming a cartel is that it is often more profitable to be outside the cartel than to be a

participant. Consequently, cartels tend to be unsustainable because their members have incentives to free-ride on the profit-enhancing efforts of the cartel. However, this is not true when the cartel is relatively small. In other words, we can find a *K* small enough such that the profits of a fringe firm when a cartel with *K-1* is active are lower than the profits that a firm would obtain being a participant in a cartel with *K* firms.[12]

A natural question that arises is whether collusion is easier to sustain with the stacked reversion or when players adopt trigger strategies. In the latter case, following Friedman (1971) these strategies are such that cartel firms adhere to the collusive agreement until there is a defection, in which case they revert forever to the static N-firm Nash equilibrium.[13] In this case, cartel firms producing *qK* in each period can be sustained as a SPNE of the repeated game if and only if, the following condition is satisfied

$$\frac{\Pi^{C}(N,K)}{1-\delta} \geq \Pi^{D}(N,K) + \frac{\delta \Pi^{N}(N)}{1-\delta}, \qquad (5)$$

where $\Pi^{N}(N)$ denotes the static *N*-firm Nash equilibrium. It is an standard exercise to verify that the minimum discount factor required for the cartel of *K* firms to be sustainable as a SPNE using trigger strategies (the minimum δ required for the condition (5) to be satisfied), that we denote by δ_K^T, is

$$\delta_K^T = \frac{(-1+K)^2(1+N)^2}{16K^3 + K^2(-15+N)(1+N) + (1+N)^2 + 2K(1+N)^2}. \qquad (6)$$

An easy comparison reveals the following result.

Proposition 2 *If K≤(N+3)/2 the cartel of K firms is more easily sustained with stacked reversion strategies than with trigger strategies. The converse is true.*

When the cartel is small enough, the incentives that fringe firms have to free-ride on the cartel's collusive efforts disappear and fringe firms make fewer profits than in the static N-firm Nash equilibrium. Consequently, the stacked reversion is a harsher punishment than reverting to the Nash equilibrium and the repeated game with stacked reversion is more effective in enforcing an agreement than with Nash reversion. Interestingly, a different form of retaliation may inflict tougher punishments and thereby allow sustaining higher collusive prices just by ignoring the deviation.

3 PRICE COMPETITION

In this section we test whether our results hinge on the assumption of quantity competition and homogeneous products by considering also differentiated products and price competition. To that extent, we assume that the industry produces non-spatial horizontally differentiated products such that the degree of differentiation between the products of any two firms is the

[12] Note that $\frac{f(N)}{N} \in (\frac{1}{2},1)$.

[13] We note that, in this case, the punishment would also consist of cartel firms losing the strategic advantage of the leadership.

same. As in section 2, we assume that in each period the cartel sets the price before the fringe. The inverse demand function exhibits a Chamberlinian symmetry:

$$p_i=a-q_i-b\sum_{j\neq i} q_j$$

where p_i denotes the price of good *i* and q_j the quantity sold of good *j*. Alternatively, we can write the demand system as

$$q_i=\alpha-\beta p_i+\gamma\sum_{j\neq i} p_j$$

where $\alpha=\frac{a}{1+(N-1)b}$, $\beta=\frac{1+(N-2)b}{(1-b)(1+(N-1)b)}$ and $\gamma=\frac{b}{(1-b)(1+(N-1)b)}$. It is assumed $a,b>0$ and $0\leq b\leq 1$. The value range for *b* (the common degree of product substitutability) implies that the products are viewed as substitutes rather than complements and that the price of each product is more susceptible to changes on its own demand rather than changes on other product demand. Hence, $b=0$ implies that the products are completely independent and $b=1$ indicates that they are perfect substitutes.

We analyze cartel sustainability extending the analysis of the previous section to price competition and differentiated products when cartel firms follow the stacked reversion strategies. It is a standard exercise to obtain the critical value below which a cartel member does not have incentives to deviate. Although the expression of this cutoff cannot be easily simplified, we have the following result.

Proposition 3 *With price competition and stacked reversion strategies, sustainable cartels exist in only two cases: i) K≤3 and ii) K=4, N=5.*

This result establishes that in the game defined above cartel sustainability is rather difficult with price competition. Thus, the implications on cartel sustainability derived in Proposition 1 do not carry over when firms compete in prices. This difference follows from the fact that reaction functions are upward sloping in price games but downward sloping in quantity games. With cartel leadership, the reaction of fringe firms reinforces the initial price increase that results from the cartel price and therefore the intuitions provided in the previous section are reversed when firms compete in prices. Intuitively, with price competition a cartel cannot (generally) be sustained with stacked reversion because the continuity of the cartel is not a severe enough punishment since fringe firms free ride on the cartel's collusive efforts to increase price. However, for a sufficiently small cartel the conferring of leadership status to the cartel allows collusion to be sustained with stacked reversion strategies. In this case, losing the leadership after defection inflicts a sufficiently harsh punishment on fringe firms in order for the cartel to be sustainable.

As in the previous section, we compare cartel sustainability with stacked reversion and under Nash reversion strategies. This leads to the following result.

Proposition 4 *With price competition a cartel is more easily sustained under trigger strategies than with stacked reversion strategies.*

Proposition 4 does not come as a surprise and it is perfectly in line with Proposition 3 and its intuition. When firms compete in prices, a Nash reversion inflicts a harsher punishment than continuing with the cartel without the defector. The intuition is that fringe firms always benefit more from a cartel than do the members and consequently, the cartel operating without the defector does not constitute (generally) an enforceable punishment.

4 CONCLUDING COMMENTS

We have developed a theoretical framework to study cartel sustainability using a more sophisticated set of strategies than the usual trigger strategies. Following Eaton and Eswaran (1998), we assume that in case of defection only the defector is ejected. We obtain that in a quantity-setting industry where the cartel produces before the fringe, when firms are patient enough collusion can be sustained with stacked reversion strategies. Furthermore, if the cartel is relatively small collusion is more easily sustained with stacked reversion than with trigger strategies. Conversely, with differentiated products and price competition, a cartel cannot generally be sustained with the stacked reversion strategies.

Our results suggest that with quantity competition and cartel leadership, a set of strategies that could be less grim than the trigger strategies should also be considered to analyze cartel sustainability. An alternative interpretation is that the existence of price wars can be limited not only by firms' ability to punish a potential deviation but also because a cartel may find optimal not to punish the deviator. In other words, usually retaliation includes temporary price wars, leading to profits below normal levels for some period of time and it may also include actions that are targeted at reducing the profits of the deviant firm. However, we proved that it may turn out to be the case that cartel firms find it optimal to leave the deviation unpunished which could be an additional explanation for the absence of price wars after retaliation.

The framework we have worked with is, admittedly, a particular one. To analyze real-world cases of cartels, firms' capacities, cost asymmetries or an extension to a wider range of demand functions should also be considered. Another natural question is also how the cartel could be able to impose its most preferred timing in a cartelistic model or considering renegotiation-proof strategies. We believe that those are subjects for future research.

APPENDIX

Proof of Proposition 1. From (4) it can be easily verified that δ_K is always positive and smaller than 1 whenever $K \leq \frac{1}{4}(5+3N-\sqrt{N(N-2)-7})$. This inequality has a real root whenever $N(N-2)-7 \geq 0$ which is true for $N \geq 4$. If $N=3$ and $K=2$, $\delta_K = 0.18$. ■

Proof of Proposition 2. From (4) and (6), it is immediate to verify that $\delta_K < \delta_K^T$ whenever $K \leq (N+3)/2$. ■

Proof of Proposition 3. The minimum discount factor required for the cartel of K firms to be sustainable as a SPNE using stacked reversion with price competition can be easily obtained but cannot be simplified (the expression and further details are available from the author upon request). We denote this cutoff by $\delta_K^P(b,K,N)$.

Easy calculations show that $\delta_K^P(0,K,N) = \frac{1}{2}(K-1)$.

Consequently, since $\left.\frac{\partial \delta_K^P(d,K,N)}{\partial d}\right|_{d=0} = \frac{1}{4}(K-3)(K-1) \leq 0 \ \forall K \leq 3$, there exists $\delta_K^P(b,K,N) \in (0,1)$ for $K \leq 3$.

On the other hand,

$$\delta_K^P(b,K,N)=\frac{9(1-2N)^4(N-3)^2(2+N)^2}{2628+N(-7332+N(11977+N(-9978+N(4089+4N(-182+N(-30+N(-78+25N)))))))}$$

which is only smaller than 1 if $N<\frac{1}{22}(11+\sqrt{55(109+6\sqrt{345})})\simeq 5.505$. Consequently, for *K=4*, only when *N=5* there exists $\delta_K^P(b,K,N)\in(0,1)$. Finally, for the case of *K>4* since $\delta_K^P(b,K,N)$ is decreasing with *b*, it is enough to verify that $\delta_K^P(1,K,N)\geq 1$. To that extent, it is easy to prove that $\frac{\partial\delta_K^P(1,K,N)}{\partial N}>0$. Thus, it suffices to check the value of $\delta_K^P(b,K,N)$ for the minimum possible value of *N* which is *K+1*. Then, we have that

$$\delta_K^P(1,K,K+1)=\frac{(1+2K)^4(1-3K+2K^2)^2}{1+K(1+2K)(-2+K(-5+2K(9+K(7+2K(-11+K(-1+6K))))))}$$ which is increasing in *K* and larger than 1 when *K>4.609* which proves that if *K>4*, the cartel is not sustainable. ■

Proof of Proposition 4. Like in the proof of Proposition 3, the minimum discount factor required for the cartel of *K* firms to be sustainable as a SPNE using trigger strategies with price competition can be easily obtained but cannot be simplified. We denote this cutoff by $\delta_K^{P,T}(b,K,N)$. Then, $\delta_K^{P,T}(b,K,N)\equiv\frac{\prod^D(N,K,b)-\prod^C(N,K,b)}{\prod^D(N,K,b)-\prod^N(N,K,b)}$, where we denote the Nash equilibrium profits by $\Pi^N(N,K,b)$. Since obviously, $\Pi^D(N,K,b)>\Pi^C(N,K,b)$ and $\Pi^D(N,K,b)>\Pi^N(N,K,b)$ $\delta_K^{P,T}(b,K,N)>0$. On the other hand, it can be easily checked that $\delta_K^{P,T}(b,K,N)\leq 1$ because $\Pi^C(N,K,b)>\Pi^N(N,K,b)$ which implies that $\delta_K^{P,T}(b,K,N)\in(0,1)$. Then, it suffices to check that $\delta_K^{P,T}(b,K,N)<\delta_K^P(b,K,N)$ when $\delta_K^P(b,K,N)\in(0,1)$. From Proposition 3, we only have two cases where the latter is true. It is immediate to verify that $\delta_K^{P,T}(b,4,5)<\delta_K^P(b,4,5)\ \forall b$ since $\delta_K^{P,T}(b,4,5)=\delta_K^P(b,4,5)$ has 3 different real roots, *b=-(1/3)*, *b=-(1/4)* and *b=-0.285714*. Also from the continuity of $\delta_K^{P,T}(b,K,N)$ and $\delta_K^P(b,K,N)$, we have that $\lim_{b\to 0}\delta_K^{P,T}(b,K,N)=\frac{1}{2}$, $\delta_K^{P,T}(1,4,5)=0.708$ and $\lim_{b\to 0}\delta_K^P(b,4,5)=\frac{3}{2}$, $\delta_K^P(1,4,5)=0.95$ which proves that $\delta_K^{P,T}(b,4,5)<\delta_K^P(b,4,5)$.

On the other hand, if *K≤3*, it can be verified that $\delta_K^{P,T}(b,K,N)=\delta_K^P(b,K,N)$ has no real roots in $b\in(0,1)\ \forall N,K=2,3$. We have that $\lim_{b\to 0}\delta_K^P(b,K,N)=\frac{(K-1)}{2}$ which implies that $\delta_K^P(0,K,N)>\delta_K^{P,T}(0,K,N)\ \forall N,K=2,3$. Finally,

$$\delta_K^P(1,3,N)=\frac{(1-2N)^4(-2+N)^2(1+N)^2}{88+N(-308+N(568+N(-580+N(372+N(-192+N(109+N(-70+17N)))))))}$$

$$>\delta_K^{P,T}(1,3,N)=\frac{(1-2N)^4}{76+N(-108+N(124+N(-132+41N)))}$$ and

$$\delta_K^P(1,2,N)=\frac{(1-2N)^4(-1+N)N^2}{-16+N(80+N(-205+N(325+4N(-91+N(71+N(-37+9N))))))}>$$

$$\delta_K^{P,T}(1,2,N)=\frac{(1-2N)^4}{73+4N(-47+N(60+N(-44+13N)))}$$, which proves that

$\delta_K^P(b,K,N) > \delta_K^{P,T}(b,K,N)$ $\forall N,K=2,3.$ ■

REFERENCES

-Abreu, D. (1986). "Extremal Equilibria of Oligopolistic Supergames". *Journal of Economic Theory 39 (1), 191-225.*

-Abreu, D. (1988). "On the Theory of Infinitely Repeated Games with Discounting". *Econometrica 56, No.2, 383-396.*

-d'Aspremont, C. Jacquemin, A., Gabszewick, J.J. and Weymark, J.A., (1983). "On the Stability of Collusive Price Leadership". *Canadian Journal of Economics 16 (1), 17-25.*

-Bos, I. and Harrington, J. Jr. (2010). "Endogenous Cartel Formation with Heterogeneous Firms". *The RAND Journal of Economics 41 (1), 92--117.*

-Donsimoni, M.-P. Economides, N. and Polemarchakis, H., (1986). "Stable Cartels". *International Economic Review 27, 317-327.*

-Eaton, C. and Eswaran, M., (1998). "Endogenous Cartel Formation". *Australian Economic Papers 37, 1-13.*

-Escrihuela-Villar, M. (2008). "Partial coordination and mergers among quantity-setting firms". *International Journal of Industrial Organization 26, 803-810.*

-Escrihuela-Villar, M. (2009). "A note on cartel stability and endogenous sequencing with tacit collusion". *Journal of Economics 96(2), 137-147.*

-Friedman, J.W. (1971). "A Non-cooperative Equilibrium for Supergames". *Review of Economic Studies 28, 1-12.*

-Fudenberg, D. and Tirole, J. (1991). "Game Theory". *The MIT Press.*

-Genesove, D. and Mullin, W.P. (2001). "Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case". *The American Economic Review, 91(3), 379-398.*

-Harrington, J. Jr. (2006). "How Do Cartels Operate?" *Foundations and Trends in Microeconomics 2, 1-105.*

-Levenstein, Margaret C., and Valerie Y. Suslow (2006). "What Determines Cartel Success?" *Journal of Economic Literature 44(1), 43--95.*

-Rothschild, R., (1999). "Cartel stability when costs are heterogeneous". *International Journal of Industrial Organization 17, 717-734.*

-Selten, R. (1973). "A Simple Model of Imperfect Competition, where 4 a Few and 6 are Many". *International Journal of Game Theory 2 (3), 141-201.*

-Shaffer, S., (1995). "Stable Cartels with a Cournot Fringe". *Southern Economic Journal 61, 744-754.*

-Stigler, George J., (1950). "Monopoly and Oligopoly by Merger," *American Economic Review Proceedings, 40, 23-34.*

ODRŽIVOST KOLUZIJE PRI „STACKED REVERSION“

SAŽETAK

Razmatramo multiperiodni model oligopola kako bismo analizirali održivost kartela u kojem je podskup koluzivnih tvrtki dan egzogeno. Pretpostavljamo da u slučaju prevare samo varalica biva izbačen iz kartela te se koluzija nastavlja bez prevaranta. Na našem modelu pokazujemo da kad tvrtke konkuriraju u količinama a kartel je dovoljno malen, Stackelberg leader kartel može uvijek biti održan ako su tvrtke dovoljno strpljive. Nadalje, u ovom slučaju, koluzija se jednostavnije održava nego kad tvrtke igraju „grim trigger“ strategije. Obrnuti se rezultat dobiva u cjenovnom nadigravanju s diferenciranim proizvodima.

***Ključne riječi**: koluzija, „stacked reversion“, „trigger“ strategije (strategije „brzog odgovora“).*

***JEL klasifikacija:** L13; L40; L41*

Vojislav Marjanović*
Jelena Gavrilović**
Nenad Stanić***

UDK 657.47:629.331>(73:430)

US AMERICAN VERSUS GERMAN ACTIVITY-BASED COSTING. EFFECTS ON BUSINESS DECISIONS MANAGEMENT IN THE AUTOMOTIVE INDUSTRY

ABSTRACT

This paper seeks to describe the role that activity-oriented cost accounting systems, i.e. US American and German activity-based costing, play in the establishment of supply-chain-networks in the modern automotive industry. These cost accounting systems are the subject of analysis for two reasons: they relatively successfully describe the causality principle between cost drivers and cost objects and represent two different approaches of cost calculation within the activity-oriented concept. Also, the paper attempts to show that the efficiency of these systems is contingent on the value chain activity to which the systems were applied. Although based on identical conceptual frameworks, these systems do not have the same cost allocation purposes in automotive industry.

Keywords: *cost accounting, activity-based costing, automotive industry, supply chain network*

JEL Classification: *M41, M49*

1. INTRODUCTION

The automotive industry is an important component of the economy due to its organisation complexity. Furthermore, it is an extremely capital intensive branch of industry which employs a great number of people and represents a leader in terms of technical innovations. In the past two decades this industry has been characterised by a series of significant processes, two of which have gained great importance. Firstly, there is a strong competition for markets, which has led to a reduction in the number of stand-alone Original Equipment Manufacturers (OEMs). Secondly, due to an increasing focus on their core competences, all OEMs transfer value chain activities to suppliers. As a result of these processes, a tight cooperation between OEMs and suppliers in the form of a supply-chain network has been established. Therefore, the main tasks of OEMs consist of recognising the customers' needs and optimal coordination of the suppliers (Håkanson and Lind, 2004, p. 53). Due to their low participation in creation of value, the OEMs are forced to identify and realize potential cost reduction opportunities, in order to maintain competitive advantage. With regard to this, the OEMs often use different cost accounting systems to make basic framework of open book accounting, which requires exchange or at least disclosure of largely detailed and relevant costs and other information on the side of the suppliers (Kajüter and Kulmala, 2005, p. 187, Cooper and Slagmulder, 2004, p. 13). This concept is extremely important for an OEM in the relationship to system and module suppliers. This paper aims to explain the main areas of application and distinctions of two activity-oriented cost accounting

* Ph.D., Assistant Professor, Singidunum University, Department of Business Administration, Danijelova 32, 11000 Belgrade, Serbia, e-mail: vmarjanovic@singidunum.ac.rs
** Teaching Assistant, Singidunum University, Department of Business Administration, Danijelova 32, 11000 Belgrade, Serbia, e-mail: jgavrilovic@singidunum.ac.rs
*** M.Sc., Teaching Assistant, Singidunum University, Department of Business Administration, Danijelova 32, 11000 Belgrade, Serbia, e-mail: nstanic@singidunum.ac.rs

systems in automotive industry: US American activity-based costing (hereafter US ABC) and German activity-based costing (hereafter German ABC). The paper is based on theoretical and practical research of these cost accounting systems. The main aim of the research was to provide some insight into the relationships between partners within supply chain networks and to facilitate the decision-making processes at all management levels with help of mentioned cost accounting systems. Moreover, a great deal of attention was paid to providing evidence that these cost accounting systems need to be implemented in different business areas. The paper consists of five sections. The section immediately following the introduction is devoted to the reviews of main literature on cost accounting systems which are based on activities and processes. The third section explains the proposed hypotheses and research methodology of the paper. The fourth section discusses the main results of the research on the cost accounting systems in question. The paper closes with comprehensive conclusions.

2. LITERATURE REVIEW

All cost accounting systems which are based on activities and processes have one significant common feature - the overhead cost allocation is carried out with help of certain cost drivers, which represents the repetition coefficient of chosen activities that cause overhead costs (Schweitzer und Küpper, 1998, p. 333). In management business administration literature which refers to cost accounting issues, these cost accounting systems were given various titles, such as activity-based costing, activity accounting, transaction costing, cost driver accounting system, operation costing etc. (Däumler and Grabe, 1998, p. 226).

Early concepts of cost accounting systems based on activities and processes were developed in the sixties and seventies by General Electric in the USA (McConville, 1993) as well as Siemens and Schlafhorst in Germany (Ziegler, 1992; Wäscher, 1987). However, the first comprehensive and theoretically structured concept of these cost accounting systems was introduced in the eighties (Kaplan and Bruns, 1987, 204-28) through the establishment of US ABC. The underlying idea of this cost accounting system was the establishment of activities, which represent the small groups of homogeneous tasks in all departments within a factory. The activities represent mediators for cost allocation on final products and must be demarcated, so that the consumption of ressources caused by one activity, could be determined through a single cost driver. Once the cost driver is determined, all activities can be classified into one of four mutually exclusive categories (Zimmerman, 2003, p. 546): unit level, batch level, product level, or production sustaining. Using specific indicators for every selected activity, a cost pool will be constituted, which consists of all costs that arise from the performance of considered activities. With help of previously specified cost drivers, all of these cost pools apart from research and development costs, will be subsequently allocated to every single product. The cost of three hierarchical categories of activities (unit level, batch level and product sustaining) are directly traceable to final products by selected cost drivers. But, for the cost allocation of facility-sustaining activities, it is neccesary to find some additional value-based cost drivers. According to this overall cost allocation procedure, US ABC is regarded as an absorption cost accounting system.

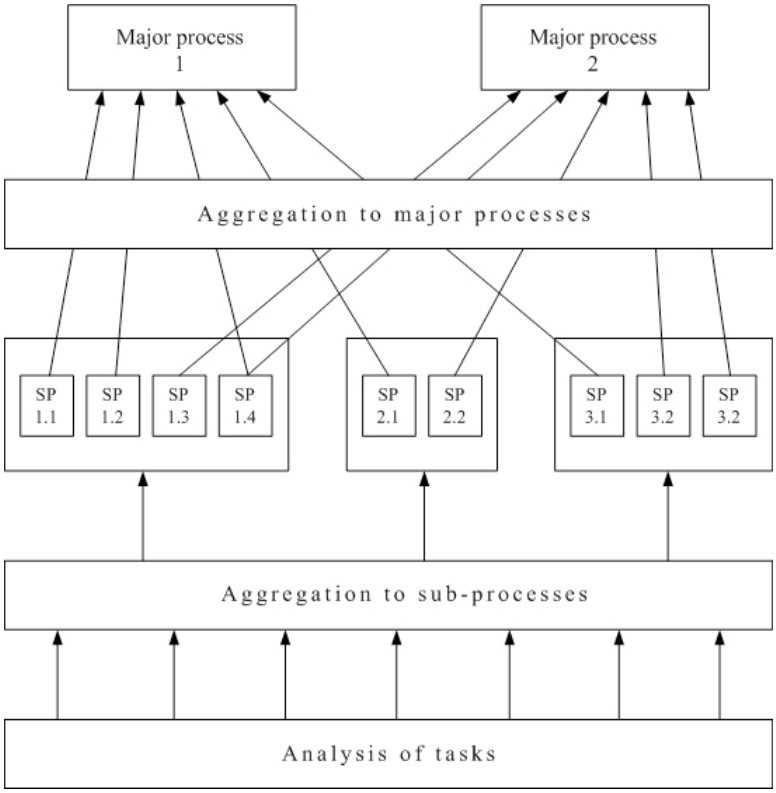


Figure 1. Aggregation of processes in German ABC (Horváth, 2008, p. 491)

German ABC (Ger. *Prozesskostenrechnung*) also represents a cost accounting system based on activities and processes and was introduced at the end of the eighties in Germany, by Horváth and Mayer (Horváth and Mayer, 1989). Despite the fact the term for this cost accounting system was literally translated into English as "process costing", they have nothing in common. Process costing is not based on activities and processes and allocates cost according to equivalent units which represent the amount of work actually performed on products which are not yet completely transformed into the work required to complete an equal number of overall units (Hilton et al., 2003, p. 308). This feature makes process costing an extreme case of cost averaging (Zimmerman, 2003, p. 464). US ABC has often been criticised for failing to provide a causual relationship between cost objects and cost drivers, which had been used for overhead cost allocation. It is also emphasized that marginal planned cost accounting (Ger. *Grenzplankostenrechnung*), as the most important Germany's cost accounting system, (Friedl et al., 2009, p. 39) provides higher causality level in cost allocation of the direct performance creation than US ABC. Marginal planned cost accounting is based on the principle of direct costing, which means that only variable costs are allocated to cost objects. The main features of marginal planned cost accounting include cost-type accounting, cost centre accounting, product cost accounting and contribution margin accounting (Friedl et al., 2009, p. 39). One of the distinct features is the separation of fixed and variable costs at the level of cost-type accounting. In contrast to that, fixed costs are used only in the profit-and-loss statement because they contain no information for short-term decision making. In this regard, the profit-and-loss statement in marginal planned cost accounting is intended to serve mainly for internal purposes and is called contribution margin accounting. German ABC deals with processes (major and sub-processes) which are aggregated from a string of activities (Figure 1). The major processes represent a chain of homogenous activities which have been the subject of the same cost factor. The activities can be regarded as homogenous if there is no substantial difference between the expenditure of work and the usage of resources. The overhead cost will be allocated to the cost objects through the cost drivers of major processes. On the other hand, the sub-processes were

defined as a chain of homogenous activities of one cost centre. They can be aggregated into one or more major processes. The main criterion for the aggregation is not the established cost drivers, but an essential belonging to one major process (Schweitzer and Küpper, 1998, p. 339). According to the repetition frequency of the cost centre performance, there is a distinction between value-added and non-value-added processes. The cost drivers will be determined just for the value-added processes because their output is a contingent on the cost centres performance. The unit rate for the cost allocation of non-value-added processes arises from their relationship to the cost of value-added processes (Coenenberg, 1999, p. 232).

The implementation of US and German ABC shows significant differences in automotive industry. They have resulted from different experiences in the development of cost accounting systems. With the increasing integration of automotive suppliers in the process of value creation, the purchase management has gained a strategic significance (Large, 2009). This has emphasized the necessity for transfer of information in the chain of value creation and the implementation of Supply Chain Controlling (Westhaus, 2007). Hence, the responsibility in terms of the purchase decisions is extended along the entire value chain. (Seuring and Koplin, 2010). This has resulted in a comprehensive understanding of the purchase and supplier management, which also takes into consideration risk management and suppliers' performance (Seuering and Müller, 2008). US and German ABC are also characterised by different areas of application in automotive industry. US ABC was mainly focused on the activities of the production and production-supporting cost centres, such as stamping press, assembly of bodywork, varnishing process and paintwork, final assembly, etc. German ABC is not concerned with these processes because they are the focus of marginal planned cost accounting, which has provided satisfactory results in cost calculation, especially in the mass production. The main differences between US and German ABC are illustrated in Figure 2.

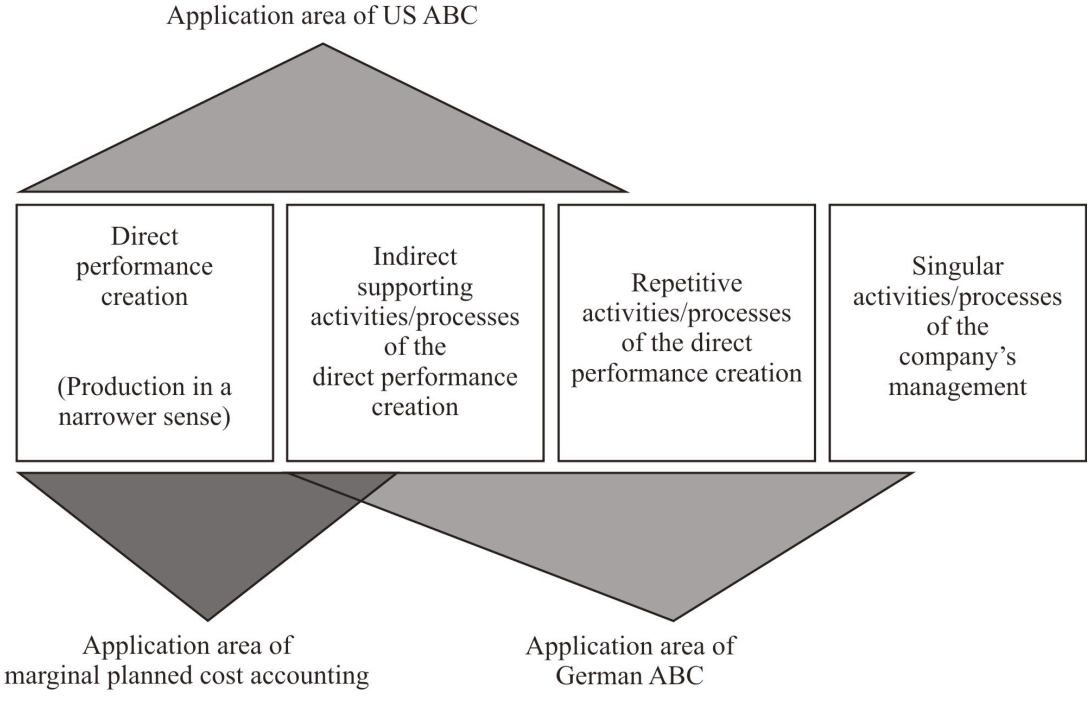


Figure 2. Application areas of US and German ABC (Horváth, 2008, p. 490)

Unlike US ABC, German ABC is mainly concerned with sub-processes and processes after the production area, in which large amounts of overhead costs arise. These sub-processes and processes in modern automotive industry include tasks such as design, logistics, sale and maintenance. Recent developments in the automotive industry also showed, that almost all OEMs have shifted their attention from the direct production to the

so-called non-core competences, such as banking, insurance or retail, due to the fact that higher profits can be made in this business area. As a result of this, the OEMs transfer more and more activities related to the direct production and assembly to the suppliers in order to gain additional room for more profitable activities.

US and German ABC also show significant differences in the quality and quantity of cost information provided, although both of them use cost centres and cost drivers. There are more activities in US ABC than processes in German ABC, but significantly less than sub-processes (Kellermanns and Islam, 2004, p. 37). German ABC uses tasks which are combined with sub-processes and both are linked to just one cost centre. US ABC combines tasks with activities, which is similar to combining sub-processes with processes, but activities relate to only one larger cost centre (Gaiser, 1998). That said, the level of aggregation of sub-processes in German ABC is partway between tasks and activities of US ABC.

In addition to the different areas of application, US and German ABC have a different role in decision making processes and supply-chain management in automotive industry. For the purpose of cost allocation to the objects, US ABC excludes only research and development costs, which makes it an absorption costing system. For the same purpose, German ABC deals only with the cost of value-added processes, whereas the costs of non-value added processes are collected in a cost pool and allocated using a proportional rate. In other words, the cost allocation of non-value added processes is not activity-oriented. For that reason, German ABC is often considered a variable costing system (Küting and Lorson, 1991, p. 1426). Some authors also argue that all entities use absorption costing for external financial statements and tax purposes (Horngren et al., 1997, p. 309). However, there are two versions of US ABC related to an absorption costing system: traditional and super US ABC (Reinstein and Bayou 1997, p. 492). If used simultaneously, they would not show the same income after the calculations, because the diversity of their methods produces different inventoriable costs. This fact creates some difficulties with integration of US ABC into the external accounting report system. Therefore, it is advisable to use international financial reporting standards (IFRS) and generally accepted accounting principles (GAAP) for external accounting and inventory evaluation and to use US ABC for other purposes. In contrast to US ABC, German ABC only provides information for internal accounting. Due to its features and being a variable costing system, German ABC draws a clear distinction between relevant and irrelevant costs, which makes it more suitable for decision making processes.

As stated earlier, one of the main features of modern automotive industry is the arrangement of tight connections between OEMs and suppliers through supply chain network. Since the OEMs and suppliers have become increasingly integrated, it remains to be considered how activity-oriented cost accounting is or can be designed and used to assist in the formulation, implementation and realization of strategies for achievement of competitive advantages (Ramos, 2004, p. 134). In that sense, US and German ABC are considered to be a part of open-book accounting concept, which implies that all participants in the supply chain network show willingness to allow the access to their internal accounting data. This concept facilitates development of more trusting and harmonious relationships between OEMs and suppliers. Thus, US and German ABC can identify additional services provided by suppliers, such as product design and development, which add value to the final product and reduce internal costs such as inspection and handling (Ramos, 2004, p. 135).

3. HYPOTHESES AND RESEARCH METHODOLOGY

Focus on cost centres and overhead cost types is one of the main prerequisites for the application of activity-oriented cost accounting systems. Cost centres represent a specific pool of collected homogeneous activities and processes, which play a role of mediators in overhead cost allocation. However, the quantity and structure of the cost centres have

different significance in US and German ABC. As stated earlier the process in German ABC is combined with sub-processes which are linked to single smaller cost centres, while the activity in US ABC relates to one larger cost centre. Higher number of cost centres also implies a lower average number of employees per cost centre and enhanced cost control. Therefore, the following hypotheses are proposed:

H1: Applicability of German ABC rises as the number of cost centres rises.
H2: Applicability of German ABC rises as the number of overhead cost types rises.

The choice between different software solutions is also an important factor in the implementation of activity-oriented cost accounting systems. SAP-System with its software solutions SAP ERP, Customer Relationship Management (SAP CRM), Product Lifecycle Management (SAP PLM) Supplier Relationship Management (SAP SRM) and Supply Chain Management (SAP SCM) offers a great deal of possibilities for cost centres control, especially through a clear demarcation between planned, actual and budget costs (Brück and Raps, 2004, p. 261), which is, to a large extent, embedded in German ABC. Moreover, SAP-System is suitable for structuring of different development projects in automotive industry (Kohlhoff, 2005). The development and design tasks are mostly placed outside direct production area, which makes German ABC eligible for allocation of their overhead costs. With regard to this, the proposed hypothesis would be:

H3: Applicability of German ABC rises as SAP-System is implemented.

US and German ABC often practice multilayered contribution margin accounting for different purposes. German ABC is regarded as a supplement to marginal planned cost accounting which is concerned with the overhead costs calculation in the direct performance creation. It uses contribution margin accounting in the indirect performance creation area, producing information which facilitates decision-making process about additional order acceptances, make-or-buy problems, short-term price decisions etc. On the other side, US ABC is directly concerned with the cost calculation in the direct performance creation but unlike marginal planned cost accounting, its separation into fixed and variable costs, direct and indirect costs is not so clear. For that reason, US ABC's contribution margin accounting is more suitable for long-term decisions concerning product-mix, improving-business processes or managing cost structures. Following these theoretical explanations, the proposed hypotheses concerning the contribution margin accounting would be:

H4: Applicability of German ABC rises as contribution margin accounting is used more for the purposes of short-term decision support.
H5: Applicability of US ABC rises as contribution margin accounting is used more for the purposes of long-term decision support.

The research on applicability of US and German ABC was conducted in two separate periods. The first research was done over a nine-month period in 2006 and 2007 including 123 automotive companies in Germany (Bavaria and Baden-Wurttemberg) and Austria. In the second period from October 2009 to March 2010, the survey was supplemented with data from thirty-one Serbian companies that have directly or indirectly been involved in automotive industry. Although this research was conducted in two separate periods, it was based on the same methodology of which the original purpose was to provide some practical insight into the most frequent cost accounting systems in automotive industry. A chosen group of companies was firstly contacted by telephone and, after obtaining either a verbal or written statement for cooperation a seven-page questionnaire was sent to the cost accounting departments of the chosen companies. The questionnaire consisted of thirty-three questions including twenty-seven questions with answers offered in advance, which

considerably facilitated statistical analysis. Most of these multiple-choice questions were designed to identify the level of respodents' agreements with the offered answers on the scale from one (do not agree) to five (fully agree). Among 123 automotive companies in Germany and Austria, which represented a return rate of 19.12 percent of filled in questionnaires, 0.06 percent were OEMs, 19.63 percent system integrators, and 42.03 percent system specialists. The remaining 38.28 percent consisted of suppliers of parts and components. The structure of surveyed companies in Serbia was quite different. Thirty one analysed companies in Serbia represented a return rate of 38.81 percent of filled in questionnaires. There was one OEM (3.22 percent) involved in the survey, 6 system integrators (19.35 percent) and 24 suppliers of parts and components (77.42 percent). Eight questions were chosen from the questionnaire as the most relevant for assessing the influence on applicability of US and German ABC under specific business circumstances (Table 1). In order to assess this influence, elements of the descriptive statistics, such as statistical mean, standard deviation and correlation coefficient, covariance and mathematical expectation, were used. Furthermore, structural equation model and Pearson's coefficient were used for the testing the hypotheses.

Questions	Abbreviation	Offered answers
Number of cost types (CT)	CT1	<100
	CT2	100-300
	CT3	>300
Number of cost centres (CC)	CC1	<10
	CC2	10-30
	CC3	>30
Average number of employees per cost centre (ACC)	ACC1	<5
	ACC2	5-10
	ACC3	>10
Systematic separation of fixed and variable cost (FVC)	FVC1	Yes
	FVC2	No
Systematic separation of direct and indirect costs (DIC)	DIC1	Yes
	DIC2	No
Contribution margin accounting (CMA)	CMA1	Yes
	CMA2	No
Software solution for cost accounting (SS)	SS1	SAP (R/3, BW, ERP)
	SS2	Oracle
	SS3	Peoplesoft
	SS4	SAS
	SS5	Other
Cost accounting purposes (CAP)	CAP1	More transparency of indirect costs
	CAP2	Short-term decision support
	CAP3	Long-term decision support
	CAP4	Higher level of efficiency for cost control
	CAP5	Motivating employees

Table 1. Questions for assessing the influence on applicability of US and German ABC

4. RESEARCH RESULTS

The research has shown that 42.3 percent of all surveyed automotive companies clearly stated that they use US ABC as a stand-alone cost accounting system, while German ABC is represented with 38.2 percent. However, these cost accounting systems are used to a large extent as supplements to other related cost accounting systems. In fact, the application level of these cost accounting systems is much higher. Many of the surveyed companies which explicitly stated that they do not practice US or German ABC actually meet all

necessary conditions for their implementation. The key element for assessing the influence on applicability of US and German ABC is structural equation model. This model was used to estimate the influence (β) of specific factors from the chosen questions on the above mentioned activity-oriented cost accounting systems (Figure 3). According to correlation analysis, correlation coefficients were calculated for each of variables that were considered relevant for the applicability of US and German ABC (Table 2). From the listed variables, number of cost types, number of cost centres, systematic separation of fixed and variables costs and systematic separation of direct and indirect costs were considered to have stronger influence on German ABC than on US ABC. On the other side, average number of employees per cost centre was regarded as more relevant for US ABC. All other variables had twofold effect on US and German ABC.

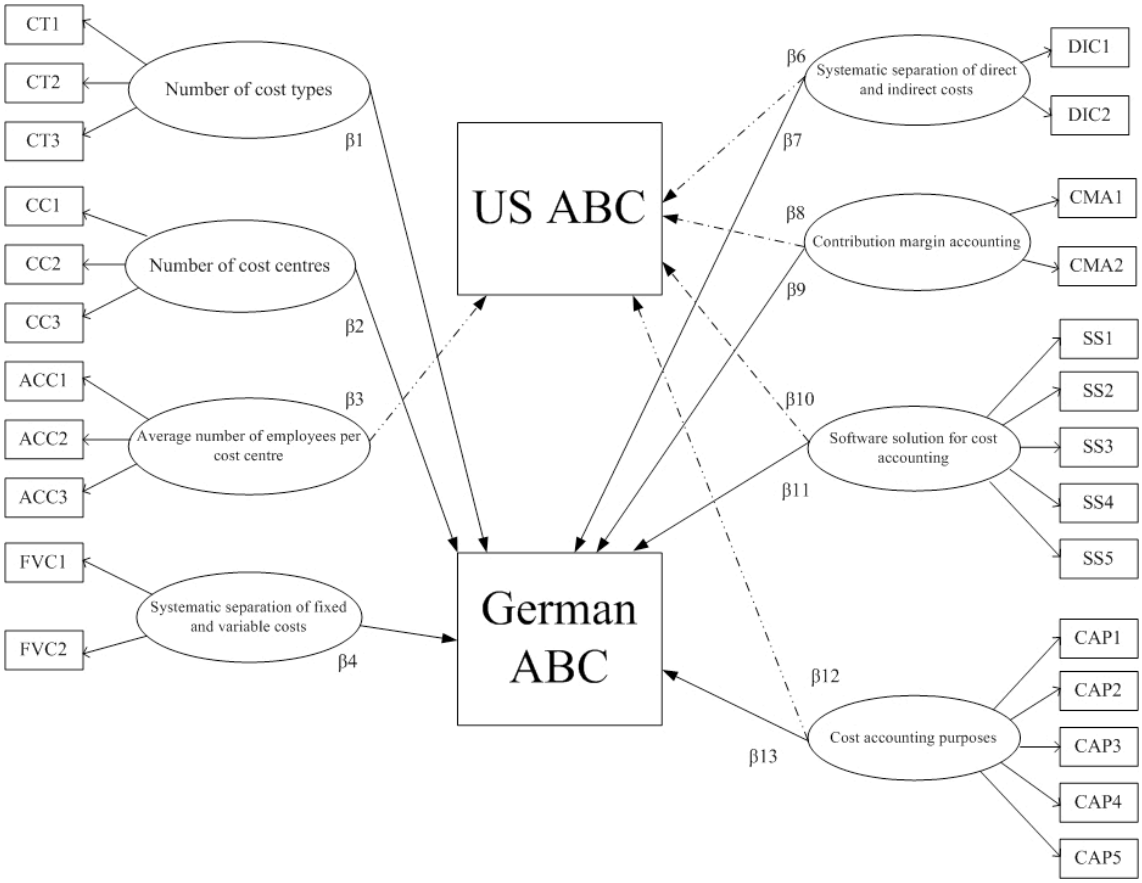


Figure 3. Structural model of the effects on US and German ABC

All of the surveyed automotive companies use cost type accounting and 84.6 percent practiced cost centre accounting. The average number of cost type accounts was 136.54 and of cost centres 36.2. With regard to this, the average number of employees per cost centre was 12.6. Moreover, fifty percent practice separation of fixed and variable costs. The results have also showed that 73.1 percent of all respondents differentiated between direct and indirect costs. If they practiced separation of direct and indirect costs to a larger extent they would fulfil more prerequisites for the implementation of German ABC (C=0.74). Non-separation of direct and indirect costs better was better suited for US ABC (C=0.89). The application of contribution margin accounting has almost equal influence on US ABC (C=0.64) and German ABC (C=0.69) but with rather different accounting purposes. Focus on long-term or strategic decisions would favour the implementation of US ABC. One of the most important factors for the application of cost accounting systems in Serbian companies is information technology support which refers to different brands of used software solutions. In fact, almost 50 percent of the surveyed companies “fully” or “largely” agree that software

was the main determinant in the adoption of their present cost accounting system. However, the results showed that only 7.7 percent of all surveyed automotive companies used SAP-System for the implementation of German ABC (C=0.12). In the case of US ABC, this brand of software was with 8.9 percent slightly more represented (C=0.14). The implentation of these activity-oriented costs accounting systems was rather supported by other software solutions.

Variables	ABC	PrKR
CT1	0.80*	0.24*
CT2	0.16*	0.36*
CT3	0.31*	0.72*
CC1	0.21*	0.09*
CC2	0.56*	0.23*
CC3	0.44*	0.51*
ACC1	0.21*	0.46*
ACC2	0.75*	0.25*
ACC3	0.85*	0.37*
FVC1	0.42*	0.56*
FVC2	0.63*	0.61*
DIC1	0.24*	0.74*
DIC2	0.89*	0.39*
CMA1	0.64**	0.69**
CMA2	0,11**	0.08**
SS1	0.14**	0.12**
SS2	0.26**	0.22**
SS3	0.10**	0.18**
SS4	0.21**	0.15**
SS5	0.29**	0.58**
CAP1	0.69**	0.84**
CAP2	0.63**	0.12**
CAP3	0.45**	0.18**
CAP4	0.27**	0.31**
CAP5	0.10**	0.05**

* *Statistically significant at the level of p<0.01*
** *Statistically significant at the level of p<0.1*

Table 2. Correlation coefficents (C) for variables

Cost accounting purposes also play an important role in the implementation of US and German ABC. Automotive companies who chose *more transparency of indirect costs* as the most important cost accounting purpose have reached it almost completely by implementing both US ABC (C=0.69) and German ABC (0.84). The purpose of *short-term decision support* was better satisfied by using German ABC (C=0.63) than US ABC (0.12). On the other hand, *long-term decision support* was to a larger extent achieved by US ABC (0.45) than by German ABC (0.18). German ABC was more relevant (C=0.31) for the purpose of *higher level of the efficiency for the cost control* than US ABC (0.27), while the purpose of *motivating employees* could hardly be met by either of these cost accounting systems.

Hypotheses	Coefficients	Conclusion
H1	0,18*	Supported
H2	0,34*	Supported
H3	0,03*	Not supported
H4	0,11*	Supported
H5	0,27*	Supported

Statistically significant at the level of p<0.05

Table 3. Pearson's coefficient for testing the hypotheses

The structural model was designed to extract all the relevant factors from the questions used for assessment the influence on applicability of US and German ABC. Eight questions with their twenty-five of variables were isolated from the whole questionnaire on the cost accounting systems in automotive industry. Six questions with eighteen variables were used for specification of five hypotheses. Pearson's two-tailed correlation coefficient was applied for the test and the calculated values are given in the Table 3. Apart from the hypothesis three, all others had been statistically supported. The main reason for proposing the hypothesis on positive relationship between SAP-System and German ABC consisted in the previously described fact that German ABC is not concerned with the direct performance creation which is threated by marginal planned cost accounting. This system was originally embedded in SAP software solutions and that leaves plenty of room for German ABC to be engaged in the indirect performance creation area, which gradually becomes the main point of interest of modern automotive industry. However, the research results have revealed that other software solutions can significantly contribute to the application of activity-oriented cost accounting systems.

5. CONCLUSIONS

The purpose of this paper was to offer insights into the role and effects of US and German ABC in the decision-making processes of modern automotive industry. This industry has experienced drastic changes over the two last decades with respect to supply-chain management. As a result of the takeover actions and concentration on the core competences, the number of stand-alone OEMs has been decreasing. This new situation has resulted in new close relationships between OEMs and suppliers. The newly-formed supply chain network facilitates decision-making process at all management levels. However, the maintenance of such networks requires different information from the cost accounting systems. A great deal of required information can be provided from US and German ABC. Both of these cost accounting systems are activity-oriented and have a common purpose in the automotive industry - to improve the cost calculation and optimize the overall creation of value. But, they also have a series of differences with respect to the area of application. US ABC is primarily focused on the direct performance creation, whereas German ABC is mostly connected with the supporting area of production in a narrower sense and with the company's management. Considering the automotive industry, German ABC is more suitable for the application by the OEMs, because they have transferred to the suppliers the large part of activities related to the direct production. German ABC is also perceived to be more detailed and provides greater quantity and quality of information, from which one would expect a high perception of usefulness (Kellermanns and Islam, 2004, p. 43). Despite the provided quantity of information, German ABC has often been criticised for use of the costs of value-added processes as a base for the allocation of costs of non-value-added processes. This decreases the causality level between the cost drivers and cost objects. However, the precision of cost allocation is not a primary goal of German ABC, but the process optimisation is. US ABC was designed to avoid the problem of cost averaging by using activities for the allocation of all types of costs. Although it represents a great advantage for the practice of cost accounting, many automotive companies have been reluctant to implement US ABC, and this system has already been described as "yesterday's hope" (Thomson and Gurowka, 2005, p. 28). These cost accounting systems show a great effectiveness when combined with other systems, such as target costing for the product design or traditional and variable costing for the regulation of the inventory level. However, the biggest advantage of German ABC has been shown in the making of so-called special decisions, i.e. make or buy, sell as it was or process furthermore, drop or keep a product, accept or not a special order, etc. The underlying concept for these decisions is the application of multi-stage contribution margin accounting process, which

presupposes a gradual deduction of fixed costs. Some authors argue that US ABC is a much more widely published and discussed topic than German ABC, due to the dominance of US business literature and the English language (Kellermanns and Islam, 2004, p. 44). Instead of one-sided solution, more effort should be done to reconcile US and German ABC, since merger or acquisition cases on the automotive market may create significant potential conflicts.

REFERENCES

1. Brück, U., Raps, A. 2004. *Gemeinkosten-Controlling mit SAP*, Bonn, Galileo Press.
2. Coenenebrg, A., 1999. *Kostenrechnung und Kostenanalyse*. 4. Auflage, Landsberg am Lech, Moderne Industrie.
3. Cooper, R., Slagmulder, R., 2004. Interorganizational Cost Management and Relational Context. *Accounting, Organizations and Society*, 29(1), 1-16.
4. Däumler, K.D., Grabe, J., 1998. Kostenrechnung 3. Plankostenrechnung. 6. Auflage, Berlin, Neue Wirtschafts-Briefe.
5. Friedl, G., Küpper, H.U., Peddel, B., 2005. Relevance Added: Combining ABC with German Cost Accounting. Strategic Finance, 3(6), 56-61.
6. Friedl, G., Hammer, C., Pedell, B., Küpper, H.U., 2009. How do German Companies run their Cost Accounting Systems?. Management Accounting Quarterly, 10(2), 38-52.
7. Gaiser, B., 1998. Prozesskostenrechnung und Activity-based costing (ABC). Prozeßkostenmanagement, München, Horváth & Partner.
8. Håkanson, H., Lind. J., 2004. Accounting and Network Coordination. Accounting, Organizations and Society, 29(1), 51-72.
9. Hilton, W.R., Maher, W.M, Selto, H.F., 2003. Cost Management. Strategies for Business Decisions. 2th Edition, New York, McGraw-Hill.
10. Horngren, C.T., Foster, G., Datar, S.M., 1997. Cost Accounting. A Managerial Emphasis. 9th Edition, New Jersey, Prentice-Hall
11. Horváth, P., 2008. Controlling. 11. Auflage, München, Vahlen.
12. Horváth, P., Mayer, R., 1989. Prozesskostenrechnung. Der neue Weg zu mehr Kostentransparenz und wirkungsvolleren Unternehmensstrategien. Controlling, 1(4), 214-219.
13. Kajüter, P., Kulmala, H.I., 2005. Open Book Accounting in Networks. Potential Achievements and Reasons for Failures. Management Accounting Research. 16(2), 179-204.
14. Kaplan, R.S., Bruns, W., 1987. Accounting and Management. A Field Study Perspective, Harvard Business School Press, Boston MA.
15. Kellermanns, F.W., Islam, M., 2004. US and German activity-based costing. A critical comparison and system acceptability propositions. Benchmarking. An International Journal, 11(1), 31-51.

16. Kohlhoff, S., 2005. Produktentwicklung mit SAP in der Automobilindustrie, Bonn, Galileo Press.

17. Küting, K., Lorson, P., 1991. Grenzplankostenrechnung versus Prozeßkostenrechnung. Quo vadis Kostenrechnung?. Betriebs-Berater, 46(1), 1421-1433.

18. Large, R., 2009. Strategisches Beschaffungsmanagement-Eine Praxisorientierte Einführung. Mit Fallstudien, Wiesbaden, Gabler.

19. McConville, D.J., 1993. Start with ABC. Industry Week. 242(17), 33-35.

20. Ramos, M. M., 2004. Interaction between Management Accounting and Supply Chain Management. Supply Chain Management. An International Journal, 9(2), 134-138.

21. Reinstein, A., Bayou E.M., 1997. Product Costing Continuum for Managerial Decisions. Managerial Auditing Journal, 12(9), 490-7.

22. Schweitzer, M., Küpper, H.U., 1998. Systeme der Kosten- und Erlösrechnung. 7. Auflage, München, Vahlen.

23. Seuring, S., Koplin, J., 2010. Nachhaltigkeit im Beschaffungsmanagement der Volkswagen AG. Wirkungen auf mittelständische Unternehmen. Betriebswirtschaftliche Forschung und Praxis (BFuP), 62(1), 71-81.

24. Seuering, S., Müller, M., 2008. From a Literature Review to a Conceptual Framework for Sustainable Supply-Chain Management. Journal of cleaner Production, 16(15), 1699-1710.

25. Thomson, J., Gurowka, J., 2005. Sorting Out the Clutter. Strategic Finance, 87(2), 27-33.

26. Wäscher, D., 1987. Gemeinkosten-Management im Material- und Logistik-Bereich. ZfB, 57(3), 297-315.

27. Westhaus, M., 2007. Supply Chain Controlling-Definition, Forschungsstand, Konzeption,. Wiesbaden, Gabler.

28. Ziegler, H., 1992. Prozeßorientierte Kostenrechnung im Hause Siemens“ BfuP. 42(4), 304-18.

29. Zimmerman, J.L., 2003. Accounting for Decision Making and Control. 4th Edition, New York, McCgraw-Hill.

AMERIČKA VS NJEMAČKA ACTIVITY-BASED COSTING (ABC) METODA. UČINCI NA UPRAVLJANJE POSLOVNIM ODLUKAMA U AUTOMOBILSKOJ INDUSTRIJI

SAŽETAK

Ovaj rad nastoji opisati ulogu koju sustavi obračuna troškova utemeljeni na aktivnostima, tj. američki i njemački sustav obračuna troškova po aktivnostima imaju u uspostavljanju mreže opskrbe u suvremenoj automobilskoj industriji. Ovi sustavi obračuna troškova su predmet analize iz dva razloga: relativno uspješno opisuju princip uzročnosti

između izazivača troškova i mjesta troškova i predstavljaju dva različita pristupa obračuna troškova unutar koncepta koji se temelji na aktivnostima. Teži se dokazivanju da učinkovitost ovih sistema ovisi od toga u kom području lanca stvaranja vrijednosti su primjenjeni. Iako se temelje na istom konceptualnom okviru, oni nemaju iste ciljeve obračuna troškova u automobilskoj industriji.

***Ključne riječi**: obračun troškova, activity-based costing, automobilska industrija, lanac opskrbe*

***JEL klasifikacija**: M41, M49*

Selma Kadić, MBA[1]
Mirko Palić, PhD[2]
Muris Čičić, PhD[3]

UDK 339.371.2:65.012.4>:629.331>(497.6:497.5)

IMPROVEMENT OF PERSONAL SALES EFFICIENCY IN AUTOMOTIVE RETAILING

ABSTRACT

Sale has existed for centuries, but the principles behind it have changed over time. Contemporary theory emphasizes increase of customer productivity through strategic sales organization that converges with marketing in order to achieve greater strategic importance in the organization. This study uses a qualitative research method (in-depth interviews) with the aim of determining the relationship between sales and marketing, as well as learning about the perceptions of sales managers on the definition of sales strategy and atmosphere design as variables that affect sales efficiency in automotive retailing. Research results from Bosnia and Herzegovina and Croatia were compared and they indicate a lack of coordination between marketing and sales efforts/departments, and the inadequate use of atmospheric factors in both of the observed markets.

Key words: *personal selling, marketing, retail atmospherics, qualitative research*

JEL Classification: *M31, M53*

1. INTRODUCTION

Sales evolution begun with transactional sales, where there are many repeated purchases that are made routinely, automatically, and based on knowledge of the characteristics of the product/service being purchased. This approach was very efficient and fast, but with disadvantages: the buyer focuses mainly at the price and it can be very difficult to launch new products or to encourage switching to new products with added value. In the late 60's and early 70's, due to increased competition and complexity of leads, customers realized that transactional purchase provided non-optimal choice. The solution is consultative customer-oriented sales, with a strong recognition of needs based on the successful communication between buyer and seller (Manning and Reece, 2008). During the 80's, sales moved to strategic sales, in which companies began strategic sales planning. In the 90's, partnerships were created as strategic long-term relationships that solve the customer's problems by building relationships based on a win-win philosophy (Paparoidamis and Guenzi, 2009). Manning and Reece (2008) recommended a move towards the creation of value for the customer, where each sale has added value that strengthens the customer's experience through an improved creative selling process using technology and other contemporary tools.

Generating superior value for customers is based on building of a competitive advantage. The company creates superior value by offering benefits that are greater than the

[1] Senior Teaching Assistant at the School of Economics and Business, University of Sarajevo; Trg oslobodjenja – Alija Izetbegovic 1; e-mail: selma.kadic@efsa.unsa.ba
[2] Assistant Professor at the Faculty of Economics and Business, University of Zagreb; Trg J.F. Kennedya 6; e-mail: mpalic@efzg.hr
[3] Professor at the School of Economics and Business, University of Sarajevo; Trg oslobodjenja – Alija Izetbegovic 1; e-mail: muris.cicic@efsa.unsa.ba

costs that buyers must pay for a product or service. Superior value creates a competitive advantage and allows company to finance benefits from trade in the long term, which means that it creates such value more effectively and efficiently than its competitors (Slater and Narver 2000). A strong market-oriented organizational culture is an effective way to create superior value for customers (Day, 1994).

In today's competitive market, a sale is a critical element in an effort of modern companies to achieve organizational success based on consumer satisfaction, loyalty, and profitable sales volume (Anderson, 1996). In order to meet the needs of consumers, sales staff must work in cooperation with consumers as "helpers" rather than competing with them as the "persuaders" (Kirman and Campbell, 2004). This is why Manning and Reece (2008) considered that development of sales philosophy should involve three basic assumptions: the adoption of marketing concepts, evaluation of personal selling, and assumption of the role of problem solver, or a partner who helps the customer in making decisions.

As customers become more sophisticated and better informed (McDonald et al., 2000), the sales process was much less about the sale of products, and much more about the creation of relationships, which leads to involving the sales force in strategy creation through field information gathering and intelligence systems. Futurell defines the sales process as a sequential series of the seller's actions that leads towards buyers' desired activity and finishes with various forms of post-selling services that ensure the satisfaction of buying (Tomašević Lišanin 2010). It follows that sales management can be defined as the process of planning, implementing and monitoring the functions of personal selling (Cummings, 2004), which includes planning, finding and training staff, budgeting, developing compensation plans, and evaluating the effectiveness of the sales team (Manning and Reece, 2008).

Although different authors have observed the sales process differently and there are differences in relation to industry or types of products/services, most definitions include: identifying customers and their needs, preparing and arranging sales meetings, disclosure of customer needs and presenting solutions, handling objections, and possibly on the basis of a win-win philosophy, concluding the sale and post-selling services (Tomašević Lišanin 2010). The sales process is not time limited, and it is possible to successfully conduct all its phases in a short period or it can last for months or even years. Regardless of the stage in the process, the seller's goal is to successfully finish the current phase and to move to the next stage.

The strategic business plan of a company should be a guide for sale strategy. The strategic business plan includes strategies that the company uses for positioning in relation to the potential buyer before the sale even starts. On the other hand, the sales strategy is the plan of sales activities through methods of reaching clients, competitive differences and resources (Storbacka et al., 2009). Gosselin and Heena (2003) argue that a strategic deal with clients reported changes in the sales strategy, while Homburg and Pflesser (2000) go a step further by claiming that the new sales strategy leads to organizational changes in the company.

2. DESIGNING ATMOSPHERE TO IMPACT BUYING BEHAVIOR

Organizational changes in the company should consider broader usage of consumer perception knowledge with the aim to improve sales efficiency. In year 1973 Kotler wrote about the atmosphere at the selling point as one of the influential factors in creating the perceptions of consumers, as well as creating a stimulating retail environment for sales efficiency improvment. Kotler (1973) defines the atmosphere as a conscious space design that has effects on customers and clients. The atmosphere and environment of the retail space should be designed in a way that positively affects consumers' perception of the purchase and operation.

In the past thirty years several studies on the impact of atmosphere on buying behavior have been conducted. Donovan and Rossiter (1982) argue that emotional reactions are shaped by the environment at the point of sale in order to influence individuals to spend

more money than they originally planned. According to the Mehrabian - Russell SOR model (stimulus - organism / condition - reaction), shown in Figure 1, stimuli from the environment influence consumers' emotional states, which in turn influences the behaviors of approach and avoidance. The outcome behavior of approach refers to the movement toward and behavior to avoid refers to the movement of different stimuli from the environment.

Figure 1: Model SOR (stimulus - organism / state - response)

Source: Mehrabian, A., Russell, J.A. (1974) *An Approach to Environmental Psychology*: MIT Press, Cambridge, pp. 126

Knowing that 75% of emotions are associated with the sense of smell, and not with eyesight or hearing, it is clear that a brand, in order to survive, must provide special relations to the customer through complete sensory and emotional experience and connections to the products. Lindstrom (2005) proved that, if companies want to survive, brand management must evolve from a two-dimensional approach to a five-dimensional concept because more sensory information in the memory triggers a stronger connection to a brand.

The question is how to positively affect all of the human senses. Using music can attract the buyer's attention; shape and image can arouse certain emotions that will increase the possibility of buying. Studies (Gobe, 2006; Broekemier, Marquardt and Gentry, 2008) proved that music at the selling point, consistent with business objectives and target segment, may affect the buyer's desire to purchase and the time spent at the point of sale.

The sense of smell is the only sense that cannot be "turned off", because every breath leads to smell and this happens on average 20,000 times a day. Marketers believe that the key factor in brand sensing management is association with emotions. Some companies do not want to admit application of scent in their retail facilities because they are afraid of being accused for using subliminal perception, causing subconscious association to the brand and manipulating the selling environment with the aim of activating Pavlovian reactions among consumers (Trivedi, 2006). The sense of taste is closely linked to the sense of smell. For most companies it is difficult to incorporate taste into their brand. The most engaging sense that dominates in comparison to other senses is a vision (Lindstrom, 2005). Lighting can be considered a sales tool and can be used to attract customers, to focus on the product and to guide a purchase. Colors also affect the perception of the passage of time, as it seems that time slows down and that objects act bigger under a red light. In contrast, time passes quickly and items seams smaller under blue light.

All mentioned researchers indirectly confirmed that better cooperation between marketing and sales, regarding sharing knowledge on atmosphere design at the sales point, will improve sales efficiency.

3. THE CONFLICT BETWEEN MARKETING AND SALES

Although significant research efforts have been dedicated to the cooperation of functional departments in organizations, relatively little attention is paid to the inter-relation between sales and marketing. One reason for the neglect of this relationship is that customers typically see sales and marketing as functions serving the same goal (Cespedes, 1994; Le

Meunier-Fitzhugh and Piercy, 2007). However, in practice it is often the case that when the sales are disappointing, marketing blames the sales force for a poor execution of the marketing plan. The sales team, however, argues that marketing set prices too high, and used an excessive budget. In this situation, sales departments believes that marketers are not in touch with what is really happening on the market, while marketers believe that the sales force is short-sighted, insufficiently aware of the market and blind to the future.

Many studies of the cross-functional relationship between sales and marketing have agreed upon existence of many negative characteristics of this relationship (Dawes and Massey, 2005), including poor coordination especially in planning and setting goals (Rouzies et al., 2005), strong misunderstanding, mistrust, poor co-operation and conflicts (Anderson, 1996). One reason for the lack of collaboration between sales and marketing is the variety of philosophies and backgrounds of the members in these teams (Ruekert and Walker, 1987; Cespedes, 1994, Kotler et al., 2006). Sales staff are often guided by intuition, while the marketers tend to be more analytical, data-oriented and focused on projects for building competitive advantages for the future. Front line staff, on the other hand, spend time talking to existing and potential customers, and they not only have an awareness of the customers' needs, but assume which products will experience maturity and which will not be successful. Sales staff earns money by closing the sale and it is much easier to determine success or failure, while the marketing budget is divided into programs, and more time is needed to determine the effects of the creation of long-term competitive advantage.

Kotler et al. (2006) states that any company can and should improve the relationship between sales and marketing. However, this sentence shows one of the most contentious relationships within organizations and there is a general consensus that the relationship between sales and marketing is poorly coordinated, and that improving this can be beneficial to the organization (Cespedes, 1993; Dewsnap and Jobber, 2007; Rouzies et al., 2005, Kotler et al., 2006). Tjosvold (1988) found that cooperation between these departments leads to improved productivity and competitiveness.

Kotler et al. (2006) conducted a research on relationship between sales and marketing, with the aim of identifying best practices that could help strengthen joint performance. They identified existence of four different types of relationships: undefined, defined, harmonized and integrated, where each has specific repercussions to the organization and business performance. The same methodological tool was used in this research to test one of the hypotheses.

When the relationship is undefined sales and marketing evolve independently, and each department is busy with its own tasks and programs. Each function does not know much about the activities, tasks and plans of the other function until the moment when conflict arises. Meetings between the two groups occurs ad hoc only with the aim of resolving conflicts, not because of proactive cooperation.

In a defined relationship the two functions establish processes and policies for the prevention of disputes through orientation that is reflected under the motto: "A good fence makes good neighbors." Retailers and marketers know who needs to do what and stick to their tasks. Groups begin to find common ground in potentially controversial areas. Meetings are more reflective, and the staff members ask questions of mutual cooperation and work together on large events.

A aligned relationship is characterized by clear but flexible boundaries between functions. Both functions take part in joint planning and training. The sales team understands and uses marketing terminology, while marketers work with vendors on the important matters.

When sales and marketing are integrated, the boundaries become blurred, and both functions redesign relationships to gain a larger share in the structure and systems of reward in companies. Marketing and sales focus on strategic tasks, which are oriented in advance. Marketers are deeply involved in the management of key customers, and jointly develop and implement performance metrics. Budgets become flexible and less controversial.

4. AUTOMOTIVE RETAIL MARKET IN BOSNIA AND HERZEGOVINA AND CROATIA

During 2009, every major car brand reported a drastic sales decline, announced a reduction in production and cut down work force. A large part of the crisis in the automotive industry is associated with an increase in oil prices during the first part of 2008 and then by the raging financial and economic crisis that has swept the world. Part of the decline in demand is a result of the decline in consumption, caused by the impaired ability of banks to issue loans, and reduced consumer confidence. However, in early 2010, economists and experts predicted that market would overcome those difficult moments, and that the best producers would adjust and balance their businesses to the current demands, and that in the coming months they would be the best placed in global rankings of business.

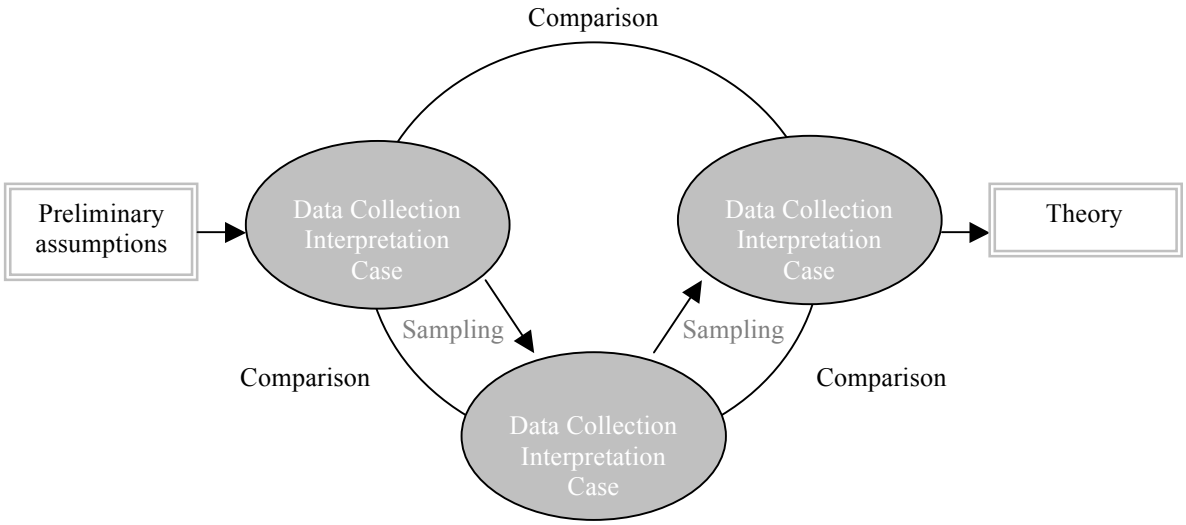
In the first three months of 2010 there was a total increase in new car registrations in Europe by 9.2% (approximately 3.7 million new sold cars) compared to the first quarter of the previous year. Compared to the same period in 2008 these results represented a decline of 9.4% (Santini, 2010). In the first ten months of 2009, in Croatia 38,506 new passenger cars were registered, which is 50% less compared to the same period in 2008 (Suvremena.hr, 2009). In the first half of 2009 the decline in sales in the B&H market was around 30%. The reasons for the fall in demand in both countries are: the economic crisis, uncertainty and fear. Drop in car sales, according to the field's experts, are closely related to the fact that in the domestic market which depended on availability of financing, a new tougher restrictions imposed by financial institutions and higher interest rates dramatically reduced number of newly approved loans.

5. RESEARCH METHODOLOGY

The research focuses on the question whether it is possible to improve sales effectiveness through the improvement of relations between marketing and sales, specifically through defining sales strategy in the creation of atmosphere at the point of sale. The aim of this paper is to determine opportunities for improving the efficiency of sales, to investigate the perceptions of sales managers about working with their marketing departments, and to determine the importance of the atmosphere at the sale point using qualitative research.

In contrast to the linear research approach, which includes development of a research model before the research is carried out, a circular approach suggests that the theory should not be applied to the subject under investigation, but that it should rather be "discovered" and formulated during the research (Flick, 2009). This approach, characteristic of qualitative methods, forces researchers to continuously take care of the entire research process, and not simply the phase presently preoccupying the researcher.

Figure 2: Preview of the circular research approach



Source: Flick U. (2009) *An Introduction to Qualitative Research*, 4th ed.: SAGE Publications: pp. 95

In accordance with the object and purpose of research, the paper used a circular approach to qualitative research, considering in this analysis the most important motives and attitudes of sales managers. The following hypotheses were set:

- H1: Companies in the car retail sector in Bosnia and Herzegovina and Croatia do not have a retail strategy.
- H2: The relationship between marketing and sales departments in companies in the car retail sector in Bosnia and Herzegovina and Croatia ranges from "undefined" to "defined" and has all the repercussions of that particular relation to business operations.
- H3: Sales managers who are familiar with the influence of the atmosphere at the sales point to customers design stores so that customers can experience a sensory and emotional experience.
- H4: Sales managers believe that investments in the retail atmosphere will result in higher customer satisfaction and therefore increase sale efficiency.

The circular approach model for the research process suggests that the sample should be selected with regard to the relevance of certain persons for the issue, and not obtained as a statistically representative sample (Flick, 2009). Therefore, the selected respondents were the sales managers in ten companies in Bosnia and Herzegovina and Croatia engaged in automotive retail. Attention was paid that the sample represents companies that have a significant market share. Examining the data obtained from agencies involved in the retail car market research in both countries (Puls Promotion Agency Zagreb and Sarajevo Plus) showed that market shares for the same brands are similar in both countries. The selected representatives work in companies that sell the same five cars brands in both countries, as follows: three companies that are in both countries ranked between 1^{st}-7^{th} within the market share and two companies that were ranked between place 8^{th} -20^{th}. It was ensured that the sample reflects the opinions of companies that have a total of around 40% of the market share in both countries.

Interviews were conducted through pre-formulated questions based on the data from the theoretical part of the paper and lasted between 60 and 120 minutes. The starting point for defining the research plan was the investigation of the electronic products sector conducted by Mallalieu and Nakamoto (2008). Questions were adapted to the car retail market, taking into account the significant changes that this market experienced during the recession.

During the interview, all participants were also asked to fill out questionnaire

concerning relations between the sales and marketing departments in their company. The questionnaire used was developed by Kotler et al. (2006) as an objective tool for defining the relationship between the marketing and sales departments in a company. The questionnaire consists of 20 statements that can be ranked on the Likert scale of five degrees. By summarizing answers expressed numerically a result is obtained that reflects the relationship of these two departments in the company. According to Kotler (2006) it is possible to identify four types of relationships: undefined (questionnaire result between 20-39), defined (from 40-59); harmonized (from 60-79), and integrated (from 80-100).

6. ANALYSIS AND PRESENTATION OF RESULTS

Distrust was noted among respondents regarding certain issues and topics. Distrust was increased with the fact that all conversations were recorded, and on the insistence of the respondents frequent interruptions occurred. However, respondents were willing to give answers to some questions on the condition that the recording stops. It was noted that respondents calculated with the questionnaire while giving grades for individual claims that were contradictory to what they previously said in the interview. In order to protect the anonymity of respondents a coding was made (Table 1).

Table 1: Coding

Brand / Country	**Croatia**	**Bosnia and Herzegovina**
Brand 1	HR1	BH1
Brand 2	HR2	BH2
Brand 3	HR3	BH3
Brand 4	HR4	BH4
Brand 5	HR5	BH5

The different economic and political structures of B&H and Croatia, and their situations regarding EU integration, (entrance of Croatia is expected in 2012, while the status of Bosnia and Herzegovina is uncertain even in the next five years) was considered during the analysis.

Despite the differences in GDP per capita (for 2009: B&H $6,300 and Croatia $17,600) (CIA, 2010) it can be concluded that there was no significant difference in the way that companies are organized in those two countries, while differences were obvious between companies that compete in different price segments in the same country. The differences are even more obvious when it comes to attitudes towards the importance of the atmosphere at the point of sale.

The first part of the conversation with respondents concerned the companies' presentation, market share, and explanation of how the sales function is organized within the company. No respondents, except for HR1 and BiH1, could confirm that the company has a document called a "corporate business strategy", and therefore "retail strategy" was nonexistent as well. This document is usually replaced by a plan of sale, which is expressed quantitatively and financially. The realization of the plan is monitored monthly, quarterly and annually, and the plan is revised in line with market trends, as well as in line with global trends.

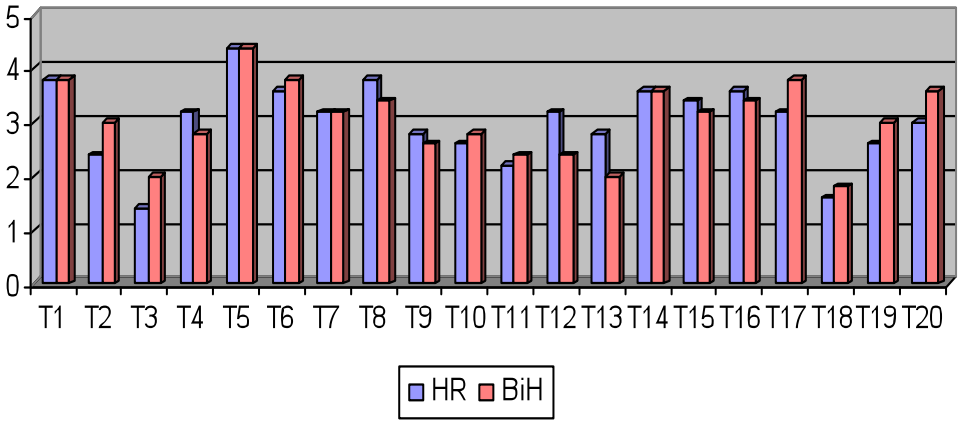
The next group of questions referred to the relationship between marketing and sales departments. Respondents showed a dose of fear in giving answers. Respondents defined cooperation as "correct" while some isolated responses clearly indicated that such qualification is far from the truth:

- The flow of information is not smooth and bilateral, and it is often reduced to the submission of the required information on one or the other side, where there is usually no feedback from the receiving side;
- Marketing departments do not fully understand needs of the target segment;

- There are reports (weekly, monthly) that the sales department delivers to the marketing department (number of visitors, the number of calls, where potential buyers find information about the current promotion, etc.), but sales departments do not feel that the information collected is used properly.

All respondents completed a questionnaire that examines the relationship between the sales and marketing departments in the company (Annex 1). Results show that in six companies the relationship between sales and marketing achieved rank of "defined" (HR2, HR3, HR5, BH2, BH3, BH4). On the other hand, in four companies it seems that this relationship tends to be "aligned" (HR1, HR4, and BH1 BH5). It is interesting that there are no major differences between the arithmetic mean of respondents' regarding the country (Figure 3).

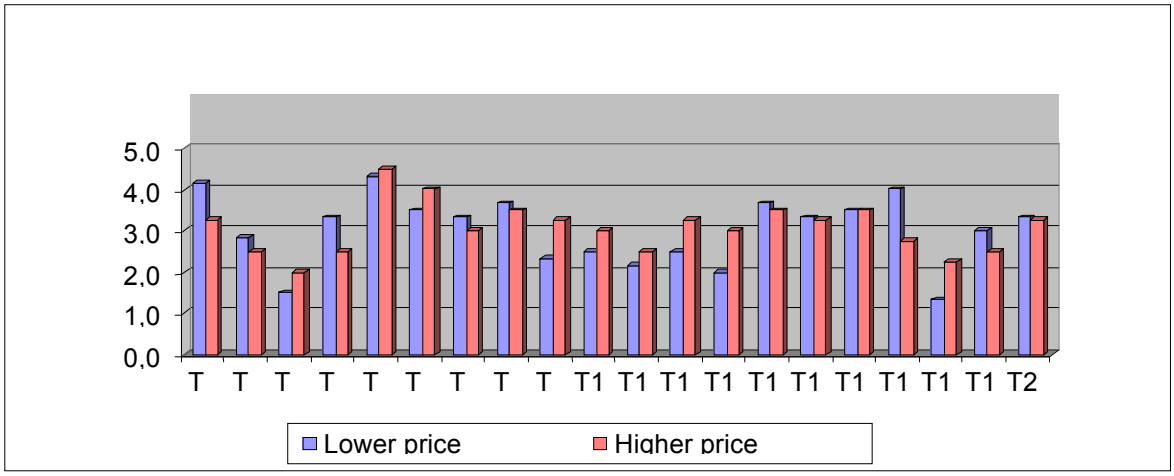
Figure 3: Arithmetic mean of the sales managers' responses by country



A fully balanced, relatively high average value supported the following claims: that the sales figures are usually at the level of the plan (T1 = 3.8), that the sales force believes that assistance from the marketing department is an important tool for achieving better sales results (T5 = 4, 4), that the marketing and sales departments can tell that they "speak the same language" (T7 = 3.2), and that sales and marketing manage their activities using a process developed in common, as well as procedures to increase the range of business, from initial identification to serving users (T14 = 3.6). However, the greatest discrepancy between the responses of sales managers in B&H compared to Croatia was observed in claim 12: that sales and marketing discuss and use the same measures for evaluating performance results, where HR managers graded the claim 3.2 while B&H managers graded it 2.4.

Significant difference in the respondents' answers was observed by comparing answers coming from managers whose companies sell cars in different price segments in both countries (Figure 4).

Figure 4: Answers sorted by price segments companies compete in



The answer values are, in most cases, higher among managers in a “added value” segment. The biggest difference is in claim 13, that the marketing department is actively involved in defining and executing sales strategies for individual sales managers, where sales managers in a higher price segment graded it 3.0, while managers in a lower price segment graded it 2.0 on average.

The last part of the interview referred to the atmosphere of the place of purchase. For all respondents the term "atmosphere at the sale point", evokes the atmosphere created by its sales staff. After the explanations, all respondents confirmed the presence of interior standards and exterior decoration of the dealership defined by the principal. The minimum standards are that the company must ensure the upgrading and improving of the interior is left to each partner. This was an obvious division between brands belonging to higher and lower price segments. Respondents dealing with a lower price class completely ignored environmental factors, emphasizing the atmosphere created by sales staff as crucial. The theory of fifth-dimension brand management was viewed as irrelevant to sales efficiency, while they agreed that investment in training sales staff, and designing an atmosphere for the sale point may result in increased customer satisfaction. On the other hand respondents HR4, BH4, HR5 and BH5, gave great importance to fifth-dimension regulation of the sale point and gave an example in which the right atmosphere at the point of sale (environmental factors) led to the positive completion of the sale process. Special importance was placed on creating personal relationships with customers, as well as on organizing special events for different customer segments. However, it is necessary to mention that, although they emphasized the importance of music at the sales point, only one of them had music on (through local radio stations).

The respondents whose companies engaged in retailing of higher price class brands pointed out that it was possible to increase sales in the last five years (especially in 2007 and 2008) by virtue of research into the atmosphere at the sale point, as well as the training of sales staff in the field of sales tactics. On the other hand, respondents that manage sales of brands that are lower priced believed that sales growth may be primarily due to price, quality of product and the service of sales staff.

7. TESTING AND EXAMINATION OF HYPOTHESES

None of the respondents explicitly confirmed the existence of a document called "retail strategy", which was substituted with occasional sales and financial plans. Therefore, if a retail strategy does not exist, then it is impossible to speak of the specifics of this strategy in terms of combining sales tactics and the atmosphere of the place of purchase. It follows that H1 is confirmed.

The second hypothesis concerns the relationship between marketing and sales departments in retail automotive companies in Bosnia and Herzegovina and Croatia. The

conclusions were very contradictory. Results obtained from the questionnaire point toward conclusion that the relationships between sales and marketing range from “defined” to “coordinated.” At the same time, during the interviews respondents cited a variety of communication problems with their marketing departments which were manifested through the following: obtaining the right information at the right time, obtaining a response to reports submitted, and not knowing the characteristics of the target segment. All of the difficulties mentioned should not be present in organizations whose marketing and sales departments have a defined relationship that tends to be “aligned.” Knowing that respondents were slightly calculated when answering the questionnaire, the authors believe that it is necessary that all responses from the questionnaires should be corrected i.e. downgraded. Such a move would support the second hypothesis. However, it is impossible to claim that H2 is fully confirmed.

Some respondents (HR3, and BiH3, BiH5) were aware of the influence of atmosphere on customers, and they tried to design it in a way that customers are provided with positive sensory and emotional experience. Other respondents considered it completely irrelevant or less important in their particular industry. Respondents agreed that investment in sales staff training and designing sales point atmosphere probably would result in customer satisfaction, while opinions on the growth of sales caused by this investment were diametrically opposed, and range from absolute confirmation (companies in a higher price segment) to absolute denial (companies in a lower price segment). It was obvious that no company has a plan to invest in an atmosphere at the sale point, nor to measure the effects of these investments on satisfaction and sales. It can be concluded that H3 is confirmed while H4 is partly confirmed.

8. CONCLUSION

In a dynamic environment, competition increases and market conditions make knowledge one of the most valuable resources for a company. It is obvious that companies in the automotive retail sector are facing the worst crisis in its history. However, this study confirmed the hypothesis that in such circumstances, enterprises in both Bosnia and Herzegovina and Croatia completely lack retail strategies. Research has shown that the information flow between marketing and sales is jammed. This applies to the fact that the sales managers were afraid to negatively evaluate the work of marketing departments in the survey. They shared a fear that colleagues from marketing department will see their survey. All those lead to the termination that relationship between marketing and sales is far from "integrated", as it should be. The relationship tends to be in the middle from “defined” to “aligned”.

Through the research the existence of two types of marketing strategies has been identified. One is based on price and the benefits that the buyer gets by the purchase of certain vehicles, which is irrelevant to other variables such as the sales point atmosphere. Nevertheless, such companies are working on building brands through the transfer of benefits that the customer achieves by cost savings in other spheres, and this strategy has proven effective. In contrast are companies that based their marketing strategy on balanced investment in other variables in order to create and improve brand image in the minds of consumers and charge a premium for the brand offered.

The scientific validity of this research is associated with a significant growth trend regarding importance of cross functional cooperation in the retail companies. The paper can serve to retail businesses in assessing the practical application of theoretical statements, allowing managers to think creatively and enhance their and knowledge of their employees. Research findings can be used for future work on a similar topic as a basis for qualitative research because of the lack of literature and data related to this area, particularly in Bosnia and Herzegovina and Croatia. In order to obtain full analyze, larger number of respondents from both countries is required. This would enable quantitative test of the hypotheses for both

countries and a more detailed analysis of the differences in set hypotheses between the two countries.

Having in mind all current difficulties for retailers in the automotive sector the best companies will remain those that demonstrate flexibility in creating the atmosphere of the sales point as result of sharing knowledge between marketing and sales department

LITERATURE

Analiza tržišnog učešća na tržištu novih automobila u BiH 2008 – 2009. (2010), Agencija Plus, Sarajevo

Analiza tržišnih udjela na tržištu novih automobila u Hrvatskoj 2006 – 2009. (2010), Promocija Puls, Zagreb

Anderson, R.E. (1996) Personal selling and sales management in the new millennium. *Journal of Personal Selling and Sales Management*, Vol. 16, Fall, pp 1-16.

Broekemier, G., Marquardt R., Gentry, J.W. (2008) An exploration of happy/sad and liked/disliked music effects on shopping intentions in a women's clothing store service setting, *Journal of Services Marketing,* Volume: 22; Issue: 1

CIA, (2010) The World Fact book [online] Dostupno na: https://www.cia.gov/library/publications/the-world-factbook/[01. svibnja 2010.]

Day G.S. (1994) The capabilities of market-driven organizations, *Journal of Marketing*, 58 (4) pp. 37–52.

Dewsnap, B., Jobber, D. (2007) Testing a model of marketing-sales relations. *Academy of Marketing Conference*. July, Kingston, UK

Donovan, R.J., Rossiter, J.R. (1982) Store Atmosphere: An Enviromental Psychology Approach; Journal of Retailing; Vol. 58 pp. 34-57.

Flick, U. (2009), *An Introduction to Qualitative Research*, SAGE Publications Ltd; Fourth edition

Gobe, Mark (2006) Emocionalno Brendiranje; *Mass Media International Beograd*, ISBN: 86-906723-3-8

Gosselin, D., Heene, A. (2003) A competence-based analysis of account management: Implications for a customer-focused organization. *Journal of Selling and Major Account Management*, Vol. 5 No. 1, pp 11-31.

Homburg C., Pflesser C. (2000) A multiple-layer model of market oriented organizational culture: measurement issues and performance outcomes. *Journal of Marketing Research*, 37. pp 449–62.

Kirmani, A., Campbell M.C. (2004) Goal Seeker and Persuasion Sentry: How Consumer Targets Respond to Interpersonal Marketing Persuasion. *Journal of Consumer Research*, 31, pp 9–31.

Kotler, P. (1973) Atmospherics as a marketing tool. *Journal of Retailing*, Vol. 49 No.4, pp 48-61.

Kotler, P., Rackham, N., Krishnaswamy, S. (2006) Ending the war between sales and marketing. *Harvard Business Review*, Vol. 84 No. 7/8, pp 68-78.

Le Meunier-Fitzhugh, K., Piercy, N.F. (2007) Exploring collaboration between sales and marketing. *European Journal of Marketing*, Vol. 41 No 7/8, pp 939-55.

Leigh, T.W., Marshall, G.W. (2001) Research priorities in sales strategy and performance. *Journal of Personal Selling and Sales Management*, Vol. 21 No. 2, pp 83-93.

Lindstrom, M. (2005) *Brand Sense*. Free Press. New York.

Manning, G.L., Reece, B.L., (2008) *Suvremena prodaja: stvaranje vrijednosti za kupca.* Mate. Zagreb.

Mallalieu L., Nakamoto K. (2008) Understanding the Role of Consumer Motivation and Salesperson Behavior in Inducing Positive Cognitive and Emotional Responses During a Sales Encounter. *Journal of Marketing Theory and Practice*, vol. 16, no. 3, pp 183–197.

McDonald, M., Rogers, B., Woodburn, D. (2000) *Key Customers: How to Manage them Profitably*. Butterworth-Heinemann. Oxford.
Mehrabian, A., Russell, J.A. (1974) *An Approach to Environmental Psychology*. MIT Press. Cambridge.
Piercy, N.F. (2006) The strategic sales organization. *The Marketing Review*, Vol. 6, pp 3-28.
Paparoidamis N.G., Guenzi P. (2009) An empirical investigation into the impact of relationship selling and LMX on salespeople's behaviours and sales effectiveness. *European Journal of Marketing,* Vol. 43 No. 7/8, pp 1053-1075
Ruekert, R.W., Walker, O.C. Jr (1987) Marketing's interaction with other functional units: a conceptual framework and empirical evidence. *Journal of Marketing,* Vol. 51 No. 1, pp 1-19.
Santini M. (2010) Stižu crni dani za europsko tržište automobila [online]. Jutarnji list.Dostupno na: http://www.jutarnji.hr/stizu-crni-dani-za-europsko-trziste-automobila/730141/[01. svibnja 2010.]
Slater S.F., Narver J.C. (2000) Intelligence generation and superior customer value. *J Acad Mark Sci*, 28(1), pp 120–127.
Suvremena.hr (2009) Hrvatska: Pad prodaje automobila za 50 posto [online] Suvremena.hr Portal za trgovinu. Dostupno na: http://www.suvremena.hr/13054.aspx[01. svibnja 2010.]
TomaševićLišanin, M. (2010) *Profesionalna prodaja i pregovaranje*. HUPUP, Zagreb.
Trivedi, B. (2006) Recruiting smell for the hard sell [online] NewScientist magazine. Dostupno na: http://www.newscientist.com/article/mg19225821.800-recruiting-smell-for-the-hard-sell.html?full=true; [03. March 2010.]

Annex 1: Arithmetic mean of the questionnaires' answers

#	**Claim**	HR 1	HR 2	HR 3	HR 4	HR 5	BH 1	BH 2	BH 3	BH 4	BH 5	Avr age
1	Our sales figures are usually close to the sales forecast.	5	3	5	2	4	4	4	4	3	4	**3,8**
2	If things go wrong, or results are disappointing, neither function points fingers or blames the other.	1	2	3	4	2	5	3	3	2	2	**2,7**

3	Marketing people often meet with key customers during the sales process.	2	1	1	2	1	3	1	1	2	3	**1,7**
4	Marketing solicits participation from Sales in drafting the marketing plan.	5	2	3	4	2	5	2	3	2	2	**3**
5	Our salespeople believe the collateral supplied by Marketing is a valuable tool to help them get more sales.	4	4	5	5	4	5	3	5	4	5	**4,4**
6	The sales force willingly cooperates in supplying feedback requested by Marketing.	3	4	3	4	4	5	3	3	5	3	**3,7**
7	There is a great deal of common language here between Sales and Marketing.	5	3	2	3	3	5	3	2	3	3	**3,2**
8	The heads of Sales and Marketing regularly confer about upstream issues such as idea generation, market sensing, and product development strategy.	5	4	3	4	3	5	2	3	3	4	**3,6**
9	Sales and Marketing work closely together to define segment buying behavior.	2	3	2	4	3	3	2	2	3	3	**2,7**
10	When Sales and Marketing meet, they do not need to spend much time on dispute resolution and crisis management.	3	2	2	4	2	3	3	2	2	4	**2,7**
11	The heads of Sales and Marketing work together on business planning for products and services that will not be launched for two or more years.	4	1	1	4	1	4	2	1	1	4	**2,3**
12	We discuss and use common metrics for determining the success of Sales and Marketing.	5	3	2	4	2	1	2	2	3	4	**2,8**
13	Marketing actively participates in defining and executing the sales strategy for individual key accounts.	2	3	2	4	3	1	2	2	3	2	**2,4**
14	Sales and Marketing manage their activities using jointly developed business funnels, processes, or pipelines that span the business chain – from initial market sensing to customer service.	5	3	3	3	4	5	3	3	3	4	**3,6**
15	Marketing makes a significant contribution to analyzing data from the sales funnel and using those data to improve the predictability and effectiveness of the funnel.	4	3	3	4	3	4	3	3	4	2	**3,3**
16	Sales and Marketing share a strong “We rise or fall together” culture.	5	3	3	4	3	5	2	3	3	4	**3,5**
17	Sales and Marketing report to a single chief customer officer, chief revenue officer, or equivalent C-level executive.	5	2	4	3	2	5	4	4	4	2	**3,5**
18	There’s significant interchange of people between Sales and Marketing.	1	2	1	2	2	1	2	1	2	3	**1,7**
19	Sales and Marketing jointly develop and deploy training programs, events, and learning opportunities for their	2	2	3	4	2	5	3	3	2	2	**2,8**

	respective staffs.											
20	Sales and Marketing actively participate in the preparation and presentation of each other's plans to top executives.	3	3	3	4	2	5	3	3	3	4	**3,3**
	SUMA	**71**	**53**	**54**	**72**	**52**	**79**	**52**	**53**	**57**	**64**	

UNAPREĐENJE UČINKOVITOSTI OSOBNE PRODAJE U SEKTORU MALOPRODAJE AUTOMOBILA

SAŽETAK

Prodaja postoji od početka robne razmjene, ali su se principi s vremenom promijenili. Suvremena prodaja u fokus stavlja povećanje produktivnosti kupca kroz strateško prodajnu organizaciju koja konvergira s marketingom kako bi dobila veći strateški značaj u organizaciji. Kvalitativnim istraživanjem, metodom dubinskog intervjua, cilj rada je utvrditi odnos između prodajnog i marketing odjela kao i saznati više o percepciji prodajnih menadžera o definiranju strategije prodaje i oblikovanja atmosfere prodajnog mjesta kao jedne od varijabli koja može utjecati na efikasnost prodaje, na slučaju maloprodaje automobila. Komparirani su rezultati istraživanja u Bosni i Hercegovini i Republici Hrvatskoj, te oni općenito ukazuju na nedovoljnu povezanost između marketinga i prodaje, te na nedostatno korištenje atmosferskih čimbenika pri prodaji na oba promatrana tržišta.

Ključne riječi: *osobna prodaja, marketing, atmosfera na prodajnom mjestu, kvalitativno istraživanje*

JEL klasifikacija: *M31, M53*

GIANFRANCO CAPODAGLIO,
MATTEO SANTI,
IVANOE TOZZI [1]

UDK 657:006(100)>:334.722>(450)

THE INTRODUCTION OF INTERNATIONAL ACCOUNTING STANDARDS IN THE ITALIAN SMALL AND MEDIUM SIZED ENTITIES

ABSTRACT

Considering the fact that the Italian government is ready to review both tax regulations and civil regulations governing financial statements, the paper try to summarize the considerations of Italian accounting science on introduction of International Accounting Standards in the national system. Accounting changes seen in recent years are significantly impacting the financial statements of the European companies: on one hand, listed companies are adopting International Accounting Standards, and on the other hand, SMEs are facing new regulations that are about to be reviewed by the Italian Government.

Keywords*: IFRS, SMEs, Micro-Entities, EU Directives*

1. THE GRADUAL INTRODUCTION OF INTERNATIONAL ACCOUNTING STANDARDS IN THE ITALIAN SYSTEM: FROM NATIONAL PRINCIPLES TO IAS/IFRS.

It appears useful to summarize the stages through which the IAS/IFRS principles took on a significant role in the Italian legal system, also considering that the Italian legal system assigned (and still assigns today) a role of certain importance to national accounting principles within the framework of regulations traditionally concerning civil law.

Within the Italian legal system, the national accounting principles have the function of support and interpretation of the regulations on the subject of financial statements; implicit reference is made to the national accounting principles in the general clause of Article 2423 of the Italian Civil Code when it introduces the so-called "integration obligation" and "exception obligation". In addition, the concept of "technical discretion" contained in the report accompanying Decree no. 127 of 1991 refers to the national accounting principles.

The International Accounting Standards were only marginally important before Law no. 306/2003 went into effect on 31 October 2003; up to that date, the international principles could only be used in the event of an issue not be addressed by the national principles.

However, as it is known, the firms that dominate the international scenario have imposed their practices and traditions on the world's most important financial markets, and if the firms belonging to different traditions and different economic, political and cultural systems intend to compete, or even just to survive in such markets, they must align themselves to the methods and procedures dictated by the dominant parties.

The European Union's initiative originated from this. After having found a minimum common denominator in the EU-based companies' preparation of financial statements during the 1990s, the EU subsequently issued "directives" through which it stated the need to oblige the Member States to adopt regulations that were initially compatible in substance with those prevailing on the main international markets, and later, increasingly coincident with the

[1] Gianfranco Capodaglio is Full Professor of accounting, Department of management, University of Bologna, Matteo Santi and Ivanoe Tozzi are Associate Professors of accounting, Department of management, University of Bologna, Via Capo di Lucca, 34 - 40126 Bologna (Italy) e-mail: m.santi@unibo.it Paragraphs 1 and 7 is to be attributed to Gianfranco Capodaglio, paragraphs 2 and 3 is to be attributed to Ivanoe Tozzi, paragraphs 4, 5, 6 is to be attributed to Matteo Santi.

regulations on the international markets.

The referenced regulations were formulated and developed in the Anglo-American countries, whose cultural and legal traditions are very different from those of most countries in Continental Europe.

With a series of regulations[1], the European Union obligated the Member States to adopt the IAS/IFRS[2] as of 1 January 2005, for the preparation of the consolidated financial statements of companies whose securities were traded on regulated markets.

With Law n. 306/2003 and Decree-Law n. 38/2005, Italy agreed to endorse the accounting harmonization promoted by the European Union, moving beyond the obligations imposed by the aforementioned series of regulations, and provided that IAS/IFRS were also to be adopted by:

- publicly traded companies in the preparation of their financial statements (non-consolidated);
- companies issuing financial instruments sold to the public, in the preparation of their consolidated and non-consolidated financial statements;
- banks and financial intermediaries subject to the oversight of the Bank of Italy, in the preparation of their consolidated and non-consolidated financial statements;
- insurance companies in the preparation of their consolidated financial statements, and if publicly traded, in the preparation of their non-consolidated financial statements in the event of their not preparing consolidated statements.

The Italian regulations also provide the option of preparing financial statements according to International Accounting Standards for all other firms that are not authorized to elect an abbreviated form for their financial statements[3].

2. THE GRADUAL INTRODUCTION OF INTERNATIONAL ACCOUNTING STANDARDS IN THE ITALIAN SYSTEM: THE ROLE AND THE PURPOSES OF IAS/IFRS.

The expression "IAS/IFRS" is normally translated in Italian as "international accounting principles", however, the use of these terms can be misleading.

An *"accounting standard*" (or *"financial reporting standard"*) is not actually an "accounting principle" as defined by accounting doctrine and Italian practices, but is a simple "empirical rule" susceptible to ongoing changes, depending on how the prevailing practice perceives the changes of the economic situations in which the rule needs to be adopted.

In particular, the Standards are not associated with any general system of reference; they instead exist with respect to the "Framework" which applies only if it is not conflicting with the content of the individual rules[4]. It is interesting to note that the Framework was not ratified by the European Union, even though frequently cited by the individual standards.

If we consider the role of the IAS/IFRS in the countries where they originated and were developed, we can see that their significance is totally different from that of the Italian accounting principles. As it is known, in countries where common law is in effect, the legal system is based not on legal codes, but rather on laws developed through the precedent of jurisprudential decisions. Such decisions are based on "best practices", which, in the case of financial statements, are represented by "generally accepted accounting principles".

The IASs/IFRSs clearly have a very significant role, considerably limiting the actions of anyone involved in the subject, including, obviously, decision-making authorities of every order and degree.

The significant innovation for the Italian system is not limited to this alone: as stated, the Standards are continuously changing and are issued by a private entity which thus substitutes the national legislature.

A reading of the Framework (which, as indicated, does not prevail over the Standards)

is nonetheless interesting because it indicates, among other things, the purposes of financial statements according to the IAS/IFRS (Onida, 1951)[5]: the financial statements are prepared to meet the information needs of many users, but, among them, the category of "investors" is the most prevalent; investors are described as[6]: "persons who supply risk capital and their consultants who are interested in the risk inherent to their investment and the related return. They need information that helps them to decide if to buy, maintain or sell. Shareholders, moreover, are interested in making use of the information that puts them in a position of evaluating the entity's capacity to pay dividends."

It is appropriate to note that the reference to "financial statements" is normally to consolidated financial statements in the environment in which IAS originated, with the financial statements for the individual company considered an almost "internal" document, and thus, one of less importance. The situation in Italy is obviously very different: the concept of "financial statements" is unanimously related to the reference reporting for the individual company, except for particular cases specifically identified in regulations, doctrine and practices. The question of the distinctiveness of the financial statements for all types of firms, regardless of their size, has been debated in doctrine in past decades, with the discussion leading to the almost unanimous conclusion of the existence of a distinct ordinary financial statement, i.e. a unique document whose information-reporting purpose is represented by the "earnings result for the period and by the related working capital." It is obvious and generally acknowledged that, if documents that have other information-reporting purposes are called "financial statements", then it is inevitable to acknowledge their existence, without however being able to label them as "ordinary", even though they may be referable to individual periods.

In the Framework, on the contrary, a decisive statement is made regarding the intention to sacrifice the needs of all other users, if they are not compatible with the needs of investors: "since the investors supply risk capital to the entity, financial statements that satisfy their needs for information will also satisfy more of the needs of the other users of the financial statements."

It is almost superfluous to note that the interests of the shareholders are not all equal, and that those of the majority investors of the present, and more importantly, the majority investors of the future, may be very different from those of the minority shareholders, creditors, employees, and especially - in our case - the state (in its role as the collector of taxes).

The Standards are accordingly aimed at mainly representing information useful for shareholders (present and future), rather than information useful to enterprise: in other words, financial statements must offer the most information possible about the probability of a shareholder to collect the sum invested over a foreseeable time period, namely, the result of the investment, through the distribution of dividends and the subsequent resale of the securities acquired. For this purpose, from the shareholder's perspective, the receipt from the company of income available, and likely set aside, for distribution or an increase in the price of the securities held has the same value.

The consequences of this choice on establishing principles for the preparation of financial statements are particularly important[7]; there are only two principles: the matching principle and the going-concern principle. The "qualitative characteristics" of the financial statements are a follow-up to such principles, and are defined as "aspects that make the information contained in the financial statements useful for the users"; such aspects include intelligibility, significance, reliability and comparability.

The requisites for reliability include not only neutrality, completeness and the prevalence of substance over form, but also prudence, which, as is known, constitutes the first of the principles currently in effect for preparing financial statements, as dictated by Article 2423 bis of the Italian Civil Code.

The application of this orientation is seen in the individual Standards: it is interesting to note, for example, the definition of "revenue" as provided by IAS 18, according to the

Italian text ratified by Italian lawmakers: «*i ricavi sono flussi lordi di benefici economici conseguenti l'esercizio derivanti dallo svolgimento dell'attività ordinaria dell'impresa, quando tali flussi determinano incrementi del patrimonio netto diversi dagli incrementi derivanti dagli apporti degli azionisti*» *["revenue is the gross inflow of economic benefits achieved during the year arising from the ordinary operating activities of the enterprise, when such inflow determines increases in shareholders' equity other than increases arising from shareholders' contributions."]*

Even more suggestive is the passage that indicates the conditions that make it possible to book the revenue: the passage includes a provision according to which revenue may be booked when «*è probabile che i benefici economici derivanti dall'operazione saranno fruiti dall'impresa e i costi sostenuti, o da sostenere, riguardo all'operazione possono essere attendibilmente determinati*» *["it is probable that the economic benefits arising from the transaction will be enjoyed by the enterprise, and the costs sustained, or to be sustained, regarding the transaction may be reliably determined"]*.

A provision contained in Paragraph 29 of IAS 16 is completely in line with the objectives indicated, and makes it possible to value plant, property and equipment at fair value, even if above book value. Fair value is defined (Paragraph 7 of IAS 18) as "the amount for which an asset could be exchanged, or a liability settled, between knowledgeable, willing parties in an arm's length transaction."

3. THE GRADUAL INTRODUCTION OF INTERNATIONAL ACCOUNTING STANDARDS IN THE ITALIAN SYSTEM: THE PROCEDURE FOR ADOPTION IN THE EUROPEAN UNION

The adoption of the International Accounting Standards in the Member States of the EU is subject to a ratification process (a process which has occurred with the regulations indicated above). Thus, unlike the situation in the countries where the principles originated, the adoption of the International Accounting Standards in Europe does not automatically occur as the principles are issued by the IASB, but rather is subject to examination and specific approval.

As soon as they are adopted at a European level, however, the IAS/IFRS become fully valid in the Member States, including in the absence of any special ratification.

In Italy's case, the Decree-Law no. 38/2005 adopted these principles, extending the sphere of application thereof; we need to emphasize, however, that Article 5, Paragraph 1 of such decree introduces a regulation that would seem to limit the indiscriminate use of the international principles, inasmuch as it specifies that "if, in exceptional cases, the application of a provision provided by the IAS is incompatible with the true and correct representation of the earnings, financial position and capital, the provision is not applied. In the financial statements any earnings arising from the exception are booked to a reserve whose distribution is restricted, unless in an amount corresponding to the value recovered."

The sphere of application of such provision is not clear: indeed, the phrase "true and correct representation" seems to refer to the content of Article 2423 of the Italian Civil Code, as interpreted up to now in the Italian legal system; such interpretation identifies the requisite of "correctness" in the application of the principles for preparation provided by Article 2423-bis[8] and ranks at the top of a body of regulations that are an alternative to the international Standards.

At this point, the Member States, such as Italy, which have extended the regulations regarding the preparation of consolidated financial statements to the non-consolidated financial statements for certain categories of larger businesses, have two distinct sets of regulations in relation to the same matter (the non-consolidated statutory financial statements) which are in effect to two different categories of businesses, which are principally distinguished from one another by their size.

4. THE PROBABLE EXTENSION OF THE IAS/IFRS LOGIC TO FINANCIAL STATEMENTS FOR MOST COMPANIES: THE ITALIAN ACCOUNTING ENTITY'S PROPOSAL FOR THE IMPLEMENTATION OF EU DIRECTIVES 2001/65 AND 2003/51 WITH AMENDMENTS TO THE ITALIAN CIVIL CODE.

On 25 October 2006, the executive committee of Organismo Italiano di Contabilità (the Italian Accounting Entity or O.I.C.) approved a proposal that is aimed at modifying the current regulations in the Italian Civil Code, and in our opinion, this is considered positive because it excludes the simple extension of the obligation of adopting the international principles to the entire array of Italian companies. We feel it is worth noting however what was observed in relation to the different role played by accounting principles in Italy, compared with what occurs in countries with common law systems: the inclusion of the principles in the provisions of the Civil Code in Italy would substantially change their nature and could have material adverse consequences on the equilibrium of the entire system of company regulation.

A new article contemplated is particularly important in that it would take the place of the current Article 2423-bis about the framework of application of the new regulations, excluding only the companies that are obligated by law to adopt the International Accounting Standards, with the consequence that all other firms would be prohibited from adopting such principles.

The changes can be summed up in several key points:

- The principle of the prevalence of substance over form is very clearly stated, compared with a current situation in which it is expressed in an ambiguous manner and inconsistently applied to different types of cases (as in the case of leasing transactions).
- Thought continuing to figure as a factor in the preparation of financial statements, the principle of prudence loses its place as the guiding principle, both because it is no longer ranked in first position, and more importantly, due to the possibility provided of an exception to the historical cost criterion; however, this does not eliminate the enormous gap existing with the meaning assigned to the term by the IAS/IFRS, inasmuch as the O.I.C. fully preserves the principle of “asymmetry” of the Italian tradition; indeed, for the positive components of income, there is still the need for “reasonable certainty”, in place of the “probability” incorporated into the international principles; instead, probability remains a sufficient condition for the accrual of risk-related charges (the asymmetry between the negative and positive components of income can also be found in Article 2423-ter no. 5 of the Italian Civil Code which proposes anew the currently prevailing regulation).
- In relation to the preceding point, the proposal provides for the option of using fair value as an alternative to historical cost for the valuation of certain elements, with historical cost no longer being allowed for financial derivatives instruments; similarly, the possibility of valuing commissioned work in process at cost is excluded, except when the amount is accrued with reasonable certainty.
- The concept of the “amortized cost” of receivables and payables (2426 bis, Paragraph 2) and the assumptions for discounting are introduced.
- The concept of “financial-statement continuity” (the provision in Article 7 of Decree-Law 87/92, which ratifies Article 31, letter f) of the EC Directive IV, according to which “the opening balance sheet for a period must correspond with the closing balance sheet of the prior period”) has been dismissed. The proposal provides that the effects of the change in the valuation criteria must be directly booked to shareholders’ equity, without flowing through the profit and loss statement.

* The content of the financial statements is enriched by the cash-flow statement and the statement of changes in shareholders' equity. The latter is necessary because of the change outlined in the preceding point. With financial-statement continuity no longer required, the relationship between the balance sheet and income statement would be impossible to understand without an additional statement.

As previously indicated, in our opinion, the O.I.C. proposal needs to be favourably embraced in general as a compromise between opposing needs, particularly the sections that continue to place the priority on the production of information useful for the traditional user of the financial statements. This does not mean overlooking the critical elements associated with the possible acceptance of the proposal: one of the most significant critical elements in our opinion is the elimination of the requirement for financial-statement continuity, in order to adjust to the new version of IAS 8 which establishes the retroactive application of changes in valuation criteria, with the consequent adjustment of the opening balance of the corresponding account of shareholders' equity. This practice is subject to criticism on two fronts: first, from the standpoint of the law, and second, from a business economics perspective. From a legal viewpoint, the establishment and use of unrestricted reserves needs to be approved by the shareholders under Italian law; therefore, any changes to the opening balance of shareholders' equity would be a clear-cut violation of such law. It would be possible to get around this limitation by submitting a proposal to the shareholders (along with the proposal for approval of the financial statements) asking for approval of the use or the increase of reserves. This response, however, is not convincing: the shareholders have absolute power over deciding the allocation, meaning they could reject such a proposal, with the consequence that the planned adjustment of the opening balances (which would be moreover obligatory) could not be adopted in such case.

Equally important (and perhaps even more important) is the business economics argument. As known (Ferrero, 1995, 179)[9], the changes in net capital can be "direct" or "indirect"; while the former originate from outside of the company (increases and reductions of share capital, share issuance premiums, and reimbursements), the latter are associated with earnings, income that flows continually (Ferrero, 1995, 14)[10] during the life of the firm, despite the merely conventional need for assigning part of it to any short period. The fundamental principle of final reconcilement is based on the considerations set out above, and according to such principle, the algebraic sum of the periodic income reported must coincide with total or overall income, i.e. the amount referable to the entire life of the firm (Campanini, Capodaglio, 1988, 92)[11].

Adopting the proposal referenced herein, the algebraic sum of the results of different periods would no longer correspond to the indirect change in net equity, thereby undermining the definition of income itself. Of no use in mitigating the criticism is the argument that there would be a dual violation of the matching principle by booking the retroactive effect of the change in valuation criteria to the period in which the change gets adopted. First, the effect of the change does not necessarily have to be retroactive, with the effects being able to be valid *ex nunc*; second, the allocation to one period of income components related to prior periods is rather common and gives rise to extraordinary components of income, which are used for reporting these types of phenomena. A violation of the matching principle could instead be represented in the tendency to eliminate the characteristic distinction between ordinary and extraordinary components of income, provided by the recent versions of IAS/IFRS.

The explicit affirmation of the principle of the prevalence of substance over form surely constitutes an aspect of greater clarity vis-à-vis the current situation, but it does not wipe away the problems inherent to the limits of such principle, in a legal system such as that in Italy which also bases its equilibrium on the rigorous respect of formal aspects. Article 2423-ter n 2 of the Italian Civil Code actually proposes the principle, "barring other provision of the law." The assumption of implementation does not add anything else: thus, the doubt remains as to whether it needs to be interpreted in the sense that the principle is always

applied, unless there is a regulation that exists that explicitly provides that a fact must be indicated in the financial statements according to its formal aspect, even if contrary to the substance of the transactions, or whether the clause makes reference to all of those transactions for which the form is essentially important from a legal standpoint, such as the obligation for the written form in certain contracts, or the presence of reasons for the nullity or voidability of a deed, etc.

5. THE PROBABLE EXTENSION OF THE IAS/IFRS LOGIC TO FINANCIAL STATEMENTS FOR MOST COMPANIES: THE "IFRS FOR SMALL AND MEDIUM-SIZED ENTITIES".

In February 2007, the IASB handed down an exposure draft for the adoption of new accounting principles to be used by smaller sized entities; compared with the initial orientation that was aimed at mere simplification or partial application of the IFRS, the prevailing opinion is that is it essential to come up with a set of specific Standards (stand-alone document) for small- and medium-sized entities (SMEs)[12].

The document does not supply a quantitative definition of a SME, but in the presentation notes and request for comments, it acknowledges that the IASB has made reference to businesses with about 50 employees in the development of the Standards. The qualitative definition elected in paragraph 1.1 identifies a small/medium-sized business as one which does not have "public accountability", meaning the presentation of financial statements to entities controlling the regulated financial markets. In addition, the SME can be distinguished in terms of the recipients of financial-statement information: minority shareholders and creditors. We need to observe that in Italy the 99% of total number of firms has less than 50 employees, according to Italian Statistical Office (ISTAT).

At this point, one would expect a "distancing" from the IFRS, which, as already indicated, identify investors (especially future investors) as the recipients of financial-statement information; instead, as we shall better specify hereunder, this does not happen.

It is interesting to note in this regard that the Framework has been substituted by *pervasive principles*; though being presented as an alternative to the Framework, such principles have preserved most of the basic characteristics thereof.

The financial-statement objectives that have been established include information regarding financial position, performance, and the cash flows of the business, all of which is useful for the economic decisions of anyone who is not in a condition to be able to request specific reports aimed at satisfying their needs for information. There is no specification of what is meant by "information useful for economic decisions", but from the development of the Standards, it is inferred that this information does not correspond to the information useful for the protection of third-party creditors and useful for prudence in the estimation of distributable income, but is very similar to that contemplated by IFRS.

Turning to the pervasive principles, the financial-statement disclosures need to be capable of addressing the economic decisions of users (*relevance*), and they must be quantitatively significant (*materiality*) in that sense that, if omitted or erroneous, they can negatively influence the economic decisions. The principles of reliability, the prevalence of substance over form and prudence then follow, with the last of them understood as a certain degree of caution in judgments inherent to valuation; this concept is very different from that of the "disparity of treatment" set out in the Italian Civil Code. The other principles regard completeness, comparability, timeliness of the information, and finally, the "cost-benefit" comparison in the preparation of the information. This last point is perhaps the only that effectively differs from the IFRS Framework, it having been expressly introduced in order to "simplify" the adoption of the Standards by small businesses and to reduce the costs of the necessary administrative compliance.

Particularly significant is the chapter inherent to the conditions for the recognition of the elements of financial statements: with regard to the assets, the concept in the IFRS is

repeated, whereby the condition sufficient for booking assets to the balance sheet is that it is probable that the benefits connected with the assets will be acquired by the business[13] and that their value is measurable in a reliable manner. It is also specified that positive components of income related to the assets may be booked to the financial statements on the basis of the same criteria.

As far as the valuation criteria are concerned, the basic orientation is also that of the IFRS, with fair value being an option for certain items, and mandatory for others.

All in all, the Standards proposed for the SMEs adopt almost entirely the same principles as the current IFRS, but they offer a simplified and reduced version thereof contained in a volume of just under 200 pages.

6. THE PROBABLE EXTENSION OF THE IAS/IFRS LOGIC TO FINANCIAL STATEMENTS FOR MOST COMPANIES: THE CONSIDERATIONS OF THE O.I.C. ON THE "QUESTIONNAIRE ABOUT THE ACCOUNTING PRINCIPLES AND THE ACCOUNTING DIRECTIVES FOR THE SMES".

In responding to a special IASB questionnaire in February 2007, the O.I.C. stated that the so-called "small enterprises" in Italy, when compared with businesses in the most important Member States of the European Union, are more often "micro businesses", whereas the definition of the "average" enterprises in such countries would likewise be identified in Italy as medium/large-sized companies.

Several very important considerations emerge from reading the responses to the questionnaire:

- Though the O.I.C. was not able to supply reliable quantitative data, it has reasonable certainty that small enterprises represent the large majority of businesses operating on the Italian market[14];
- The users of the financial statements of the SMEs are generally shareholders/owners of the companies, lenders and creditors, tax authorities, and employees.

The O.I.C.'s concluding note is fundamental: "in our opinion, the IASB's exposure draft in relation to Accounting Standards for SMEs does not currently appear to be a solution that can be shared for representing, from an accounting standpoint, the problems typical of the small businesses in Italy. The Standard proposed is too "close" to the IAS/IFRS overall; it presents few examples that guide the drafters of the financial statements; and it continues to be focused mainly on the benefit of the investor (in the meaning other than that of a business owner), whereas the group of users is different for the majority of small businesses." Of course we agree with this idea, even in the light of the final draft of IFRS for SMEs.

An other recent proposal of the Commission of the European Union would be the introduction of the concept of "micro-entity", in order to simplify the accounting system (inside the "small business act" and the idea of "think small first"). After defining the micro-entity as firm in which we have less than 10 employees, total revenues less than 1 million euro, and total assets less than 500.000 euro, one of the proposal of this project is to abolish the financial accounting to be kept mandatory; we think that this idea will get worse and worse in terms of knowing the performance of the single entity and of the whole economy, particularly in a moment of strong economic and financial crisis.

7. CONCLUDING OBSERVATIONS

As a conclusion, the recent broad financial crisis make many economic operators wondering the possibility that some assumption previously considered "non changeable" can be reconsidered. We refer to the opinion, very well known in Europe and North America, that

the engine of the development in developed countries is the finance and the increasing or entities' dimension is considered the main success's element.
The consideration that some national system, such as Italian, have faced the great crisis in a less dangerous way can suggest that the interpretation believed as true maybe were not be so true. In this perspective we can understand the difficult adoption of some Standards, such as IAS 32 and IAS 39: their impact on financial statement has made vaster the effects of some facts that characterized these last years: the speculative bubbles and consequently the enormous falling of all the values.
Everybody knows the strong intervention the European Union has to make on the IASB in order to obtain the change of some consideration contained in the Standards and the subsequent difficulties that still make their application very difficult.

---

[1] 1606/2002; 1725/2003; 707, 2086, 2236, 2237, 2238 of 2004; 211, 1073, 1751, 1864, 1910, 2106 of 2005; 108, 708, 1329 of 2006.
[2] It is interesting to note how the change of the name attributed to the standards (from IAS to IFRS) is symptomatic of a developmental process that tends to separate the objectives of the financial statements from the needs for the "statement" of the general accounting and the perspective financial-reporting needs.
[3] Company which, in actual Italian Civil Code, are entitled to deliver abbreviated accounts to the Register of Company are those who, in the first year or for two consecutive year, do not superate two out of three of theese limits:
- total assets, euro 4.400.000
- total sales euro 8.800.000
- people occupied in the year: 50

[4] *Framework*, purposes and role, paragraph 3.
[5] PIETRO ONIDA, at the beginning of the 1950s (*Il bilancio d'esercizio nelle imprese*, Financial statement, Giuffrè, 1951, page 4 and pages thereafter) argued that it might be "already obvious in the doctrine that financial-statement valuations and, more specifically, the criteria for these valuations, can change depending on the purposes for which the financial statements are prepared, or better, depending on the information that one wants to get from the financial statements. Without specifying the purposes, as defined in this regard, it is not possible to prepare or to interpret any financial statements logically." As a consequence, if the different purposes are not compatible with one another, it is futile to want to reach the same objectives with the same financial statements, with the further corollary that, in the case of users who do not need distinct solutions, "one can debate about valuations till eternity, . . ., looking for the so-called 'the company's real situation or the real equity', which can be said to be . . . an aggregate as powerful as it mysterious." The author notes that at the time, and despite the level reached by doctrine, "the implicit or explicit pretence of having the same financial statements serve for the most disparate purposes was still very widespread, especially at the less sophisticated experts: for determining, for example, the income of a given period or the unitary economic value of the shares making up share capital . . .". Later (pages 109 and 110), he draws conclusions about the question, admitting the distinctiveness of the ordinary financial statements, whose purpose is represented by the determination of the income for the period, it being understood for this particular meaning as the result valid for the effect of any distribution or consumption of the earnings. Other information-reporting purposes normally need different valuation criteria, unless they are perfectly compatible with the first objective recognized for the ordinary financial statements.
[6] *Framework*, Paragraph 9.
[7] *Framework*, paragraphs 22 and the paragraphs thereafter.
[8] It is interesting to note that the purposes of the financial statements and their "postulates" serving as the basis of IAS/IFRS (mainly, the matching principle and the business-continuity principle) are indicated by the Framework, which, as stated, was not ratified by the EU.
[9] The issue has been addressed by most authors; among the explanations that are clearest and most rigorous, we suggest GIOVANNI FERRERO, *La valutazione del capitale di bilancio*, Giuffrè, 1995, page 179 and the pages thereafter.
[10] GIOVANNI FERRERO, *La valutazione del capitale di bilancio*, Giuffrè, 1995, page 14.
[11] COSTANTINO CAMPANINI, GIANFRANCO CAPODAGLIO, *Introduzione alla Economia aziendale (Introduction to economia aziendale)*, CLUEB, Bologna, 1988, page 92.
[12] It's important to note that this part of the paper was referred to the Exposure Draft of the IFRS for SMEs. Having a brief look at the final version we can observe that many chianges have been put in the new document, maybe following the observations made by O.I.C. Even the European Union has decided to postpone the implementation of the directive which will reform the subject, after a new consultation period.
[13] *"An entity shall recognise an asset in the balance sheet when It is probable that future economic benefits will flow to the entity and the asset has a cost or value that can be measured reliably"*.
[14] A recent article appearing in Il Sole-24 Ore stated that only 2,5 of every 1.000 Italian companies have revenues of more than EUR 50 million.

References

Alexander D., Nobes C., Caruso G.D., Ferrari E.(2008), *Financial accounting, il bilancio secondo i principi contabili internazionali*, Pearson Mondadori.

Campanini C., Capodaglio G. (1988) *Introduzione alla economia aziendale*, CLUEB, Bologna.

Capodaglio G. (1988), *Autofinanziamento, investimenti e variazione dei prezzi*, CLUEB, Bologna.

Ceriani G. (2001) L'inflazione e la significatività dei valori di bila*ncio,* in *Le rivalutazioni di bilancio,* Rirea.

D'Ippolito T. (1955), *I principi direttivi delle valutazioni di bilancio ed i limiti di massimo e minimo del capitale di funzionamento*, Lilla, Palermo.
Di Pietra R. (2005), *La comunicazione dei comportamenti aziendali mediante i dati contabili. Il ruolo della ragioneria internazionale*, Cedam.
Epstein B.J., Mirza A.A., (2004). *IAS. Interpretation and Application of International Accounting and Financial Reporting Standards,* Wiley.
Ferrero G.(1995), *La valutazione del capitale di bilancio*, Giuffrè.
Gray, S.J. (1984) *International Financial Reporting in 30 Countries*, Macmillan, London.
Guatri L., Bini M. (2005*)., Effetto Ias anticipati sui gruppi*, Il Sole 24 Ore.
Husmann S., Schmidt M. (2008), The Discount Rate: A Note on IAS 36, *Accounting in Europe*, Volume 5, Issue 1 June
Matacena A., *Il bilancio d'esercizio. Strutture formali, logiche sostanziali, principi generali*, Clueb, Bologna, 1993.
Nobes, C.W. (1981), An Empirical Analysis of International Accounting Principles: a Com*ment*, *Journal of Accounting R*esearch.
Onida P. (1951), *Il bilancio d'esercizio nelle imprese*, Giuffrè.
Parker R.H. (1996) Harmonizing the notes in the UK and France: a case study in de jure harmonization, *European Accounting Review,* V.
Quagli A. (2003), *Bilancio d'esercizio e principi contabili*, Giappichelli, Torino.
Rocchi F. (2006), Accounting and taxation in Italy, *European Accounting Review*, special edition, V
Saita M. (1988)., How to Integrate Italian and International Accounting Information Systems, *Economia aziendale*, n.1.
Tozzi I. (2001), Note in margine al tema dell'integrità economica del capitale, in *Le rivalutazioni di bilancio,* Rirea.

UVOĐENJE MEĐUNARODNIH RAČUNOVODSTVENIH STANDARDA U TALIJANSKA MALA I SREDNJA PODUZEĆA

SAŽETAK

S obzirom na to da se talijanska vlada sprema revidirati poreznu i civilnu financijsku regulativu, ovaj rad želi sažeti promišljanja talijanske računovodstvene znanosti i o uvođenju Međunarodnih računovodstvenih standarda u državni sustav. Promjene u računovodstvu koje vidimo posljednjih godina značajno utječu na financijske izvještaje europskih tvrtki: s jedne strane, tvrtke kotirane na burzi uvode Međunarodne računovodstvene standarde, dok su s druge, mala i srednja poduzeća suočena s novom regulativom koju se talijanska Vlada sprema preispitati.

***Ključne riječi**: IFRS, Malo i srednje poduzetništvo, mikro entiteti, direktive EU*

Šaban Gračanin[1]
Edin Kalač M.Sc.[2]

UDK 657:336.71(4+73)"20"

THE IMPACT OF FAIR VALUE ACCOUNTING ON THE CRISIS IN BANKING SECTOR OF EU AND USA

ABSTRACT

Authors who criticize fair value accounting (FVA) claim that the use of fair value accounting as a measurement attribute had essential impact on the origin, spreading and strengthening of actual global financial crisis. Similar outlook towards the possible impact of fair value accounting on the stability of global financial system and real economy was previously expressed by the European Central Bank (ECB), in its notes and assumptions. In the light of above mentioned criticism, of great number of researches and ongoing debates over the FVA issue, in the center of the world accounting and financial community currently are requirements for a deep reform and even withdrawal of FVA standards. In this work we analyze the correctness of such statements and the impact of fair value accounting on the EU and USA banks financial results before and during the crisis. We will try to answer the questions - Whether the use of fair value accounting contributed to originating and strengthening of the current financial crisis? Would the market have reacted differently if the banks hadn't shown financial losses during 2008? Should regulatory institutions still insist upon the use of fair value?

The analysis is based on the secondary data. Sources of secondary data for this topic are surveys, organizational records and data collected through qualitative research in literature.

***Keywords**: Fair value accounting, financial crisis, banks in EU and USA, FVA reform*

1. FAIR VALUE ACCOUNTING AND FINANCIAL CRISIS FROM 2008

What is the main lesson of the current financial crisis? Whenever regulatory bodies and existing accounting standards and legal acts do not strictly demand from financial institutions to timely present and face their losses, the losses can sharply increase. If, on the contrary, the Banks are by clearly defined regulations forced to timely present losses, and write-down assests due to them, they are therefore instigated to promptly take corrective actions, limit high-risk loans, which eventually limits the severity of the crisis.

In its pure form, fair value accounting understands reporting assets and liabilities at fair values, and recognizing changes in fair values in the form of gains or losses in income statement. The use of fair value is justified by the fact that it represents market value, and as such provides the users of financial reports a more complete, relevant and trustworthy information for business decision-making.

The most FVA critics are based on claims that FVA contributes to excessive write-downs of assets during the crisis. One most often speaks about the fact that the use of FVA during the crisis increases volatility of bank profitability represented in financial reports, and with that pro-

[1] Graduate economist, PhD student in Economics, University of Kragujevac, Serbia, and Head of BC Novi Pazar of Corporate Banking Division Eurobank EFG a.d. Beograd. +381 648853289 , saban.gracanin@eurobankefg.rs
[2] Head of Local economic development office - Novi Pazar City administration. Novi Pazar, Serbia +38163236-237 edin.kalac@novipazar.org.rs

cycality and limitation of Banks credit activities, which in the end deepens and prolongs recession periods and negative impacts of the crisis [7, p.7-9 ; 19, p.3 ; 14, p.109-111].

In its monthly bulletin [8, p.76-78], back in 2004, the Central European Bank (ECB) stated 4 possible scenarios, which describe the possible ways how application of FVA can cause problems to banking and overall financial system – decrease in quality of Banks loan portfolio, sudden changes of interest rates, crisis in the real estate market, sudden changes of prices of securities. Although the analysis of ECB mentions different market disorders, the mechanism most often mentioned through which FVA can contribute to origin and spreading of financial crisis is the bond that exists between accounting and mechanisms of control of banks obligatory capital reserves.

Market price changes can occur due to influence of great number of factors. If the changes of market prices are related to assets to which FVA is applied, banks are forced to write-down and decrease the book-value of those assets. That further leads to capital draining and forcing the banks to sell their assets in the market at lower (fire sale) prices, in order to acquire additional capital. If, after that, other banks accept those new discounted market prices as a representative measure of value, FVA leads to new write-downs, problems with maintaining obligatory capital reserves, liquidity problems, and spreading of Crisis as well - for these lower (fair) prices in the market become relevant for other banks too. Downward liquidity spirals occur, financial markets froze up, which finally results in banks bankruptcies.

Activation of downward cycles and spreading of crisis can occur in situation when Banks management is focused on short-term goals - short-term profit. This problem was particularly pointed out by a great number of authors and analysts, because bonuses in banking sector are usually based on achieved annual profits. Along with the fist signs of market disorder and fall in assets prices, the banking management mostly goes for selling relatively illiquid assets at lower prices, in order to precede their competition, and thus avoid higher losses and selling them at even lower prices. This further leads to spreading panic in financial sector, and later to unavoidable crises overflow into real economy flows [25, p. 4-6].

Above mentioned stresses that potential problems with application of pure FVA and its everyday use in financial reporting were seen long before the crisis. However, currently adopted accounting standards do not require use of pure FVA, and allow resigning from fair value in some cases.

2. FAIR VALUE ACCOUNTING CONCEPT

In practice, for preparing primary and secondary financial reports, two standard groups are most often used – EU International Financial Reporting Standards (IFRS) and Generally Accepted Accounting Principles of the USA (USA GAAP).

Companies quoted on US stock exchange are required by Security and Exchange Commission (SEC) to prepare and publish regular financial statements. These financial reports are prepared using US GAAP. The jurisdiction over reviewing and announcing GAAP was transferred from SEC to Financial Accounting Standards Board (FASB).

Companies quoted on the EU stock exchange prepare their financial reports by applying International Financial Reporting Standards (IFRS), prescribed by International Accounting Standard Board (IASB).

Both regulatory bodies, FASB and IASB, support the use of FVA.

GAAP concept of FVA - The definition and use of fair value in the USA[3] was arranged by Statement of Financial accounting standard 157 (FAS 157 – Fair value Measurements), published by FASB in September 2006 – today known as Topic 820 [34]. Fair value is defined as a price that would be received to sell an asset, or paid to transfer a liability in an orderly transaction between market participants at the measurement date. Standard defines hierarchy and order of information (inputs) entities must use to determine fair value of asset or liability.

The FAS 157 defines 3 Levels of information (inputs) which are used for determining fair value. [11, p. 22-32].

Level I – inputs are current, quoted prices at active markets, for identical assets and liabilities, that the reporting entity has the ability to access at the measurement date. If these prices come from regular (orderly) business transactions at active markets, they **must be used** for measuring fair value. FAS 157 clearly defines that orderly market transaction is not the same as "fire sale" or "forced liquidation" transaction. In case of fire sale or forced liquidation transactions, management should not use distorted prices in determining fair value. If Level I inputs are not available, one uses information of Level II.

Level II – inputs include all directly or indirectly available and observable market information not comprised by Level I. They include: market prices of similar assets and liabilities at active markets, market prices of same assets and liabilities from non-active markets (markets with small number of transactions, prices from previous periods, with low level of information released publicly), other information on assets/liabilities available in the market in regular intervals (interest rates, credit risk, default rates, prepayments spreads, income curves), as well as all other relevant information observable on the market [12, p. 1-2]. If an entity does not have these information at its disposal, the final option are Level III inputs.

Level III – inputs are unobservable inputs which are used if market inputs are not available to reporting entity, and they represent typical assumptions generated by company by using different prediction Models. These information represent an entity's assumptions on value of assets/liabilities that can be acquired in the market. Prediction models must use the best possible information available in the existing circumstances, taking into account the current situation.

Since there are significant differences in information quality used to define fair value by above mentioned three Levels, FAS 157 also requires disclosure of following information: [11, p. 32-35].

1. For all assets comprised by FVA – data on which Level was used for defining fair value
2. For each balance sheet item estimated by using Level 3 information – data on initial and ending values; changes in fair value during calculating period; gains/losses shown in income statement; gains/losses recognized in equity; all reclassifications of assets or liabilities in or out from this group.

These disclosures are very important for tracking Banks financial position and results, because investors are given a deeper insight and additional information for decision making.

IASB concept of FVA – the principles for acknowledging and valuation of financial assets and liabilities are defined by International Accounting Standard 39 (IAS 39) – *Financial instruments: acknowledgement and measurement.* The use of fair value in IFRS is not defined by

[3] FVA concept was not first introduced to GAAP with the publishing of SFAS-157. This concept is deeply integrated into the accounting system of USA. Many other, earlier published SFAS use this concept and provide guidance for reporting of assets and liabilities at Fair values. Some of the most important were – SFAS 107 – Disclosures about Fair value of Financial instruments; SFAS 115 – Accounting for certain Investments in Debt and Equity Securities; SFAS 119 – Disclosure about Derivate financial instruments and Fair value of Financial instruments; SFAS 133 – Accounting and reporting for derivate financial instruments and Hedging activities

this standard only, but is dispersed through a several other standards, which makes its use very complexed.[4]

As with US GAAP, the most reliable way to determine assets fair value according to IFRS is the use of quoted prices on active markets (Level I). If these prices are not available, fair value of a financial asset is determined by using Level II information (market prices of similar assets and liabilities in active markets, market prices from inactive markets and all other information on assets/liabilities available and verifiable in the market) , or by using appropriate assessment techniques - Models (level III). The entities are required to publish information on financial assets comprised by Level III inputs in their financial statements, changes of these values during a reporting period, and all reclassification from and into this group of assets.

IAS 39 allows an entity to make its own decisions which financial instruments will be balanced at fair values at initial classification. By initial classification, all financial assets are classified into one of the following five groups: Investment (securities) held till maturity, financial assets (securities) available for sale, held for trading assets (securities), securities valued at fair value through Income Statement, loans and receivables. After the initial classification of financial instruments in some of five above mentioned groups, additional reclassification was not possible up until the financial crisis.

Amendments of IAS 39 in October 2008, allowed reclassification of financial instruments from the "trading securities" group into group of available for sale securities (paragraph 50B) in rare circumstances, as well as reclassification of financial instruments from group of "trading" and "available for sale" securities into Loans (paragraph 50D) if an entity has intention and ability to hold that asset for the foreseeable future or until maturity [17, p. 4-6].

3. ACCOUNTING TREATMENT OF KEY ITEMS IN BANKS BALANCE SHEETS

Introduction and application of fair value concept had the biggest influence on financial statements and results of the banking sector. Since the financial instruments make a dominant fraction in Banks balance sheets, over 90% share with some banks, accounting treatment of these items represents a critical area for their business and success.

The accounting standards and rules for reporting financial instruments are based on mixed-attribute model [8, p. 76]. Treatment of each balance sheet item depends on type of an asset, and the way it will be used in regular business. *What is the share of total assets of Banks which was really balanced at fair values before the crisis? What possibilities are left to banks and other financial institutions to circumvent the use of FV for some assets during the crisis?*

3.1. ACCOUNTING TREATMENT OF KEY ASSETS IN BALANCE SHEET OF BANKS IN THE EU

Table 1 – provide data on averages of key financial assets of EU banks from 2005 to 2007 – before financial crisis. ***How were individual assets treated?*** The data are divided into two groups of banks, Large bank holdings (assets of over $2 trillion) and Small bank groups (assets below $1.5 trillion).

[4] I.e. requests regarding financial instruments presentation and reporting are defined by *IAS 32 – Financial instruments-presentation*. Requests regarding disclosures of financial instruments information are defined by *IFRS 7 – Financial Instruments- disclosures* + IFRS 9

Table 1

Key Balance Sheets Assets of EU Banks from 2005 to 2007

(as a fraction of total assets for 2005-2007 period)

Large Bank Holdings		***Small Bank Holdings***	
Trading assets	36,89 %	Trading assets	14,95 %
Net trading assets	3,06 %	Net trading assets	1,36 %
Securities at Fair Value trough Income Statements	5,49 %	Securities at Fair Value trough Income Statements	10,39 %
Available for sale securities	6,02 %	Available for sale securities	9,41 %
Loans and Leasing	42,57 %	Loans and Leasing	59,42 %
Financial instruments	**94,03 %**	**Financial instruments**	**95,54 %**
Total Assets	100 %	Total Assets	100 %

Notes – *Amounts represent average participation of financial instruments in total assets for Banks for 2005-2007 period in the EU. The data are divided into two groups, and as source we used data on banks ranking by total assets since the end of 2007 [4]. The data for big banking groups are related to banks with total assets of over $2 trillion, calculated on a sample base of five biggest banking groups in the EU as of late 2007: Royal Bank of Scotland, Deutsche Bank, Barclays PLC, HSBC Holdings plc, BNP Paribas Group. The sample for small banking groups is related to banks with assets less than $1.5 trillion and comprises 10 banks: (DZ Bank AG, Dansk Bank, BBVA Group, Fortis Bank, Santander Bank Group, Societe Generale Group, Unicredit Bank Group, Lloyd TSB Bank, Credit Suisse Group and Dexia Bank Group).*
Within each analyzed Bank group and year, observations are weighted by total assets. For all analyzed Banks, numbers were taken from the official annual reports and financial statements.

Securities *such as state bills and bonds, bonds issued by other state institutions, shares, equities and financial derivates can be classified in held for trading securities, available for sale securities and held to maturity securities. The classification of these assets is done by management of reporting entities.*
1 - Held for trading securities- *are purchased and kept to be sold in the near future. This position also includes derivates (not in a qualifying hedge position). These securities are reported at their* ***fair value****. Any consequential change of fair value of these securities is recognized in Income Statement as profit/loss. According to data for EU banks, these assets make 36, 89% of total assets for Large Bank Holdings, and 14.95% of Small Bank Holdings total assets.*
2 – Securities valued at fair value through Income Statement *are reported at* ***fair values****. Any consequential change of fair value is calculated in income statement as profit/loss.*
3 - Available for sale securities *are balanced at* ***fair values****. Impairment losses and exchange losses are recognized in income statement, using effective interest method. Other changes of fair values of these securities are recognized in a separate component of shareholder's equity, ?unit disposal, when the accrued gain/loss is recognized in Income statement. IFRS do not make a difference between temporary and other-than-temporary values changes of securities, as with US GAAP.*

4- Loans and leasing *represent the most important asset class of the most banking groups, and in the most cases these assets comprise over 50% of banks' total assets. All loans are divided into two groups: held for trading and held for maturity (investment) loans.*

- ***Held for trading Loans*** *are reported at* ***the lowest price principle or fair value****. Losses on the basis of decrease of fair value of these loans are recognized in Income Statement. The share of held for trading loans in total loans and leasing is very small, and they are included in trading assets class in Table 1.*

- ***Held to maturity (Investment) loans*** *are reported on the principles of* ***historical costs accounting****. Loans and leasing are reported at amortized value, by using the method of effective interest rate. These amounts are basis for possible impairments and write-downs in case of problems with loan repayments. Loan impairment occurs in case that it is possible that a Creditor bank will fail to charge the full amount of loan receivable - if "probable and predictable" loan losses exist. Loan amounts are written-down and their book value is decreased to current value of expected cash flow in the future*[5]*. Moreover,* ***financial institutions are required to disclose estimate of fair value of Loans in their financial reports.***

IFRS does not distinguish between investment in held-to-maturity securities and investment into loan agreements. Assets are classified as ***held-to-maturity*** *if they have fixed or determinable payments, a fixed maturity, or a bank has an intention and possibility to hold them in its portfolio until maturity. Regardless of form of the investment, investment with fixed or clearly defined repayments are generally sorted into Loans group, if they are not traded in the market and the investor does not plan their sale in near future.*

According to Table 1, for the observed banking groups in the EU from 2005 to 2007, 43.8% of total assets are balanced at fair values (cash equivalents, held for trading securities, securities at fair value through Income Statement and available for sale securities). Fair value of additional 51% of financial assets (loans and leasing) is published in Notes to the financial reports.

3.2. ACCOUNTING TREATMENT OF KEY ASSETS IN BALANCE SHEETS OF THE US BANKS

US GAAP for financial instruments most significantly differs from IFRS in the following respects [29, p. 31-32] :

1 - GAAP distinguishes between investments that are in the form of debt securities and those that are investments in loans.

2 - Prior to IAS 39 amendments in October 2008, IFRS had more restrictive requirements than US GAAP about transferring certain financial assets.

3 - Under IFRS, the trigger for recognizing impairment differs from U.S. GAAP, resulting in the potential for differences in the timing of when an impairment charge is recorded.

4 - Measurement of impairment losses differs under IFRS for HTM securities, which are written down through income under both U.S. GAAP and IFRS. However, under U.S. GAAP, these securities are written down to fair value; under IFRS, they are written down only for incurred credit losses.

5 - IFRS has greater restrictions on the use of the option to elect fair value accounting.

---

[5] *With US GAAP, if "probable and predictable" loan losses exist , loan amounts are impaired to their current fair value, and not to the value of expected cash flow in the future, as with IFRS.*

Research of C.Laux and C.Leuz [5, p. 98-100], provides data on averages of key financial assets the US banks had at their disposal from 2004 to 2006 - before the crisis in US. ***How were these assets treated?*** According to this research, during 2004-2006 period, only ***36% of total assets for Large Banks in US were balanced at fair value*** (trading assets, available for sale securities, and REPO agreements). Fair values of additional 50% of assets (Held to maturity loans and securities, and leasing) are subject of disclosures in Notes to the financial reports. For small bank holdings in USA, this fraction is even smaller. For investment banks, the fraction of balance sheets assets reported at fair values tends to be higher as they have a large trading book portfolio.

Data acquired by SEC in the USA, regarding the same issue [29, p. 47-49], point out that the percentage of total financial assets balanced by fair value decreased after the beginning of crisis. According to the research of SEC, at the end of the first quarter of 2008, ***only 13% of total assets*** of the US banks were balanced at fair value. The same research showed that 45% of total write-downs refer to fair value, 25% of which were directly recognized in Income Statements. The rest 55% of write-downs do not refer to use of fair value accounting, and are mostly related to Loans impairments!

4. DID THE FVA HAVE AN IMPACT ON THE CRISIS ORIGIN?

The financial crisis that began in the fall of 2007 proved to be one of the longer lasting periods of financial disruption in decades [26, p. 283]. Fall of real estate prices, delays and problems in loans repayment by debtors, terminations of financing contracts and mortgage sales, various cases of mortgage frauds and manipulations, decrease of credit rating – caused breakdown of "mortgage bubble" and huge problems with mortgage loans and financial investments in mortgage securities.

Uncertainty occurred on how these assets should be valued, and the anxiety of investors about the reliability of information and the quality of mentioned assets as well as of quality of exposure based on them. During the first phase of the financial crisis the US press was filled with reports from policy makers that financial markets had frozen up [26, p. 284]. It all led to a sudden withdrawal of capital from the markets and bankruptcy of great number of banks in the US, where the crisis originated. [22, p. 281]

According to data of the American Corporation for deposit insurance (FDIC), [13] before the crisis, in the period from 2000 to 2007, a total of 32 banks went bankrupt. After the beginning of the crisis, during 2008, 25 banks went bankrupt, and the number soared to 140 in 2009. The same trend was continued in 2010, where until September 17 another 125 banks pronounced bankruptcy. That this has been the worst crisis since the Great Depression, was confirmed by data of bankruptcy by decades (Table 2), from which we can clearly see that the number of banks that have proclaimed bankruptcy in 3-year period of 2008-2010 has surpassed the number of bankrupt banks in the six years period after the Great Depression.

Table 2

Proclaimed bankruptcies by decades in the USA

Period		Number
2000-2010	**:**	**322**
1990-1999	:	925
1980-1989	:	2036
1970-1979	:	79
1960-1969	:	44
1950-1959	:	28
1940-1949	:	99
1934-1939	**:**	**312**

Source: FDIC/Failed Banks

But, the fall of real estate prices and sudden downturn in quality of banks loan portfolio led to draining of the market due to the reasons which are not connected in any way to accounting concept of financial reporting. *Would the market have reacted differently if the banks had not had shown financial losses?*

If the market reaction would have been the same, it would then be very difficult to claim that the FVA by itself caused the crisis. During 2006 sub prime mortgage lending and the securitization of those loans accelerated [26, p. 285]. The fact is that the Banks were heavily reliant on mortgage financial arrangements before the crisis [30]. But the amount of debt/receivable that can be collected from mortgage agreements depends on market collateral prices, and not on the book ones. Thus, after the first signals that point out to problems in real estate market, investors would become worried for business and value of shares and equity of banks with great mortgage exposure, even if the banks refused to write-down value of mortgage-based assets, and continued to report these assets at their historical costs. Always cautious investors would certainly become worried on first signs of crisis in real estate market. Investors would certainly react!

What is the influence of FVA on such sequence of events? Empirical evidence does not support critics that FVA is the main cause of the crisis – of enormous assets write-downs and loss admittance of banks due to market prices disorder, and then liquidity problems and spreading of crisis to the real sector.

Some data for the biggest investment banks in the US , show that at the beginning of the crisis, book values of the banks' assets were far higher than their market (fair) values, but even that was not enough to improve investors confidence and stability of global financial sector. For example, empirical data for Merrill Lynch show that accounting values of mortgage-based assets were 65% higher than their exit prices, achieved on the market [24].

A very important fact missed by many is that the biggest share of write-downs by banks is not related to use of FVA, but to Loans write-downs and losses [33]. And the loans are item in balance sheets which is reported by historical cost accounting principles. For example, Bear Sterns published high mortgage losses in early June 2007 in its two hedging funds, and in December 2007 for the first time in its history reported quarterly loss. In January 2008 Bank of America took over one of the biggest creditors and distributors of mortgage loans in America – Countrywide, in order to prevent it from going bankrupt [26, p. 288].

Also, the greatest financial giants before the crisis based their businesses on strategy – borrow short-term, lend long-term. The data in 2007 annual reports of American International Group (AIG) show the fact that $75 Billion, or 14% of total liabilities were financed by short-term sources. This strategy, together with provisions and loses on additional reserves due to

quality decrease in loan portfolio, are the main reasons for banks high losses and write-downs. *Thus*, *FVA did not have essential influence on origin of crisis, because it is not the basic cause of market disorder.*

5. DID FVA HAVE IMPACT ON SPREADING AND STRENGTHENING OF CRISIS?

In order to determine if FVA have influenced on spreading and strengthening of crisis, it is necessary to explain the impact of FVA on behavior of banks management and the structure of financial statements during the crisis.

The standpoint of American Banking Association (ABA) [1] is that FVA suits only to positions of trading financial assets. Their position is that reporting of loans and leasing at fair values, as well as available for sale and held to maturity securities, is not appropriate and can lead to wrong business decisions, especially in periods of crisis [2]. ***However, presented data for EU and USA indicate that the greatest fraction of Assets in the balance sheet of bank holdings were not reported at fair values. When banks do apply fair values, used rules of values estimation obviously deter from pure FVA. This is happening because the current standards allow banks to deter from fair value in some situations. How?***

1 - Position with largest share in Banks total assets, portfolio of investment (held to maturity) loans, is not comprised by FVA in the balance sheet, but is a subject to impairments by the rules of historical cost accounting.

Banks with large share loans and leasing in total assets, which is mostly the case (Table 1), can avoid effects of FVA, by classifying loans as held for investment. The empirical data for 31 banking group that bankrupted during 2007, or was capitalized by US government, show that loans made up around 75% of total assets, and that share of held for trading assets within that position was very low. The data for 10 small banking groups in the EU show that share of the loans position went as high as 70%. Likewise, held to maturity securities are not reported at fair value.

Share of mortgage loans for construction financing and residential premises purchase was above 47% of total assets, for most of the banks in the US that went bankrupt from 2008 to 2010 period. The main cause for such high share of investment mortgage loans is their high profitability for the banks, especially after 2000. The average growth of income for most of the banks that went bankrupt during the crisis was approximately 18% a year, from 2003 to 2006. Average annual growth of income for the banks with share of investment loans above 30% was 53% in the same period. When in the early 2008 the quality and collection efficiency of these loans started falling sharply, the pressure on banks started rising quickly. According to the latest data of FDIC, in the first quarter of 2010, the share of investment loans with delays of more than 90 days was 16.82%. With banks that went bankrupt in 2010, the share of loans with delays longer than 90 days and non-repayable loans in total loan portfolio was almost 40% [30]. It is important to repeat – this reporting position is not under direct influence of FVA

The banks are required to publish fair values for position of loans and leasing. One can not, however, claim that disclosure of fair values of these financial assets in Notes had an impact on sharpening and spreading of the crisis. Since the issue of mortgage loans and mortgage bubble was discussed long before the crisis, prudent investors were already worried for banking business and results, regardless of the fact that the fair value data were or were not disclosed. "***For years, it has been clear that American growth was not sustainable. It was based on a real estate bubble, which sustained a consumption boom"*** [22, p. 281]**.** Instead, publishing these data would suppress panic in the market. Moreover, it would disable banks to neglect current and

potential problems, and these information would actually be early signals for undertaking corrective actions. All that would limit high risk credit activities and severity of disorder that the crisis brought to global financial system.

2 - With analyzed banks in the EU (Table 1), the position of trading securities has an average share up to 10%. With banks in the US this position is second largest by share in total assets. However, FVA in the US GAAP has only limited impact on this position – changes of fair value are recognized in financial statements only if entity management estimates that fair value changes of these securities are "other than temporary" (OTTI rule). Profit/loss due to fair value change is in only in that case recognized in Income statements. If on the contrary, an entity's management estimates that value changes of these securities are only temporary, they are not recognized in income statement, but are shown as a separate component of shareholder's equity. If a bank has an intention and ability to keep these securities long enough until the market and the prices recover, it can treat these losses as only temporary and thus avoid effects of FVA on success and reserves. However, the question is how long banks can justify temporality of such losses in the market, during the crisis and its spreading.

During 2008, as effects of financial crisis grew stronger, and its overleaping on real sector occurred, banks found it increasingly difficult to give arguments and prove that losses on securities were only temporary. This is precisely the period when political pressures on IASB and FASB started, in order to soften standards which impose the use of fair values.

These pressures resulted in IASB reacting, and softening the use of fair value, with intention to decrease the negative effects of crisis, on October 13th 2008. Pronounced amendments of IAS 39 understood possibility of reclassification of financial instruments, in special situations such as financial crisis for example, which was not allowed previously [15]. Paragraph 50B allowed reclassification of trading securities into group of held for sale securities and paragraph 50D reclassified instruments from trading and available for sale into group of loans, thus avoiding balancing them at fair values.

FASB also amended FAS 157 in May 2009. As of June 15, 2009, all other than temporary changes of fair value are divided into two groups: 1) Loan losses and 2) other losses. Only other than temporary changes and Loan losses are recognized in Income Statement.

3 – The remaining positions, directly under the impact of fair value accounting, are trading assets and securities at fair value in Income statement. All relevant regulatory factors, including ABA, agree that the treatment of these positions by fair value is a correct one. But, only the biggest bank holdings are entitled to occupy large amounts of these liquid assets, which they use for financing their investment activities. For example, JP Morgan and Citigroup in the US had a 19% and 16% share of trading securities in total assets. The share of trading assets in total assets of Royal Bank of Scotland and BNP Paribas is even 50%. Although recognized losses of these banks during 2008 regarding trading securities are not small, they are not the main cause of origin and spreading of crisis.

Moreover, for trading securities, both GAAP and IFRS do not require strict use of FVA. US GAAP and EU IFRS have several measures of protection against use of fair value, in case of distorted market prices, which limits impact and influence of accounting on crisis spreading.

First – FAS 157 explicitly impose that fire sale prices or liquidation prices should not be used when estimating fair value. If these prices are acquired on the market, banks are not required to define value of other assets positions in accordance with them. It is very hard in practice to define which prices have come out from fire sales – but this rule gives banks the right to discard extreme prices appearing on the market.

Second – banks alone determine how to classify its securities (according to FAS 115 and IAS 39), and thus by defining the amount of trading securities, determine which share of assets

will be under impact of FVA. Besides, under rare circumstances (such as crisis) defined by standards, a bank can reclassify securities. For example, Citigroup reclassified trading securities of $60B in the last quarter of 2008, defining them held to maturity. By doing that, this group limited the impact of market prices decrease on business results and share equity. Notes to the financial reports reveal that during 2008, Royal bank of Scotland reclassified €59.729 mil from position of trading and available for sale securities into group of loans. Deutsche Bank in the same way reclassified €34.424 mil, and BNP Paribas €7.077 mil, during the same year.

Third, when markets become inactive, and quoted market prices are no longer at analysts' disposal, banks are not required to use data and distorted prices from non-liquid markets, and FAS 157 and IAS 39 explicitly allow banks to use models (unobservable Level III inputs) for estimation of fair values of financial instruments. Table 3 data confirms that Banks seized this option, allowed by accounting standards.

Table 3

Structure of assets at fair value by level of valuation

	Investment Banks in US		***Large Banks in US***		***Large Banks in EU***	
	Dec-07	Dec-08	Dec-07	Dec-08	Dec-07	Dec-08
Level I	25,20%	15,50%	29,30%	18,60%	36,60%	23,40%
Level II	64,60%	70,20%	59,60%	68,20%	59,80%	73,60%
Level III	10,20%	14,30%	11,10%	13,20%	3,60%	3,00%
FV/Total assets	48,00%	51,80%	32,00%	29,90%	57,46%	63,46%

Notes – *data in the table are related to period from late 2007 to late 2008. For all analyzed Bank Groups, numbers and data are taken from their official annual financial reports and Notes to the financial statements.*

Sample for big bank holdings in the US includes four banks data - JP Morgan, Bank of America, Citigroup and Wells Fargo. Sample for US investments Banks comprise - Goldman Sachs, Morgan Stenley and Merill Lynch. Sample for banking groups in the EU comprises 3 banks - Royal Bank of Scotland, BNP Paribas and Deutsche Bank

Of total financial assets the biggest banks in US and EU reported at fair value in 2007, only 30.4% of financial assets on average were valued at fair values using Level I information. At the end of 2008, this percentage was decreased by 11.2% on average. This reduction of Level I inputs usage was compensated by increased share of fair values determined by using information of levels II and III. Share of financial assets whose fair value were determined using models (information of level III), increased from 9% to 13% on average!

What is even more important is the information that reclassifications were conducted in the very beginning of the crisis. In the EU, transfers into level III category, go up to 50% for the largest five banks: Royal Bank of Scotland, BNP Paribas, Deutsche Bank, HSBC Holdings plc and Barclays Bank. In the US, from the last quarter of 2007 to the last quarter of 2008, total assets transferred into category of level III go from 40 to 80 % of total financial assets comprised by FVA. This percentage is highest with banks that suffered the strongest negative crises effects.

Table 3 data clearly point out that the current standards gave the banks enough possibilities and options to avoid negative effects of sudden decline of market prices, and to use Models (unobservable information) and their own estimations when determining fair values, even at the very beginning of the crisis. Share of assets whose fair values were determined directly by using information/quoted prices from the market (Level I) constantly decreased as of early 2007, after the first signals of the crisis.

An SEC research [29, p. 61-62]**.** shows that at the end of the first quarter of 2008, only 7% of fair values were determined by using quoted market prices, and on the other side 82% and 11% by using less reliable Level II and Level III information. Share of used Level III inputs grows during 2008. *Banks financial reports are significantly protected from effects of fair value changes.*

6. EMPIRICAL ANALYSIS OF FVA IMPACT ON EU AND USA BANKS

FVA issues have been in focus of financial community for the last 18 months, due to numerous debates and announced FVA reform. All this led to many researches on FVA impact on financial reporting and crises of EU and USA banking sector. Academic works on impacts of FVA are just appearing in science magazines. Researches show that the effects of fair value accounting on incomes and required capital reserves of the EU and USA banks were far less than often claimed. During the crisis, most banks used allowed maneuver space in applying FVA.

The fact that the banks greatly avoided use of fair value is confirmed by mentioned research of SEC from 2008. Moreover, US GAAP allows entities possibility to report some assets voluntarily at fair value (fair value option) through several standards (FAS 155, 156, 159) [29, p. 31-32]**.** These options refer only to hybrid financial derivates and services. When an entity goes for reporting these assets positions at fair values, return to previous measurement attribute is no longer possible, by standards. The research has shown that only 4% of total assets with allowed "option" were reported at fair value. [29, p. 55]**.**

Additional evidence on Banks refusal to recognize and report losses due to decrease of assets fair value, is the accounting treatment of goodwill position. Banks, by taking over and acquisitions of smaller banks, acquire the right of activation of purchased goodwill. This position can be written-down, in case when fair value of goodwill is lower than the book one. However, Banks did not impair the value of goodwill, although real business value of banks decreased steeply in the beginning of crisis. Our analysis for 4 major bank holdings in EU in the early 2008 shows that only Royal Bank of Scotland in 2008 had significant write-downs of goodwill, £30.062 mil, while in period of 2005-2007 there was no write-downs. With other largest bank groups (Deutsche Bank, Barclays PLC Bank and BNP Paribas bank) there was no significant write-down of goodwill from 2005 to 2008. Research performed by an independent investment company Disclosure Insight [6] shows that 50 major banks in the US before the crisis had a great number of acquisitions and takeovers of smaller banks, but 35 did not write-down the value of goodwill, although real business value of banks decreased steeply in the beginning of crisis.

Opposite to these researches, some authors (Kalin Kolev; Song, Tomas and Yi; Goh, Ng and Young,) [32; 23; 36] reached to opposite conclusions in their analysis, that FVA is the main cause of starting and spreading the crisis. They point out to the facts that reported and disclosed fair values of financial assets do not differ much from their book values – i.e. that write-downs of financial funds did exist and that they are the cause of crisis. Other researches (Issak, Forbes, and ABA) emphasize negative FVA effects on financial system and crises effects, but they don't provide clear evidence which supports these stands.

Taking all into consideration, pro and con FVA, there is insufficient evidence to prove that FVA had a strong impact on EU and USA banking sector, and thus contributed to origin of financial crisis. How correct our conclusions are, will be shown by great number of researches on this topic, which are certainly necessary to bring a final conclusion and define steps for further development and reform of accounting standards. *What are the main features of ongoing reform of current FVA standards in the US and EU?*

7. FVA REFORM – BASIC ISSUE

In the light of mentioned critics, large numbers of works on this issue and ongoing debates, in focus of world accounting and financial community currently is announced reform of FVA standards. Basic issue of further reforms is should regulatory institutions decide for softening or sharpening of FVA standards. One should take into account a great number of trade-offs regulatory bodies will meet during the very changes.

"***Softening" FVA standards,*** on one hand, would give entities management more space to avoid problems that fair value reporting can cause during crisis. But, on the other hand, it opens huge possibilities for manipulation and blurring financial reports, which would lead to decrease in reliability of financial statements. Transparency decrease of financial reports would condition even higher precaution of investors during the crisis, and therefore faster spreading of crisis, than a strictly defined use of FVA.

"***Tightening" FVA standards*** and requests for their strict use during the crisis, for example in case of market disorder and price decrease, can contribute to faster spreading of crisis and sharpening of its negative effects. However, in such situations, it is necessary to compare these negative effects with positive effects from "timely "recognized losses, and corrective actions taken on the basis of them.

The application of FVA at the same time brings both positive and negative effects [7, p. 7-13] – it basically can speed the downturn of economy activities during the crisis, but on the other hand it can contribute to faster identification of losses during the crisis and timely corrective actions by regulatory institutions. If the banks are forced by standards to show losses, and write-down assets value due to fair values decrease, they are thus stimulated to take immediate corrective actions, and to limit high risk crediting, which eventually limits the severity of the crisis. This further means that demands for withdrawing current FVA standards that require are over exaggerated. If the final goal of such request is to prevent economy pro-cycality, probably it is better to use some of the instruments of monetary and fiscal policies – i.e. regulating the Banks obligatory reserves.

That this stand is correct, was confirmed by the latest decision of Basel Committee for Bank Supervision and adopting of ***Basel III agreement*** by representatives of 27 member states of this committee. Only several years after the adoption of Basel II agreement in 2004, when it was deemed that the formula for limiting high risk crediting of banks and higher stability of financial system was found, the world's greatest financial giants went bankrupt and world financial crisis occurred. The new Basel III agreement was adopted on September 12, 2010 and it foresees banks to put aside a €7 reserve on each €100 disbursed loans, instead of €2. The goal of these measures is to prevent breakdowns in markets in future. The penalties for banks whose reserves fell below 7% will be implemented through decrease in dividends and bonuses to management [20, p.46-47]**.** The deadline for adjusting to new rules is the beginning of 2019.

8. MAIN PURPOSE OF FVA REFORM – UNIQUE FVA STANDARD

What is the main purpose of current FVA standards reform? The goal is defining and publishing a unique, global Fair value accounting standard, which will represent sole source of rules and instructions of fair value accounting. It should: provide fair value definition, define all positions comprised by fair value, provide clear frames for fair value reporting, and precisely define requirements for disclosing fair values for some positions in financial reports [18, p.5]**.** Thus all vagueness in applying fair value that currently exists will be eliminated, due to the fact that use of fair value in IFRS and GAAP intertwines through a great number of standards, what makes its use very complex.

How important precise defining of FVA use in regular annual reporting and establishment of unique FVA standard is, shows complemented Memorandum of understanding of FASB and IASB on standards development (MoU). Mutual project of both regulatory bodies, FASB and IASB is directed toward increase of transparency and decrease of complexity of financial instruments balancing. Activities are divided into three phases: phase I – classification and valuation of financial instruments, phase II – methodology of financial instruments value impairment, phase III – financial derivates reporting [9, p.4-5].

Basic principles on which cooperation of the two Boards is based are: increasing comparability of information and usefulness of information for investors; for trading financial instruments used timely information on fair values are very important and necessary; classification of financial instruments must be less complex than the current one; instruments and methods of establishing fair value must be less complex than the current ones; principles and rules of impairment must be the same for all financial instruments.

Declaring of a unique fair value standard will mean an end to a long-lasting debate between two sides – pro and con FVA. One of main advocates against wider use of FVA is American Bankers Association (ABA). Treatment of financial instruments is the key area for banking sector. This is why the way of revising standards by FABS and IASC is very important for ABA.

The main point of misunderstanding is the issue of accounting treatment of loans – reporting values in balance sheet and methods of impairment. While IASB, on one hand, is firmly standing on its current positions, FASB on the other hand presses IASB to wider use of fair value. The FASB intentions are that banks should report all Loans and long-term financial assets in assets, in balance sheet at fair values. But that is something that currently doesn't exist as request in US GAAP.

Published drafts of announced reform of current FVA standards (IAS 39 and FAS 157), in July 2010, clearly points that both FASB and IASB support further use of FVA. The impression is that FASB pushes IASB to wider use of fair value. Publishing of unique FVA standard is planned for first quarter of 2011.

9. CONCLUSION

Based on empirical data, as well as literature on this topic, there is insufficient evidence to prove that FVA had strong negative impact on EU and USA banking sector, and that it contributed to origin of financial crisis, and later on to greatest world crisis since the great depression of the 30-ies of the 20th century. Always cautious investors would certainly become worried on first signs of crisis in real estate market, as well as for businesses of Banks with high

exposure to mortgage loans of poorer quality of portfolio, no matter whether official financial reports of banks contain data on fair value of assets. Investors would certainly react!

FVA did not have large influence on spreading and severity of the crisis effects. FVA generally had very small impact on structure of Banks financial reports, as well as on obligatory reserves and losses, except for small number of banks in the US and EU that had greater share of trading assets in balance sheets. Besides, current standards contain different measures of protection and give banks a significant freedom of choice that enables them to avoid balancing assets to deformed market prices. Researches show that banks greatly used this allowed flexibility during the crisis.

This further means that requests for withdrawing current FVA standards are also rash. More research on this issue is necessary in order to determine effects of FVA impact during the crisis and its expansion, and to establish further FVA reform, which is certainly necessary. Regardless of critics, contents of new amended FVA standards, planed for publishing in first quarter 2011, will confirm the stands and opinion of most authors – that FVA is not the main cause of origin and spreading of negative effects of the financial crisis.

REFERENCES:

1. **American Bankers Association** „Fair Value and Mark to market Acounting", at www.aba.com/issues/issues_FairValue.htm, 2009.
2. **American Bankers Asociation** – "Mark-to-market accounting:" The battle will take place this year", July 2010. at http://www.ababj.com/aba/mark-to-market-accounting-the-battle-will-take-place-this-year.html
3. **American Bankers Association** "Letter to SEC", 2008,
4. **Credit Write-Downs** "The Top 25 European banks by assets" available at http://www.creditwritedowns.com/2009/02/the-top-25-european-banks-by-assets.html
5. **C. Laux and C. Leuz** - "Did Fair-Value Accounting Contribute to Financial Crisis?", Journal of Economic Perspectives, Volume 24 (1), 2010.
6. **Disclosure Insight** – "Bank Goodwill Impairment Study "research from March 2009, at http://www.scribd.com/doc/14454673/Banks-Disclosure-Insight
7. **European Central Bank** - "Fair Value Accounting and Financial stability", Occasional Paper Series no 13, April 2004
8. **European Central Bank** – "The impact of FVA on the EU banking sector" Monthly Bulletin, February 2004.
9. **FASB and IASB** - Reaffirm Commitment to Memorandum of Understanding, A joint Statement of FASB and IASB, November 5, 2009
10. **Financial Accounting Standards Board –** "Fair Value Measurements and Disclosures" at http://www.fasb.org/cs/BlobServer?blobcol=urldata&blobtable=MungoBlobs&blobkey=id&blobwhere=1175820873224&blobheader=application/pdf , June 2010
11. **Financial Accounting Standards Board** – FAS 157 (as amended), Available at http://www.fasb.org/cs/BlobServer?blobcol=urldata&blobtable=MungoBlobs&blobkey=id&blobwhere=1175820931833&blobheader=application%2Fpdf
12. **Financial executives international** – „FASB issues FAS 157: Fair Value Measurement", FEI Summary, Sep 2006.
13. **Federal Deposit Insurance Corporation** – Failed Bank List since October 2000, at http://www.fdic.gov/bank/individual/failed/banklist.html

14. **International Monetary Fond** - "Fair Value Accounting and Procyclicality " Global stability report, October 2008, at http://www.imf.org/external/pubs/ft/gfsr/2008/02/pdf/chap3.pdf
15. **International Accounting Standards Board** – "IASB amendments permit reclassification of financial instruments" , Press Release, 13.10.2008
16. **International Accounting Standards Board** – IAS 39 Implementation Guidance
17. **International Accounting Standards Board** – "Reclassification of Financial Assets - Amendments to IAS 39 Financial Instruments: Recognition and Measurement and IFRS 7 Financial Instruments: Disclosures", at http://www.iasb.org/NR/rdonlyres/BE8B72FB-B7B8-49D9-95A3-CE2BDCFB915F/0/AmdmentsIAS39andIFRS7.pdf
18. **International Financial Standards Board** – "Developing common fair value measurements and disclosure requirements in IFRS and US GAAP" , Project Summary, July 2010
19. **Isaac, W** – "Prepared remarks for SEC roundtable on mark-to-market accounting", 2008 at - http://www.sec.gov/comments/4-573/4573-79.pdf
20. **Ivana Radmila Djurdjic** - " Nova Bankarska pravila ", Ekonom:east magazine, Sep.2010, no 539.
21. "**Memorandum of Understanding**" - November 5th , 2009. At http://www.fasb.org/cs/ContentServer ?c=Document_C&pagename=FASB%2FDocument_C%2FDocumentPage&cid=1176156535882
22. **Joseph E. Stiglitz,**– "The current Economic Crisis and Lessons for Economic Theory", Eastern Economic Journal , 2009, 35, p 281-296
23. **Kalin Kolev** - „Do Investitors perceive Marking to market as marking to Myth? Early evidence from FAS 157 Disclosure" ,at http://ssrn.com/abstract=1336368
24. **Keunon Bradley and Christine Harper** – "Merrill has $5,7 billion Write-down's, Sales Shares" , Bloomberg.com, July 2008
25. **Plantin G., Sapra H. and Hyun S. Shin** – "Marking to market: Panacea or Pandora's Box" , Journal of Accounting Research, 46(2), 2008
26. **Robert A. Eisenbeis** – "The Financial Crisis: Miss-Diagnosis and Reactionary Responses", Atlantic Economic Journal, 2010, 38: 283-294
27. **Ramanna Karthik and L.Watts** – „Evidence from Goodwill non-impairments on the Effects using Unverifiable estimates in financial reporting", Harvard Business School, Unit working paper 09-106, 2009
28. **Savez Racunovodja Srbije** – „Praktična primena MSFI u Republici Srbiji" , Savez Računovođa Srbije, Second Edition, 2008
29. **Securities and Exchange Commission** – "Report and recommendations pursuant to Section 133 of the Emergency Economic Stabilization Act of 2008: Study on Mark-to-Market Accounting" , 2008
30. **Standard Poor's** – "Credit trends: An increasing number of Failed Small U.S. banks is taking a toll on local Economies", 2010, http://www.standardandpoors.com/products-services/articles/en/us/?assetID=1245225768756
31. **Steve Forbes** - "End mark-to-market" , March 2009 , at http://www.forbes.com/2009/03/20/steve-forbes-mark-to-market-intelligent-investing-market.html
32. **Song , Chang J, Wayne B Tomas and Han Yi** – „Value relevance od FAS 157 Fair Value Hierarchy Information and the Impact of Corporate Governance Mechanisms" at – http://ssrn.com/abstract=1198142

33. **The Economic Minute**: "Mark-to-market Accounting – Transcript of John Sizer letter to Economic Club of Phoenix members " 2009, at http://knowledge.wpcarey.asu.edu/article.cfm?articleid=1771
34. **www.fasb.org**
35. **www.sec.gov - Securities and Exchange Commission** – "Final Report of the Advisory Committee on improvements to Financial Reporting to the SEC", August 2008
36. **Wee Beng, Goh, Jeffrey Ng, Kevin Ow Yong** –„Market pricing of Banks Fair value Assets Reported under SFAS 157 during the 2008 Economic Crisis", at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1335848

UTJECAJ RAČUNOVODSTVA FER VRIJEDNOSTI NA KRIZU U BANKARSKOM SEKTORU EU I SAD

SAŽETAK

Autori koji kritiziraju računovodstvo fer vrijednosti (Fair Value Accounting – FVA) ističu da je upotreba „fer vrijednosti" kao mjernog atributa utjecala na nastanak, širenje i jačanje efekata globalne financijske krize. Slične stavove o mogućem utjecaju računovodstva fer vrijednosti na stabilnost globalnog financijskog sustava i realnu ekonomiju, dosta ranije iznijela je i Europska centralna banka (ECB), iznoseći svoja zapažanja i pretpostavke. U svijetlu spomenutih kritika, velikog broja radova na temu FVA i debata koje su u tijeku, u centru pažnje svjetske računovodstvene i financijske javnosti trenutno su zahtjevi za dubokom reformom pa čak i povlačenjem FVA standarda. U ovome radu analiziramo točnost ovakvih tvrdnji kao i utjecaj "računovodstva fer vrijednosti" na poslovanje banaka u EU i SAD prije i tijekom krize. Pokušat ćemo dati odgovor na pitanja – da li je upotreba "fer vrijednosti" pridonijela nastanku i jačanju efekata financijske krize? Da li bi tržište reagiralo drugačije, da banke tijekom 2008. nisu prikazale financijske gubitke? Treba li i dalje inzistirati na široj upotrebi "fer vrijednosti"?

Analiza je zasnovana na sekundarnim podacima. Izvori sekundarnih podataka za ovu temu su istraživanja, evidencije organizacija i podaci prikupljeni putem kvalitativnih istraživanja u literaturi.

***KLJUČNE RIJEČI**: Računovodstvo fer vrijednosti, financijska kriza, banke u EU i SAD, reforma FVA.*

PhD Tihomir Jovanovski[1]
M.sci Mehmed Muric[2]

UDK 336.7:65.015.2

THE PHENOMENON OF LAG IN APPLICATION OF THE MEASURES OF MONETARY POLICY

ABSTRACT

This paper discusses the theoretical aspect of the phenomenon of lag in the application of the measures of monetary policy. Monetary and fiscal policy faces the phenomenon of lag. One of the controversial and pressing questions of monetary policy is the nature and length of lag between the application of the measures of monetary policy and the effects on macroeconomic aggregates. While the monetary strategy points to several possible reasons for the lag, there is no general agreement on the length of the lag. The study of the phenomenon of lag imposes two questions: what does the lag of the application of monetary processes and policy imply and why are the asset holders not capable of immediately adjusting their portfolios at the time of disbalance? There are two categories of lag known in economic literature: the inside lag (which encompasses the recognition of the problem and the implementation of measures) and the outside lag (which encompasses the reaction of macroeconomic aggregates to the applied measures of monetary policy). The paper descriptively notes the different identifications of time lag and provides schematic representations of the effects of the observed phenomenon.

Key words: *Monetary Policy, measures of monetary policy, time lag, problems of lag, consequences.*

INTRODUCTION

The application of monetary politics and achieving the goals of monetary politics would be significantly facilitated if complete information about the effects of transitional mechanisms of monetary processes, as well as the actual relations between the instruments of monetary policies, financial and real variables existed. In such a case, the only thing left to the bearers of monetary policy to do is to fulfill the task at hand to the best of their capabilities, in respect to the handling of available instruments of monetary regulation in a way that results in known and wanted effects on the level and extent of change of final goals of monetary policies.

Within the process of formulation and application, monetary policy is marked by a far greater lack of information than most of the other economic policies. When discussing other economic policies, such as fiscal or foreign currency policies, it is possible to appropriately evaluate the situation and the effect of the changes brought on by one of the instruments of monetary policies on the variables of the ultimate macroeconomic goals.

As far as monetary policy is concerned, what remains unknown are the full effects of the change in the reference interest rate of the central bank or the change in reserve

[1] European Center for Peace and Development (ECPD) – University for Peace established by United Nations.
[2] PhD student. Societe Generale Bank Serbia, mesamuric@gmail.com

requirements with respect to inflation, employment or any other ultimate outcome variable of that policy, or even when that effect (i.e. with what time span) will be visible. The reason for this can be found in the fact that there is no complete information concerning the relations between the monetary sector of one economy and its real sector. The empirical experiences of the transfer of monetary impulses onto real variables and the feedback of the effects from the real towards the monetary sector of economy are still unsatisfactory [Živković & Kožetinac, 2008, 291].

Monetary and fiscal policies face the phenomenon of lag. One of the controversial and pressing questions of monetary policy is the nature and length of lag between the application of the measures of monetary policy and the effects on macroeconomic aggregates. While the monetary strategy points to several possible reasons for the lag, there is no general agreement on the length of the lag [Cagan & Gandolfi, 1969, 277].

The goal of this paper is to explain the phenomenon of lag in application of the measures of monetary policies from a theoretical aspect.

1. THE PHENOMENON OF LAG

The theoretical and empirical research into lag of the effects of monetary policies, as well as the phenomenon of lag itself, is needed because these measures are not visible immediately and fully, but after some time and appear inconsistently.

Essentially, there is no single occurrence which could be referred to as lag of the effects of monetary policies. If we presume that the effect of a single momentary monetary change to national income could be fully isolated, it would doubtlessly be seen that the effect immediately starts rising to its high point and then gradually decreases and does not stop for a certain period of time. Thus, the term of distributed temporal lag can be discussed. When we talk about a specific kind of lag, we usually refer to something as the average interval between the application and effect of a certain measure [Friedman, 1973, 262].

The application of monetary policy, but also macroeconomic policies in whole, is complicated by the presence of lag in the ultimate application of undertaken measures; the lag between the action of the policy and its effects on macro aggregates (production, employment, income, etc.); the lag between recognizing the need for action and measure implementation based on the prognosis of future economic conditions. The presence of change lag of political variables is burdened by the risk of the nature of the problem, which caused the change in political actions, being completely altered during the application of policies [Cooper et al., 1983, 540].

Monetary policy produces effects to economic motions in the situation of portfolio imbalance (real and financial property) of the transactor, which effects the purchase and sale of financial and real assets to restore balance, when the marginal unit of each asset will reject equal income, or when there will be no advantage of changing the possession of different forms of assets.

Further study of the lag phenomenon imposes two questions: what is implied by lag during monetary processes and why are the asset holders unable to immediately adjust their portfolios in the conditions of imbalance?

It is a fact that the transactors do not immediately adjust to altered circumstances for two reasons: (1) they need time to become aware of the conditions and (2) when they become aware of the imbalance in their portfolios, they need time to correct them [Crockett, 1979, 74-75].

Therefore one of the main insecurities in the application of monetary policies is the lag between the precise need for action and the effect of that action to the ultimate goals of monetary or macroeconomic policy.

The time frame for political action depends on the fundamental stability or instability of the economy. If the inner strength of a stable economy is powerful, short term behavioral changes in the economy can swiftly immobilize autonomous powers to restore balance (McCarty, 1982, 436). This is the reason why the economy can absorb initial changes without any political changes, while political interventions would even be harmful at this point. In case the inner strengths for stability are weak, the required time for autonomous adjustment will be unacceptably long. Political interventions must be passed quickly so the process of stabilizing the economy is as short as possible.

When the information is gathered slowly and is incomplete, it is difficult to come to a swift political decision. Short term goal values can be wide-ranged when compared to fundamental economic conditions. A lag of the effects of political actions causes problems in determining clear goals for immediate action [Živković & Kožetinac, 2008, 315-316].

2. COMPONENTS OF LAG

In time lag analysis, we can differentiate between several of its components [Živković, 1993, 76-79]: (1) Implementation lag which represents the time interval between the moment when monetary actions need to be applied and the moment when the action is implemented and (2) operation lag which is the time interval between the moment of implementation of the instruments of monetary policies and the moment when their effects become apparent on the ultimate goals.

Other such divisions can also be found in economic literature. Inside lag encompasses the following:

- Recognition lag – the time span between the moment when the need for implementation of monetary policies arises and the moment when the monetary authorities act [Živković & Kožetinac, 2008, 318]. For instance, before any action is undertaken, the existence of any problem has to be identified. The identification of a problem implies gathering and analyzing economic data. Data concerning unemployment and inflation is usually available for the previous month. Let us presume that the unemployment rate for January is available in February. The GDP data is collected quarterly and is accompanied by a long delay. The GDP data for January, February and March, for instance, are available in April or even May. When the data is collected, it is necessary to analyze it and determine with certainty the outcomes of the problems. This analysis often requires data gathered over several months to establish the trends and eliminate temporary statistical digressions.

- Action lag – is the lag between the moment when the monetary authorities act and the moment when the banking system faces altered conditions.

- Outside lag – is the time span between the moment when the banking system functions under altered conditions and the moment when the companies and households, i.e. the non-banking sector, face altered monetary mass and credits [Živković & Kožetinac, 2008, 318].

If we presume that the time span between the moment when the need for implementation of monetary action arose and the actual implementation of monetary measures is close to zero, then the inside lag is the result of recognition lag [Kraken & Slow, 1963, 3-7].

Wrighstman (1971) differentiates the following forms of lag of effects of monetary politics: (a) first time lag between the need for action and the implementation of monetary action; (b) second time lag between the implementation of monetary action and the effects of the actions to changes in the financial sphere (altered interest rates, monetary mass and other financial variables), or within the real sphere (changes in actual income, production, etc.).

Crockett (1979) believes it is important to differentiate the various kinds of lag based on the moment when the political changes become necessary and the moment when political actions give results. He distinguishes: (a) information lag which is the time needed for gathering the information and presenting it to the policy makers, and which point to future changes in the political course; (b) implementation lag is the time needed to formulate the appropriate policies in terms of the new conditions, and which lasts from the moment the policy makers become aware of the changed circumstances; (c) instrument response lag appears when the immediate variables do not respond to the changes in the instrument variables; and (d) reaction lag representing the lag which includes the reaction of target variables to the change of immediate variables.

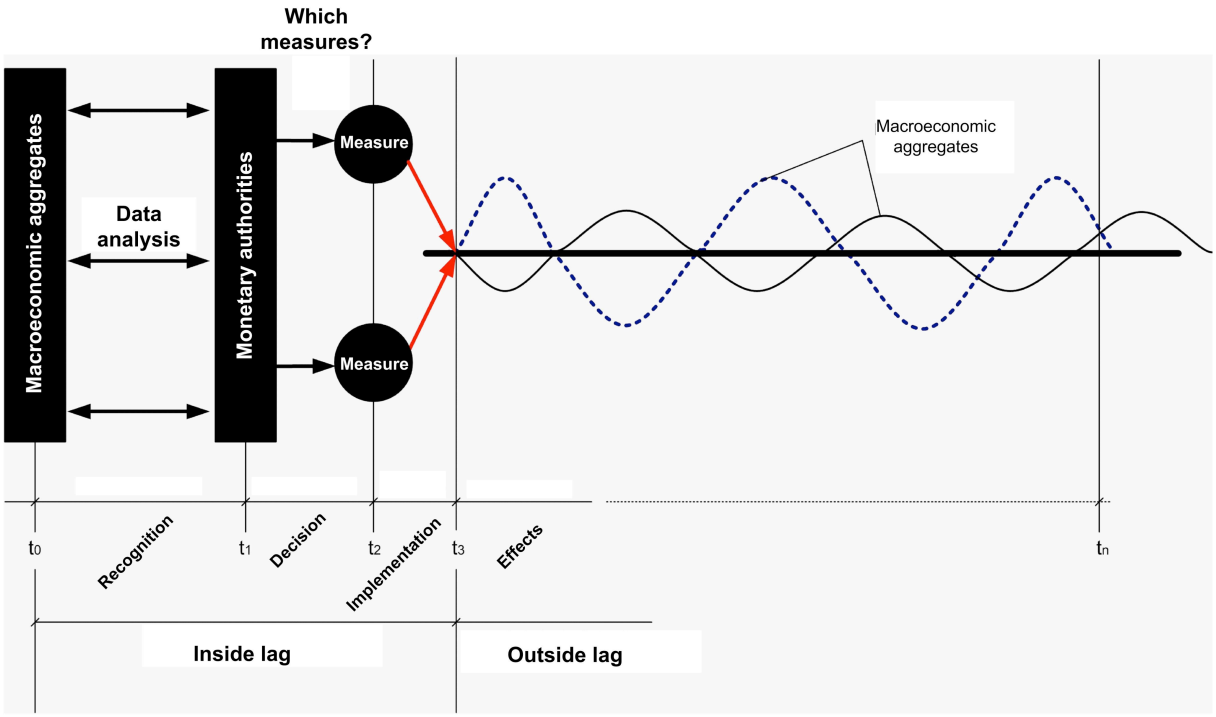
Struthers and Speight (1986) came to the following conclusions in their research: the total lag consists of several distinct sections, which can be classified within inside and outside lag. The inside lag can be further split into recognition lag (the time needed to decide which action to implement). The largest portion of this lag is the information lag which arises from the following reasons: (a) information which the economic policy and especially monetary policy makers require to make an accurate picture of the state of the economy, is statistical in nature and needs time (at least a month, and in some cases a quarter or longer) to be gathered and processed. From this it can be concluded that the policy makers always deal with outdated and incomplete information, which can have serious repercussions, depending on whether the situation is changing and how fast; (b) all the variables which are mutually conditioned on the same change tracks cannot be encompassed statistically; (c) some of the short term changes in track movements of a certain variable can statistically point to a new trend in the movement of the observed phenomenon, even though there is a reversible deviation from that trend.

The next inside lag is the implementation lag or administrative lag. It consists of decision lag which is the time needed to establish what is to be done, and action lag which represents the time needed for the decided action to be taken.

Outside lag is the period between the application of political measures and the effects of these changes on the ultimate outcomes of economic policy. Different financial lags should also be mentioned here: interest rate lag for changes in the amount of money; time for the banks and other financial institutions to adjust their portfolios; transactor's reaction lag dealing with money or exacting credits caused by changes in the interest rate; response lag of the new securities on financial markets to changes of interest rates and the general economic climate.

Pierce (1984) states three causes of lag in political actions, and these are: (1) recognition lag; (2) implementation lag; and (3) response lag of the economy to political changes. These three types of lag determine the total duration between the need for a change in the political course and the ultimate effects on the national economy, with the duration of the lag significantly decreasing the ability of monetary politics to stabilize the economy in a short time [Živković & Kožetinac, 2008, 318-320].

Diagram 1. Time lag of the measures of monetary politics

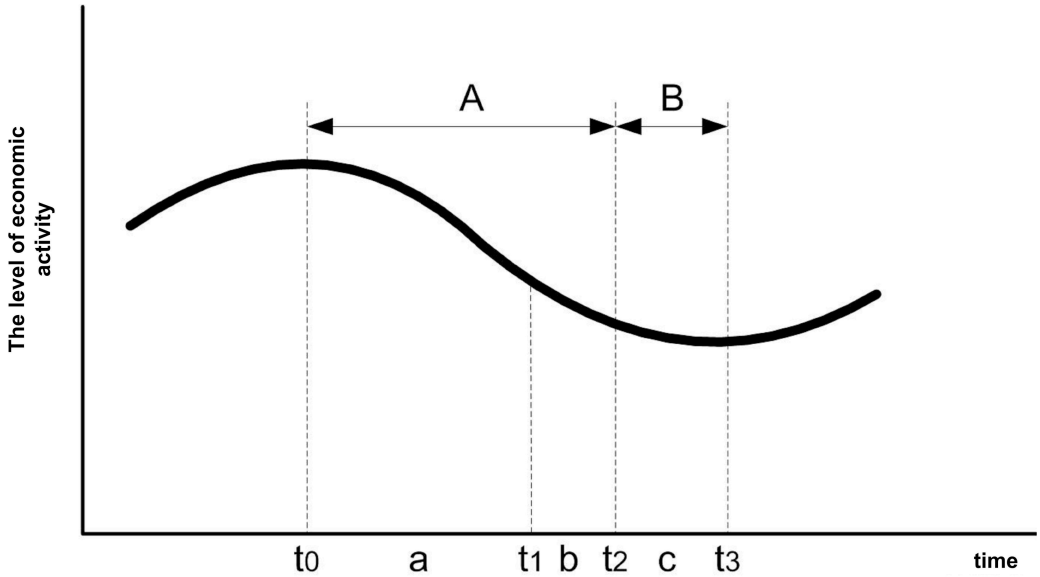


Source: The authors

Empirical evidence in developed market economies shows that the effects of fiscal policy are felt sooner than the effects of monetary policy. Fiscal policy directly influences aggregate demand and income. Monetary policy influences aggregate demand and financial income indirectly, through changes in interest rates and wealth. Fiscal policy is also changed relatively rarely, but its changes seem permanent to the economy subjects. Monetary policy is changed more frequently and the economy transactors find it hard to differentiate between transitory and permanent changes. In addition, the measures of fiscal policies affect aggregate demand sooner, where the shorter effect lag of fiscal policy is not the result of the longer implementation lag of the measures of fiscal politics.

Fiscal and monetary politics are similar in the way of recognition lag; however, while monetary politics have a shorter implementation lag, fiscal policy has a shorter effect lag. Coming to a decision on the plan of monetary politics is faster than that is the case with fiscal politics, which has a longer coordination lag and takes longer to pass the necessary laws. Due to these characteristics, monetary politics is the key instrument of macroeconomic politics in implementation of economic stability [Živković & Kožetinac, 2008, 321-322].

Diagram 2. Time lag of monetary politics and the cyclic movement of economic activity



NOTICE: **A = Inside lag** **B = Outside lag**

a) Recognition lag
b) Administration lag
c) Operation lag

Source: [Poindexter & Jones, 1980, 496-501]

Three phases of lag can be seen on the curve of the business cycle shown in diagram 2. [Poindexter & Jones, 1980, 496-501], and these are the following:

- Recognition lag (t0 – t1) can amount to several months before the stabilization politics bearers become aware of the need for action;
- Administration lag (t1 – t2) shows the interval between recognizing the need for political action and the moment in time when action is taken;
- Operation lag (t2 – t3) is the time interval during which the effects of implemented political measures become evident.

What needs to be stressed is that the main effect on economic activity becomes apparent during time interval t3. It is because of this that the outside lag is more important than the inside lag, because it shows the postponed effects of implemented monetary policies. Empirical research does not provide solid evidence for the length of this kind of lag and ranges anywhere between a month and up to two or three years, when the actions of monetary politics begin to significantly affect the economy [Živković & Kožetinac, 2008, 322].

3. MEASURING THE LENGTH OF MONETARY POLICY MEASURES LAG

The problem of lag is much more complex than it would seem based on the aforementioned descriptions of this economic phenomenon. The process of lag is manifested in the following way: the reaction to some of the monetary variables is spread during the period of implementation and the effects of political action can accumulate and reach a peak of maximum value sooner or later during the process; they can manifest in the form of amplitudes with several different values.

The lag in monetary politics comes from two different kinds of problem: the first, if the lag lasts very long – for two or three years – expansive monetary politics can have consequences, maybe even the main consequences to the economic trends in the time when the situation is completely changed and when the restrictions of demand and deceleration of monetary growth become necessary, and vice versa. The second problem is if the lag is unstable and constant, the inclusion of the time factor can become overly difficult, and the effects of that action are slightly better than simple guessing [Struthers & Speight, 1980, 311-313].

Establishing the needed time span to achieve the greatest or complete effect, or producing a given percentage of the complete effect should be included in the terms of measuring the lag of the manifestation of certain monetary variables or measurements.

Postponing the reactions can vary from one immediate variable to another, from one ultimate goal to another. This is the case, for instance, when the short term interest rates are more flexible than long term interest rates, and banking rates being more flexible than interbanking rates. The response lag of some transactors varies in length depending on different activities. Finally, the length of lag varies with respect to the way of formulating the monetary policies and the choice of monetary variables. This is the case in the USA, where the length of the lag depends on the choice of operative monetary variables. While the total bank reserves, the primary money of monetary mass have short term lag of four to five quarters, the unborrowed reserves manifest fully only after a lag of two and a half years [Struthers & Speight, 1980, 312].

The most important question connected to the implementation of monetary policies is the length of lag. Even though a great deal of empirical research on this subject has been carried out, it is extremely difficult to precisely determine the length of the inside and outside lag above all because that length is variable.

For monetary politics, the insides lag amounts to an average of three to six months (three to four months for recognition and a month to two for action). While the inside lag is longer and highly variable for fiscal policies, the outside lag for monetary policies amounts to anywhere between twelve to eighteen months, and only a few months for fiscal policy. [Willes, 1968, 67-73].

Table 1. The estimated ranking of average time lags for monetary and fiscal policies (in months)

Policy	Inside lag		Outside lag	TOTAL lag
	Recognition lag	Action lag		
Monetary policy	3	0	1 – 20	4 – 23
Fiscal policy	3	1 – 15	1 – 3	5 – 21

Source: [Willes, 1968]

4. EMPIRICAL EVIDENCE OF TIME LAG

Although the concept of lag of monetary policies has deep historical roots in the literature of monetary economy, there are only a small number of studies which are focused on studying this economic phenomenon. In their earlier studies, Cagan and Gandolfi (1969) used a time sample of monetary effects on the interest rates as an indirect measure for the trends of expenses and incomes. Their findings indicate that the effect of lag on the actions of monetary policies on incomes sums up to between six months and two years. Friedman (1972), continuing one of his previous works and using the data from the USA and UK, published empirical evidence of the lag of actions of monetary policies (the amount of money, for instance) and its result in the way of inflation, and by doing so confirming his hypothesis from the past. Tanner (1979) tested the variability of lag between the actions of monetary policies and the results in the change of production. His results showed that the length of lag is highly variable. Duduay (1994), using the data from Canada, estimated that the lag of monetary policies with respect to production amounts to twelve to eighteen months and that the lag connected to inflation is from eighteen to twenty-four months. Batini and Nelson (2001) confirm Friedman's (1972) empirical results. Applying the data from the USA and UK from the period from 1953 to 2001, they established that the length of lag of the effects of measures of monetary politics sums up to twenty-five months for the USA and thirteen months for the UK. They also estimated the length of lag by utilizing various samples from this period and they found little proof that the length of lag shortened in the recent years, which led them to the contrary conclusion that the length of lag became longer in the recent period. Hafer et al. (2007) re-examined the role of money and established a statistically strong relationship between lag in the change of the amount of money and the production gap.

From this empirical evidence, the phenomenon of lag in monetary policy becomes obvious, both in the past and the present and remains one of the key factors influencing the management of monetary policy [Nishiyama, 2009, 2].

Due to the variability of the calculated time lag, the importance of the question of efficiency of operations of macroeconomy is growing, and especially in its concept of monetary policy. In the conditions where the length of lag and/or the length susceptible to change, i.e. the variable size, cannot be precisely determined, it is increasingly difficult to manage an efficient stabilization policy. The time needed for monetary policy to influence the key macroeconomic variables is of crucial importance for its benefits as an instrument of macroeconomic stabilization. If the lag of monetary policy is simultaneous and of long and variable duration, the discretional politics can act towards destabilization. [Cooper et al., 1983, 390-391].

In case the total lag lasts longer than the validity period of reliable economic prognosis of target variables, the fulfillment of political goals can have a destabilizing effect. Similarly, the high variability of lag limits the range of monetary policy [Uselton, 1974, 11].

Efficient monetary policy or stabilization policy must be formulated based on reliable economic predictions. The key importance is given to recognition lag rather than prognosis. If the time horizon of reliable economic prognosis is longer than the total lag of the effects of monetary politics, then the only thing that can jeopardize the results of the preferred monetary policy is the unpredictability of the lag [Živković & Kožetinac, 2008, 327].

CONCLUSION

The phenomenon of lag in the implementation of measures of monetary politics is a controversial and cumbersome problem of monetary politics because the stabilizing or even destabilizing effect of these measures depends on the length of the lag. Research has shown

that if the average lag is extremely long and highly variable at the same time, monetary politics can destabilize the situation. Monetary policies must respond eventually, because it takes time for them to become active and begin to affect the economy trends. For successful monetary policy management, it is not enough only to know the length of the lag but also the lasting of the effects of the undertaken monetary actions. In the case of monetary policies, it is essential to possess the right prognosis beforehand to ensure the precise timing and direction of monetary actions. However, when the lag varies, it is difficult to reach a decision in the right direction of the actions, which could lead to contributing to the instability of the economy, if the decisions were made using incomplete and improper information. When discussing this problem, Friedman stated at one point that the empirical evidence persuaded him that it is much more important to prevent monetary changes to contribute to the instability, than it is their exact use to neutralize the other forces at work.

Due to the existence of time lag in the implementation of monetary policy it is not realistic to expect the monetary policies to generate immediate on-line effects. Time needs to pass until the economic transactors adjust to the newly created circumstances and financial conditions, or until the real expenses adjust to the existing amount of money in the monetary system. This must be taken into account while planning and implementing monetary policies, because their efficiency will be greater in the measures which are capable of predicting the future trends ex ante, and with high likelihood, so that the activated instruments of monetary politics in the previous period are appropriate for the upcoming conditions.

LITERATURE

1. Batini, N., Nelson, E. (2001) "The lag from monetary policy actions to inflation: Friedman revisited", *International Finance*, 4(3): 381 - 400.
2. Cagan, P., Gandolfi, A. (1969) "The Lag in Monetary Policy As Implied by the Time Pattern of Monetary Effects on Interest Rates", *American Economic Review,* 59: 277 - 284.
3. Cooper et al. (1983) *Money, the Financial System and Economic Policy*, Addison-Weseley Publishing Company.
4. Crockett, A. (1979) *Money – Theory, Policy and Institutions*, Second Edition, Nelson.
5. Duguay, P. (1994) "Empirical Evidence on the Strength of the Monetary Transmission Mechanism in Canada - An aggregate approach.", *Journal of Monetary Economics,* 33: 39 - 61.
6. Friedman, M. (1972) "Have monetary policy failed?" American *Economic Review,* 62: 11 - 18.
7. Friedman, M. (1973) *Teorija novca i monetarna politika*, (Beograd: Rad).
8. Hafer, R.W., Haslag, J.H., Jones, G. (2007) "On money and output: Is money redundant?" *Journal of Monetary Economics,* 54: 945 - 954.
9. Kraken, I., Slow, R., (1963) *Summary of Part I, Lags in Monetary Policy, Stabilization on Policies Commision on Money and Credit* (Prentice-Hall).
10. McCarty, M., H. (1982) *Money and Banking – Financial Institutions and Economic Policy,* (Addison - Wesley Publishing Company).
11. Nishiyama, S. (2009) *Monetary Policy Lag, Zero Lower Bound and Inflation Targeting,* (Bank of Canada).
12. Pierce, J., L. (1984) *Monetary and Financial Economics*, (John Wiley and Sons).
13. Poindexter, J., C. & Jones, C., P. (1980) *Money, Financial Markets and the Economy*, (West Publishing Company).

14. Struthers, J. & Speight, H., (1980) *Money, Institutions, Theory and Policy*, (Longman).
15. Tanner, J.E., (1979) "Are the lags in the effects of monetary policy variable?" *Journal of Monetary Economics,* 5: 105 - 121.
16. Uselton, G., C. (1974) *Lags in the Effects of Monetary Policy*, (Marcel Dekker).
17. Willes, M., A. (1968) "Lags in Monetary and Fiscal Policy", *FRB of Philadelphia Business Review,*: 67 – 73.
18. Wrightsman, D. (1971) *An Introduction to Monetary Theory and Policy*, (New York: The Free Press).
19. Živković, A., & Kožetinac, G. (2008) *Monetarna ekonomija*, (Beograd: Centar za izdavačku delatnost Ekonomskog fakulteta).
20. Živković, A., (1993) *Analiza efikasnosti monetarne politike u Jugoslaviji*, (Beograd: Ekonomski fakultet).

FENOMEN VREMENSKOG POMAKA („LAG") U PRIMJENI MJERA MONETARNE POLITIKE

SAŽETAK

Ovaj rad se bavi teorijskim aspektom fenomena vremenskog kašnjenja u djelovanju mjera monetarne politike. Monetarna i fiskalna politika suočavaju se s fenomenom kašnjenja. Jedno od kontroverznih i opterećujućih pitanja monetarne politike jest priroda i duljina kašnjenja između izvršenja monetarnih mjera i efekata na makroekonomske agregate. Dok monetarna strategija ukazuje na nekoliko mogućih razloga kašnjenja ne postoji opća suglasnost u vezi duljine kašnjenja. Proučavanje fenomena kašnjenja nameće dva pitanja: šta implicira kašnjenje u odvijanju monetarnih procesa i monetarne politike i zbog čega u uvjetima neravnoteže vlasnici aktiva nisu sposobni prilagoditi svoje portfolije neposredno? U ekonomskoj literaturi poznate su dvije kategorije kašnjenja: unutarnje kašnjenje (koje obuhvaća prepoznavanje problema i provedbu mjera) i vanjsko kašnjenje (koje obuhvaća reakciju makroekonomskih agregata na provedene mjere monetarne politike). U radu su deskriptivno navedene različite identifikacije vremenskog kašnjenja a dati su i shematski prikazi djelovanja promatranog fenomena.

***Ključne riječi**: monetarna politika, mjere monetarne politike, vremensko kašnjenje, problemi kašnjenja, posljedice*

Lovre Božina, PhD Full professor
Juraj Dobrila University in Pula
Department of economics and tourism "Dr. Mijo Mirković" Pula

BOOK REVIEW

Monetarna ekonomija

Author: Branko Matić, PhD Full professor at the Faculty of Economics in Osijek
Publisher: Josip Juraj Strossmayer University in Osijek
Faculty of economics in Osijeku

Monetarna ekonomija (engl. Monetary economics) is a scientific and academic book that extensively explains all issues in the field of monetary policy starting from its heritage, i.e. the development and role of money, up to its modern trends. It is well known that institutions dealing with money issuing are closely bounded with social conditions, therefore in this book the author integrates in a very natural and successful way such a phenomena and gives emphasis on the monetary system in Croatia and the European union.

Relying on his rich experience in both the scientific and educational workfield, the author shapes this work in the way it can be not only an academic textbook intended for students researching monetary issues but also a literature for a broader audience of scientists and specialists interested in learning the way and reasons of the evolution of money together with different practical skills in monetary relations and politics.

Monetarna ekonomija is a multilayered read: it not only reflects a scientifically valuable work but also a work that provides a logical course of development of forms of money and monetary institutions, defined by economic, legal, cultural and social features during the evolution of social relations in which money is used from the earliest human history up today.

The book is written in Croatian and gives the reader an insight into the richness of many concepts that help undergraduate students, graduate and postgraduate specialist as well as doctoral students to enter into the field of monetary economics through different sections as follows: Non-monetary economies, Monetary Economics, The functions of money, Monetary Systems, Monetary theory, Changes in the value of money, Monetary aggregates, Money and the world trade, Money and Banking.

In the first chapter entitled *Non-monetary economies* the author analyzes the flow of trade in order to show the economic advantages carried out by the use of money as the general form of value. The author concludes this section by pointing out that although non-monetary economies do not exist today, disturbances *"[...] (wars, natural disasters, hyperinflation, etc.) and instability in modern economic systems can lead to disruptions in their financial systems. It can be a sporadic and short-term effect that individual economies take the characteristics of a non-monetary economy. In these situations, money can be suppressed partially or fully in the national economy."*

In the second chapter (*Monetary Economics*), the author introduces the term and role of money as an intermediary in the exchange process from different points of view, which is then continued in the third chapter and followed till the ninth chapter.

The chapter starts with the author's discussion on various forms of money that emerged during its evolution, including money as a cultural, economic, social and legal phenomenon. The author gives a clear chronological survey of major determinants that influenced the rise of different forms of money dependent from the social and economic circumstances of the time. Particular attention is paid to technical details of monetary units, technical characteristics of coins and banknotes, which gives the reader a detailed knowledge about the importance of the apparition of money. Money is not only an economic, but also a cultural, phenomenon, which the author highlights to the reader when explaining the technical aspects of the "production" of money. Such an approach permits the author to establish a firm connection between different forms of money, from wrought coins to deposit money.

In the third chapter the author discusses the functions of money. It is worth pointing out that besides the four traditional functions the author explains a fifth function too, i.e. the "Numismatic and notaphily function of money", which was not considered in other works in this field in Croatia. In this chapter the author also mentions the effects of conducting the coin and notaphily policy within the emission policies.

In the fourth chapter the author deals with monetary system, the basis of which is money - its manifestations and functions. The author points out that the development of a monetary system is highly linked with the growth of real spheres thus indicating the high correlation between real and monetary economics.

The author continues with a logical line by explaining in the fifth chapter the theories of money and giving to the reader a theoretical framework for understanding the evolution of money, while pointing out to other authors and references that explain the issue in greater detail.

The four remaining chapters are dedicated to the presentation and the explanation of the relationship between changes in money value and monetary aggregates, the link between money and international trade and the interaction between money and banking.

From the book *Monetarna ekonomija* it can be concluded that the use of money allows and facilitates labour division and specialization, keeping in mind the assumption that specialization fully relies on the availability of trade on the market. Trade is important not only for economists, but also for those who think that money reflects the connections among people that arise during trade. That is why money is interesting from an economic, social and cultural perspective, as well as from a legal, moral, ethical and theological standpoint. Despite the development of its different forms, the role of money in respect of all these approaches did not change with the development of the society. Similarly, modern integration flows do not significantly change its function, but they introduce rules that aim to preserve confidence in various forms of money and its symbols.

Lovre Božina, PhD Full professor