Vol. 4 (2) 2007 • UDK 33 • ISSN 1820-4570

megatrend review

The international review of applied economics

Content

Megatrend University, Belgrade

Megatrend Review

The international review of applied economics

Vol. 4 (2) 2007

Megatrend University, Belgrade

Megatrend Review

The international review of applied economics

Vol. 4 (2) 2007

Published by:
Megatrend University

ISSN 1820-4570

UDK 33

The review is published twice a year in Serbian and twice a year in English language.

All papers have been reviewed.

Adress:
Megatrend Review
Obilićev venac 12, 11000 Belgrade,
Serbia & Montenegro
Tel: +381 11 32 86 730, ext. 931;
Fax: +381 11 32 86 747
e-mail: imilutinovic@megatrend.edu.yu
imilutinovic@megatrend-edu.net

For publisher:
Nevenka Trifunović

Publishing director:
Dragan Karanović

CONTENTS

Statistics and econometrics

Economy of regions

Original scientific paper
Received: September 9th, 2007
UDK 005.52:330.43

Professor Branko Dragović, PhD
Institute of Physics, Belgrade, Serbia
Professor Dušan Joksimović, PhD
Megatrend University, Belgrade, Serbia

ON POSSIBLE USES OF p-ADIC ANALYSIS IN ECONOMETRICS

Abstract: *This paper considers the possible uses of p-adic analysis in economics. There is a short overview of p-adic numbers, p-adic analysis and the possibilities of their usage in modern econometrical methods is implied. The paper aims to initiate systematic research of p-adics usage within economic sciences.*

Key words: *econometrics, p-adic numbers, p-adic analysis, p-adics usage*

1. Introduction

Econometrics researches based on the methods of classical mathematics, with the particular overview of probability and statistics, are in the last couple of decades very frequent, and naturally very usable. These researches have certain limitations which are the consequences of using only "ordinary" defined Euclidean metric, which of course in our space-time environment is well justified. It was natural to expect the assumed usability of this metrics in the space of states of various econometrical variables, i.e. that classical mathematics represents the best way to describe metrical characteristics amongst different states, both in time-wise invariable and time-wise different moments. Is it so?

In the last two decades, there is very interesting development and application in mathematical physics of one, one can say, brand new mathematics, in which the metric between states (points, numerals,…) is not determined using classical Euclidean metrics, but through the so called *p*-adic metrics. Together with physics this *p*-adic mathematics is gradually finding its use in some other sciences, particularly biology. Generally speaking, practically in all complex systems whose state space has hierarchical structure, *p*-adics is used as a natural method of describing the properties of these systems. In relation to usability of *p*-adics it must be emphasized that the results of all measure-

ments are rational numbers, and they belong to both set of real numbers and all sets of *p*-adic numbers.

Our opinion, as well as the opinion of a certain number of colleagues dealing with *p*-adics in the world, is that economics is the science in which *p*-adic structures will shine in their full glory and significantly improve some of the existing models, and probably define new models with better econometric performances.

This paper is an attempt to, at least in some econometric segments (analysis of time series, regression and correlation analysis, dynamic analysis of the securities price formation, behaviour during stock market crash), slightly open the doors of *p*-adic mathematics usage in economics. We hope that this paper will serve as an incentive to the authors themselves as well as other researchers to focus their interests towards the usage of *p*-adics in economics.

For reasons of presentation clarity we will first give an overview of basic terms related to *p*-adic numbers and their functions, and then we will show the possibilities of their application in current econometrical trends.

2. *p*-Adic numbers

Before introducing *p*-adic numbers it is useful to remind ourselves what different kinds of numbers exist, that people use in everyday life, as well as in solving various problems in applied mathematics. The first type of numbers that people come across is 1, 2, 3…, i.e. the numbers that we use to count objects in our environment. All numbers of this type are called the set of natural numbers, symbolised by N. In order for someone to be able to solve simple problems represented by the equation $x+a=b$, where a and b are any given natural numbers while x is an unknown number, it is necessary to expand the above set by adding a 0 to it, and the same natural numbers but with a negative sign, i.e. with the opposite meaning to regular natural numbers. This set of numbers is denoted by Z. Frequently it is necessary to use not only integer numbers but their fractions as well. So in order to solve e.g. simple equation $ax=b$, where a is a natural number and b some given integer number, we need to introduce the so called rational numbers denoted by $\frac{a}{b}$. The set of all these numbers is symbolised by Q. If we were to replace x with x^2 in the above equations new types of numbers must be introduced: irrational, imaginary and complex. The example of an irrational number follows from the equation $x^2=2$, and imaginary from $x^2=-1$. The set of rational and irrational numbers is called a set of real numbers R. Every algebraic equation $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$ has its solution in the set of complex numbers C which can be represented in the form $x=a+ib$, where a and b are real numbers while i is the so called imaginary unit, $i=\sqrt{-1}$. Listed sets

satisfy the relation $N \subset Z \subset Q \subset R \subset C$. One can state that the main role in this series of sets is played by the real number set. Every real number can be represented in the form of infinite sum

$$x = 10^m (x_0 + x_1 10^{-1} + x_2 10^{-2} + \cdots), \quad m \in Z, \tag{2.1}$$

where x_i are digits $0,1,\cdots,9$. We can notice that for every irrational number there is an infinite expansion (2.1) without repeating a set of numbers different from zero, and that for a convergence of this infinite sum a notion of metric is important.

Between any two numbers inside of any of the sets $N,\cdots,C$ there is a properly defined standard metric with the absolute value function $d(x,y) = |x-y|$. This metric satisfies the property $d(x,y) \le d(x,z) + d(z,y)$, where x, y, z are three arbitrary points. There are quaternions and octonions as complex number extensions, but they are not of interest for our further consideration.

Is there another metric between natural, integer and rational numbers, however unnatural it may look to us on first glance? The answer is positive: In relation to every prime number p $(p = 2,3,5,7,11,\cdots)$ a metric can be defined. With regard to that lets introduce the notion of p-adic absolute value for an integer number m in relation to prime number $p : |m|_p = p^{-k}$, where k is a degree of divisibility of number m by p. For example, let's say that $m = 360$. Then $|360|_2 = 1/8, |360|_3 = 1/9, |360|_5 = 1/5$ and $|360|_p = 1$ for $p \ge 7$, because $360 = 2^3 \cdot 3^2 \cdot 5$. One can easily conclude that $|m|_p \le 1$ for an arbitrary integer number m and any prime number p. p-Adic absolute value of some rational number m/n is $|m/n|_p = p^{-k}$, where now k is an integer equal to the difference of the degree of divisibility of m and n with p. p-Adic metric is introduced analogously to ordinary, but using p-adic absolute value instead of ordinary, i.e. $d_p(x,y) = |x-y|_p$. For this metric the following applies

$$d_p(x,y) \le \max\{d_p(x,z), d_p(z,y)\} \le d_p(x,z) + d_p(z,y). \tag{2.2}$$

Because of the first inequality in (2.2), this metric is significantly different from the standard one and it is called non-Archimedean or ultrametric distance.

In relation to p-adic metric the infinite sums of the form

$$x = p^{\nu} (x_0 + x_1 p + x_2 p^2 + \cdots) = p^{\nu} \sum_{i=0}^{+\infty} x_i p^i, \quad \nu \in Z, \tag{2.3}$$

where $x_i = 0,1,\cdots, p-1$ are digits, converge and uniquely represent p-adic numbers. It is interesting to notice that all p-adic numbers are represented by sums of positive members. Negative whole numbers are represented by a sum of infinite number of positive members. For example,

$$-1 = (p-1) + (p-1)\, p + (p-1) p^2 + \cdots = (p-1) \sum_{i=0}^{+\infty} p^i. \tag{2.4}$$

For every prime number p there is a corresponding infinite set of p-adic numbers Q_p for which the operations of addition, subtraction, division and multiplication are valid, along with the usual associativity and commutation properties, i.e. Q_p, like R, is a number field. Thus, starting from the field of rational numbers and possible non-trivial metrics on it, Q can be expanded in such a way that we get a field of real numbers R and infinite number of p-adic number fields Q_p. It should be said that all these number fields differ amongst themselves and that there are not any other completions of the field Q.

From the mathematical viewpoint p-adic numbers are also natural as real numbers; they are even simpler then real numbers. However, as we live in a world best described in real and not p-adic numbers, real numbers appear more natural to us.

German mathematician Kurt Hensel first introduced p-adic numbers in mathemtacis in 1897 and he used them in solving Diophantine equations. p-Adic numbers are employed in various mathematical fields, such as algebraic geometry, theory of representations and number theory. There is also a well developed theory of functions and mathematical analysis with p-adic numbers. For any further details on p-adic numbers and p-adic analysis we recommend the books listed in Bibliography under 1-4.

Since 1987 p-adic analysis has been actively used in the modelling of various physical and other systems (see, for example, monographs and papers listed in the Bibliography under 4-9). This very attractive area of the implementation of p-adic analysis is known as "p-adic mathematical physics". Many relevant topics from this field are covered in information on the third international conference on p-adic mathematical physics (see the conference website p-adic mathphys. 2007 listed in the Bibliography under 10).

Let us mention that there is a rich structure of algebraic extensions of all levels. There is also a p-adic analogue of the set of complex numbers from the real case and it has a structure of an infinite vector space; it is metrically complete and algebraically closed. It is also useful to mention, though it will not be directly used in this paper, that there are so called adeles whereby all real and p-adic numbers are unified in a whole.

3. p-Adic analysis econometry-relevant areas

p-Adic analysis was developing with greatest intensity in the latter half of the 20th century; it has been constantly developing ever since and represents an active field of the contemporary mathematics. There are various interesting aspects of it but we are going to deal here only with what seems interesting from the viewpoint of its applicability in economy.

It is possible to introduce mapping between various fields of p-adic numbers and various analyses can thus be observed and studied. However, from purely mathematical point of view and its possible applications, there are two types of mappings and relevant analyses which are mostly in use. These are: 1) analysis with functions of the type $y = f(x)$, where function $f(x)$ and independent variable x have p-adic values, and 2) analysis with functions of the type $y = \varphi(x)$, where values of the function $\varphi(x)$ are real or complex, whilst the independent variable x has p-adic value.

In the first case $(y = f(x))$ all known functions from the real analysis, which have the form of $y = \sum_{n=0}^{\infty} a_n x^n$ with rational coefficients a_n, can be converted into p-adic functions by means of treating x as a p-adic variable. For example, in such a way the following exponential function can be introduced: exponential function with p-adic values $\exp_p x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, which is not convergent for all the values of the p-adic variable x but only for those values that satisfy the inequality $|x|_p < |2|_p$. Second way of introducing of functions with p-adic values is through the method of interpolation. This method is based on the fact that the set of natural numbers is dense in the Z_p set of p-adic integers. Namely, if the function is given at the set N then there is only one way how it can be interpolated to the whole Z_p domain.

In the second case $(y = \varphi(x))$, when the function has either real or complex values, it is usually considered that functions depend only on several first terms in the expansion of the p-adic number. Such functions retain the same value in the neighbourhoods of each p-adic point and are said to be locally constant. Some more basic examples of such functions are: $|x|_p$, $\exp 2\pi i\{x\}_p$ and $\Omega(|x|_p)$, where $\{x\}_p$ is a fractional part of x. Function $\Omega(|x|_p) = 1$ if $|x|_p \le 1$ and $\Omega(|x|_p) = 0$ when $|x|_p > 1$. The relation $|x| \prod_p |x|_p = 1$ holds for rational numbers $x \neq 0$, and $\exp(-2\pi i x) \prod_p \exp 2\pi i\{x\}_p = 1$ for any rational number x.

We are going to mention here several areas of the p-adic analysis which seem to be suitable candidates for the illustration of its application in econometry. These are: interpolation polynomials, pseudo-constants, pseudo-differential operators and path integrals.

Interpolation polynomials. Due to the influence of various factors, variables that feature in econometry are usually time-dependant in a very complex way. Such time-dependency is not precisely defined and what needs to be done is to create as approximate a function as possible based on a finite number of its past values. Usually the real function $f(t)$, where t is time, is approximated by an appropriate polynomial. If function values of $f(t)$ are known in $m+1$ number of equally distributed time units, then in the p-adic analysis a role of the approximation function is played by the following interpolation polynomial:

$$F_m(t) = \sum_{n=0}^{m} a_n \binom{t}{n}, \quad t \in Z_p, \quad \binom{t}{n} = \frac{t(t-1)\cdots(t-n+1)}{n!}, \qquad (3.1)$$

where is

$$a_n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f(k), \quad n = 0,1,2,\cdots,m \quad . \tag{3.2}$$

This interpolation polynomial (known as Mahler polynomial) (3.1) is the only of all possible polynomials $P_n(t)$ of the degree $n \le m$ for which the following is valid: $P_n(k) = f(k), k = 0,1,2,...,m.$ The second important feature of such $F_m(t)$ polynomial is that it approximates the function given in infinite number of points in the best possible way

$$f(t) = \sum_{n=0}^{\infty} a_n \binom{t}{n} \tag{3.3}$$

in relation to the set of all $P_n(t)$ polynomials of the degree $n \le m.$ It should also be noted that time value t in (3.1) and (3.3) can be any integer number. Thus, based on the time series $f(t)$ for the values $t = 0,1,2,...,m$ the value of the function $f(t)$ is determined in the time $t = m+1$ and later. The function $f(t)$ defined in (3.3) is the only p-adic function on Z_p that is continuous in the case of the input values $f(t)$, n= 0,1,2,... and $\lim_{n\to\infty} a_n = 0$. It is also differentiable in all integer number points if and only if the condition $\lim_{n\to\infty} \frac{a_{k+n}}{n} = 0$ is satisfied for each $k = 0,1,2,\cdots$.

The time series $f(t), t = 0,1,2,...m$ can be modelled by the polynomial $F_m(t)$ if the input values $f(0), f(1),\cdots, f(m)$ are rational numbers, which is a condition practically always fulfilled, and if $|a_n|_p < |a_{n-1}|_p$ is satisfied for at least one prime number p in several latest time points. In this case we have an interpolation polynomial with rational coefficients so that its applicability can be studied not only by means of the real but also p-adic analysis. This statement results from the fact that rational numbers are common for the set of real numbers and all sets of p-adic numbers. More detailed information on p-adic interpolation can be found under items 2 and 3 of the Bibliography.

Pseudo-constants. The basic difference between the analysis of functions $y = f(x)$ of a real argument and p-adic functions of a p-adic argument lies in the existence of p-adic pseudo-constants. Namely, there are such $y = C(x)$ p-adic functions whose first derivative equals zero though they are not constants, that is $C'(x) = 0.$ A special case of pseudo-constants are ordinary constants that are some p-adic numbers. Sum, difference and product of two pseudo-constants is again a pseudo-constant. A simple example of a psedo-constant is the following:

$$C(x) = x_0 + x_1 p^{w_1} + x_2 p^{w_2} + \cdots,$$

where $x_i = 0,1,\cdots, p-1$ are digits of the p-adic independent variable form x (2.3) for $\nu = 0$, whilst exponents satisfy the inequality $1 \le w_1 < w_2 < w_3 < \cdots$ and $\lim_{n\to\infty}(w_n - n) = \infty$. The theory of pseudo-constants illustrated by examples is in the best way described in a book by Mahler (Mahler, Item 3 in the Bibliography).

Pseudo-differential operators. In the case of functions $y = \varphi(x)$, where $\varphi(x)$ has real or complex values for p-adic x, well defined usual derivatives do not exist as the function and the independent variable exist in considerably different spaces. In this case, the role of a derivative is played by a pseudo-differential operator that is defined through an integral. Integrals with real or complex values are well-defined for such functions $y = \varphi(x)$. Vladimirov pseudo-differential operator is most commonly in use. The theory of p-adic pseudo-differential operators is well-described in a monograph by Vladimirov, Volovich and Zelenov (Bibliography, Item 4).

Path integrals. In order to describe p-adic Brown motion and quantum-mechanical processes one can use path integral formalism. (for details see Items 4 and 7 in the Bibliography).

4. Contemporary econometric trends and ways of applying p-adic analysis in them

4.1 Possibilities for the application of p-adics in the analysis of econometric time series

First, we are going to state some contemporary methods of predicting econometric time series. A time series is a sequence of data with a certain chronological ordering. It can be also understood as a realisation of a random process.

Namely, let $\{X_t, t \in R\}$ be a given random process whose realisation in certain moments is:

$$t = 1,2,\ldots,n \qquad X_1, X_2, \ldots X_n$$
$$t = 1,2,\ldots \qquad X_1, X_2, \ldots$$
$$t = 0,\pm1,\pm2,\ldots \quad \ldots X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots \quad .$$

Each of these strings is a time series. So in this sense, a time series is a string (finite or infinite) of random variables.

One of the most important tasks when studying time series is to predict their future values, both in the near and distant future. Due to a widespread and always topical need and call for the answer to such a question different methods have been developed which, depending on the nature of a time series, are all trying to produce the best possible answer to the following question: what is the value of a time series in a certain moment in the future

Some of the most modern approaches in the sphere of predicting time series in the future are:

a) If a time series is stationary, that is, if

$$E(X_t) = const, \qquad t = 1,2,3,...$$
$$\sigma^2(X_t) = const \qquad t = 1,2,3,...$$
$$Cov(X_t, X_{t-k}) = f(k) \qquad t = 1,2,3,...$$

then the time series analysis, and its prediction in the future, is done by either auto-regressive model of the type $\boldsymbol{p}$ (AR(p)), or the model of moving average of the type $\boldsymbol{q}$ (MA(q)), or, alternatively, by a combined auto-regressive model of the type $\boldsymbol{p}$ with moving average of the type $\boldsymbol{q}$ (ARMA (p,q)).

The choice of a model, as well as the choice of coefficients within a model, is done by a well-known Box-Jenkins methodology that is combined with the method of the mean-square error minimising, in relation to the realised values of a time series in the past. (Ways of applying of these methods can be found under 11-13 in the Bibliography).

b) if a time series is non-stationary but in such a way that its $\boldsymbol{d}$-th difference is stationary, then the prediction of such a time series is done through the ARIMA (p,d,q) model, where the choice of coefficients is also performed by means of the well-known Box-Jenkins methodology combined with the method of the mean-square error minimising, in relation to the realised values of a time series in the past. (Ways of applying of these methods can be found under 11-13 in the Bibliography).

All the above-listed models are so called linear models.

The most famous modern non-linear models are ARCH (p) and GARCH (p,q) models in which the selection of coefficients is done through the maximum likelihood model (Ways of applying of these methods can be found under 14 and 15 in the Bibliography).

One of the possibilities of the application of p -adics in the time series analysis, and consequently in the prediction of time series in the future, can be obtained by analogy with the so called exponential smoothing that is very often used for the prediction of a non-stationary time series in the near future.

If a time series $\{X_t, t = 1,2,...,n.\}$ is given, then its predicted values are reached by means of employing the exponential smoothing theory, and through the following recurrent formulas:

$$\hat{X}_1 = X_1$$
$$\hat{X}_{i+1} = \alpha \cdot X_i + (1-\alpha) \cdot \hat{X}_i \qquad 0 < \alpha < 1 \qquad i = 1,2,...,n.$$

It is obvious that in case of $i = n$ we get a predicted value of a time series in the near future. Also, one can easily show that it is as follows:

$$\hat{X}_{n+1} = \sum_{j=0}^{n-2} \alpha \cdot (1-\alpha)^j \cdot X_{n-j} + (1-\alpha)^{n-1} \cdot X_1 = \tag{4.1}$$
$$= \alpha \cdot X_n + \cdot\alpha \cdot (1-\alpha) \cdot X_{n-1} + \alpha \cdot (1-\alpha)^2 \cdot X_{n-2} +$$
$$+ \alpha \cdot (1-\alpha)^3 \cdot X_{n-3} + \ldots + \alpha \cdot (1-\alpha)^{n-2} \cdot X_2 + (1-\alpha)^{n-1} \cdot X_1.$$

Analysis of the equation (4.1) indicates the fact that the influence of the past time series values on its near future value decreases with the *(1-α)* factor, which is altogether to be expected in the case of fast-changing time series.

The implementation of p-adics in the time series prediction could be referred to as p-adic smoothing and it consists of the application of the interpolation polynomial as described in the previous section.

4.2. Possibilities for the application of p-adics in the regressive-correlation analysis

Classic type of the simple correlation analysis has been dominant for a number of years in the attempts to establish the dependency rules between a dependent variable (***Y***) and an explanatory variable (***X***), based on the ***n*** number of sample pairs (x_i, y_i). (Bibilography13).

General form of the simple regressive curve is: $Y_i = f(X_i) + \varepsilon_i$, where ε_i is a stochastic term whereby deviation of the real value of the dependent variable from its value determined by the regressive curve $f(X_i)$ is determined.

Basic assumptions concerning the possibilities of the application of the classic correlation analysis refer to the stochastic term ε_i. These are:

1) Mean value of the stochastic error
   1. ε_i is zero, 1. $E(\varepsilon_i) = 0$
2) Stochastic errors are homoscedastic
   $(\forall i) \quad \sigma^2(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2 = const$
3) Stochastic errors are not inter-correlated
   $(\forall i, j)(i \neq j) \quad Cov(\varepsilon_i, \varepsilon_j) = 0$
4) Stochastic error has normal distribution,
   $\varepsilon_i \prec N(0, \sigma^2)$.

In those cases when $f(X_i)$ is a polynomial of the **m**-th order, the abovementioned assumptions require that polynomial coefficients have to be determined by minimum of the mean square error method in relation to the sample pairs (x_i, y_i), *i=1,2,...,n.*

p-Adic regressive-correlation analysis would then consist of introducing instead of the $f(X_i)$ function a p-adic interpolation polynomial and also treating the stochastic term ε_i as a stochastic pseudo-constant.

4.3. Possibilities for the application of p-adics in the dynamic analysis of securities' prices

Contemporary approach of the dynamic securities' prices analysis is such that it determines the price of securities in every moment in time, that is, it defines the function of the value of securities as a continuous time function. One of the most common models describing the rule of the securities' price fluctuations in time is a generalised Ito model (Bibliography 16 and 17) according to which the following applies to the securities' price marked as X:

$$dX = a(X,t)\cdot dt + b(X,t)\cdot dW ,$$

where $a(X,t)\cdot dt$ describes the deterministic part of the change, that is the fluctuation trend X (the form of dependency is determined by the function $a(X,t)$), whilst $b(X,t)\cdot dW$ describes the stochastic part of the change which defines any random variations of the value X, where W is a well-known Wiener process.

In those cases when Wiener process is presented as Brownian geometric motion, and taking into consideration the Paul Levy theory according to which every Wiener process is at the same time and roughly speaking standard Brownian motion, we obtain the following relation:

$dX = a\cdot X\cdot dt + b\cdot X\cdot dW$ where is $dW = \varepsilon\cdot\sqrt{dt}$, and ε is an stochastic error distributed by the rule of $\varepsilon \sim N(0,1)$.

Coefficient a represents a rate of change ($a\cdot X$ is a drift coefficient), whilst the change variance is $b^2\cdot X^2$ ($b\cdot X$ is diffusion coefficient). (This is directly applicable in the cases of non-stationary time series described as random walk where $X_{t+1} = X_t + \varepsilon_{t+1}$). Graphic presentation is given in Figure 1.

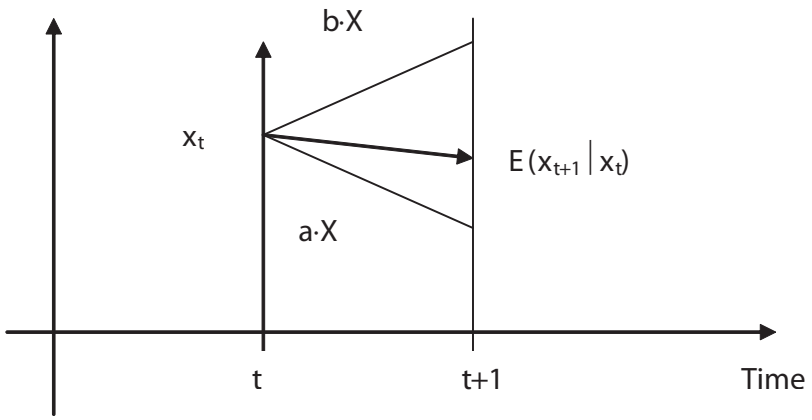


Figure 1. *Drift and diffusion*

In this case the application of p-adics on the process of Brownian motion is performed through a pseudo-differential operator.

4.4. Possibilities for the application of p-adics in critical moments of drastic fluctuations in securities' prices (stock market crash)

Attempts at predicting major instabilities on the securities market, as well as the behaviour of securities' prices during such periods of instability, represent so far insufficiently explored parts of the financial analyses. Some papers written on the subject indicate a huge level of the hierarchal structure in dynamic models that endeavour to describe such processes. (Bibliography, Item 18).

One way of p-adic description of such a process was the subject of the work of Bikulov, Zubarev and Kajdalova (Bibliography 19). They used p-adic analysis with the Vladimirov pseudo-differential operator to show that during longer periods of time preceding major stock market crashes, price logarithms behave like power low multiplied by a sum of logoperiodic harmonics.

5. Conclusion

This is our first paper dealing with the application of p-adics in econometry. In this paper we just indicated possible directions of the implementation of p-adics in econometry (Bibliography 20). In our future papers we plan to elaborate further this p-adic approach with more focus on some specific models.

References

[1] Gouvea, F.Q.: *p-Adic Numbers: An Introduction*, (Universitext), Springer, Berlin, 1993.
[2] Schikhof, W.: *Ultrametric Calculus*, Cambridge University Press, Cambridge, 1984.
[3] Mahler, K.: *Introduction to p-Adic Numbers and Their Functions*, Cambridge University Press, Cambridge, 1973.
[4] Vladimirov, V. S. – Volovich, I. V. – Zelenov, E. I.: *p-Adic Analysis and Mathematical Physics*, World Scientific, Singapore, 1994.
[5] Brekke, L. – Freund, P. G. O.: *p-Adic Numbers in Physics,* Physics Reports 233, 1993, pages 1-6.
[6] Khrennikov, A.: *Non-Archimedean Analysis: Quantum Paradoxes, Dynamical Systems and Biological Models,* Kluwer Academic Publishers, Dordrecht, 1997.

[7] Dragovich, B.: *p-Adic and Adelic Quantum Mechanics*, Proceedings of V. A. Steklov Inst. Math., 245, 2004, pages 72-85; arXiv:hep-th/0312046.
[8] Dragovich, B.: *p-Adic and Adelic Cosmology: p-Adic Origin of Dark Energy And Dark Matter,* p-Adic Mathematical Physics, AIP Conference Proceedings *826*, 2006, pages 25 - 42; hep - th/0602044.
[9] Dragovich, B. – Dragovich, A.: *p-Adic Modelling of the Genome and Genetic Code*, received for publication in Computer Journal, 2006; arXiv:0707.3043v1[q-bio.OT].
[10] p-ADIC MATHPHYS.2007: http://p-adicmathphys2007.mi.ras.ru/index_e.html .
[11] Box, G. E. P. – Jenkins, G. M.: *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden-Day, San Francisco, 1976.
[12] Brockwell, P. J: *Introduction to Time Series and Forecasting*, Second Edition, Springer Texts in Statistics, 2002.
[13] Gujarati, D. N.: *Basic Econometrics*, International Editions, McGraw-Hill, Inc., 1995.
[14] Bollerslev, T.: "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, April 31:3, 1986, pages 307-327.
[15] Engle, R. F.: "Autoregressive Conditional Heteroskedasticity with Estimates Of the Variance of UK Inflation", *Econometrica* 50:4, 1982, pages 987-1007.
[16] Choudhry, M.: *The Bond and Money Markets: Strategy, Trading, Analysis*, First Edition, Butterworth-Heinemann, 2001.
[17] Dumas, B. – Allaz, B.: *Financial Securities: Market Equilibrium and Pricing Methods,* First Edition, Chapman-Hall, 1995.
[18] Sornette, D. – Johansen, A.: "A Hierarchical Model of Financial Crashes", *Physica* 261A, 1998, pages 581-598.
[19] Бикулов, А. Х. – Зубарев А. П. – Каидалова Л. В.: *Иерархическая динамическая модел финансового ринка вблизи точки обвала и р-адический Математический Анализ*, private statement in written form, 2006.
[20] J. Q. Trelewicz and I. V. Volovich, "Analysis of business connections utilizing theory of topology of random graphs", in: *p-Adic Mathematical Physics, 2nd International conference on p-adic mathematical physics*, Belgrade, Serbia and Montenegro, 15-21 September 2005, AIP Conference Proceeedings, 826, 2006, pp330-344.

Original scientific paper
Received: August 29th, 2007
UDK 336.748(4)"192/193"

Professor Nikolay Nenovsky, PhD
University for National and World Economics, Sofia
Professor Kalina Dimitrova, PhD
"St. Kliment Ohridsky" University, Sofia

EXCHANGE RATE AND INFLATION: FRANCE AND BULGARIA IN THE INTERWAR PERIOD*

Abstract: *The objective of this paper is twofold. First, we compare the model of financial stabilization in the interwar period in France (a country in the "core") with that in Bulgaria (a peripheral country). Second, by applying modern econometric techniques (VAR models), we would like to "test "whether the theory designating a dominant role of the exchange rate on inflation (in comparison to that of money in circulation) holds and can be empirically proven by the actual movement of the monetary variables and the direction of their causality. Going back to the history of the stabilization in France and Bulgaria in the interwar period and studying it through theoretical ideas at the beginning of the XX century would provide us not only with new elements in the analysis of the present-day problems of monetary stabilizations but also strengthen the arguments regarding the crucial significance of the exchange rate and monetary rules for the efficiency and credibility of the monetary regime.*

Key words: *economic history, modeling, France, Bulgaria.*

1. Introduction

The fall of communism in the Eastern European countries at the end of the 20th century raised the issue of financial stabilization in the region. When we look back into history, we can see some similar cases of financial stabilization

* This version of the paper was prepared during the stay of Nikolay Nenovsky at ICER, Torino. Earlier version was presented for the first annual conference of the South-East Europe Monetary History Network (SEEMHN) on "Monetary and Fiscal Policies in South East Europe: Historical and Comparative Perspectives" held on 13-14 of April 2006 in Sofia. The discussions and comments with researchers were very fruitful and helpful for completing the study. Among them we have to mention Enrico Colombatto, Bertrand Blancheton, Henri Bourguina, Dominique Torre, Alain Raybaut, Philippe Saucier, Cécile Hagnauer, Michel Lelart, Martin Ivanov, Roumen Avramov, Daniel Vachkov, Atanas Leonidov, Darina Koleva, Yuri Golland, Vladimir Mau and Sophia Lazaretou.

throughout Europe as a consequence of World War I (the Great War). In fact, the majority of European countries undertook radical measures during the 1920s in order to restore the monetary and financial stability that had characterized the European countries' economies during the time of the gold standard.

The study of the history of monetary regimes can provide us with some interesting insights. First, from the theoretical aspect, it provides us with material for the present-day debate about the "best" monetary regimes, not only in terms of obsolete and forgotten theoretical postulates but also in terms of empirical historical facts. Second, the present-day development of statistics and econometrics allows for the *ex post* testing of different theoretical models. And third, an analysis of economic history can provide us with valuable practical information, which can then be integrated in one way or another into the decision-making process in the present and the future.

Following a brief overview of the theoretical discussions that were current during the 1920s in Chapter 2, we will perform a comparative analysis of the financial stabilizations in France and Bulgaria following the Great War. In Chapter 3, we will test a theory by which the exchange rate has a dominant influence on inflation (Aftalion's Theory), using modern economic techniques (the VAR model and tests of causality). After summing up the results in accordance with the analyses of that time, we will offer some concluding observations about the factors that lay behind some of the similarities and differences between these two stabilization processes.

2. The exchange rate and inflation – a short journey through the theoretical debates of the 1920s

A theoretical analysis and empirical observations regarding the dynamics of price variation, money flows and exchange rates in several European countries at the beginning of the 20th century led the prominent French economist Albert Aftalion (1874-1956)[1] to certain doubts about the validity of the so-called Quantitative Theory of Money – QTM, and Purchasing Power Parity – PPP. Aftalion was not the first to examine the shortcomings and problems connected with QTM and PPP (Keynes, Hawtrey and Nogaro (1924), among others, questioned the validity and usefulness of these theories), but he was the first to propose a well-founded and systematic theory as a possible and constructive alternative. Despite its eclectic character, Aftalion's theory is characterized by elements of integrity and logical structure.

[1] Albert Aftalion was born in Rousse (Rustchuk). For biographical details and more information about his theory of exchange rate and money, see N. Nenovsky: *Exchange rate Inflation. The Contribution of Albert Aftalion (1874-1956)*, editions of the Bulgarian National Bank, 2006.

In brief, this is the way of thinking and the exposition of Aftalion's Theory. He begins with observations connected with the movements of the main variables in the QTM and PPP in different countries and different intervals (usually short periods of tome). Using basic statistical techniques (without taking into account whether they can also be interpreted as tests of causality), he arrived at the conclusion that neither of these two theoretical postulates can be supported by facts, nor can they, by way of testing, offer an explanation through (i) presenting the strongest possible arguments in favor of the development of monetary variables – income-based monetary theory, and, later, through (ii) an elaboration of the fundamental theory of income by adding certain expectations in regard to the behavior of monetary variables, before all the exchange rate ("the psychological theory of money and the exchange rate").

Aftalion held that causality in the context of QTM and PPP (chains of relationships between cause-and-effect) differs by country (nine were included in his empirical studies) and periods, since the exchange rate has a more dominant role in the price-setting process than money in circulation or the money supply. Taking the example of France, he estimated that the scale of the link (i.e. synchronized movement) between money in circulation and prices weakened – being still pretty strong in 1914-1919, then becoming pretty weak during 1919-1920, before finally disappearing altogether. At the same time, this weak causal connection was replaced by another, quite strong causality between prices and exchange rates – i.e. the devaluation of the franc directly resulted in an increase in prices. Moreover, Aftalion observed that growth in the level of money supply during 1927 and 1928 was not followed by a growth in inflation.[2] The movements of these variables were not the same in other examined countries, and Germany and Austria provide good examples. That is why Aftalion speaks of the "hegemony of the exchange rate influence" on inflation in the 1922-1924 period.[3]

Aftalion similarly criticizes the PPP, which is a logical continuation of the QTM in the context of international monetary relations. Here we remind the reader that, according to the PPP theoretical postulate, the nominal exchange rate is determined on the basis of the difference in prices between two countries, while this difference is again determined by the amount of money in circulation. Aftalion thinks that neither the PPP approach nor the approach based on current accounts is sufficient to explain exchange rate behavior. On the basis of empirical research, Aftalion rejects the causal relationship that starts from money in circulation and goes, by way of prices, to the exchange rate, claiming that the direction is, in fact, reversed, i.e. from the exchange rate toward prices.

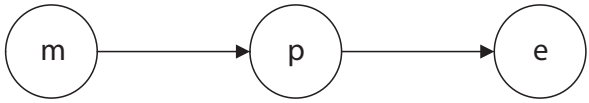
---

2 According to Aftalion, an expected stabilization of the exchange rate leads to an increase in the money supply (Aftalion, 1927, pp. 98, 109). A. Aftalion, *Monnaie, prix et change. Expériences récentes et théorie*, Recueil Sirey, Paris, 1927, pp. 98 and 109.

3 A. Aftalion, p. 109.

Such causal chains analyzed in the context of the QTM framework and Aftalion's theoretical proposals can be illustrated by Drawings 1 and 2, where *m* is the money in circulation, p is the price level and *e* is the exchange rate. In Drawing 1, the causal connection (arrow) from money toward prices is defined through the QTM, and the other, from prices toward the exchange rate, is the basis for the PPP.

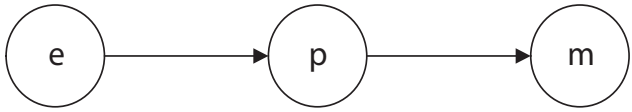
Drawing 1. *Causal chains in QTM and PPP*



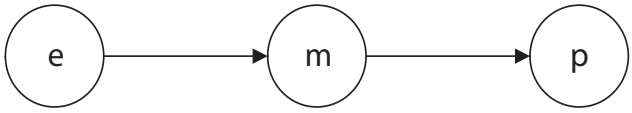
According to Aftalion's findings, the French experience can be illustrated through the first causal chain in Drawing 2, while the second chain is more reflective of the development of indicators in Germany. Still, in both countries he finds empirical proof that "the exchange rate has hegemony over prices."

Drawing 2. *Chains of causality in Aftalion's Theory*

French case

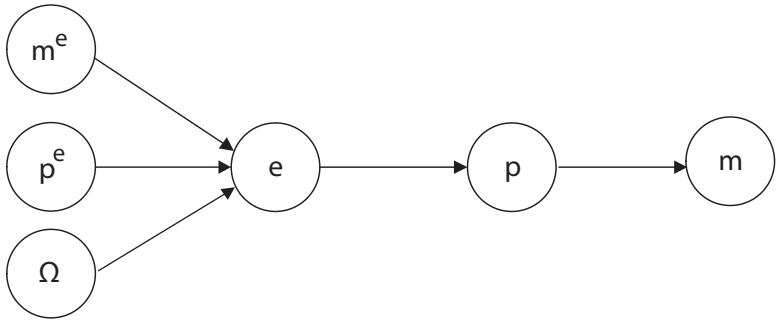


German case



In a more elaborated version of Aftalion's Theory (Drawing 3), the author assumes that inflationary expectations and projected future developments in the field of money supply (marked by ^{e}) influence exhange rate fluctuations. Moreover, he adds a large set Ω of other macroeconomic and political factors (situation tied to public finances, balance of payments, tax and tariff policy, political, international and war news, etc.) that influence the behavior of the "core" exchange rate variable.

Drawing 3. *Chains of causality in Aftalion's elaborated theory*



Whichever of these configurations we take, the same follows: Aftalion sums up the story with the claim that the exchange rate is the main and direct cause of inflation: "internal currency devaluation moves more or less in parallel with the exchange rate; it is the satellite, not the driving force of the exchange rate."[4] He, thus, concludes that the exchange rate has great significance as an "anchor" in securing the stability of money. He then proceeds to offer a practical proposal, i.e. that monetary stabilization should begin with the stabilization of *external purchasing power* (exchange rate), which will automatically lead to the stabilization of *internal purchasing power* (price level). The exchange rate *directly and immediately* sets expectations from economic factors, which is precisely why the battle against inflation should start with the exchange rate and not the money supply. Such a claim has direct bearing on policy, i.e. on the making of choices between stabilization that targets the exchange rate, monetary aggregates, or price stability.

3. Stabilization in France and Bulgaria – a chronological overview

The monetary and financial stabilizations after the Great War allow us not only to consider the complexity of the monetary and the exchange rate phenomena (economics, politics, ideology, diplomacy, nationalism and other aspects) but to draw parallels with the present, when monetary stability is on the table both at the core and at the periphery of the global economy.[5] During the examined time, France was a country at the core of the international financial system, a World War I victor, while Bulgaria was the total opposite, a peripheral

[4] A. Aftalion, p. 794.

[5] For an example, see: C. Kindleberger, *Histoire financière de l'Europe occidentale*, Economica, Paris, 1990. [1984]; C. Kindleberger, *La crise mondiale 1929-1939*, Economica, Paris, 1988. [1973, 1986]; B. Eichengreen, *L'expansion du capital. Une histoire du système monétaire international*, L' Harmattan, 1997. [1996].

country in which the stabilization issue was urgent.[6] Bulgaria was on the losing side in World War I and, like Germany, was forced to pay reparations.[7]

The Sargent Committee, established in 1926, brought a three-step plan toward the coordination of the establishment of international monetary stabilization: 1) a preparatory period in which inflation was to be controlled, while the exchange rate could fluctuate; 2) *de facto* stabilization once the central bank takes over responsibility for keeping the exchange rate at a certain level through the buying and selling of gold on the basis of gold points; and 3) the final step is when the exchange rate is set *de jure* (on the basis of gold).[8] Thus, it turns out that, despite the above-mentioned differences between the two countries, the stabilization process went through similar phases in both France and Bulgaria.[9] As result of the specific implementation of the prescribed phases of stabilization and their differing international financial positions, France and Bulgaria were among the countries that succeeded in preserving the stability of their currencies during the long period following the devaluation in Great Britain (1931) and the US (1933). France abandoned the fixed excange rate in 1936. By way of specific control of the exchange rate, Bulgaria officially withdrew from the rules on convertibility in 1941.[10]

---

6 The definitions of the center and the periphery differ; nevertheless, here we shall accept the distinction made by B. Eichengreen; he claims that in the period of the gold standard (1870-1914) the "core" was made up of Great Britain, the US, France and Germany, which were creditor countries, while the peripheral countries were characterized as debtors. For more details regarding questions on the exchange rate from the global and historical aspects, see: M. Bordo, M. Frandreau, "Core, Periphery, Exchange Rate Regimes, and Globalization," in: M. D. Bordo, A. M. Taylor, J. G. Williamson, Editors, *Globalization in Historical Perspective*, NBER conference held on May 4-5, 2001. Expected to be published by The University of Chicago Press.

7 Bulgaria entered the war as an enemy of France. The two sides battled in Macedonia at the end of 1915, after the Anglo-French landing on the Thessalonica Front.

8 R. Hawtrey, *The Art of Central Banking*, Longmans, Green and Co., London, 1932, p. 10.

9 For more details about the logical phases of the stabilization process, see: A. Aftalion, *L'or et la monnaie. Leur valeur. Les mouvements de l'or*, Les éditions Domat-Montchrestien, Paris, 1938; Vallance *Histoire du franc*, Champs, Flammarion, Pariz, 1998, [1996]; and Ch. Rist, *Essais sur quelques problèmes économiques et monétaires*, Recueil Sirey, Pariz, 1933. [1925]. Rist claims that monetary stabilization goes hand-in-hand with the stabilization of public finances, consisting of two components – stopping excessive (without backing) emission of money and balancing the budget. Stabilization is more a question of credibility and it usually begins once a sufficient quantity of foreign reserves is accumulated.

10 See: M. Ivanov, "Could We Devaluate? What is there behind the Orthodox Bulgarian Answer to the Great Depression?" *Historical Review*, no. 3, 4, 2005 (in Bulgarian)

3.1. France – Poincaré's stabilization

World War I brought extremely negative consequences to the public finances and the economies of the developed European countries, and, by extension, to their monetary stability.[11] The lack of will among the developed countries to reach a compromise resulted in chaotic and inconsistent measures that undoubtedly postponed all stabilization attempts.[12] After the restoration of the gold standard to the pre-war level in Great Britain in April 1925 [stabilization measures were also applied in Austria (1923), Germany (1924), Poland (1924), Switzerland (1924), Hungary (1925), Belgium (1925), Canada (1926), Finland (1926), Czechoslovakia (1926) and even Russia (1922), within the context of differing ideologies], the political will to at last solve the dilemma about how to achieve stabilization – by revaluation (i.e. deflation) or devaluation – finally appeared in France.[13]

Here we remind the reader that the stable franc dated all the way back to the days of Napoleon (*le franc Germinal*), with an unchanged gold content since March 27, 1803. As a result of currency over-emission during the Great War (as a comparison, in 1913 there were six billion francs in circulation, while their number reached 35 billion in 1919), prices increased significantly and a great number of new economic agreements were signed according to these prices. This made it difficult and almost "immoral" to restore the pre-war exhange rate by way of deep deflation and reduction of money in circulation. Another consequence of World War I were France's large domestic and foreing debts, further aggrevated by short-term debts, the so-called floating debts (*dette flottante*), which consisted of short-term treasury bills and, especially, war bonds (*bons de défense*). In spite of the widespread conviction at the time that the pre-

[11] On the consequences of the Great War, particularly in France, see: A. Sauvy, *Histoire économique de la France entre les deux guerres*, Economica, Paris, 1984, who estimated that the country lost in 15 months the earnings or wealth it had accumulated during the previous 11 years.

[12] Each country accused its partners of egotism and selfishness. For example, France did not want to make any political or economic concessions to Germany and its allies, as it greatly depended on the reparation damages from the defeated countries. On the role of reparations, see two different schools of thought expressed by: J. M. Keynes, *A Tract of Monetary reform*, MacMillan, London, 1923, and J. Bainville, *Les conséquences politiques de la paix*, Gallimard, Paris. 1920 [2002].

[13] Rueff enriched this dilemma, which was personified by Poincaré's life, with the "tones of an antique drama, where the heart (i.e. sentiments favoring the restoration of the old exchange rate) is fighting with reason (in favor of devaluation as a result of an irreversible price jump during the war)." Vallance, op. cit., p. 250. At that time, in most developed countries, and especially in France, there was a "pathological devotion to monetary stability and orthodox ways of thinking." T. Kemp, "The French Economy under the Franc Poincare," *Economic Historical Review*, New Series, vol. 24, no. 1, 1971, p. 82. Keynes analyzed in detail this dilemma on the "deflation-devaluation" relation, reducing it to a choice between price stability and exchange rate stability. J. M. Keynes, ibid.

war exchange rate should be restored (one of the chief proponents was Baron Rotschild), experts and representatives of the main interest groups gradually agreed that this was not possible and that a new, "cheaper" franc was needed. Even though the debate about the fixed exchange rate was important (one only need remember Keynes' criticism of Churchill regarding the restoration of the pre-war exchange rate of the British pound[14]), in this analysis we are interested only in the process of stabilizing the exchange rate, the restoration of currency convertibility and the gold standard behind the money in circulation, which was repealed on August 15, 1914.

After several currency crises caused by mounting evidence that Germany would not pay the expected reparations, and former president Raymond Poincaré's (1860-1934) ascendance to the prime ministry in January 1924, measures were taken in the direction of stabilization and the balancing of public finances. Still, not long after, Poincaré lost his office and, although the new, leftist-oriented government headed by Herriot initially tried to continue his financial policy, they soon lost power and brought France to the brink of economic collapse, to a situation on the verge of a currency crisis and domestic default. In the period of March 3, 1924 to April 5, 1925, the weekly statements of Banc de France were falsified through various accounting manipulations in order to conceal the significant rise of money in circulation. The break through the "ceiling" of 41 billion French francs became obvious on October 2, 1924. Under the pressure of advisers in Banc de France, the truth was revealed on April 9, 1925, which immediatel crashed the credibility of the franc, voided all expectations and increased the likelihood of a currency and financial crisis, on top of which Herriot's government lost power.[15]

Poincaré regained power in the July 1926 elections, formed a broad coalition (including Herriot himself) and immediately undertook radical reforms with the aim of "paying off the war bills" and stabilizing the franc. An expert group was formed for the purposes of discussing the technical details of the stabilization, within which Jacques Rueff was given a "special mission" (*chargé de mission*)[16] to find the optimal exchange rate for the franc. Expectations

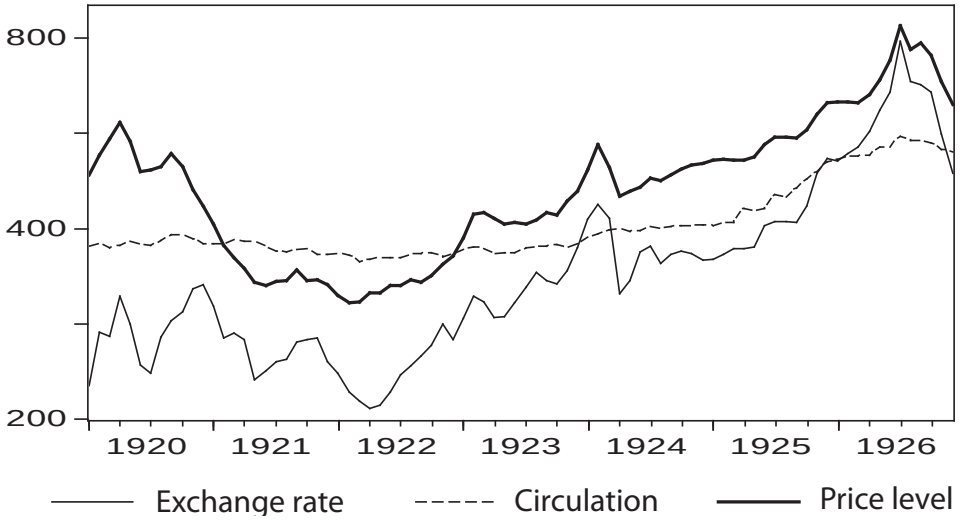
[14] Keynes was proposing a stabilized level of the franc's exchange rate, which was contrary to the British decision (see the collection of Keynes' articles and pamphlets on the franc from 1928).

[15] For more details about the falsification of the statements of the Banc de France, see: B. Blancheton, *Le Pape et l'Empereur. La Banque de France, la direction du Trésor et la politique monétaire de la France (1914-1928)*, Albin Michel, Paris, 2001; and D. Jacob, *La Banque centrale et le secret, Thèse pour le doctorat ès sciences économiques*, Université Montesquieu-Bordeaux, 1996.

[16] Rueff put his experience in the field of stabilization to concrete use during the stabilization of the franc in 1958, which he carried out under the patronage of president de Gaulle, Jacques Rueff led the financial mission of the League of Nations in Bulgaria, Greece and Portugal in the 1927-1930 period.

regarding the stabilization rose and strengthened as a result of constructive reforms in the public finance sector (balancing the budget) and, mainly, by cutting costs, increasing taxes and converting short-term into long-term debt. This process was accompanied by an inflow of capital that led to an accumulation of foreign reserves at the Banc de France and the restoration of demand for the domestic currency (in real quantities).

Figure 1: *France (1920-1926): Price levels, money in circulation and exchange rate (franc-US dollar) (logarithmic scale)*



In August 1926, a new upper limit for money in circulation was introduced, while in February 1927 the emission of treasury notes stopped. The nominal value of the franc began to grow relative to other currencies that opted for the next step of *de facto* stabilization. Banc de France began to intervene on the foreign currency market in order to lessen the fluctuations of the exchange rate (just as it protected the franc from sudden appreciation in certain periods). The next step was the removal of restrictions on the outflow of capital (January 10, 1928) and the introduction of the Law on the Devaluation of the Franc, through which the value of the franc reached about 80% of its pre-war value. In this phase of the stabilization, the cenral bank held a significant quantity of gold reserves. From June 1928 to December 1932, foreign reserves increased to 55 billion francs, i.e. from 8% to 27% of global gold reserves,[17] which allowed *de facto* stabilization. This was aided by the fact that profits from the revaluation of gold were used to strengthen the balance of the

[17] K. Mouré, "Undervaluing the Franc Poincare," *Economic Historical Review*, New Series, vol. 49, no. 1, 1996, pp. 137, 138.

central bank (according to the new rate, 1,700 tons of gold brought an increase in Banque de France's foreign reserves from 5.6 billion to 26 billion francs).

Subsequently, following the devaluations in Britain and the US in 1931 and 1933, France remained isolated in the so-called gold block. And, finally, when even the members of the "gold block" devalued their currencies one by one (Belgium, Switzerland), France was compelled, under the leftist government of Leon Blum, to renounce Poincaré's franc on September 26, 1936. Thus did the gold franc live for the span of about seven years, making France's monetary stabilization the longest among the core countries of the financial system. We can say that Poincaré's stabilization was a clear illustration of the role of the fixed exchange rate,[18] while convertibility and discipline in the field of public finance are a classic example of how to build confidence in the national currency.

3.2. Bulgaria – the stabilization of the Bulgarian lev

The stabilization in Bulgaria followed the stabilization logic of the developed countries, augmented by specificities of the peripheral and less developed countries.[19] Bulgaria's economic and financial losses from World War I piled on top of the losses it sustained in the exhausting Balkan Wars of 1912-1913 (these three wars are generally referred to as the Great War). According to Kiril Nedelchev, the daily costs of waging the Balkan Wars amounted to about one million levs, climbing to two million levs during World War I.[20] According to the same author, excepting territorial losses, Bulgaria's overall costs during World War I can be rounded off to three billion gold levs. The public finances were in absolute turmoil; during the 1916-1918 period, the budget deficit equaled about 1.5 billion gold levs, while the central bank (the Bulgarian National Bank – BNB) financed almost all the government's war costs.[21] As a result, the amount of money in circulation drastically increased (by a

[18] K. Mouré says that certain critical remarks regarding the QTM and PPP were recorded in connection with discussions about the technical details of Poincaré's stabilization, as well as in connection with the balanced exchange rate (this is why the terms *depreciation and appreciation* are rarely mentioned), and even regarding the significance and popularity of Aftalion's psychological theory of the exchange rate. K. Mouré, ibid.

[19] On the way in which the gold standard functioned and its characteristics at the core and the periphery of the global economy during the pre-war years, see: P. Whale, "The working of the prewar gold standard," in: B. Eichengreen, M. Flandreau Eds., *The gold standard in theory and history, Routledge*, London and New York, 1997, and for the interwar period: B. Simmons, *Rulers of the game: central bank independence during the interwar years*, International Organization, vol. 50, no. 3, 1996.

[20] K. Nedelchev, *Monetary Issues: Bulgaria 1879-1940*, Knipegraph Printing House, Sofia, 1940, pp. 76, 77 (in Bulgarian).

[21] A. Ivanov, *BNB governor Asen Ivanov's report at the five-year celebration of the founding of the BNB*, Central Bureau of Public Records, Bulgarian National Bank: Document collection, vol. 3, 1915-1929, document 22, 1929, p. 139 (in Bulgarian).

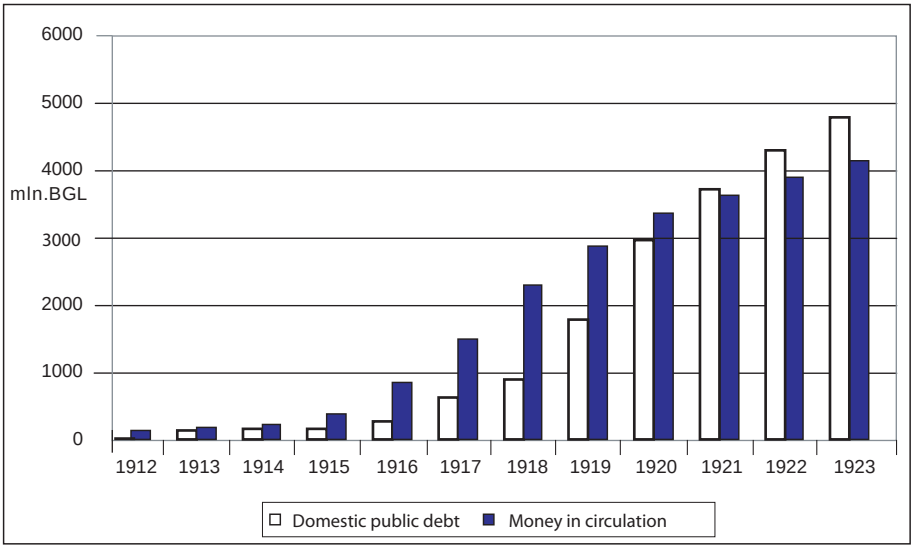
rough factor of 14), while the backing of the gold currency fell to 3.2% and of the silver currency to 5.9% (Table 1). The public debt, and especially the floating debt, reached dangerous heights (Figure 2).

Table 1: *The backing for money in circulation in Bulgaria (1912-1918)*

Year	Money with gold backing (in mil. of levs)	Gold reserves (mil. of levs)	Gold backing (%)	Money with silver backing (mil. of levs)	Silver reserves (in mil. of levs)	Silver backing (%)
1912	139.6	51.1	36.6	24.7	16.8	58.0
1913	166.0	55.3	33.3	22.8	23.4	102.6
1914	198.9	55.1	27.7	27.7	28.5	102.9
1915	304.8	61.4	20.1	65.1	22.5	34.6
1916	577.1	68.2	11.8	256.8	17.2	6.7
1917	1,176.0	62.9	5.3	316.8	16.9	5.3
1918	1,969.4	64.0	3.2	329.2	19.4	5.9

Source and notes: K. Nedelchev, "Monetary Issues: Bulgaria, 1879-1940," 1940, p. 77. In examining the data we found that, in the course of calculating the gold and silver backing for the money in circulation, Nedelchev used the maximum amount of money in circulation achieved during the year in question, as well as the gold and silver reserves as of December 31 of the same year.

Figure 2: *The public debt and money in circulation (1912-1918)*



Source and notes: K. Nedelchev, "Monetary Issues: Bulgaria, 1879-1940," 1940, p. 81. "The total public debt for 1922 and 1923 includes treasury notes, which for these two years equaled 150 and 300 million levs, respectively."

In the period between the end of 1918 and the end of 1922, before reparation payments began on October 1, 1923, 112 million gold levs went toward the servicing of the foreign debt, which equaled 16.3% of total public spending.[22] To this debt were then added, on November 27, 1919, according to the Neuilly-on-the-Seine Agreement, the reparation payments, which all came to 2,250 million gold levs, with an annual interest of 5% for the next 37 years, plus the occupation costs, which equaled one-fourth of the national wealth. Out of the total Bulgarian foreign debt, about 26% was to France, followed by Italy (25%), Greece (12.7%) and Romania (10.55%). The foreign debt made up 96% of the public debt, and reparations covered 9/10 of the total foreign debt.[23] Despite this difficult situation, Bulgaria made great efforts to maintain its reputation of a "good" debtor, who not only carried the debt burden on his own shoulders but did not enjoy any preferential treatment in terms of debt relief.[24]

The convertibility of the lev was *de facto* abolished at the beginning of the wars (October 10, 1912), while unconditional wartime government financing was ended in January 1919.[25] It was assumed that the interruption in convertibility would be temporary, akin to typical gold standard disruptions during wars or other extreme events (the exit clause rule[26]). As a result, the lev was devalued in the 1915-1918 period by a factor of 16.4[27] and, for the extended 1912-1923 period, by a factor of 26.65.

As prescribed by international financial experts, the stabilization in Bulgaria also had three phases and the following chronology.[28] The first, initial phase began in 1922, with the introduction of the Law on Limiting the Emission of Money

22 For more on the development of the Bulgarian foreign debt as well as the overall situation in the Bulgarian economy after the Great War, see the essential and yet unpublished research work on the history of the Bulgarian government's foreign debt: Vatchkov, Ivanova, Todorov and Koszul, *Les efforts de restauration financière de la Bulgarie (1922-1931),* Félix Alcan, Paris, 1932, and N. Stoyanov, *Reparations and Inter-Allied Debts. Bulgarian Government Debts*, Sofia, 1933. (in Bulgarian).

23 J. Koszul, p. 40.

24 M. Ivanov, *Political Games with the external Debt. Bulgarian Scenarios of Economic Crises and Upsurge*, 1929-1934, Zlatyo Boyadzhiev Publishing House, Sofia, 2001 (in Bulgarian); M. Ivanov, "Could We Devaluate? What is There behind the Orthodox Bulgarian Answer to the Great Depression?," *Historical Review*, no. 3, 4, 2005. (in Bulgarian).

25 BNB, Document Collection, vol. 3, Sofia, 2001, pp. 55-56

26 M. Bordo, F. Kydland, "The Gold Standard as a Commitment Mechanism," in: T. Bayoumi, B. Eichengreen, M. Taylor, ed., *Modern Perspectives on the Gold Standard*, Cambridge University Press, 1996.

27 G. Toshev, *Foreign Currency and Foreign Exchange Policy. A Scientific Inquiry into the Reasons of the Economic Crisis,* Franklin Cooperative Printing House, Sofia, 1928, pp. 116, 172 (in Bulgarian).

28 For more details, see: Zh. T. Burilkov, *The Monetary Reform*, S. M. Staikov Printing House, Sofia, 1928 (in Bulgarian); G. Toshev, ibid.; A. Ivanov, ibid.; K. Nedelchev, *Monetary Issues: Bulgaria 1879-1940*, Knipegraph Printing House, Sofia, 1940 (in Bulgarian); L. Berov, *The Bulgarian National Bank at 120, unpublished document*, 1977 (in Bulgarian).

(with an upper limit of about 5.5 billion levs). A subsequent amendment to this law required that the money in circulation and other constant obligations and debts of the Bulgarian National Bank must not grow larger than double the value of the gold and reserve currencies when it came to the bank's available resources.

Measures were also taken to limit public finances by limiting all advance government crediting on the part of the BNB to the amount of 4,700 million levs (or up to 5,400 million levs in special cases). Still, under such harsh budgetary limits and large debt burden (yearly payments came to about 132.5 million gold francs, or more than 50% of yearly budget revenues[29]) government financing by way of monetary emission was the key issue. Thus did the state of public finances have the key role and influence on variations of the exchange rate and the price levels?

The expectations that accompanied each round of debt negotiations greatly influenced the dynamic of the exchange rate variations. The Bulgarian lev reached its lowest international buying power in 1921 (184 levs to one US dollar), while in June 1923 there was a sudden jump of the lev to 75 to the dollar, or an appreciation of 245%. The events that followed, i.e. the French invasion of the Ruhr region and the accompanying political and economic consequences for Germany, once again set the international exchange rate system on the path of devaluation. As a result of the drastic devaluation of the reichsmark, Bulgaria (the BNB) lost a huge amount of money that was denominated in reichsmarks and blocked in German banks.[30] These resources were the backing of the Bulgarian lev and their devaluation delivered a powerful blow to the Bulgarian lev's buying power. This once more compelled the BNB to introduce a monopoly on currency exchange transactions (the first currency exchange monopoly was introduced in December 1918, and repealed in May 1920). The currency exchange market was closed (December 11, 1923) and the BNB began to set the buying and selling rates for foreign currencies. According to the law of May 2, 1924, the BNB began to manage the exchange rate (on a scale where the selling rate was 139 and the buying rate 137.2 levs to the dollar), which was interpreted as an act of effective stabilization.[31]

The Law on the BNB of November 20, 1926, is considered to be the next step in the process of stabilizing the lev, and it consisted of the creation of convertibility, which strengthened the transition to a gold-foreign currency standard. On the basis of Article 8 of this law, the backing for this money was to be 33 1/3%, with a target level of 40%. Although the law defined the backing for the money in circulation, the exchange rate against gold was not fixed, i.e. the gold content in the lev was not yet determined.

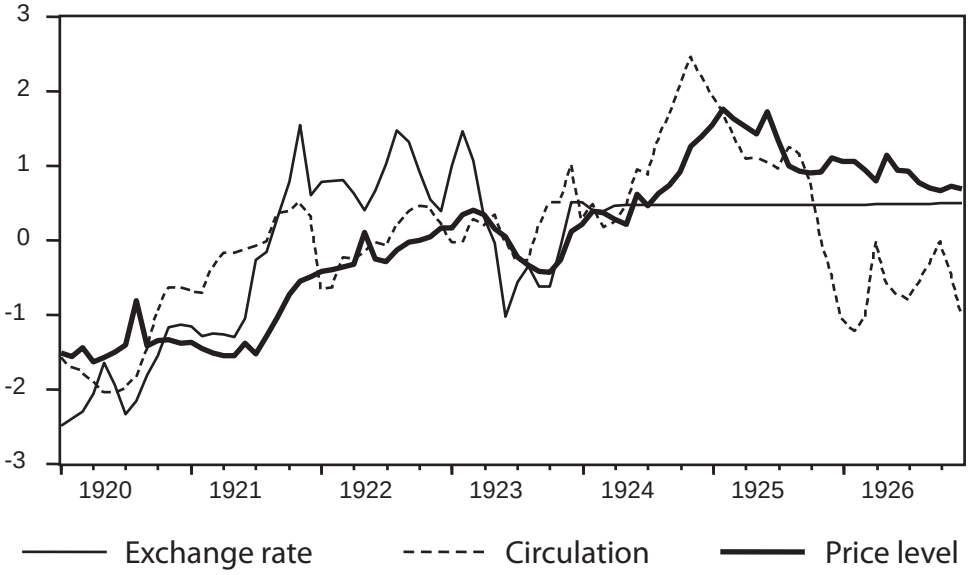
---

29 J. Koszul, ibid.

30 Bulgaria entered World War I on the condition that it would receive credits and financial aid from Germany and Austria-Hungary in the amount of about 200 million gold francs.

31 A. Ivanov, *BNB governor Asen Ivanov's report at the five-year celebration of the founding of the BNB*, Central Bureau of Public Records, Bulgarian National Bank: Document collection, vol. 3, 1915-1929, document 22, 1929, p. 141 (in Bulgarian).

Figure 3: *Bulgaria (1920-1926): Price level, money in circulation and exchange rate (lev-US dollar) (normalized scale)*



With the Law on Stabilization (of December 3, 1928), the lev was at last legally tied to gold, as Article 1 of the law set the exchange rate at "92 levs for 1 gram of pure gold." Even more specifically, which explains the BNB's commission, the rate of 139 levs for one US dollar meant 139 levs for one-and-a-half gram of gold (which was the amount of gold in one dollar). From the very start, Bulgaria's statesmen emphasized the key role of the lev's stabilization as the basis of an overall financial and economic stabilization. In the name of the central bank, deputy director Burilkov linked the stabilization of the lev to the restoration of morals to business relations.[32]

Having in mind the loss of resources in terms of foreign currencies, the stabilization of the lev was accompanied by a deflationary monetary policy.[33] The reduction in the money supply was a subject of debate and criticism that came mainly

[32] Zh. T. Burilkov, *The Monetary Reform*, S. M. Staikov Printing House, Sofia, 1928, p. 23 (in Bulgarian).

[33] In the period following the exchange rate crisis of 1924-1927, the money supply (money in circulation) fell dramatically due to the BNB's restrictive monetary policy. For example, Toshev estimates that the reduction came to about one third. G. Toshev, op. cit., pp. 176-177, while, on the other hand, prices did not fall by the same amount. According to some other authors, in order to achieve balance, the prices should have fallen by an even bigger percentage (about 40%) than the percentage reduction of the money in circulation. Yurii, *Monetary Crisis Liquidation (the Appreciation of Our Lev)*, Pridvorna Printing House, Sofia, 1923, p. 28. (in Bulgarian)

from academic economists such as Toshev,[34] Yurii,[35] Nikolov,[36] Totev,[37] Boshnyakov,[38] Chapkunov,[39] Sarailiev,[40] and Monchev,[41] among others, who considered it the root of the economic crisis in the country.

Even though the Great Depression exacerbated the economic crisis in the country,[42] Bulgaria maintained its fixed exchange rate and the convertibility of its currency even after the devaluation of the US dollar in 1933. The lev remained fixed to the French franc.[43] After the devaluation of the franc in 1939, the BNB governing board continued to maintain a fixed exchange rate until the end of the year, stating that "we are not directly affected by this devaluation and there is no need at the moment for any meliorative measures, our exports will continue to grow, as they have until now."[44]

The chronology of the stabilization processes in France and Bulgaria shows obvious parallels between events. This parallel movement of the monetary stabilizations was preordained by the decisions brought at two international conferences, in Brussels (1920) and, especially, in Genoa (1922). Moreover, in a way, France (as the main creditor) was, for Bulgarian politicians

34 G. Toshev, ibid.

35 Yurii, ibid.

36 G. Nikolov, *Stabilization: Monetary and Economic (Thoughts on the Monetary and Economic Crisis),* Elisei Petkov Printing House, Sofia, 1927 (in Bulgarian).

37 T. Totev, *The Money. Coins – Banknotes – Paper Money*, Zadruzhen Trud Printing House, Sofia, 1932 (in Bulgarian).

38 D. Boshnyakov, *Prices and the Monetary Problem*, Razvitie Printing House, Sofia, 1936 (in Bulgarian).

39 A. Chapkunov, *Crises and Foreign Currency*, Hudozhnik Printing House, Sofia, 1936 (in Bulgarian).

40 G. Sarailiev, *Monetary Devaluations and Their Effect at Home and Abroad*, Sofia, 1937 (in Bulgarian).

41 B. Monchev, *The Monetary Problem after the World War*, Sofia, 1939 (in Bulgarian).

42 For more details on the influence of the Great Depression on the Bulgarian economy and for theoretical discussions regarding that period, see: M. Ivanov, *Political Games with the External Debt. Bulgarian Scenarios of Economic Crises and Upsurge*, 1929-1934, 2001, Zlatyo Boyadzhiev Publishing House, Sofia (in Bulgarian); M. Ivanov, "Could We Devaluate? What is There behind the Orthodox Bulgarian Answer to the Great Depression?," *Historical Review*, no. 3, 4, 2005 (in Bulgarian).

43 BNB, document collection, vol. 4, Sofia, 2004, p. 419.

44 BNB, document collection, vol. 4, Sofia, 2004, pp. 557-562. According to Monchev, there were two tendencies: the creditors (including France) sought to devaluate their own currencies in order to improve their balances of payments, while the debtors (mainly agricultural countries, including Bulgaria) opted to preserve the purchasing power of their currencies in order to reduce their debts (they frequently resorted to protective tariffs). B. Monchev, ibid. A historical analysis as to why Bulgaria did not devaluate is offered by: M. Ivanov, "Could We Devaluate? What is There behind the Orthodox Bulgarian Answer to the Great Depression?," *Historical Review*, no. 3, 4, 2001 (in Bulgarian).

and economists, an example (institutional benchmark) of a sound monetary policy, which they went to great lengths to "imitate." As a result, France and Bulgaria began preparations for the stabilization of their national currencies at almost the same time, in 1924. Nevertheless, due to the fact that the French stabilization was interrupted (Poincaré's first stabilization) for a period of two years due to political reasons and that it was resumed in 1926, when Poincaré returned to power, the *de facto* stabilization of the Bulgarian lev was carried out before the stabilization of the French franc. Subsequently, both the stabilizations reached their "legal basis" together, in 1928.

3.3. Two different ideological schemes

It should be noted at the beginning that most Bulgarian authors who dealt with the period preceding the stabilization noted that there were empirical inconsistencies in the traditional QTM and PPP postulates, for example Petkof,[45] Kemilev,[46] and Yurii.[47] If we return to the war years of 1915-1918, Berov summed up these disagreements in the following way: he explained the increase in money in circulation by a factor of 6.2, the increase in prices by a factor of 5.5 and the devaluation of the lev against the Swiss franc by a factor of only 1.5, by the state's strong intervention in the economy during those years.[48] According to Toshev, "the barometer that informs us when inflation is starting to accelerate is the exchange rate."[49] Many other authors,[50] along with Bulgarian economists acquainted with French literature on the subject,[51] who analyzed the period before the stabilization, agreed that Aftalion's theoretical framework offered the most logical explanation for the failures of the QTM and PPP during the interwar period.

If we take the classic interpretation of stabilization as a *credibility effect* (or confidence effect) and a *discipline effect*,[52] we can claim that, contrary to the dom-

45 J. Petkof, *Prix, circulation et change en Bulgarie de 1890 a 1924*, Jouve & Cie Editeurs, Paris, 1926.

46 A. Kemilev, "Money Circulation in Bulgaria," *Journal of the Commercial College,* Varna, vol. I, Edition 4, 1936, pp. 287-300 (in Bulgarian).

47 Yurii, op. cit.

48 L. Berov, *The Bulgarian National Bank at 120*, unpublished document (in Bulgarian), 1997, p. 71.

49 G. Toshev, op. cit., pp. 114-116.

50 J. Koszul, *Les efforts de restauration financière de la Bulgarie (1922-1931)*, Félix Alcan, Paris, 1932.

51 P. Ilief, *La banque nationale de Bulgarie et l'histoire de sa politique monetaire*, Lyon, 1930; J. Petkof, ibid.

52 Raybaut and Torre give an interesting analysis on the topic of the effects of credibility and discipline. A. Raybaut, D. Torre, *Discipline, confiance et stabilité des régimes de caisse d'émission en transition vers l'Euro*, paper presented in Marrakesh, Morocco, March 16-17, 2005.

inant ideology of the stabilization of the franc in France, the role of the credibility effect was relatively underestimated in the stabilization of the lev. In other words, Bulgarian economists and politicians constantly emphasized that the main motive behind the stabilization of the lev was actually to manage the supply of money in circulation, and not the achievement of convertibility and a fixed exchange rate that could be trusted. Exceptions were rare, and it was mostly the BNB economists[53] who gave more weight to the credibility effect achieved by the stabilization of the lev, without rejecting the significance of the discipline effect.

Nevertheless, some Bulgarian authors totally neglected the role of the credibility effect produced by stabilization. For example, in criticizing Adolph Wagner's theory of credibility, Toshev pointed to the existence of certain objective economic laws, in the absence of which "scientific arbitrariness" would rule the field.[54] He claims that the amount of money in circulation is more important than the rules of convertibility because "convertibility is just an empty word,"[55] while "an estimate of the lev's gold content is an arbitrary act." [56] Like Toshev, Nikolov stated on numerous occasions that the main issue was not the backing of the money or whether the backing should be in a reserve currency or in gold, but that quantity was the crucial factor. He said that there was a "harmful psychological tendency toward a gold currency instead of a reserve currency (or fiduciary money based on confidence)."[57] In his opinion, "the same money backing is not equally appropriate at all times and in all countries."[58] Toshev also agreed that the main question was: "how much money is needed for the foreign exchange market?"[59] And, even though he mentions the role of "money with credibility" (as the second factor influencing the value of money, in addi-

[53] On numerous occasions, BNB economists have emphasized the role and significance of the factor of confidence in monetary stabilization. For example, in BNB's annual report for 1929, within an analysis of the Bulgarian economic situation at the beginning of the Great Depression, it is underlined: "Neither the situation regarding the exchange rate nor the measures taken on the part of BNB management give any reasons for concern regarding the stabilization of the lev. Unfortunately, some disturbing rumors in this regard caused a great stir in our public opinion, which is famous for always worrying about all sorts of invented threats; these rumors also reached the outside world and caused great damage to our finances. The Bank's management did everything to dispel such fears and doubts and to convince the public that, despite the unfavorable economic development, it is capable of maintaining the stability of the lev." BNB, document collection, vol. 3, Sofia, 2001, pp. 55-56

[54] G. Toshev, op. cit., p. 99

[55] Ibid, p. 178

[56] Ibid, p.199

[57] G. Nikolov, *Stabilization: Monetary and Economic (Thoughts on the Monetary and Economic Crisis)*, Elisei Petkov Printing House, Sofia, 1927, pp. 32, 33 (in Bulgarian).

[58] Ibid, p. 34

[59] G. Toshev, op. cit., p. 52

tion to quantity), he insists, "in the process of stabilizing the monetary system, the quantity of money is more important than the backing."[60]

Thus were formed two different "ideological schemes" of monetary stabilization – in France, where the credibility effect is dominant, and in Bulgaria, where the discipline effect predominated, and these two schemes to a certain extent preordained (of course, along with some other factors) the different approaches to the implementation of the stabilization processes in these two countries.[61] While the stabilization in France was more "market oriented" and based on the rebuilding of credibility, the Bulgarian stabilization was more administrative in character and regulated by the state (the monopoly on foreign currency exchange imposed by the BNB, control over capital flows), and its goal was the establishment of financial discipline.

In spite of numerous interpretations, Bulgarian economists were united in emphasizing the role of the psychological factor and of expectations during the setting of the exchange rate and the movement in prices. Similarly to the expectations tied to the influence of Germany's reparation payments on the French franc, the value of the lev was greatly influenced by expectations regarding the results of debt negotiations. The exchange rate of the franc was additionally influenced by the expected introduction of new taxes announced by the leftist government in July 1926, while the lev's exchange rate reflected the expected increase in customs tariffs of 1921.

Moreover, we could draw a direct parallel between the influence of capital flows on the exchange rate and price movement. There was a great outflow of capital from France until the middle of 1926, but after the Poincaré government's second mandate, capital returned to the country. Similarly, there was a great outlow of capital from Bulgaria between July 1923 and April 1924, when high capital profit taxes were expected, while a quite contrary tendency occurred in the wake of the strong and explicit motivation on the part of the stabilization authorities. If we take Aftalion's scheme, the sum of factors Ω that influenced the exchange rate was made up of: 1) the expected optimistic outcome of reparations negotiations; 2) a tendency in the direction of balancing the budget; 3) the development of a positive trade balance; and 4) many other psychological factors, which were mentioned by Chapkunov.[62]

---

60 Ibid, 102-105

61 The different characteristics of the stabilization process when Bulgaria is compared with the other "gold bloc" countries are the subject of Sarailiev's analysis, which focuses on the balancing between "the devaluation of the lev and the increase in customs duties" and comes out for the latter in order to improve Bulgaria's foreign balance. According to him, the first approach is like a "leap in the dark." Readers will find in Sarailiev's book some arguments regarding the presence of the "initial sign" in the peripheral countries. G. Sarailiev, op. cit., pp. 27, 32.

62 A. Chapkunov, *Crises and Foreign Currency*, Hudozhnik Printing House, Sofia, 1936, p. 39 (in Bulgarian).

4. An econometric test almost one century later

Even though they seem simple from the standpoint of the present, Aftalion's methods of correlation and standard errors within time series of the main variable indexes, and his pedantic comparisons of their dynamic were quite progressive for his time. In the context of present-day theories, he mistakenly, or at least in an oversimplified manner, interpreted changes in variable x that come before the change in variable y as a causal connection ($x \rightarrow y$). Today we know that problems connected with causality are much more complex, even with causality as understood by Granger-Sims.[63]

Before we move on to econometric tests of Aftalion's psychological theory, it is necessary to comment certain methodological characteristics. First, the tests will be applied to the period beginning with the end of World War I and ending with the stabilization – since the stabilization is being treated as a reaction to the dominant role of the exchange rate in the explanation of price movements and in inflationary expectations. The tests end with the *de facto* stabilization, so that the last observation in the case of Bulgaria dates from May 1924, and, in the case of France, from July 1926. This automatically shortens the examined time period to about two years.

Second, it is important to take into consideration the market "purity" characteristics in the period under study, i.e. to what extent the movement of the variables is driven by market forces and principles and to what extent the state regulates the prices or determines the development of other indicators (regulating the deficit by a coupon system or exchange rate controls). The more "clearer" the market is, the more obvious the effect of the psychological theory. No doubt, that during the war years the intrusion of the state in both countries is very strong. As we have already mentioned, the Bulgarian stabilization is considerably less market-based than the French one, and respectively the period in which we can detect Aftalion's interrelations between the variables is shorter (this could be one of the reasons why he has not included Bulgaria in the sample of countries which he studied).

Third, it is important not to lose sight of the quality of statistics in both countries, for example the extent to which the given price indexes reflect the actual price movements or the extent to which the reported money in circulation matches the true state of affairs. We merely wish to emphasize that; similarly to the case of France (May 1924 – April 1925), there were reports of accounting manipulations regarding the balance of payments of the Bulgarian central bank,[64] in order to conceal the increase in the money in circulation and

[63] Variable x causes y within the given set of data if the momentary value of y can be better predicted by adding the earlier values of x than without them.

[64] This occurred outside of the examined period, during the 1930s, under BNB governor Dobrim Bozhilov. Subsequent accusations connected with falsification were directed

violations of the law. Toward this purpose, we will provide a detailed description of data regarding the dynamic of the indicators that reflect the characteristics of data quality and the influence of politics, which will be taken into consideration in the econometric analysis.

Econometric estimates of Aftalion's psychological theory for the pre-stabilization period in both countries (France and Bulgaria) will come down to several tests of the dynamic of the exchange rate, money in circulation and prices. The first step includes an analysis of the development of the indicators and their basic characteristics with a focus on making consistent statistics regarding time series, inconsistencies and correlation.[65] In order to identify the chains of causality in the second step, we will apply Granger's causality test to each pair of variables. Finally, the econometric analysis of the relationship between the indicators and their mutual dependence will be extended to tests of "shock replies" and a total probability analysis based on VAR models.[66]

4.1. The econometric test for France (1920-1926)

The examined pre-stabilization period is the same one examined by Aftalion (1920-1926) and, even more precisely, it ends with July 1926, when it became clear that Poincaré would launch a "second" wave of stabilization reforms. Thus, we are using the same data that Aftalion used.[67] Here we would like to direct the reader's attention to a very interesting methodological case.

As already said, the Banc de France's weekly balance sheets were falsified in the period between May 1924 and April 1925, in order to conceal the over-emission of money (Blancheton examined this case in detail).[68] The weekly analyses of the balance sheets reveal that, by way of these falsifications, the central bank succeeded in holding to the legally established maximum of 41 billion francs in the fourth week of each month until August 1924. In the last week of September 1924, the violation of the rules could no longer be hidden, and the whole affair came to the surface on October 2. Even though neither the period in which the falsifications took place nor the scope of the falsified information were overly large (they, nevertheless, served as symbols of rule-

at Kiril Gunev, deputy governor in the same period. BNB Document Collection, vol. 4, Sofia, 2004, pp. 945-952

65 Similarly to what Aftalion and Petkof did in the case of Bulgaria, by way of "manual" calculations. A. Aftalion, ibid.; J. Petkof, ibid.

66 The analytical steps are similar to those taken by Spanos. A. Spanos, E. Andreou, G. Syrichas, *A VAR model for the monetary sector of the Cyprus economy*, University of Cyprus and the Cyprus Central Bank, 1997.

67 A. Aftalion, op. cit., pp. 58-64.

68 B. Blancheton, *Le Pape et l'Empereur. La Banque de France, la direction du Trésor et la politique monétaire de la France (1914-1928)*, Albin Michel, Paris, 2001.

breaking), we encounter a methodological case based on the fact that Aftalion built his theory on the official values of the observed indicatiors.

It would, thus, be logical to ask ourselves some questions. First, is there any sense in using true, credible information when the expectations tied to the factors and the decisions about them were made on the basis of falsified data? Second, is it still not better to use real data, which can explain the fundamental relations within the QTM and PPP in a better way? Third, would it be right to reject Aftalion's theoretical relations based on data he did not even use? And, finally, why didn't he (or other economists, such as, for example, Rueff) begin to use real data when the falsifications were uncovered in April 1925 and when it turned out that this was the direct cause of the exchange rate crisis of 1926?[69] The answers to these questions could serve as the topic of another discussion.[70] In this work, we shall use the same data used by Aftalion to build his psychological theory of money and the exchange rate.

The statistical characteristics and, before all, the unmatching variables (either in levels or in the first difference of the logarithms) immediately show that the dynamic of the exchange rate is the most volatile (0.07), followed by price movements (0.04) and money in circulation (0.01). Moreover, the correlation matrix (Table 2) shows a stronger interdependence of movement between the exchange rate and prices, and between the exchange rate and money in circulation than between money in circulation and prices. We should recall that Aftalion divided this period into small samples (up to one year) and calculated the correlation for each of these sub-samples. Nevertheless, we think that this approach is not overly useful and informative, due, among other things, to the problem of interpreting correlations as causalities. Hense, we prefer to test Aftalion's theory for the entire period.

Table 2: *France: matrices of variable correlations*

In levels	Exchange rate	Money in circulation	Price level
Exchange rate	1.000000	0.925512	0.874854
Money in circulation	0.925512	1.000000	0.855251
Price level	0.874854	0.855251	1.000000

First difference of the logarithms	Exchange rate	Money in circulation	Price level
Exchange rate	1.000000	0.286006	0.713637
Money in circulation	0.286006	1.000000	0.332630
Price level	0.713637	0.332630	1.000000

69 J. Rueff, *Théorie des phénomènes monétaires*, Payot, Paris, 1927.

70 It is possible to estimate the correlation between true and falsified data for the purpose of analyzing whether these data are moving in parallel or not. A second proposal is to place an invented "control" variable inside the model.

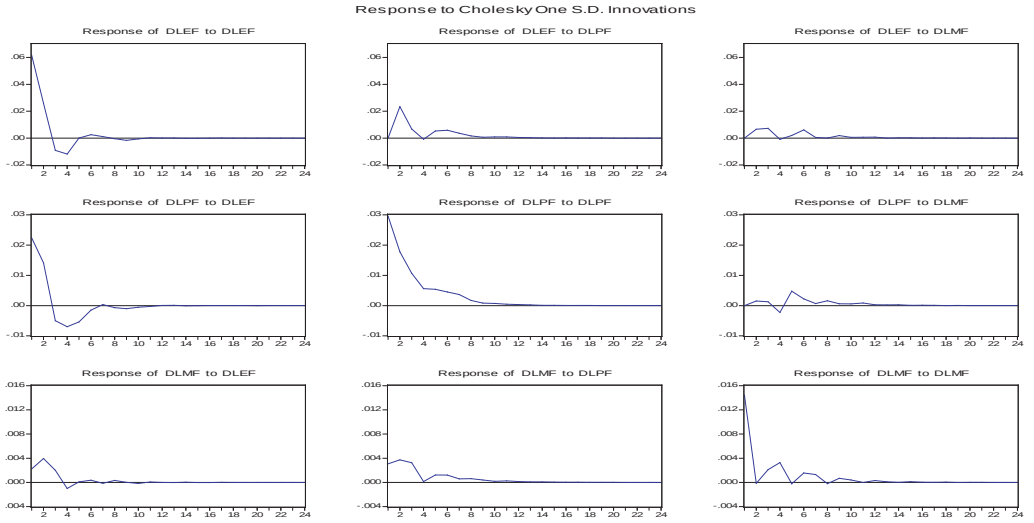
Within the second step of indentifying the causal relations, we examine Granger's test of causality in pairs by taking a probability threshold of 0.05 as a criterion for rejecting the null hypothesis (Table 3). We can thus sum up the results and give them the following economic interpretation: 1) it is more probable that the exchange rate dynamic is the determinant of movements connected to the money in circulation than the other way around; 2) the exchange rate dominates the price dynamic; and 3) the causal chain between prices and money in circulation is not a clear and simple relationship. Thus we come to the conclusion that the causal direction, as Aftalion observed it and theoretically postulated it, in connection with the claim that the impulse begins with the exchange rate and then moves on toward prices and money in circulation, has essentially been confirmed by Granger's causality tests.

Table 3: *France: Granger's causality tests*

Granger's test of causality performed in pairs			
Sample: I 1920 – VII 1926			
Lags: 4			
Null hypothesis:	No.	F-Statistics	Probability
MF does not cause EF under Granger	75	2.08188	0.09308
EF does not cause MF under Granger		2.66357	0.04003
PF does not cause EF under Granger	75	2.58119	0.04513
EF does not cause PF under Granger		4.40683	0.00322
PF does not cause MF under Granger	75	4.20828	0.00428
MF does not cause PF under Granger		4.22419	0.00418

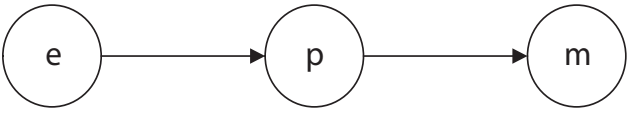
The data analysis results provide us with interesting information on the basis of which we can construct a VAR model that includes everything up to the fourth optimal lag of the variable set based on Akaike's and Schwarz's informational criteria (see Appendix). On the basis of this model's relatively acceptable statistical characteristics, in the next step of econometric analysis we create simulated impulse responses and a total probability of the simulated impulse responses (Figure 4); they confirm and "give validity" to Aftalion's theory on the hegemony of the exchange rate, since it turns out that prices react much more to exchange rate shocks than to those produced by money in circulation (column 2). Moreover, it is obvious that money in circulation has a relatively weak influence on the development and movement of prices and the exchange rate (column 3).

Figure 4: *France: Analysis of impulse responses (IR)*



In addition, the results of total probability analysis show the role of inertia and expectations in the process of determining prices and the exchange rate. For example, a total probability analysis regarding prices shows that, for about 62%, the explanation lies in past price values themselves, 36% are explained by the exchange rate, while only 1-2% are explained by money in circulation. Similarly, about 85% of exchange rate volatility happens automatically, 12% is connected to price movements, while only about 3% are a result of changes in money in circulation. A total probability analysis regarding money in circulation gives us econometric proof that the explanation for as many as 80% of the cases is in past values, about 11-12% are determined by price dynamics, while 8-9% are determined by exchange rate behavior.

The final step of econometric analysis allows us to "confirm" Aftalion's observations and theoretical arguments in connection to the causal chain between these three variables, mainly in the sense that the exchange rate dynamic is the main source of shock for the economy in the interwar period, that this is then passed on to prices and finally ends with changes and adjustments in the scale of money in circulation. These relationships can be illustrated with the following graph.



4.2. Econometric test for Bulgaria (1920-1924)

For the examined period (1920-1924) we use monthly price data from the annual statistical reports of the Kingdom of Bulgaria, which are available starting with 1922. In addition, we have analyzed data gathered from literature connected with time series of changes in the prices of food, heating and electricity in the 1922-1931 period (base idexes 1914 = 100),[71] as well as monthly price index data for the 1920-1927 period (base indexes 1913 = 100),[72] without any indication as to the contents and sources of the information. If we put them all on the same basis (1913 = 100) and compare them with official statistical data beginning with 1922, it turns out that three of the time series match, except for the year 1924, where Nedelchev's data somewhat differ. Thus it turns out that the price indicator for Bulgaria is the "index of changes in prices of food, heating and electricity in the 12 largest cities in the Kingdom of Bulgaria," since we shall be using Nedelchev's data for the 1920-1923 and official data for 1924.

It is much more difficult to find reliable data about the exchange rates of the lev and the US dollar in the examined period. Koszul uses average monthly data in the Federal Reserve Register, which has been published since 1922.[73] Using the same source of information, we managed to broaden the series back to July 1921, before which the available data covered only the minimum and maximum values. In the BNB's quarterly report there are data about the average monthly exchange rate for the period preceding the observed period – all the way up to 1919, where we can observe some differences between the two time series. As already said regarding the Bulgarian price data, if we put aside deliberate data manipulation (a typical phenomenon for the period before and after the stabilization), differences may also appear due to the rounding off of values, insufficient expertise of the statistics staff in the BNB, or printing errors.[74] In seeking to cover the longest possible period, we were faced with limitations in the use of BNB data, if we assume that the economic agents in the country truly made decisions on the basis of official sources of information. It is interesting that the BNB tried to underestimate the nominal appreciation and to emphasize the devaluation of the national currency.

---

71 J. Koszul, *Les efforts de restauration financière de la Bulgarie (1922-1931)*, Félix Alcan, Paris, 1932.

72 K. Nedelchev, *Monetary Issues: Bulgaria 1879-1940*, Knipegraph Printing House, Sofia, 1940 (in Bulgarian)

73 J. Koszul, ibid.

74 For example, the average buy rate for May 1924 was 147.2, while the maximum and ninimum rate equaled 137.2. This is obviously a printing error, which explains why the buy rate is higher than the average monthly sell rate.

And, finally, in connection with the third variable (money in circulation), the differences between the available data series are relatively insignificant, despite the numerous approaches and attempts to define this indicator. When we take into account limitations regarding data availability, we are examining only the money in circulation because, for the period prior to 1922, we did not find the monthly statistics for this component when it comes to the BNB's obligations and burdens (demand deposits and correspondent accounts). Moreover, the official data published by the BNB[75] are identical to the money in circulation data given by Koszul, which means that we did not encounter problems in creating this historical time series for the examined period.[76]

Now we can turn to the testing of statistics, following in the steps of the analysis of the French pre-stabilization period. Statistical volatility throws up to the suface precisely the movements of the exchange rate, although not as obviously as in the French case. The biggest departure from the mean rate was that of the exchange rate (0.11), followed by prices (0.04), while money in circulation was once again in the last place (0.02), if we take variables such as the first difference in the logarithms.

It is interesting that the correlation matrices, especially when the variables turn into the first difference in the logarithms, say that the correlation between the exchange rate and money in circulation is stronger (Table 4). It should also be mentioned that the correlations vary greatly for different subsamples, which reflects the high dynamics and differing characteristics of the market mechanisms, state intervention levels, etc. For example, if we shorten the observed period by eliminating the observations regarding the period preceding 1921, we get a closer correlation between the exchange rate and the price dynamics, both in terms of levels and in terms of the first difference in the logarithms.

Table 4: *Bulgaria: matrices of variable correlations*

In levels	Exchange rate	Money in circulation	Price level
Exchange rate	1.000000	0.793905	0.780339
Money in circulation	0.793905	1.000000	0.691938
Price level	0.780339	0.691938	1.000000

First difference in the logarithms	Exchange rate	Money in circulation	Price level
Exchange rate	1.000000	0.141123	0.000123
Money in circulation	0.141123	1.000000	0.019133
Price level	0.000123	0.019133	1.000000

[75] The monthly data about money in circulation for the period before 1928 was recorded and generously made available to us by Martin Ivanov.

[76] J. Koszul, ibid.

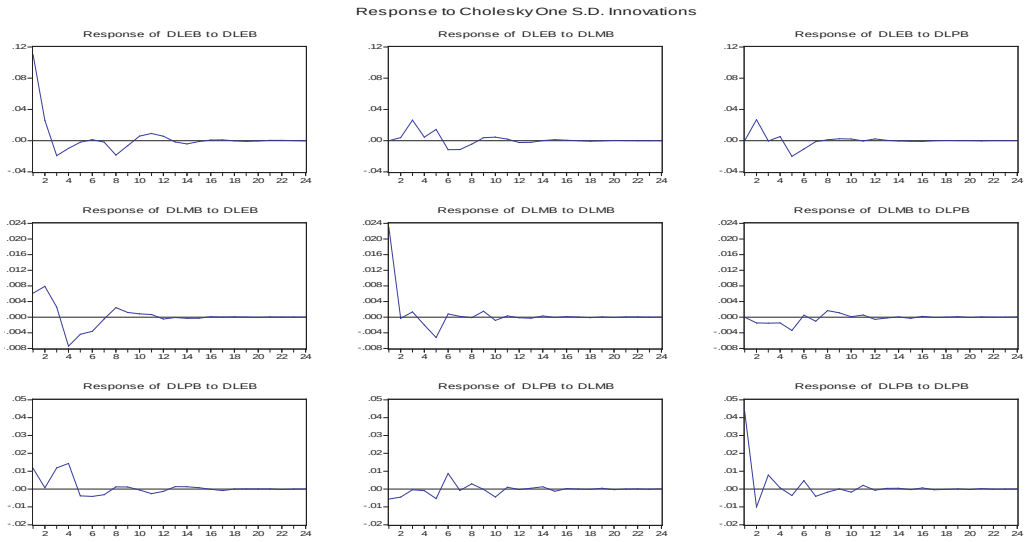
Granger's causality test also shows some differences in relation to the dynamic of the French monetary variables. If we take a somewhat more liberal starting position (applying the acceptance/rejection criterion with a probability of 0.10), we can say that: 1) it is obvious that the exchange rate is the cause of the price movement; 2) the exchange rate also reflects the money in circulation dynamic; and 3) there is a probability that the money in circulation indluences price changes. What can in no case be excluded as an option is that the dynamic of the exchange rate directly influences prices.

Table 5: *Bulgaria: Granger's test of causality*

Granger's test of causality in pairs			
Sample: I 1920 – V 1924			
Lags: 2			
Null Hypothesis:	No.	F-Statistics	Probability
PB does not cause EB under Granger	51	0.43735	0.64840
EB does not cause PB under Granger		4.95947	0.01121
MB does not cause EB under Granger	51	2.69840	0.07797
EB does not cause MB under Granger		2.29515	0.11217
MB does not cause PB under Granger	51	2.07661	0.13695
PB does not cause MB under Granger		0.07956	0.92365

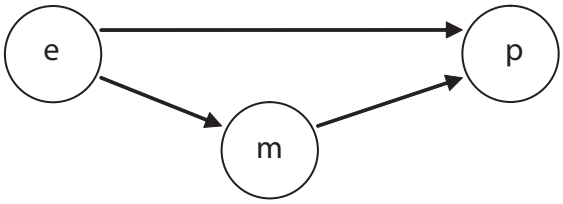
Despite the relatively small sample of 53 monthly observations (there were 79 for France), the VAR models give us pretty interesting results (see Appendix). First, the analysis of impulse responses gives an indication of the weak influence of money in circulation on the mechanism for setting prices and the exchange rage (column 2 in Figure 5). All three variables are, before all, self-determined, which implies either a high level of independence (large inertia) or disrupted relationships due to some sort of interference or intervention on the part of the administrative mechanism. What is essential in the Bulgarian case is the relatively strong response of both the money in circulation (column 1, row 2) and prices (column 1, row 1) to the shocks coming from the exchange rate, which only speaks in favor of the hegemony of the exchange rate. Still, the IR analysis shows that, on the ladder of causality, the quantity of money in circulation lies somewhere between the exchange rate and prices.

Figure 5. *Bulgaria: Impulse response analysis*



A total probability analysis provides us with quantitative estimates regarding the relationship of causality among variables. The results show that 74-75% of the price variations were determined by their past values, 19% by exchange rate volatility and 7% by movements in the money in circulation. The total probability analysis of the exchange rate (row 1), besides a pretty high level of self-induced variations (84-85%), was equally influenced (about 8%) by (expected) movements in the price levels and the money in circulation. The econometric evidence of the total probability analyis of the money in circulation show that the direct exchange rate influence came to about 25%, which is much more significant that the influence of inflation, which equaled about 3%. Thus it turns out that the money in circulation is obviously a function of the exchange rate and not the other way around. To summarize, the results show that the exchange rate was the second most important variable in explaining movements in the money in circulation and in prices, while first place was taken by their strong dependence on their own "past behavior." Of course, if we except their dependence and limitation to a small sample, a reasonable economic explanation could be found for these estimates, in terms of the pretty "non-market" determination of the exchange rate in the pre-stabilization period in Bulgaria (the BNB's monopolistic interventions in currency exchange activities), as compared to the French pre-stabilization period. Similar causality test results are achieved with VAR models when we shorten the observed period from the end of 1923, when exchange rate controls were *de facto* introduced and movements in the exchange rate began to greatly reflect the central bank's interference and interventions.

Even though the exchange rate was the cause of price movements, they were nevertheless, also determined by the money in circulation. This provides a basis for claims that the exchange rate influences prices in two ways: 1) directly and 2) indirectly, via money in circulation – and, in both cases, this occurs by way of expectations. Thus, the causal chain in the pre-stabilization period in Bulgaria can be illustratied with the following graph.



Since Aftalion's theory and statistical tests have been "applied" to Bulgarian monetary history by his contemporaries, it would be interesting to compare their results with those obtained on the basis of VAR techniques. All in all, the estimates of Koszul and Petkof[77] are very similar to our own. According to the former, the analyses of the curves and correlations connected with money in circulation give an unambiguous result, which says that, in the period between 1920-1924, the causal chains began with the exchange rate, as the significance of money in circulation grew with time. Petkof's statistical estimates are even closer to our own. That author claims that, during the entire observed sample (1920, 1921-1924), the exchange rate dictated movements in the field of prices and money in circulation, since he observed the following causal chain during the period of economic rise: "exchange rate-money in circulation-prices," while during the period of economic decline he registered two chains: "exchange rate-prices" and "money in circulation-prices."[78] Finally, Petkof summarizes by saying that there were two price determinants: the exchange rate and money in circulation.[79]

It is interesting to observe that, on the basis of out tests; the causal chains in Bulgaria are similar to those identified by Aftalion in Germany. This similarity does not surprise if we know that both countries were on the losing side in Wold War I and that it is very probable that the debt burden and the expectations connected with debt payments influenced the dynamic of the variables and led to precisely this causal relationship. Under such a state of affairs, the exchange rate influences prices by way of two channels: directly and indirectly through money in circulation.

77 J. Koszul, op. cit., pp. 120, 121; J. Petkof, op. cit., pp. 187-191

78 J. Petkof, p. 112

79 Ibidem, pp. 141-148; 361-377

5. Concluding remarks

The interwar stabilization had a great number of points in common with the modern concepts that dominate today's discussions on the efficiency of monetary regimes, especially with those based on the idea of stabilizating the exchange rate, convertibility or on rules in general instead of on discretionary decisions. Such monetary regimes have a clear advantage when it comes to raising the credibility of the monetary authorities, which leads to an increase in discipline. In this paper, we dealt with a comparative analysis of the stabilizations in France and Bulgaria. The two approaches to stabilization are logical replies to the specific pre-stabilization dynamics of the macroeconomic variables that are difficult to explain by way of conventional QTM and PPP theories.

Albert Aftalion was among the first economists to emphasize the role of psychological factors, expectations and confidence within analyses of inflatory trends, exchange rates and money in circulation. This psychological theory of money and the exchange rate is in many aspects a forerunner of the modern understanding of the process of price and exchange rate determination process (the role of expectations, overshooting effects, the *Random Walk* – the chance movement of prices without influence from the past, multiple balance dynamics, self-fulfilling prophecy, etc.). Aftalion built his theory on the basis of observations of pre-stabilization experiences of several European countries (Bulgaria was not included in his sample of empirical illustrations) and the application of basic statistical methods (which sometimes leads to imprecise interpretations of correlations, such as, for example, causality). Today's econometric techniques allow us to carry out new tests on the validity of Aftalion's theory of that time. Despite the structural differences between France and Bulgaria in the pre-stabilization period, taking into account that Bulgaria saw a high level of governmental administrative interference in monetary questions, Aftalion's theory still proves that, all in all, it fits into all the facts.

The role of expectations is very important, because money in circulation is no longer the leading indicator of prices and exchange rate movements, but is, instead, most frequently the final element in the causal chain of monetary relations. In both cases (France and Bulgaria), the exchange rate is in focus when it comes to the expectations of monetary factors ("agents"). It's *de facto* and, later, *de jure* stability, along with the convertibility of the national currency, is seemingly the starting position of the entire stabilization process. The specific characteristics of the stabilization process in the two countries are a reflection of the differing ideological paradigms in which the respective stabilizations were carried out. These two examples illustrate the significance of the effect of discipline (the control of money in circulation) in the debate regarding the stabilization in Bulgaria, as well as the dominant role of the effect of convertibility which was adamantly supported and championed by

all the politicians and economists involved in the process of stabilization in France. Despite the differences, the stabilization processes in both the cases were carred out as an answer to similar problems of monetary dependence and limitations and, in both cases; the outcome lay in the decisive role of the exchange rate and expectations.

References

- Aftalion, A.: *Existe-t-il un niveau normal du change, Revue Economique internationale*, vol. 4, 1925, pp. 423-450.
- Aftalion, A.: *Monnaie, prix et change. Expériences récentes et théorie*, Recueil Sirey, Paris, 1927.
- Aftalion, A.: *L'or et la monnaie. Leur valeur. Les mouvements de l'or*, Les éditions Domat-Montchrestien, Paris, 1938.
- Berov, L.: *The Bulgarian National Bank at 120*, 1997, unpublished document (in Bulgarian).
- Bainville, J.: *Les conséquences politiques de la paix*, Gallimard, Paris, 1920. [2002]
- Blancheton, B.: *Le Pape et l'Empereur. La Banque de France, la direction du Trésor et la politique monétaire de la France (1914-1928)*, Albin Michel, Paris, 2001.
- BNB: *Annual report for 1929*, 1930 (in Bulgarian).
- BNB: *The motives for the draft Law on the exchange rate and backing for the money of the BNB, with the draft submitted text*, The Central Bureau of Public Records, Bulgarian National Bank, document collection, vol. 4, 1915-1929, document 2, 2001, pp. 55-56 (in Bulgarian).
- BNB: *Protocol of the BNB managing council with a resolution that states that the exchange rate should be temporarily based on the French franc instead of the US dollar due to the dollar's significant depreciation on the international market,* The Central Bureau of Public Records, Bulgarian National Bank, Document Collection, vol. 4, 1930-1947, document 183, 2004, p. 419 (in Bulgarian).
- BNB: *Protocol of the BNB managing council, with a statement of governor D. Bozhilov on the policy of the exchange rate and the clearing agreements with certain countries im zemljama and a resolution on amendments to general import and export acts*, The Central Bureau of Public Records, Bulgarian National Bank, document collection, vol. 4, 1930-1947, document 295, 2004.a, pp. 557-562 (in Bulgarian).
- BNB: *Indictment against: Kiril Gunev, Dr. P. Kyosev, G. Proichev, Sv. Toshev, Dr. N. Hristov and St. Stoichev, along with a decision to prosecute according to the Law on the Prosecution of those Responsible for Draw-*

ing Bulgaria into War against the Allies, The Central Bureau of Public Records, Bulgarian National Bank, document collection, vol. 4, 1930-1947, document 566, 2004.b, pp. 945-952 (in Bulgarian).

- Bordo, M. – Kydland, F.: "The Gold Standard as a Commitment Mechanism," Bayoumi, T. – Eichengreen, B. – Taylor, M. eds.: Modern Perspectives on the Gold Standard, Cambridge University Press, 1996.
- Bordo, M. – Frandreau, M.: "Core, Periphery, Exchange Rate Regimes, and Globalization," Michael D. Bordo, Alan M. Taylor, Jeffrey G. Williamson, eds.: *Globalization in Historical Perspective*, NBER conference held May 4-5 2001.
- Boshnyakov, D.: *Prices and the Monetary Problem*, Razvitie, Sofia, 1936 (in Bulgarian).
- Burilkov, Zh. T.: *The Monetary Reform*, "S. M. Staikov," Sofia, 1928 (in Bulgarian).
- Chapkunov, A.: *Crises and Foreign Currency*, Hudozhnik, Sofia, 1936 (in Bulgarian).
- Dangel-Hagnauer, C. – Raybaut, A.: "Monetary reform in France: the French economists and the stabilization of the franc in the 1920s," Barens, I. – Gaspari, V. eds.: *Political Events and Economic Ideas*, Edward Elgar, 2004.
- Eichengreen, B. – Flandreau, M. eds.: *The gold standard in theory and history*, Routledge, London and New York, 1997.
- Eichengreen, B.: *L'expansion du capital. Une histoire du système monétaire international*, L' Harmattan, 1997 [1996].
- Hawtrey, R.: *The Art of Central Banking*, Longmans, Green and Co., London, 1932.
- Ilief, P.: *La banque nationale de Bulgarie et l'histoire de sa politique monetaire*, Lyon, 1930.
- Ivanov, A.: *Report of BNB governor Asen Ivanov at the five-year celebration of the founding of the BNB*, The Central Bureau of Public Records, Bulgarian National Bank, Document Collection, vol. 3, 1915-1929, document 22, 1929, pp. 136-144 (in Bulgarian).
- Ivanov, M.: *Political Games with the external Debt. Bulgarian Scenarios of Economic Crises and Upsurge, 1929-1934*, "Zlatyo Boyadzhiev," Sofia, 2001 (in Bulgarian).
- Ivanov, M: "Could We Devaluate? What is There behind the Orthodox Bulgarian Answer to the Great Depression?," *Historical Review*, no. 3, 4, 2005 (in Bulgarian).
- Jacob, D.: *La Banque centrale et le secret*, Thèse pour le doctorat ès sciences économiques, Université Montesquieu-Bordeaux IV, 1996.
- Kemilev, A.: "Money Circulation in Bulgaria," *Journal of the Commercial College*, Varna, vol. I, edition 4, 1936, pp. 287-300 (in Bulgarian).

- Kemp, T.: "The French Economy under the Franc Poincare," *Economic Historical Review*, new series, vol. 24, no. 1, 1971, pp. 82-99.
- Keynes, J. M.: *A Tract of Monetary reform*, MacMillan, London, 1923.
- Keynes, J. M.: *Reflexions sur le franc et sur quelques autres sujets*, Les documentaires, Paris, 1928.
- Kindleberger (1990.) [1984.]. Histoire financière de l'Europe occidentale, Economica, Paris.
- Kindleberger: *La crise mondiale 1929-1939*, Economica, Paris, 1988 [1973, 1986].
- Koszul, J.: *Les efforts de restauration financière de la Bulgarie (1922-1931)*, Félix Alcan, Paris, 1932.
- Monchev, B.: *The Monetary Problem after the World War*, Sofia, 1939. (in Bulgarian).
- Mouré, K.: "Undervaluing the Franc Poincare," *Economic Historical Review*, New Series, vol. 49, no. 1, 1996, pp.137-153.
- Nedelchev, K.: *Monetary Issues: Bulgaria 1879-1940*, Knipegraph, Sofia, 1940 (in Bulgarian).
- Nikolov, G.: *Stabilization: Monetary and Economic* (Thoughts on the Monetary and Economic Crisis), "Elisei Petkov," Sofia, 1927 (in Bulgarian).
- Nogaro, B.: *La monnaie et les phénomènes monétaires contemporains*, Marcel Giard, Editeur, Paris, 1924.
- Nenovsky, N.: *Exchange rate Inflation. The contribution of Albert Aftalion (1874-1956)*, Edotions of the Bulgarian National Bank, 2006.
- Petkof, J.: *Prix, circulation et change en Bulgarie de 1890 a 1924*, Jouve & Cie Editeurs, Paris, 1926.
- Raybaut, A. – Torre, D.: "Discipline, confiance et stabilité des régimes de caisse d'émission en transition vers l'Euro," paper presented in Marrakesh, Marocco, 16-17 March 2005.
- Rist, Ch.: "Pour la stabilisation du franc," Rist, Ch.: *Essais sur quelques problèmes économiques et monétaires*, Recueil Sirey, Paris, 1933, [1925].
- Rueff, J.: *Théorie des phénomènes monétaires*, Payot, Paris, 1927.
- Sarailiev, G.: *Monetary Devaluations and Their Effect at Home and Abroad*, Sofija, 1937 (in Bulgarian).
- Sauvy, A.: *Histoire économique de la France entre les deux guerres*, Economica, Paris, 1984.
- Simmons, B.: "Rulers of the game: central bank independence during the interwar years," *International Organization*, vol. 50, no. 3, Summer 1996, pp. 407-443.
- Spanos, A. – Andreou, E. – Syrichas, G.: *A VAR model for the monetary sector of the Cyprus economy*, University of Cyprus and the Central Bank of Cyprus, 1997.

- Stoyanov, N.: *Reparations and Inter-Allied Debts. Bulgarian Government Debts*, Sofia, 1933 (in Bulgarian).
- Toshev, G.: *Foreign Currency and Foreign Exchange Policy. A Scientific Inquiry into the Reasons of the Economic Crisis*, Franklin Cooperative, Sofia, 1928 (in Bulgarian).
- Totev, T.: *The Money. Coins – Banknotes – Paper Money*, Zadruzhen Trud, Sofia, 1932 (in Bulgarian).
- Vallance: *Histoire du franc*, Champs, Flammarion, Paris, 1998 [1996].
- Whale, P.: "The working of the prewar gold standard," Eichengreen, B. – Flandreau, M. eds.: *The gold standard in theory and history*, Routledge, London-New York, 1997.
- Yurii, *Monetary Crisis Liquidation (the Appreciation of our Lev)*, Pridvorna, Sofia, 1923 (in Bulgarian).

Appendix

VAR Model in France

Estimate of the vector of autoregression			
Date: 08/08/06 Time: 14:06			
Sample (adjusted): 1920:06 1926:07			
Inclided observations: 74 after the final adjustment			
Standard errors in ( ) & t-statistics in [ ]			
	DLEF	DLPF	DLMF
DLEF(-1)	0.134482	0.011680	0.018543
	(0.15956)	(0.09593)	(0.03860)
	[0.84283]	[0.12175]	[0.48037]
DLEF(-2)	-0.266052	-0.222034	-0.010792
	(0.15529)	(0.09336)	(0.03757)
	[-.71324]	[-2.37816]	[-0.28727]
DLEF(-3)	0.043869	-0.024333	0.012586
	(0.15030)	(0.09036)	(0.03636)
	[0.29187]	[-0.26928]	[0.34614]
DLEF(-4)	-0.047258	-0.092423	0.008085
	(0.14425)	(0.08673)	(0.03490)
	[-.32761]	[-1.06569]	[0.23167]
DLPF(-1)	0.742948	0.589630	0.127069
	(0.26025)	(0.15646)	(0.06296)
	[2.85480]	[3.76848]	[2.01830]
DLPF(-2)	-0.418530	-0.017684	0.007236
	(0.28566)	(0.17175)	(0.06911)
	[-.46511]	[-0.10296]	[0.10471]
DLPF(-3)	0.045958	0.158871	-0.077449
	(0.29074)	(0.17480)	(0.07033)
	[0.15807]	[0.90889]	[-1.10115]
DLPF(-4)	0.134019	0.012063	0.013832
	(0.27123)	(0.16307)	(0.06562)
	[0.49412]	[0.07398]	[0.21080]
DLMF(-1)	0.458827	0.106168	-0.009510
	(0.54824)	(0.32961)	(0.13263)
	[0.83692]	[0.32210]	[-0.07170]
DLMF(-2)	0.366661	0.021685	0.124091
	(0.53723)	(0.32299)	(0.12997)
	[0.68251]	[0.06714]	[0.95480]
DLMF(-3)	-0.092247	-0.126553	0.213472
	(0.52357)	(0.31478)	(0.12666)
	[-.17619]	[-0.40204]	[1.68537]
DLMF(-4)	0.247051	0.502510	-0.000416
	(0.52339)	(0.31467)	(0.12662)
	[0.47202]	[1.59694]	[-0.00328]
C	0.009187	0.004769	0.003164
	(0.00835)	(0.00502)	(0.00202)
	[1.09986]	[0.94964]	[1.56608]
R-squared	0.323921	0.398689	0.234344
Adj. R-squared	0.190921	0.280399	0.083723
Sum sq. resids	0.232151	0.083914	0.013587
S.E. equation	0.061691	0.037090	0.014924
F-statistic	2.435508	3.370420	1.555849
Log likelihood	108.2825	145.9337	213.2996
Akaike AIC	-2.575202	-3.592802	-5.413503
Schwarz SC	-2.170434	-3.188034	-5.008735
Mean dependent	0.013873	0.005674	0.005169
S.D. dependent	0.068584	0.043723	0.015591
Determinant Residual Covariance		6.94E-10	
Log Likelihood (d.f. adjusted)		465.2493	
Akaike Information Criteria		-11.52025	
Schwarz Criteria		-10.30595	

VAR Model in Bulgaria

Estimate of the vector of autoregression			
Date: 08/08/06 Time: 14:11			
Sample (adjusted): 1920:06 1924:05			
Inclided observations: 48 after the final adjustment			
Standard errors in ( ) & t-statistics in [ ]			
	DLEB	DLPB	DLMB
DLEB(-1)	0.152272	0.045539	0.076154
	(0.15511)	(0.06360)	(0.03339)
	[0.98172]	[0.71605]	[2.28052]
DLEB(-2)	-0.313069	0.109705	0.015853
	(0.15772)	(0.06467)	(0.03396)
	[-1.98493]	[1.69640]	[0.46685]
DLEB(-3)	-0.191620	0.151034	-0.036271
	(0.16091)	(0.06598)	(0.03464)
	[-1.19086]	[2.28923]	[-1.04701]
DLEB(-4)	-0.147989	0.012393	0.029794
	(0.17155)	(0.07034)	(0.03693)
	[-0.86265]	[0.17619]	[0.80669]
DLMB(-1)	0.320343	-0.252603	-0.021176
	(0.73011)	(0.29936)	(0.15719)
	[0.43876]	[-0.84381]	[-0.13472]
DLMB(-2)	1.263009	-0.053182	0.015976
	(0.69362)	(0.28440)	(0.14933)
	[1.82088]	[-0.18700]	[0.10698]
DLMB(-3)	0.160999	-0.081902	-0.210192
	(0.71850)	(0.29460)	(0.15469)
	[0.22408]	[-0.27801]	[-1.35881]
DLMB(-4)	0.963770	-0.470828	-0.292166
	(0.70101)	(0.28743)	(0.15092)
	[1.37484]	[-1.63808]	[-1.93588]
DLPB(-1)	0.622371	-0.234208	-0.034306
	(0.39255)	(0.16095)	(0.08451)
	[1.58547]	[-1.45515]	[-0.40593]
DLPB(-2)	0.053459	0.089589	-0.092002
	(0.37463)	(0.15361)	(0.08066)
	[0.14270]	[0.58323]	[-1.14068]
DLPB(-3)	0.272731	-0.000270	-0.058845
	(0.35342)	(0.14491)	(0.07609)
	[0.77168]	[-0.00187]	[-0.77336]
DLPB(-4)	-0.259082	-0.207514	-0.068658
	(0.35277)	(0.14464)	(0.07595)
	[-0.73442]	[-1.43467]	[-0.90401]
C	0.008887	0.008602	0.007394
	(0.01820)	(0.00746)	(0.00392)
	[0.48832]	[1.15278]	[1.88726]
R-squared	0.290444	0.314350	0.309344
Adj. R-squared	0.047168	0.079269	0.072548
Sum sq. Resids	0.425355	0.071509	0.019715
S.E. equation	0.110241	0.045201	0.023734
F-statistic	1.193887	1.337202	1.306374
Log likelihood	45.31574	88.11014	119.0323
Akaike AIC	-1.346489	-3.129589	-4.418010
Schwarz SC	-0.839706	-2.622806	-3.911227
Mean dependent	0.014162	0.008986	0.004814
S.D. dependent	0.112936	0.047106	0.024645
Determinant Residual Covariance		1.20E-08	
Log Likelihood (d.f. adjusted)		233.4311	
Akaike Information Criteria		-8.101296	
Schwarz Criteria		-6.580945	

Original scientific paper
Received: September 19th, 2007
UDK 004.087.5

Assistant Professor Dušan Bulatović, PhD
Informatika a.d., Belgrade
Professor Milan Tuba, PhD
University of Belgrade, Faculty of Mathematics
Professor Dana Simian, PhD
Lucian Blaga University, Sibiu, Romania

VULNERABILITY OF SMART CARDS AND METHODS TO DEFETE ATTACKS

Abstract: *Smart cards are increasingly used to store confidential information, primarily secret digital keys. Financial institutions and commercial companies, as well as the end users, rely on smart cards security,. This paper gives a survey of possible attacks on smart cards, and its intention is to show that declared smart cards security should be taken with reserve. A classification of possible attacks is proposed, based on approach, required resources and probability of success. Attacks based on power supply analysis, differential analysis of errors, encryption time measurement and hardware attacks are considered. It is shown that the design and testing procedures of smart card security should be improved, and that application security should not depend on security of an individual card. Proposed measures to defeat possible attacks are classified according to degree of successfulness and implementation complexity.*

ACM *Subject Classification: C.3, E.3, K.4.4, K.6.5*
AMS *Subject Classification: 68P25, 94A60*
Keywords: *Smart cards, Code breaking, Security, Cryptography, Data encryption*

1. Introduction and definitions

Hundreds of millions of smart cards are being used today, in thousands of applications, including mobile telephony, paid TV programs, controlling computer access, storing medical information, identification cards, credit cards etc. As a rule, these cards are used for the execution of cryptographic transformations carried out under the control of secret keys that have been built into their protected memory. The goal of attackers is to extract these secret keys from a physically protected card, for the purpose of:

- modifying the card's contents;

- creating a duplicate card; or
- generating an unauthorized transaction.

As a rule, smart cards are cheap, easy to use, have a small memory, some basic computing capabilities and some input/output ports through which they receive power and synchronizational impulses (clock).

Ideally, smart cards are supposed to be resistant to all attempts to access their contents (tamper-proof), except by way of permitted access to cryptographic operations via the input/output ports, so that secret data from the card's memory cannot be extracted. The attacker can be:

- someone who has stolen a smart card;
- its legitimate user; or
- someone trying to access the card through ports.

In practice, there have been many successful attacks on smart cards. Here are some examples of scenarios of smart card use and the consequences in case the card is compromised and its protected contents exposed.

- A bank checks the authenticity of an ATM user, or a merchant checks the identity of a smart credit card user. The smart card holds the user's secret key and participates in the authenticity-checking protocols through the use of specific cryptographic operations. One of the threats is that an attacker steals a smart card and attempts to gain the legitimate user's secret key in order to withdraw money from his account. A second risk is that a malicious ATM, interacting with the card, attempts to get the key.
- When paying for a TV program, the raw TV signal is encrypted, while the smart card contains the secret key and the decryption algorithm code. This way it is easy to change the key and the decryption algorithm easily and cheaply. In this case, the card's end user himself can be the attacker and seek to access his own card's secret key. This would enable him to sell cloned cards on the black market, which actually happens in practice, and it is possible to obtain such cards.

There are two distinct types of attacks on smart cards.

1) **Active attacks**, in which the chip from the smart card can be extracted, modified, measured and partially destroyed, or used in a different environment. An active attack leaves clearly distinguishable traces behind, which means that this kind of attack is usually carried out on stolen cards or in situations when the card owner himself seeks to compromise its security. Active attacks include fault attacks, probing attacks, chip microsurgery with focused ion beam devices, etc. They usually require substantial amounts of time, sophisticated equipment

and detailed knowledge of the chip's physical design. Such attacks are exceptionally effective in revealing information about the smart card's system but, due to their cost, are rarely used as a way of obtaining an individual user's secret data.

2) **Passive attacks** in which the card can only be viewed externally during its normal interaction with the (perhaps modified) smart card reader. Such an attack is preferred when the smart card owner seeks to preserve its security, for example in financial applications. In case of discovering this type of abuse, the owner returns the smart card to the issuer for examination. Passive attacks include timing attacks, glitch attacks, and power analysis. These attacks require minimal investment and relatively scant acquaintance with the card, and can be launched against a large number of smart cards by using a small number of modified card readers.

The attackers themselves can also be classified:

Class I (clever outsiders)

These attackers are often highly intelligent but perhaps do not possess sufficient knowledge about the system. They have access only to moderately sophisticated apparatus and tools. Most often, they seek to exploit some known weakness of the system, but don't try to generate a new one.

Class II (insiders with technical knowledge)

These attackers have significant technical and specialized education and experience. They have varying degrees of familiarity with the system's components but have access to the majority of such information. They often have highly sophisticated tools and instruments for analysis.

Class II (organizations with large funds)

Such organizations can gather teams of specialists with complementary knowledge, supported by significant material resources. They are capable of making detailed system analysis, design sophisticated attacks and use the most up-to-date tools for analysis. They can also use Class II attackers as part of the team.

The critical question when it comes to smart card attacks is whether the attacker has unsupervised access to the card. If the answer is no, then fairly simple measures may prove sufficient. However, in a growing number of applications, the opponent is able to have totally unsupervised access, and not only to one cryptographic card but to many. That is why such systems are the targets of a growing number of well-prepared attacks.

2. Analyzing the card's power supply

Analyzing the card's power supply is a passive attack, very easy to execute but very hard to avoid. It is based on the observation that a detailed curve of power use by a typical smart card (which shows how the external power supply changes over time) contains a great deal of information about the operations being performed on the card. With sufficiently sensitive measuring equipment it is possible to observe the exact sequence of events (in the form of individual electrical circuits that open or close) during the execution of each instruction's microcode. For example, the profiles of power consumption during adding and subtracting operations are totally different, and the power consumption in the course of writing ones or zeroes into the memory can also be registered; thus, it is, for example, possible to visually obtain the secret key of an RSA operation on the basis of knowing when modular squaring and when modular multiplication are being performed. There are two types of attacks: Simple Power Analysis – SPA and Differential Power Analysis – DPA.

2.1. Simple Power Analysis

The idea of executing this attack is simple and consists of a passive observation of the card's power consumption over time and reaching a conclusion on the basis of this observation about the operations being performed on the card. For example, DES has 16 rounds and the power consumption curve will show 16 jumps. Each round uses a different portion of the secret key. The placing of bit 1 into the register has a different power consumption than the placing of bit 0. On the basis of this, the amplitude of the jump on the consumption curve will provide a great deal of information about the secret key that was used, and sometimes even the entire key.

Figure 1. *A card's power trace*

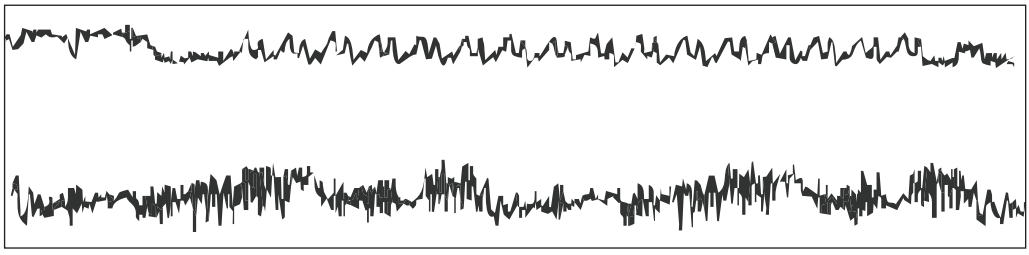


Figure 1 shows a card's power trace during the execution of a DES operation. The upper trace shows the entire DES operation, including the initial permutation, 16 DES loops and the final permutation.

For RSA and Diffie-Hellman-based cryptosystems, there is a very easy attack too: in these systems, we often take a givien **c** and raise it to **d** modulo **n**, where **d** is the secret key. The exponentiation algorithm is typically "repeated squaring," which cycles through the bits of **d** and conditionally multiplies by a running accumulator, then squares the accumulator:

```
// compute c^d mod p, for d >= 0
exponentiate(c, d, p):
total = 1
s = c
while(d > 0)
if d is odd:
total = total * s
d = d – 1
s = s^2
d = d / 2
return total
```

Each conditional multiplication in the algorithm represents an additional operation that demands extra power consumption, which means that secret data item ***d*** can be directly obtained by way of power supply analysis.

A countermeasure that would prevent such an analysis could be the unconditional execution of both the operations, after which the superfluous data item is discarded. The discarding operation always has the same power consumption regardless of the contents of the data being discarded. Shamir[1] also gave a hardware solution to this problem, in which one does not have to worry about the method of implementing the algorithm in order to avoid passive power supply analysis.

2.2 Differential power analysis

Differential power analysis is much more dangerous: the attacker analyzes a great number of power supply traces recorded during multiple algorithm executions with different inputs, and by using statistical analysis on enough measurements; the secret key can be recovered.

The differential power analysis method was discovered by Kocher, Jaffe and Jun.[2] It can also discover secret data in cases when the data about the algorithm obtained by way of power supply analysis are represented in a much less

1 A. Shamir, "Protecting Smart Cards from Passive Power Analysis with Detached Power Supplies." http://jya.com/dfa.htm , 2000.

2 P. Kocher, J. Jaffe, B. Jun, "Introduction to Dfferential Power Analysis and Related Attacks." http://www.cryptography.com/dpa/technical/index.html , 1998.

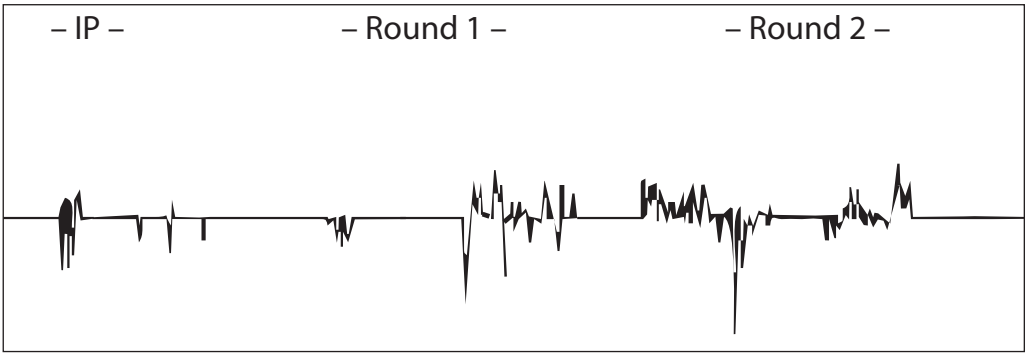
clear way. The idea behind DPA is to feed many slightly different inputs (say, differing in 1 bit) to the smart card and observe the differences in power usage over time for the different inputs.

Since the cryptographic algorithm is known to the attacker, he can monitor how the variation in one bit spreads through the algorithm and interacts with parts of the secret key. By using statistical analysis of the results of a large number of measurements, one can get the secret key. Moreover, the attack can be automated and generally requires only a thousand sample measurements.

Figure 2 shows the difference in power consumption by changing one bit of the source text. In the initial permutation (IP), this bit is entered into the register and its influence is felt regarding the power supply of the first algorithm round, and pretty strongly regarding the second round. The result was obtained after 1,000 measurements, although the variations could be observed much earlier.

The authors of this attack publicly announced that, through the use of DPA technique, they succeeded in cracking practically all types of smart cards that had been used by financial institutions up to that point.

Figure 2. *Differential power analysis*



2.3 Techniques of protection against power analysis attacks

While protection from simple power analysis is possible by using certain simple and generally successful countermeasures, it is much harder to protect hardware from DPA attacks. Attempts to produce a smart card with absolutely uniform power consumption by changing its design have not succeeded, for even small power supply variations can be detected with the help of a sensitive digital oscilloscope. In addition, forcing the same number of transistors to turn on for each instruction is a totally unnatural operation, which increases total consumption and slows down operations.

One proposed solution consists of adding a capacitor to the card's power supply lines that would mitigate the power traces. However, physical limita-

tions to the size of the capacitors prevent them from having sufficient capacity, which means that enough non-uniformity of power supply is left to allow a DPA attack to occur.

A related technique is to add to the smart card chip a sensor which measures the actual current supplied to the chip, and tries to actively equalize it by controlling an additional current sink. However, the local changes in the power supply curve are so rapid that any compensation technique is likely to lag behind and leave many power spikes clearly visible.

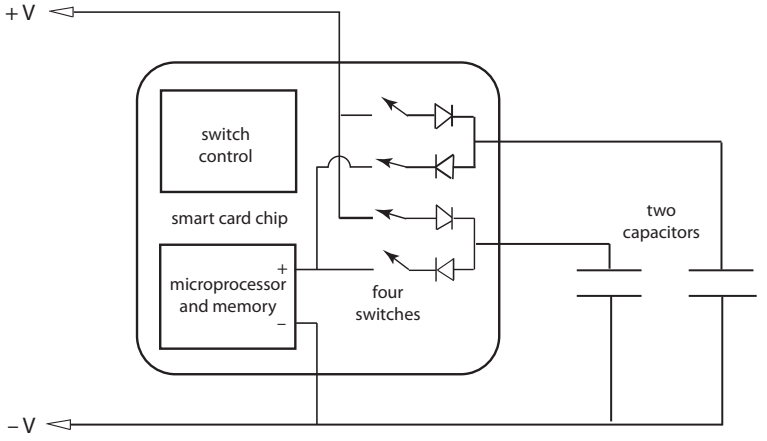
Other techniques propose the use of hardware random noise generators, the execution of unusual instructions, parallel execution of multiple instructions, etc.

However, hardware noise can be eliminated by observing multiple power traces and taking the average value, while software techniques to not help if individual instructions can be followed, which means that the effect of these measures is also limited.

A solution requiring an exchange of an external power source for a battery-powered card would ensure that power supply could not be measured from the outside. However, the card is only 0.76 mm thick, which means that such a thin battery would be too expensive, too short-lived and hard to replace, making it impractical. An alternative variant using a battery which is recharged each time the card is placed into the reader also has shortcomings, because such batteries easily dry out and during of the card's normal use – which is occassional and random – the recharge time would be intolerably long and the battery would have to be replaced after only several hundred uses.

Shamir[3] proposed a solution by which the power source would be totally separated from the card by way of two capacitors (Figure 3).

Figure 3. *A smart card with a separate power supply*



[3] A. Shamir, ibid.; http://jya.com/dfa.htm

During the first half of the cycle, capacitor 1 is charged by the external power supply, while capacitor 2 is discharged by charging the card, while in the second part of the cycle their roles are reversed. This solution prevents a DPA attack outside the card, but if the chip is physically opened and measured (more on these techniques later), then attack is still possible.

3. Differential fault analysis

The fault analysis model was introduced by Boneh,[4] and subsequently analyzed by others. It assumes that – if a chip is exposed to low-level ion radiation or some other influence with a similar effect – that the fault generated in one bit can compromise sensitive material, especially the material of the key.

Anderson and Kuhn[5] described how such an attack can be successfully mounted even by amateur hackers on, for example, pay-TV cards. The idea is to apply a rapid transient of the power supply or clock (a much faster clock than normal) to the chip, which will affect the functioning of only some of the chip's transistors. By varying the parameters, the CPU can be brought into a state of executing totally wrong instructions, sometimes even instructions not supported by the microcode. Although it cannot be known in advance which transient will cause what error, systematic fault analysis can be carried out in a relatively simple manner.

A typical subroutine found in security processors is a loop that writes contents of a limited memory scope on a serial port:

```
1 b = answer_address
2 a = answer_length
3 if (a == 0) goto 8
4 transmit(*b)
5 b = b + 1
6 a = a - 1
7 goto 3
8 ...
```

The idea is to find a transient that normally increases the program counter, but also transforms the conditional jump on line 3 or the loop decriment on line

[4] D. Boneh, R.A. DeMillo, R.J. Lipton, "On the Importance of Checking Computations." preprint, 11/96.

[5] R. J. Anderson, M.G. Kuhn, "Tamper Resistance - a Cautionary Note." The Second USENIX Workshop on Electronic Commerce Proceedings, 1996.

6 into something else. After this, by repeating this same transient, the entire contents of the memory can be read.

Output loops are only one of the targets of a transient attack. Other goals can be the checking of the password, right of access, or of the protocol itself, where the compromising of only one instruction can completely wipe out the protection. When the attack is applied to the algorithm itself instead of the control code, it can prove to be highly efficient, as will be seen below. An attack based on injecting a fault into the instruction code is more easily executed and gives the attacker more information than an attack on the data.

3.1. Attack on the RSA algorithm

One variety of attack can take place in the following way. Let's assume that the card computes the RSA digital signature **S** of message **M** over the module **n = pq**, by computing it over the module **p** and **q** separately, and then combines the result by using the Chinese Reminder Theorem. If the fault is injected into any of the last two computations, then **n** can be factored practically immediately. If **e** is a public exponent and if digital signature $\mathbf{S = M^d (mode\ pq)}$ is correct for module **p**, and incorrect for module **q**, then **p** is easily obtained as:

$$p = \gcd(n; S^e - M) .$$

This is an ideal situation for an attack by way of a power transient or impulse transient. As the card spends most of its time in computing digital signature ***mode p*** and ***mode q***, any transient that affects the result will do the job; the attacker doesn't have to make sure about selectively using the transient on some chosen instruction. The digital signature is used in communication only once, which means this attack can be launched online as well. A previously prepared terminal, in an operation belonging to whoever is launching the attack, can apply the transient to the card at the moment of payment, factor the modules, compute the actual digital signature and complete the transaction in real time. In this way, attackers can retrieve RSA secret keys in such a way that neither the card's user nor the bank have any data directing them to the place where the secret key was cracked. If one knows that the average implementation of an EMV electronic vallet has about 10,000 different RSA keys per bank, it may be concluded that, in this way, attackers can get the opportunity to abuse the payment cards of a large user population.

3.2 Attack on DES algorithm

When the attacker has the option of choosing the instruction on which he'll induce a fault by way of a transient, this opens the way for several direct

attacks on the DES. It is possible in the last two rounds of the cipher coding loops to remove one of the 8-bit ***xor*** operations that are used to combine the loop keys with the S tables and to repeat this operation for all the key's bits. The wrong coding result obtained after this attack will differ from the original coded text in the output at the exit of usually two, and sometimes 3 S boxes. Through the use of the differential crypto-analysis technique, it is possible to get the information for five bits of the eight key bits that weren't used in the ***xor*** operation due to the injected fault. Thus, for example, six coded texts with the wrong last loop will give 30 key bits and, thus, make the keysearch simpler.

An even faster attack is to reduce the number of rounds in DES to one or two by corrupting the appropriate loop variable or conditional jump, as in the protocol attack described above. Then the key can be found by inspection. The practicality of this attack will depend on the implementation detail.

In this way, the DES can be attacked by way of one to the ten known coded texts. Still the question remains: how realistic is it to assume that the instruction chosen for attack will be chosen?

In most smart cards, the manufacturer provides a certain number of routines in ROM. Although occasionally presented as an "operative system," ROM code is primarily a library or an application kit that allows the programmer to control communications and other functions. These routines usually include the DES algorithm and, by purchasing the manufacturer's smart card development toolkit (usually at a price of several thousand dollars), the attacker obtains the complete documentation plus testing samples, which means that the possibilities for attacking individual instructions are real.

When it comes to an unusual application, there is a need for a bit more experimenting (which should be done in any case in order to obtain the necessary parameters of the transient being applied to the card). However the search space is relatively small, and on looking at a few DES implementations it becomes clear that we can usually recognise the effects of removing a single instruction from either of the last two rounds.

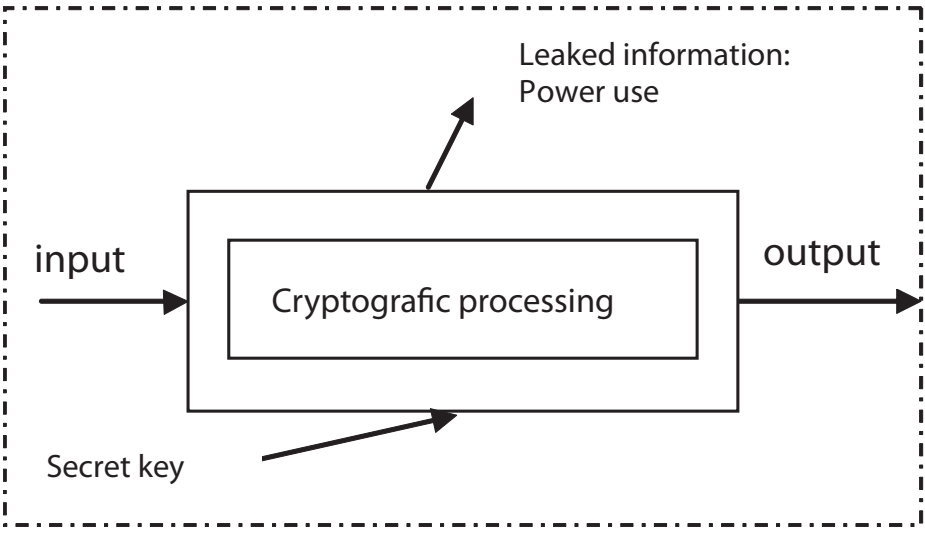
Power or impulse transients can always be applied as long as simple statistical tests don't suddenly show a great dependence between input and output bits of cryptographic transformation, which is an indication that there has been a reduction in the number of encryption loops. Such an approach can be practical even when the details of the implementation are unknown, which points to the conclusion that it is possible to use this technique for the purpose of reconstructing an unknown algorithm, such as Skipjack, without the need to use expensive chip testing equipment.

4. Timing attacks

Crypto-systems most often need different amounts of time to process different inputs. The reasons for this are many and depend on the existence of branches or conditional instructions in the algorithm, memory access, processor instructions (such as multiplication or division) that are not executed at a fixed time, and various other reasons. As a rule, performance depends on the key and the input data (e.g., plaintext or ciphertext). The time channel provides certain information outside the cryptographic card that may be sufficient to be used by the attacker in order to, with their help and performing measurements on sensitive systems, crack the entire secret key.

A timing attack is an attack outside the card (side-channel attack), as it is power analysis or analysis of emitted electromagnetic radiation (tempest). A cryptographic device (a card or a token) does not give only the result of the cryptographic operation at its output, but also other information that could prove useful as a basis for launching an attack (Figure 4).

Figure 4. *Real available information*



The basic idea of a timing attack is to measure the time a certain cryptographic operation lasts. Each cryptographic information that requires a different amount of time depending on the secret data that was used is a potential source of information about that data. For example, RC5 rotates bytes by a certain number of positions, which dependes on the secret key. If the smart card does not have a barrel shifter, the rotating operation must be executed like a set of one-place rotations in sequence. The time measurement gives the number of performed rotations, which allows the cracking of the secret key.

A second possible target of a timing attack is discovering the factoring of the RSA module $n = pq$. In case of a successful factoring of this module, the security of the RSA algorithm is broken and the secret key easily cracked. For efficiency, when operating on a value **y**, an implementation will typically reduce **y mod p** and **y mod q** and compute with those two values separately, then combine them later using the Chinese remainder theorem. These initial modular reductions can be vulnerable to timing attacks: choose an input **y** that is near **p**, and use the timing measurements to determine whether **y** is greater than or less than **p** (a division/subtraction step will be required if **y** is greater). When the smaller interval which contains number **p** is determined, using binary search one can learn **p** and discover **n**'s factorization.

Time measurement can also be used in the Diffie-Hellman key exchange protocol or in a full RSA protocol. In order for an attack to be possible, the victim must first execute several operations of modular exponentiation of the message over the module, during which the attacker, by way of passive surveillance, should get the message being exponentiated and the time necessary for the exponentiation. The algorithm of exponentiation which is, as a rule, "repeated squaring," examines bits of the key and conditionally multiplies them with the accumulator. Based on the time of each iteration one can get if bit from which depends whether the multiplication operation in the iteration is being performed is 0 or 1.[6]

The most obvious way to prevent timing attacks is to make all operations take exactly the same amount of time. Unfortunately this is often dificult. Making software run in fixed time, especially in a platform-independent manner, is hard because compiler optimizations, instruction timings, and other factors can introduce unexpected timing variations. If a timer is used to delay returning results until a pre-specified time, factors such as the system responsiveness or power consumption may still change detectably when the operation finishes.

Another approach is to make timing measurements so inaccurate that the attack becomes unfeasible. Random delays added to the processing time do increase the number of ciphertexts required, but attackers can compensate by collecting more measurements.

There are more efficient solutions. Techniques used for blinding signatures can be adapted to prevent attackers from knowing the input to the modular exponentiation function. Before computing the modular exponentiation operation, choose a random pair ($\mathbf{v}_i$; $\mathbf{v}_f$) such that $\mathbf{v}_i^{-1} = \mathbf{v}_f^x$ **mode n.** Before the modular exponentiation operation, the input message should be multiplied by $\mathbf{v}_i$ **(mod n)**, and afterward the result is corrected by multiplying with $\mathbf{v}_f$ **(mod n)**. The system should reject messages equal to **0 (mod n)**.

[6] P. Kocher, "Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems." *Proceedings of Crypto 96*, Springer-Verlag, 1996.

5. Hardware attacks on smart cards

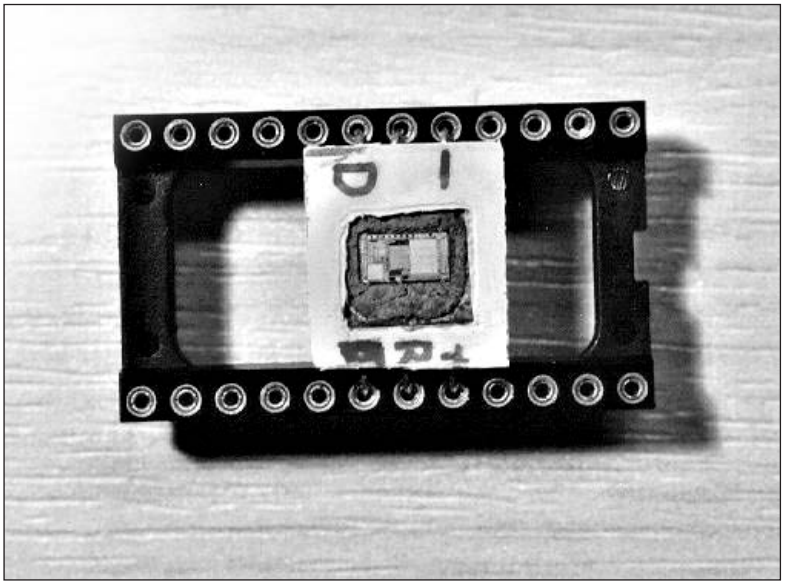
Physical resistance to attack (tamper resistance) is not absolute: an opponent with a semiconductor testing device can obtain data from a smart card key via direct observation and manipulation of the chip components. It is generally believed that, with sufficient investment, the security of any chip-sized device can be broken.

The physical attacks on some microcontrollers are almost trivial. For example, the lock bit of certain devices with a built-in EPROM can be erased by focusing UV light on the security cell, which is usually sufficiently distant from the rest of the memory.

Contemporary smart cards are somewhat harder to attack, but not too hard. As a rule, they have a bit of protection that prevents direct access to the silicone. A trained attacker with a large budget can accomplish a great deal. However, even a class I attacker, by using various chemical techniques, can remove the epoxy resin from the chip, remove the upper protection layers, cut the wires or directly access the internal wires with an instrument. Figure 5 shows a totally functional microprocessor from a smart card, whose plastic cover has been removed for the purposes of micro-experimenting. The tools necessary for preparing the chip in such a way were obtained in a pharmacy for $30.

As can be seen in the figure, the entire interior of the chip is exposed to observation and testing while the chip continues to function normally. Such an attack is discouraging: none of the data used by the microprocessor can remain secret.

Figure 5. *Physical attack*



By using more advanced techniques, a class III attacker can peel away one by one layer of print from the chip, scan the chip with a strong microscope and get to its entire design. There are also less invasive methods: under light of a certain wavelength, silicone becomes transparent and reveals the metal wires inside the chip.

It is very difficult to protect against such an attack. Basically, it is very hard to protect a chip's design from a motivated opponent with enough time and a large budget. Still, there are some strategies to make the attack more difficult:

- place the chip in a pressurized container and program the printed circuits to deactivate if loss of pressure is detected;
- embed sensors into the epoxy resin coating of the chip, which would detect damage to the resin and disable the chip;
- install a self-destruct mechanism that detects attempts to physically penetrate the chip.

A modern semiconductor laboratory equipped for reverse engineering of chips can cost millions of dollars. Still, as will be seen below, there are ways to attack smart cards open to attackers with much smaller budgets.

5.1. Attack on the chip's contents

Wherever the implementation of the algorithm on the card is known, there are multiple ways of extracting the keys from the card by targeting specific circuits or fuses, or by rewriting some of the memory contents. Bovenlander has described breaking smartcards by using two microprobe needles to bridge the fuse blown at the end of the card test cycle, and using the re-enabled test routine to read out the memory contents. Even where this method is not possible, the memory cells can still be successfully attacked, and this can be accomplished with a relatively small budget.

5.2. ROM overwrite attack

Single bits in a ROM can be overwritten using a laser cutter microscope, and where the DES implementation is well known, we can find one bit (or a small number of bits) with the property that changing it will enable the key to be extracted easily. The details will depend on the implementation but we might well be able, for example, to make a jump instruction unconditional and thus reduce the number of rounds to one or two. We can also progressively remove instructions such as exclusive-or's of key material.

Where we have incomplete information on the implementation, ROM overwriting attacks can be used in other ways. For example, DES S-boxes in ROM can be identified and a number of their bits overwritten such that the

encryption function becomes a linear transformation; we can then extract the key from a single plaintext/ciphertext pair.

5.3 EEPROM modification attack

When the algorithm is located in the EEPROM, the targeted bit can be set and reset with the aid of two microprobes.[7] This technique can also be used in the previously described attacks, but the ability to set and reset the bit opens up additional possibilities.

Since the DES algorithm uses a key with odd-even checking, if the number of key bits is not odd, correct implementation should report the error. Let's suppose that the location of the DES key in the EEPROM is known, but that it cannot be directly accessed due to a lack of necessary equipment. In that case, the attack can be mounted in the following way. The first bit of the EEPROM is set to 1 (or 0, the choice is irrelevant), and then if the algorithm continues to function properly, the bit of the key is 1. Then one moves on to the next bit and once again checks the functioning.

The tools necessary for this attack are two microprobes and a laser cutter, which can often be found in university laboratories, where even students can access them without authorization. Class I attackers can purchase such apparatus for several thousand dollars.

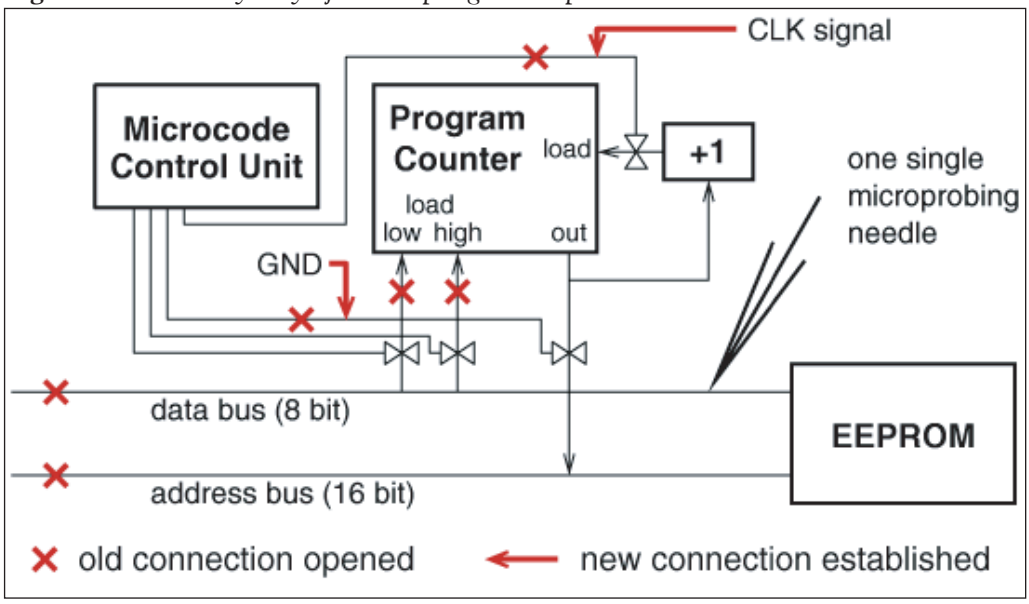
5.4. Attack by interrupting the chip's electrical circuits

Attacks mounted by pirates on certain pay-TV cards used a work station with a focused ion ray. This device can cut the wires on the metallized layer of the chip and form new wires or isolation layers. Such machines cost several million dollars, but attackers with a small budget can access them by renting them from various semiconductor companies.

Thanks to such tools, smart card attacks are becoming simpler and much more powerful. A typical attack includes the separation of almost the entire processor from the rest of the chip, leaving only the EEPROM and CPU components, which can generate read access. For example, the program counter can be left connected in such a way as to enable access to memory locations in sequence when a clock is applied to the device (Figure 6).

[7] O. Kocar, "Hardwaresicherheit von Mikrochips in Chipkarten." *Datenschutz und Datensicherheit*, vol. 20, no. 7. July 1996.

Figure 6. *Attack by way of interrupting the chip's electric circuits*



When this is done, the attacker needs only one microprobe or optical probe to read the EEPROM's entire contents. This allows a much easier program analysis than in passive attacks, because only one-time measurement is necessary. This also avoids mechanical difficulties tied to the simultaneous application of multiple probes on the chip, where the width of the wires, for example, is measured in microns.

6. Conclusion

Most smart card attacks, such as side-channel attacks (for example, power analysis or differential fault analysis), are *ad hoc*, just like the countermeasures taken against such attacks.[8] Such a state of affairs is unsatisfactory because we can never be sure whether the data in the card (or token) are at all protected.

We have seen that it is possible, with the aid of inexpensive and widely available devices, such as microprobes or laser cutters, to mount attacks on the chip surface, after which it is easy to obtain data about the secret key from the smart card processor, which we previously thought to be physically secure.

[8] M. Volar, D. Bulatović, "Primer implementacije VPN-a i smart kartica u zaštiti međubankarskih komunikacija." XVI Scientific-Professional Gathering INFO-TECH 2002, proceedings on CD,Vrnjačka Banja, 2002; D. Bulatović, D. Velašević, "Sistemi za detekciju i sprečavanje napada." Stručna knjiga, Beograd, June 2001, p. 202.

In addition, the redundancy of the key, such as DES parity bit, makes such attacks even easier.

Due to such circumstances, engineering practice should strive to avoid situations in which a security breech in one location compromises the security of the entire system (single points of failure). This is especially important due to the fact that the hardware resistance of smart cards and other commercial products was for a long time left unchecked. Hence, all claims about their security should be taken with reserve.

A new approach to physical security that takes into account the above-mentioned facts about smart card security should lie not in a demand to achieve absolute data security in the hardware but in ensuring that hardware that is secure to a certain extent will be used in a provably secure way. In any case, experience shows that the design of security systems must be executed very carefully, without neglecting their testing.

References

- Anderson, R. J. – Kuhn, M. G.: „Tamper Resistance – a Cautionary Note“, The Second USENIX Workshop on Electronic Commerce Proceedings, November 1996.
- Boneh, D. – DeMillo, R. A. – Lipton, R. J.: „On the Importance of Checking Computations“, preprint, 11/96.
- Bovenlander, E.: invited talk on smart card security, Eurocrypt 97.
- Bulatović, D. – Velašević, D.: „Sistemi za detekciju i sprečavanje napada“, Stručna knjiga, Beograd, June 2001, p. 202.
- Chaum, D.: „Blind Signatures for Untraceable Payments“, Advances in Cryptology: Proceedings of Crypto 82, Plenum Press, 1983.
- Kocar, O.: „Hardwaresicherheit von Mikrochips in Chipkarten“, *Datenschutz und Datensicherheit*, vol. 20, no. 7, July 96.
- Kocher, P. – Jaffe, J. – Jun, B.: „Introduction to Differential Power Analysis and Related Attacks“, http://www.cryptography.com/dpa/technical/index.html , 1998.
- Kocher, P.: „Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems“, Proceedings of Crypto 96, Springer-Verlag, 1996.
- Shamir, A.: Protecting Smart Cards from Passive Power Analysis with Detached Power Supplies, http://jya.com/dfa.htm 2000.
- Volar, M. – Bulatović, D.: „Primer implementacije VPN-a i smart kartica u zaštiti međubankarskih komunikacija“, XVI Scientific-Professional Gathering INFO-TECH 2002, proceedings on CD, Vrnjačka Banja, 2002.

Original scientific paper
Received: August 27th, 2007
UDK 339.56(5-15:6-17)

Assistant Professor Nataša Stanojević, PhD
Geoeconomics Faculty, Megatrend University, Belgrade

THE DISPROPORTIONS OF REAL AND POTENTIAL INTRA-REGIONAL TRADE IN THE REGION OF THE MIDDLE EAST AND NORTH AFRICA*

Abstract: *In each Middle Eastern and North African country a very low level of export to other countries in the region has been observed. A fuller use of export capacities could represent an adequate means of accelerating the growth of these economies. Through the use of the gravity model, as one of the more significant econometric models of external trade, this work seeks to determine the export potentials of each country in the region and to link them with the real volume of exports. This would determine the most desirable export directions, i.e. the regional markets with the most "space" for the products of a given state's economy.*

Key words: *external trade, market, geographic distance, export potential*

1. Introduction

In geo-economic divisions, the Middle East is most often viewed as a whole together with North Africa, due to their great natural, social, religious, political and cultural similarities. The region of the Middle East and North Africa consists of 18 countries, extends over an area of over 11 million square kilometers, has a population of about 300 million and is characterized by mixed economies. It, therefore, represents a large market, whose uneven and often slow economic growth could be stimulated by an increase in intra-regional trade among its countries.

The subject of this work is to determine the intra-regional export potentials of the countries in the region, primarily in the sense of more precisely determining their export direction and export quantities. By applying the gravity model on the mutual trade between each of the region's countries and

* At the request of the reviewers, this work has undergone a revision.

the others, we shall determine the export potential, i.e. the export value that the individual economies can absorb from the country whose export potential is being determined. A comparison of potential and real exports will clearly indicate the export directions, i.e. mark the countries with the freest "space" for imports from the analyzed export country.

Even though it is clear, even in the absence of a more detailed analysis, that these potentials are unused, results showing potential exports as being substantially greater than real exports would indicate that the countries in the region should more closely examine the needs of their neighboring markets and develop their production in accordance with them.

2. Theoretical postulates on the effects of geographic distances on economic development

Geographic distances have a great influence on transportation costs, whose reduction bears a positive influence on industrial exports and long-term economic growth. It is much more difficult for countries with high transport costs to advance their export-oriented development, even if they reduce customs duties, remove quantitative restrictions and adapt a macroeconomic policy of thrift.

An especially important influence on transport costs is waged by:

1) the proximity of transport routes and transport hubs, and
2) the proximity of large markets, i.e. economic centers.

The first assumption, on the influence of the proximity of transport hubs on economic growth, is based on the empirical results of extensive research done by Gallup and Sachs at the Harvard Institute for International Development.[1] The sample consisted of 92 developing countries and 15 countries with the greatest growth in (non-primary) exports in the world in the 1965-1990 period. The research emphasized the importance of having a geographical position near seas and rivers. Firstly, in the 92 developing countries 45% of the population lives near coastal areas, while in the 15 countries with a high export rate, it turned out that in all except one the percentage of the population living near coastal areas was much greater than average. Second, among the 15 most successful export countries, not a single one lacked an exit to the sea. Third, almost all the successful export companies are located on important water routes, in port cities or in export zones located near ports.

The second assumption, about the influence of large markets and economic centers, is also born out by much empirical evidence. Especially vis-

[1] John Luke Gallup and Jeffrey D. Sachs, *Geography and Economic Development,* Columbia University National Bureau of Economic Research (NBER), 1998.

ible are the negative ramifications of a geographical position outside of the existing economic centers. The costs of distance from established centers are a result, in the first place, of costs of trade with the center and of costs of receiving information and technologies from the center.

In their second research, Gallup and Sachs (1999) analyzed the influence of distance from existing centers. On a sample of 83 countries, they researched the relationship between earnings per capita in certain countries and their distance from one of three central regions (they measured distances from New York, Rotterdam and Tokyo). The results showed that as the distance doubled earnings dropped by about 25%.

Studies based on the gravity model, which we shall apply to the countries of the Middle East and North Africa, show that bilateral trade between countries depends significantly on geographic factors.[2] Geographical distance is negatively related to import and export volumes. Gravitational estimates of bilateral trade use distance as a representation of transport costs; thus, a common border is considered to be a great advantage to mutual trade. These studies show that the elasticity of trade volume in terms of distance amounts to between -1 and -1.3, which points to the great influence of geographical position on the direction of economic development and the level of development.

Limao and Venable[3] combine the results of the application of the gravity model with estimates of the elasticity of transport costs in relation to distance and conclude that the elasticity of trade volume in relation to transport costs equals about -2.5. Thus, with the doubling of transport costs, the volume of trade is reduced by about 80%. The results also show that the trade volume of countries without an exit to the sea is 40% lower than that of coastal countries.

The effects of trade reduction are the strongest in transport-intensive branches – those dependent on exports or imports of semi-finished goods. Radelet and Sachs (1999) discovered that an increase in transport costs in a given country from 12% to 17% reduces the long-term growth of the share of non-primary exports in the gross domestic product by about 0.2% per year.

Even when transport and communications costs fall with time, distance continues to wage considerable influence on local earnings. Improvements in transport and communications (the development of canals, roads, the airport network, the postal service, telegraph and telephone) do not remove the significance of distance. New technologies, just like trade policy liberalization, can certainly alleviate some of the transport costs, but many of their aspects nevertheless remain.

---

2 Frankel Jeffrey, Romer David, "Does Trade Cause Growth?", *American Economic Review*, vol. 89, no. 3, 1999.

3 Limao, Venable, *Infrastructure, Geographical Disadvantage, and Transport Costs*, World Bank Policy Research Working Paper No. 2257.

The standard framework for analyzing the directions of trade and, even more precisely, the potentials of trade flows, includes the application of the gravity model. In spite of numerous theoretical controversies, the gravity model has shown itself to be a quite applicable instrument for the explanation and prediction of bilateral trade. Trade theory can explain the causes of trade, but not its volume. This is why the gravity model takes into consideration a number of factors in order to explain trade volume as one of the aspects of international trade directions. It is based on economic sizes (most often taking into account GDP) and the distance between two countries or locations.

The model often includes variables such as price levels, common language, tariffs, colonial history (if country 1 had ever colonized country 2). The gravity model has often been used to test the efficiency of trade agreements and organizations such as NAFTA or WTO.

According to the principle of gravity convergence, bilateral trade between two regions or countries is directly proportional to their gross domestic product and inversely proportional to the distance between them. The application of this model to international trade dates back to Tinbergen (1962)[4] and Lineman (1966). Lineman added more variables and went further in the theoretical justification of the model. Lineman found that there are three main factors that should be taken into account in applying the gravity model in international trade. These are:

1) total potential supply (or exports) of a given country to the world market;
2) total potential demand (or imports) of a country from the world market;
3) factors resisting trade and those supporting trade intensity. This usually includes tariff barriers and transport costs.

Prominent among newer works are Winters' and Wang's analysis of Eastern Europe's potential trade directions ("Eastern Europe's International Trade") [5], as well as the works of Feenstra,[6] Deardorff,[7] and Frankel.[8]

4 Jan Tinbergen, *International Economic Integration*, Elsevier, London, 1954.

5 L. A. Winters, Z. K. Wang, *Eastern Europe's international trade*, Manchester University Press, Manchester and New York, 1994.

6 Robert C. Feenstra, James R. Markusen, Andrew K. Rose, "Using the Gravity Equation to Differentiate among Alternative Theories of Trade", *The Canadian Journal of Economics*, vol. 34, no. 2, 2001.

7 Alan V. Deardorff, "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?", in: J. A. Frankel, *The Regionalization of the World Economy*, Chicago: University of Chicago Press, 1998.

8 Jeffery A. Frankel, *Regional Trading Blocs: In the World Economic System*. Washington, DC: Institute of International Ecoomics, 1997.

To be sure, it should be emphasized that all factors affecting trade volume between two countries are not exhausted by those included in the gravity model (distance, population, GDP, common border, and joint membership in economic integrations). Trade is also affected by other economic as well as non-economic factors.

An important trade factor is the complementarity of economic systems, which, in this region, taking into account the similarity of structures in oil-based economies, does not contribute to mutual trade. On the other hand, the monolithic nature of oil-based companies would not present a problem if oil and gas were not exported in their raw states. If these countries were to develop the processing of oil and gas, which are raw materials for thousands of industrial products, this would create the conditions for an increase in trade in accord with the determined potential export directions.

On the other hand, there exist numerous non-economic factors that are often the main reason for the disproportion between real and potential exports. During periods of war and economic sanctions, to which the countries in the region are often exposed, the economic indicators do not reflect the true state of affairs, which means that data for those periods are usually excluded from the conclusions of performed econometric analyses.

3. The Most Important Characteristics of the Economies in the Region of the Middle East and North Africa

As regards types of economies, there are two distinct groups of countries in this region: the first group is made up of countries with large oil and gas reserves, which represent their main export articles and the bases of their economies, whereas the second group is made up of countries without natural resources.

The non-oil economies of the Middle East and North Africa have the usual problems faced by developing countries: slow economic growth (expressed in slow GDP growth) and weak integration into the global economy, i.e. a small volume of trade. From these stem a number of other weaknesses: unemployment, large external debts, dependence on foreign investments, low levels of modern technology use, poor industrialization, poor productivity, etc. From the beginning of the 1990s to today, the average increase in GDP in these countries has been below 4%, while GDP per capita increased by less than 2%, which puts these countries in the group of the world's slowest-growing economies.

Neither do most of the region's oil-based economies fall into the category of developed countries. On the contrary, their growth (measured by GDP growth) is often slower than that of countries without significant natural resources. Still, it cannot be said that these countries consistently suffer from low economic growth, as this parameter often reaches quite high levels, even for periods last-

ing several years. The main weaknesses associated with oil-based economies are in fact unstable growth and distorted sector structures, as a consequence of their dependence on raw oil and gas exports. If we add to this the fact that their energy reserves will be depleted in four to five decades, it becomes incumbent to find new bases for these countries' economic development.

Due to numerous natural-geographic and economic-geographic specificities, the countries in this region are not receptive to conventional development projects to a significant degree. The first thing to have in mind is that the lack of water and fertile land prevents agricultural development. In addition, the natural conditions are not favorable for most industrial branches, which also require great quantities of water for the production process itself. The oil industry in these countries has never taken root; instead, oil and gas are only exported, with the region's countries most often drawing only the rent from the oil wells exploited by companies from the developed West.

It is truly difficult to find an appropriate catalyst for economic growth in this region. Due to the previously mentioned shortcomings and limitations, we think that it is impossible to overrate the importance of increasing trade between Arab countries. The region's countries have a greater potential for intra-regional exports than for exports to developed European industrial countries, to which they can offer little outside of raw oil and natural gas.

The sum data for the region indicate that non-oil exports from the region during the past several decades have grown much more slowly than those in other regions. As a result, the region's total share in global exports has fallen by more than one half during the last two decades of the 20th century. The countries without oil and natural gas reserves have consistently highly unfavorable foreign trade balances.[9]

The small share of the non-oil-based economies in the region's total exports is something to be expected, as oil is the region's main product and has a greater value than other products coming from the region's countries. And the data about intra-regional exports[10] aren't encouraging either! According to some estimates,[11] were oil to be excluded from data concerning the region's exports, intra-Arab trade would make up 30% of their total external trade. According to UN data,[12] non-oil-based economies produce only a fifth of intra-regional exports, while their share in intra-regional imports equals 30% (on the level of the entire region).

Trade with Europe makes up three fourths of Algeria's total foreign trade and two thirds of Tunisia' and Morocco's trade. The same applies for Middle Eastern exports to Europe. The Levant countries – Israel, Jordan, Syria, and Lebanon, primarily trade with Europe and the sea routes connect them. For those Arab coun-

9 International Trade Center UNCTAD-WTO.

10 UN, ESCWA, *Great Arab Free Trade Area*, New York, 2001.

11 Ibid.

12 Ibid.

tries that do not produce oil, the region is an important, albeit unexploited export market. The non-oil-based economies' share of exports into the countries of the region in relation to their total exports is, on the average, only 17.7%. Jordan's exports to the countries in the region come to 47%, Lebanon's 45%, Syria's 20%, while in Egypt the figure is only 9.6% and in Yemen less than 2% (Table 1).

Table 1: *Intra-regional trade of Arab countries (1997)*

	Intra-regional exports			Intra-regional imports		
	Value in $ millions	Share within region %	Share in total exports	Value in $ millions	Share within region %	Share in total imports
Middle East region	10 465	100	8.6	11 122	100	10.8
GCC member-countries	8 291	79.1	7.6	8 251	69.2	8.5
Bahrain	516	4.2	10.2	2 147	3.8	8.9
Kuwait	391	3.8	2.7	1 067	11.4	12.9
Oman	762	7.3	10.2	1 473	15.8	29.3
Qatar	354	3.4	8.6	523	5.6	14.5
Saudi Arabia	4.381	42.2	8.3	1 638	17.5	6.0
U. A. Emirates	1 887	18.2	7.5	1 403	15.0	5.0
Diversified economies	2 174	20.9	17.7	2 871	30.8	9.3
Egypt	374	3.6	9.6	564	6.0	4.3
Jordan	714	6.9	47.5	918	9.8	22.4
Lebanon	289	2.8	45.0	619	6.6	8.3
Syria	785	7.5	20.0	269	2.9	6.7
Yemen	39	0.4	1.7	501	5.4	24.6

Source: ESCWA, Survey of Economic and Social Development in the ESCWA region, 1999

Considering its size and population, the region of the Middle East and North Africa has the potential to form a regional market. The governments of the countries in the region are only partially aware of the development potentials of exporting to neighboring countries. Although there have been several attempts to cash in on those potentials through the formation of regional integrations and bilateral agreements, no country has invested meaningful efforts in truly strengthening mutual ties and increasing the volume of trade within the said economic integrations. Various trade agreements between Arab countries have been signed, without, however, securing sufficient levels of integration.

4. The Export Reality and Potentials of the Countries of the Middle East and North Africa – Applying the Gravity Model

By using the gravity model we can determine the potential volume of mutual trade for each of the region's countries, i.e. their individual export potentials. By comparing the obtained results with the existing volume of trade, it is possible to identify potential directions for the development of their mutual trade.

We shall apply the gravity model to the bilateral trade of 18 countries in the region of the Middle East and North Africa. The aim is to determine each individual country's export potential to each of the region's other countries, and then to compare it to current exports. More simply put, the goal is to determine whether and by how much each of the region's countries can increase its exports to other countries.

The basic gravity model applied to international trade is:

$$\ln X_{ij} = \alpha + \beta_1 \ln Y_i + \beta_2 \ln N_i + \beta_3 \ln Y_j + \beta_4 \ln N_j + \beta_5 \ln D_{ij} + \beta_6 \ln A_{ij} + \sum_k \gamma_k \ln P_{ijk} + u_{ij}$$

Where:

X_{ij} = value of trade from country *i* to country *j*;
Y_i, Y_j = aggregate GDP of countries *i* and *j*,
N_i, N_j = total population of countries *i* and *j*,
D_{ij} = distance between countries *i* and *j*,
A_{ij} = *dummy* variable representing the common border of countries *i* and *j*,
P_{ijk} = *dummy* variable representing preferential arrangements, i.e. the fact that countries *i* and *j* belong to the same economic grouping;
u_{ij} = the model's random error.

The data on trade and GDP of the countries of the Middle East and North Africa are expressed in millions of US dollars. UN and World Bank data were used.[13] In order to avoid the effects of current imbalances and shocks, the three-year average gross domestic product was used, according to data for 2002, 2003 and 2004. The population is expressed in millions,[14] and the distances in kilometers.[15]

[13] GDP of countries in the region – World Bank – *Data – Key Development Data & Statistics*, data on mutual trade between the countries in the region – *United Nation Statistics Division – unstats.un.org/unsd/comtrade/default.aspx*

[14] Population data – UN Population Fund, The State of World population, New York, 2004.

[15] Data on distances between economic centers – *geography.about.com*.

Table 2: *Gross domestic product in the countries of the Middle East and North Africa* (Millions of US dollars)

	2002	2003	2004
Algeria	55,914	68,019	84,649
Bahrain	8,448	9,607	...
Egypt	75,148	82,427	89,854
Iraq	58,000		89,800
Iran	113,729	137,144	162,709
Israel	104,206	110,227	117,548
Yemen	9,900	11,002	12,834
Jordan	9,448	9,946	11,196
Qatar	19,707	20,426	20,426
Kuwait	35,180	41,748	41,748
Lebanon	18,379	19,895	21,768
Libya	19,346	23,465	29,119
Morocco	36,093	43,727	50,055
Oman	20,309	21,698	21,698
S. Arabia	188,551	212,581	250,557
Syria	19,949	21,470	23,133
Tunisia	21,024	25,037	28,185
U.A.Emirates	70,960		70,960

Source: World Bank, Data – Key Development Data & Statistics

Since the evaluation of parameters is a complicated and demanding process, both in terms of material and human resources, which is why it is usually performed by institutions devoted to economic research and much more rarely by individuals, only several of the works published here have applied the gravity model. Moreover, in the works by domestic authors in which the gravity model was applied, the parameters were taken from foreign authors. The work "Economic Integration on the Territory of Former Yugoslavia," published as part of the volume "Analysis of Potential Directions of Trade between the Former Republics of Yugoslavia",[16] used the gravity model for the purposes of determining potential trade flows between former Yugoslavian republics (FYR). The parameters used were those of L.A. Winters and Z.K. Wang,[17] which will be used in this case as well. The cross-section analysis employed by these authors included 76 countries, of which 19 were industrial. The gravity model was also applied in the master's thesis of Dejan Gajić, "The Eastern Expansion of the

16 *Ekonomska integracija na teritoriji bivše Jugoslavije: na putu ka evropskoj integraciji*, ed. Marta Bazler-Madžar, Ekonomski institut, Beograd, 2000, pp. 57-70.

17 Videti: L. Alan Winters, Zhen Kun Wang, *Eastern Europe's international trade*, Manchester University Press, Manchester and New York, 1994.

European Union and the Position of Serbia and Montenegro,"[18] which analyzes the potential trade directions of the Federal Republic of Yugoslavia.

The model was evaluated through the method of ordinary least squares (OLS). The obtained results show that the coefficient evaluations are statistically significant and logical.

Table 3: *Evaluated coefficients for the year 1998*

lnXij	Coefficient	St. error	t	P>|t|	95% Conf. Interval	
lnYi	0.930705	0.0310027	30.02	0.000	0.8698706	0.9915393
lnNi	0.0246654	0.0386548	0.64	0.524	-0.0511841	0.100515
lnYj	0.6829835	0.0310023	22.03	0.000	0.6221499	0.7438171
lnNj	0.0473422	0.0386642	1.22	0.221	-0.0285259	0.1232102
lnDij	-1.222391	0.0598855	-20.41	0.000	-1.3399	-1.104882
Aij	0.6401405	0.1234944	5.18	0.000	0.3978161	0.8824648
Pij	0.5789241	0.1033485	5.60	0.000	0.3761305	0.7817176
cons.	-4.655665	0.5933937	-7.85	0.000	-5.82004	-3.49129

According to the gravity model's given formula, potential exports from country *i* to country *j* shall be determined in the following way: all the known values – the export country's average three-year GDP, the import country's average three-year GDP, population of the export and import country and the distance between them, are multiplied by the appropriate parameters of Winters and Wang. Artificial variables receive the value of 2, if the countries share a common border or belong to the same economic grouping, or 1 – if that is not the case. These are also multiplied by the appropriate parameters and then added to the sum total.

A characteristic problem for cross-section models is that of heteroscedasticity, due to the nature of the data itself and certain shortcomings of the model. On the one hand, there is the problem of great data variations within the course of a single year and, on the other, what is even more important and characteristic of economic phenomena, it is impossible to wholly explain dependent variables by way of independent variables, regardless of the extent to which we include them in the model. Unexplained factors are included in the model by way of independent variables, and accumulate in the residuals. With the fluctuations in independent variables, the variances of chance errors (*iij*) also fluctuate and are not equal, which are the conditions of heteroscedasticity.[19] The problem of eliminating heteroscedasticity in cross-section models is beyond the scope of this paper. It is certain that the gravity model contains certain inconsistencies, but these are consciously accepted (and mostly eliminated), since the analytical

[18] Dejan Gajić, *Istočno proširenje Evropske unije i položaj Srbije i Crne Gore*, Beograd, 2005.

[19] Milena Jovičić, *Ekonometrijski metodi*, Ekonomski fakultet, Beograd, 1989, pp. 132-141.

worth of the obtained results is manifold, which is why the gravity model continues to be readily used in empirical research.

The result of such an application of the gravity model is the potential export from country *i* to country *j*. This result is then related to current exports in order to determine the "space" for an eventual increase in current exports. This operation is repeated for all import-export bilateral combinations between 18 countries in the region. A tabular presentation of these results in the text is provided only for the three most distinct cases.

We shall exclude two countries – Iraq and Israel – from our consideration and conclusions. Their data are included in the gravity model but, due to the significant scale of non-economic influences, the ratio of real and potential exports for these countries is almost always close to zero.

Exports to Israel are the smallest in all the countries, but this is due to specific political conditions. When it comes to Israel's intra-regional trade, political factors have the decisive role. As the only non-Muslim country in the heart of the Islamic world, with disputed territories, and a pronounced pro-American orientation amidst countries exposed to US pressures and sanctions, Israel's trade volume with the rest of the countries in the region is negligible, due to obviously non-economic factors. Data for Israel that were obtained by applying the gravity model shall not be considered as part of the conclusions. An exception to the extremely large difference between potential and real intra-regional trade in the case of Israel are Jordan's exports to Israel, which make up more than one half of potential exports (56%).

We shall also not take into account data for Iraq's exports in this period when making our conclusions. They indicate almost totally unexploited export potentials, which is a consequence of the political situation in the country over the past several decades. From 1980 to today, Iraq has gone through wars with Iran (1980-1988) and Kuwait (1990-1991), conflict with the US (the Gulf War intervention, periodic Anglo-American bombings during the 1990s, the invasion of Iraq in 2003-2005), rigorous UN sanctions and internal conflict, which is why the entire economy, and not just the country's exports – is in a state of collapse.

Algeria, as a large oil producer, very poorly exploits its export potentials on average. Its export potential is least fulfilled vis-à-vis the distant countries of Bahrain, Oman and Qatar (near 0). Especially surprising is the large disproportion between potential and real exports to Saudi Arabia – only 0.001%, since this large, wealthy country could be a significant market for Algerian goods. Algeria's export potential is fulfilled to a large degree when it comes to trade with Egypt – 0.553%, although this is only one half the value that could be achieved.

Bahrain achieves surprisingly large exports to North African countries: as high as 58% of potential exports to Algeria, almost 40% to Libya, 30% to

Morocco, 21% to Tunisia. What is especially interesting is that imports from these countries into Bahrain are negligibly small. Discovering the causes of this phenomenon would require additional analysis, which would probably have to do with the operation of large companies in Bahrain. Also surprising is the low export level to the large Saudi Arabian market, which comes to only 0.093% of potential exports.

Egypt, in the best case (exports to Morocco) achieves less than 20% of its export potential. Exports to Yemen and the Emirates equal only 10% of the potential, while exports to other countries in the region are almost non-existent. Especially surprising is the small export volume (3% of potential) to Tunisia, which is relatively nearby. Egypt is not a significant oil exporter, which further underlines the possibilities of increasing exports to other countries in the region.

Iran. It is generally concluded that Iran exploits its export potentials within the region of the Middle East and North Africa to a very small degree, which can largely be explained by the large share of oil and gas in its total exports, which, in Iran's case, is the largest among all the countries in the region (about 90%). Exports to Middle Eastern countries come to only 2-7% of the potential, with the exception of the United Arab Emirates, where they are much larger – over 22%.

What draws attention in the case of Iran is the fact that its export potential is exploited to its fullest extent vis-à-vis Morocco – 68% of the potential, while exports to other North African countries – Tunisia, Algeria and Egypt – come to only 0.006% of the potential. There is no obvious reason for this disproportion.

Yemen. Its most significant export destinations are Kuwait and the United Arab Emirates, to which it exports 68% and 40% of possible export value, respectively. Export capacities vis-à-vis other countries are totally unexploited. Exports to North African countries amount to between 0.004% and 7% of possible value, while Yemen's exports to Middle Eastern countries range between only 1-6%. Yemen's extremely small trade volume does not apply only to exports into the region: the country's overall exports are generally weak. In any case, this is the least developed country in the region.

Kuwait. Kuwait's exports to Yemen are one of the rare examples in the Middle East where the full export potential is exceeded – 1.387. What is also interesting is that Kuwait, as an almost exclusive oil and gas exporter, most approaches full potential exports when it comes to certain North African countries – Morocco 19.7%, Tunisia 12.3%, Egypt 5.7%. As is to be expected, exports to other oil-producing states in the region are significantly below potential. Except for exports to Yemen, the ratio of real and potential exports among export destinations in the Middle East is the most favorable when it

comes to Jordan – over 10%, followed by Syria – almost 8%, Lebanon and Oman – about 6%.

Lebanon. Since Lebanon has no mineral resources, its export potential to countries in the region is in the greatest degree fulfilled. The most significant are exports to the Emirates – 0.358, followed by Kuwait – 0.241, Algeria – 0.222, Qatar – 0.124, and Bahrain – 0.10, while those to Syria are in the last place – only 0.007.

Jordan, as a relatively diversified economy, fulfills its export potential to a greater extent than the other countries in the Middle East and North Africa – over 35% - and this percentage does not fall below 2% vis-à-vis any other country. Only exports to Egypt, Morocco and Lebanon fall below 10% of potential (0.021, 0.037 and 0.056). Its exports to Iraq are one of the rare instances in the region where real exports are greater than potential – 1.211. Thus, Iraq is a market already saturated with Jordanina goods, while Algeria and the United Arab Emirates are close to the threshold (0.889 and 0.668). On the other hand, the toher countries in the region still represent significant potential export destinations for Jordan. There is much free "space" for Jordanian exports to Egypt, to which Jordan exports only 2% of its potential, as well as to Morocco and Syria – to which it exports about 3% of possible trade volume.

Table 4: *Jordan's real and potential intra-regional exports*

	Real exports in millions $ (1)	Potential exports in millions $ (2)	Ratio of real to potential (1:2)
Jordan-Algeria	57.394105	64.529738	0.889
Jordan-Bahrain	19.462255	53.588793	0.363
Jordan-Egypt	25.693133	1241.047477	0.021
Jordan-Iraq	626.415064	517.292441	1.211
Jordan-Iran	30.008005	253.633727	0.118
Jordan-Israel	120.295093	211.655072	0.568
Jordan-Yemen	22.785918	45.387096	0.502
Jordan-Qatar	26.146629	76.250627	0.343
Jordan-Kuwait	43.312852	108.610671	0.399
Jordan-Liban	55.370703	994.147137	0.056
Jordan-Lebanon	23.165790	78.963394	0.293
Jordan-Morocco	2.159624	58.522093	0.037
Jordan-Oman	16.023328	63.072280	0.254
Jordan-S.Arabia	175.818931	1278.379152	0.138
Jordan-Syria	106.456035	3088.111473	0.034
Jordan-Tunisia	8.482917	70.187151	0.121
Jordan-U.A.Emirates	112.838535	168.867809	0.668

Qatar. This country's economy is almost totally reliant on oil and natural gas exports. Thus, the low exploitation of its export potentials to other countries in the region is not surprising. Its export destinations are South and West European countries. Its ration of existing and potential exports is the lowest vis-à-vis Bahrain – only 0.014, followed by Saudi Arabia – 0.025, and Oman and Iran – about 0.03. The closest to fulfilling their potential are exports to Egypt, which, nevertheless, equal only 0.277 or only a little above one fourth of the potential, then to Morocco – about 20% of potential, Jordan – 16%, and the Emirates – about 11%.

Saudi Arabia. Similarly to other large Middle Eastern oil producers – Kuwait and Iran, Saudi Arabia fulfills more of its potential exports vis-à-vis the countries of North Africa than those of the Middle East. In all three cases, the largest exports are to Morocco, to which potential exports are exceeded, equaling 105%, followed by Egypt (14% of potential) and Tunisia (12.5%). As expected, the least exploited export potentials are those to the other large oil producers in the region – Libya – 0.005, Kuwait – 0.029 and Qatar – 0.021.

Libya is the country that least uses its export potentials within the region – only 4.6% on average. It fulfills the greatest percentage of its potential exports vis-à-vis neighboring countries: Morocco – almost 20% and Tunisia – about 15%. Libya's fulfillment of potential exports to the countries of the Middle East is practically 0%, as shown in the table.

Table 5: *Libya's real and potential intra-regional exports*

	Real exports in millions $ (1)	Potential exports in millions $ (2)	Ratio of real to potential (1:2)
Libya-Algeria	2.614923	1863.288133	0.00140
Libya -Bahrain	0.001968	41.777235	0.00005
Libya -Egypt	45.685668	1118.941351	0.04083
Libya -Iraq		224.498173	0.00000
Libya -Iran		353.283993	0.00000
Libya -Yemen	0.025222	52.822666	0.00048
Libya -Jordan	12.686549	97.592340	0.13000
Libya -Qatar	0.5848	62.864306	0.00930
Libya -Kuwait	0.078196	122.935498	0.00064
Libya -Lebanon	16.73255	141.186409	0.11851
Libya -Morocco	63.376968	318.739534	0.19884
Libya -Oman	0.060868	63.412746	0.00096
Libya -S.Arabia	0.303906	481.687521	0.00063
Libya -Syria	0.000031	170.462373	0.00000
Libya -Tunisia	280.559082	1919.225894	0.14618
Libya -U.A.Emirates	0.000016	156.204872	0.00000

Morocco. As Morocco is not an oil producer able to export this product to industrial countries, what surprises is its very low level of exports to other countries in the region, with the highest achieved export level being only 16% of potential. Especially surprising is its low level of fulfilled potential exports to Algeria, which would, due to its chiefly oil-based economy (deficient in many goods and services) and relatively large population, be an excellent market for Morocco's relatively diverse economy.

Table 6: *Morocco's real and potential intra-regional exports*

	Real exports in millions $ (1)	Potential exports in millions $ (2)	Ratio of real to potential (1:2)
Morocco-Algeria	28.595293	1795.432996	0.016
Morocco -Bahrain	0.622287	45.045601	0.014
Morocco -Egypt	22.812432	426.812236	0.053
Morocco -Iraq	4.861528	219.822133	0.022
Morocco -Iran	32.277393	214.658370	0.150
Morocco -Yemen	5.113581	59.289240	0.086
Morocco -Jordan	12.380315	80.493997	0.154
Morocco -Katar	0.845391	68.718935	0.012
Morocco -Kuwait	2.957824	71.364630	0.041
Morocco -Lebanon	12.490002	115.316232	0.108
Morocco -Libya	49.062916	354.723246	0.138
Morocco -Oman	1.204018	73.900381	0.016
Morocco -S.Arabija	55.121012	500.490618	0.110
Morocco -Syria	23.000504	142.027764	0.162
Morocco -Tunisia	50.783942	443.132434	0.115
Morocco -U.A.Emirates	10.000152	177.313032	0.056

United Arab Emirates. The United Arab Emirates fulfill their export potential to a greater degree than most of the other countries (after Jordan) in the region – on average almost 25%. This percentage falls below 10 only in the case of two countries – Tunisia (9.6%) and Saudi Arabia (6.9%). The highest percentage of fulfilled export potential is achieved with Iran – over 88%, followed by Jordan – almost 42%, Morocco – 24% and Lebanon – about 21%.

The Emirates' previously dominant oil-based economy has in the last decade significantly diversified into other economic branches, which is one of the causes of their increased exports to other countries in the region.

Oman has significantly approached the fulfillment of its export potentials when it comes to its trade with Iran – 81%, Jordan – 67%, Libya – 45% and Yemen – 31%. As Oman's economy is primarily oil-based, this represents a

significant trade achievement, especially having in mind the region's average. Oman's lowest export potential fulfillment is with some of the North African countries: Morocco – 0.015, Algeria – 0.021, and Egypt – 0.032.

Syria. Syria's exports approach their potential only in the case of its trade with Algeria (73%) and Saudi Arabia (about one third of the potential – 0.342). Exports to Kuwait come to almost 30% of the potential, and to Yemen and Libya about one fourth of potential exports, more precisely 26%. These are quite poor results, considering the fact that Syria is not an oil exporter and that, within the context of the region, it has relatively favorable conditions for agricultural production, which could meet the needs of the neighboring markets.

5. Conclusions

On the basis of data obtained by applying the gravity model, it can be concluded that the potentials of each of the countries of the Middle East and North Africa for intensifying exports to other countries in the region – are quite large. The only exceptions to this are Jordan's exports to Iraq, the Emirates' to Yemen and Saudi Arabia's to Morocco – all of which exceed potential volume.

The countries of the Middle East have achieved a stronger regional integration than the North African countries, which are fulfilling less than 7% of the potential value of intra-regional trade. The reason for this is their large volume of trade with the countries of the European Union, which is a consequence of geographical proximity (the south of Europe is closer than the Middle Eastern countries), historical ties based on the metropolis-colony model (which is a legacy of trade relations) and stability (trade and security).

The low level of fulfillment of regional export potentials is fairly uniform among the North African countries, with Algeria achieving about 7%, Egypt 6%, Libya 5%, Morocco 7%, and Tunisia 8% of potential exports within the MENA region.

Even if we exclude the Middle Eastern countries as partners of the North African countries, we are still left with a huge unexploited mutual trade potential among the latter. For example, Algeria significantly approaches its potential trade volume only in its exports to Egypt (about 55%), while, in the case of exports to other North African countries, the figure amounts to only 4-8% of the potential. Its biggest export potential is vis-à-vis Libya, with which export value comes to only 0.5% of the potential.

The countries closest to achieving the maximum of their intra-regional potential are Jordan with 35% and the United Arab Emirates with 28%.

The more diversified economies of the Middle East *are not* closer to the maximum potential volume of intra-regional trade than the oil-based ones.

Oman is achieving one fifth of its potential exports to the region, Syria 16%, Bahrain 15%, Kuwait 14%, Saudi Arabia 13%, Yemen 10%, Lebanon 9%, Qatar 8%, and Iran 7.5%. This is not an expected result, since the oil-based economies' exports are almost totally directed toward developed industrial countries, since the region's oil needs are quite small.

Political, geopolitical and military-strategic factors bear a strong influence on the direction of trade flows, which limits the development of the regional trade. This particularly applies in the case of Israel, as the only Jewish state in a Muslim region. Israel's exports to neighboring states are, in some cases, only a thousandth of the value the country could potentially achieve from intra-regional trade. Its average ratio of potential to real exports is less than 2%. None of the region's other countries is so far removed from fulfilling its export potentials within the region.

The data on the real exports of Iraq to Iran in the selected period (2002, 2003, 2004) do not reflect the true state of this country's external trade, due to the war that started in March 2003. Iraq's exports to the region's other countries average only about 5% of possible exports; however, in the rare periods of relative peace, they are much greater.

The application of the gravity model has confirmed that the intensity of trade is much stronger between countries that are closer in the cultural, linguistic and geographical sense.

Geographical distances, i.e. the costs of transport to the economic center, have such a great role that they can negatively affect trade volume even when the trade partners have a common border. A prominent example is Saudi Arabia's imports from its neighbors. The exports of Oman (7%), Bahrain (9%), Qatar (2.5%), Kuwait (1%) and the Emirates (7%) to Saudi Arabia do not even remotely approach their possible volume, while those of Yemen and Jordan are substantially closer to the potential (31% and 14%, respectively). The reason for this is that the first group of Saudi Arabia's neighbors, on its eastern rim (northeast, southeast), is more distant from its economic centers – Mecca and Jeddah. On the other hand, although it is the capital, Riyadh has no great economic significance because it is located in the desert, despite the fact that it lies relatively close to the first group of countries. Mecca and Jeddah are located on the Red Sea, to which Yemen and Jordan have access, which allows for much cheaper sea transport. In this case it's not just a matter of distance but also of the poor transport infrastructure in the Arabian desert.

References

- Deardorff, Alan V.: "Determinatns of Bilateral Trade: Does Gravity Work in a Neoclassical World?", in: J. A. Frankel, *The Regionalization of the World Economy*, Chicago: University of Chicago Press, 1998.
- *Ekonomska integracija na teritoriji bivše Jugoslavije: na putu ka evropskoj integraciji*, redaktor Marta Bazler-Madžar, Ekonomski institut, Beograd, 2000.
- Feenstra, Robert C. – Markusen, James R. – Rose, Andrew K.: "Using the Gravity Equation to Differentiate among Alternative Theories of Trade", *The Canadian Journal of Economics*, vol. 34, no. 2. 2001.
- Frankel, J. – Romer, D.: "Does Trade Cause Growth?", *American Economic Review*, vol. 89, no. 3, 1999.
- Frankel, Jeffery A.: *Regional Trading Blocs: In the World Economic System*. Washington, DC: Institute of International Economics, 1997.
- Gajić, D.: *Istočno proširenje Evropske unije i položaj Srbije i Crne Gore*, Beograd, 2005.
- Gallup, J. L. – Jeffrey, D. S.: *Geography and Economic Development*, Columbia University National Bureau of Economic Research (NBER), 1998.
- *Great Arab Free Trade Area*, UN, ESCWA, New York, 2001.
- International Trade Center UNCTAD-WTO.
- Jovičić, M.: *Ekonometrijski metodi*, Ekonomski fakultet, Beograd, 1989.
- Limao & Venable: "Infrastructure, Geographical Disadvantage, and Transport Costs", World Bank Policy Research Working Paper No. 2257.
- Tinbergen, J.: *International Economic Integration*, Elsevier, London, 1954.
- UN Population Fund, The State of World population, New York, 2004.
- UN, ESCWA, *Great Arab Free Trade Area*, New York, 2001.
- *United Nation Statistics Division – unstats.un.org/unsd/comtrade/default.aspx*
- Winters, L. A., Wang, Z. K.: *Eastern Europe's international trade* Manchester University Press, Manchester and New York, 1994.
- World bank – *Data – Key Development Data & Statistics*

Expert article
Received: August 13th, 2007
UDK 519.866:330.101.542

Professor Ivica Stojanović, PhD
Faculty of Business Studies, Megatrend University, Belgrade

ABSTRACT MATHEMATICAL MODELS AND DEDUCTIVE METHODS IN THE MICROECONOMIC ANALYSIS OF REAL ECONOMIC PHENOMENA

Abstract: *Economists like to think that their discipline is "the most scientific" of all the social sciences, which argument is supported by the abstract mathematical models and deductive methods that are currently prevalent in the profession. However, the inability of accepted economic theory to come to grips with the world's problems from the 1970s to the present has been quite disturbing for this profession. Economists' reliance on the said mathematical models and deductive methods as a substitute for studying economics in the real world and its social institutions, has contributed to their failure.*

Key words: *microeconomic analysis, "neoclassical synthesis," neoclassical economics, abstract mathematical models, deductive methods*

1. Introduction

Modern science and its method came about as a result of the battle against the medieval way of thinking. Galileo Galilei, who lost the battle but won the war,[1] thought that ignorance was the mother of malice, envy, anger and all other vices, of horrible and vile sins.

Instruments for observation, measurement and experimenting became the chief tools of science. Theory was no longer just contemplation but, before all, mathematical construction and mathematical discourse, which must have its own constant experimental interpretation and its own exact, sensory-observational justification.

Modern scientific thought began its triumphant campaign against the metaphysical way of thinking by battling for a new method. Descartes, the

[1] Homa Katouzjan, "Evolucija ekonomskog metoda: od političke ekonomije do pozitivne ekonomike." *Gledišta*, no. 5-6 for 1985, p. 27.

founder of the famed Cartesian[2] rationalism, thought that the old knowledge framework did not answer the demands of the age. Through the simple formula "I think, therefore I am," he at last succeeded in liberating philosophical speculation from the medieval obsession with proving the reality of the objective world. On the other hand, British philosophers resisted this Cartesian subjectivism, claiming that sensory experience and observation of the external world were primary in comprehending the world. According to them, the human mind was not the source but a storehouse of knowledge initially gained through sensory experience and then systematized by the ability of the mind. This opened the way for the establishment of two great traditions within modern philosophy - rationalism[3] and empiricism.[4] These traditions exerted great influence on the beginnings and the development of economics as a science, as well as on the beginnings of microeconomic analysis – all the way up to the present. The use of abstract mathematical models and the deductive method in microeconomic analysis trace their roots to the rationalistic philosophical school. On the other hand, the roots of the analysis of real economic phenomena and social institutions in the real world can be found in empiricism.

What is it that actually happened with economic theory as a whole and, by extension, to microeconomic analysis itself?

2. Neoclassical economic thought and "neoclassical synthesis"

During the 1960s it seemed as though the centuries-old problem of sudden economic growth and sudden economic decline had been solved. Differences in economic theory had converged into a neoclassical economic school of thought better known as "neoclassical synthesis." "Neoclassical synthesis" had reconciled the classical economics of Adam Smith's "invisible hand" with the theory of John Maynard Keynes. In the introduction to his well-known university textbook, "Economics,"[5] Paul Samuelson wrote that neoclassical economic thought had been thoroughly accepted, stating that the system of modern economic analysis was increasingly moving in the direction of a synthesis of its two basic fields: 1) the macroeconomic analysis of scales of production,

[2] Cartesianism is the philosophy of the founder of modern rationalism and dualism, René Descartes (Lat. Renatus Cartesius, 1596-1650).

[3] Rationalism is the philosophical school that considers reason and thought to be the only sources of cognition and the only way to arrive at the truth. Philosophers of this school are: Plato, Descartes, Spinoza, Leibnitz, Fichte, Schelling, etc.

[4] Empiricism is, in fact, a philosophy of experience, i.e. a gnoseological school of thought that deduces all cognition from experience. Methodologically speaking, empiricism is a principle according to which all sciences must be founded upon experience. The founder of gnoseological empiricism is John Locke, and of methodological empiricism Francis Bacon.

[5] P. A. Samuelson, *Economics*, McGrow-Hill, New York, 1948, 7th ed. 1967, p. 196.

employment and price stability and 2) microeconomic analysis, i.e. the analysis of prices, markets and costs, as well as the distribution or formation of incomes: personal incomes from labor (wages), profits, rents and interests. Macroeconomic analysis and microeconomic analysis appeared as the two basic components of a general system of modern economic analysis. Samuelson referred to this general system of modern economic analysis as "neoclassical synthesis."

Macroeconomic analysis mainly represents the further development of Keynes' analysis. Microeconomic analysis is engaged in analyzing prices and distribution and represents an improved, significantly corrected and further developed system of post-classical traditional, i.e. Marshallian economics. While the general system of post-classical traditional economics mainly encompassed only the microeconomic analysis of prices and distribution, modern microeconomic analysis appears merely as one of the two basic fields of a general system of modern economic theory, i.e. of "neoclassical synthesis." The Marshallian system of microeconomic analysis was corrected and improved by later economists, especially Hicks and the theorists of monopolistic and incomplete competition. As for some of its basic arguments and suppositions, the Marshallian system has undergone changes, being brought into harmony with the modern system of macroeconomic analysis. In this changed and supplemented version, in the form of modern post-Marshallian microeconomic analysis, it has become one of the two basic, integral components of a general system of economic theory, which Samuelson has dubbed as "neoclassical synthesis."

3. The "perfect competition" model

The concept of "perfect competition," just like the ***model of "perfect competition,"*** is not a new thing in the history of economic thought. "Perfect competition" was propounded by the economists of the physiocratic school of economics, followed by economists belonging to the school of classical economic thought and the economists of the school of marginal economics. For Samuelson,[6] "perfect competition" was a state in which no producer or owner of means of production is able to exert individual influence on market prices. Otherwise, we wind up with a greater or lesser degree of imperfect competition.

The model of "perfect competition" in its pure form did not exist even in the period of liberal capitalism of the 19th century.[7] A "complete" or "perfect competitor" is one who buys or sells a certain good but cannot influence market prices. These could be small farmers, who individually produce a negligible share of overall agricultural production.

6 Paul A. Samuelson, *Economics*, 7th ed., McGraw-Hill Book Company, New York, p. 41.

7 Ibidem

The model of "perfect competition" requires "perfect information." Consumers must be sufficiently informed so that they can compare products. Workers must be aware of alternative possibilities, while capitalists must be aware of eventual competitive investments. In the absence of this, sellers can demand higher prices than the competition and get away with it, while workers can demand wages higher than the balanced wage. In that case, this stops being the model of perfect competition.

"Perfect competition" demands "the perfect mobility of the factors of production." Workers must have the freedom to demand the highest price they can get for their work, and capitalists the freedom to invest their capital where it will bring maximum profit. In the absence of this, we will have a situation where equal factors of production dictate different prices, and that is not the model of perfect competition. The economic proponents of the perfect competition model argue that monopolistic prices or wages cannot last too long, as there will always be entrepreneurs who will take advantage of the opportunity, appear on the market and force a restoration in the price balance.

The "perfect competition" model defines the best of all worlds, but also a world that is not real. It is not real because, among other things, such a model presumes that the market is made up of numerous sellers and buyers, each of whom has too little market power to dictate prices. The seller is the one who accepts the balance price. If not, then this is no longer a model of perfect competition.

The "perfect competition" model, as already mentioned, is nothing new in the history of economic practice, and especially not in the history of economic thought. However, what is new is the fact that, during the 1970s, "neoclassical synthesis" was seriously shaken up. Macroeconomic analysis was pushed aside within the framework of "neoclassical synthesis," while post-Marshallian microeconomic analysis pretty much regained its dominance. In addition, what was especially significant, the "perfect competition" model once more regained currency and significance. It was Baumol who formulated the theory of "perfect competition," i.e. the competitive market, in 1982.[8]

4. "Neoclassical synthesis," stagflation and "perfect competition"

During the 1970s, the economies of the world's industrially developed countries were not only faltering but doing so in a way that puzzled the accepted "neoclassical synthesis," i.e. the neoclassical economic theory. By the end of the 1970s, countries with developed market economies had an increasingly difficult time in achieving the three basic goals of development: 1) stability and dynamic growth accompanied by an optimal rate of economic growth, 2) the

[8] W. Baumol, J. Panzar, R. Willig, *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Jovanovich, New York, 1982.

full employment of all factors of development and 3) if feasible, a sufficiently balanced balance of payments. The basic dilemma was: either economic and price stability or dynamic economic development burdened by inflation. Either high employment accompanied by instability or stabilization of the economy accompanied by high unemployment and a lower rate of economic growth. Economists referred to this economic situation as stagflation. Since it combined mass unemployment and high rates of inflation, economists dubbed stagflation as the "two-headed monster." [9]

The sharpening of stagflationary contradictions brought into crisis the macroeconomic analysis of the scale of production, employment and price stability, i.e. Keynesian economic thought. Economists gained a bad reputation and lost orientation. This also led to a sharp polarization between the ruling conceptions of economic policy.

Under conditions of growing budget deficits, rising unemployment and inflation, unsatisfactory results of deflationary policy and income policy, as the two main methods of the anti-inflationary policy of the 1960s and 1970s, Keynesian economic policy began attempting to introduce elements of monetarist recipes in order to try to contain inflation, without abandoning its basic orientation.[10] Modern conservatism began as a result of a critique of the New Deal social state, Johnson's Great Society and Nixon's "imperial presidency." [11]

The consensus known as the "neoclassical synthesis," which existed during the 1960s, split into: a) the so-called Chicago "monetarists" headed by Milton Friedman, [12] Jeffrey Sachs and others, b) the post-Keynesians[13] and c) the school of "rational expectations."[14] The post-Keynesians are viewed as a dissident school, the weakest and, seemingly, the most despised of all. The post-Keynesians are seeking to save Keynes from neoclassical synthesis, which Keynes' pupil Joan Robinson has dubbed as "bastard Keynesianism." Neoclassical economists treat the neo-Keynesians with great contempt. Those claiming the mantle of

9 J. E. Miede, *Stagflation: Wage-Fixing*, vol. I, George Allen and Unwin, London, 1982.

10 S. Nikitin, "Inflacija narube`i 80-h godov." *Voprosy ekonomiki* 3/82, pp. 115-124.

11 Collection of works, *Konservativismus, Eine Gefahr fur die Freiheit?*, Piper und Co. Verlag, Munchen, 1983.

12 Milton Fridman, *Teorija novca i monetarna politika*, Izdavačko preduzeće "Rad." Belgrade, 1973.

13 G. A. Akerlof, "Labor contracts as partial gift exchanges," *Quarterly Journal of Economics*, November 1982, P. Arestis, P. Scott, "Conflict, wage determination, and hysteresis in U.K. wage determination," *Journal of Post Keynesian Economics*, no. 3, 1993; P. Davidson, "The elephant and the butterfly," *Journal of Post Keynesian Economics*, no. 3, 1993.

14 R. Lucas, "On the Mechanics of Development." *Journal of Monenary Economics*, no. 1, 1986; T. Sargent, *The Ends of Four Big Inflations, Rational Expectations and Inflation*, Harper & Row, 1986.

Keynes' true successors are viewed as some sort of economic Anabaptists,[15] a historically peculiar theological misconception that just won't go away.

After 1970, an entire series of serious and ideologically diverse articles and books announced that economics had fallen into a deep crisis. Some of the titles reflect this state of affairs: "The Crisis in Economic Theory,"[16] "Economists at Bay,"[17] "What's Wrong with Economics,"[18] "The Irrelevance of Conventional Economics," "Why Economics Is Not Yet a Science," "Dangerous Currents: The State of Economics," etc.

For Keynesians, inflation appears when aggregate demand is greater than the supply of goods and services. Anti-inflationary policy strives to reduce excess demand by raising taxes and interest rates and reducing public works.

For monetarists, inflation is a monetary phenomenon.[19] Inflation, recession, unemployment and crisis in the balance of payments are the results of a faulty monetary policy that leads to a growth of money in circulation greater than the growth in production.[20] Friedman thought that the growth of investments above savings, the pressure of wages, profits and material costs, and the growth of budget deficits produced inflation. Divergence between the quantity of money issue and the demand of real money stocks led to inflationary processes. Anti-inflationary policy should reduce public spending, taxes and public loans, enforce strong control over growth in the money supply and abandon income-based policy, i.e. the policy of price and wage control. The monetarist theory of inflation is founded on the quantitative theory of money. A change in the quantity of money in circulation determines aggregate demand and produces changes in the level of prices. *This theoretical model on the causes of inflation presumes an implicit assumption of the existence of a market of "perfect competition."*

Keynesians claim that monetary policy is inefficient in bringing cost inflation under control. This inflation can be prevented through fiscal policy and income policy.

The monetarists hold the opposite: only monetary policy can stabilize the economy, because the sole goal of budget policy is the reallocation of economic

15 Anabaptism is a religious movement requiring a new baptism in adulthood.

16 Publication Information: *The Crisis in Economic Theory*, Daniel Bell – Irving Kristol eds., Basic Books, New York, 1981, p. 226; http://www.questia.com/PM.qst?a=o&d=99875505, June 14, 2007.

17 Robert Lekachman, Susan Previant Lee, "Why the Experts Will Never Solve Your Problems." *Political Science Quarterly*, vol. 91, no. 3, 1976), pp. 527-528; http://links.jstor.org/sici?sici=0032-3195(197623)91%3A3%3C527%3AEABWTE%3E2.0.CO%3B2-D, June 14, 2007.

18 Fred E. Foldvary, ed., *What's Wrong With Economics?*, 1998; http://www.progress.org/fold23.htm, June 14, 2007.

19 L. Yeager, *Monetary Policy and Economic Performance*, Washington, 1972, p. 13.

20 M. Friedman, "The Counter-Revolution in Monetary Theory." *Occacional Papers*, no. 33, 1976, p. 6.

resources. Fiscal policy has a secondary significance in achieving short-term price stabilization. An increase in fiscal revenues does not necessarily have to lead to a reduction in investments and demand.

When it comes to unemployment, the Keynesians' starting point is the so-called Philips Curve.[21] Philips conducted an empirical analysis of the relationship between unemployment and the rate of change in nominal wages. His goal was to prove the stability of the long-term relationship between unemployment and inflation. The stability of the Philips Curve allows economic policy to make a choice between unemployment rates and inflation. Neo-Keynesian J.E. Meade points to the fact that Keynes analyzed unemployment as a phenomenon of weak total demand during the time of the great economic crisis of the 1930s. The stagflation unemployment of the 1970s was different. This was structural unemployment. This is why Mead thinks that unemployment should be treated by sectors.

In considering the problem of unemployment, the monetarists reject Philips' evidence regarding the stability of the long-term relationship between unemployment and inflation. There is no long-term and constant choice between unemployment and inflation. Pointing to the instability of the Philips Curve, *the monetarists have returned to the theory of the "perfect competition" of the labor market.*

All the way up to the mid 1970s, labor unions in Western Europe and the US were powerful and the influence of the state was quite high. The activities of these two powerful and well-organized participants in the labor market were directed, before all, toward maintaining the high rate of real wages and the protection of workers' rights when it comes to work dismissals. At the same time, unemployment compensation was also high, which satisfied the demand for social justice.

However, the 1970s brought high unemployment accompanied by low production growth to almost all the Western European countries. The political parties (conservative) that came to power in Western Europe and the US at the end of the 1970s and the beginning of the 1980s, held that wage rigidity amidst decline were the basic cause of growing unemployment. Namely, high and constantly rising real wages were turning labor into an expensive factor. Thus, under conditions of economic hardship, the substitution of labor with capital was inevitable. In addition, companies were resisting the hiring of new workers, because workers' rights were protected to such a degree that, in times of crisis and company adjustment to new conditions, it was very difficult to reduce the number of employees. Having in mind the above facts, it was thought that the problem of unemployment could be resolved in the following ways: a) reduce the protection of employees and the unemployed, b) weaken the labor unions and c) deregulate. The expected result was strengthened competition in the labor market and a reduction in real wages. Cheaper labor and

[21] A. W. Philips, "The Relation Between Unemployment and the Rate Change of Money Age Rates in the United Kingdom 1962-1975." *Economica*, vol. 25, 1978.

greater flexibility regarding the number of employees, i.e. a reduction in barriers to dismissing workers, would pressure workers into accepting lower wages and should stimulate entrepreneurs into hiring more workers.

The view that wage rigidity was the cause of growing unemployment grew out of the neo-classical theory of employment, which champions the "perfect competition" model on the labor market as well. This model presumes the unfettered influence of competition on the labor and product markets, which also presumes the free movement of prices and wages in both directions. This is a necessary precondition for markets to reach a state of balance in the long term. On the labor market, the balance is not determined by full employment but is characterized by the existence of a certain dose of "voluntary" unemployment – the so-called natural rate of unemployment. However, if there are obstacles to the free activity of market forces (such as, for example, the organization of workers into unions, state interventionism, etc.), which prevent the free movement of prices and wages depending on supply and demand, unemployment will outstrip the balance level; in other words, forced unemployment will appear alongside voluntary unemployment.

Thus, neoclassical theory and the model of "perfect competition" allow for a reduction in unemployment through the weakening of labor unions and the withdrawal of the state from the labor market. The main weakness of neoclassical theory and the model of "perfect competition" on the labor market lies in the fact that this theory and this model start from the premise of unfettered market activity, i.e. that it is possible to create a "perfect competition" model in the economy – which is something that has never occurred in practice and never will.

Hence, differently from the Keynesians, the monetarists think that inflation is more harmful than unemployment. They believe that *"perfect competition" on the labor market*, without state intervention, leads to full employment. They also believe in the rule whereby a reduction in the growth rate of the money supply limits the potential inflation rate, because this denies the economy the additional liquid means that cause inflation. Keynesians regulate the size of aggregate demand through interest rates and fiscal policy.

What has remained of the starting theoretical premises of neoclassical theory are the standpoints of Smith's classical "invisible hand"; at the end of the twentieth and the beginning of the twenty-first century the projections of "perfect competition" and "general balance" are once again dominant. It is thought that a "general balance" characterizes the entire economy. The model of perfect competition is spoken of almost axiomatically, while Keynes' thoughts are treated as though they don't exist.

Formally, neoclassical economics are still the ruling school. Its theoretical premise still presents the possibility of connecting the starting premises of Smith's classical invisible hand – the principle of the self-regulating economy – with Keynes' idea that it is necessary for government to achieve macroeconomic stabilization through state intervention. However, the said theoretical starting premise of

neoclassical economics has simply been neglected, while economic theory has been flooded by analyses *dominated by projections of perfect competition and "general balance." It is thought that a "general balance" characterizes the entire economy.*

Due to the current situation: 1) the disintegration of "neoclassical synthesis" and 2) the further survival of neoclassical economics as the ruling school of economic thought, which is, however, dominated by projections of "perfect competition," which, since they are not real but are founded on theoretical textbook assumptions, must be proven by way of abstract and deductive mathematical and econometric methods, economist Joan Robinson rightly concludes that "economic theory in the West is in very bad shape. It is highly confusing and the basic ideas are divided into various segments, which cannot survive together. The production function is an effective means of stultification. The political economy student is given the formula Y = f/c,k/, after which one moves on to the next question in the hope that the student will forget to ask in which units 'k' is measured. On the other hand, before the student even asks this question, he will become a professor himself; thus are the habits of spiritual laziness passed on from one generation to the next."[22]

Neoclassicists are opposed to Keynes' state interventionism. They believe that state interventionism measures disrupt certain aspects of economic life. Even though they "realize" the shortcomings of the "perfect competition" model, they are nevertheless – starting from assumptions instead of real-life facts – constructing a model of "perfect competition" that defines the best possible world. However, since their model is based on assumptions and not on real and attainable facts, that world does not represent a real world that can actually be achieved.

5. Abstract mathematical models, deductive methods, empiricism and microeconomic analysis – The traditions of rationalism and empiricism –

Since the proponents of neoclassical economics do not construct the "perfect competition" model on the basis of real-life facts but on the basis of assumptions, they must use abstract and deductive mathematical and econometric methods in order to prove their starting premises and hypotheses. In this way, neoclassical economics creates the illusion about the best of all possible worlds, which it equates with the real world, and believes that the real world, which is completely different, must accommodate to their models of "perfect competition." Therefore, the neoclassicists simply overlook the fact that the world of reality is that which is, while the world of illusions is that which isn't.

[22] Džoan Robinson, *Ekonomska filozofija*, Istraživačko-izdavački centar SSO Srbije, Belgrade, 1981, p. 58.

Neoclassical economic theory uses deductive reasoning,[23] and starts from axioms. It *a priori* takes the view that perfect competition is at the same time a description of the optimal world and a useful approximation of the real world. When it is pointed out that high unemployment or a fragmented labor market, oligopolistic corporations, national economic development strategies, large public sectors, controlled banks, protected agricultural markets and the logic of social organization all speak of a world far removed from the textbook picture of perfect competition and automatically correcting markets, neoclassical economists claim how the world would be a much better place if it was built on the basis of textbook lessons. They prove this by using deductive methods in their textbooks. However, "the deductive method in the field of economics is replacing empiricism. Those who possess true empirical curiosity and the ability to understand the functioning of banks, corporations, production technologies, labor unions, as well as a desire to become acquainted with economic history or individual behavior, are dismissed as *ad hoc* empiricists, literary historians or sociologists, and are relegated to the margins of the profession. Their place is taken by a new generation of excellent experts in abstract mathematics, who are, however, ignorant when it comes to the realities of economic life."[24]

They use statistical and mathematical models in a variety of ways. Some mathematical models are purely theoretical and use algebra only in order to manipulate the premises. This also the case with the theory of general balance, as well as with many less significant questions about which they write articles in periodicals.

The use of mathematics in econometric modeling can also be quite empirical.[25] Commercial consulting forms, such as "Wharton Econometrics," "Chase Econometrics," and "Data Resources, Inc." have come up with complex economic models, which can use even a thousand equations connected by some computer program. Gathering data over a long period of time and producing correlational tables, the econometric modeler seeks to predict the ways in which variables will affect one another in the future. At the end of the 1960s, there was a rapid growth in the popularity of constructing large econometric models of economic prediction. The problem with such models is that there are always more variables than the model assumes, and that the past does not always predict the future.

During the 1970s, when the main influences on macroeconomic stability were exerted by external variables, the so-called external shocks (for example, the rise in the OPEC countries' price of oil) and structural changes in eco-

[23] The deductive method presumes deduction, concluding about the specific from the general, a method of thought by which specific laws are arrived at from a general law.

[24] Robert Kuttner," O stanju u oblasti ekonomije." *Pregled* magazine, p. 236 for 1986/87, Belgrade, taken from *The Atlantic* magazine, p. 41.

[25] Ibid., p. 48.

nomic systems, the large econometric models of economic prediction stopped being all that popular.

A different, even more controversial use of econometric technique consists of the testing of algebraically modeled hypotheses. One mathematical method, known as multiple linear regressions, allows the testing of a larger number of variables, which might have cause-and-effect connections, by considering them one at a time.

By manipulating time lags, a good econometric expert can "prove" just about anything. Besides, although many economists argue that an honest way of checking a theory consists of specifying the hypothesis and a single execution of regressive equations, the usual practice is to play with the equations over a period of time, during which one manipulates time lags, time periods and other variables, until the equations more or less confirm the hypothesis. Of course, certain correlations can be a pure coincidence, while other apparent correlations can mask the true causes that were overlooked.

When it comes to models, Friedman's neoclassical model of employment has gained prominence. Normally, as already mentioned, in considering the problem of unemployment, Friedman and the monetarists reject Philips' evidence about the stability of the long-term relationship between unemployment and inflation. They think that there is no long-term, constant choice between unemployment and inflation. Having in mind the instability of the Philips Curve, *the monetarists are returning to the theory of "perfect competition," i.e. the model of balance on the labor market.*

For Milton Friedman[26] it is not important whether a certain assumption is empirically accurate as long as it is internally consistent and the facts don't negate the model. Thus, even in a world in which competition is far from being perfect, it is justified to construct a certain model of economic activity according to the model of "perfect competition" and the axioms of neoclassical economics. That is, for instance, how Friedman constructed the model of the "perfect competition" of the labor market in the first place, even though he knew that it was founded exclusively on assumptions that do not exist in real economic life.

With his model of "perfect competition" on the labor market, i.e. his employment model, Friedman[27] sought to prove that "perfect competition" and a long-term balance without forced unemployment characterize the labor market. To be sure, Friedman admits that a certain percentage of unemployment appears in a state of long-term balance, but adds that these are exclusively workers looking for better jobs, i.e. that this is voluntary unemployment. The unemployment rate that exists within a state of long-term balance is called the natural rate of unemployment.

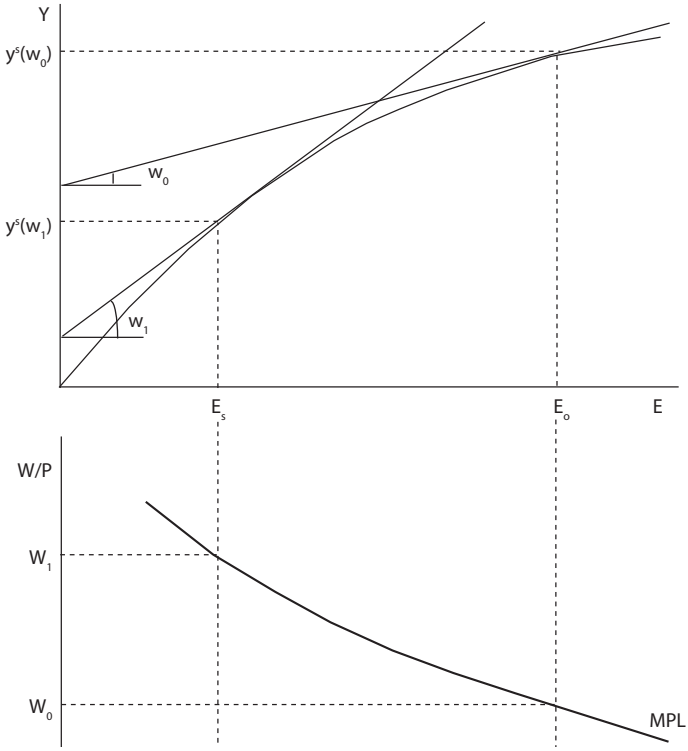
[26] Milton Friedman, *Essays in Positive Economics*, New York, 1953.

[27] Milton Friedman, "The Role of the Monetary Policy." *American Economic Review*, 58, 1968, pp. 1-17.

The basic characteristic of this model is that employment and production are defined on the labor market as the result of labor demand and supply. In accordance with this, the appearance of involuntary unemployment can be only temporary. The unemployed exert pressure in the direction of a reduction in real wages, which – if the price of labor is flexible – truly leads to a fall in wages and a reduction in the number of unemployed workers.

The model assumes that the labor and production markets are perfectly competitive. Under such conditions, a firm will continue to employ new labor until marginal income reaches the level of the real wage, which is exogenously determined for each firm. The scale of engaged labor, as well as capital determines marginal income. Thus, in order to eliminate the influence of capital on employment rates, one starts from the short-term production function, which assumes an unchanging capital stock. In that case, the marginal labor product depends exclusively on labor input. A newly engaged work unit brings a certain increase in production, which represents the marginal income.

Figure 1. *Labor demand under conditions of perfect competition*



Y = f(E): Short-term production function
E0 (W): Labor demand function

The way in which labor demand is formed is shown in Figure 1. The short-term production function y = f(E) is given in the upper part of the figure, where "y" represents production and E employment. As already mentioned, production appears exclusively as a function of the number of employees. Marginal productivity of labor (MPL) is defined by the slope of the production function *dy/dE = f'(E)*, while the level of employment depends on the real wage, which is represented in Figure 1 as the slope of the tangent on the production function.[28] Namely, the firm will continue hiring new workers until marginal productivity becomes equal to the real wage, which is exogenously determined. In other words, the firm cannot influence the wage level, but accepts it as a given. Thus, if the real wage, for example, grows from Wo to W1, profit maximization will require for marginal productivity of labor to increase by the same amount, which is, in this case, possible to achieve only by reducing the number of employees from E to E1. And, vice versa, employment can grow only if real wages fall. The lower part of Figure 1 shows labor demand as a function of the real wage.

In short, a firm's production is a function of the number of employed workers ys = ys(Ed), while the number of employees is functionally dependent upon the real wage Ed = Ed (w).

Thus, as already mentioned, for Milton Friedman it is not important whether an assumption within the "perfect competition" model of the labor market is empirically accurate. What is important is that the assumptions are internally consistent and that the data do not refute the model.

For example, one characteristic article[29] poses the question about how wealth, income and other variable factors influence the making of a decision about whether to buy or rent a home. *The model used in the analysis starts, among other things, from the premise that the housing market is "balanced."* The model algebraically, in a theoretical way, projects various possibilities of consumer behavior under various market circumstances. The authors then reach certain conclusions, such as, for example: "Renting becomes attractive if house construction is dependent on some random monetary gain or loss, but consumers may also invest their money in the capital markets according to a set rate of return." In other words, consumers would rather choose something

[28] The reason why the maximum profit for given wage level w is achieved at the point of intersection of the tangent and the production function can be explained in the following way. If total real profits achieved by a firm are $\pi = y - wE$, and $w = wo$, then $y = \pi = y + woE$. Since π has a constant share in y for each line of slope wo, it is in the firm's interest to have the line set as high as possible. As the combination of y and E must be within the production function, the highest common point of the line and the production function is the one in which the line only touches y, i.e. where it appears as the tangent of the production curve.

[29] J. V. Henderson, Y. M. Ioannides, "A Model of Housing Tenure Choice." *The American Economic Review*, 1984.

certain than something uncertain. What is surprising in this and other, similar articles is that they don't contain any data, or even any indication of whether the authors had ever studied the real housing market. *The article is a pure manipulation of assumptions and conclusions aided by mathematical logic.*

Among the sharpest critics of the excessive use of mathematics in economics is one of the most prominent mathematicians, Wassily Leontief).[30] In his acceptance speech to the presidency of the American Economic Association (AEA) in 1970, Leontief – as the first economist who worked with computers and the inventor of the "Input-Output" mathematical analysis, which brought him the Nobel Prize in 1973 – condemned the increasing obsession with the imaginary and the hypothetical instead of with empirical reality, and described the "Darwinian" process by which pure theorists are displacing those who study real economics.

On the other hand, as a departure from the use of mathematical logic, abstract mathematical models and the deductive method, it is possible for those engaged in the study of real economics in an empirical way to investigate when certain assumptions of the standard perfect competition model should be applied and when not. For example, it is possible to analyze 1) How and in what way, and under which assumptions of the standard perfect competition model will technological and institutional changes influence economic growth? 2) Which institutional circumstances justify public intervention? 3) What are the connections between economic performance and cultural and political values? 4) Which markets behave like those from the textbooks? 5) What truly explains the great disparities in the level of technological success achieved by various countries during various historical periods? These questions are rarely considered in economic periodicals, except at inaccessible levels of abstraction and assumption. And, should any of them bring theory into question, then true empiricism becomes a thorn in the side of the deductive method.

Harvard University economists Richard Freeman[31] and James Medoff[32] have gathered data that show that the real influence of labor unions on labor

[30] Wassily Leontief was born in the USSR, and became a professor of economics at Harvard in 1946. His most significant work is the analysis of interdependence within the framework of an economic system, especially in the production sector, according to a method he called the "Input-Output Analysis." In such works as *Studies in the Structure of American Economy* from 1953, and *Input-Output Economics* from 1966, he broadened the interactive models of Kenney and others into an advanced mathematical model that showed the interdependence of the elements of the economic system. Leontief applied this technique to the American economy and obtained interesting results in the areas of international trade and the economy of natural resources. The technique became a basis for planning in many non-market economies. Leontief received the Nobel Prize for Economics in 1973.

[31] Richard Freeman, *Can Labor Standards Improve Under Globalization?* Harvard University, 2003; *What Workers Want*, Harvard University, 1999.

[32] James Medoff, *Labor's Capital, Business Confidence, and the Market for Loanable Funds*, Harvard University, 2004; http://econweb.fas.harvard.edu/hier/2004papers/HIER2042.pdf , 7.6.2007.

productivity depends on the specificity of institutions. By allowing workers voting rights and rights to just complaints, some labor unions obviously contribute to an increase in performance. Others exert a negative influence. When all this is reduced to concrete cases, neoclassical economic thought overlooks the relevant variable factors of worker motivation and the interaction between workers and management. Studies such as those by Freeman and Medoff are usually dismissed in the following way: "This is very interesting but – it isn't economics."

According to the prevailing definition – that economics is the study of the distribution of scarce resources, namely the study of ways in which a free and perfect market directs scarce resources so that they can be used in an optimal way and secures the maximum profit, standard neoclassical economic literature offers little useful insight into the sources of technological innovations. There are historical examples indicating situations in which a free and perfect market did not direct scarce resources to where maximum profits were to be obtained. For example, economic resources in the US on the eve of World War II were insufficiently used because of high unemployment. Demand and consumer purchasing power had declined. Thus, firms did not have buyers for their products, their capacities were unused and, quite normally, the result was non-rational capital investment. In addition, the war that came forced the government to increase taxes and public loans and to abruptly invest almost a third of the gross national program into war production. The result were: 1) a fall in unemployment from 15% to 2%; 2) the average annual economic growth rate equaled 11% throughout the war; and 3) new and improved technologies were developed, of which many had post-war, peacetime applications and trained whole generations of qualified industrial workers. US economic development during World War II is a "textbook example" of the positive effects of redirecting certain economic activities that are characteristic of an economy with so-called perfect competition. In place of the mechanisms of the so-called perfect market economy, economic flows were regulated by way of huge state planning and the state control of investments, wages and prices. However, neoclassical economists do not devote too much attention to wartime economies since wartime economic experiences disturb the equations of their econometric models of a perfect competition economy. And therein appears a highly interesting and, at first sight, paradoxical situation. Namely, as already mentioned, neoclassical economists do not devote particular attention to war years. However, since World War II, there have been about 260 wars of various intensities and lengths of time. In more than 220 of them, the United States was directly engaged with its military forces and its products, and in almost all the others with its arms or in some other, indirect way.[33] Thus, when it comes to the real world, the wielders

[33] Božidar Raičević, *Ratovi kao motor privrednog razvoja SAD*, Kosovo@yurope.com (a private forum dealing with the NATO aggression against the Federal Republic of Yugoslavia).

of economic power, especially in the US, are very well acquainted with wartime economic experiences and these experiences, as we have seen, have been very aptly applied in the US, especially during World War II. However, at the same time, the application of abstract mathematic models and the deductive method continue to be taught at universities. Quite simply the real world is going in one direction, and economic theory in another. One gets the impression that a deliberate accent is being put on the further improvement of abstract mathematical models and deductive methods, while the real world is being marginalized. It is no accident that Joan Robinson concluded that the production function is an effective method of stultification, where the political economy student is given the formula Y = f/ c,k/, after which the discussion moves on to the next question in the hope that the student will forget to ask in which units *k* is measured.

Lately it seems that a trace of doubt has infiltrated the economic periodicals themselves. In a remarkably rich article that appeared in the "Journal of Economic Literature," entitled "The Rhetoric of Economics," Donald McCloskey[34] of Iowa University wrote that economics is teeming with metaphors that have no application in empirical reality but which are accepted literally, because they are expressed in mathematical language. McCloskey quotes literary critic I.A. Richards and says: "To say that markets can be represented with 'diagrams' of supply and demand is no less a metaphor than to say that that 'the west wind is the breath of autumn itself.' (…) Each step in economic reasoning, even in the reasoning of official rhetoric, represents a metaphor. It is said that the world 'resembles' some complex model, and that its measurements are similar to some easily measurable variable close at hand." [35]

Quoting Richards, McCloskey says that modern economics is in many ways a metaphor that has spun out of control. It not only stylizes and wrongly formulates what is occurring within the narrow confines of economics, but also stretches its theory of the rational "homo economicus" into those spheres of life in which values other than material maximization prevail, about which, however, there are no data. One such example are the economic theories about the family, according to which even children are material goods.

[34] Donald McCloskey, *Rhetoric of Economics*, Univ. of Wisconsin Pr, SAD, 1998.

[35] Ibidem, p. 232.

6. Instead of a conclusion

As Thomas Kuhn observed in his book, "The Structure of Scientific Revolutions," scientific conflict is always the most intense along the boundaries of some established scientific paradigm that abnormal observations bring into question. The "overthrow" of a scientific paradigm, says Kuhn, is very similar to a political revolution that seeks "to change political institutions in a way which those institutions forbid." *Suspicion toward empiricism in economics is a reliable sign of a regime's uncertainty.*

Thomas Kuhn's most oft-repeated thought is that no paradigm can be displaced by proof, but only by another paradigm. And it seems that there is no dissenting paradigm that is able to take root in economics. In this way, economic orthodoxy is strengthened by ideology, the profession's sociology, and the policy regarding whose works are published, who is promoted and whose research will be financed. In the economic profession, the free market of ideas is just one more market that doesn't function like the model.

References

- Baumol, W. – Panzar, J. – Willig, R.: *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Jovanovich, New York, 1982.
- Freeman, R.: *Can Labor Standards Improve Under Globalization?*, Harvard University, 2003.
- Freeman, R.: *What Workers Want*, Harvard University, 1999.
- Friedman, M.: "The Counter-Revolution in Monetary Theory", *Occacional Papers*, no. 33.
- Friedman, M.: "The Role of the Monetary Policy", *American Economic Review*, 58, 1968.
- Friedman, M.: *Essays in Positive Economics*, New York, 1953.
- Friedman, M.: *Teorija novca i monetarna politika*, Izdavačko preduzeće "Rad", Belgrade, 1973.
- Henderson, J. V. – Ioannides, Y. M.: "A Model of Housing Tenure Choice", *The American Economic Review*, 1984.
- Katouzjan, H.: "Evolucija ekonomskog metoda: od političke ekonomije do pozitivne ekonomike", *Gledišta*, no. 5-6, 1985.
- Kun, Tomas S.: *Strukturata na naučnite revolucii,* Magor, Skoplje, 2002.
- Kuttner, R.: "O stanju u oblasti ekonomije", *Pregled* no. 236 for 1986/87, Beograd, taken from *The Atlantic* magazine, p. 41.
- McCloskey, D.: *Rhetoric of Economics,* Univ. of Wisconsin, USA, 1998.
- Medoff, J.: *Labor's Capital, Business Confidence, and the Market for Loanable Funds,* Harvard University, April 2004;
- http://econweb.fas.harvard.edu/hier/2004papers/HIER2042.pdf, 7.6.2007.
- Miede, J. E.: *Stagflation: Wage-Fixing*, vol. I, George Allen and Unwin, London, 1982.
- Nikitin, S.: "Inflacija narube`i 80-h godov", *Voprosy ekonomiki 3/82*
- Philips, A. W.: "The Relation Between Unemployment and the Rate Change of Money age Rates in the United Kingdom 1962-1975", *Economica*, vol. 25, 1978.
- Robinson, Dž.: *Ekonomska filozofija*, Istraživačko-izdavački centar SSO Srbije, Belgrade, 1981.
- Samuelson, P. A.: *Economics*, McGrow-Hill, New York, 1948, 7th edition, 1967.

Original scientific paper
Received: June 22nd, 2007
UDK 339.543.624(7)

Assistant Professor Biljana Prlja, PhD
Geoeconomics Faculty, Megatrend University, Belgrade

THE NORTH AMERICAN FREE TRADE AREA (NAFTA): POTENTIAL LESSONS FOR SERBIA

Abstract: *Since the fall of the Berlin Wall and the defeat of the state-socialist development model, the concept of liberal economics has emerged as victorious in international economic relations. This victory of the liberal concept stands as one of the reasons for the formation of the North American Free Trade Association (NAFTA). The formation of such a zone would have been unthinkable as late as 1988. The victory of the Conservative Party in Canada and the changes in Mexico's foreign policy in 1989 paved the way for the formation of the Free Trade Zone between the US and Canada, which subsequently expanded to include Mexico. For the first time in history two developed global economies and an underdeveloped economy were a part of the same trade zone. Mexico's experiences from its accession to NAFTA may hold lessons for Serbia and its process of integration into regional free trade zones and its future integration into the European Union.*

Key words: *Trade liberalization, NAFTA, Serbia, CEFTA, European Union*

1. Introduction

Since the fall of the Berlin Wall and the defeat of the state-socialist development model, the concept of liberal economics has emerged as victorious in international economic relations. According to its definition from the end of the 18th century, "classical liberalism" implies the rule of the law, individual freedoms, private property and free trade. With the breaking out of World War I in 1914, and especially starting with the 1920s, the global liberal economic order ceased to exist. It may be said that the period between the start of World War I and the fall of the Berlin Wall was primarily marked by economic and financial nationalism, despite the attempts of nation-states to establish global economic institutions such as the International Monetary Fund and the World Bank, and the signing of the General Agreement on Tariffs and Trade (GATT).

The fall of the Berlin Wall and the defeat of the state-socialist concept of development marked the victory of the liberal-democratic model, which has become the only accepted development concept. The vacuum on the global eco-

nomic and political scene formed by the defeat of the socialist bloc still exists and, thus, it cannot be said that an alternative development model currently exists.

The basic components of the liberal economic model are:[1] a market economy, a free market of labor and other factors of production, a market-oriented monetary system, free banking and free trade. Ideally, the import and export of all products and services should be freed of tariff barriers, subventions or any other government activity. The government should not forbid or regulate the free flow of people and money from one country to another.

This victory of the liberal concept and, within it, the concept of free trade, stands as one of the reasons for the formation of the North American Free Trade Association (NAFTA). The formation of such a zone would have been totally unthinkable as late as 1988. NAFTA represents the first case in the history of economic relations in which a single free trade zone has connected two of the world's developed economies (the United States and Canada) with a Third World economy (Mexico). In this sense, it represents an extremely interesting topic of study, especially in terms of the lessons it may offer to other underdeveloped countries and their strivings to connect with more developed, market economies. Thus, the first part of this study will provide a more precise definition of the concept of the free trade zone and of classical economic theories representing the concept of free trade. The second part of this work will explain the strategic context that led to the signing of NAFTA. The third part will analyze the contents of the agreement on forming a zone of free trade in North America. The fourth part of the work shall deal with the consequences of NAFTA's signing for Mexico's economy, which will lead to the final portion of the study, where possible lessons for Serbia in the course of its own integration into regional free trade zones will be formulated.

2. The concept of a free trade zone and classical economic theories on free trade

There are various forms of economic integration: a system of tariff preferences, a free trade zone, a customs union and an economic union.[2] It is obvious that this division has as its starting point the role and function of duties, customs and the trade regime, which represent the most significant defense mechanisms in a free trade system. Still, it may be said that such a division represents an attempt to rationally present something that is, in reality, a mere difference in degree.

[1] R. Ebeling, *A Liberal World Order*, http://www.fff.org/freedom/0991b.asp (11.6.2007)

[2] R. Tamames, B. Huerta, *Estructura económica internacional (20a edición)*, Alianza Editorial, Madrid, 2003, p. 207.

Free trade zones are most easily defined as something between tariff preferences and a customs union.[3] While preferential systems[4] are a thing of the past, free trade zones are often unstable formations, have a transitional character and are most often either transformed into customs unions or disappears.

A free trade zone is an area formed by two or more countries that remove trade or customs barriers between each other either at once or gradually, but who retain their own customs systems or trade regimes toward other countries. According to Article XXIV of GATT, countries that establish a free trade zone must remove most of the trade barriers between them and this must be done within a reasonable time period.[5]

Theorists who favor economic integration most often point to the following advantages:[6] the establishment of economies of scale; increased competition on the newly established market; improvement in balance of payments imbalances due to an increase in savings in convertible currencies; the possibility of developing activities difficult to organize under isolated market conditions (primarily because of the individual markets' insufficient size); increased bargaining power in relation to other regional groups; the waging of more balanced national economic policies; the introduction of mid and long-term structural reforms that, when the market is left to itself, are quite difficult to carry out, as well as the possibility of achieving accelerated economic development (rather than just growth of the gross national product – GNP).

The first theorist of liberal economics was Adam Smith who, in his classic work, "The Wealth of Nations," wrote about the "invisible hand of the

[3] A customs union is a form of maximal integration that can exist between the economies of two or more countries which were previously separate. The introduction of a customs union assumes an immediate or gradual removal of all customs and trade barriers to the flow of goods and services between the countries forming the union. Differently from free trade zones, a customs union assumes the erection of common customs system vis-à-vis third countries. It is important to keep in mind that it is difficult for a pure customs union to occur in practice. The establishment of free trade without customs barriers or quantitative barriers cannot but create huge problems that are a result of the existence of differing monetary, fiscal, transport and other systems. As a result, parallel with the establishment of a customs union it is necessary to harmonize all the previously mentioned elements, which all together make up the institutional framework of each economy. In other words, when efficiently consolidated, a customs union most often winds up, out of pure need, becoming an economic union. This was the experience of the European customs unions of the 19th century, especially Germany and Italy (in which process their nation-states came into being).

[4] The system of tariff preferences is very specific type of partial "integration." What it basically means is that several customs territories mutually approve a series of non-transferable tariff concessions. With the signing of GATT, all member-countries have accepted the most-favored-nation clause for all the club's members, which has lessened the attraction of signing tariff preferences.

[5] R. Tamames, B. Huerta, p. 211.

[6] Ibid., p. 212.

market."[7] His theory of absolute advantage is based on the presumption of free competition and the mobility of the factors of production and of final products on the international market. Smith thought that the basis of trade between nations lay in the advantages of specialization and of the division of labor on the international level. Countries rationally consent to specialize in the production of the goods they can produce at a lower cost and then export, while at the same time importing goods whose production would cost them more relative to the cost of production of that same good in the country from which they are importing them.[8] For this reason it is necessary to allow the free flow of goods and services between countries.

David Ricardo posited a theory by which international trade is determined by the differences in relative production costs instead of in absolute costs. All countries will participate in international trade regardless of absolute advantages. In this way, the country that achieves absolute advantage in the production of all products will specialize in the production of the product for which it achieves the greatest efficiency in production and for which it has comparative advantage, while a country that possesses no absolute advantages will specialize in the production of products it can produce at the relatively lowest cost. By way of specialization on the basis of comparative advantages, the country that has absolute advantages in all goods achieves the maximum gain, while the country with no absolute advantages would still acquire a greater fund of available goods through trade than it would have in the absence of trade.[9] Thus, in accordance with this, it is not difficult to see that these two classical theories of international trade are energetic champions of trade liberalization or free trade.

It may be said that the application of Keynes' theory in practice (1945-1970) marked the abandonment of liberal economic theory. In 1936, Keynes published his work "The General Theory of Employment, Interest and Money," certainly one of the most influential books of the twentieth century. In the course of more than three decades that followed, the ideas put forward in that book formed the basis of macroeconomic policy in the US as well as other developed countries.[10]

The book put forward the following theses: 1) there is no natural tendency in a given economy that would restore its balance following a period of recession; investment demand may be insufficient to take advantage of all the population's savings under conditions of full employment; 2) when an economy has less than full employment, monetary policy alone will not be able to reanimate investment and contribute to the restoration of full employment, and 3) if the goal is full employment following a recession, it may be necessary to control the

7 A. Smith, *The Wealth of Nations*, Modern Library Edition, Random House, New York, 1937.

8 J. Kozomara, *Tehnološka konkurentnost*, Ekonomski fakultet, Beograd, 1994, p. 16.

9 J. Kozomara, p. 19.

10 G. Walton, H. Rockoff, *History of the American Economy, 10th Edition*, Thomson, South-Western, USA, 2005, p. 545.

level of private investments, along with organizing government-financed public works, as needed. This last point Keynes' students Alvin Hansen, Abba Lerner and Paul Samuelson interpreted in such a way as to claim that the economy could be kept in balance if state spending was increased or if tax rates were lowered, or if both these policies were carried out during a recession. Of course, if full employment is achieved and if, as a result, the threat of inflation appears, these measures should be reversed.[11] It truly seemed during the entire post-war period that increased state spending, in place of trade liberalization, could be the key impulse for economic development, all with the goal of reducing unemployment. It seemed, therefore, that Keynes' recipe for ending the Great Depression was the correct one. Stagflation[12] represented a fundamental challenge to Keynes' theory. As a result, economists and politicians in power began to pay more attention to the ideas of Milton Friedman. He was an economics professor at the University of Chicago and championed the idea of the free market, which was a return to the liberal theory of economic development. Friedman presented several theses that left a fundamental mark on the economic policies of all developed countries from the end of the 1970s to the end of the 20th century: 1) that the relationship between inflation and unemployment, as described by Keynes, was of a temporary character (Keynes claimed that low inflation rates ensured the disappearance of unemployment, which was brought into doubt by the appearance of stagflation); 2) that inflation was a monetary problem that could be solved by an increase in the money supply, but only by way of a stable, slow-growing rate, and 3) that flexible currency rates were better than fixed ones. These ideas of Friedman's were called "monetarist" ideas while the school of thought came to be known as the Monetarist School. The application of monetarist ideas in practice became the basis of monetary and fiscal policies during the 1980s and early 1990s. Inflation was halted during the early 1980s, but at the price of an exceptionally sharp recession. Even though inflation was never reduced to zero, it was still held under control for the remainder of the twentieth century.[13] Friedman's theory was responsible for restoring all economic theory and practice to the classic liberal postulate of no state intervention in the economy and the return to the championing of the advantages of free market functioning, on both the national and international levels.[14]

11 Ibid., p. 546.

12 Stagflation is the simultaneous appearance of inflation and economic stagnation, which was impossible according to Keynes' model.

13 G. Walton, H. Rockhoff, p. 55.

14 From the end of the 1980s, the International Monetary Fund and the World Bank started to apply the so-called Washington Consensus, which was initially formulated by British economist John Williamson. It is thought that the Washington Consensus represents a return, in both the national and the international context, to the postulates of classical liberal theory. The ten points of the Washington Consensus are: budgetary discipline, restrictive tax policy (with the goal of reducing public spending), market-defined inter-

3. The strategic context of the signing of NAFTA

The signing of the NAFTA agreement was made possible by a change in the strategic context of the economies of the three member-states. NAFTA is an agreement on the creation of a free trade zone between Canada, Mexico and the US, which came into force on January 1, 1994, creating a trade bloc with almost 330 million people and an overall GNP of 11.6 trillion dollars (2003). NAFTA eliminated the trade barriers between Canada, the US and Mexico – some immediately and others during the next 15 years. In this way, the free trade concept prevailed for the long term, in the entire region.

With the fall of the Berlin Wall and the disappearance of an "alternative development concept," political discussion died down. In the NAFTA signatory countries, it was the economists that gained a monopoly on defining and promoting the advantages of the signing of such an agreement. Their predictions were mainly founded on classical economic theory, which insists on the advantages of international specialization and on trade liberalization. NAFTA critics were mainly not economists and they didn't succeed in offering adequate, methodologically explained counter-arguments. Mainstream expert opinion often described them as "ignorant panic-mongers."[15] NAFTA US advocates presented themselves as cosmopolitans and optimists fighting for a better future. Their opponents were often described as myopic, ignorant nationalists.[16]

Orthodox trade theory offered NAFTA advocates the main justification for signing the agreement: when a country specializes in the production of that which it produces the most successfully, overall economic gain increases and productivity rises. Profits grow because of a more efficient international division of labor, as each country produces those products through which it achieves comparative advantages, while gaining cheap imports in return. This sort of explanation also allows for the possibility that trade liberalization might also produce losers (for example insufficiently competitive workers or entrepreneurs) but its proponents consider that the overall gain is sufficient to compensate the losers. Thus, theoretically speaking, liberalization's winners could compensate the damages suffered by the losers if the state were to define an adequate transfer mechanism.

The champions of this theory thought that Mexico would gain more from NAFTA than the US, since trade between the two made up a large portion of

est rates, market-defined foreign currency rates, trade liberalization, the full openness of economies for the purposes of stimulating foreign direct investments, privatization of the economy, deregulation of the economy, protection of private property and the introduction of the principles of freedom and democracy.

15 D. Arsen, *The NAFTA debate in retrospect: U.S. perspectives, in: K. Roberts, M. Wilson, (eds), Policy Choices: NAFTA – Free Trade Among Nations*, Michigan State University Press, Institute for Public Policy, SAD, 1996, p. 38.

16 Ibid.

Mexican exports but only a small portion of American exports. It was also thought that immigrational pressure from Mexico into the US would decrease since NAFTA would enable Mexico to develop.

However, in the US, as well as Canada and Mexico, NAFTA main opponents could be defined as "cosmopolitan opponents."[17] They didn't oppose free trade or economic integration but were primarily worried about the principles by which North American integration was to take place. They thought that overall social gain could not be achieved through the actions of the private business sector.[18] The way they saw it, NAFTA was conceived as an elitist business strategy that increased the power of transnational companies (TNCs) while decreasing the power of the workers and governments. Hence, NAFTA could be defined as a continental "bill of rights" for the TNCs. The cosmopolitan opponents thought that, even though NAFTA secured a continental market, it did very little in terms of ecological protection or workers' rights.

The critics concentrated on underlining the unequal distribution of NAFTA benefits among the three member-states. The way they saw it, by locating their production in Mexico, the TNCs would achieve lower production costs while retaining free access to the North American market. In the same way, the option of being able to move capital to Mexico would have a consistently negative effect on wages and business regulations, both in the US and Canada. Critics were also arguing that NAFTA would not bring higher wages to Mexico. According to them, Mexico's labor surplus, powerless labor unions and efforts to maintain macroeconomic stability would prevent the growth of wages in Mexico, even if foreign direct investments (FDI) were to increase. The resulting decreased costs would not bring lower prices for the final consumers but only higher profits to the TNCs. In this way, NAFTA would benefit those on the high end of the US earnings scale, while its costs would be borne by those on the lowest rungs of the ladder.

As far as the strategic context in the US is concerned, NAFTA represented a reply to the economic challenges coming out of Europe and Japan and a means of increasing American power in the Western hemisphere. For the first 25 years after the end of World War II the United States led the world in growth of labor productivity, which allowed it to build an internationally competitive economy and retain a position of unchallengeable import in international institutions. As Europe and Japan began to reduce the differences in labor productivity, American producers began to encounter serious competition both on the domestic and the international market. As a result, during the last two decades the American business sector has consistently sought to increase its competitiveness, both through increasing labor productivity and by lowering costs. A clear strategy for increasing the competitiveness of American produc-

[17] For a detailed discussion of the term "cosmopolitanism," see: A. Đilas, *Najteže pitanje – Eseji*, Artist, Subotica, 2005, p. 196-229.

[18] D. Arsen,, in: K. Roberts, M. Wilson, (eds), p. 40.

ers was the establishment of production facilities in low-wage countries, where goods for the American market were to be produced. Among such countries, Mexico had a significant advantage due to its geographical location.

However, all developed countries always need to compare potential lower costs against the tendencies of all governments in underdeveloped countries to regulate and direct FDI flows by way of legislation. In this sense, NAFTA enabled the US to establish rules that would protect the rights of foreign investors as well as to remove the rules established by the Mexican government in the domain of regulating FDI flows within the Mexican economy. At the same time, the United States sought to open up those Mexican sectors that were previously closed to FDI (primarily the financial and energy sectors).

The advantages brought by NAFTA to American producers in relation to producers from any other country are great. It is important to emphasize that Japan is not a member of any regional free trade zone, As for the European Union (EU), it has no members with wages as low as Mexico's. In addition, the EU compels its members to introduce extremely expensive social harmonization measures. Differently from the EU, NAFTA is not a customs union, allowing the US a high degree of autonomy in its relations with all non-NAFTA countries. By way of rules about origin, the United States ensured that products from other developed countries could not fall under NAFTA and, thus, take advantage of the free trade zone.

As far as Canada is concerned,[19] its first free trade experiences with the US are tied to the 1960s and the agreements on the auto industry and common defense. From the Canadian standpoint, both these agreements functioned extremely well. The automobile and defense industry producers in both countries received duty-free access and domestic treatment from the other country, within the framework of previously agreed quantities, contents and quality standards. Both trade and investment grew significantly in both directions.

Nevertheless, the success of these two agreements and the strong multilateral desire to expand free trade dating back to the establishment of GATT in 1947 did not allay the worries of various Canadian governments at the prospect of broadening free trade with the US, fearing cultural domination and loss of sovereignty. However, during the mid 1980s the Canadian economy was in pretty bad shape and the Mulroney government thought that a more drastic change was needed. As a result, the main theme of the 1988 elections was free trade with the US – for or against. The Conservatives won and a free trade agreement between Canada and the US was signed as an added means of pressuring the Canadian economy to increase its productivity, combined with a desire to establish a conflict resolution system that would prevent Canada from being victimized by aggressive American trade measures.

[19] D. Wismer, *Perspectives from Canada*, in: K. Roberts, M. Wilson, p. 7.

It's quite certain that the said free trade agreement resulted in greater Canadian productivity and that it brought benefits both to Canada and the US. However, the agreement didn't make a clear distinction between dumping and subventions, while Canadians were often angry and powerless before the actions of American business. A great number of disputes arose in the areas of wheat, iron and light lumber.

When Mexico and the US announced the signing of a free trade agreement, the Canadians felt that they were not ready for something similar but that, at the same time, they couldn't allow themselves the luxury of being excluded. There was a risk that the US would agree to a series of bilateral relief measures with Mexico and thus become a so-called hub, with Canada being turned into just one of many spokes in the wheel.[20] At the same time, Canada wanted to encourage Mexico's economic development, recognizing the great potentials of an 85-million strong market. As a result, Canada decided to join NAFTA, which wound up being a positive experience for it.

Canadian exports to the US grew significantly, along with exports to Mexico. However, Mexican exports to Canada grew even more. Still, during the first six month after the agreement was signed, the scope of Canada's trade with both countries drastically increased. During 1994, exports and imports with the US grew by 18% compared to 1993, while exports to Mexico grew by 24% and imports by 36%.[21] Canada did not suffer any significant delocalizations, while its auto industry profited exceptionally.

Nevertheless, the main reason for Canada's decision to join NAFTA lay in the formation of working groups that were to deal with subsidies and dumping. Canada's economy is extremely dependent on exports to the US. One third of all US FDIs are in Canada, where American FDIs control about 50% of Canada's industry, including 75% of its oil industry, 90% of its automobile industry, 90% of its film industry, 85% of its music industry and 60% of its television industry.[22] Thus, by acceding to NAFTA Canada was seeking security from American protectionism. The trade barriers between the two countries were already negligible at the time of NAFTA signing. However, Canada was seeking to secure itself in the area of non-tariff barriers (in the first place US anti-dumping measures).

Anti-dumping laws allow domestic production to benefit from the introduction of penalty duties on imports whose price is judged to be lower than the "normal" price charged in the exporter's market. Non-tariff barriers are

[20] The term "hub and spokes development model" is frequently used; it draws its name from the bicycle wheel, which consists of several spokes emanating from the same center. In a transitional sense, it means that one geographical location has been chosen as the center, with the spokes being defined by the points of supply and delivery.

[21] D. Wismer, *Perspectives from Canada*, in: K. Roberts, M. Wilson, p. 9.

[22] A. Alvarez Bejar, S. Martinez Aguilar, "Significado del Tratado del Libre Comercio de América del Norte para México", ICE, no. 795, November-December 2001, p. 64.

placed on imports considered to be subsidized by the government of the country of export. Section 301 of the American Law on Trade of 1974[23] gives the government great leeway in adopting punitive measures against any foreign government that practices "unreasonable or discriminatory measures so as to impede American trade." Section "Super" 301 relates to the Law on American Competitiveness of 1988, which further expands the actions allowed by Section 301 (for example, allowing the introduction of trade sanctions in case of violation of the rights of American intellectual property or against countries with constant trade surpluses in relation to the US).[24]

The US had already undertaken such actions against Canada, most often in the areas of agriculture, fishing, lumber and natural resources. It was thought that a free trade agreement between Canada and the US would remedy these measures, which, however, didn't happen. Therefore, it was hoped that Canada would achieve through NAFTA what it didn't manage to achieve through the agreement on the establishment of a US-Canada free trade zone.

As for Mexico, the main reason for signing NAFTA lay in its large need for foreign capital. The debt crisis of 1982 was the catalyst for all the waves of liberalization that followed. During the 1980s, the Mexican economy stagnated due to its large debt-servicing burden and restrictive monetary policies. Still, the idea of advancing toward something as big as the idea of NAFTA seemed totally impossible as late as 1988. Having in mind Mexico's historical inertia and its subjection to strong state interventionism, as well as its tendencies toward high protectionism, the signing of an agreement on a free trade zone with the US and Canada looked more like an illusion. However, the internal and external liberalizations of 1989 carried out by Mexican president Salinas de Gortario, opened the way for Mexico's acceptance of George Bush Sr.'s counsel regarding the benefits of such an agreement.[25] The year 1989 saw a great change in Mexico's foreign policy, as well as structural changes in the institutional system: the state's role in the economy was redefined (by limiting the state's role in the economy and drastically reducing the participation of para-state companies); there was a strong process of deregulation of economic activity as well as of opening up Mexico's economy. This opening caused, on the one hand, a reorientation of the production apparatus toward export-oriented industrialization and, on the other, the elimination of protection barriers enjoyed by the industrial and financial sectors since the 1940s.

Through NAFTA Mexico sought to stimulate the inflow of foreign investment as well as to turn the country into an export power in the industrial sector. In addition, the Mexican government thought that, besides the inflow

[23] D. Arsen, in: K. Roberts, M. Wilson, (eds), p. 57.

[24] Ibidem

[25] R. Tamames, B. Huerta, *Estructura económica internacional (20a edición)*, Alianza Editorial, Madrid, 2003, p. 286.

of foreign capital, liberalization would lead to greater employment and the modernization of Mexico's backward industry.

Within the framework of NAFTA negotiations, the US position in relation to Mexico and Canada was favorable due to several factors:[26]

- In the first place, Mexico needed foreign investments. Mexico had obligations both toward the IMF and the World Bank. In this sense, in order to strengthen its relations with these institutions, Mexico decided to carry out the reforms demanded of it, first of all by unilaterally removing trade barriers, even before signing NAFTA. As a result, prior to signing on to NAFTA, Mexico had significantly reduced its trade tariffs (from 23% in 1985 to 6.5% in 1990), and then reduced them additionally prior to NAFTA implementation. American tariffs for Mexican products were even lower. Since a free trade agreement between the US and Canada was already in existence, it may be said that free trade was introduced to the North American continent even before NAFTA came into being.
- Sharp internal political divisions additionally strengthened the US negotiating position. While American trade negotiators are extremely sensitive to the opinions of Congress and powerful interest groups, the Canadian parliament is much less susceptible to special interest influence, meaning that it had very little influence on the negotiating process. As long as Canada's prime minister could control his parliament, this trade agreement could be passed on the wings of traditional party discipline. In Mexico, the PRI (the Institutional Revolutionary Party) had controlled the political system for the previous 70 years, which means that Mexico's congress was not a problem either. The negotiators were totally free of domestic pressure. This of course, proved to be a weakness rather than an advantage. The Mexican negotiators could not hide behind "domestic political pressure" in order to gain concessions in the negotiation process. [27]
- As the largest economy, the US was in a position to demand a great number of concessions. The significance of a strong multilateral system with clearly defined rules of the game is greater for smaller and less developed countries than for large, developed countries, since large countries are more easily able to protect their interests in an unregulated environment.

The compensations demanded by the United States as a condition to signing NAFTA had to do with the following: [28]

[26] D. Arsen, in: K. Roberts, M. Wilson, (eds), p. 43.

[27] The theory that explains this situation is Putnam's "two-level game" theory.

[28] D. Arsen, in: K. Roberts, M. Wilson, (eds), p. 44.

- the introduction of special clauses for the purposes of protecting import-sensitive branches of the American economy (fruits, vegetables, sugar) as well as establishing rules of origin for many industrial products (textiles, metallurgy, automobiles and auto parts);
- domestic treatment for American TNCs operating in Mexico and Canada, so that American companies could share in the greatest part of the growth benefits in these two countries; the main demands had to do with eliminating maximum limits for the percentage of invested foreign capital, minimum limits for local investment, and the removal of various legal procedures for foreign investors;
- sufficient ambiguity in defining terms such as "subsidies" and "unfair trade behavior," so that the conflict-resolution mechanism could allow the US to apply these measures in relation to Canada and Mexico if it so decided.

Although the positive sides of NAFTA were felt in all three countries, it's clear that the American negotiating position ensured that the greatest benefits would be felt within the US itself. All US companies received domestic treatment on the entire continent and gained open access to the Canadian and Mexican banking sectors. In addition, the United States tried to use NAFTA within the GATT framework so as to ensure the adoption of the same rules within GATT in the fields of investments, finance, intellectual property and agriculture. The US succeeded in its intention of transforming its economy into a regionally based economy, in order to more successfully enter global competition.

4. The contents of NAFTA

The NAFTA agreement came into force on January 1, 1994. It consists of 8 parts and 22 chapters, through which all three signatory countries express their desire to deepen their mutual relations in various areas: creating larger and more secure markets for products and services; working on the protection of intellectual property; creating adequate conditions for the liberalization of investments; there are also two annexes relating to environmental protection and the protection of workers' rights. [29]

[29] The NAFTA agreement came into being following a great debate in the US, which compelled US president Clinton to append two protocols to the basic NAFTA agreement: one dealing with ecological policy and the other with work conditions. The goal of these annexes was to silence the remaining voices of caution in the American Congress, which were primarily an expression of protectionist reflexes and the fear of immigration.

The essence of NAFTA can be expressed in several points: [30]

- following the model of free trade between Canada and the US, NAFTA includes an agreed time strategy for removing customs barriers in order to achieve an absolutely free circulation of goods;
- it introduces the free transit of business people through all three countries without even once raising the issue of removing barriers to the free circulation of workers;
- it provides for the harmonization of a complicated area dealing with the mutual recognition of professional diplomas;
- it grants the right of establishing any sort of financial institution, according to the rules of the country chosen by investors, along with defining a specific conflict resolution system for all institutions founded in this way;
- Mexico's energy sector has remained closed to foreign investments (which is provided by the constitution of the country); America's immigration policy has remained in place while Canada has excluded all its "traditional" economic branches from NAFTA.

The Commission, NAFTA main controlling organ, has its seat in Washington. It consists of a Council of Ministers, an International Secretariat for Coordination, and three national administrative commissions. In accordance with free trade zone regulations, all three NAFTA signatory countries retain their customs systems vis-à-vis the rest of the world.

Within the framework of the main parts of NAFTA are, in the first place, rules tied to "market access." These rules provide for the lowering of customs rates, the removal of quantitative restrictions on the flow of goods and services, and they contain all the regulations tied to the rules on product origin. Rules about origin are not uniform for all sectors covered by NAFTA. In addition, the main portion of NAFTA contains rules dealing with opening up to foreign investment. Having in mind that customs duties on the flow of goods between the three signatory countries were removed even before the adoption of NAFTA, it may be concluded that the most important part of the agreement is tied to investments, especially from the American point of view. The entire NAFTA agreement has about 2,000 pages, with the greatest part dealing with the regulation of investments instead of trade. Within these rules, the primary criterion is that of domestic treatment for investments from all three signatory countries, as well as mandatory international arbitration in case of dispute. In addition, this part contains the rules dealing with the opening up of the financial services sector, allowing the opening of insurance companies, banks and investment funds in all three member countries, in which all actors in

[30] R. Tamames, B. Huerta, *Estructura económica internacional (20a edición)*, Alianza Editorial, Madrid, 2003, p. 287.

the marketplace are granted domestic treatment. This part also provides for a repeal of all barriers to the free circulation of capital as well as the repeal of rights to expropriate foreign-owned property. The third important portion of NAFTA contains rules dealing with intellectual property, having to do with authors' rights and brand registration, as well as with industrial secrets, industrial design rights and even rights connected with new plant types (of great importance to agriculture). The fourth aspect, of no less importance, deals with the mechanism for the resolution of disputes and with all anti-dumping measures, compensational rights, the review of all local judicial actions and the establishment of bi-national judicial panels. The establishment of such panels is often pointed to as a great novelty brought by the NAFTA agreement. Through NAFTA, all investment flows between the signatory countries are fully liberalized, investments receive domestic treatment, while, in case of dispute, in lieu of national laws, disputes are settled by way of bi-national judicial panels, which opens up the way for the US to impose its will within the US-Mexico relationship. This part also encompasses the important area of public tenders. Finally, the annexes define the criteria having to do with certain aspects of protecting workers' rights and the environment.

5. NAFTA effects on the Mexican economy

Between 1993 and 2003, Mexico's foreign trade increased 1.9 times.[31] In analyzing the elements of such an increase it is first of all important to emphasize the change in the contents of the said foreign trade. Oil products, which between 1982 and 1985 represented an average of 60.3% of Mexico's exports, came to represent only 9.6% of exports in the 1993-2003-time period. In the same period, industrial periods took the place formerly held by oil products in Mexico's exports – between 1982 and 1985 industrial product exports grew at a rate of 32.7% while between 1993 and 2003 they grew by 86.6%.[32] In this way, Mexico had turned into an American "workshop."

In the same way, there has been a noticeable change in the dynamics of foreign capital flows into Mexico since NAFTA has come into force. Before NAFTA, foreign direct investments were at an annual level of 3.7 billion American dollars, while in the first decade of NAFTA existence they raised to 12.8 billion dollars annually. [33] It's important to emphasize the fact that the onset of NAFTA coincided with a period of great economic expansion in the

[31] J. Zabludovsky, "El TLCAN y la política de comercio exterior de México: una agenda inconclusa", *ICE*, no. 821, Marzo-Abril 2005.

[32] A. Alvarez Bejar, S. Martinez Aguilar, *ICE*, no. 795, November-December 2001, p. 64.

[33] Ibid.

US, which was reflected in greater import demand. [34] In addition, the depreciation of the Mexican peso that occurred amidst the crisis in Mexico's balance of payments during 1994, combined with the right of preferential access for Mexican products on the markets of Canada and the US greatly contributed to an improvement in the relative prices of Mexican products, which significantly boosted Mexico's export activity.

NAFTA played a key role in strengthening Mexico's export dynamism as well as in the great inflow of foreign investment into Mexico. Also, NAFTA contributed to a change in the nature of trade between Mexico and its North American partners, by significantly strengthening intra-industrial trade flows. Integration in North America has resulted in significant changes in Mexico's business cycle. Not only was there a convergence with the rhythm of US economic activity but also Mexico's business cycle instability fell by almost 30%, while incoming speculative investments fell by about 40%.[35]

It is, thus, clear that accession to NAFTA significantly contributed to Mexico's achieving macroeconomic stability and to the strengthening of its institutional framework. NAFTA has taken over the role of "anchor" or "initiator" of modernization. Having in mind NAFTA contribution to the consolidation of Mexico's economic liberalization vis-à-vis the world's largest economy, it may be said that it has turned into a source of predictability and stability. In the same way, it has stimulated the adoption of modern policies as well as the creation of institutions of a true market economy, the promotion of competitiveness, the protection of intellectual property rights and, indirectly, a general modernization of regulatory rules in the economy.

It could be said that NAFTA had the effect of an initial spark for Mexico's general economic modernization, not only because of its role of promoter of the liberalization of Mexico's economy but, before all, because it compelled Mexico's economy to adopt the rigorous discipline necessitated by the existence of a true market economy and because it strengthened the institutional

[34] The increase in work productivity was achieved through investments in new technologies and the spread of the so-called new economy within the American economy. The symbols of the new economy were the dot.com companies, which totally changed the way in which business in the US – and, subsequently, in the rest of the world – was done. The rhythm of technological progress increased significantly. Two centuries earlier, the world had already gone through the Industrial Revolution, which caused the center of the economy to move from agriculture to industry. The appearance of the new economy called for a similar kind of change: the center of the economy shifted from the production of products to the production of ideas. The new economy is primarily characterized by large investments in the telecommunications and information technology sectors. The new economy encompasses: electronic systems of classification, selection, calculation, decision-making, design, animation, automated production processes and robotization; a better use of human resources through better knowledge management; electronic trade, as well as a total revolution in agricultural technologies.

[35] J. Zabludovsky, *ICE*, no. 821, Marzo-Abril 2005, ibid.

framework by introducing a number of permanent rules. In this sense, it is not an exaggeration to say that NAFTA represents one of the first and most important instruments of Mexico's long-term economic policies. Taking into consideration the fact that NAFTA broadened the horizons of economic planning and the fact that the introduction of NAFTA significantly increased the costs of any eventual repeal of trade liberalization, it may be said that this agreement also contributed to the achievement of a political transition in Mexico, because it reduced the uncertainties that have always accompanied transfers of political power in the country.

As far as the problems of Mexico's economic stability are concerned, it's important to underline that one of the main sources of concern during the 1980s was the great instability of the currency exchange rate, which even caused some economists to argue that the only solution to this problem lay in a permanent and absolute dollarization of Mexico's economy. In Mexico, periods of the peso's overvaluation and undervaluation relative to the dollar constantly alternated, which, in turn, stimulated either exports or imports. It's clear that the signing of NAFTA greatly remedied this chronic problem of Mexico's economy.

However, although it was claimed that NAFTA would contribute to the improvement and stabilization of the Mexican economy's growth, the real evolution of the GNP does not evidence any drastic changes: the annual growth of the real GNP according to 1993 prices, in the periods of 1980-87, 1987-94 and 1994-99, were 1.3%, 4.5% and 2.4%, respectively. [36] It may thus be said that Mexico as a country benefited from NAFTA, as distinct from the majority of its population.

In addition, the center of Mexico's extremely dynamic export center is occupied by a number of multinational companies of US origin along with some domestic companies. Mexico's entire exports are concentrated in a mere 312 companies, both domestic and foreign. If "maquiladora"-type companies are included, the number grows to a bit more than 3,000 – out of a total of more than 3 million companies in Mexico. [37]

Maquiladora is a term that cannot be literally translated, referring to companies engaged in the assembly of imported parts, which, following assembly, are then re-exported. It's obvious that American companies establish assembling companies in Mexico in order to take advantage of its cheap labor force. It's also obvious that American firms have concentrated on using cheap, unqualified labor in Mexico, specifically in the industrial sector. Such an orientation has brought a reduction in the wage levels of all employees across all sectors in all of North America. What this shows is that the "cosmopolitan critics" were right. The general minimum wage level in the entire region fell by 20%

[36] A. Alvarez Bejar, S. Martinez Aguilar, *ICE* no.795, November-December 2001, ibid.

[37] Ibid.

in the 1990-2000 period. This piece of data gains further significance once it's revealed that about 15% of all those employed receive the minimum wage. [38]

The industrial products registering the best export results are, in the first place, consumer goods such as automobiles and electronics. A third sector scoring good results is the textile sector, which is dominated by multinational companies. Analyzing by sectors, if the growth of intra-industrial trade and its concentration are analyzed, it's easy to see that Mexican-US relations have specialized in the production of auto parts and automobile assembly, with these sectors taking up 70% and 85%, respectively, of total trade within the automobile industry. In 1990, exports in these two sectors totaled 4.7 billion US dollars while in 1998 they totaled 21 billion dollars.[39] It is clear that Mexico is specializing in the production of several mid-size automobile models as well as in the production of agricultural machinery. In addition, a third of total exports are manufactured auto parts with a high degree of standardization. Thus, automobiles are Mexico's main export item, followed by auto parts and motors. These sectors are dominated by TNCs.

The electronic industry is a second sector enjoying dynamic export growth since the signing of NAFTA. Mexico has turned into the main supplier of the US market in this field, taking up the positions previously held by companies form Japan and Canada. As far as the textile industry is concerned, it's important to emphasize that, between 1993 and 1997, it enjoyed a more dynamic export growth rate relative to Mexico's total export growth as well as to the growth in industrial goods imports.[40] However, this export is dominated by production from maquiladora-type companies, which can be found primarily in the clothing sector but also in the textile industry and the chemical fibers industry.

As for NAFTA's effects on Mexico's regional distribution of production, it's clear that the great concentration of economic activity in Mexico's capital city that occurred during the 40 years of applying the "import substitution" model has ended. NAFTA has contributed to the consolidation of several highly specialized but territorially unconnected regional centers. Some of these are located in the north, while the others are nearer to the country's center.

Historically speaking, companies engaged in parts assembly have been located in the north, along the border with the US. Since the onset of NAFTA, some of these have relocated to the center-north region. These companies have specialized in the production of a limited group of products within a small number of industrial branches, specializing, naturally, in semi-finished goods, where wages are low.

Still, it's quite clear that the trade relations promoted by NAFTA are leading to a consolidation or formation of regions, sectors and industrial branches in a

38 J. Zabludovsky, *ICE*, no. 821, Marzo-Abril 2005, ibid.

39 Ibidem

40 Ibid.

way reminiscent of the creation of an "enclave-based" economy such as existed in the final third of the 19th century and the beginning of the 20th century. This was an economy with a small number of dynamic centers linked to foreign countries, which had no possibility of stimulating the growth of the remaining parts of the economy. One could speak of this being a typical instance of a so-called dual economy. As already emphasized, NAFTA is better defined as primarily an agreement on investments. The adopted rules grant "domestic treatment" for all investments originating with any of the NAFTA member-countries, while "most favored nation" treatment is accorded to investors from all other countries. Moreover, the principle of "minimal intervention" ensures investors from NAFTA member-countries the protection and security of international law, including the paying of damages in case of nationalization of export companies and products. It may be concluded that NAFTA increases investor rights and restricts national powers.

For Mexico, NAFTA means a trade balance of 200 billion dollars.[41] About 80% of Mexico's total trade is with the US and Canada. For Mexico, NAFTA has eased the process of economic structural adjustment, which international financial institutions had been recommending since the 1980s. Mexico's economy has transformed from a closed to an open economy. The import substitution model has been exchanged for an export-oriented one. At the same time, the economy, in which the state used to play a significant role, is today subjected to a low degree of intervention. NAFTA has contributed to a large inflow of FDIs to Mexico, which has contributed to the adoption of new mechanisms of production and the modernization of the financial and trade sectors. However, the main portion of FDI inflows is concentrated in only a few of the country's economic branches, sectors and regions, and these are dominated by investments from the US.

6. Lessons for Serbia

Since the onset of NAFTA, internal trade in the North American region has grown rapidly, reaching 570 billion dollars in the year 2000. This means that the scale of regional trade grew by 100% in a six year period. The main characteristics of these trade flows are:[42]

The United States is the main trade partner both for Mexico and Canada;

- Canada is the US's main trade partner, with 352 billion dollars in total, two-way trade;

[41] R. Tamames, B. Huerta, *Estructura económica internacional (20a edición)*, Alianza Editorial, Madrid, 2003, p. 288.

[42] R. Tamames, B. Huerta, 2003, p. 289.

- Mexico is the US's second-ranked trade partner, with 215 billion dollars worth of two-way trade;
- Canada is Mexico's second largest export market (9.4 billion dollars);
- During NAFTA first six years of existence, Mexico-US trade grew by more than 150%, or 17% annually;
- in the same period, total trade between Canada and the US also grew, at an annual rate of 10%;
- trade between Mexico and Canada has grown by 15% annually.

However, it should be taken into account that Mexico is showing tendencies of extreme specialization in the field of parts assembly, of sacrificing the activities of the agricultural sector, which support millions of Mexicans, and of reducing the role of large state concerns to that of raw materials exporters. All these tendencies have led to perceptions inside Mexico that NAFTA as an agreement has strengthened the asymmetry of power that already existed between the member-countries, thus exacerbating Mexico's social inequalities, which were already large even before the signing of NAFTA.

As for eventual lessons for Serbia, it's of course difficult to compare things that might be non-comparable, because Serbia is a small country in the European southeast, while Mexico is a large country with a 2,000 km long border with the world's most developed economy. Nevertheless, despite these differences, it is possible, through an analysis of the Mexico-NAFTA case, to gain a more realistic overview of the advantages to be gained by Serbia through EU membership.

Presently, Serbia has membership only in the Central European Free Trade Area (CEFTA), which was formed in the early 1990s along with the Baltic Free Trade Agreement (BFTA). Keeping in mind that EU expansion to former Eastern Bloc countries seemed like a distant dream at the time, it was decided that the creation of these areas would help prepare the economies of the countries of Baltic, Central and Eastern Europe for future full-fledged EU membership. It may be said that CEFTA and BFTA had two goals:[43] to test the preparedness of Central and East European countries for mutual cooperation by way of trade agreements, and to remedy the growing dependence of all the regions' countries on trade with the EU by reestablishing regional trade flows.

As for the first goal, it could be said that, during the 1990s, CEFTA and BFTA became the main instruments of the liberalization of regional trade flows. CEFTA grew from its initial three members to encompassing all of Central Europe. As for BFTA, although its membership didn't increase, in time the scope of the agreement itself broadened to include certain highly sensitive areas such as agriculture and fishing. Speaking of the second goal, both these agreements only

[43] R. E. Baldwin, "A Domino Theory on Regionalism," NBER Working Paper, no. 4465.

partially reduced the regions' dependence on trade with the EU. Still, both the agreements generated a positive influence on regional bilateral trade.

The countries of Southeastern Europe, including Serbia, are in a position similar to that held by CEFTA and BFTA member-countries in the early 1990s. By way of Stabilization and Association agreements, the EU is offering Southeastern Europe preferential trade agreements along with the prospect of full membership, just as it did with the European Agreements in Central and Eastern European countries. Still, EU countries are displaying great skepticism regarding the ability of Western Balkan countries to build long-lasting and harmonic mutual economic and political relations.

In addition to the great advantages they offered, the European Agreements had a great shortcoming: they did little to spur intra-regional Central European and Eastern European trade. Many economists have concluded that asymmetric trade liberalization can lead to "hub and spoke" problems.[44] It was thought that the European Agreements would form trade flows within which EU countries would become the "hubs" while the Central European countries would become the "spokes." This would mean that the Central European economies would be turned into satellites circling around the EU, with negligible levels of mutual, intra-regional trade. In a situation in which trade would be mainly dependent on the EU, Central European countries would become highly sensitive to negative shocks affecting the EU. In addition, there were also fears that the lack of trade liberalization between Central European countries would limit intra-regional competition.[45] It could be said that the dangers of the "hub and spoke" development model compelled the countries of Central Europe to more quickly adopt and apply plans for the development of regional trade integration.[46]

The European Agreements, CEFTA and BFTA brought a period of exceptional growth in exports from the Central and Eastern European countries. Between 1993 and 2001, total export earnings in BFTA countries rose by about 400%, while CEFTA countries' earnings rose by 250%. Between 1993 and 2001, BFTA tripled its exports into the EU, while CEFTA doubled them. Also, both CEFTA and BFTA succeeded in increasing their EU market shares. During the same period, the BFTA countries' market share rose from 0.11% to 0.26%, while that of the CEFTA countries rose from 2.13% to 4.08%. Intra-regional trade grew both within CEFTA and BFTA. Between 1993 and 2001, intra-regional trade within CEFTA doubled, while market share fell from 11%

44 See footnote 19. R. E. Baldwin, "Towards an Integrated Europe," *CEPR*, London, 1994.

45 Such competition is always positive, as companies in the region are on the same development level. Through mutual competition they can prepare for competition with EU companies.

46 A. Adam, T. Kosma, J. McHugh, "Trade Liberalization Strategies: What Could Southeastern Europe Learn from CEFTA and BFTA?", International Monetary Fund, Working Paper 03/239.

to 10%. Nevertheless, the results were not the same for all the CEFTA members. The Czech Republic and Slovakia scored a significant decline in market share in intra-regional trade. This was in the first place a result of a decrease in bilateral trade following the disintegration of Czechoslovakia. Hungary, Poland and Romania increased their share in intra-regional trade, while Bulgaria and Slovenia mainly retained their share of the intra-regional market.[47] The BFTA countries saw increases in both intra-regional trade and market share.[48] Between 1993 and 2001, intra-regional export earnings rose fourfold, while intra-regional market share rose from 8.4% to 9.6%.[49]

Even though both CEFTA and BFTA developed intra-regional trade, they didn't succeed in preventing a growing dependence on the EU market. In 1993, exports from CEFTA into the EU made up 56% of its total exports, while in 2001 this had risen to 68%. In 1993, 50% of total BFTA exports were going to the EU; in 2001, the figure was 55%. It could be said that the asymmetrical trade liberalization that formed the essence of the European Agreements was overly significant and did not succeed in preventing growing dependence on the EU on the part of Central and Eastern European countries.

As far as Southeastern European countries are concerned, including Serbia, Stabilization and Association Agreements (SAA) are a substitute for the European Agreements. The SAA form an institutional framework for closer political and economic integration. As in the case of the European Agreements, the SAA establish asymmetrical trade liberalization. Such a trade liberalization regime can contribute to the formation of a "hub and spokes" development model, as in the case of Central and Eastern European countries. Thus, the Southeastern European countries might come to the conclusion that the strengthening of their mutual trade relations could accelerate EU expansion to the south, as was realized by the Central and Eastern European countries when they were signing CEFTA and BFTA. A more intense linking of economies at similar levels of development, as a prelude to linkage with more developed European economies is not only desirable but necessary, since Mexico's experience clearly shows what can happen when the economy of an underdeveloped country links with two of the world's most developed economies. Regional linkage as a step on that road represents the best possible preparation and protection from the sharp competition that exists in the EU marketplace.

It's important to underline one more advantage on which Serbia may count, which can only be a distant dream for Mexico. The chief comparative advan-

47 Ibidem.

48 Intra-regional market share is the percentage of regional imports coming from the countries in the region. For example, in the case of CEFTA, the intra-regional market share is the ratio between CEFTA exports in relation to other CEFTA countries as a percentage of total CEFTA imports.

49 Ibid.

tage of the EU in relation to all other and different regional integration processes is the existence of structural and cohesive funds. The European Union is the only integration in the world that has developed such a financial instrument. Its existence is of decisive importance for all underdeveloped economies in its vicinity aspiring to future full membership. A great portion of the negative consequences of Mexico's accession to NAFTA would have been annulled had NAFTA developed a similar financial instrument.

Structural and cohesive funds represent the main financial instrument of the EU regional policy, whose aim it is to reduce the differences in degrees of development between all the regions of the Union's member countries. It may be said that the funds actively participate in the achievement of economic, social and territorial cohesion goals.

For 2007-2013, the EU budget for regional policy equals almost 348 billion euros, of which 278 billion euros are made up of structural funds, while the remaining 70 billion euros are cohesive funds. Both these funds make up about 35% of the Union's total budget and represent its larges financial outlay, right behind funds slated for the common agricultural policy.[50]

There are two kinds of structural funds:[51]

- the European Regional Development Fund (ERDF) is the largest fund; since its introduction in 1975, its use has been directed toward infrastructure building and investments in entrepreneurial activities;
- the European Social Fund (ESF), established in 1958, whose aim is the inclusion of the unemployed and discriminated population segments into the labor market, primarily through the organization of training programs.

In addition, the EU established cohesive funds in 1994 in order to accelerate the economic, social and territorial convergence between all the member countries' regions. These funds are open to all countries whose GNP per capita is less than 90% of the Union's average.[52] The goal of the cohesive funds is the financing of transportation and ecological infrastructure construction projects. Still, the use of moneys from the cohesive funds is partly conditioned. If the public debt of a country seeking to use money from the cohesive fund is greater than 3% of its GNP, no new projects will be approved until the deficit is brought under control.

From all the above it may be concluded that Serbia may count on a series of comparative advantages compared to Mexico, stemming from the very fact of its being located in Europe. Differently from Mexico, Serbia can and must insist on

50 EUROPA – Glossary – Structural Funds and Cohesion Funds: http://europa.eu/scadplus/glossary/structural_cohesion_fund_en.htm (15.6.2007).

51 Ibid.

52 Ibid.

regional trade liberalization as a means of strengthening the Serbian economy prior to entering direct competition with EU companies on the EU market. This is the only way for the monopolistic or oligopolistic market structures that presently characterize the Serbian economy to weaken. In the same way, differently from Mexico, Serbia can eventually count on structural and cohesive funds that will prevent its economy from turning into a so-called dual economy, and allow it to achieve a balanced regional development. All delays or slowdowns on the assigned path will only mean that Serbia doesn't possess sufficient strength to win the battle to improve its population's living standard, before all through a fight with the oligipolistic and monopolistic structures that make up the chief characteristic of its economy today.

References

- Adam, A. – Kosma, T. – McHugh, J.: "Trade Liberalization Strategies: What Could Southeastern Europe Learn From CEFTA and BFTA?", International Monetary Fund, *Working Paper* 03/239.
- Alvarez Bejar, A. – Martinez Aguilar, S.: "Significado del Tratado del Libre Comercio de América del Norte para Móxico", *ICE*, no.175, novembar-decembar 2001.
- Bacon, D.: *The Children of NAFTA – Labor Wars on the US / Mexican Border*, University of California Press, Berkeley, Los Angeles, 2004.
- Baldwin, R. E.: "A Domino Theory on Regionalism", *NBER*, Working Paper, no. 4465.
- Clement, W. (ed.): *Understanding Canada – Building on the New Canadian Political Economy*, McGill-Queen's University Press, Montreal, 1997.
- Đilas, A.: *Najteže Pitanje – Eseji*, Artist, Subotica, 2005.
- Ebeling, R.: "A Liberal World Order", http://www.fff.org/freedom/0991b.asp
- EUROPA – Glossary – Structural Funds and Cohesion Funds, http://europa.eu/scadplus/glossary/structural_cohesion_fund_en.htm
- Kozomara, J.: *Tehnološka konkurentnost*, Ekonomski fakultet, Beograd, 1994.
- Roberts, K. – Wilson, M. (eds): *Policy Choices: NAFTA – Free Trade Among Nations*, Michigan State University Press, Institute for Public Policy, SAD, 1996.
- Smith, A.: *Wealth of Nations*, Modern Library Edition, Random House, New York, 1937.
- Tamames, R. – Huerta, B.: *Estructura Económica Internacional (20a edición)*, Alianza Editorial, Madrid, 2003.

- Walton, G. – Rockhoff, H.: *History of the American Economy*, 10th Edition, Thomson, South-Western, SAD, 2005.
- Zabludovsky, J.: "El TLCAN y la política de comercio exterior de México: una agenda inconclusa", *ICE*, no. 821, Mart-April, 2005.

Scientific review article
Received: August 31st, 2007
UDK 005.71-022.51

Professor Božidar Leković, PhD
Faculty of Economics, Subotica, Novi Sad University
Slobodan Marić, MA
Faculty of Economics, Subotica, Novi Sad University

SMALL COMPANIES IN A GLOBAL ENVIRONMENT

Abstract: *Only flexible companies can hope to achieve business success in a changing environment, being flexible in organizational structure, business policy, system of management, etc. On the path to accomplishing their goals, companies are faced with a series of limitations, including falling market needs, intense competition, legal limitations, their home country's inadequate economic policy and many others. One of the most important limitations certainly lies in existing markets that are either stagnant or declining. This means that the existing markets (national) are becoming tight and insufficient and that companies are compelled to seek out new ones. As a result, an increasing number of factors is orienting companies toward foreign markets and the expansion of their business beyond national borders; in other words, processes of business internationalization and globalization are becoming unavoidable. A global marketplace assumes equal business conditions for all participants, offers great development chances, especially when it comes to increasing sales capacities, the securing of new resource markets, and improvements in competitive advantages. Despite their size, small companies – if they choose the right strategy – have great chances of business success even outside of their national boundaries.*

Key words: *small companies, business environment, globalization*

1. Introduction

One of the most significant factors for the success of small businesses is the very context, i.e. environment in which they operate. The environment shapes not only the role of small companies but also their strategy and their success. There are several dimensions to the environment of small companies, from local to global, to which attention will be devoted in the part of the work that follows.

The business environment of small companies is a set of factors that secures the survival, functioning and development of these economic subjects. States which invest even minimal efforts in ordering the business environment by

way of certain rules and regulations, securing tax breaks and other kinds of stimuli such as training and consulting services to entrepreneurs, achieve enviable results in this economic sector, which are reflected in larger investments in risky business projects (first-time businesses). In addition, the securing of financial resources, the necessary infrastructure, and the existence of universities open to cooperation with the business sector in the form of offering training and research services will significantly contribute to the expansion and development of the small business sector. An accent is being placed on the environment due to the fact that, as an external factor, it represents the first barrier to the successful start of an individual business and the formation of small companies. If these barriers are minimal, chances grow that small companies will "live to an old age" or develop into larger ones. The more attention is devoted to the business environment, the better the results in the small company sector will be; in addition, individuals will be encouraged to start new businesses if they perceive that attention is being devoted to entrepreneurial initiatives. The environment conditions and determines the success of entrepreneurial activities; in other words, environmental factors can be either a positive or a negative force in the shaping of entrepreneurial intentions, and are classified among external processes beyond the control of individuals. From the aspect of market dimensions, we may divide the business environment into two segments: the global segment and the prevailing conditions within it for the existence of small companies outside of their national markets; and the segment connected to the domestic market and its political-economic conditions.

2. The global environment for small companies

Market expansion and globalization have not, as was expected, moved in line with the expansion of the role of large companies. Newer data on small companies show the important and expansive role played by this type of company in the global economy, which varies by sector depending on the way in which competitive advantage is secured.

Do larger markets favor larger companies? Will new or small companies play a role in the global economy that's certain to come? If we are guided by experiences from the past, the answer would be positive – large markets favor large companies, and negative – the role of small and medium companies (SME) in the future global economy will be limited. However, a significant number of analyses and events suggests not only that constantly developing large markets present an opportunity for small companies but that the dynamic small company sector represents a critical, fundamental base for the growth and development of such markets. More so than the weakened, vulnerable existing sector that demands subsidies, the small company sector has

the capability of not only advancing but of becoming an essential component of greater economic prosperity.

Much has been spoken about the growth of large companies in large markets during the past several centuries. The history of large and successful economies is a story of the advantages of size, which strengthens certain portions of the economic structure through the formation of oligopolies or monopolies. In the majority of cases, such trends have required state intervention in order to ensure competitiveness. In the same way, interventions may be directed to the SME sector with the aim of providing the necessary support for the creation of favorable conditions for the formation and survival of small companies, which are threatened by already existing large companies.

Market-expanding changes can be viewed as two types of changes:[1]

1) Technological advance in communications and transport;
2) The harmonization of the regulation environment in which economic activity is taking place.

The first reduces costs while the second reduces risks in international business. Basic improvement is evidenced in the reduction of costs along with an increase in the reliability and speed of communications and transport. Many of the transaction costs tied to the delivery of products and services to consumers are thus reduced. The harmonization of the regulatory environment results in the standardization of the institutional and regulatory context, which will increase the confidence of the economic actors and ensure that all the sides in business transactions are dependable when it comes to fulfilling contract obligations.

Such thinking is present in a wide spectrum of programs developed within the European Union during the integration of the European market. This large market secured basic advantages for large companies, which brought concern about the survival and sustainability of the small company sector. The same discussions and fears appeared regarding the SME sector in the US and Canada after the onset of NAFTA (North American Free Trade Area). Were these fears justified? It seems that the answer is negative, since the number of workplaces and the total economic activity in this sector remained stable or even increased in some cases. An analysis of employment in the member countries, of profits and sales earnings on the basis of company size showed that small companies had increased their share in all the pertinent measures at the beginning of the 1990s.

The fear of the domination of large companies in the global marketplace is a prominent and well-represented topic, which has resulted in increased attention being paid to the role of small companies in international business.

[1] P. D. Reynolds, "New and Small Firms in Expanding Markets", *Small Business Economics*, 1997, p. 79.

An analysis of the SME sector within international business processes was recently completed by the OECD (Organization for Economic Cooperation and Development), working group for SME. Its attention was focused on 18 OECD member-countries and 8 Asian countries (Indonesia, South Korea, Malaysia, PR China, Singapore, Taiwan, Thailand and Vietnam). The analysis established that in the early 1990s the SME sector participated significantly in international business. About 10% of all SME (mostly in the industrial sector) were involved in direct foreign investments. Within the OECD framework 26% of direct exports went through SME. It was estimated that there were about 35,000 transnational corporations in the world, of which 20,000 had fewer than 500 employees. In the OECD countries only 1% of the world's industrial corporations could be considered global, which comes to 30-40,000 companies. In addition, 10-20% of the manufacturing SME took up 10-40% of export sales or a total of 300-600,000 companies within the OECD countries. Finally, about 40% of all SME in OECD member-countries are considered to be isolated from globalization.

On the basis of the presented data we can say that there is solid evidence that small companies are direct participants in international business, with a tendency of increase.

Strong entrepreneurial leadership often characterizes small companies. However, as small companies grow, there is an increasing need for the introduction of formal structures, systems, procedures and controls. A small company goes through numerous difficulties during its organizational transitions. The transition from a small entrepreneurial company to a large corporation is characterized by numerous internal and external changes. Internal changes in the first place include the professionalization of management in order to create the necessary conditions for governing a growing complex organization. Internal changes such as increasing sales, the number of employees, and the complexity of the organization change the role of the small company owner within the company. The search for growth often compels the owner/manager to learn new skills and acquire new knowledge, which brings changes to his role, transforming it from that of a mere executor to a managerial one. Often, the main limiting factor in the growth of small companies is that managers don't manage to carry out this transformation. During this development process, the manager often becomes alienated from the employees, which results in significant difficulties in the functioning of the company.

3. The process of internationalizing business operations

The time and the environment in which we live and work are, by their nature, highly volatile. Changes occur on a daily basis, and challenges are present in all elements of economic and social life. The continuity of the past has been replaced by today's discontinuity.

The task of company management in such business conditions is to create a strategy of the company's reactions to opportunities and dangers generated by the business environment, in order to achieve the company's basic goals – survival, growth and development. The strategy is connected to the choice of directions, methods and instruments for the achievement of the company's goals, as well as to the establishment of relations between the environment, the competition and the company's own capacities, in order to secure successful business operations – which are dependent on success achieved on the chosen market.

Only flexible companies can hope to achieve business success in a changing environment, i.e. those flexible in organizational structure, business policy, system of management, etc. On the path to accomplishing their goals, companies are faced with a series of limitations, including falling market needs, intense competition, legal limitations, their home country's inadequate economic policy and many others. One of the most important limitations certainly lies in existing markets that are either stagnant or declining, for the following reasons:

1) Advancing technology is in a position to induce constantly the production of increasing quantities of products;
2) In existing markets, the growth rate of the national income has declined, which results in reduced consumer purchasing power.

This means that the existing markets (national) are becoming tight and insufficient and that companies are compelled to seek out new ones. As a result, an increasing number of factors are orienting companies toward foreign markets and the expansion of their business beyond national borders; in other words, processes of business internationalization and globalization are becoming unavoidable. A global marketplace assumes equal business conditions for all participants, offers great development chances, especially when it comes to increasing sales capacities, the securing of new resource markets, and improvements in competitive advantages. The process of internationalizing business operations begins with closer and more secure markets and simpler forms of appearance on the market, moving on to more distant and riskier markets and more complex forms of market appearance.

In the majority of cases, the first appearance on a foreign market is done by way of exports, as the simplest form of internationalization, characterized by minimal investments and low risk levels. Today's issues connected with enter-

ing a foreign market are much more complex. For these reasons, the process of globalization demands a strategic approach, which means that entrance into a foreign market must not be improvised, for such attempts can be qualified in advance as being unsuccessful. Entering a foreign market requires a planned approach, which means that each step to be taken in that direction needs to be well thought out in advance. Companies planning to enter the international market must be able to recognize opportunities, the ways of exploiting competitive advantages, must be able to evaluate their own capabilities – all with the goal of reducing potential risks tied with appearing on a foreign market to the lowest possible level.

Under present-day conditions, all companies who've seen even the smallest opportunity to do so have entered the process of internationalizing business operations. The reason is quite understandable: since business risk is at an extremely high level, expanding operations over several markets disperses risk and reduces the chances of its occurrence. Thus, a company's appearance on foreign markets is one of the key moves contributing to its ultimate business goals; this, of course, is not of equal significance for all companies, nor for all countries, as this depends on their size and strength. For most small countries, their foreign trade relations form the backbone of their economic development. This is confirmed by the fact that in the majority of small countries more than 50% of the total gross domestic product (GDP) is tied to their foreign trade. Thus, for example, 42% of Switzerland's 1999 GDP was tied to its foreign trade.[2] Small countries such as Holland, Belgium, Luxembourg, New Zealand, etc. have to export most of what they produce in order to secure the proper conditions for economic development.

As much as the process of internationalization is significant for national economies, it is no less important for individual companies. In order to achieve its goals, a company must have the ability to take its business outside its national boundaries. It's up to each individual company as to how it will enter a foreign market: through export as the simplest form or by a more complex path, such as co-production, technology transfer, franchising, joint venture or business alliance. The choice depends on the company's goals, available resources and opportunities offered by foreign markets.

The carriers of the globalization process are multinational companies (MNC), which have accumulated such financial power by taking advantage of the advantages offered by this process that they have become stronger than the governments of certain countries. Even more revealing are data showing that multinational companies control one half of the world's manufacturing, two thirds of its trade and three fourths of international technology transfers. The first MNC was formed in 1929, with the merger of the Dutch Margarine Unie Company and the British Leveer Brothers Company, resulting in a single

[2] S. Vasiljev, "Marketing", Economics Faculty, Subotica, 1999, p. 335.

company named Unie lever, which is now one of the world's largest, with 500 branches in about 60 countries.[3] During the initial period of MNC formation, the leading position was held by European companies; during the 1980s, US companies gained primacy, while today the situaion is quite different, as Japanese MNCs have also gained many of the top spots. Most often, companies are classified as MNCs according to the following criteria:

1) That they are principals in at least six different countries;
2) That at least 20% of the company's business comes from doing business in countries outside the home country;
3) Globally-oriented management;
4) The company's resources are distributed regardless of national borders;
5) Their personnel are transferable worldwide.

In recent years, MNC expansion has been taking place in the developing countries, which offer new sales markets, important raw materials, cheap labor and other significant resources. Large barriers to MNC operations are certainly presented by legal regulations regarding ownership share, use of earned profits, as well as by political and economic stability. Since in the course of doing business they come into contact with various social and cultural milieus with differing customs, beliefs and values, it is important that they learn them and adjust to them so as to avoid business failure in specific geographical areas.

4. The global strategy of small companies

Without doubt, the process of globalization is influencing economic activity of any kind. In the 21st century, the Swedish system of manufacturing and distribution has developed according to the same model by which national markets developed out of their local and regional entities in the 19th century. In each developed country international trade and foreign direct investment significantly influence the level of the gross domestic product.

Globalization affects an entire network connecting states, societies and organizations of the existing global economic system. It creates new structures and relations, as a result of which business decisions and moves made in one part of the world have significant bearing in other regions. Globalization trends are magnified by rapid changes in the technological environment, especially in the fields of biotechnology, telecommunications and information technologies. The development of these segments allows the management of research, marketing and production all over the world. The development of markets on the global level stimulates competition and forces national governments to adapt

[3] B. Đorđević, "Menadžment", Priština, 2003, p. F-102.

market-oriented policies. Economic policies that seek to exclude global participation by way of various instruments and measures are in danger of being made obsolete, denying their countries their full share of prosperity.

Over the past few decades, two trends have dominated the international business environment:

- Market integration, with the formation of a global market in many industries;
- The appearance of large international corporations as the main carriers of these processes.

Most of the research dealing with international business has focused on the multinational companies and the ways in which they are managed, without a great deal of interest being devoted to small and medium companies.

The basic concept of international trade is that of comparative advantages, which are based on assumptions that have not changed that much since the times of David Ricardo and Adam Smith. Each country has comparative advantages in some sphere of economic activity. However, it's not the countries that do the trading but the companies! If a country has comparative advantages, then companies have competitive advantages, regardless of whether we are speaking of small, medium or large companies.

Since World War II, international trade and foreign direct investments have been growing. Goods, money, information, labor and services overcome national borders, and companies that previously enjoyed domestic market protection must now face competition. All the above trends contribute to interdependence on the global level: technological advance and diffusion, market integration, lower const of information and transport, and the liberalization of trade barriers. All these phenomena are seen as the advantages of a global marketplace, especially in the case of large companies, at the same time reducing or even bringing into question the role and capabilities of small companies. In accordance with these observations, the following question can be asked: what are the possibilities for small companies in the global marketplace?

The revolution in transport and the movement of information at the start of the 19th century marked the beginning of multinational competition. The revolution in information and communications created the global competition we are familiar with today. The reason for globalization is usually sought in the reduction of fixed costs. As survival and advancement on the global marketplace require an infrastructure and an adequate marketing approach in the marketplace, many see in these opportunities for large companies, leading them to disregard the role and the position of small companies. On the other hand, there are small companies which do not enjoy the protection of special government measures, but are nevertheless present in the global marketplace.

The basis for achieving competitive advantages on the global market certainly lies in the following: favorable climate, wealth in mineral raw materials and cheap labor. These are, in turn, more easily taken advantage of by large companies. However, one of the basic sources of the competitive advantage enjoyed by small companies are innovations with which they can parry the large companies and, thereby, secure a strong position in the global marketplace.

Company ***size*** appears ***as an advantage in securing a favorable competitive position***, for the following reasons:[4]

- they have no resource limitations;
- they control most of the market and can more easily capture new ones;
- they possess greater business experience;
- they have a better negotiating position with buyers and suppliers;
- they can more easily withstand mistakes and failures.

One of the ***strategies*** employed by small companies when it comes to global competition is ***avoidance***. One of the dominant schools of thought when it comes to the small companies' business strategy on the global market is that small companies should avoid direct competition with the giants, and there are several ways of achieving this:[5]

- focusing on specific activities;
- focusing on specific market segments not targeted by large companies;
- taking risks on the basis of specific managerial attributes;
- using the advantages of flexibility and adaptation to changes in the environment;
- entering into associate relations with large companies;
- permanent innovation with the goal of becoming large.

Both on the domestic and the global market advantages are gained through cost reduction – which results in competitive prices. Small companies can be successful in this regard only if they focus on a specific part of the market, small enough not to attract the large companies. The greatest opportunity for small companies lies in producing for a small number of sophisticated buyers who demand quality before all, or in producing for buyers who demand frequent product changes in accordance with their needs and are ready to pay a high price for this privilege. Problems that can appear in connection with the market are either its non-development or overly rapid development – which might attract large competitors. Technologies used by small companies must also be specific, so that they either aren't available to large companies or do not interest them; otherwise, they would be considered as barriers to entry into a specific business segment.

4 T. Agmon, R. Drobnick, "Small Firms in Global Competition", Oxford University Press, New York, 1999, p. 11.

5 Ibid., p. 13.

Large multinational companies truly have advantages which smaller companies cannot match, but there is a tendency of allowing small companies to retain their autonomy, in which case the large companies take the responsibility for the decision-making and for making moves that are in the general interest, while small companies specialize and have an operative role – an associate relationship. If small companies do not wish to become parts of large corporations, they alone must take the risk of such a decision. This risk is greatly reduced if their national governments, due to the specific role played by small companies in the national economy, decide to protect the domestic market from international competition by way of special measures of economic policy. The risk of independent business ventures can be reduced through the linking of small companies into different types of associations, securing them with certain advantages by enabling them to jointly compete on the market.

One possible strategy for securing the place of small companies in the global marketplace is an orientation toward innovation. Even though small companies spend less on research and development than large companies, they nevertheless "produce" almost twice as many innovations per employee. Thus, the conclusion seems to be that small companies are superior when it comes to creating innovations while the large companies are superior in appropriating their return effects, either through purchasing property rights or taking over the companies themselves or the results of their research. The main reason for this phenomenon is the lack of copyright protection within large companies, which means that new inventions, regardless of the accomplishment of the inventors, are appropriated by the organization. This, of course, reduces the interest of the employee in creating and inventing for the needs and benefits of the company. Differently from those employed by large companies, independent inventors are able to completely protect their rights and are not burdened by risk or bureaucracy, all of which makes small companies more efficient innovators.

5. Conclusion

On the basis of the above analysis, we know that small companies must face great difficulties, which become ever larger as the market increasingly integrates (requiring large investments, distribution, service support, capital, management). Many small companies developed within the framework of a static domestic market and did not recognize the importance of market analysis and the accommodation of products to the needs of consumers. These problems can be overcome by a quick market reorientation, since, within a fragmented economy or specialized industries; business experience is not of key significance. In order to achieve international success, it's necessary to make accommodations in the direction of having a small number of transactions, focusing

on a specific market segment or business operation and having great flexibility. Small corporations should develop their exports through risk distribution arrangements, relying either on their own abilities or on government help.

As for the choice of a company's growth strategy – in our case, that of a small company – we are talking about a number of mutually conditioned decisions that are to be made simultaneously in order to secure the most favorable position in an environment influenced by social, economic, technological and market fluctuations. In choosing from a wide array of growth methods, small companies most often decide on external growth, which offers them much greater chances of achieving their development goals in a shorter time period, because small companies are mostly burdened by limiting factors (financial issues, employees, installed capacities, knowledge) that compel them to pursue their business goals with the help of others' resources.

The avoidance of direct competition with the "big boys" is certainly the optimal decision toward achieving business success in the global marketplace.

References

- Agmon, T. – Drobnick, R.: *Small Firms in Global Competition*, Oxford University Press, New York, 1994.
- Drucker, F. P.: *Menadžment za budućnost*, Grmeč, Beograd, 1995.
- Hisrich, R. D.: *Small Business Solution*, McGraw Hill, New York, 2004.
- Kirby, D. A.: *Entrepreneurship*, McGraw Hill, London, 2003
- Lacković, Z.: *Malo poduzeće u uvijetima tranzicije*, Revnost d.o.o., Osijek, 2000.
- Lambing, P. A. – Kuehl, C. R.: *Entrepreneurship*, Prentice Hall, New Jersey, 2003
- Straub, J. –, Attner, R.: *Introduction to business*, International Thompson Publishing, 1994.
- Tan, J.: *Regulatory Environment and Strategic Orientation in Transitional Economy*, Entrepreneurship Theory and Practice, 1996.

Original scientific paper
Received: August 26th, 2007
UDK 336.71(497.11)"2001/2007"

Assistant Professor Predrag Kapor, PhD
Faculty of Business Studies, Megatrend University, Belgrade

BANKING IN SERBIA: THE LONG ROAD TO THE EUROPEAN UNION

Abstract: *Serbia's financial system has traditionally relied on a dominant banking sector. After 2001, foreign banks began to enter the Serbian financial market, initially forming new banks (affiliates) and then buying domestic socially (state) or privately owned banks. As a result, out of a total of 37 banks operating in Serbia at the end of 2006, 22 were foreign-owned, with a 78.7% participation in the total sum balance of the Serbian banking sector.*

The arrival of foreign banks in Serbia, helped along by newly brought regulations harmonized with international practice and standards, has to a large extent returned confidence to the banking system, along with professionalizing the industry, improving the array of banking services and easing capital access to those economic actors able to fulfill the newly set standards of creditworthiness. However, this development of the banking sector was not accompanied by appropriate changes in other segments of the financial system, and even less so in the real economic sector.

The great expansion of credit in Serbia over the last three years, to a great degree founded in the foreign-owned banks' borrowing from their parent banks abroad, but also in direct foreign borrowing on the part of Serbian economic actors, in addition to contributing to the country's overall foreign debt balance, is generating substantial inflationary pressure and contributing to both the foreign trade deficit and the current accounts deficit. At the same time, the linking of the credit obligations of domestic end users to foreign currencies and international market interest rates brings additional risks to the domestic financial system. The Serbian National Bank's occasional restrictive credit-monetary measures produce limited effects, but it is an open question as to how much and how long the Serbian National Bank will be able to successfully control such a financial system, in which the key role is played by the Serbian banking sector, with its dominant foreign component.

Key words: *financial system, banking sector, banks, credit, crediting the general populace, interest rates, foreign debt*

1. Introduction

The financial system has the key role in the functioning of each market economy, since it enables the transfer of money between economic subjects, both within the bounds of the national economy and on the international level. The financial system may be defined as a part of the total economic system, which represents a group of institutions and instruments through which the collection, concentration, transfer and allocation of financial resources is performed.[1] Theory has generally and broadly accepted the division of financial systems into bank-based or bank-oriented financial systems and market-based or market-oriented systems.[2]

The model of the bank-based financial system is characteristic for continental Europe, as opposed to the model of the market-based financial system, which is characteristic for the Anglo-Saxon countries. In the first model, the dominant role in the transfer of financial resources is played by the banks as financial intermediaries, organized, as a rule, as universal banks, while in the second model this role is played by the financial market itself, which must be at a high level of development, with a large number of multifaceted and liquid financial instruments and a broader circle of financial actors, especially those from the ranks of institutional investors, as well as a significant role on the part of investment banks, which are separated from the commercial banking sector.

Over the last several decades, the differences between these two types of financial systems have lessened, which means that one can no longer talk about a 'clean' model of a bank-based or market-based system but, rather, about financial systems that are to a greater or lesser degree dependent on either banks or the financial market.[3] Nevertheless, it seems that the dominant role is gradually being taken by a financial system based on the money and capital markets, while banking institutions are increasingly adjusting to the mechanisms of the financial market. In both transition countries and developing countries, however, banking institutions still play the dominant role in the sphere of financial mediation, since the necessary conditions for a stronger role of the money and capital markets are still in the process of creation.

---

1 P. Trifunović, "Banke – organizacija, poslovanje i politika", "Službeni list SFRJ," Beograd, 1980, pp. 7-8.

2 Asli Demirguc-Kunt, Asli and Ross Levine, *Bank-based and Market-based systems: cross-country comparisons*, World Bank, Washington, University of Minnesota, June 1999, p. 2; John A. Goddard, Philip Molyneux, John O.S. Wilson, *European Banking: Efficiency, Technology and Growth*, John Willey & Sons, Ltd. Baffins Lane, Chichester, England, 2001, p. 302; Lawrence S. Ritter, William L. Silber, Gregory F. Udell, *Principles of Money, Banking & Financial Markets* (10th Edition), Addison-Wesley Longman Inc., Reading, Massachusetts, 1999, p. 263; Jordi Canalis, *Universal Banking*, Claredon Press, Oxford, 1997, p. 40.

3 David T. Liewellyn, "Globalisation and convergence on the shareholder value model in European banking," BIS Paper No. 32, *Financial globalization*, December 2006, pp. 20-29.

This is why it is extremely important for transition countries to have a healthy and efficient banking sector, which is working in the function of their general economic development. It is generally thought that the arrival of foreign banks into transition countries and developing countries brings the following benefits to their financial and banking systems: 1) increased competition and improved efficiency in the providing of financial services; 2) better quality and availability of banking services; 3) contribution to transfers of technology and management skills, especially risk-management skills; 4) improved credit allocation; 5) stimulates the development of the local financial market; 6) opens the way to integration into the global financial market and easier access to capital; 7) improves the 'health' of the domestic financial system and has a stabilizing role in crises; 8) reduces sensitivity to the country's credit cycles.[4] However, the arrival of foreign banks and their domination of the banking market can expose the financial systems of transition countries and developing countries to certain risks.

The aim of this work is to, on the basis of analyzed data, identify the key changes in Serbia's banking sector in the period between 2001 and mid 2007, in the first place in its structure and way of doing business, along with the broader economic influence of these changes – including on Serbia's foreign economic relations – as well as to point to certain problems that have appeared in connection with them.

2. Structural changes in Serbia's banking sector

Serbia's financial system has been traditionally based on a dominant banking sector, which is also true for the present phase of Serbia's transition to a genuine market economy, while the significance of the other parts of the financial sector and forms of financial mediation (insurance, investment and pension funds, brokerage-dealership companies), and of the financial market itself, has been relatively small, due to a lack of institutional order as well as the scant choice of financial instruments.

It is correct to say that the financial infrastructure of Serbia's banking sector at the end of 2000 was institutionally unsatisfactory, due to the small number of 'good' banks, its limited scope of operation (the banks were banks for

[4] BIS, "Foreign direct investment in the financial sector - experiences in Asia, central and eastern Europe and Latin America," Committee on the Global Financial System, SGFS Paper No. 25, June 2005, Basel, pp. 1-2; Dietrich Domanski, "Foreign banks in emerging market countries: changing players, changing issues," *BIS Quarterly Review*, December 2005, Basel, pp. 69-81; IMF "Global Financial Stability Report," September 2006, Washington D.C., p. 46; BIS, Committee on the Global Financial System, "Foreign direct investment in the financial sector of emerging market economies," March 2004, Basel, pp. 10-15.

their shareholders only), its structural aspects (only domestic banks existed). In other words, Serbia's banking sector was generally disorganized, inefficient and expensive.[5]

Foreign banks began to enter Serbia in 2001, initially forming their own affiliates, which were primarily oriented toward private individuals and small and medium companies, as well as toward foreign payment operations. As part of the program of rehabilitating Serbia's banking sector, the number of banks was drastically reduced – from 91 at the end of 2000 to only 48 in 2002.[6] At the beginning of 2002, the four previously largest banks (socially-state owned) were liquidated, the official evaluation being that their rehabilitation would be several times more costly than their liquidaton.

Foreign banks continued their further expansion on the territory of Serbia, now basing their entrance strategy (conditioned on receiving permission from the National Bank of Serbia) on their acquisition of domestic banks. There were also takeovers, initiated within Serbia itself (HVB takeover of Eksimbanka, EFG Eurobank takeover of Nacionalna štedionica, NBG takeover of Vojvođanska banka), and concentrations on that basis (OTP bank merged together Kulska banka, Zepter banka and Niška banka), as well as external takeovers (with UniCredito takeover of HVB). As a result, at the end of 2006, Serbia's banking sector was made up of 37 banks (with 2,158 organizational units), with 22 being foreign-owned (i.e. with majority or absolute foreign ownership in most cases; these will be treated in this work as foreign-owned banks), with a 78.7% share in the total balance of the entire Serbian banking sector, while 7 banks were majority-owned by domestic private individuals and companies, with a 6.5% share in the total balance, and 8 were under majority ownership of the Republic of Serbia, with a 14.8% share in the total balance.[7] A more detailed analysis of the ownership structure of the remaining banks under majority domestic ownership (private as well as state) would show that a respectable percentage of their owners (shareholders) are foreign companies (foreign investment funds and foreign banks), but that these investments have the character of portfolio investments. What is also interesting is that, as a rule, when foreign banks acquire domestic banks, once they acquire a majority share they go on to acquire the remaining shares and form closed ownership structures, which means that only the shares of banks with a majority domestic ownership are traded on the stock market.

The total balance of Serbia's banking sector has grown considerably in the last few years, totaling 1,169,271 million dinars by the end of 2006, or almost

5 Ž. Ristić, D. Vuković, "Bankarski sistem Srbije," *Jugoslovensko bankarstvo*, no. 3-4/2002, pp. 3-19.

6 Ibidem.

7 NBS: "Bankarski sektor u Srbiji – Izveštaj za IV kvartal 2006. godine," april 2006, p. 4.

14.8 billion euros according to the official exchange rate at the end of 2006.[8] This is a significant amount for Serbia, relative to its economic potential measured by gross domestic product, but relatively insignificant on a global banking scale, having in mind that, for example, Deutsche Bank alone reported a total balance of 1,126.2 billion euros at the end of 2006.[9]

Following the sales of the state's share in the remaining banks, and the acquisition of the controlling share packages of several banks with a domestic majority ownership (private individuals and companies), it is to be expected that foreign-owned banks will own more than 90% of Serbia's banking sector, as measured by its total balance. This is not unusual for a transition country in Eastern and Central Europe, as transition countries throughout the region are faced with a similar phenomenon. Thus, the foreign-owned banks' share in the banking sector's asset balance (according to data from 2005) is the following for the countries in the region: Hungary – 94.5%, Czech Republic – 94.5%, Slovakia – 97.3%, Bulgaria – 80%, Albania – 100%, Bosnia and Herzegovina – 91%, Croatia – 91.3%.[10] What should be kept in mind is that, in the majority of the said countries, this 'ownership transformation' of the banking sector took 10-15 years, while in Serbia everything was accomplished in the space of only five years. It should be pointed out that, in the most developed countries, with a few exceptions, the share of foreign-owned banks in the total balance of the domestic banking sector is mostly under 15% (i.e.: Germany – 5.4%, France – 13.9%, Italy – 5.1%, Sweden – 1.7%, Japan – 2.1%, USA – 8.5%).[11]

The arrival of foreign banks to Serbia, in tandem with the passage of new regulations in accordance with international practice and standards, has undoubtedly, to a large measure, restored confidence in the banking system, professionalized the sector, improved the array and quality of banking products and eased access to capital to economic actors who fulfill the required standards of creditworthiness. However, the said development of the banking sector has not been accompanied by the necessary changes in other segments of the financial system and the real economic sector. This has resulted in structural problems that could prove to be limiting factors for the further development of the financial system, even of the economy as a whole. The banks' business policy, which is mostly dictated by the same foreign banks throughout the Eastern and Central European region, is contributing to the creation of macroeconomic risks, which could play out very negatively in the coming period.

8 Ibidem.

9 www.deutche bank.de

10 Raiffeisen Zentralbank Oestereich AG: "CEE Banking Sector Report," September 2006, Vienna (various pages).

11 Bayraktar Nihal, Wang Yan, "Bank Sector Oppenness and Economic Growth," World Bank Policy Research Working Paper 4019 (WPS4019), October 2006, p. 27.

The sale of state and privately owned domestic banks in Serbia to foreign banks has brought an inflow of foreign capital in the form of foreign direct investments. Out of the total net inflow of foreign direct investments in the 2001-2006 period, which, according to the data of the National Bank of Serbia, totaled about 9 billion US dollars, about 2.5 billion US dollars were invested in the banking sector.[12] Foreign banks acquired domestic banks at prices that were mostly higher than book value, but they were, in fact, de facto buying the market, and it is quite probable that, due to the high interest margins in Serbia, their investments will be returned relatively quickly.

For comparison's sake, in December 2005, Austria's Erste Bank paid 3.75 billion euros for 61.88% of the state's capital in the Commercial Bank of Romania (BCR), whose share in the total balance of Romania's banking sector was 26.3% (2005).[13] This sum was greater than the total capital of Serbia's banking sector at the end of 2006 (2.7 billion euros) and substantially greater than the total capital of all the foreign-owned banks in Serbia (1.9 billion euros). Thus, it can be said that the inflow of foreign capital into Serbia in the form of foreign direct investments or privatization earnings was not all that great, at least in comparison to some other countries.[14] On the contrary, the impression is that foreign banks did not pay a high price for almost four-fifths of Serbia's banking market, in which net profits on the basis of interest, compensations and commissions for 2006 alone equaled about 900 million euros.[15] Incidentally, in 2006, foreign-owned banks had a lower share in the total capital of the entire banking sector (68.1%) than in its total balance (78.8%), which means that domestically-owned banks were more capitalized in relation to their market share.

3. The role of banks in the growth of Serbia's foreign debt

Banking operations are traditionally based on the collection of freed up monetary resources from business and individuals and their investment into finance-lacking economic subjects. However, if domestic savings and the deposit base are insufficient for financing credit expansion, banks are forced to mobilize resources from other sources, including external borrowing. The significant increase in Serbia's foreign debt (despite the debt write-offs granted by creditors within the Paris and London Clubs) is a problem that may greatly influence its further economic development. However, this time, the generators of the new foreign borrowing are banks and companies, not the state, as was the case previously.

12 NBS: "Statistički pregled," mart 2007, p. 51.

13 Raiffeisen Zentralbank Oestereich AG: "CEE Banking Sector Repor," September 2006, Vienna, pp. 38-39.

14 NBS: "Godišnji izveštaj 2006," p. 128.

15 NBS: "Bankarski sektor u Srbiji – Izveštaj za IV kvartal 2006," april 2007, p. 12.

Under the conditions of relatively modest domestic savings, the deposit base was not sufficient to finance the credit expansion. As a result, in order to achieve the ambitious goals of their founders and take advantage of the existing market possibilities for profitable investments, the foreign-owned banks in Serbia began to increasingly borrow abroad, either from their founding banks or within their business groups. According to NBS data from mid 2006, an average of 64% of total credits in the Serbian banking sector were covered by sources of foreign origin, with the average being 87% for foreign-owned banks and only 7% for domestically-owned banks.[16] At the same time, direct foreign borrowing by banks in Serbia, which, according to NBS data, equaled 746.9 million US dollars at the end of 2000, grew to 5,099 million US dollars by the end of 2006, and then slightly fell by the end of June 2007 to 4,850.8 million US dollars.[17]

Table 1: *Serbia's foreign debt according to debtor types (at the end of 2000 and 2006, and in June 2007, in millions USD)*

	2000	2006	June 2007
Public sector	8,797.3	8,456.8	8,406.4
Private sector	2,032.4	11,148.9	12,815.1
of which:			
banks	746.9	5,099.0	4,850.8
companies	1,285.5	6,049.9	7,964.3
Total:	10,829.7	19,605.7	21,092.0

Source: NBS: *The Republic of Serbia: external debt according to type of debtor*; www.nbs.yu

At the same time, the foreign debt of companies in Serbia has increased from 1,285.5 million US dollars at the end of 2000 to 6,049.9 million USD at the end of 2006, only to grow even further, reaching 7,964.3 million USD by the end of June 2007. The large growth in direct foreign borrowing by business has been especially noticeable since 2005, as a way of getting around the domestic banking system in order to avoid the high costs resulting from restrictive measures set by the monetary policy. Thus, the domestic economy is rapidly increasing its direct foreign borrowing, most often through foreign-owned banks, which direct their clients in Serbia toward direct borrowing from their parent banks abroad, in order to avoid the limitations and costs of borrowing from foreign sources faced by banks in Serbia.

This trend will probably continue in 2007, as there are more and more categories of clients getting access to such, for them, relatively cheap credits. In the beginning, these were only large, reputable companies, while by 2006

[16] NBS: "Ekonomski pregled," jul 2006, p. 22.

[17] NBS: "Republika Srbija: stanje spoljnog duga po vrsti dužnika"; www.nbs.yu

direct foreign credits had become accessible to small and medium firms, and even entrepreneurs.

In this context, it should be born in mind that there has also been substantial growth in foreign borrowing on the part of leasing companies. At the end of 2006, within the liabilities in the balance sheets of financial leasing companies, long-term foreign-originated credits amounted to 83.6%, or 56.6 billion dinars.[18] This means that these companies' foreign debt came to more than 900 million US dollars. This is indicative, since the development of the financial leasing sector in Serbia began, along with the passage of the Law on Financial Leasing of 2003, mostly with the support of banks (domestic with foreign capital and foreign banks) whose affiliates were the first financial leasing companies that were founded. Of the 15 current financial leasing providers, 14 are owned by foreign legal entities or domestic majority foreign-owned legal entities, primarily banks, while only one is wholly domestically-owned (in 2006, it had a 4.4% share in the total balance of financial leasing companies).[19]

4. The credit expansion of banks in Serbia: causes and effects

The last several years in Serbia have seen a large credit expansion, mainly driven by foreign-owned banks, which is a reflection of their dominant position. This has also brought substantial changes in the structure of the borrowers, i.e. the 'client target group,' since foreign-owned banks are increasingly pushing business with private individuals and small business, i.e. retail banking, which is characteristic of the business strategy of foreign-owned banks pretty much in the entire region of Central, Eastern and Southeastern Europe, as well as quite normal, as all these banks, with several exceptions, are parts of the business operations of the same banks from several Western European countries.

The best indicator of this strategy, along with the credit card business and various commission-based operations, are credit placements. In the period since 2003, bank receivables on the basis of approved credits to the retail sector have practically doubled each year. While the structure of credit placements is still dominated by legal entities (companies), it is evident that their relative share in total credit placements by banks has been falling significantly year by year, which can be seen in Table 2.

The basic reasons for such a policy on the part of the banks in the course of their credit expansion lie primarily in the opportunity to gain high profits through interest spreads, as well on the basis of various commissions tied to credit approval and servicing, at an acceptable risk level. As an illustration, the average yearly interest rate on newly-approved short-term (up to 1 year) Euro-

[18] NBS: "Nadzor finansijskog lizinga – Izveštaj za IV tromesečje 2006," pp. 3-4.

[19] NBS: "Nadzor finansijskog lizinga – Izveštaj za IV tromesečje 2006," p. 1.

zone consumer credits in December 2006 was 7.56%, while the average weighted active interest rates of banks in Serbia for short-term retail credits was 29.50% per year, and 14.54% per year for consumer credits.[20] In June 2007, the average interest rate for newly approved short-term consumer credits in the Eurozone was 8.08% per year, while the average weighted interest rate for short-term retail credits was 30.31% per year, and 16.74% per year for consumer credits.[21]

Table 2: *Accounts receivable of commercial banks in Serbia from selected sectors, 2000 – May 2007**

Year/month	Public companies	Other companies	General public
millions of dinars			
2000	32,499	180,554	2,849
2001	36,987	209,900	5,277
2002	7,335	144,291	16,139
2003	9,890	185,552	29,333
2004	15,948	248,229	66,356
2005	19,171	351,220	131,860
2006	16,866	363,529	203,318
June 2007	18,662	433,476	254,803

* Note: state at the end of the period (year/month)

Source: NBS: *Statistical Bulletin*, June 2007, p. 23.

On the other hand, the interest rates for long-term individual housing loans in Serbia are much closer to the interests for housing loans in the Eurozone; thus, the average weighted interest rate for new housing credits in Serbia was 7.03% per year in June 2007, while in the Eurozone it was 4.79% (for new housing credits over more than 10 years).[22]

However, these comparative indicators do not give a complete picture. Namely, a good part of retail credits in Serbia, the so-called non-earmarked cash credits, are approved for periods much longer than one year (even up to 10 years), which means that interest rates for short-term retail credits are not a sufficiently representative indicator of the price of credit for this debtor category. The same is true for the interest rate on the so-called consumer credits, when it is compared to the consumer credit category in the Eurozone. Another reason why it is not representative is that, according to the methodology of processing statistical data on the part of NBS 'the table of the active weighted interest rate charged by banks shows the average weighted interest rates on the

20 ECB: "Euro Area MFI Interest Rate Statistics: June 2007," 7 August 2007.

21 Ibidem

22 NBS: "Statistički bilten," jun 2007, p. 44-46, and ECB: "Euro Area MFI Interest Rate Statistics: June 2007," 7 August 2007.

annual level, according to which banks agreed on the interest on credits put into circulation for a given month.'[23] Thus, we are dealing with agreed upon interest rates, not with the real price of credit expressed by way of the effective interest rate in accordance with NBS regulations, which would, in that case, be significantly higher.[24] Thus, according to NBS data from April 2006, effective interest rates on standard consumer credits, as well as on cash credits issued by banks that record their cash credits as consumer credits, ranged between 10% and 28.9% for Euro-indexed credits, and between 17% and as high as 47.06% for non-indexed credits.[25]

To be sure, the real price of credit and, with it, the effective interest rate that truly expresses the debtor's full indebtedness, cannot be calculated ex ante but only ex post, i.e. – when the entire credit is paid off, complete with expenses, including those tied to exchange rate margins (which debtors in the Eurozone do not have), in the case of credits indexed with a currency clause – which are the most numerous among credits approved to the retail sector. This also applies to long-term credits, such as housing credits, which have an additional cost factor – the agreed reference international market interest rate (EURIBOR as a rule, but also LIBOR on CHF of late), plus a fixed spread (usually 3% to 4%), which means that, under conditions of a rise in interest rates on the international capital markets, which has been the case during the last two years, the costs of servicing these credits also grow, and their total costs can be calculated only when all the obligations are paid out, after 10-30 years.[26]

As an illustration, in its statistical monitoring of interest rates in the Eurozone, the European Central Bank (ECB) gives the following figure: 'annual percentage rate of charge' – APRC, which covers the total cost of credits and includes the interest rate component, along with other related costs such as costs of credit application processing, credit administration, document preparation, guarantees, etc.).[27] Calculated in this way, total average annual costs for consumer credits in the Eurozone in December 2006 were 7.72%, and for housing loans 4.80%, while in June 2007 these were 8.25% for consumer credits and 5.14% for housing loans.[28]

---

23 NBS: "Statistički bilten," maj 2007, p. 84.

24 Uputstvo o primeni odluke o jedinstvenom načinu obračuna i iskazivanja efektivne kamatne stope na kredite i depozite, "Sl. glasnik RS" br. 11/05, 28/05 and 108/05.

25 NBS: "Ekonomski pregled," april 2006, p. 23.

26 In the case of these, mostly long-term housing credits, the effective interest rate ranges between 7.93% and 11.68% in euros and between 4.75% and 6.51% for CHF, with terms of payment from 60 to 360 months, and security instruments such as promissory notes, mortgages, real estate insurance and life insurance of the credit customer. NBS: "Ekonomski pregled," januar 2006, p. 21.

27 ECB: "Euro Area MFI Interest Rate Statistics: June 2007," 7. August 2007.

28 Ibid.

In any case, interest rates for bank credits in Serbia for the entire period since 2001 have been at a relatively high nominal level and, with the reduced inflation over the last two years, at a high real level (see Table 3), substantially higher than that in the countries of the Eurozone, and even in neighboring countries, although it must be emphasized once again that they, due to the data-processing methodology, do not express the full price of credit paid by borrowers in Serbia.

Table 3: *Weighted active and passive interest rates of commercial banks in Serbia, 2000 – June 2007 (in %)**

Year/month	Active weighted interest rate on short-term retail credits	Consumer credits	Passive weighted interest rate on total individual deposits
2000	29.89	58.17	5.69
2001	23.23	30.40	6.42
2002	23.80	17.79	3.83
2003	24.66	20.27	2.14
2004	23.88	27.37	2.03
2005	25.24	19.71	1.54
2006	29.50	14.54	1.90
June 2007	30.31	16.74	2.34

* Note: state at the end of the period (year/month)

Source: NBS, "Statistical Bulletin", June 2007, pp. 44-47.

The question is, is it realistic to expect that the interest rate for retail credits, as well as for other credit categories, will fall and that they will at least approach those offered in neighboring countries? This has been announced for years, but has never occurred. It is not very probable that this will take place as long as the current ratio of requested and approved credits is maintained, since total potential demand for dinar credits in June 2007 was 36.4% larger than the total sum of approved dinar credits, and as long as banks have no major loan collection problems, as the average percentage of collected loans in June 2007 was 91.8%, with the rate being 93.8% for private individuals and 89.8% for businesses.[29]

The credit boom in Serbia, stimulated by the banks' battle to conquer the largest possible chunks of the market, which was the main spark for the development of banking and the entire financial sector over the last several years, has had a destabilizing effect on the macroeconomic plane, stimulating aggregate demand and, thus, creating undesirable inflationary pressures. The monetary authorities, especially in the second half of 2006, when fiscal policy was

[29] NBS: "Anketa o poslovnim aktivnostima i namerama banaka" (performed in mid June 2007), www.nbs.yu

loosened, which raised aggregate demand, have targeted bank credits in Serbia. Their growth was brought under control near the end of 2006, precisely thanks to the restrictive monetary policy of the National Bank of Serbia. However, the first half of 2007 has seen a continued high growth of credits (with 116 billion dinars in approved credits, compared to 107 billion dinars for the entire 2006), especially cash credits, which rose to 95 billion dinars by the end of June 2007, which made new restrictive measures on the part of NBS inevitable.[30]

As a result, in August 2007 NBS mandated that banks must adjust their gross credit placements to the public by the end of 2007, so that they will equal 150% or less of the value of their capital assets (the previous limit was 200%), and that non-earmarked cash credits can be approved in the future with a maximum repayment term of up to two years (they had been approved for terms even exceeding 10 years, which was a way to get around the limit by which credit indebtedness cannot exceed 30% of the debtor's income).[31]

Out of the total debt taken on by Serbian citizens, cash credits, consumer credits, credit cards and negative running account balances take up about 75%, and housing credits about 25%. This corresponds with the observed trend in other transition countries in Eastern and Central Europe, where various consumer credits make up about 70% of total approved credits to the public sector, while housing credits, mortgages and other similarly earmarked credits make up about 30% of the total, which is opposite of the structure of credit indebtedness in developed countries.[32]

In the sector of Serbia's economic-financial foreign relations, banks (as well as leasing companies) have a significant influence on the creation of creditworthy demand for imports through the approval of credits, especially to the public sector, which are to a large degree used for buying foreign durable consumer goods, or for financing the acquisition of leasing items from abroad. This is in accordance with the dominant business strategy of banks in Serbia, based on the expansion of retail banking.

There are no exact data as to the degree in which the banks' credit expansion and direct foreign borrowing on the part of business have contributed to the increase in imports, the foreign trade deficit and the deficit in the balance of current accounts of Serbia. The fact that the total foreign trade deficit for

30 NBS: Press conference of the Governor of the NBS, August 17, 2007. At the same time, according to data of the Bank Association of Serbia, Serbian citizens are mostly taking out cash credits, whose total amount at the end of May 2007 reached 115 billion dinars (with a monthly growth of 15% in the first 5 months of 2007), while consumer credits amounted to 51 billion dinars and housing credits 66 billion dinars ("Firme i građani u minusu 721 milijardu", *Pregled*, 6.6.2007).

31 NBS: Press conference of the Governor of the NBS, August 17, 2007. www.nbs.yu; Odluka o usklađivanju bruto plasmana odobrenih stanovništvu sa osnovnim kapitalom banaka, "Sl. glasnik RS" br. 39/2006, 49/2006, 64/2006, 31/2007 and 77/2007.

32 NBS: "Statistički bilten," jun 2007, pp. 52-53.

the 2001-2006 period equaled 31.7 billion US dollars, while the deficit in the balance of current accounts equaled 11.7 billion US dollars should be reason enough to force a detailed analysis of all the causes of such trends.[33] This especially because the foreign trade deficit for the first 6 months of this year equaled 4.2 billion US dollars (which represents a 33.4% growth compared to the same period in the previous year), which makes this problem even more pressing.[34]

However, it is well known that excessive credit expansion carries numerous potential risks for macroeconomic stability, which has already been observed in the cae of other transiton countries. Thus, the IMF's economic analysts point out that rapid credit expansions in Bulgaria and Romania are a signficant factor of their worsening foreign trade deficits (by increasing creditworthy import demand), where it has been determined that each credit increase in the amount of 1% of GDP resulted in an increase in the deficit of goods and (non-factor) services (with a 3-month lag) of 0.4% of GDP in Bulgaria and 0.7% of GDP in Romania.[35]

The other side of the banks' credit expansion is the enormous growth of companies' 'internal' indebtedness, and especially of the sector of the general public, vis-à-vis the banks, all of which could prove to be a limiting factor of development at some future point. This phenomenon is either not receiving sufficient attention, or is being relativized by way of comparisons to the state of affairs in other transition countries, without getting into the crux of the problem.[36]

The banks' policy of indexing credits by way of currency clauses, or of tying interest rates on long-term credits to reference international market rates, exposes domestic debtors to the risk of fluctuating currency exchange rates and foreign interest rates, while little attention is being paid to the broader possible negative consequences of this exposure to risks of international character. There is a real danger that certain over-indebted segments of the population and the economy, whose earnings are exclusively in dinars, will not be able to service their current obligations should these additionally increase by a significant amount, all of which can bring broader economic as well as social consequences.

The potential magnitude of these risks, which are practically systemic, is testified by the fact that, out of the total dinar credit activity between the banks and the general population in June 2007, indexed dinar credits made up 84.1%,

33 Data of the Republic Bureau of Statistics: www.statserb.sr.gov.yu

34 IMF: "Global Financial Stability Report," September 2006, Washington D.C., p. 53.

35 Duenwald Cristoph, Guerguiev Nikolay, Schaechter Andrea, "Too Much of a Good Thing? Credit Booms in Transition Economies: The Cases of Bulgaria, Romania and Ukraine," IMF Working Paper WP/05/128, June 2005. Washington DC, p. 20.

36 According to the statement of the General Secretary of the Bank Association of Serbia, Dr. Veroljub Dugalić, average per capita debt in Serbia rose in the first 5 months of 2007 by about 19%, reaching 31,455 dinars or 380 euros. He stated that individual debt is not high when compared to neighboring countries, pointing out that the average per capita debt in Croatia is about 3,000 euros, while the average monthly wage is about 600 euros ("Firme i građani u minusu 721 milijardu," *Pregled*, 6.6.2007).

of which 96.4% were credits indexed via a foreign currency clause, while the rest were indexed to the inflation rate, as well as by the fact that 76.5% of business credits were indexed, of which 83.7% were indexed via a foreign currency clause.[37] In addition, as many as 92% of cash credits are, in fact, long-term loans, with 97% of them having been approved with a foreign currency clause, which means that 600,000 people in Serbia, or one out of three employees (this is the number of people who've taken these credits), are also exposed to foreign currency risk, as the great majority of them have exclusively dinar-based earnings.[38]

Commercial banks have especially intensified their credit expansion during 2007, as part of their battle with the competition and race for profits. This is quite normal business behavior in a situation in which interest spreads in Serbia are substantially higher than those in the neighborhood and offer the opportunity for extra-profit taking. In addition, as foreign investors, they are also able to take advantage of tax breaks, which are an additional source of earnings.[39] Also indicative is the fact that more and more credits to refinance previous credits are being offered, which helps continue the dangerous debt spiral.

Although, as a result of competition, interest spreads will probably tend to decrease over the long term, their current level is the result of certain other factors, which may prove decisive in determining the prices of credits in the short and medium term. These are, in the first place, interest rates on the European money market (EURIBOR and LIBOR), especially in the wake of the great US mortgage market crisis, the risk premium for investing on the Serbian market, interest rates on deposits and measures taken by the domestic monetary authorities. European interest rates, which the foreign banks use as the basis for their own interest rates (due to the foreign borrowing), had an upward trend during the previous year, while, in 2007 and 2008, it is expected that global market interest rates will continue to grow, eventually stabilizing at a significantly higher level relative to that of 2002-2003.[40] For comparison's sake, the six month EURIBOR stood at an annual average of 2.33% in 2005, while in June 2007, it had grown to 4.28% annually.[41]

When it comes to the country risk premium (which normally ranges between 1.5%-2%), which is incorporated into the costs of the banks' foreign

37 NBS: Survey on the Banks' Business Activities and Intentions (performed in mid July 2007); www.nbs.yu

38 NBS: Press conference of the Governor of the NBS, August 17, 2007; www.nbs.yu

39 According to the Law on the Corporate Income Tax ("Sl. glasnik RS" br. 25/01, 80/02, 135/04, 62/06, 65/06), Article 50: "The taxpayer who invests in his capital assets, or in whose capital assets a second party invests more than 600 million dinars, who uses these means within a registered business activity in the Republic and who, in the investment period, additionally employs at least 100 persons for an indefinite time period, is freed from paying corporate income tax for a period of 10 years, in proportion to the said investment."

40 IMF: "World Economic Outlook," April 2007, Washington D.C., p. 270.

41 ECB: "Statistical Pocket Book," June 2007, p. 32.

borrowing and which is also 'built into' the interests approved to credit customers within the country, it is certain that it will not fall in the foreseeable future, having in mind the way in which the main creditors continue to assess non-commercial risk in Serbia.[42]

The costs that arose in the wake of NBS measures in August 2007 (the obligation to reduce, by the end of 2007, cash credits to 150% of each bank's capital assets), will probably be, as in earlier instances, in the greatest part 'passed on' to the citizenry in various ways, which means that there is little probability, under conditions of continued high credit demand, that interest rates will fall.

5. The 'bank-centricity' of Serbia's financial system

As a result of its development during the past 5-6 years, Serbia's existing 'bank-centric' financial system has been significantly strengthened, being dominantly based on banks, with a still relatively weakly developed financial market and a modest offering of quality and liquid financial instruments, where institutional investors and other market financial middlemen are still on the margins, or only in initial phases of positioning.[43] The result: at the end of 2006, in the total balance of Serbia's financial sector, the banks' share was 90.4%, that of leasing companies was 5.3%, while the share of insurance companies was 4.3% (voluntary pension funds and investment funds are not included here, because they were only beginning to operate).[44]

A second important characteristic is that foreign-owned banks, leasing companies and insurance companies already have a more than 75% share in the total balance of Serbia's financial sector, and that this share, in the absence of significant changes in the conditions for foreign capital investments, before all in the growth of the country risk, will continue to grow with time. New institutional investors in the form of pension funds and investment funds, are mostly founded by foreign financial institutions (banks, insurance companies, companies that manage pension and investment funds), or domestic financial institutions with majority foreign ownership, which means that a similar situation can be expected in this segment as well. This also means that most of the key strategic decisions tied to the operations of these financial institutions will be made outside of Serbia, in the context of global risk assessments and the earnings of large foreign financial groups to which these entities belong.

---

42 "Country Risk Classification of the Participants to the Arrangement on Officially Supported Export Credits," valid as of 03 April 2007, OECD, Paris.

43 Milko Štimac, "Bankocentrična tranzicija," *Ekonomist magazin*, br. 345/346, 1.1.2007, pp. 27-29.

44 NBS: "Sektor osiguranja u Srbiji – Izveštaj za 2006. g.," p. 4.

This phenomenon, which exposes a country's financial system to great and unpredictable influence of external factors, has already been observed in certain transition countries and developing countries.[45] The problem of the "vulnerability" of national financial systems with large foreign-owned components demands special attention, a constant analysis of all the potential economic and even political repercussions, as well as strengthened direct monitoring of potential risks, as part of a timely conception of a set of preventive and corrective mechanisms.

6. Conclusion

A national economy's integration into international financial flows brings numerous advantages but also limits the capacity of national governments to use fiscal, financial and monetary policy in order to influence the domestic economy's performance.[46] This is the reality being faced by the creators of Serbia's economic policy today. It can be asked whether, more or less consciously and willingly, the best model for the financial integration of Serbia's economy and development of its financial sector, and especially the banking sector, has been chosen. However, this is of lesser importance at the moment, since the existing state and structure of the banking sector – and, tomorrow, Serbia's entire financial sector, which by the looks of it is developing according to the model of the banking sector – is presently a given, which determines the maneuvering space and possible directions of activity.

Other European transition countries have also encountered, and still are still encountering, this same problem. Thus, the simplest solution may be to study the measures taken by these countries and the effects they produced, if nothing else – in order to at least avoid making the same mistakes. Something should be done as soon as possible, and this is not a task only for the National Bank of Serbia, which has been left pretty much isolated in its activities, but for all the other relevant organs and institutions of the state.

In any case, the possibilities for using foreign credits and deposits for financing the banks' credit expansion should become even more restrictively limited by way of an appropriate level of required reserves or in some other way. In addition, measures of controlling direct foreign borrowing by companies should also be taken as soon as possible. Even though this is a matter of so-called private debts, these nevertheless affect the level of the foreign debt and are a potential source of danger for internal financial stability should a change in exchange rates and interest rates worsen the conditions for the servicing of these debts.

45 IMF: "Global Financial Stability Report," April 2006, Washington D.C., p. 120.

46 Kenneth Kletzer, "International Financial Integration, Sovereignty, and Constraints on Macroeconomic Policy," IMF Working Paper 06/79, March 2006, p. 3.

Although measures such as limiting bank placements and the indebtedness of businesses and private individuals, or limiting interest rates are unpopular, it should not be forgotten that many countries, even the most developed ones, have resorted to them when they judged that there was a danger to the domestic financial system and its stability. What should be reexamined is whether the indexation of credits in foreign currencies, which is allowed by the Law on Foreign Exchange, is a good solution, both from the aspect of confidence in the domestic currency and of dangers faced by those who, in taking on such debts, take on great risks, which cannot be anticipated when they initially sign their credit agreements, and cannot be controlled thereafter, since they are totally dependent on external factors.

The IMF and the Bank of International Settlements (BIS), the two institutions within the Global Financial Stability Forum project most competent to monitor and analyze phenomena within the international financial system, are increasingly calling attention to the potential dangers to transition countries and developing countries caused by excessive growth in retail crediting, the approval of foreign currency credits or the tying of credits to a foreign currency, and the approval of credits according to variable interest rates tied to international market interest rates – all this in a relatively weak macroeconomic environment lacking adequate regulations and a legal and institutional infrastructure.[47]

Further improvements in prudential regulations, as well as in NBS cooperation with bank oversight bodies in other countries, especially those in which the banks with majority ownership in banks in Serbia are based – are imperative. Tardiness in bringing a new Law on Banks (passed at the end of 2005, came into effect only in the second half of 2006) and efforts to order Serbia's banking sector through numerous amendments of the Law on Banks and Other Financial Organizations (of 1993) and of the by-laws that stemmed from it – were certainly not the best of moves.[48] On top of this, EBRD experts concluded in 2005 that, with its banking regulations, Serbia was among those transition countries least harmonized with the Basic Principles of Effective Supervision of the Basel Committee of 1997.[49]

In any case, foreign banks are no strangers to cartelization even 'at home,' much less in less regulated environments such as the Serbian one, where they

[47] IMF: Global Financial Stability Report, September 2006, Washington D.C., pp. 46-70; IMF: Global Financial Stability Report, April 2007, Washington D.C., p. 49; Ramon Moreno, Agustin Villar, "The increased role of foreign banks in emerging markets," BIS, BIS Papers No. 23, 2005, Basel, pp. 9-16; BIS: "Foreign direct investment in the financial sector - experiences in Asia, central and eastern Europe and Latin America," CGFS Paper No. 25, June 2005; M. Arena, C. Reinhart, F. Vasquez, "The Lending Channel in Emerging Markets: Are Foreign Banks Different?," IMF Working Paper No. 47, February 2007, Washington D.C. p. 4.

[48] Zakon o deviznom poslovanju, "Sl. glasnik RS" br. 62/2006 and Zakon o deviznom poslovanju, "Sl. list SRJ" br. 23/02 and 34/02.

[49] Gian Piero Cigna, "Law in transition online 2005," EBRD, October 2005, London, p. 6.

feel much more at liberty in taking advantage of current market conditions.[50] Optimistic claims to the effect that 'Serbian banking has already entered the EU' are true from the aspect of bank ownership in Serbia.[51] However, when it comes to bank operations in Serbia and the conditions being offered to domestic customers, there is still much that remains to be done in order to achieve the level of the European Union.[52]

References

- Arena M. – Reinhart, C. – Vasquez, F.: "The Lending Chanel in Emerging Markets: Are Foreign Banks Different?", IMF Working Paper No. 47, February 2007, Washington D.C.
- Bayraktar, N. – Yan, W.: "Bank Sector Oppenness and Economic Growth", World Bank Policy Research Working Paper 4019 (WPS4019), October 2006.
- BIS, Committee on the Global Financial System: "Foreign direct investment in the financial sector of emerging market economies", March 2004, Basel.
- BIS: "Financial globalisation", BIS Paper No. 32, December 2006, Basel.

[50] Eight Austrian banks that cover the entire territory of that country were penalized at the beginning of 2006 in the amount of 124.26 million euros, when the European Commission discovered that the banks regularly met in order to fix the level of deposits, credit interest rates and other interest rates to the detriment of private citizens and firms who used their services. Tanja Jakobi, "Ne tako liberalni," B&F, 16.2.2006, pp. 43-45); "Stranci" nemaju razlog da smanje kamate," *Pregled*, 9.5.2007; Dr Đorđe Đukić, "Opasna spirala zaduživanja," *Finansije Top 2006*, special issue of the magazine *Biznis & Finansije*, jun 2007, pp. 22-25.

[51] "During 2006, through the ownership transformation of the banking sector, either through the takeover or the privatization of state-owned banks, the process of privatizing majority state-owned banks has been completed. The NBS insisted on the application of strict, "fit and proper" criteria in the evaluation of the creditworthiness of the purchasers of shares or ownership. The banking sector of Serbia has already "entered" the European Union, since 80% of the sector is now in the hands of owners from those countries." (Source: NBS, Godišnji izveštaj 2006, p. 4).

[52] Statement of NBS Governor, R. Jelašić: "Perhaps it was unrealistic, but I was hoping that foreign banks would bring to our market practices from their own countries. However, they are acting as though they don't understand what, for example, transparent service means when it comes to crediting (...) I have nothing against banks approving credits with high interest rates, but they must have a clear and comprehensible credit price policy. Consumers in Serbia unfortunately do not know enough about banking. I was hoping that foreign banks would bring us positive practices and self-regulatory mechanisms, that they would strengthen the confidence in banking by helping their clients understand it better, but it's obvious that, to them, quick profits are more important than the long-term building of confidence." ("Banke skrivaju kamatne stope," *Pregled*, 17-19.3.2006).

- BIS: "Foreign direct investment in the financial sector – experiences in Asia, central and eastern Europe and Latin America", Committe on the Global Financial System, SGFS Paper No. 25, June 2005, Basel.
- Canalis, J.: *Universal Banking*, Claredon Press, Oxford, 1997.
- Cigna, G. P.: "Law in transition online 2005", EBRD, October 2005, London.
- Cottarelli, C. – Dell'Ariccia, G. – Vladkova-Hollari, I.: "Early Birds, Late Risers and Sleeping Beauties: Bank Credit Growth to the Private Sector in Central and Eastern Europe and the Balkans", IMF Working Paper No., 213, November 2003, Washington D.C.
- Demirguc-Kunt, A. – Ross, L.: "Bank based and market – based systems: cross-country comparisons", World Bank, Washington and University of Minnesota, June 1999.
- Domanski, D.: "Foreign banks in emerging market countries: changing players, changin issues", *BIS Quarterly Review*, December 2005, Basel, pp. 69-81.
- Duenwald, C. – Guerguiev, N. – Schaechter, A.: "Too Much of Good Thing? Credit Booms in Transition Economies: The Cases of Bulgaria, Romania and Ukraina", IMF Working Paper WP/05/128, June 2005. Washington D.C.
- Đukić, Đ.: "Opasna spirala zaduživanja", *Finansije Top 2006*, *Biznis & Finansije*, jun 2007, pp. 22-25.
- ECB: "Euro Area MFI Interest Rate Statistics: June 2007", 7 August 2007.
- Fries, S. – Taci, A.: "Banking reforms and development in transition economies", EBRD Working Papers No. 71, June 2002, London.
- Gelos, G. R. – Rolodos, J.: "Consolidation and Market Structure in Emerging Market Banking Systems", IMF Woring Paper No. 186, June 2002, Washington D.C.
- Goddard, J. A. – Molyneux, P. – Wilson, J.: *European Banking: Efficiency, Technology and Growth*, John Willey & Sons, Ltd. Baffins Lane, Chichester, England, 2001.
- IMF: "Global Financial Stability Report", April 2006, Washington D.C.
- IMF: "Global Financial Stability Report", September 2006, Washington D.C.
- IMF: "Global Financial Stablity Report", April 2007, Washington D.C.
- IMF: "World Economic Outlook", April 2007, Washington D.C.
- Kapor, P.: "Strana ulaganja u domaće banke", *Poslovna politika* br. 12/2000, pp. 29-33.
- Kapor, P.: "Uticaj strukturnih promena u finansijskom sektoru na ekonomsko-finansijske odnose Srbije sa inostranstvom", Zbornik sa okruglog stola "Ekonomsko-finansijski odnosi Srbije sa inostranstvom"

održanog 26.6.2007, Naučno društvo ekonomista, Akademija ekonomskih nauka, Ekonomski fakultet, Beograd, pp. 145-156.
- Kapor, P.: *Bankarstvo*, Megatrend univerzitet primenjenih nauka, Beograd, 2005.
- Liewellyn, D. T.: "Globalisation and convergence on the shareholder value model in European banking", BIS Paper No. 32, *Financial globalization*, December 2006, pp. 20-29.
- Moreno, R. – Villar, A.: "The increased role of foreign banks in emerging markets", BIS, BIS Papers No. 23, 2005, Basel, pp. 9-16.
- NBS: "Anketa o poslovnim aktivnostima i namerama banaka" (sprovedena sredinom jula 2007. g.); www.nbs.yu
- NBS: "Bankarski sektor u Srbiji – Izveštaj za IV kvartal 2006. g.", april 2007.
- NBS: "Ekonomski pregled", januar, april i jul 2006.
- NBS: "Godišnji izveštaj 2006"
- NBS: "Nadzor finansijskog lizinga – Izveštaj za IV tromesečje 2006"
- NBS: "Republika Srbija: stanje spoljnog duga po vrsti dužnika"; www.nbs.yu
- NBS: "Sektor osiguranja u Srbiji – Izveštaj za 2006. g."
- NBS: "Statistički bilten", maj i jun 2007.
- Odluka o usklađivanju bruto plasmana odobrenih stanovništvu sa osnovnim kapitalom banaka, "Sl. glasnik RS" br. 39/2006, 49/2006, 64/2006, 31/2007 i 77/2007.
- OECD: "Conutry Risk Classification of the Participants to the Arrangement on Officially Suported Export Credits", Valid as of 03 April 2007, Paris.
- Raiffeisen Zentralbank Oestereich AG: "CEE Banking Sector Report", September 2006, Vienna.
- Ristić, Ž. – Vuković, D.: "Bankarski sistem Srbije", *Jugoslovensko bankarstvo*, br. 3-4/2002, pp. 3-19.
- Ritter, L. S. – Silber, W. L. – Udell, G. F.: *Principles of Money, Banking & Financial Markets* (10th Edition), Addison-Wesley Longman Inc., Reading, Massachusetts, 1999.
- Štimac, M.: "Bankocentrična tranzicija", *Ekonomist magazin*, br. 345/346, 1.1.2007, pp. 27-29.
- Trifunović, P.: "Banke – organizacija, poslovanje i politika", "Službeni list SFRJ", Beograd, 1980, pp. 7-8.
- Uputstvo o primeni odluke o jedinstvenom načinu obračuna i iskazivanja efektivne kamatne stope na kredite i depozite, "Sl.glasnik RS" br. 11/05, 28/05 i 108/05.
- Zakon o deviznom poslovanju, "Sl. glasnik RS" br. 62/2006.
- Zakon o deviznom poslovanju, "Sl. list SRJ" br. 23/02 i 34/02.
- Zakon o porezu na dobit preduzeća, "Sl. glasnik RS" br. 25/01, 80/02, 135/04, 62/06, 65/06.

Scientific review article
Received: August 21st, 2007

UDK 316.77:339.727.22(497.11)

Assistant Professor Tatjana Cvetkovski, PhD
Faculty of Business Studies, Megatrend University, Belgrade
Violeta Cvetkovska-Ocokoljić, M.A.
Faculty of Culture and Media, Megatrend University, Belgrade

INTERCULTURAL COMMUNICATION AND THE DEVELOPMENT OF SERBIA

Abstract: *Considering Serbia's opening to the world and the accompanying increasing investment in it, special attention should be paid to the development of adequate systems of communication with the environment and with investors. However, it should not be forgotten that Serbia has passed through a great crisis over the past several decades, characterized by a fear for existence in every sense of the word. Conditions of life imposed a different kind of thinking and communication, in which there was no room for kindness and courtesy. With the belief that the bad times are a thing of the past and relying on development through increasing investments in Serbia, it is necessary to adjust to the changes and, before all, relearn the language of successful social communication, both in Serbia and in global relations.*

Key words: *Communication, development, environment, and investments*

1. Introduction

Communication is a means of exchanging meanings. It exists when there is an exchange of messages, in which the sender has understood the message sent to him by the sender. Communication encompasses verbal messages – words, as well as non-verbal messages – facial expressions, gesticulation, tone of voice or the way something is said, and the like. Also, it encompasses consciously sent messages as well as unconscious and unplanned messages sent by the sender.

All business activities require communication. Effective communication is significant in doing business on the domestic market, where the work force is relatively homogenous. However, motivation, decision-making, negotiations and other activities require the manager, when we speak of business globally, to successfully communicate with colleagues, clients, and partners from other cultures. Due to the mixing of cultures that comes as a result of necessary business expansion, there arises a need to get to know and to harmonize hab-

its, communication, and behavior with those with whom business is being done. Naturally, the greater pressure is on the organization beginning to do business in a certain country to adjust to the latter's rules; however, the host countries themselves also have a great responsibility to attract foreign companies to invest in them. Certainly, we will neither want nor be able to change our own tradition, values and everything else that is deeply ingrained, as these have been woven in through time and history. However, legal regulations, ways of communication, the development of infrastructure are all something that we can work on, if we want to attract foreign investors. On the other hand, demands for business expansion have created a need for managers trained for global management, who possess skills of working with people from other countries. They need to be able to understand the behavior of people who belong to other cultures and to know how to work together with organizations, i.e. employees from those cultures.

These are the problems that characterize today's business environment. However, for Serbia, communication with existing investors is especially important, as well as with those organizations interested in investing here. Generally speaking, it is important to create a good business climate.

All communication has a message sender and a message receiver. In order to understand what good communication means and where problems can arise, it is necessary to review the entire process of communication.

2. The communication process

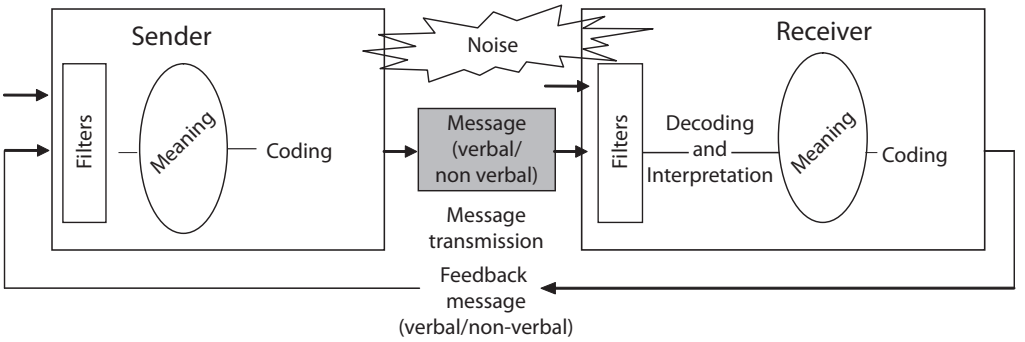
The following can be identified as the basic elements of the communication process.[1]

Meaning – which is formed in the message sender's consciousness on the basis of data he is receiving through his five senses. In fact, all information from the environment is received through seeing, hearing, touching, smelling or tasting. Damage to any one of these senses can be an obstacle to the reception of the said information. However, even when there is no damage to the senses, due to their limited powers of reception from the environment, we always choose the stimuli to which we will pay attention, because it is impossible to pay attention to everything that is happening around us.

Filters – can be philosophical and psychological. All data coming from the environment are filtered through beliefs, values, and the person's current physical or emotional state. A filter is something that limits the ability of the person to sense or observe a stimulus.

[1] T. Cvetkovski, V. Cvetkovska-Ocokoljić, *Poslovna komunikacija*, Megatrend univezitet, Beograd, 2007, p. 14.

Figure 1. *Elements in the communication process*



Source: T. Cvetkovski, V. Cvetkovska-Ocokoljić: *Poslovna komunikacija*, Megatrend univezitet, Beograd, 2007.

Communication begins from the level of intrapersonal communication, when the message sender decides which idea or which meaning he wishes to transmit to the message receiver, and crosses to the level of interpersonal communication when the sender codes the message and chooses the channel by which he will transmit it to the receiver.

By way of ***coding***, internal thoughts are transformed into a form understandable to the message receiver. This is necessary, since people do not have the capability of reading another's thoughts. Thus, thoughts have to be changed into a form that is observable to the message receiver. The way of coding depends on whom the message is being sent to and the message's purpose. Since the symbols used in the course of message coding are sometimes ambiguous, it is necessary to select them carefully. Erroneous message interpretation often arises when the message receiver is not close to the sender's language and culture.

Message – represents a concrete physical product of coding. This can be text, image, arm movements, head shaking, facial expressions, speech.

Channels are means by which the message travels from the message sender to the message receiver. When choosing channels one should pay attention to how complex the transmitted message is and whether it is being transmitted to one person or a greater number of people. The channel also becomes a part of the message. Attention should be paid that the choice of channels matches the message receiver's reception abilities. For example, one should not use visual channels in communications with the blind or audio channels in communications with deaf persons.

Decoding is the opposite process of coding. The sent message arrives to the message receiver. The latter, just like the sender, observes the message through

his senses. This means that, during the message's reception, the message is filtered through the receiver's physiological and psychological filters. The receiver now decodes the received message in order to asign it a specific meaning.

In order to understand the message, it is not enough to merely decode; the decoded symbols must also be asigned a meaning. The receiver attempts to reconstruct the message sender's idea, assigning meaning to the symbols and ***interpreting*** the message as a whole. Communication appears only when the receiver has received the message and understood it in the way intended by the sender.

We find out how the message has been understood through feedback, i.e. the ***return message***. That is the receiver's reply to the message. In the absence of a reply, such as a return e-mail or a nod of the head, the sender cannot know how successful the communication was. In the course of sending the return message, the roles of the sender and the receiver of the message are changed, and, in order to send the return message, it is necessary to once again code the meaning into the message.

Noise appears in the course of communication. Noise is everything that interferes with communication.

- *Physical noise* can be factory noise, a blot on printed text, the inability to hear the lecturer due to a faulty microphone, and the like. *Physical noise includes all interference that appears in the course of the message's transmission from the sender to the receiver, i.e. from its source to its target.*
- *Semantic noise* is everything that interferes with the message receiver to understand the message in the way wanted by the sender. The cause can lie in the message sender's and receiver's different experiences, views, beliefs, values, knowledge, individual communication skills, prejudices, belonging to different social classes or cultures. The essence of successful intercultural communication lies precisely in the discovery and avoidance of semantic noise. *Semantic noise causes misunderstanding between the message sender and receiver, even if the message has been exchanged.*

Communication is successful only when the message receiver has understood the message in the way intended by the sender.

As can be seen, successful communication demands time, patience and a desire to understand the collocutor. It does not happen by itself. Unfortunately, often when misunderstanding occurs, neither the message sender nor the message receiver are wholly conscious of what went wrong and why. By analyzing each individual step of the communication process, one can discover where the problem in communication arose. However, a frequent problem lies in a total lack of communication or its insufficiency.

2.1. Communication, perception and culture

Hence, it is evident that communication is tied to perception. Communication is the process of transmitting and understanding messages. Since man is a social being, strongly influenced by the environment, human thoughts, feelings, conclusions and behavior are a result of interaction with the environment.

By way of their senses, people observe their environment and what is happening in it, and assign meanings to what they see. The receive stimuli from the environment through their senses. Everyone pays selective attention to some aspects of his environment and selectively ignores other aspects. The selection process encompasses internal and external factors, the filtering of perception and the choosing of the stimuli to which the greatest attention will be paid. The person then organizes the selected stimuli into forms that give them meaning.[2]

The way in which people interpret what they observe varies. Hand waving can be interpreted as a friendly gesture or as a threat, depending on the circumstances and a person's state of consciousness in a given situation. The way in which a person interprets stimuli will lead to the reply: open (action) or concealed (motivation, views and feelings), or both. Each person selects and organizes stimuli differently and, thus, differently interprets them and replies to them.

Experience and age, i.e. previous observations, influence the individual's formation into a specific personality, and this previous experience determines the person's further thoughts and behavior. Because communication is connected to perception, many actions or reactions of people around us may seem funny or unjustified to us, and the reason for this is that we don't understand their way of looking at things, i.e. the background of their reactions.

When we speak of communication and perception, the question of the relationship between communication and culture inevitably arises. The individual's communication style is tied to the culture in which he was born and raised. Culture determines the way of thinking and behavior of the people that belong to it. We develop it through the symbols by which we communicate, by which we transmit our common thoughts and expectations. The most obvious influence of culture on communication is reflected in the fact that the members of certain groups live on the same territory, speak the same language or write in the same script. The most obvious difference between people who originate from different cultures lies in the fact that they speak a different language and share different beliefs and expectations.

Today's business environment is full of everyday cultural encounters. They characterize the business environment as well and, whether we like it or not, influence our lives. In Serbia, this influence is felt even more strongly, since, after

[2] M. Jovanović Božinov, M. Živković, T. Cvetkovski, *Organizaciono ponašanje*, Megatrend univezitet, Beograd, 2003, p. 71.

years of isolation, we once again have foreigners in our environment, and are once more doing business, cooperating, socializing and coexisting with them.

3. Intercultural communication

Intercultural communication occurs when a person belonging to one culture sends a message to a person belonging to another culture. In this case as well, misunderstanding occurs when the person from the other culture does not receive or does not comprehend the message in the way intended by the sender. The greater the difference between the message sender and receiver's cultures, the greater the chances of intercultural misunderstanding. Misunderstandings are more pronounced during intercultural communication, since people from different cultures see, interpret and evaluate events differently. A specific problem arises in the course of interpreting non-verbal messages. Language can be learned but, in order for people to understand each other, it is also necessary to understand non-verbal messages properly. Since views, beliefs and communication styles vary with different cultures, verbal and non-verbal business messages should be designed with these differences in mind. For example, our proximity to our conversation partner will depend on the culture to which we belong. In Latin America and, partly, in Europe and the Arab countries, it is natural to stand very close to the other person during the course of conversation, while in American culture this proximity causes discomfort.[3]

Perception is a process by which stimuli from the environment are selected, organized and evaluated. Perception depends on culture, while the problems arise in the course of interpretation.

Through interpretation, the individual assigns meaning to what he has seen. It helps us organize our experiences and directs our behavior. On the basis of our experiences we make suppositions regarding what we have seen, so that, when we find ourselves in a similar situation, we will not be rediscovering the meaning each time.[4]

Cultural differences influence organizations and their business primarily through their employees. The culture in which people are raised influences the way in which their personalities are formed. Parents transmit to their children tradition, rules of behavior, values; teach them how to behave, while, on the other hand, personality formation is also influenced by experience accumulated through life. Different cultural systems produce different styles of behavior, which are shaped by the culture, geography, history and political system

[3] T. Cvetkovski, V. Cvetkovska-Ocokoljić, *Poslovna komunikacija*, Megatrend univezitet, Beograd, 2007, p. 13.

[4] N. Adler, *International Dimensions of Organizational Behavior*, South-Western (a division of Thomson Learning), Canada, 2002, p. 79.

of each nation.[5] As already formed individuals, people work in organizations, transmitting and building a part of their personality into the organizational culture. On the other hand, people's expectations regarding the work itself are different. They are partly determined by conditions of life, interests or the culture in which they were raised, and this will doubtlessly influence variations in the behavior of employees across different cultures.

Intercultural communication is inevitable. People look at things from their own angle, and tend to seek what is bad in other cultures and to denigrate them. However, instead of concentrating on differences, it is better to create cultural synergy for, as emphasized by many, diversity is a base for creativity and creativity is the basis of success and progress. This is just one more reason why we need to ask ourselves why it is important to have good communication with our environment.

4. Investments in Serbia and the significance of communication

If we take into account the fact that an increasing number of countries is investing in Serbia, and that communication and business communication are skills to which no particular attention has been devoted in Serbia, then it becomes clear why it is important to deal with communication. Thus far, Serbia has seen investments from companies such as "Phillip Morris," "Tetrapak," "Henkel," "Agrocor," "US Steel," "Ball Corporation," "Merkator" and many others. According to the data of the Croatian National Bank, in 2006 Croatian companies invested the most in Serbia and Montenegro, Bosnia and Herzegovina and Switzerland. Slovenian companies have so far invested more than 900 million euros in Serbia. As for foreign direct investments, in 2006 they mostly came from Norway, Germany, Greece, Great Britain, Croatia, and Bulgaria. Foreign investors invested the most in telecommunications, banking and insurance, industrial processing, etc., and somewhat less in the production of food and beverages, and in the tobacco and textile industries. It is obvious that Serbia is becoming interesting for foreign investors; the only question is how well we understand each other and where and how we can help one another.

Having already mentioned "Ball Corporation" ("Ball Packaging Europe"), it is important to point out that Serbia is this company's center of regional operations, which produces about 2 million cans each day. Anthony Barnett, the director of the Zemun-based plant, says that the main reason why the most modern plant in "Ball Packaging's" entire system was built in Serbia was its geographical location. Serbia is located in the central portion of the region, which reduces the transport costs of the company, whose biggest buyers are

[5] M. Jovanović Božinov, A. Langović-Milićević, *Interkulturni izazovi globalizacije*, Megatrend univerzitet, Beograd, 2006, p. 8.

located in the region and in Serbia itself. Geographical position is certainly one of Serbia's advantages. However, other extremely important factors are things such as investment incentives, available work force and availability of the necessary infrastructure. It is interesting that many companies point to the work force as an important investment factor: expertise, readiness for further training, good communication ability, knowledge of foreign languages.

The main complaints of investors already doing business in Serbia are: unresolved property ownership issues, lack of contract security and over-administration.[6]

5. Basic directions of communication development in Serbia

When speaking about communication in Serbia, the problem is that there is almost no true business communication (in which the message sender and receiver understand each other and wish to communicate) to speak of. That is why it is important to distinguish and separately examine three basic directions of communication development, which are mutually intertwined:

1) ***The development of good communication and, especially, business communication within the domestic organizations.*** This is perhaps also one of the ways in which the rules of cultured behavior and courtesy can once again return to everyday life. Good communication between an organization's employees would certainly improve business efficiency and influence the development of a positive organizational culture and climate. Just how important organizational culture is can be understood if we take into account the fact that people are spending more and more time at the workplace. It is a question how productive they will be if they are unhappy with their environment, possibilities, communication, and the like. The trend of the development of communication skills among employees has been imposed by new business conditions. Companies with foreign capital initiated seminars for the development of employee communication skills in Serbia, but this trend is now spreading to other companies as well.
2) ***The establishment of good intercultural communication and cooperation with countries and organizations that invest or are interested in investing in Serbia.*** This is an especially significant means of communication for our country at this time. Attracting and keeping investors, knowing what satisfies them, where problems arise, and cooperating with them is possible only through communication. In the process of attracting and executing investment projects, one of the most important roles is that of the municipal authorities and constant communi-

[6] http://www.siepa.sr.gov.yu/attach/InvesticijeDecembar2006.pdf

cation with them. For this reason, in 2005, SIEPA (Serbian Investment and Export Promotion Agency) started a project of establishing permanent communication with municipal authorities throughout Serbia and created a network of about 170 professionals, economic advisors, city managers or municipal presidents with whom it maintains permanent communication and with whom it is successfully dealing in the process of carrying out various projects.[7] Another important organization is the Council of Foreign Investors, established in July 2002, which represents its members before the Serbian state organs, provides them with information about the business climate in Serbia and helps in harmonizing domestic laws with international business practices. In this way, the Council helps facilitate better communications between existing investors and the Serbian authorities, stimulates the arrival of new investors, in addition to which it can play the part of mediator in the establishing of cooperation between domestic firms and its members.[8] This is especially important if we take into account that the Council's includes members such as "British American Tobacco," "Delyug," "HVB Bank," "OMV," "Siemens" and other large companies, and that their numbers are growing and currently number more than 100 members. Especially significant for Serbia is the *White Book* issued by the Council annually, which contains the views of members regarding the business environment in Serbia and recommendations for removing the shortcomings. It would be useful to consider these recommendations and use them toward the establishment of a better business climate.

3) ***The development of good intercultural communication with countries to which Serbia (i.e. its firms) exports its goods or in which it seeks to invest capital.*** Serbian companies are only beginning to think about investing abroad. Those that have already done so are few, while many emphasize the limitations they face in terms of taking capital out, competition, political pressure and market conditions. The largest investment was that of "Telekom Srbija," which purchased 65% of the state capital of "Telekom Srpska" for 646 million euros at the end of 2006. The decision to sell "Telekom Srpska" to "Telekom Srbija" was made by the government of Republika Srpska, the Privatization Council and the Tender Commission.

"Sintelon" and "Hemofarm" have also invested abroad. However, these companies, which were until recently domestic, are now majority foreign-owned.

7 http://www.serbioanfurniture.org/srpski/aktuelno/siepa/investicijeinv_dec_06/budite_dobri_prodavci.htm.

8 http://www.siepa.sr.gov.yu/attach/exporter_8.pdf.

For 2008, "Delta Holding" is planning to open a chain of "Costa Coffee" cafes in Croatia, in which it will invest four million euros during the first five years. The same company will invest about 27 million euros in building the first shopping center in Montenegro, "Delta City Montenegro."[9]

India's ambassador in Belgrade, Ajay Swarup, has stated that his embassy will soon open a business center, which will provide all those interested in doing business with India with information and contacts about the markets and the companies that interest them.[10]

Even though domestic investment outside of Serbia's boundaries is still small, it should be born in mind that – just as the country in which money is being invested must strive to establish good communication and work conditions for foreign investors – it is equally important for the investing firms to leave a good impression and gain a good reputation in the country in which they are investing. On the one hand, communication with the employees is extremely important. It will show that the organization is interested in the views of its employees, which will help establish good relations, motivate the employees and, by extension, help the company successfully execute its projects. As for relations with the surroundings, many companies use donations or sponsorships to gain the sympathies of their surroundings and build good relations with them. Relations with employees are especially important when buying existing companies. Each change, and especially a change of ownership, causes employee resistance. The reason for this is that changes bring uncertainty, and if there is no communication with employees, uncertainty and fear are even greater. This could cause future strikes and profit losses – which could be avoided.

The organization must have developed communications both internally and externally with the environment, in order to be able to react quickly to all eventual changes, problems and crises. Poor relations with the local community can seriously imperil the company's business. This is why corporate communications should be an integral part of business strategy and policy. The organization should communicate with customers, clients, business partners, employees, banks, government institutions and non-governmental organizations. Internal and external communications are of equal importance. Communication is the means of establishing trust between partners, which reduces uncertainty, i.e. the turbulence of the environment, and there is a lesser probability that someone will come between the company and its environment, and impose himself as a new partner or inflict some other sort of damage to the company.

[9] http://www.siepa.sr.gov.yu/attach/exporter_8.pdf.

[10] http://www.siepa.sr.gov.yu/attach/exporter_8.pdf.

6. The significance of good communications – regional and local aspects

Many organizations that invest in Serbia complain about inadequate communications with the municipal authorities, while all analyses of foreign investment inflows and factors on which they depend place special accent on the efficiency of the functioning of the state and the local authorities. It is not certain whether the problem lies in the fact that the latter don't know how, do not desire it or lack the necessary resources, but foreign investors emphasize that improvements are still small and that things have essentially not moved past expressions of good will to improve work conditions and cooperation on various projects.

It is a fact that we still have complex bureaucratic procedures that obstruct successful business, which demands quick reactions and flexibility. The municipalities do not have persons who deal with investment projects. Kamil Beffa, director of the "Lafarge" cement factory in Beocin, says that municipal officials should be more available to investors and need to develop a proactive approach.[11] Some investors point to negative aspects such as the complicated legal regulations, insufficiently defined roles and competencies of the various organs of government, an absence or insufficient understanding of national interests on the part of local government organs and the resulting administrative inefficiency that slows down the investors' business.

The question that comes up is: how can municipalities influence the scale and types of foreign investments in Serbia? One approach is to organize presentations at investment fairs. Another is the use of promotional materials that show the advantages of investing in a particular municipality. For those who have already invested, what is important is an active municipal administration, which has good communications with the investor and is prepared to provide support and aid when necessary. On the other hand, the municipalities should not simply wait for the investors to address them, but should be active in finding and attracting good investors. For that it is necessary to find the right channels and adjust the communication content. It is not sufficient to use only one channel for sending messages. The greater the number of channels, the greater the probability that the message will be received and understood. Also, it is necessary to adjust the channel to concrete situations and concrete investors. If we are addressing a specific investor, we should use the communication to underline the benefits he would gain by investing in Serbia, i.e. in the concrete municipality. Good communication is of vital significance, as large global companies that invest money seek to establish long-term relationships with domestic institutions since, among other things, they are engaged in large and long-term projects. Even though Belgrade was chosen by "FDI Magazine" (a specialized magazine for foreign direct investments, published

[11] http://www.siepa.sr.gov.yu/attach/InvesticijeDecembar2006.pdf.

by the "Financial Times") as the city of the future in southern Europe, much more needs to be done in order to attract good foreign investors and choose and carry out the appropriate investments. The categories and criteria on the basis of which the magazine panel made its decision were economic potential, costs of doing business, human resources, transport, information technologies and telecommunications, quality of life for foreign investors, as well as the best promotional strategy for attracting foreign direct investments.[12]

The mere fact that Belgrade found itself among cities such as London ("the city of future for Northern Europe"), Paris ("the city of future for Western Europe"), Brno ("the city of future for Central Europe") and Baku ("the city of future for Eastern Europe"), speaks enough about the fact that Serbia's outlook is good. However, the way in which that will be taken advantage of in practice depends on the readiness of the country and the municipal organs to create the necessary climate through mutual cooperation and better communication, and to attract investors by way of concrete moves.

A higher level of business culture and communication has a positive influence on investments, because it is an indicator of the state of business consciousness. Since profit is what pushes people into action, we can count on the fact that, if no other stimuli succeed in compelling us to think more seriously about the importance of good communication, the possibility of making a profit certainly will. On the other hand, we should also know what we want, have a strategy and work on developing a good business climate, but not at our own expense, for true partnership with investors is possible only when the benefits are mutual.

[12] http://www.siepa.sr.gov.yu/srp/news/beograd_nagradavesti.htm

References

- Adler, N.: *International Dimensions of Organizational Behavior*, South-Western (a division of Thomson Learning), Canada, 2002.
- Cvetkovski, T. – Cvetkovska-Ocokoljić, V.: *Poslovna komunikacija*, Megatrend univerzitet, Beograd, 2007.
- Cvetkovski, T. – Langović, A.: "Komunikacija u uslovima globalnog poslovanja", *Proceedings*, international academic gathering "Radikalne promene u preduzećima i privredi u uslovima globalizacije," Belgrade, November 28, 2003.
- Cvetkovski, T. – Langović, A.: "Novo radno okruženje i korporativna kultura", *Proceedings*, VIII International Symposium "Menadžment u novom okruženju" – SIMORG 2002, Zlatibor June 2-5, Faculty of Oranizational Sciences, Belgrade, 2002.
- Cvetkovski, T. – Langović, A.: "Uticaj kulture na vođenje projekata", *Proceedings*, V Project Management International Symposium "Upravljanje projektima – nove tendencije" – YUPMA 2001, Zlatibor 28-30.5.2001, Yugoslavian Project Management Association, Belgrade, 2001.
- Cvetkovski, T. – Langović-Milićević, A. – Cvetkovska-Ocokoljić, V.: "Koliko kvalitetno zaposleni obavljaju svoj posao?", IV gathering of entrepreneurs and scientists, "Menadžment tehnologije i inovacija – ključni faktor superiornih operacija i konkurentnosti," November 2-3, 2006, Faculty of Oranizational Sciences,\ – Institute for Operative Management, Belgrade, 2006.
- Cvetkovski, T. – Langović-Milićević, A. – Cvetkovska-Ocokoljić, V.: "Pregovaranje između projektnih timova – specifičnost interkulturnih pregovora", XI Project Management International Symposium, "Projektno upravljanje organizacijama – novi pristupi" – YUPMA 2007, Zlatibor, May 15-17, 2007, Project Management Association of Serbia and Montenegro, 2007.
- Cvetkovski, T. – Langović-Milićević, A. – Cvetkovska-Ocokoljić, V.: "Uticaj karakteristika ličnosti zaposlenih na organizaciono ponašanje i poslovanje organizacije", Symposium of Economic Research and Design 2006, iipp; Editorial board of the Magazine for Economic Research and Design, Faculty of Mechanical Engineering, Belgrade, Academy of Engineering Sciences of Serbia and Montenegro (under the patronage of the Ministry of Science and Environmental Protection of the Republic of Serbia), 2006.
- Cvetkovski, T. – Langović-Milićević, A. – Cvetkovska-Ocokoljić, V.: "Uzroci i posledice loše komunikacije u projektu", XI Project Management International Symposium "Projektno upravljanje organizacijama

– novi pristupi" – YUPMA 2007, Zlatibor, May 15-17, 2007, Project Management Association of Serbia and Montenegro, 2007.
- Cvetkovski, T. – Langović-Milićević, A.: "Problemi socijalizacije i orijentacije zaposlenih", SymOrg 2006, X International Symposium "Promene u organizaciji i menadžmentu – Izazovi evropskih integracija", *Proceedings*, Zlatibor, June 7-10, 2006, Faculty of Organizational Sciences, Belgrade, 2006.
- Cvetkovski, T. – Marinković, D.: "Analiza posla kao prvi korak ka poboljšanju performansi preduzeća", International Scientific Gathering "Poboljšanje performansi preduzeća – uslov poslovanja u evropskom okruženju", Megatrend University, Belgrade, December 10, 2004.
- Grgurević, S. – Cvetković, N. "Perspektive i okviri regionalne ekonomske saradnje na Balkanu", u zborniku skupa "Privredna kretanja i ekonomska politika 2003/2004", *Ekonomski anali*, NDEJ, Faculty of Economics, Belgrade, December 17 and 18, 2003, pp. 275-282.
- http://www.serbioanfurniture.org/srpski/aktuelno/siepa/investicijeinv_dec_06/budite_dobri_prodavci.htm
- http://www.siepa.sr.gov.yu/attach/exporter_8.pdf
- http://www.siepa.sr.gov.yu/attach/InvesticijeDecembar2006.pdf
- Jovanović Božinov, M. – Kulić, Ž. – Cvetkovski, T.: *Menadžment ljudskih resursa*, Megatrend University of Applied Scineces, Belgrade, 2004.
- Jovanović Božinov, M. – Langović-Milićević, A.: *Interkulturni izazovi globalizacije*, Megatrend univerzitet, Beograd, 2006.
- Jovanović Božinov, M. – Živković, M. – Cvetkovski, T.: *Organizaciono ponašanje*, Megatrend univerzitet, Beograd, 2003.
- *Marketing*, no. 298, February 2, 2006.

Expert article
Received: August 21st, 2007
UDK 316.334.2:339.13(485)

Professor Ana Langović-Milićević, PhD
Faculty of Business Studies, Megatrend University, Belgrade

CULTURE AND MANAGEMENT
One Experience: Sweden

Abstract: *When engaged in business activity it is very important to avoid beginners' mistakes related to deeper layers of communication. The new age brings technical and technological advances, improved communications and the shrinking of gaps that exist between various cultures. Economics crosses the borders of the home country, creating the world of the global economy. This global economy gives rise to new obstacles related to the harmonization of various business practices in various cultures. All cultures are guided by their economies, and the values of each culture are reflected in the way that those who belong to them conduct their business dealings.*

Several studies have indicated the influence of culture on the choice of negotiation strategy in international business negotiations. Another set of studies indicated that concessions and agreements in negotiations are also influenced by culture. The understanding of cultural differences is the first step toward successful negotiatons and subsequent successful business cooperation, as well as the creative management of multicultural teams. Differences in style do not have to present obstacles but can be complementary. In literature dealing with international business, most empirical works have dealt with negotiations outcomes influenced by culture, while theoretical works, such as this one, are more oriented toward an explicit study of cultural differences between countries.

Key words: *culture, international business, cultural differences*

1. Introduction

The new world of business brings expansion of business activities beyond the borders of the home country, which makes it necessary to learn the culture and customs of one's business partners. Successful business also requires a new approach on the part of the management, with the goal of gaining a better understanding of national influences and cultures on a particular company's

business. The new business environment demands networking, opening and global activity, and an increasing accent on intercultural communications.[1]

The way companies do business is influenced by culture, which means that the neglect of this important factor brings unfavorable business results. In order to establish a business relationship, the first step is to negotiate, and this is when the necessity of being acquainted with the business partner's culture becomes evident. Regardless of the differences that exist between the Germanic countries and the European south and southwest, for example, there is always a desire to get to know the potential business partner well before entering a business relationship, for the sake of establishing a successful relationship. Successful communications require respect for differences and a readiness to accept and understand them. The reasons why people behave according to certain rules can be explained by way of four cultural dimensions:[2]

- *distance of power* – indicates an acceptance of the fact that the distribution of power is unequal (distance of power is **great**: there is blind obedience to superiors on the part of subordinates; distance of power is **small**: the employees participate in decision-making and lower-level employees are highly qualified);
- *avoidance of insecurity* – members of this cultural dimension do not like insecurity, there is a markedly strong faith in experts and their knowledge; the tendency is to structure all activity and establish a large number of written rules;
- *individualism* – indicates a tendency of people to take care of themselves and their families; company employees seek to solve only their own tasks, without expecting help from others or a desire to extend help;
- *masculinity/femininity* – in the cultural dimension of *masculinity*, the dominant values are: success, money and material goods; *femininity* is a dimension opposite to the dimension of *masculinity*, dominated by social values, care for those who are close and quality of life.

Members of various cultures can variously interpret the negotiating process: the approach to negotiations can be deductive (from general postulates to details), as opposed to members of some other cultures who begin with the details and then move on to more general issues. The goal is to show the importance of being familiar with the characteristics possessed by members of certain cultures, to understand them and respect them during negotiations and, of course, subsequently, in the course of the business relationship.

[1] T. Cvetkovski, V. Cvetkovska-Ocokoljić, *Poslovna komunikacija u savremenim uslovima poslovanja*, Megatrend University, Belgrade, 2007.

[2] G. Hofstede, *Culture and Organizations: Software of the mind*, McGraw-Hill, London, 1991.

2. Business style

Sweden is a country with a rich cultural heritage and great natural beauty, which has taken advantage of its natural resources (ores, forest potentials…). Sweden is striving to turn its name into a brand of a producer of clean, healthy food. It is planning to develop and use new technologies, which will engage all structures of society, since everyone's future is dependent on it. Forest and ore products make up a significant part of Swedish exports. However, the real miracle has, in fact, happened in the engineering industry, with firms such as "Electrolux," "Ericson," Atlas Copco," "ASEA/ABB," and "Alfa Laval," which are successfully competing in the global marketplace. It should be noted that two of the strongest automobile industry brands in Europe are Swedish – "Volvo" and "Saab."

The export dependence of Swedish industry is a weakness that has, on the other hand, oriented Swedish companies to constantly bring out new products, innovate, restructure, increase efficiency and plot long-term strategy. This approach is especially prevalent in companies in the **information technology** and **chemical-pharmaceutical** sectors, where intensive research work is emphasized. The best examples of this are "Ericson" (mobile telephones) and "Astra" (pharmaceutical company).

A specific business philosophy, whereby modern functionality and design are brought closer to broad categories of consumers, has stood behind "Ikea" and "Hennes&Mauritz" (H&M) and their successful development of global networks. The "Swedish style" has become famous, a culture of furnishing living spaces with an air of elegance, functionality, modern design and simplicity. Traditional Swedish cuisine and the "Swedish table" (all-you-can-eat) are especially prominent. Music holds a very important place in this culture, and Sweden is known as a large music exporter. In fact, culture and tradition influence all segments of Swedish life including, naturally, the business culture.

It can be said that the basis of the humane Swedish society was laid in the 16th century, when almost the entire Swedish population converted to Protestantism, which gave rise to the development of a capitalist economy. The Protestant work ethic influenced the Swedish mentality and accelerated the development of the cities and industries. It is interesting to note that economic development was accompanied by a new influx of immigrants – Catholics and Jews – in 1780. This is evidence of the fact that the Swedes were never ethnically or religiously intolerant and that multiculturalism and tolerance have existed in their society for centuries.

One specific feature of Sweden is the fact that it has the best organized labor unions. According to official statistics, six out of seven employees belong to a labor union. About 2.3 million workers are members of one of the 24 labor unions associations of the Central Organization of Workers Unions, while

1.3 million employees are members of the Central Labor Union Organization of Employees. Academically educated employees are members of the Central labor Union Organization of Academically Educated Employees. Swedish business proprietors are also well organized; the Association of Private Employers has about 50,000 companies. Employers in the state, tax and municipal sector have their own organizations. Laws and labor agreements very well protect the workers, and it is very difficult for empoyers to dismiss them. All problems between employer and employee are solved peacefully, based on a mutual agreement formed back in 1930.

Sweden is a country with a pronounced *feminine* cultural dimension, which greatly cares about its fellow citizens, their economic security and human rights. Swedes are taught from childhood that all people have a right to happiness, that one should also care for others, that all people are equal, that discriminatin of any kind is a great sin. In pre-school, children are taught housework and to make decorative objects from wood, to help their parents and other people who need help. There is no division between male and female work. Swedish society develops the work culture from the earliest childhood, along with a sense of belonging to social groups, as well as a sense of common purpose. The ruling thought in Sweden is that individual happiness is dependent on a happy society. Such an upbringing greatly facitlitates the subsequent establishment of a corporate culture, which, according to many management theorists, forms the basis of competitive advantage today.

Education has a special significance in Sweden, and the educational standards are exceptionally high. In fact, Swedish society is a learning society. Nine-year elementary education is compulsory, and pupils are graded only in the eighth and ninth grades, on which basis selection is made for enrollment in secondary school, which lasts three years. There are schools in Sweden for the education of people of all ages, and there is a large number of available instructors for educational and professional orientation, whether for pupils trying to choose their secondary school, immigrants trying to adapt to society or people seeking to change jobs. As Swedish society is a learning society, on-the-job education is encouraged in accordance with each individual's desires and personal abilities. Cooperation between universities and business is an established practice in Sweden, and the former are able to organize continued professional education in the companies themselves.

Swedish higher education produces managers who are able to apply their theoretical knowledge in practice, who "know how to think" and allow their workers to think. Swedish society is extremely advanced, technologically, educationally, and culturally – both in terms of culture of living and of organizational culture.

Another specific feature and another Swedish advantage compared to other countries is the existence of a unique right to spend time in nature and

enjoy its bounty. Swedes are highly devoted to the protection of animals and plants, and their ecological standards are extremely high.

Two key concepts are often used to describe the Swedish national character: egalitarianism and collectivism. The fact that no one, man-woman, old-young, is considered more valuable than anyone else is what allows the Swedish approach to life. Swedes have become accustomed to focusing on their own work, which they perform industriously, and to accepting responsibility, without boasting. Elitism, extravagance and a life that stands out even a little bit from the norm are still considered as taboos among many Swedes. On the other hand, an open spirit, an ability to totally enjoy celebrations and their free time, characterizes Swedes. It can be said that their ambition, devotion to work and controlled behavior during working hours stand in opposition to their wild, open spirit and great love for entertainment during free time.

One of the negative sides of the national character is the tendency toward excessive self-criticism, and there is an impression that Swedes are strong anti-nationalists. In essence, the tendency to self-criticism is an indication of a new national characteristic – perfectionism, i.e. the conviction that all that is imperfect can be made better, while everything that is good is taken for granted.

The Swedes are an exceptionally peacable people, guided by a "do not meddle" motto, evidenced by the historical fact that they have not been at war for almost two centuries. This characteristic is also reflected in their corporate culture, and the avoidance of conflict situations within company confines as well as in dealings with business partners.

3. Business contacts

If there is an intention to establish a good business relationship, it is necessary to respect certain rules. Namely, the appointed time is greatly respected in Sweden, which means that tardiness is not acceptable. A business meeting should be arranged two weeks in advance. In case of cancellation, it is necessary to inform the other party in advance, and provide a very convincing and justifiable explanation.

As Swedes consider their free time to be "sacred," it would be wrong to expect them to work during weekends, holidays or vacations. If they don't complete their task at work, it is customary for them to take it home. It is acceptable to call a Swedish colleague or business partner at home, if necessary, if he or she has given you their home number and permission to call, and if the problem cannot wait until the next morning.

When visiting and doing business in this country, it is necessary to be aware of the numerous holidays that are celebrated in Sweden. Many of the holidays are preceded by a pre-holiday day, which, as a rule, is counted as a

Saturday; on this day banks, offices and other institutions are closed, while stores close earlier than usual. Its owners determine a store's working hours; many stores that work Sundays are also open during holidays, but this is not the case with all of them. During summer vacation time, in July, many factories, workshops and some stores are closed, and some post offices have shortened working hours. Schools and universities are closed between June 10 and August 20. Thus, one should avoid setting business appointments for June, July and August, as well as for the end of February and beginning of March, as this is when Swedes most often go on vacation. When writing dates, they first write the day, followed by the month and the year. Working hours in Sweden start at 8:30 or 9:00 and last until 17:00, while the workweek lasts from Monday to Friday. There is little overtime work in Sweden, since Swedish labor laws allow only 5 overtime hours per week, and no more than 200 per year. The minimum vacation is five weeks long; most people take their vacations in July, and this should be kept in mind when planning to do business in Sweden.

A second precondition for a successful relationship is respect for their style of business attire. Business dress in Sweden is conservative. Suit and tie for men and suit or skirt for women are recommended for business meetings. Swedes are familiar with fashion and follow trends, but avoid displaying wealth through dress, i.e. the members of different social classes dress pretty similarly. It is possible for men to wear sneakers to work, although not of the overly sporting variety, while women do not practice this. Most restaurants, except the elite ones, do not require overly formal attire (coat and tie).

Due to the specific weather conditions in Sweden (winter can last up to nine months), employees are sometimes allowed to come to work in boots, and then to change into more comfortable office dress and shoes. Light colors at work are usually avoided; dark suits and silk ties are recommended for men and dark skirts and blouses for women, except during summer months.

One more guideline to establishing a good business relationship is to avoid asking Swedes about their family, job and personal matters in general during conversation. One should avoid trying to outshout the conversation partner and, as Swedes often make pauses in their conversation, one should not rush to interrupt them during them. It is desirable to avoid all comparisons, especially between local regions. Knowing the distinctions between Denmark, Finland, Norway and Sweden and their inhabitants is highly respected. They are particularly concerned about nature and like to talk about its protection. They do not like the use of superlatives or lies.

Desirable conversation themes are traveling, Swedish culture, hockey, Swedish history, current events, etc., while themes to be avoided are criticism of the Swedish authorities and economy, talk about salaries, family Olof Palme's assassination, personal history and anything having to do with status, position or wealth. Swedes like informality in conversation but not exces-

sive intimacy. Once a good relationship is established with a Swede, closeness grows with time. If one speaks any Swedish at all, one should freely use it. Still, if their business partners prefer or are better acquainted with English, Swedes will switch to the more universal language.

In certain cultures, gift giving is a part of the business style. In Sweden, gifts are not common between business associates. It is best not to make gifts, unless one receives a gift from a Swedish colleague. Although exchange of gifts is not desirable at the start of a business relationship, a practical gift should be made once the agreement has reached its end – for example, a book about one's own country or a small practical item for the office (appointment book, paper and pen holder, or similar).

What are acceptable or, better said, recommended are holiday greetings, especially before the major holidays, in the first place Christmas. Flowers are usually brought as a gift to the hostess of a private party one attends, but one should bear in mind that in Sweden chrysanthemums and white lilies are avoided as celebration gifts, as they are considered to be flowers more appropriate for funerals. Also, one should avoid orchids and red roses, as these are considered as signs of romantic intentions.

If a business partner invites one to dinner, it is best to bring wine, a liqueur, flowers, a cake and chocolate for the children to a Swedish home. Liqueurs are especially valued gifts, since they are pretty expensive. Especially respected are quality liqueurs or wines from the United States.

If one is staying with a Swedish family, it is recommended to give a high-quality product from one's own country upon parting, one that is difficult to obtain in Sweden. An excellent gift is food delicacies or a handbook about one's country or city. One should not give things that are readily available in Sweden. What one receives from Swedes is a small red wooden horse, a house decoration uniquely made in Sweden.

4. Cultural heritage and management

In each country, managers strive to develop their own style of planning, organization, motivation and employee control. With Swedes, a successful manager's important characteristics are the ability to organize teamwork and delegate power to those below. In comparison to other countries, teamwork is highly important. Good managers are those who successfully deal with processes such as team integration, consensus and cooperation. A second important characteristic is being worker-oriented, which is reflected in the choice of managers and in delegating responsibilities.

Successful managers should possess the ability to build, integrate and coordinate teams. Workers value a manager who can make them a part of

a team of players, in which the leader is more a coach than a referee. Generally speaking, a good manager is someone who will make it so that the team members give it their all for the achievement of a given goal and demonstrate a personal interest in the development of their own performance and career. If a team member gives his all, then the team will perform in the best possible way. Teamwork is seen as an integral part of cooperation. This is in contrast to other models of teamwork, where, for example, the task of the players is to help the manager give his all and lead the team to victory. In Sweden, two essential things appear in connection with the organization of teamwork: consultations and consensus in decision-making.

The consultation process begins with the providing of information about a certain project and a question about whether people wish to participate in it. A meeting is organized and the participants have a right to give their opinion or professional advice. The participants' opinions are usually heard, despite their place in the overall hierarchy. In this phase, the participants are considered to be equally competent to suggest good ideas, and they are viewed as equal members of the team. An individual's leadership function (or the person responsible for the team's result) is considered to be distinct in the context of the necessary distribution of tasks, and not on the basis of individual qualities. It is truly customary in Sweden for an individual's authority to be tied with their role or task within the organization. In other words, power is an "instrument." In other countries, for example Belgium or France, the director's authority is usually connected with his personality. Authority is personal, which in practice means that the director's opinon does not have to be decisive in an environment where power is seen as an "instrument." It will be heard (and perhaps even questioned) just like the opinions of others, which may become wrongly interpreted by members of a different culture as humiliating, especially among members of cultures where power is seen to be personal.

In the consultation phase, everyone brings out his or her opinion. This is in contrast to the second step, in which the team strives to reach consensus. The consultation phase can showcase many different opinions, but it is in the decision-making phase that a leader's cooperative skills come into play. The ideal leader should leave the impression with the other team members that everyone has received a hearing and that each opinion is being taken into consideration to the greatest possible extent. Swedish managers tend to be group and decision-oriented, making use of consultation that transforms into a conclusion. The decision is oriented toward reaching a consensus, which the team members are supposed to accept and apply regardless of the opinion they expressed in the course of consultation. Again, this may seem contradictory to people from other culturological contexts. How can people first express their opinion and then accept a compromise that does not correspond to what they had in mind? Actually, the consultation phase is the time when workers can

show their individuality. Once they express their opinion, they concentrate on the team and a consensus that can contribute to the team's advantage. One of the consequences of this is that conflicts are considered to be very negative. In some countries conflicts are considered to be unavoidable elements of the teamwork dynamic. They are seen as a way of strengthening leadership or simply as the moment of imposing loyalty upon the team members. In Sweden, conflicts in the course of teamwork are considered dangerous, especially in the decision-making phase. And, truly, if the members focus on the team instead of on individuals, conflict is a sign that the team is no longer functioning. This team dynamic significantly differs from the prevailing dynamic in most countries. In fact, in some countries, expressing an opinion contrary to that of the manager is unthinkable or insulting and, on the other hand, taking into consideration the opinions of employees is seen as the showing of weakness on the part of the manager. Among Swedes, changing or not respecting a decision brought in the name of the group is considered an insult to the consultative and consensual process.

In Sweden, the task of the management is to empower the employees. The broadening of competencies is carried out by the delegation of authority and responsibilities. It is believed that employees have capacities and potentials and those managers should value their initiative and accept their advice. The result of this in practice is that the decision-making process is more shared than it is in other countries. There is a notable tendency in countries such as Sweden, where authorizations are instrumental, that the employees seek greater authorizations and often get them. Interestingly, managers do not give direct orders to the employees but leave them, perhaps only superficially, a certain amount of freedom. On the other hand, in Germany and Switzerland authorization is carried out with a much greater degree of management influence. They delegate responsibilities but also exert strict control. In Sweden, good leaders are described as people who provide certain information without a directive. Order giving is considered insulting, as it presumes disrespect for another person's feelings of autonomy. This is contrary to the state of affairs in other environments, where giving orders means being efficient or represents proof of one's authority. In Sweden, frequent communications between employees and their direct supervisor is considered essential. The favorite way of communication is eye-to-eye meetings. These personal conversations allow employees to interpret their visions in their own way and to undertake independent actions. The goal of frequent communication is to develop mutual understanding and the personal interpretation of tasks on the part of the employee. Some experts say that the transfer of authority is a successful strategy, as respect is distributed and much is gained in terms of employee initiative.

Besides the said advantages of this style of management, certain limitations can also be observed, as noted by researchers and foreign observers in

Sweden. First, the importance given to teamwork unavoidably lengthens the process. Some people (probably less used to the consultative way of management, belonging to other cultures) describe the decision-making processes in Sweden as overly lengthy. They also think that groups are often inflexible.

Another criticism is focused on a certain kind of "group pressure." It has been observed that the calling of a meeting has important implications for the future decision. Workers will occasionally decide not to attend the meeting because they do not agree with a particular person's vision, only in order to avoid conflict. And, on the other hand, they may attend the meeting because they agree with the leader's basic postulates. On the basis of this interpretation there arises the question as to the true extent of diversity of thought in the consultation phase. It has also been observed that, during the consultation phase, the participants are expected to be fully supportive and to express constructive criticism and thinking, but also – if employees are opposed to a certain project – to occasionally leave the team or be thrown out of it. If homogeneity of thought is assumed within a given team, the delegation of authority has a different meaning, i.e. if the superiors know that the employees strive to have the same opinion as he/she, then the delegation of authority is no longer a sign of confidence but an excellent way of reducing one's own work load. Also, managers have a tendency of ensuring that the employees think the same as they by adapting a specific style of communication. It has already been noted that the favorite way of communication is eye-to-eye, which, again, can have a different connotation, as a way of developing a personal relationship. With the aid of personal ties, managers can wage greater influence upon their employees. Still, it is evident that this style of communication is efficient.

If the evaluation is that the above Swedish specificities such as teamwork and the distribution of authority make the Swedish concept of management unique, the Dutch might reply that processes marked by consultations and consensual decision-making are the current practice in Holland as well. Depending on where they come from, people will note other similarities or differences between the dominant concept in their own country and the one in Sweden. However, the management style in Sweden is unique to the extent that it combines various aspects according to a unique recipe. Consciousness about what makes a good manager can be of use when it comes to the question of their motivation, as well as to the question of how to understand their reactions to various styles. For example, in the course of a joint venture, a French and a Swedish company soon arrived to the question of cooperation and tensions appeared. The decision-making processes in these two companies were so different that they encountered a crisis at the very beginning of their joint business endeavor. Tension and disturbance among the Swedes appeared because the French directors were behaving "dictatorially," modifying without consultation decisions made in work groups. On the other hand, many of the

French were confused when they saw at the very start that the Swedes were "inflexible" and slow in making decisions. The French did not comprehend why so many people needed to be consulted and why the managers were not doing their work and making decisions. In this case, the French and the Swedish company did not take into consideration the fact that their management practices were fundamentally different. The Swedes could not understand that the organization of work in France includes processes that are less consultative but, in a certain sense, faster. The changes made by the directors to the decisions brought within a group were perhaps carried out on the basis of strategic information they possessed. The French employees were perhaps less included in decision-making than their Swedish colleagues, but their managers' arguments convinced them to make modifications. This did not happen with the Swedish employees, who felt more included in the teamwork and the decision-making. They thought that the changes were dictatorial and that the team wasn't being respected. Thus, they did not succeed in establishing communication founded in their organizational traditions. Also, the Swedes considered the lack of authority transfer on the part of the French directors as a lack of trust, although this was standard practice in France. Each team was starting from its own organizational culture and practice.

5. Conclusion

It is useful to have in mind that management practices vary from country to country, just like the concept of the good manager, i.e. the capabilities a successful manager should possess. Understanding this is the first step toward creative management. Differences in style do not have to be obstacles, they can be complementary. In the previously mentioned example, the French directors should have been included in the consultation phase. They should have had the opportunity to present what they thought were the strategic challenges connected to their task. In this way, a joint decision could have been reached. Management in a multiculturological situation perhaps depends on a successful combination of differences into something culturologically new. The project members can develop new references, having in mind the way in which things have to be accomplished, progressively complementing their own culturological framework with new elements that strengthen the basis for multicultural cooperation.

Hence, if one is making plans with Swedes, one has to make an extensive study of their rich culture, long history and national character in order to avoid misunderstandings. If a meeting is appointed, it is considered that half the work has already been done, because Swedes would rather cancel a meeting than allow a conflict to occur during it. The decision-making process

demands time, but Swedes are thorough, precise and clear. They will quite readily talk about their cultural heritage but will also be interested in the culture of their business partner.

References

- Ake Daun: *Swedish Mentality*, Penn State Press, 1996.
- Charlotte J. Dewitt: "Sweden – Culture Smart! A Quick Guide to Customs and Etiquette", Synopses & Reviews, 2007.
- Cvetković, N.: "Globalna slika stranih ulaganja sa osvrtom na zemlje u tranziciji," *Ekonomski vidici*, no.1-2, vol. VI, Beograd, June 2001, pp. 35-46.
- Cvetković, N.: "Menadžment – evolucija i tekući trendovi razvoja," *Ekonomski anali*, br. 41, godina III (part I), 1997, pp. 37-43; and no. 42, year III, (part II), 1997, pp. 46-49.
- Cvetkovski, T. – Cvetkovska-Ocokoljić, V.: *Poslovna komunikacija u savremenim uslovima poslovanja*, Megatrend University, Belgrade, 2007.
- Cvetkovski, T. – Langović, A. – Cvetkovska-Ocokoljić, V.: "Pregovaranje između projektnih timova – specifičnost interkulturnih pregovora," XI international symposium on project management, "Projektno upravljanje organizacijama – novi pristupi," YUPMA 2007, Zlatibor 15-17. May 2007, Udruženje za upravljanje projektima Srbije i Crne Gore, 2007.
- Cvetkovski, T. – Langović, A.: "Komunikacija u uslovima globalnog poslovanja," International Academic Conference "Radikalne promene u preduzećima i privredi u uslovima globalizacije," Megatrend University of Applied Sciences, Belgrade, 2003.
- Cvetkovski, T. – Langović-Milićević, A. – Cvetkovska-Ocokoljić, V.: "Uzroci i posledice loše komunikacije u projektu," XI international symposium on project management, "Projektno upravljanje organizacijama – novi pristupi," YUPMA 2007, Zlatibor 15-17. May 2007, Udruženje za upravljanje projektima Srbije i Crne, 2007.
- Hofstede, G.: *Culture and Organizations: Software of the mind*, McGraw-Hill, London, 1991.
- Jovanović, M. – Langović-Milićević, A.: *Interkulturni izazovi globalizacije*, Megatrend University, Belgrade, 2006.
- Jovanović, M.: *Interkulturni menadžment*, Megatrend University, Beograd, 2006.
- Langović, A. – Cvetkovski, T. – Langović, Z.: "Uticaj kulture – realizacija projekta," XI international symposium on project management, "Projektno upravljanje organizacijama – novi pristupi," YUPMA 2007, Zlatibor 15-17. May 2007, Udruženje za upravljanje projektima Srbije i Crne Gore, 2007.

- Langović, A. – Cvetkovski, T.: "Tajna uspešne saradnje," SymOrg 2006, X international symposium "Promene u organizaciji i menadžmentu – Izazovi evropskih integracija," Proceedings, Zlatibor 7-10. June 2006, Faculty of Organizational Sciences, Belgrade, 2006.
- Langović, A. – Cvetkovski, T.: "Uspešno pregovaranje u međunarodnom poslovanju," International Academic Conference, Megatrend University, Belgrade, December 2006.
- Langović, A. – Cvetkovski, T.: "Uticaj kulture na pregovaranja," IX international symposium "Menadžment – ključni faktori uspeha," Proceedings, Symorg 2004, Zlatibor, Faculty of Organizational Sciences, Belgrade, 2004.
- Langović, A. – Cvetkovski, T.: "Uticaj kulture na strategijski menadžment," VI international symposium on project management, "Primena projekt menadžmenta u reformama tedencije," Proceedings, YUPMA 2002, Zlatibor, Jugoslovensko udruženje za upravljanje projektima, Belgrade, 2002.
- Langović, A. – Cvetkovski, T.: "Uticaj kulture na vođenje projekata," V international symposium on project management, "Project management – nove tedencije," Proceedings, YUPMA 2001, Zlatibor, Jugoslovensko udruženje za upravljanje projektima, Belgrade, 2001.

Book review
Received: September 13th, 2007

Professor Mlađen Kovačević, PhD
Economics Faculty, Belgrade University

THE ASIAN "ECONOMIC MIRACLE"

An essay on the book "The Economy of Asia"
by Professor Blagoje Babić, PhD, Megatrend University, Belgrade, 2007

I

Asia's social and, particularly, economic development, especially that of the continent's largest countries, has been the subject of much research and various academic disciplines throughout the world. Due to various circumstances, little has been written about Asia's development during the last two decades in Serbia. Thanks to Prof. Blagoje Babić's several decades of work, as well as to Megatrend University, this void has been completely filled with the monograph "The Economy of Asia." This voluminous work is the only complete monographic work on Asia's economy in Serbia. It is the first book in the Serbian language that contains an all-encompassing insight into the economic development strategy, economic policies and economic achievements of Asia's most important countries.

The study of social and, especially, economic issues is of huge importance. Until recently, Asia was a typical underdeveloped region. At the start of the 21st century, however, it is poised to once again become the center of the world's economy, as was the case until the beginning of the Middle Ages. Economic history knows of no other region that maintained such high rates of economic growth over such a long time interval. At the same time, in all the economically successful countries of the Far East, the highly dynamic economic growth was accompanied by a just distribution of national income. For all these reasons, the World Bank has referred to these countries as "economic miracles."

There is a high degree of consensus among top researchers throughout the world that the 21st century will be marked by the impressive development of Asia and the continent's most important countries.

It was Japan that led the way for the Asian "flock of birds," with the Republic of Korea, Taiwan, Hong Kong and Singapore "taking off" in its path. This was, in turn, followed by the rapid ascent of a number of other countries

– Thailand, Malaysia, Indonesia and Vietnam. At the end of the last and the beginning of this century, it was India's economic growth that took the lead among all the other Asian countries, with the exception of China.

It is understandable that most of the world's attention has been focused on the rise of the three most powerful Asian economies – the Japanese, the Chinese, and the Indian. These three Asian giants have taken different paths to the same goal – rapid economic development. Thus, it is natural that the book has made a comparative study and detailed analysis of the three paths of development in Asia's three most important countries, which have resulted in three "economic miracles": the Japanese, the Chinese and the Indian.

II

The book consists of two parts – a general and a specialized part. The general part carries the title "The Economy of Asia – Main Features," while the specialized part consists of "The Economy of India," "The Economy of China" and "The Economy of Japan."

A – The general part starts with a definition of the idea of Asia, which isn't as simple as it seems. The chapter "The Psychological Relationship with Asia," in which the author explains the origins and the negative effects of our own negative perception of Asia, follows this.

The many years of research spent by the author, a lone figure in domestic economic scholarship, in seeking to empirically and theoretically explain the stagnation of Asia's economy between the mid Middle Ages and the end of World War II are summarized in the chapter "The Political Economy of Stagnation."

The chapter "The Asian Economies' Paths of Development" lays out the most important factors in the ascendance of Asian economies after World War II. This part of the book is rounded out with chapters entitled "The Role of Asia in International Economic Relations," "Competitiveness in the Strategies of Development of India and China," and "Causes and Consequences of the Asian Financial Crisis."

B – The "Specialized Part," shaped in the form of integral wholes – "The Economy of India," "The Economy of China," and "The Economy of Japan," analyses the economic development paths of Asia's three largest countries. Of special interest to the reader will be segments devoted to the following three issues: Asian economic thought, differences in paths of socio-economic transformation and similarities in strategies of economic development.

a) In analyzing Asia's development, the author always gives precedence to the views and findings of Asian analysts, assuming that they are in the best position to understand the civilizational peculiarities of the world from which they come.

In the parts devoted to India and China, the initial chapters are devoted to the history of economic thought in these two ancient civilizations. The author points out the Euro-centric misconception demonstrated in the claim that the history of economic thought begins with ancient Greece. The fact is, however, that autochthonous economic thought developed in Asia several thousand years ago. Many economic ideas that have become a lasting legacy of the world's economic thought originated in China and India. The author's achievement lies in the fact that he is the first among us to provide an overview of the economic thought of these two old civilizations.

The author of this monograph also provides an insight into the development of Japanese economic thought, which was especially significant in the second part of the 19th century and during the 20th century. He divides Japanese economists into: "romantics," who explain Japan's economic ascendance by the uniqueness of Japanese culture; "realists," who attempt to explain it by way of standard economic analysis; and "synthesists," who take into account both the cultural uniquenesses of Japanese society and standard economic analysis.

In short, with this book, we have gained, for the first time in our economic literature, an all-encompassing insight into the development of economic thought in Asia.

Especially interesting is the author's explanation of the role of Confucianism in China's economic development. Many authors assume that this teaching lies at the foundation of the "economic miracles" of all the countries of the Confucian tradition. On the other hand, some well-known Chinese authors, including the most prominent historian of the Chinese economy, Hu Jichuang, claim that, in fact, Confucianism, along with Taoism and Buddhism, was the basic cause of the Chinese economy's slow development through the centuries. The author concludes by saying that both sides err in excluding one another.

Confucianism was a teaching devoted to the preservation of a social order based on pre-industrial technology, i.e. the so-called agrarian society. Confucius' notions about industriousness, order, modesty, thrift and learning served the preservation of "agrarian society." When the social order was changed, it turned out that Confucianism had prepared Eastern societies for industrial work: modesty, thrift, order, work and learning are the basis of the rapid industrialization of all the countries of the Far East.

According to the author's opinion, it was Singapore economist Lim Chong Yah who explained Asia's development in the most picturesque way, with his EGOIN (Enterprise, Government, Ordinary workers, Investments, Natural resources) theory. According to him, the key to the Asian economies' success lies in their use of the main factors of development – entrepreneurship, the state, ordinary labor, investments and natural resources. The least successful countries in this regard he calls "turtles," the rapidly developing ones are referred to as "horses," while the "mature" economies are "elephants."

b) Three economically highly successful Asian societies – Japan, China and India – have different historical experiences. Japan was never a colony. China had a semi-colonial status, while India was a colony. Japan entered the new era with a capitalist economy based on a "guided market" and private property. China began its modern development as a socialist society with a centrally planned economy and collective property. Upon gaining independence, India established a mixed economy by linking planning and markets, and private and state property.

c) All three of these large countries accepted and carried out similar strategies of economic development. Their leaderships started from the premise that they could defeat the chief social ill – unemployment – only through industrialization. All three began their rapid post-World War II development with a strategy of "forced development" supplemented by a strategy of "import substitution." With time, all three abandoned this strategy, which favored heavy industry, was highly capital intensive and demanded a relatively small labor force, while producing a low profit rate. They changed to a strategy of using so-called comparative advantages – the development of light industry, which employed a large, cheap labor force, required low capital and had a high profit rate. This is where the Asian countries achieved high competitiveness and large accumulations of capital, which enabled them to subsequently return to the development of heavy industry.

The author emphasizes Nehru's view that the most important characteristic of Easter Asia is its "openness of spirit," in which all forms of development in all types of political systems can succeed.

The author of this book correctly points out that Japan underwent a crisis at the start of 1990, from which it is gradually recovering, while China and India are achieving consistently high growth rates. Some writers figuratively compare India with an elephant and China with a dragon, emphasizing that the elephant moves slowly but surely, while the dragon moves vigorously, and takes more risks. The dragon wins the race in the short term, while it's a question as to who will win the marathon.

III

Thanks to the author's broad learning and his decades-long study and consultation of a huge amount of literature in a number of languages, the monograph "The Economy of Asia" is, without doubt, the highest quality work in this field written in the Serbian language. In addition, the book is written in a nice, clear style.

The text is appropriately supported by statistical overviews and graphs. The author used the latest literature in several languages. Thanks to this

monograph, which is, without doubt, more than welcome, future researchers of Asia's economic and social development will have at their disposal an extensive inventory of sources, which will greatly aid their research.

The monograph can be useful to all those interested in Asia's economy – students preparing for exams, masters' theses and doctoral dissertations, the carriers of economic policy and businessmen working with Asian economies.

Under conditions in which neo-liberalism is playing the dominant role in the economic science of developed countries, most transition countries and developing countries, including Serbia, it is of great significance that this monograph convincingly demonstrates that a series of Asian countries that did not accept this economic philosophy have achieved impressive economic performances and have laid the basis for continued long-term dynamic economic growth and development, with a substantially more just distribution of the national wealth than is the case with a great number of countries that have accepted and faithfully carried out the neo-liberal concept.

With the publishing of this work, Megatrend University can proudly claim farsightedness and justifiability in introducing the detailed study of geo-economics as a discipline, and especially the subject of the Economy of Asia.

Book review
Received: June 26th, 2007

Assistant Professor Svetlana Vukotić, PhD
Union University, Belgrade

THE MANAGEMENT OF ENTERPRISE DEVELOPMENT

An essay on the book "The Management of Enterprise Development" by Professor Gordana Komazec, PhD and Professor Sonja Petrović-Lazarević, PhD Megatrend University, Belgrade, 2007

In their book, "The Management of Enterprise Development," Professors Gordana Komazec and Sonja Petrović-Lazarević have introduced us to an inspiring, inexhaustible and progressive topic.

The challenges that contemporary trends put before researchers – to gain an insight into the essence of *companies that create knowledge and organizational systems that teach*, which are the modern parameters of enterprise development – have also inspired this tandem of authors. This topic requires a multidisciplinary approach, especially having in mind the complex nature of things such as management principles, strategic management, globalization trends, information technologies, innovational processes, corporate culture, the layered approach to building competitive advantages, and the process of focusing on people and the management of human resources. In any case, along these lines, the authors themselves confirm this in their introduction, with the following statement: *in order to successfully manage an enterprise's development, one must follow the trends of globalization, apply modern information-communication technologies and discover specific, unique development programs that will enable the offering of products different from those offered by other producers.* All these enumerated points form a part of an integral whole entitled *The Management of Enterprise Development*, written by the "sure hand" of Professors G. Komazec and S. Petrović-Lazarević.

Besides the fluid and well-presented essence, the graphic and pictorial applications additionally contribute to the book's clearness and documentation, and bring it closer to the reader, making it at the same time studious due to the relevance of the topic and also quite *readable* and attractive due to its luxurious illustration.

The review of the book was signed by two respected names: Prof. Zdenka Djurić and Prof. Kazimir Kurij.

The Management of Enterprise Development is thematically divided into five chapters.

In the first part, development is broadly treated as a phenomenon of natural and social systems. In addition to examining the *ambient dispersion* of development, a retrospective of its evolution is also given. The distinction between growth and development is also considered in detail, since these concepts are often interchanged, as the authors underline, unless approached in a systematic way. This chapter is made up of three consistent wholes: the first, which deals with evolution; the second, which deals with enterprise development; and the third, which gives priority to enterprise management. In this, the first chapter, the determinants of development are presented generally, which is in accordance with the methodology of gradually leading the reader into uncharted categories.

In the second chapter, entitled *Significant Factors of Enterprise Development*, again by way of three sub-chapters, the authors lead the reader down the road of an enterprise's development. The first determining factor is the environment, in which context the unavoidable globalization trends are also considered. The next factors, bound together within an integral part entitled *Science and Technology as Factors of Enterprise Development*, serve to confirm yet another modern tendency implied by technical-technological development. Here it is important to point out the emphasis placed on cooperation between scientific research organizations, universities and companies, which is a topic that seems to have been insufficiently researched with us. Again, as in the first chapter, this chapter ends with research and development within the process of enterprise development.

The third chapter, *The Enterprise and its Environment*, is richly nuanced into seven segments, and the entire approach in this portion is based on the principles of strategic management, which is founded in long-term actions taken by a given enterprise in order to achieve superior performance. The starting premises are based on the enterprise's mission as the reason for the enterprise's existence. After providing a conceptual definition of the mission, the authors turn to the actors of strategic management, the so-called stakeholders. Strategies and organizational structure are the next steps along the way to the strategic management process, manifested in planning and the intended strategy. The choice of strategy is considered among four different types: functional strategy, business strategy, corporate strategy and global strategy. The seventh

part of the third chapter deals with organizational changes, first outlining the types of organizational changes and then the phases in the process of change.

Development Programs is both the title and the content of the next chapter, differentiated into four sectors: *Development Programs in the Value Chain, Innovations, Innovational Processes* and *Re-Engineering*. Innovation is examined in detail, not only from the theoretical aspect but also from the standpoint of its implementation into development programs. Significance is given to innovation both as a category and as an innovational process. The concept of Total Quality Management (TQM), is also given due attention. The authors especially focus on re-engineering as a specific feature in the domain of development programs.

The fifth chapter, *The Management of Enterprise Development*, defines the enterprise as an adaptive organizational system and places an accent on the process of managing the enterprise's development. This last chapter delves into the very process of managing enterprise development, which, being complimentary to strategic management, is made up of the following phases: the planning of development programs, choosing programs, organizing the execution of programs and program monitoring and control.

The authors also provide an explanation of why they used the term "enterprise," since this term is associated with the concept of undertaking action in the Serbian language. They have justified themselves by their subtle exposition of this heterogeneous subject matter, whose bounded dynamics may be sublimated in the following sentence: *Enterprise development is like a wagon wheel, which force is constantly pushing and moving into the distance, even as it perpetually revolves around its won axis and spins.*

The book will be useful for academic circles researching the field of managing company development, graduate students and students of business.

Book review
Received: September 12th, 2007

Ana Jovancai, M.A.
Geoeconomics Faculty, Megatrend University, Belgrade

THE ECONOMY OF SERBIA PERCEPTION, LESSONS AND PROSPECTS

Essay on the book "The Economy of Serbia," by Professor Dragana Gnjatović, PhD, Megatrend University, Belgrade, 2007

In this period of accelerating globalization, as Serbia seeks further integrations and attempts to overcome various crises, it seems that there is always a missing link. The young are interested in living abroad, in becoming acquainted with new cultures, exotic lands and global powers. However, the question that inevitably comes up is how much do they actually know about the overall economical reality of their own country?

Prof. Dragana Gnjatović's book, "The Economy of Serbia," has arrived at just the right time, as one of the first successful, well-grounded attempts to present and analyze this topic on the Serbian academic scene. The book is primarily intended for students of the economic sciences, but also for all others who wish to become better acquainted with the national economy. Very few works of scholarship have taken on the subject of examining the country's overall economic reality and have studied the conditions, states and changes that have occurred during Serbia's long economic history in such an all-encompassing and many-faceted way. In her book prof. Gnjatović examines the economic system, the economic structure and the growth of Serbia's national industry, emphasizing the specifics of its economic development. This volume successfully traces the development of economic systems from the time of the first civilizations and quite skillfully and knowledgably intertwines historical and economic development.

This book of a high print quality is 307 pages long, richly illustrated, and contains almost 30 tables, which contribute to an easier understanding of the subject matter. The book's contents are organized into sixteen chapters. The bibliography contains a list of more than 100 works.

The first chapter introduces the reader to the study of Serbian economic history and acquaints him with the scientific methods of studying the national

economy. "The basic scientific methods of studying the national economy have been posited by the postulates of four schools of economic thought: the historical school, Marxist economic thought, the theory of comparative economic systems and contemporary macroeconomic analysis."

The second chapter quite clearly and concisely considers the essential characteristics of pre-industrial economic systems by way of providing answers to four key questions:

1) Which form of property over the means of production is the dominant one?
2) Who owns the means of production?
3) Who and in what way appropriates the results of economic activity and in what form?
4) Who and in what way appropriates the results of economic activity and means of production?

The author introduces us to the definitions of the basic economic categories necessary for the study of economic history.

The third chapter summarizes in a quite concise way the period between the Roman conquest of the Balkans in the 3rd century BC and the end of the period of Turkish feudalism, which had basic characteristics of its own.

Next comes the fourth chapter and the most controversial of all economic systems – capitalism, which begins its development in the 18th century, "when the industrial, machine or factory system was introduced into production for the first time." The author also draws our attention to the "foundation of the capitalist system" and a very important moment in economic history – the onset of banking and stock exchange institutions.

Thanks to great technical-technological changes, this period brought great aggregations of capital and the creation of "companies of dimensions unprecedented in history" – or, more precisely, the onset of monopolies.

In the fifth chapter the author deals with the appearance of capitalism in Serbia. Following its final liberation from the Turks, Serbia also liberated itself from the feudalist economic system. The Serbian peasant became the owner of the land he cultivated, but also a taxpayer "on whose back rested the tax revenue strength of the Serbian state."

As in other European countries, market institutions were also formed in Serbia: the Privileged National Bank of the Kingdom of Serbia (1884) and the Belgrade Stock Exchange (1894).

The author then proceeds to acquaint us with Serbia's trading relations with Austria-Hungary, which were – due to the fact that 80% of Serbia's exports went to Austria-Hungary – of exceptional importance during this period.

The sixth, seventh and eighth chapters cover the period form the end of World War I to the end of World War II.

With the end of World War I in 1918 came the formation of the Kingdom of Serbs, Croats and Slovenes. The main priority was to create the conditions for a normal functioning of the economy. This period was marked by another agrarian reform, the unification of the tax systems and currencies and the further expansion of stock exchange activity.

The great global economic crisis did not circumvent the Kingdom of Yugoslavia either. The author first acquaints us in detail with the causes of the crisis, its influence on Serbia's economy, and the dinar fluctuations during the crisis and the efforts of the state to solve the newly appeared problems. This period also saw a significant fall in foreign currency inflows to the current account of the balance of payments of the Kingdom of Yugoslavia, which meant that, on the eve of World War II Serbia, was an underdeveloped agricultural land.

Serbia shifted to a wartime economy; production became subordinated to the wartime needs of the German Reich, while the currency and capital exchanges were blocked. The reader is acquainted in great detail with the distribution of the state's property, the withdrawal of the Kingdom of Yugoslavia's dinar from circulation and the reduced workload of the National Bank in emigration.

In the ninth chapter we are acquainted with the economic system of state socialism developed in Russia after the October Revolution in 1917. One of its basic characteristics is the change of ownership over the means of production and the appearance of state property as the dominant form of property (as high as 90% in the U.S.S.R.). Economic activities were organized according to plans, and the fundamental task that was set down was the industrialization of the economy. The author acquaints us with the appearance of unique phenomena, such as: *planned profit, non-refundable financing and the policy of planned losses.*

Soon after the war, capitalism in Yugoslavia was abolished and a new system, self-managing socialism, appeared in its place in the 1950s. This is the topic of chapters ten and eleven.

After World War II, the "battle against domestic and foreign private property" spread to Yugoslavia as well. Nationalization, agrarian reform and confiscation drastically changed the property ownership balance in favor of the state. In this chapter the author also deals with one of the most significant watersheds in the history of Yugoslavia, its refusal of U.S.-funded aid, i.e. the so-called Marshall Plan, and its definite alignment with the Eastern Bloc countries.

A new economic system of self-managing socialism was established in Serbia. In place of the state, the proclaimed sovereign became "the working people," who wielded the political power and governed social and state affairs via self-management organs, going upwards from the level of company to the level of municipality, republic and the federation. Workers' councils were introduced in the companies while workers' social and managing rights gained pri-

macy over political and civil rights, which were treated as leftovers from the bourgeois period.

The twelfth and the thirteenth chapters follow the transformation of socialist economies into mixed market economies. In the twelfth chapter we are acquainted with the mixed market economy. "The system of a mixed market economy allows the existence of both a private and a state sector of property, companies under varying types of ownership conduct business in the market, while both public and private institutions participate in the control of economic flows."

The transition economy was nothing more than a change from a centrally planned to a market economy. Advocates of gradualism as well as radicalism appeared. In this chapter, the author puts the accent on the privatization of the state sector of the economy and describes three different privatization methods: internal privatization, external privatization and voucher privatization.

In the fourteenth chapter we are acquainted with the efforts of FR Yugoslavia (i.e. the successor to the former SFR Yugoslavia, consisting of Serbia and Montenegro) to build a new economic system entailing: change in property relations, change in the form of company management and the development of an integrated market. The central bank received institutional independence and conditions were created for the revival of stock exchange activity. Social property was transformed in two phases.

The next, fifteenth chapter deals with the crises during the 1990s. This, the darkest period of newer Serbian economic history, whose effects are still felt even today, began in 1990. One by one republic began to exit Yugoslavia's formerly unified financial system: Slovenia, Croatia, Macedonia and, finally, Bosnia and Herzegovina, ending their obligations in the economic, financial and monetary spheres. After these events Serbia was faced with unprecedented macroeconomic crises, including a hyperinflation – about which the author writes separately – that became included in economic textbooks throughout the world as a negative example in the extreme. The author then goes on to analyze the introduction and results of the Program of the Monetary Reconstruction and Economic Recovery of Yugoslavia.

Just how hopeless the situation truly was is evidenced by the fact that Serbia actually qualified for international economic aid. As a result of the NATO intervention Serbia sustained damages of nearly 30 billion dollars and qualified for the first time for loans earmarked as economic aid for the world's least developed countries in 1999.

Finally, in the last chapter, the author lays out in detail the latest macroeconomic trends in our country. The first six years of the 21st century have seen progress in terms of growth of GNP and real wages, and the reduction of inflation. The author also reviews the privatization models, which have secured significant budget revenues. A number of reform laws designed to introduce

order into the fiscal, monetary and banking systems have been passed, and the foundations for an accelerated process of joining the European Union have been laid. Serbia has entered the Stabilization and Association process within which it is preparing itself to fulfill high European standards in the functioning of a market economy, the functioning of a legal state, democratization and human rights, as the key conditions for full-fledged membership.

This chapter also gives a clear overview of macroeconomic trends in the period of 2001-2006, providing the reader with the latest data and predictions.

The last six decades of the 20th century have certainly been among the most turbulent in the last two hundred years of Serbian history. We have lived through one world war and four internal ones. Out of these sixty years, a full fifteen were spent in war. Serbia entered World War II as an integral part of monarchist Yugoslavia and came out of it as one of six republics forming the Yugoslav socialist community. Together with the other five republics it shared all the advantages enjoyed by Yugoslavia throughout the world, only to become the subject of scorn in that same world after the country broke up. In spite of all this, Serbia is courageously facing the legacy of the preceding period, characterized, as we can conclude, by a backward economic structure and the non-market behavior of economic participants – all a consequence of the decades-long self-managing economic system. On top of everything came a harsh economic and social crisis, which was reflected by a drastic fall in the gross national product (GNP) and the standard of living during the 1990s, amidst the disintegration of the former SFRY, international sanctions, wars and numerous mistakes in economic and overall policy.

Thus, today, in order to help Serbia find the right path and successfully overcome its problems, it is unusually important for us to have a picture of the overall economic history. It seems that, in our strategies, plans, proposals and desires to learn as much as possible about the successful economic policies of developed countries – we have forgotten our own. It's logical to seek examples among those more successful than us, but who today can truly be called an economist who has not, in the course of his academic career, thoroughly examined all the economic periods through which his country has gone!

This book represents an exceptional contribution to scholarship; it is very clearly written, leading us through the centuries in a manner reminiscent of a historical novel, and acquainting us with the development of the economy, the founding of market institutions, the building of the infrastructure and the monetary system. It is, thus, intended for all readers interested in understanding the functioning of a highly specific economic system, encouraging them to ask themselves when Serbia will at last learn something from its own mistakes.

Book review
Received: September 21st, 2007

Jana Kovačević
Faculty for Business Studies, Megatrend University, Belgrade

MODERN STRATEGIC MANAGEMENT

An essay on the book "Modern Strategic Management" by Professor Momčilo Milisavljević, PhD, Megatrend University, 2007

The greatest danger in turbulent times isn't the turbulence, but conducting business according to yesterday's logic

Five years after the appearance of the first edition, Megatrend University has published the sixth amended and expanded edition of Prof. Dr. Momčilo Milisavljević's book "Modern Strategic Management."

Strategic management is a modern concept that came about in the process of the development of management systems. It is important for companies doing business under the conditions of globalization and which are faced with continuous changes in their own business environment. An understanding of modern strategic management should help managers in dealing with new non-structured problems in their business and in successfully solving them. The purpose of this work is to improve the strategic thinking, decision-making and actions of present and future managers. The book is intended for professionals, both future and present. At the Faculty for Business Studies in Belgrade it is used as the basic textbook for students of management striving to become top experts in their field. At the same time, the book is of special significance both for the academic and the professional public.

The work makes a strict selection of material that is a basic contribution to theory and, at the same time, relevant to business practice. Changes were made in the systematization of the material, so that the number of chapters has been reduced in relation to previous editions. The sixth edition has been adapted to standards prescribed by the Bologna Declaration in the domain of university teaching. Specific to this edition, the aims of each chapter are given at the start, while a summary along with key words and questions for discussion are given at the end. This provides for a simplified and modernized way of presentation that brings students closer to the material, helping them to better understand and master the modern concept of business.

The book is organized into four main parts. The first part deals with the dimensions of strategic management; the second explains strategic analysis, while the third deals with strategic choices. The fourth part is devoted to the study of the dimensions of strategic change and it has specific practical use.

The first part, consisting of four chapters, explains the concept of strategic management necessary to enable a company to face the dangers and possibilities of the global economy. The modern concept of strategic management contributes to the improvement of the system of planning – producing a flexible planning process applicable to companies in an environment of discontinuity. Strategic management is a continuous and interactive process by which a company's harmony with its environment is maintained. The concepts of company mission, vision and goals are then clarified. It is well known that for those who have no goals nothing is relevant! Strategy is explained as the basic means of achieving a company's goals. This is followed by an explanation of the process of its creation, of the hierarchy of strategies that exists in companies with a broader production program and, finally, of the concept of comparative advantage and the significance of its creation and preservation in the modern global economy.

The basic goal of strategic analysis is for company management to gain an understanding of the company's strategic position. This is the subject of the second part of the book, which is divided into four chapters. Strategic analysis focuses on the company's environment (possibilities and dangers), sources and capabilities, as well as its strengths and weaknesses. This is followed by an analysis of the competition and the main competitors from the company's strategic group. A definition is also given for the relatively new term of essential competence, used in explaining the basis for creating a company's competitive advantage. Strategic analysis is needed so that a company may gain a genuine understanding of events in the environment in which it's conducting its business activity, and so that it may objectively evaluate its sources and capabilities.

Strategic analysis forms the basis of strategic choice, to which the third part of this "volume of knowledge" is devoted, consisting of eight chapters. It is suggested that, in order to choose a rational strategic option, it is necessary to have a clear vision, which allows the choice and consideration of realistic alternative directions of action, and the best choice in given circumstances. Detailed explanations are given of three generic strategies, growth strategies and strategies of contraction and turnaround. Since competitive advantage is increasingly based on intangible assets that the competition cannot obtain or successfully imitate in the short term, attention is also devoted to the study of strategies of knowledge. As technological changes are one of the important factors contributing to strategic changes in both the national and the global economy, an explanation of the technological strategy is also important. Globalization is characterized by the increasing interdependence between national

economies and the global economy, which means that the study of international strategies is also essential.

The goal of strategic analysis and strategic choice is strategic change, and this is the subject of the fourth part, which consists of five chapters. If the application of the chosen strategic direction does not bring the desired strategic position, business activity cannot be evaluated positively. Most works in the field of strategic management are primarily concerned with strategic analysis and strategic choice. As the execution of strategy is the most critical component of strategic management, special attention is paid to strategic change – which is a significant characteristic of this book. Strategic analysis and strategic choice are only preparations for strategic change. This is one of the main messages of "Modern Strategic Management."

This work's unique approach allows us to gain an understanding of this scientific concept as well as of the best ways in which companies may accomplish their goals and conduct their business activities, especially under the conditions of globalization and accelerating technological changes.

We may conclude that this essential work dealing with the modern world of business represents a great contribution to both the academic and the professional community, and a significant achievement when it comes to literature in the field of strategic management.

The Library of Congress Catalog

Megatrend review: the international review of applied economics.

LC Control No.: 2007201331
Type of Material: Serial (Periodical)
Uniform Title: Megatrend Revija. English.
Main Title: Megatrend review : the international review of applied economics.
Published/Created: Belgrade : ill. ; 24 cm.
Description: v. : Megatrend University of Applied Sciences, [2004]-
Year 1, no. 1 ('04)-
ISSN: 1820-4570
CALL NUMBER: HB1 .M44

CIP - Katalogizacija u publikaciji
Narodna biblioteka Srbije, Beograd

33

MEGATREND review : the international review of applied economics / editor-in-chief Nataša Cvetković. - Year 1, 1 (2004) - Belgrade: Megatrend University of Applied Sciences, 2004 - (Belgrade: Megatrend University of Applied Sciences). - 24 cm

Dva puta godišnje. - Je izdanje na drugom jeziku: Megatrend revija = ISSN 1820-3159

ISSN 1820-4570 = Megatrend review
COBISS.SR-ID 119185164

"Megatrend Review" is registered as the scientific review of national interest (R 62), by settlment of the Board of Social sciences, Ministry of Science of Republic of Serbia, on May 29th 2007.