HUNGARIAN STATISTICAL REVIEW

JOURNAL OF THE
HUNGARIAN CENTRAL
STATISTICAL OFFICE

VOLUME 84 2006. SPECIAL NUMBER 10

CONTENTS

ISSN 0039 0690

Published by the Central Statistical Office
Editor in charge: dr. Péter Pukli
Executive editor: dr. László Hunyadi
Printed by the Akadémiai Nyomda
4555 – Martonvásár, 2006
Director in charge: Lajos Reisenleitner

Managing editor: Orsolya Dobokayné Szabó
Editors: Andrea Polyák
Technical editors: Éva Bartha, Ágnes Simonné Káli

Editorial office: Budapest II., Keleti Károly utca 5–7. Postal address: P.O.B. 51. Budapest, 1525.
Phone: (361)-345-6546, Telefax: (361)-345-6594,
Internet: www.ksh.hu/statszemle E-mail: statszemle@ksh..hu
Publishing office: Central Statistical Office, Budapest II., Keleti Károly utca 5–7.
Postal address: P.O.B. 51. Budapest, 1525. Phone: (361)-345-6000
The publication can be purchased at the Statistical Bookshop:
Budapest II. Keleti Károly utca 10. Phone: (361)-212-4348

Poverty in Hungary with special reference to child poverty

Ödön Éltető,
retired deputy head
of department of the HCSO
E-mail: odon.elteto@ksh.hu

Éva Havasi,
chief adviser of the HCSO
E-mail: eva.havasi@ksh.hu

Database of income surveys carried out by the HCSO on 25 percent subsamples of the 2 percent microcensuses in 1996 and 2005 with reference years 1995 and 2004 are outstandingly suitable to investigate the presence and the size of poverty in Hungarian society as well as its causing factors. In this study the authors present important results of their analyses concerning poverty issues in the years referred and the changes through the examined period. Child poverty and its underlying causes are discussed in some more details. In conclusion some basic characteristics of child poverty in Hungary are enlighted.

KEYWORDS: Social statistics.
Poverty and social deprivation.

A great advantage of combining of microcensuses and income surveys is based on the fact that to answer to the former was obligatory by law. Consequently, for all households and persons not responding to the income surveys we disposed plenty of data relevant in respect of the income situation (age, sex, marital status, family status, educational attainment, economic activity, occupation, economic branch, etc.) from the microcensus. This made it possible to *impute* their incomes either by hot deck or cold deck imputation techniques. In cases where detailed and reliable data were available either from macro statistics or from large scale sample surveys (e.g. the annual earning survey covering several hundred thousands employed earners) or from tax authorities, generally cold deck imputation techniques were applied using *microsimulation methods*. For other income sources hot deck imputation was applied by selecting randomly one of the responding households (or persons) having similar characteristics as the household (person) in question and imputing his/her/its corresponding income item to the non-respondent. Thus for all selected households of the income surveys we had income data irrespective of whether the household in question cooperated in the income survey or not. Correspondingly, information on the income distribution and poverty indicators are based on data of the whole subsamples of the microcensuses, i.e. on 18120 and 18880 households, respectively.

In addition to imputations *corrections* on certain income items were also made, because in many cases people tend to underreport or forget their incomes. In respect of such income items correction procedures were carried out, for which reliable macrostatistical or large scale sample survey data were available. Obviously, quite a lot of income items could be found, which were not corrected at all, the reported data were processed. More detailed descriptions of the imputation techniques and correction procedures can be found in *Kesztehelyiné* [2006a] and *Keszthelyiné* [2006b]. The following chapter introduces the basic principles of the concept and measuring of poverty.

1. Income poverty and poverty measures used

Poverty can be characterized by various aspects: income, consumption, housing conditions, earning possibilities, educational facilities, deprivation, etc. When primarily the income situation of households is considered, we focus on the *income* *poverty. I*n the following most of our analysis concentrates on income poverty, however, other aspects of poverty are also investigated.

In the study poverty rates and characteristics of the poor are investigated using two different poverty thresholds (k):

1. relative poverty (k_1), and
2. subjective poverty (k_2).

Relative poverty is the usual way of defining who can be considered as poor. In this study we define it as 60 percent of the median equalized income using the original OECD1 equivalence scale. In most of the analyses this threshold is used in the paper. In both income surveys households were asked not only to report their incomes, but also to estimate how much money a household similar to their own in size and composition would need to reach various levels of living standards. The lowest of these indicates conditions when a household hardly can make both ends meet. The weighted means of these amounts was considered as the threshold of *subjective poverty*.

Poverty rate (PR) is the proportion of households or persons living below a given poverty threshold. If n denotes the number of all households or persons and p the number of the poor, then PR in percent is

$$PR\% = \frac{p}{n} * 100 .$$

Poverty gap(R) is the average distance of the equalized incomes of poor households or persons from the poverty threshold. If $I(i)$ denotes the equalized income of the i^{th} poor household or person, then

$$R = \frac{\sum_{i=1}^{p}\left(k - I(i)\right)}{p} = k - \overline{x}_p$$

where $\overline{x}_p$ is the average income of the poor.

Income gap ratio (Rr) is the ratio of the poverty gap to the poverty threshold, i.e.

$$Rr = \frac{R}{k} = \frac{k - \overline{x}_p}{k} = 1 - \frac{\overline{x}_p}{k} .$$

Rr is generally given in percentage form. It measures the intensity of poverty. Its limits are between 0 and 1. The closer *Rr* is to 1, the deeper the poverty of those living below the poverty threshold k.

Poverty risk is defined as the ratio of proportion of a population subgroup within the poor to the proportion of this subgroup in the total population.

2. Poverty in Hungary in years 1995 and 2004

According to data of the respective income survey in 1995 10.4 percent of the households and 13.0 percent of the population could be considered as poor in Hungary, lived below the relative poverty threshold based on the OECD1 equivalence scale. The corresponding percentages were 12.7 percent and 13.9 percent, respectively, in 2004. It means a slight increase in poverty from 1995 to 2004. The rate of the poor is generally greater than that of poor households, because most poor households are of a larger size than the average household size. The respective household size figures were 3.27 and 2.61 in 1995 and 2.85 and 2.61 in 1995.

As households tend to overestimate the amount of money a household similar to their own would need to reach a low level of living standard, the poverty rates based on subjective thresholds are higher than those based on relative thresholds. The corresponding poverty rates were 27.0 percent for households and 31.4 percent for persons in 1995, while 23.0 percent and 24.5 percent in 2004.

The values of the income gap ratio – 27.5 percent in 1995 and 21.8 percent in 2004 – indicate that poverty, although significant, is not really deep in Hungary. Moreover, there was a remarkable decrease in this poverty indicator from 1995 to 2004. The lower decile of the income distribution was only by 9 percent smaller than the relative poverty line in 2004 and even the lower 5 percent quantile of the distribution almost reached its three-quarter in this year.

The size of the household is a good characteristic of poverty risk. For households of size greater than three the risk to become poor exceeds the average and in case of households with six or more members the risk is 2.2fold of the average. Among poor households the young ones represent a much greater proportion than the average as well as households where the head has low educational attainment. Further factors having relatively high poverty risk are on the one hand when there is no active earner in the household, and, on the other, when unemployed person(s) can be found among the members. Table 1. shows some more detailed figures.

There were considerable regional differences in the poverty rates in 2004. While only 7.3 percent of the households belonged to the poor in the Central Hungary region (including the capital), in the Northern Plain region the poverty rate exceeded 18 percent. The regions Southern Transdanubia, Northern Hungary and the Southern Plain can be considered as the poorer parts of the country with poverty rates of households 17.1, 16.2 and 15.3 percent. In Central and Western Transdanubia, on the other hand, only about one tenth of the households could be considered as poor in 2004. The size of the settlement also considerably influences poverty. In the smallest villages (with inhabitants less than 1000 persons) the poverty rate was nearly 22 percent, in 2004, and even in settlements with inhabitants between 1000 and 4999 persons it exceeded 15 percent. In the largest cities (with inhabitants more than 50 thousands) somewhat more than 9 percent of the households could be considered as poor in 2004, and even less, 6.5 percent in the capital. The respective risk indicators show similar features. If a household lived in the capital in 2004, its chance to be poor was only half of the country average and three-quarter of it in the large cities, but it amounted to 1.7 fold of the country average if they lived in one of the smallest villages.

Table 1

Poverty risk of poor households with various characteristics, 2004

Household characteristics	Relative	Subjective
	poor households	
1 member	1.15	1.15
5 members	1.55	1.48
6 or more members	2.27	2.20
With no child	0.82	0.85
1 child	1.11	1.08
2 children	1.45	1.34
3 or more children	2.66	2.39
No active earner	1.54	1.44
Three or more active earners	0.16	0.24
Unemployment is present	2.94	2.28
Educational attainment of the head is		
elementary school or lower	1.50	1.48
third degree	0.21	0.22
Age of the head is		
below 30 years	1.51	1.29
30-49 years	1.21	1.11
60-69 years	0.73	1.91
70 years or older	0.73	0.94

An outstanding number of the poor lives in dwellings of inferior quality than the average. While only 9 percent of all households shelter in dwellings with no amenities or in temporary accommodation, 23 percent of the poor households live in such dwellings. In 24 percent of the poor households the situation is made harder by the lack of indoor flushing toilet, while the rate of such dwellings among all dwellings is only 9 percent. In country total only 3 percent of the dwellings do not dispose running water within. Among poor households this rate is almost 10 percent.

To sum up the results of comparing the poverty measures obtained from the two income surveys we can conclude that not too considerable changes have occurred in the period investigated. Meanwhile other sources, e.g. data of the continuous HBSs indicate, that after 1995 the poverty rate – in parallel with a decrease in the real income of the households and an increase in the income inequality – increased to some extent, and then it has decreased again.

3. Child poverty

In poverty issues child poverty represents one of the most important and most challenging one, not only because children represent a particular population group, which is weak to influence their own economic condition or to find escape from poverty. Moreover, without the efforts and well-considered measures of governments and the society there is a high risk of poor children growing to poor adults. Attaining higher educational degrees than that of their parents is one of the most hopeful way for poor children to get out of the circle of poverty and deprivation.

Child poverty can be investigated simpliest by analysing the living conditions of households upbringing children. In this study we consider a household having a child (or children), when at least one dependent child under the age of 20 lives in the household as pupil, student or other dependant. As shown in Table 2, the poverty rate in households with children was considerably higher than the average both in 1995 and 2004.

Table 2

Poverty rates using relative and subjective poverty thresholds in households with children, 1995, 2004
(percent)

Poverty rate	Relative threshold		Subjective threshold	
	1995	2004	1995	2004
For households	17.0	18.3	37.8	30.9
For persons	18.1	19.6	39.5	32.7

As data indicate, there was a slight increase in the *relative poverty* of households with child (children) from 1995 to 2004. In 2004 the rate of relative poverty for persons living in households with children was 1.5 percentage points higher than in 1995. However, using the subjective poverty thresholds, we experienced a decrease instead indicating that the households in 2004 might have been more realistic in judging their monetary needs than in 1995. It is worth to mention that the poverty is somewhat less “deep” in poor households with children than in all poor households. The income gap ratio was 22.8 percent among the former group of the poor in 1995 as compared to 23.1 percent among all poor households, while 20.3 percent against 21.8 percent in 2004. The figures show at the same time a decrease in the income gap ratio from 1995 to 2004.

The more children a household has, the larger the probability is that the household will be poor. The relative poverty rates in the respective two years for persons living in households with different number of children are shown in Table 3.

Table 3

Number and rate of children living in poverty and relative poverty rates

Number of children in the household	Number and rate of poor children				Relative poverty rate (percent)	
	1995		2004		1995	2004
	persons	percent	persons	percent		
1 child	96 062	13.5	81 408	12.8	12.7	13.2
2 children	184 594	16.0	146 262	14.9	15.9	18.6
3 children	109 470	28.1	103 554	23.3	28.3	29.0
4 or more children	107 259	59.3	90 311	43.3	59.5	51.1
Together	497 385	20.4	421 535	18.6	18.1	19.6

Note that while in contrast with the general tendency the total number of children in households with 3, 4 or more children did not decrease but increased (by 14 and 15 percent) from 1995 to 2004. This was not characteristic to the poor children. The rate of poor children in these two groups of households with children considerably, decreased (by 11 and 15 percent) from 1995 to 2004. However, in spite of this decrease the rate of the poor remained very high, more than 23 and 43 percent, respectively, in these two groups of households with children even in 2004. The last two columns of Table 3 require some explanation. While the poverty rates in the first three groups of households with children increased to a less or more extent from 1995 to 2004, it markedly decreased in the fourth group, in households with four or more children (though it was still very high, more than 51 percent in 2004). We think that this positive change can mainly be attributed to the so called regular child-welfare assistance introduced in the period considered. Those households are entitled to apply for this assistance, where the per capita available income is less than the amount of the current minimal pension. As primarily households with four or more children meet this condition, the amount per children (22 percent of the current minimal pension) obtained in the form of this new assistance significantly improved the income situation of many of the households with four or more children in 2004.

Regional differences in the risk that a household with children becomes poor increased from 1995 to 2004, but the change was radical in two of the regions only. In county Pest the risk decreased from 1.1 to as low as 0.4. Meanwhile in the South Transdanubia region, it increased from 0.8 to 1.5. There were radical decreases in the poverty risk and poverty rate of households with children living in Budapest. The risk to become poor decreased in their case from 0.9 in 1995 to 0.5 in 2004. While in 1995 15 percent of the poor living in households with child (children) could be found in the capital, in 2004 only 7.6 percent. Similar to the general tendency, for households with children, the risk to become poor increased when the population size of a settlement is smaller. For instance in small villages (with population less than 1000 persons) the risk becoming poor was 1.6 in 2004 for persons living in households with children. As data of Table 4 suggest, the distribution of persons living in poor households according to the type of settlement is significantly different from that of the total population both in the case of all households and households with children.

Table 4

Distribution of persons living in all households and in households with children, 2004

Type and size of the settlement	All households		Households with children	
	the average	the poor	the average	the poor
Budapest	17.4	8.4	14.2	7.6
Large cities with more than 50 000 inhabitants	19.1	13.3	17.7	11.9
Settlements with 1 000-50 000 inhabitants	56.1	65.9	59.9	67.6
Small villages	7.3	12.5	8.2	13.0
Total	*100.0*	*100.0*	*100.0*	*100.0*

Since the middle of the 1990s there was a considerable improvement in the general housing conditions in Hungary. This refers also to the poor households, but many of the poor still have rather bad living conditions mostly with no amenities. One-quarter of the numbers of poor families with children lives in a dwelling where there is no in-door flushing toilet, 23 percent of them in dwellings with no amenities.

Table 5

Distribution of persons living in all and poor households with children according to their housing conditions

Characteristics of the dwelling	All	Poor	All	Poor
	households with children			
	1995		2004	
Without comfort	12.3	30.1	8.1	23.0
Block of flats	14.5	10.9	11.3	6.1
Houses without foundation	4.1	9.0	5.0	8.6
Absence of an in-door flushing toilet	12.4	31.3	8.1	24.1
Absence of running water	2.5	7.3	2.6	8.9
Other	54.2	11.4	64.9	29.3
Total	*100.0*	*100.0*	*100.0*	*100.0*

In addition to bad housing conditions poverty manifests itself also in such everyday situations when the household does not know how to make both ends meet. 50 percent of the relative poor and 44 percent of the subjective poor mentioned such problems in 2004. Another serious trouble for the poor is to pay the bills for electricity, gas, running water, sewage, district heating, etc. As can be seen from Table 6, households with children are in worse position in this respect, too.

Table 6

Proportion of all poor households and poor households with children mentioning financial worries and difficulties in paying bills

Type of the worry	Relative poor households	Subjective poor households	All households
Everyday financial worries	50.1	44.4	23.9
Difficulties in paying bills	44.4	37.5	18.9
	with children		
Everyday financial worries	56.4	50.1	29.6
Difficulties in paying bills	52.1	46.2	26.3

4. Characteristics of poor households with children

Whether we investigate the age structure of all the members of a household or the age of the head we can conclude that young households had the highest risk to become poor both in 1995 and 2004. In both years the poverty risk of persons living in households with children where the age of the head was less then 30 years exceeded double of the share of the population living in such households. The poverty risk decreases with the increase of the age of the household head.

Table 7

Poverty risk of persons living in households with children by the age of the head

Age of the head of the household	All households with children, 1995	Poor households with children, 1995	All households with children, 2004	Poor households with children, 2004
Below 30 years	1.3	2.3	1.3	2.2
30-49 years	1.6	1.5	2.1	1.9
50-59 years	0.5	0.3	0.5	0.3
60 and more years	0.1	0.1	0.1	0.1

The number of active earners in a household is an important factor influencing poverty. However, it seems that comparing to 1995 an even one more important factor which induces poverty is *unemploymen*t. If there is an unemployed in the household, then the poverty risk of the members is extremely high.

Table 8

Poverty risks according to the number of active earners and the presence of unemployment, 2004

Number of active earners and the presence of unemployment	All poor households	All households with children	Poor households with children
No active earner	1.9	0.4	1.5
1 active earner	1.2	1.3	1.5
2 active earners	0.3	1.3	0.3
3 and more active earners	0.2	0.7	0.2
No unemployed earner	0.7	0.9	0.7
Unemployment is present	3.7	1.7	4.2

The role of educational attainment in inducing poverty has also increased, especially when we consider the lowest and highest level of education. While in 1995 the difference in the poverty risks between the highest and lowest level of education was 2.8fold, in 2004 it exceeded to 8fold.

Table 9

Poverty risks according to educational attainment 1995, 2004

Level of educational attainment of the head of household	All poor households		All households with children		Poor households with children	
	1995	2004	1995	2004	1995	2004
Elementary school	1.3	1.7	0.7	0.7	1.1	1.7
Vocational school	1.2	1.0	1.3	1.3	1.2	1.2
Secondary school	0.7	0.5	1.2	1.0	0.9	0.5
Third level education	0.2	0.2	1.1	1.0	0.4	0.2

5. The role of social assistances

The most important social assistance for households with children is the family allowance. Almost all (98.4 percent in 2004) of households with children receive this benefit. However, while the per capita income of the population increased in nominal value almost to fourfold from 1995 to 2004, the value of the family allowance for a receiving household was only somewhat more than doubled (increased to 2.2fold) in this period. Available data indicate that family allowance is not an exception in this respect. None of the comparable social assistances preserved their real value in the period considered. In connection with the family allowance it must be noted, however, that a decrease in the number of children entitled to receive family allowance also contributed to the smaller increase of the nominal value of the family allowance.

Table 10

Dynamics of the rate of utilizing households and the value in case of certain social assistances 1995, 2004

Social assistance	Rate of utilization	The sum for a utilizing household
Family allowance	0.9	2.2
Orphan's allowance	1.1	3.2
Support on housing	2.4	2.6

As far as the sum of the various social assistances for a utilizing household is concerned the orphan's allowance with its relatively high sum provides an essential contribution to the living of the households concerned. Though the monthly sum between HUF5500 and HUF9100 of the family allowance per child – the amount depended on the number of children and on whether the family is a one-parent or two-parents family – which households with children received in 2004 meant also a considerable promotion to bring up their children. The sums of various social assistances and the rates of utilization are shown in Table 11.

Table 11

Rate of utilization of various social assistances and their amounts among households with children, 2004

Social assistance	Rate of utilization (percent)	Average monthly amount, HUF	
		for households with children	for utilizing households
Family allowance	98.4	11 909	12 106
Orphan's allowance	1.2	1 350	34 446
Regular allowances	7.3	2 687	11 033
Occasional allowances	1.8	236	3 914
Support on housing	2.3	238	3 044

If we consider the same data for *poor households* with children, the rate of utilization is reasonably higher in the case of regular and non-regular allowances as well as of support on housing. The monthly amounts are also remarkably larger in several cases, e.g. for family and regular allowances, but in other cases they are smaller.

Table 12

Rate of utilization of various social assistances and their amounts in poor households with children, 2004

Social assistance	Rate of utilization (percent)	Average monthly amount, HUF	
		for poor households with children	for receiving poor households
Family allowance	98.0	14 856	15 160
Orphan's allowance	3.1	794	25 975
Regular allowances	77.4	8 792	11 359
Occasional allowances	21.2	936	4 414
Support on housing	29.5	782	2 655

With the increase of the number of children the role of the family allowance in the living conditions of the household also increases. However, its amount does not reach one fifth of the income of the household even in case of five children. It means that family allowance does not cover even a significant portion of the costs of supporting the children, not even supposing a very modest provision.

6. The role of various factors in inducing poverty

Beside investigating the size of poverty in contemporary Hungary, the changes of its nature in the last decade and the characteristics of all poor households and poor households with children we made some researches to find out what are the really significant factors and household characteristics bringing about poverty. Moreover, our researches extended to quantify the importance of the various factors both in 1995 and in 2004. Naturally, the income position and living conditions of an individual household and its members are determined by lot of factors. Though we are convinced that by thorough research and with the knowledge of long experience it is possible to set up relevant models expressing the existing relationship between poverty and a number of explaining factors.

We applied the well known logistic regression model with the incidence of poverty as dependent variable.

At first step the following explanatory variables were considered:

– the number of dependant children under 20 years in the household,
– the number of active earners in the household,
– at least one member is unemployed within the household,
– the absence or presence of in-door flushing toilet.

In both years and in the cases of both all poor households and poor households with children the last two variables proved to have the largest explaining power. Table 13 shows the values of the exponential β and the pseudo R^2 of the models for both groups of the poor in 1995 and 2004.

Table 13

Parameters of the logistic models with four explanatory variables

Explaining variables of the model	Values of the parameters			
	all poor households		poor households with children	
	exponential β	pseudo R^2 (percent)	exponential β	pseudo R^2 (percent)
	1995			
Number of children under 20 years	1.89	19.4	1.59	20.5
Number of active earners	0.56		0.43	
At least one member is unemployed	2.55		1.97	
Absence of in-door flushing toilet	2.19		2.71	
	2004			
Number of children under 20 years	1.71	22.4	1.34	27.0
Number of active earners	0.37		0.25	
Number of unemployed members	3.50		2.35	
Absence of in-door flushing toilet	3.43		4.06	

Including the level of educational attainment of the household head into the explanatory variables (with the reference category: third level of education or more) the explanatory power of the model significantly increased. For the 1995 model three educational levels above the first level were differentiated, for the 2004 model five levels. It is quite interesting that the explanatory power of the educational attainment of the household head is remarkably strong if the head is a skilled worker or has secondary school attainment. The models have significantly higher explanatory power for 2004 than for 1995. The value of the pseudo R^2 is 30.8 percent for poor households with children in 2004 which can be considered as rather high. Table 14 shows more detailed data for 1995 and 2004.

Table 14

Parameters of the logistic regression models, five explanatory variables

Explanatory variables of the model	Values of the parameters			
	all poor households		poor households with children	
	exponential β	pseudo R^2 (percent)	exponential β	pseudo R^2 (percent)
	1995			
Number of children under 20 years	1.90		1.55	
Educational attainment of the head				
– elementary school	4,13	20.6	6.21	22.8
– vocational school	3.74		3.64	
– secondary school	2,80		2.89	
Number of active earners	0.60		0.47	
At least one member is unemployed	2.45		1.82	
Absence of in-door flushing toilet	1.82		1.94	
	2004			
Number of children under 20 years	1.71		1.27	
Educational attainment of the head				
– elementary school	1.57	24.7	4.70	30.8
– vocational school	1.86		1.99	
– specialized secondary school	1.71		1.15	
– secondary school	0.97		0.62	
– college	0.38		0.29	
– university	0.23		0.21	
Number of active earners	0.40		0.28	
At least one member is unemployed	3.13		1.94	
Absence of in-door flushing toilet	2.76		2.46	

Note. For the educational attainment of the head in 1995 reference category is the third level of education, while in 2004 the reference category is the Phd degree.

Note that the inclusion of the size category of the settlement into the explanatory variables does not improve noticeably the explaining power of the models, probably because this last variable is in very close correlation with the remaining ones.

7. Conclusions

The paper presents some important findings on poverty and especially on child poverty in 2004 and in the middle of the 1990s in Hungary. The data originate from two income surveys covering 0.5 percent of the private households. The analysis is based mainly on the notion of *relative poverty*, where the threshold is defined as 60 percent of the equalized median income.

There was a slight increase in the poverty rate from 1995 to 2004. The poverty for people living in *households with children* was considerably higher in both years in question. Having children in the household represents one of the primary sources of poverty in contemporary Hungary. The risk to become poor is rather high also for households where *unemployed person(s*) can be found among the members. A *low educational attainment* can also considerably contribute to poverty. Those who live in *small villages* have larger probability to become poor than those living in towns or in the capital. The various factors are correlated and influence poverty simultaneously.

The individual and common impacts of various explanatory factors on poverty are investigated applying logistic regression models. In 2004 the five factors considered explained almost 31 percent of the variations in poverty among household with children. A majority of poor households with children live not only in rather bad conditions, but they also feel and realize the difficulties in their living conditions. Social care, first of all family allowance can significantly mitigate poverty, but its amount in real value decreased in the period investigated and covers only a modest part of the cost of bringing up children.

References

CORAC, M. – SUTHERLAND, H. – LIETZ, CH. [2006]: *The impact of tax and transfer systems on children in the European Union.* 29th General Conference of the IARIW. Joensuu. Working paper.

CROSSLEY, T. F. – CURTIS, L. J. [2006]: Child poverty in Canada. *The Review of Income and Wealth*, Se. 52. No. 2. 237–260. old.

FÖRSTER, M. – N. MIRA D'ERCOLE [2005]: *Income distribution and poverty in OECD countries in the second half of 1990s.* OECD Social, Employment and Migration Working Papers. No. 22. Paris.

FREGUJA, C. – SICILIANI, I. [2006]: *Poverty and deprivation of Italian children: Evidence from the EU-SILC Survey*. ISTAT Paper submitted to the 29th General Conference of the IARIW. Joensuu. Working Paper.

KESZTHELYINÉ RÉDEI M. [2006a]: A lakossági jövedelmek mérésének megbízhatóbb módszere. *Statisztikai Szemle*. Vol. 84. No. 5–6. p. 518–551.

KESZTHELYINÉ RÉDEI M. (ed.) [2006b]: *A 2005. évi lakossági jövedelemfelvétel összefoglaló adatai.* KSH. Budapest.

TÓTH, I. GY. [2005]: *Jövedelemeloszlás. A gazdasági rendszerváltástól az uniós csatlakozásig.* Andorka Rudolf Társadalomtudományi Társaság-Századvég Kiadó. Budapest.

Fiscal expenditures and the GDP – interdependences in transition

Balázs Kotosz
PhD, assistant professor
of the Corvinus University
of Budapest
E-mail: balazs.kotosz@uni-corvinus.hu

The fall of the so-called socialist economies, as well as their transition to market economies, had one of the most interesting and long lasted economic events of the 20th century. Recently, time series became long enough to be analysed by modern econometric methods. With a simple, two-equation, linear model we can analyse and compare the fiscal policy of seventeen Eastern European countries. The empirical testing of the two versions of the model has been resulted a variegated picture about the relation of the GDP and the government expenditures. The Eastern European transition countries are possessing very different features. The economic processes are country specific and it is difficult to elaborate even a simple economic model to apply for this group of countries.

KEYWORDS:
Model building.
Financial applications, financial and stock market.
International analyses, comparisons.

The collapse of socialist economies in Eastern Europe and the former Soviet Union, as well as their subsequent transition towards market economies, was arguably one of the most considerable economic events of the 20th century. Recently, time series are long enough to analyse them with the latest econometric methods. The theoretical background of this paper is based on a paper by *Mellár* [2001] who constructed a simple, but easily verifiable economic model to investigate the Hungarian economy. The model can be transformed to a two-equation vector-autoregressive model which should be estimated on the basis of GDP and general government expenditure data. This idea opens the possibility to analyse and compare the fiscal policies of seventeen Eastern European countries. Four questions could be answered, namely whether

1. the aggregated demand or aggregated supply adjusts faster;
2. the Keynesian multiplicator effect is working;
3. Wagner's law is true; and
4. the government expenditures are limited.

In the first part of the paper, we offer a critical summary of the original Mellár's model. In the second part, we sketch the database used in the empirical estimations and tests. Then the pervious four questions are answered and we turn our attention to the stability question of this model. Originated on this phenomenon of the original model, we estimate an extended model that allows analysing a more shaded equilibrium situation. Following the economic side of the model, through the tests of integration and co-integration, we analyse the statistical characteristics of the applied time series. Finally, further research possibilities are explored.

1. The model

In this chapter, we provide a critical presentation of Mellár's model (*Mellár* [2001]). By this simple model, we can analyse the two-way relation between the GDP and the general government expenditures allowing to follow spillover effects. Because of its simplicity, the model cannot faithfully describe either the effect of different budget expenditures or the evolution of macro-processes. (For the detailed analysis of the effects of different budgetary actions in Eastern Europe, see *Purfield* [2003] and *Kotosz* [2006].) The dynamics of the GDP is based on three equations:

$$YD_t = c\left(Y_t - \tau G_t\right) + A_t + G_t \qquad 0 < c < 1, \qquad \tau > 0\,, \qquad /1/$$

$$YS_t = Y_t + \gamma G_t - \delta\left(\tau G_t\right) \quad \gamma, \delta > 0\,, \qquad /2/$$

$$\Delta Y_t = Y_{t+1} - Y_t = \alpha\left(YD_t - YS_t\right), \qquad |\alpha| < 1 , \qquad /3/$$

where Y means the GDP, c is the marginal rate of consumption, G means the budget expenditures, A represents autonomous expenditures, and t is for time.

The dynamics of the budget expenditures is as follows:

$$\Delta G_t = \beta\left(Y_t^T - Y_t\right) + \omega\left(\bar{G}_t - G_t\right) \qquad \beta, \omega > 0 , \qquad /4a/$$

where Y^T is the expected GDP, $\bar{G}$ is the practical upper limit of budget expenditures.

Additionally:

$$A_t = aY_{t-1} \qquad Y_t^T = hY_{t-1} \qquad \bar{G}_t = kY_{t-1} \qquad a,h,k > 0 . \qquad /5a/$$

Equation /1/ is a simple Keynesian demand function, and it suggests that budget expenditures are covered only by income taxes, financing can be partial (τ<1) or full ($\tau \geq 1$). Equation /2/ is a mixed supply function; the first and the third elements are Lucas-type, while the second element is a Keynesian one. Equation /3/ is not so trivial. As the sign of the α parameter is not fixed, the active role of the aggregated demand is not presupposed. Therefore in small, open economies (like some Eastern European countries) the increase of demand through the expansion of the imports and through the devaluation of the national currency can generate the decline of production. Equation /4a/ offers that the larger is the lag between expected and actual GDP the larger is the growth of budget expenditures, though this increment is reined by the upper limit. The dynamic kind of the model requires flexibility of the autonomous terms; the benchmark can be the lagged GDP (see equation /5a/).

At this point, Mellár makes three simplifications to gain a model easy to deal with. By his idea, we can replace the lagged GDP by current GDP in equation /5a/. By this manipulation, the matrix form of the model is:

$$\begin{bmatrix} Y_{t+1} \\ G_{t+1} \end{bmatrix} = \mathbf{A} \begin{bmatrix} Y_t \\ G_t \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 1 + \alpha\left(c + a - 1\right) & \alpha\left[\left(1-\gamma\right) + \tau\left(\delta - c\right)\right] \\ \beta\left(h-1\right) + \omega k & 1 - \omega \end{bmatrix} . \qquad /6/$$

Version /6/ of the model is very kind for statistical analysis, but doubtful from a theoretical point of view. Let us see what happened. *First*, the autonomous demand depends on current GDP, i.e. not autonomous. This inconsistency cannot be strained off at this level of simplicity of the model.[1] A new interpretation of equation /1/ is the following: a part of the demand is the function of the income but not of the dis-

[1] A clear solution would be the separate analysis of autonomous demand time series, but as they do not exist, the direct measuring is not possible. If we investigate relatively short time series, the autonomous demand (in real terms) can be considered as constant. In this case, equation /6/ is transformed as $a_{11} = 1 + \alpha\left(c - 1\right)$. The stability feature of the model does not change, but we have a constant in the first equation, without constant in the second one. This restriction makes trouble in econometric estimations.

posable income, so some demand is directly independent from taxes. *Second*, the expected GDP is the function of current GDP. The conflict is clear; there is no more expectation about a known measure. Furthermore, the next year budget expenditures grow accordingly as we have faulted the measurement of GDP (i.e. as the difference of current and expected GDP for year t). This conflict can be eliminated by a simple change in equation /4a/, instead of Y_t^T we use Y_{t+1}^T (equation /4b/). This form of the equation suggests that the budget expenditures are higher in year $t+1$, if the expectation of the government of GDP for year $t+1$ is higher than the current GDP in year t. This is a normal assumption, and it is sustainable even without any change in the expenditures/GDP ratio. *Third*, the practical upper limit of the budget expenditures is defined in the function of the current GDP. This change is not too rough, and it can be restored by the different use of equation /4a/, where instead of $\bar{G}_t$ we use $\bar{G}_{t+1}$ (equation /4b/). On account of the latter two variations, the model becomes more prospective, budget expenditures are planned on the basis of the future possibilities and not of the present bias. The new equations are:

$$\Delta G_t = \beta\left(Y_{t+1}^T - Y_t\right) + \omega\left(\bar{G}_{t+1} - G_t\right) \qquad \beta, \omega > 0\,, \qquad /4b/$$

$$A_t = aY_t \qquad Y_t^T = hY_{t-1} \qquad \bar{G}_t = kY_{t-1} \qquad a, h, k > 0\,. \qquad /5b/$$

The stability of the model depends on the absolute values of eigenvalues of the **A** matrix. As *Mellár* [2001] shows, calculated by economically rational parameter values, $tr\mathbf{A} \in [0, 2]$, thereby one of the necessary conditions is fulfilled ($|tr\mathbf{A}| < n$).[2] Mellár supposed that $|det\mathbf{A}| < 1$. We have doubts whether this condition is always fulfilled. If the economy is demand-directed ($\alpha > 0$), the condition on determinant normally is in order; but in a supply-oriented ($\alpha < 0$) economy, if the adjustment of government expenditures is slow (ω is low, the increase of government expenditures is based on supply expansion), the determinant may exceed 1. Based on these strongly Lucasian circumstances, the typically Keynesian model becomes unstable.

If the model is stable, and eigenvalues are real numbers (as they are by the empirical evidence), the equilibrium is stable node or saddle-point. When the eigenvalues are complex numbers, the equilibrium is stable spiral.

2. Data

To analyse the Eastern European economies, we need the description of the database taken in the following chapter. The surveyed period is 1990–2003. An earlier starting date would be pointless for the transition period; the closing date has a practical reason (comparable data has not been available for more recent years) and a

[2] On necessary and sufficient conditions of stability, see *Dameron* [2001].

theoretical one (in some countries the transition has been completed by that time). The relatively short period generates additional uncertainty to the results; thereby one should accept the economic results only with some provisions.

The unification of macroeconomic statistic measures is an old goal of different authorities and a dream of researchers, but recently, has never been reached. The 1993 SNA system and its European adaptation the ESA 95 have some measures of budget revenues and expenditures, and a relatively detailed system has been developed, created by the International Monetary Fund (IMF), the so-called Government Finance Statistics (GFS) system. Its first version was set in 1986; the new one was published in 2001.

By the GFS 1986, the fiscal operations were calculated on cash basis, while in SNA 1993 flows have been recorded on an accrual basis, so the data of the two systems were not comparable. By the new GFS, data are compiled on accrual basis, what makes SNA and GFS data comparable. It is a pity that the old and new data of GFS become incomparable because of the methodological changes; thereby longer time series could be analyzed after numerous adjustments. Otherwise, GFS is fully consolidated, but SNA is not, therefore the calculation of some relative measures (e.g. deficit/GDP) becomes inconsistent.

Even if there are certain standards, only a part of the countries uses them, and just a few in transition. It is clear now, that in the early transition period more important politico-economic tasks were emphasized than producing methodologically comparable government finance statistics, and for this period a set of data is no more reconstruable. In many cases, the analyst has to rely on estimations based on available data. These estimations can be better or worse, but the real numbers remain incognizable, therefore we call this phenomenon "fiscal data illusion".

What type of data would be optimal for our analysis? The GFS (according to SNA) divide the total economy of a country into five sectors:

1. non-financial corporations sector,
2. financial corporations sector,
3. general government sector,
4. non-profit institutions serving households sector, and
5. households sector.

For analytical purposes, each of these sectors can be divided into subsectors. The general government sector is usually divided into central, state and local government sector, but in non-federal countries where regional governments do not exist or do not have enough power, the state government level is skippable. Social security system appears on the competent level, though this element is widely different in countries. As in Eastern Europe, the structure of the general government is diverse, the central government data represent less the role of the state, so general government consolidated data would be better to use. The new *Government Finance Statistics Manual* [2001] proposes the compilation of data for the whole public sector, which has an additional information function, but in our opinion it is less expressive for fiscal policy analysis.

A uniform database for general government fiscal operations does not exist. Even the IMF, publisher of GFS rules, does not have methodologically consistent data. The researcher must scout national sources (as Ministry of Finance, Statistical Office, National Bank), rarely prepared by the same principles, to convert and estimate comparable data. In some cases, even the national authorities have no acceptable data. Further problem is the confliction of sources. The necessary exploratory work was the subject of another paper (*Kotosz* [2004]).

In this one however, a database (as far as it was possible according to GFS standards) compiled and estimated by the author from different sources will be used. For general macroeconomic data, we looked for the World Economic Outlook Database of the IMF, published in April 2004, and Financial Statistics Yearbook series of the IMF. GDP data are on 1995 prices in local currencies, while general government expenditures are estimated indirectly from the current expenditure/GDP rate.

3. Empirical evidence

The econometric estimation and testing of model /6/ are simple. We have to estimate a simple VAR (Vector autoregressive) model which can be figured out most economic software. For this paper, we used EViews.[3] The model parameters' summary and the eigenvalues can be found in Table 1. The main characteristics are the following.

Element a_{11} of the matrix of each country indicates whether aggregated demand (if $a_{11}<1$) or aggregated supply (if $a_{11}>1$) is determinant in growth. As in this case the zero point is 1, in significance tests we have to check if we can reject the null hypothesis of $a_{11}=1$. For all other parameters, the crucial value is 0. Condition $a_{12}>0$ is fulfilled if the government expenditures have adequate transmission mechanism for that Keynesian multiplicator effect could proceed. By the two parameters together, the Keynesian kind of the economy can be tested. We suppose that in an economy where the aggregated demand is dominant, some demand-based economic policy can be successful (if $a_{11}<1$ then $a_{12}>0$). Parameter a_{11} is spread in the [0.53; 1.84] interval. (The value is smallest in Estonia, the largest in Belarus.) The growth is demand-based in eight countries (but significant at 5 percent only in two cases: Estonia and Latvia, at 10 percent in Lithuania, as the Baltic states are the examples of this type of economy). The economy seems to be supply-based in nine countries (in the case of Slovenia the a_{11} parameter is 1.0075, barely different from 1), and this behaviour is significant in Belarus, Bosnia and Herzegovina, Bulgaria, Hungary and Ukraine.

The Keynesian multiplicator effect ($a_{12}>0$) can be found in nine countries (in the case of Slovenia evidently neglectable – $a_{12}=0.0674$), but it is significant only in Es-

[3] As the estimation of VAR models is based on iterative methods, some smaller differences among different methods may emerge.

tonia and Latvia. In eight countries, the increase of the government expenditures refrains the economy; this effect is significant in five countries (Belarus, Bosnia and Herzegovina, Bulgaria, Hungary and Ukraine).

Table 1

*VAR parameter summary**

Country	Dependent variables	Independent variables		Eigenvalues
		Y_t	G_t	
Albania	Y_{t+1}	0.9887 (0.1944)	0.2447 (0.6025)	1.0585
	G_{t+1}	0.0880 (0.0839)	0.7501 (0.2601)**	0.6803
Belarus	Y_{t+1}	1.8402 (0.2751)**	–1.7243 (0.5848)**	1.0289
	G_{t+1}	0.7803 (0.1206)**	–0.6296 (0.2564)**	0.1818
Bosnia and Herzegovina	Y_{t+1}	1.3204 (0.1075)**	–0.4638 (0.1963)**	1.0754
	G_{t+1}	0.4867 (0.2373)**	0.1540 (0.4331)	0.3990
Bulgaria	Y_{t+1}	1.2409 (0.0596)**	–0.5693 (0.1248)**	1.0072
	G_{t+1}	0.3604 (0.0511)**	0.1292 (0.1070)	0.3828
Croatia	Y_{t+1}	1.2689 (0.1878)	–0.4428 (0.3669)	1.0398
	G_{t+1}	0.4219 (0.1165)**	0.2245 (0.2277)	0.4536
Czech Republic	Y_{t+1}	0.5983 (0.3239)	1.0308 (0.7894)	1.0254
	G_{t+1}	0.5363 (0.4285)	–0.2689 (1.0444)	–0.6960
Estonia	Y_{t+1}	0.5332 (0.1408)**	1.3159 (0.3588)**	1.0561
	G_{t+1}	0.1127 (0.0705)	0.7727 (0.1795)**	0.2497
Hungary	Y_{t+1}	1.1695 (0.0517)**	–0.2697 (0.1015)	1.0400
	G_{t+1}	0.1925 (0.0883)**	0.6391 (0.1733)**	0.7686
Latvia	Y_{t+1}	0.7019 (0.1052)**	0.8993 (0.2698)**	1.0650
	G_{t+1}	0.0779 (0.0680)	0.8722 (0.1744)**	0.5090
Lithuania	Y_{t+1}	0.6225 (0.2455)	1.2392 (0.7541)	1.0301
	G_{t+1}	0.0307 (0.1370)	0.9368 (0.4207)**	0.5292
Macedonia	Y_{t+1}	0.9606 (0.2539)	0.1357 (0.6991)	1.0124
	G_{t+1}	0.1470 (0.1400)	0.6273 (0.3855)	0.5755
Moldova	Y_{t+1}	0.7731 (0.2537)	0.5134 (0.7277)	*0.9622*
	G_{t+1}	0.0857 (0.0558)	0.7293 (0.1601)**	*0.5403*
Poland	Y_{t+1}	0.9381 (0.1337)	0.2292 (0.3085)	1.0389
	G_{t+1}	0.1502 (0.1070)	0.6975 (0.2470)**	0.5967
Romania	Y_{t+1}	1.0806 (0.2187)	–0.1703 (0.6295)	1.0231
	G_{t+1}	0.3278 (0.0717)**	0.0529 (0.2065)	0.1105
Slovakia	Y_{t+1}	1.1905 (0.2025)	–0.3662 (0.4845)	1.0346
	G_{t+1}	0.2094 (0.0894)**	0.5426 (0.2139)**	0.6984
Slovenia	Y_{t+1}	1.0075 (0.0678)	0.0674 (0.1462)	1.0292
	G_{t+1}	0.0235 (0.1209)	0.9565 (0.2609)**	0.9347
Ukraine	Y_{t+1}	1.5435 (0.0872)**	–1.3843 (0.2173)**	1.0449
	G_{t+1}	0.3758 (0.0667)**	0.0017 (0.1660)	0.5003

* Standard errors in parentheses.
** Significant at 5 percent (one-side tests, for null hypotheses, see in the table).
Note. Italic numbers for both eigenvalues less than 1.
Source: Here and in the following tables the author's own calculation.

By the empirical evidence, the significant demand-based features and the successful Keynesian macroeconomic policy are attached. It is not a surprise, as the Keynesian policy must raise the demand. Tough the question of Keynesian transmission mechanism is not whether it exists or not, but where and when it is true. Furthermore, what kind of macroeconomic/institutional/political environment makes the economy of a country Keynesian type? This question goes over the goals and frames of this paper.

The parameter a_{21} suggests the test of Wagner's law. In this form (the higher the GDP is, the higher the budget expenditures are in the next year) the law is always true (a_{21}>0). The more absorbing version (if the reallocation rate is growing in the function of the GDP) is the matter of an extended model. We have to mention here that a_{21} is significantly greater than zero only in Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Hungary, Romania, Slovakia, and Ukraine. This set of countries has special equilibrium characteristics (see Figure 4).

Finally, the a_{22}<0 is the sign of the weak control of budget expenditures (e.g. the expected upper limit can be overpassed). This is the case in Belarus and in the Czech Republic, significant only in Belarus where the transition towards market economy is at less developed state (*Kotosz* [2005]).

The general view about equilibrium of the model is almost flat: one eigenvalue of matrix **A** is out of unit circle (except for Moldova); thereby equilibria are saddle-points (in Moldova stable node). We must remark that the stable node equilibrium attracts the country to the origin, where the economy is totally collapsed (with no GDP and no budget). It is very important to see *that the stability of the mathematical model is not equivalent to the stability of the economy*, moreover any economic growth is possible only at the unstable saddle path (balanced growth is not stationary). We know that saddle-point stability with the equilibrium of (Y=0, G=0) means that only a certain budget expenditures/GDP rate ensures the stability of the model. Let us see the phase diagrams! It is clear by economic rationality (and partly based on a previous test on Keynesian kind) that the ΔY_t=0 and ΔG_t=0 lines have positive gradient (less than 1). The ΔY_t=0 curve is $G_t = \frac{1-a_{11}}{a_{12}} Y_t$ and the ΔG_t=0 curve is $G_t = \frac{a_{21}}{1-a_{22}} Y_t$ (see on Figure 1).

The stability feature of the model in the relevant (+,+) quadrant depends on the relative position of the two curves. If the gradient of ΔG_t=0 is larger, then phase diagram can be seen on Figure 2. In this case, the phase diagram suggests an optimal long-term growth path. The (otherwise instable) saddle path results a constant expenditure/GDP ratio while the economy is continuously growing. In the positive quadrant, the dynamics is getting closer and closer to this optimal path. If the expenditure/GDP ratio is too high, the growing GDP decreases it; if the ratio is too low, the government expenditures increase faster than the GDP.

If the gradient of ΔY_t=0 is larger than the gradient of ΔG_t=0, the phase diagram looks a bit different (see Figure 3). This situation is less favourable than that on Figure 2. In these countries the too high expenditure/GDP rate can decrease only via the shut of the GDP, in extreme situation, the huge rate can result the collapse of the economy. Generally, these countries have to pay a large price for their budget consolidation by necessary depression.

Figure 1. $\Delta Y_t=0$ and $\Delta G_t=0$ curves

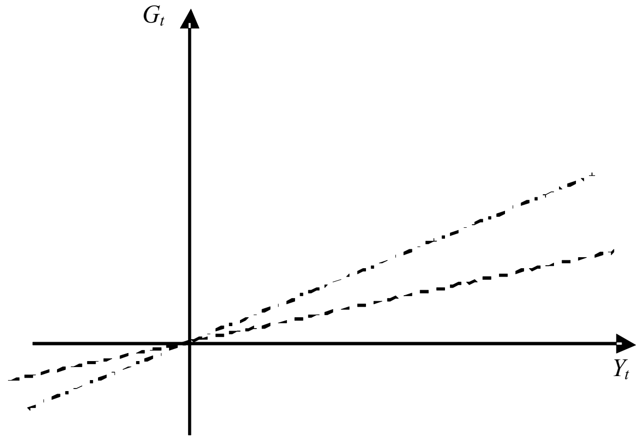


Figure 2. Phase diagrams for Albania, the Czech Republic, Estonia, Latvia, Lithuania, Macedonia, Poland and Slovenia

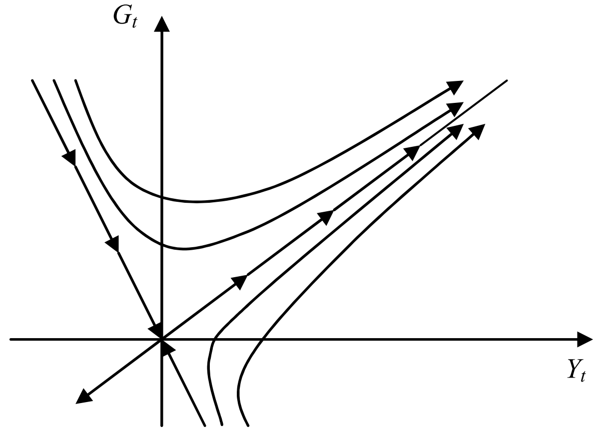
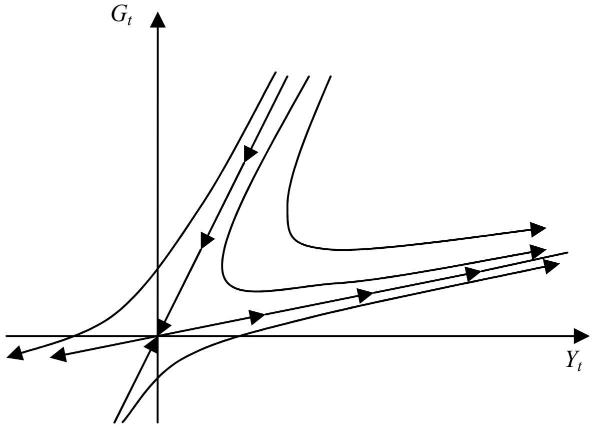


Figure 3. Phase diagrams for Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Hungary, Romania, Slovakia and Ukraine

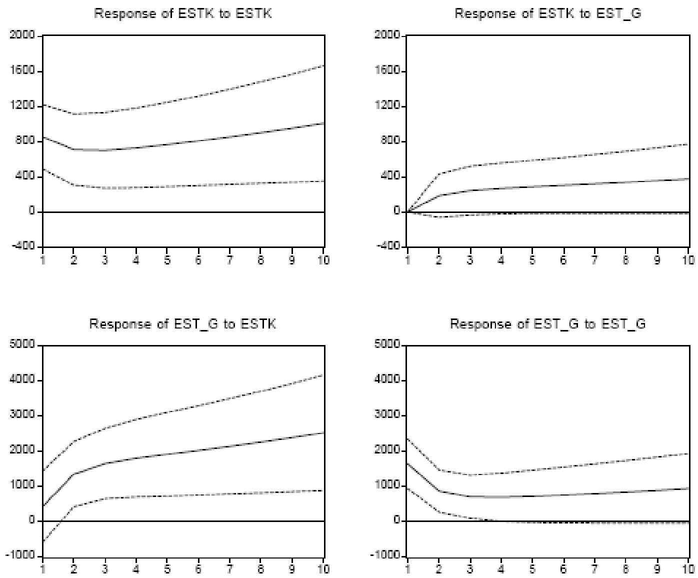


Finally, the mid-, and long-term behaviour of the economies can be described by the impulse response functions. This figure is very spectacular about the effect of different shocks (one standard deviation shock in government expenditures or of GDP). By the empirical evidence, there are two main types of the impulse response functions.

In one of these types all positive shocks have positive effects (a positive shock of the government expenditures makes the future GDP and the future government expenditures higher for all the next ten years, and a positive shock of the GDP makes the future GDP and the future government expenditures higher), but the size of the effect is different from country to country. Albania, Croatia, the Czech Republic, Estonia, Latvia, Lithuania, Moldova, Poland, Romania, Slovakia, and Slovenia are classed of this type. We turn a special attention to the Czech Republic. Because of the negative eigenvalue of the **A** matrix, all effects are sinusoidal decaying. In Lithuania, the effects of GDP shocks are positive, but insignificant all the time. In Poland, the prompt effect of the government expenditures shock is negative, but the second year positive impact balances it, then the response is continuously positive. In Slovakia, the shock of the government expenditures is non-effective for the GDP, and the effect on the government expenditures breaks down very quickly. In Slovenia, the budget expenditures are totally insensitive to GDP shocks.

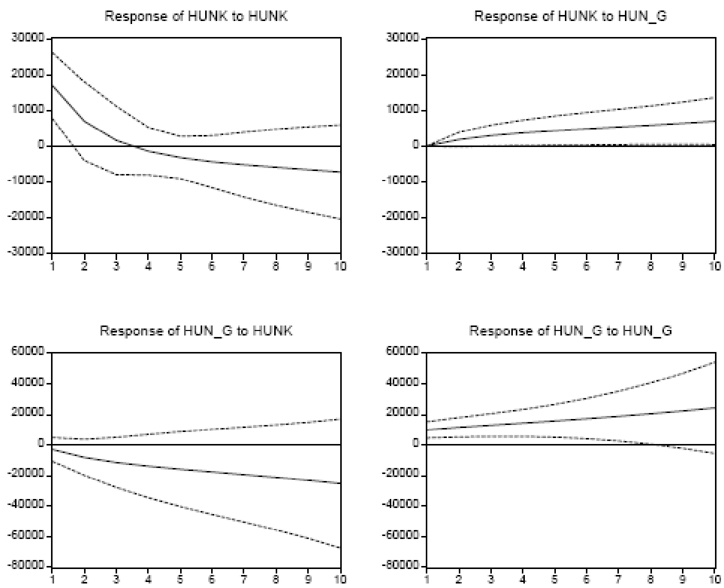
In the second type (Belarus, Bosnia and Herzegovina, Bulgaria, Hungary, Macedonia, and Ukraine), the response of the GDP and of the government expenditures are positive to GDP shocks, just like in the first type. Though in long-term, their response to government expenditures shocks are negative. The effect on budget expenditures breaks down, in the third or fourth year, it turns to be negative, and the cumulative effect in ten years time is negative. The most rigid budget is the Hungarian one (see Figure 5).

Figure 4. Response to Cholesky One Standard Deviation Innovations ± 2 Standard Errors (Estonia)



Note. ESTK for Estonian government expenditures and EST_G for Estonian GDP.

Figure 5. Response to Cholesky One Standard Deviation Innovations ± 2 Standard Errors (Hungary)



Note. HUNK for Hungarian government expenditures and HUN_G for Hungarian GDP.

4. Extension of the model

In the first part of the paper, we had a question about the possible extension of the modified model. The autonomous demand in the previously analysed model was not really autonomous. We can solve this inconsistency by assuming a constant autonomous demand (A_t=a). To be able to estimate the parameters of the new model, we need a theoretical constant in the second equation, as well. The easiest way is to hypothesize a constant (g) in equation /4b/ transformed it to equation /4c/.

$$\Delta G_t = g + \beta\left(Y_{t+1}^T - Y_t\right) + \omega\left(\bar{G}_{t+1} - G_t\right) \qquad \beta, \omega > 0 . \qquad /4c/$$

This constant means that we suppose that there is a permanent change in the government expenditures. For simplicity and coherence with the original model, we do not make any restriction about this constant. If g=0, then equation /4b/ is equal to equation/4c/.

The new VAR model is as follows:

$$\begin{bmatrix} Y_{t+1} \\ G_{t+1} \end{bmatrix} = \mathbf{A}\begin{bmatrix} Y_t \\ G_t \end{bmatrix} + \begin{bmatrix} \alpha a \\ g \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 1+\alpha\left(c-1\right) & \alpha\left[\left(1-\gamma\right)+\tau\left(\delta-c\right)\right] \\ \beta\left(h-1\right)+\omega k & 1-\omega \end{bmatrix} . \qquad /7/$$

The estimated parameters of the new model are shown in Table 2. If we compare the parameters of equation /6/ and /7/, a much diversified picture is shown. While in some countries beside the insignificant constants the parameters are practically unchanged, in others the main characteristics are totally different.

Table 2

VAR with constant – parameter summary

Country	Dependent variables	Independent variables		
		C	Y_t	G_t
Albania	Y_{t+1}	–773.9288	0.984933	0.287218
	G_{t+1}	6995.751	0.122285	0.365904
Belarus	Y_{t+1}	–1016.703	1.896046*	–1.704487*
	G_{t+1}	–1051.449	0.838088*	–0.609042*
Bosnia and Herzegovina	Y_{t+1}	636.1174*	0.890003*	–0.194015*
	G_{t+1}	179.0803	0.365533	0.229932
Bulgaria	Y_{t+1}	7.652851	1.191257	–0.533431*
	G_{t+1}	–35.31150	0.589437*	–0.036224
Croatia	Y_{t+1}	4431.195	0.924837	–0.080842
	G_{t+1}	3532.612*	0.147639	0.513008*

(Continued on the next page.)

(Continuation.)

Country	Dependent variables	Independent variables		
		C	Y_t	G_t
Czech Republic	Y_{t+1}	14238.32	0.538097	0.926526
	G_{t+1}	–10935.25	0.582486	–0.188781
Estonia	Y_{t+1}	–4296.226	0.630700*	1.271165*
	G_{t+1}	103.2788	0.110329	0.773782*
Hungary	Y_{t+1}	2660.415	1.169197*	–0.277739
	G_{t+1}	64432.32	0.185934*	0.443906
Latvia	Y_{t+1}	–38.53512	0.740275	0.860893*
	G_{t+1}	16.24386	0.061698	0.888346*
Lithuania	Y_{t+1}	187.4230	0.603447	1.249804
	G_{t+1}	615.5001	–0.031935	0.971601*
Macedonia	Y_{t+1}	45936.36	0.520185	–0.037213
	G_{t+1}	–32503.13*	0.458687*	0.749593*
Moldova	Y_{t+1}	1192.582*	0.720731	–1.177609
	G_{t+1}	102.2250	0.081263	0.584390*
Poland	Y_{t+1}	301.8661	1.038816	–0.171459
	G_{t+1}	185.6762	0.212109*	0.451035
Romania	Y_{t+1}	5976161	0.932018	–0.749964
	G_{t+1}	3529799*	0.240032*	–0.289446
Slovakia	Y_{t+1}	77060.37	0.765635	0.501389
	G_{t+1}	–18181.30	0.309665*	0.337868
Slovenia	Y_{t+1}	17662.04*	1.030718	–0.304118
	G_{t+1}	23963.88	0.054997	0.452326
Ukraine	Y_{t+1}	130.5329	1.465027*	–1.330780*
	G_{t+1}	–27.83720	0.392520*	–0.009256

* Significant at 5 percent (one-sided tests, for null hypotheses, see previously).

Let us see which countries are sensitive to the constant. Bosnia and Herzegovina seemed to be significantly supply-based economy in the original model, but it is weakly demand-based in the new one. The Keynesian multiplicator effect was significantly out of work, the significance is extincted, just as the antecedently significant Wagner's law is no more significant. In Croatia, the trend of the changes is the same. Lithuania is the first and last example of the (insignificant) falsification of Wagner's law. Macedonia follows the tendency of drastical changes of the other two ex-Yugoslavian state, as all the four parameters are greatly different in the models. With Romania, these four countries break the rule of demand-side manipulation feature, as in this model they are demand-based, but not Keynesian economies. The contrary is Slovakia, where in the model without constants the economy is supply-based and not Keynesian, in the extended version, we find a demand-based and Keynesian economy. In the case of Slovenia, the insignificantly Keynesian economy turns to be significantly non-Keynesian in the extended model. Finally, in the original version, the budget expenditures have no limit only in Belarus and in the Czech Republic (see Table 2). Bulgaria, Romania and Ukraine join to this group of weak control.

Table 3

VAR with constant – stability feature

Country	Dependent variables	Equilibrium*	Tangent of the Δ=0 line	Eigenvalues	Stability feature
Albania	*Y*	–59391.781	0.052	1.037249	saddlepoint
	G	–421.02919	0.193	0.313587	
Belarus	*Y*	–11779.637	0.526	1.018144	saddlepoint
	G	–6789.0217	0.521	0.268860	
Bosnia and Herzegovina	*Y*	2924.4154	–0.567	*0.754914*	stable node
	G	1620.7019	0.475	*0.365021*	
Bulgaria	*Y*	230.26992	0.359	*0.827022*	stable node
	G	96.907727	0.569	*0.328011*	
Croatia	*Y*	38574.444	–0.930	*0.893465*	stable node
	G	18948.369	0.303	*0.544380*	
Czech Republic	*Y*	721809.02	0.499	*0.994277*	stable node
	G	344477.58	0.490	*–0.644961*	
Estonia	*Y*	14824.323	0.291	1.083507	saddlepoint
	G	7686.5305	0.488	0.320975	
Hungary	*Y*	386727.43	0.609	1.089166	saddlepoint
	G	245170.96	0.334	0.523937	
Latvia	*Y*	–401.45997	0.302	1.056378	saddlepoint
	G	–76.355679	0.553	0.572243	
Lithuania	*Y*	15136.087	0.317	*0.787524±*	stable spiral (clockwise)
	G	4652.5996	–1.125	*±0.077641i*	
Macedonia	*Y*	92643.172	–12.894	*0.634889±*	stable spiral (counter-clockwise)
	G	39899.399	1.832	*±0.062546i*	
Moldova	*Y*	1772.1126	–0.237	*0.652560±*	stable spiral (counter-clockwise)
	G	592.45972	0.196	*±0.301744i*	
Poland	*Y*	8890.0169	0.226	*0.968541*	stable node
	G	3773.1527	0.386	*0.521311*	
Romania	*Y*	18898756	–0.091	*0.760578*	stable node
	G	6255481.3	0.186	*–0.118006*	
Slovakia	*Y*	–510711633	0.467	1.000091	saddlepoint
	G	–238876386	0.468	0.103412	
Slovenia	*Y*	–24370462	0.101	1.000190	saddlepoint
	G	–2403507.2	0.100	0.482855	
Ukraine	*Y*	3183.0571	0.349	*0.869725*	stable node
	G	1210.3732	0.389	*0.586046*	

* Equilibrium coordinates are not comparable among countries because of the use of local currency units.
Note. Italic numbers for eigenvalues of equilibrium different from saddlepoint.

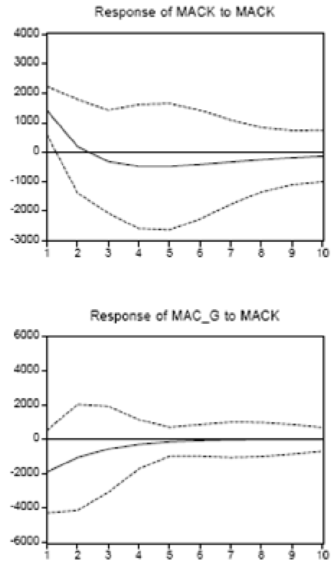
The stability feature of the extended model is different theoretically. Even the presence of the constants does not influence our statements about the relationship of matrix **A** and the stability conditions. If we ignore the constraint (e.g. the equilibrium point can be different from the origin), we have more theoretical possibilities. Beside the saddlepoint, other stability features can be rational in economic sense. For an economy with equilibrium in the positive quadrant, in mid-term the stable node or stable spiral stability can be fruitful if the equilibrium GDP is larger than the actual one. The equilibria in the negative quadrant all have saddlepoint stability.

In the countries, where the equilibria are in the positive quadrant, the equilibrium expenditure/GDP rates are computable. This rate is 55.4 percent in Bosnia and Herzegovina, 42.1 percent in Bulgaria, 49.1 percent in Croatia, 47.7 percent in the Czech Republic, 51.2 percent in Estonia, 63.4 percent in Hungary, 30.7 percent in Lithuania, 43.1 percent in Macedonia, 33.4 percent in Moldova, 42.4 percent in Poland, 33.1 percent in Romania, and 38.0 percent in Ukraine. Where the economy has a stable node or stable spiral stability feature, these expenditure/GDP rates can be declared as target of the countries. In the case of Estonia and Hungary, the particular situation of the saddle-paths is determinant, so further explorative work is necessary. Having regard to the tangent of the ΔY=0 and the ΔG=0 curves, the equilibria are the same as in the equation /6/, for Estonia (see on Figure 2) and for Hungary (see on Figure 3). It means that GDP and budget expenditures of both countries are over the equilibrium values, the dynamics assures the continual increasing of the GDP and the government expenditures. We will focus to the stability question of the expenditures/GDP rate in the next chapter of the paper.

The impulse responses are generally smaller, the effects become faster insignificant in the model with constants. For a series of the countries (Albania, Belarus, Bulgaria, Croatia, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Macedonia, Romania, and Ukraine) the impulse response functions are very similar in the two models.

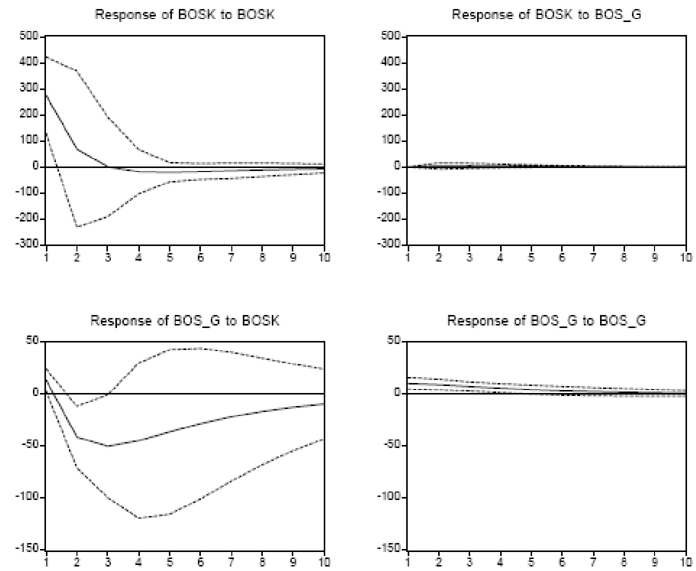
In countries, where the presence of the constants establishes stable equilibrium (instead of saddlepoint), the response functions are declining (converging to zero) after three or four years. The most conspicuous example is Macedonia (see Figure 6). Turn our attention to the special cases. In Bosnia and Herzegovina, the response to the shocks of the GDP is negligible, the self-effect of government expenditures is strongly declining, while the GDP responds to budget shocks with an enormous shut from the second year (see Figure 7). Moldova and Poland have the same characteristics as Hungary in both models (see Figure 5), even in the original model they formed part of the "every effect is positive" group. Slovakia is the counter-example of Poland, the declining hardly significant effects of the original model are replaced by the "every effect is significantly positive" case. Slovenia is another special case. Except for the autoresponse of the GDP (that is significantly positive in long term), all other responses are the same as in Bosnia and Herzegovina.

Figure 6. Response to Cholesky One Standard Deviation Innovations ± 2 Standard Errors (Model with constants – Macedonia)



Note. MACK for Macedonian government expenditures and MAC_G for Macedonian GDP.

Figure 7. Response to Cholesky One Standard Deviation Innovations ± 2 Standard Errors (Model with constants – Bosnia)



Note. BOSK for Bosnian government expenditures and BOS_G for Bosnian GDP.

5. Integration, cointegration, causality

In this analysis, we favoured the economic model and used the econometric methods ancillary. At this point, one has to check the general characteristics of the analysed time series to be sensible for the further possibilities of analysis. This section contains the tests about our time series. The results of some tests are conditions of the application of modeling techniques, while others go forward. We decided to hold together statistical tests, because we would like to separate the economic and statistical analysis of the model.

First it is important to investigate the integration of the time series. In *Mellár* [2001], the author experienced that the longer Hungarian time series both were $I(2)$ processes. Mellár comments that usually these type of time series are $I(1)$ processes. The unit root tests of these time series are problematic, because critical values are available for sample size at least 20, while our series are no longer than 14 years. It is predictable that the absolute value of small sample critical values are higher and it implicates that p values are higher than estimated for sample size of 20 (in some cases the test – adjusted to sample size of 20 – indicates the rejection of null hypothesis, but it should not be rejected, as the time series has a unit root).

Table 4

ADF test statistics

Country	Dependent variables	$I(0)$	$I(1)$	$I(2)$	Other
Albania	*Y*	0.204	–4.639**		
	G	–0.605	–3.447**		
Belarus	*Y*	–1.084	–1.254	–2.432	trend stationary
	G	–0.990	–6.604**		
Bosnia and Herzegovina	*Y*	–14.03**			
	G	–6.477**			
Bulgaria	*Y*	–1.126	–1.590	–3.477**	
	G	–5.790**			
Croatia	*Y*	–0.810	–5.035**		
	G	–4.339**			
Czech Republic	*Y*	–0.482	–4.510**		
	G	0.342	–2.096	–3.493**	
Estonia	*Y*	1.662	–3.385**		
	G	–0.802	–4.326**		
Hungary	*Y*	0.744	–7.479**		
	G	–0.829	–2.788*	–4.212**	
Latvia	*Y*	1.683	–2.690	–2.222	trend stationary
	G	–0.031	–2.825*	–5.278**	
Lithuania	*Y*	0.267	–3.774**		
	G	–0.407	–3.933**		

(Continued on the next page.)

(Continuation.)

Country	Dependent variables	*I*(0)	*I*(1)	*I*(2)	Other
Macedonia	*Y*	–2.004	–3.142*	–3.738**	
	G	0.956	–2.683	N/A	first difference trend stationary
Moldova	*Y*	0.194	–4.823**		
	G	–1.017	–3.373**		
Poland	*Y*	–2.202	–5.221**		
	G	1.054	–3.145*	–11.80**	
Romania	*Y*	–2.887*	–3.453**		
	G	0.366	–6.242**		
Slovakia	*Y*	–0.098	–4.146**		
	G	0.116	–2.934*	–4.743**	
Slovenia	*Y*	0.107	–2.224	–4.413**	
	G	–3.366**			
Ukraine	*Y*	–0.525	–1.317	–4.632**	
	G	–4.029**			

* Null hypothesis of the unit root is rejected at 10 percent.
** Null hypothesis of the unit root is rejected at 5 percent.

We have chosen the ADF (Augmented Dickey–Fuller) test for testing the unit root. The test statistics are in Table 4. In the cases when the original time series has not a unit root, *I*(1) and *I*(2) were not tested, and when the first difference of the time series has not a unit root, *I*(2) was not tested. The results differ from country to country. The most expected version that both GDP and government expenditures are integrated in first order (*I*(1)), cannot be rejected in eight countries (Albania, Estonia, Hungary, Lithuania, Moldova, Poland, Romania, Slovakia). In the case of Hungary, Poland and Slovakia at 5 percent the budget expenditures are *I*(2) time series. Bulgaria, Slovenia and Ukraine forms another group with *I*(2) GDP and stationary (*I*(0)) budget expenditures. Bosnia and Herzegovina has special feature with its stationary time series (due to data only after the war). In Belarus and Latvia, the GDP series are trend stationary (deterministic trend). Our results are far from Mellár's results, as for Hungary the second order integrated time series of the GDP are not proved, thereby his longer series are maybe *I*(2) processes because of a structural break around 1990.

Engle and *Granger* [1987] pointed out that a linear combination of two or more non-stationary series can be stationary. If such a stationary linear combination exists, the non-stationary time series are said to be cointegrated. The stationary linear combination is called the cointegrating equation and may be interpreted as a long-run equilibrium relationship among the variables (here between the GDP and the budget expenditures).

Owing to the results of the unit root tests, we can seek for cointegration equation only in eight countries. In this paper, we used the Johansen cointegration test (*Johansen* [1991], [1995]), the results can be found in Table 5.

Table 5

P-values of the cointegration tests and G/Y rates of cointegration equations*
(percent)

Country	Trace test without constant	Maximum eigenvalue test without constant	G/Y rate in CE equation	Trace test with constant	Maximum eigenvalue test with constant
Albania	4.57	9.35	-	0.01	0.00
Estonia	0.92	9.29	0.53	0.94	3.85
Hungary	0.43	8.71	2.00	1.92	10.15
Lithuania	0.47	1.58	0.34	4.79	4.78
Moldova	37.72	58.98	-	0.00	0.00
Poland	1.37	10.98	0.46	7.15	25.57
Romania	48.44	46.93	-	0.01	0.00
Slovakia	5.44	3.84	0.45	4.79	10.47

* The null hypothesis is the lack of cointegrating vector.

The results of the cointegration tests are antinomic. At 5 percent significance, only the Lithuanian GDP and budget expenditures are cointegrated without constant[4] (i.e. a fix expenditures/GDP rate can be supposed). By the trace test, we have found the two time series cointegrated without constant in Estonia, Hungary, Lithuania and Poland. In Albania, Moldova, and Romania cointegration with constant have been found. In the case of Estonia and Slovakia, a cointegration equation with and without constant can be presumed. In Table 5 the long-term relation of the two variables are expressed by *G/Y* rate, as it appears in the cointegration equation without constant. These rates generally are close to the observed rates, except for Hungary, where a theoretically (economically) impossible value has been calculated. Even the tests indicate cointegration equation also without constant, the version including the constant would be economically rational.

Table 6 contains the cointegration equations with constant, in the form of $G=\alpha Y+c$ (for the inverse form a simple re-arrangement is necessary). This form expresses a long-term – linear – relation between the budget expenditures and the GDP, the α parameter reads as the marginal reallocation rate; from 1 additional unit of national income, the government spends α unit more. Theoretically, the $\alpha=0$, $c>0$ situation means that the government spends without any regard to the GDP, while the $\alpha>0$, $c=0$ is the case of pure proportional spending, when a fix rate of the GDP is reallocated every year. As the base of the budget expenditures are the budget revenues, the two extremities imply the lump-sum and the proportional taxing system. By the empirical evidence, Albania and Romania is near to the lump-sum attitude, and the European Union members have very similar characteristics to each other.

[4] Note that including the constant in the cointegration equation is insignificant.

Table 6

*Parameters of cointegration equations with constant**

Country	α	*c*
Albania	0.181 (0.007)	8 979.4 (542.0)
Estonia	0.313 (0.028)	4 322.4 (1 345.5)
Hungary	0.277 (0.082)	168 350 (44 626)
Moldova	0.577 (0.047)	1 580.7 (82.54)
Romania	0.147 (0.011)	3 406 948 (189 531)
Slovakia	0.299 (0.078)	336 249 (99 762)

* Standard errors in brackets.

Table 7

*P-values of Granger causality tests**

Country	Dependent variables	Lags		
		1	2	3
Albania	ΔY	0.532	0.870	0.672
	ΔG	0.640	0.822	0.770
Bosnia and Herzegovina	Y	**0.005**	0.252	N/A
	G	0.538	0.988	N/A
Estonia	ΔY	0.923	0.830	0.888
	ΔG	0.770	0.384	**0.023**
Hungary	ΔY	0.733	0.916	0.743
	ΔG	0.521	0.597	0.224
Lithuania	ΔY	0.906	0.505	0.657
	ΔG	0.734	0.644	0.525
Moldova	ΔY	0.390	0.365	0.840
	ΔG	0.476	0.935	0.485
Poland	ΔY	0.640	0.692	0.978
	ΔG	0.723	0.778	**0.009**
Romania	ΔY	0.777	0.224	0.316
	ΔG	0.538	0.251	0.305
Slovakia	ΔY	0.245	0.727	0.881
	ΔG	0.752	0.974	0.308

* Null hypothesis: the independent variables do not Granger cause the dependent variables.
Note. Bold numbers for *p*-values less than 0.05.

Finally, in the countries, where the GDP and the government expenditures have the same $I(d)$ processes, the causality of the variables or their differences can be tested. The necessary condition for Granger causality test is the stationarity of the time series, thereby the Bosnian time series can be directly tested, and for the other eight countries, where both the GDP and the government expenditures can be $I(1)$ processes, the test can be executed on the first differences (i.e. the annual change of the GDP and the annual change of the budget expenditures). The main results are summarized in Table 7. Generally, the GDP cannot be declared to either cause the government expenditures, or vice versa, with any rational lags. There are three exceptions. In Bosnia and Herzegovina, the government expenditures are Granger cause of the GDP. As data are available for the after-war period, the reconstruction of the country has been based on international aids, arriving through governmental channels. In Estonia and Poland, with three lags in the test equation, the GDP change can be the cause of the government expenditures change. It is difficult to find the economic background of the three years lagged effect of the GDP changes. Otherwise, the significant VAR parameters and the lack of Granger causality suggest that the relation cannot be described in Grangerian term.

6. Conclusions

The empirical testing of the two versions of the simple dynamic model has been a result of a variegated picture about the relation of the GDP and the government expenditures. The Eastern European countries – each of them in transition from planned to market economy – have very different characteristics.

In the simplest model, clear country groups can be formed. While in some demand-based economies the Keynesian multiplicator is working, and the dynamics assure the continual growth, in others only supply-based economic policy can be efficient, and if once the government expenditures/GDP rate becomes too high, the reduction is possible only by general depression. The question is not more about the existence of Keynesian multiplicator, but about the institutional and political background of the nature of the economy. The extended model covers the equilibrium and attracts attention to the sensitivity of the equilibrium feature of certain countries. The discrepancies of the results of the two model versions turn our attention to the question if the different paths are in great part caused by different initial conditions.

The econometric analysis of the time series explores that for sophisticated analysis of this country group we need different tools. The varied order of integration opens the possibility of further analysis in certain countries, while restricts in others. In the countries of I (1) GDP and expenditure series, these two variables are cointegrated, but generally, a cause-consequence relation cannot be explored.

Finally, we have to mention that the main results of the original and the extended model for Hungary are analogous to the main findings of Mellár for a longer and mainly not transitional period. This fact strengthens our hypothesis that the deep economic processes are country specific and it is difficult to elaborate even a simple economic model applicable for this group of countries. For further analysis, it would be interesting to check if other – country-by-country – empirical works accept or deny the found fiscal rule. From modelling point of view, the application of a more detailed and/or non-linear model would be able to fortify (or refute) our hypotheses.

References

DAMERON, P. [2001]: *Mathématiques des modèles économiques*. Economica. Paris.

ENGLE, R. F. – GRANGER, C. W. J. [1987]: Co-integration and error correction: Representation, estimation and testing. *Econometrica*. Vol. 55. No. 2. p. 251–276.

Government Finance Statistics Manual 2001 [2001]. International Monetary Fund. Washington D.C.

GRANGER, C. W. J. [1969]: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. Vol. 37. No. 3. p. 424–438.

HUNYADI, L. [1994]: Egységgyökök és tesztjeik. *Szigma*. Vol. 25. No. 3. p. 135–164.

IMF [1994-2003]: *International financial statistics yearbook* series, different numbers. International Monetary Fund, Washington, D.C.

IMF [2004]: *World economic outlook database*. International Monetary Fund, Washington, D.C.

JOHANSEN, S. [1991]: Estimation and hypothesis testing of cointegration vectors in Gaussian Vector Autoregressive Models. *Econometrica*. Vol. 59. No. 6. p. 1551–1580.

JOHANSEN, S. [1995]: *Likelihood-based inference in Cointegrated Vector Autoregressive Models*. Oxford University Press. Oxford.

KOTOSZ, B. [2004]: A fiskális adatbázisok összeállításának nehézségei Kelet-Európában. *Statisztikai Szemle*. Vol. 82. No. 10–11. p. 945–961.

KOTOSZ, B. [2005]: A transzformációs reformok méréséről. *Statisztikai Szemle*. Vol. 83. No. 4. p. 365–387.

KOTOSZ, B. [2006]: Megszorítások és lazítások – A rendszerváltás fiskális politikájának szerkezetéről. *Közgazdasági Szemle*. Vol. 53. No. 2. p. 158–174.

MELLÁR, T. [2001]: Kedvezményezett vagy áldozat: A GDP és a költségvetési kiadások kapcsolata. *Statisztikai Szemle*. Vol. 79. No. 7. p. 573–586.

PURFIELD, C. [2003]: *Fiscal adjustments in transition countries: Evidence from the 1990s*. IMF Working Paper 03/36. International Monetary Fund, Washington D.C.

WAGNER, A. [1883]: *Finanzwissenschaft*. Winter. Leipzig.

The triple deficit of Hungary*

György Szakolczai
professor emeritus, Budapest College of Management
E-mail: szakolczai.gyorgy@t-online.hu

The aim of this paper is to show that the notion and analysis of the twin deficit, the deficit of the current account and of the state budget must be extended to the notion and analysis of the triple deficit, the deficit the same two deficits and the deficit or insufficiency of domestic savings. The results support the view that these three problems, although closely intertwined, are to a certain extent independent, have autonomous causes, and must therefore dealt with separately. This result is contradictory to the common view that all problems are the consequences of state overspending and all of them can be solved by reducing the budget deficit and by cutting state expenditures. The ensuing policy recommendations are therefore, that the exports of goods and particularly services, and domestic private savings must be increased along with the reduction of budget deficit.

KEYWORDS:
National accounts, Input-output analyses, GDP.

* The author would like to thank for the useful comments of *dr. Erzsébet Bablina*, *Gábor Huszár* and *Annamária Márkus*. He is also grateful to *György Muszély* for his computer support. All the remaining errors are the author's.

This paper deals with the latest results of a major study the first findings of which were published in *Szakolczai* [2005a] and *Szakolczai* [2005b]. The aim of the whole research is to analyse the fundamental disequilibria of the Hungarian economy, i.e. the triple deficit, the joint deficit of the current account, of the state budget, and of the deficit or insufficiency of domestic savings. The paper presents the most important data available and some conclusions that can be drown on the basis of their simple verbal analysis. The author hopes that the same and some further data will make possible the use of more intricate methods to corroborate and extend the results presented here.

The fundamentals of the theoretical side of the problem are simple. The triple deficit is the obvious extension of the well-known twin deficit, the deficit of the current account and of the state budget. As it is shown in *Szakolczai* [2005b], following the analysis of *Dornbusch* [1988)], the identity

$$CA = (T - G) + (S - I)$$

necessarily holds, i.e. current account necessarily equals the sum of the excess of taxes over government expenditures and of the excess of domestic private savings over domestic private investments. This equation shows that the extension of the problem of twin deficit to the third component, the insufficiency of domestic savings as compared with domestic investments, i.e. to the triple deficit, is therefore almost unavoidable. This formula by itself, however, tells us nothing about the causes and interconnections of the mentioned deficits.

These causes and interconnections pertain already to the field of applied economics or economic policy. The commonly accepted view is that the fundamental cause of the twin or triple deficit is the deficit of the state budget, and that the cure is the decrease of the budget deficit or the balanced budget. According to this view the twin or triple deficit is therefore the consequence of the overspending of the state, and all three deficits cease to exist if state expenditures are cut back. This view accepted generally, however has no theoretical foundation, as the formula tells nothing about causes and effects. This way of thinking is also inconsistent with the obvious symmetry of the problem, as well as with actual experiences. This paper tries to show that, in the recent Hungarian case, three independent primary problems exist, those *1.* of budget deficit, *2.* of current account deficit and *3.* of the deficiency of savings. These three problems, although intertwined, have their own autonomous causes, and the whole problem can only be solved if the three deficits are dealt with separately. This does not mean that one of the problems, if exists, does not aggravate the other two. The whole triple problem is made even more complicated by the ac cumulated debt of the budget and the accumulated foreign debt. These two accumulated debts would necessitate a surplus of primary budget and a surplus of trade in goods and services which can be very difficult or even impossible to attain under present Hungarian circumstances.

To show this from another angle, it is certainly true that current account deficit shrinks if taxes are increased, state expenditures are cut back, and, as a result, GDP shrinks. When current account deficit is the consequence of structural trade problems and/or of the foreign indebtedness of the country, no politically acceptable or even feasible tax increase or expenditure cut can be enough to solve the problems of current account and trade deficit. These measures may or even must lead to the decrease of domestic savings. This shows that neither simplistic analysis nor simplistic policy recommendations are acceptable. As a result, the independent or autonomous causes of the three deficits or gaps must be analysed separately, and the policy advice will necessarily be the parallel intervention on all three fields.

These problems are particularly difficult under present Hungarian circumstances. The common view is, that the origin of all three problems is state overspending, and that the solution is not only a reduction of state expenditures, but the radical decrease of the role of the state, and the retreat or even partial elimination of the welfare state. This paper tries to show that the origin of the triple deficit is not state overspending alone, and that the problem cannot be solved by cutting state expenditures, or eliminating, at least partially, the welfare state. It does not mean that budget deficits are not high and should not be cut or even eliminated in the longer run.

The final solution of this problem would call for more intricate theoretical analysis, which is beyond the scope of this paper. Our aim is therefore only to show that the simplest possible analysis of national accounts data published by the Hungarian Central Statistical Office (HCSO), supplemented by some data of the National Bank of Hungary (Magyar Nemzeti Bank), is enough to clear up the most fundamental elements of the problem. These data point also to the advisable way of decreasing the triple deficit. Some elements of these time series go back to 1995, the year following the end of transition depression in Hungary, but the income accounts of the government are only available (for the author at the time of completing the paper) for the years 2001–2004. The author is, of course, clearly unable to extend these time series to the previous years. Even these very short time series enable us, however, to deal with these issues, and to point to their solution.

1. Domestic use

Let us first deal with the real side of the problem, with domestic use or domestic excess consumption, while the financing of this excess consumption, i.e. the monetary aspects are left to the later parts of the paper. Table 1 presents the ten year time series of the most important data.

Table 1

Gross domestic product, domestic use and external balance of goods and services, 1995–2004
(million HUF at current prices)

Year	GDP	DU	FCH	FCG	GFCF	CI	EBGS
1995	5 614 042	5 616 747	3 730 258	617 700	1 125 389	143 400	-2 705
1996	6 893 934	6 862 063	4 400 359	703 619	1 475 538	282 547	31 871
1997	8 540 669	8 453 306	5 283 032	900 797	1 898 917	370 560	87 363
1998	10 087 434	10 232 425	6 297 192	1 024 579	2 384 615	526 039	-144 991
1999	11 393 499	11 703 435	7 274 153	1 156 726	2 724 532	548 024	-309 936
2000	13 150 766	13 679 267	8 334 942	1 253 347	3 099 131	971 846	-506 974
2000	13 272 167	13 809 584	8 489 615	1 352 799	3 099 131	868 038	-537 417
2001	14 989 800	15 227 436	9 723 771	1 541 268	3 499 687	462 710	-237 636
2002	16 915 259	17 312 148	11 228 255	1 849 717	3 941 489	292 686	-396 889
2003	18 650 746	19 462 419	12 816 005	2 088 844	4 156 000	401 570	-811 673
2004	20 429 456	21 155 327	13 785 221	2 189 154	4 631 205	549 767	-725 871

Note. GDP is gross domestic product; DU is domestic use, total; FCH is final consumption of households; FCG is final consumption of government; GFCF is gross fixed capital formation; CI is changes in inventories; EBGS is external balance of goods and services.

Source: National Accounts, Hungary, 2003–2004 [2006]. HCSO. Budapest. p. 13., 14–15. and 20–21.

Data for 2000 of Table 1 are presented here and also later according to the methodology of the period of 1995–2000 and 2000–2004, respectively. Data show that GDP was practically equal to domestic use in 1995, and that GDP even surpassed slightly domestic use in the following two years. In the later years the two time series diverged, and a high negative external balance of goods and services developed. The origin of this problem can be *1.* general overspending *2.* overspending of the state as it is generally assumed, or *3.* a decrease of exports with respect to imports.

These possibilities can be examined by using the data of percentage distribution of total domestic use shown by Table 2.

The first and last columns of Table 2 show the excess of domestic use over the GDP in percentages. The increase of this excess is not monotonous but pronounced, and it is obvious that even a slight overspending of a few percentages, if it is maintained for a period of almost a decade, leads to serious problems as we must see now. The data of the four central columns show that while the excess of total domestic use grows steadily there are only marginal changes in the distribution of domestic use. The share of final consumption of households decreases slightly, that of final consumption of government decreases marginally, while the share of gross fixed capital formation increases somewhat, which can be considered as advantageous. These changes are in the right direction and might have been even more pronounced but the ratios show no special private or government overspending. If there is overspending, it is general domestic, but not government overspending, as generally stated. This increasing excess of domestic use over the GDP seems therefore to be independent of any change in the composition of domestic use. This seems to support the view that its origin can be found in the field of international trade problems and not in state overspending.

Table 2

Domestic use and external balance of goods and services, 1995–2004
(percent)

Year	DU%	FCH%	FCG%	GFCF%	CI%	EBGS%
1995	100.0	66.4	11.0	20.0	2.6	0.0
1996	99.5	64.1	10.3	21.5	4.1	0.5
1997	99.0	62.5	10.7	22.5	4.4	1.0
1998	101.4	61.5	10.0	23.3	5.1	-1.4
1999	102.7	62.2	9.9	23.3	4.7	-2.7
2000	104.0	60.9	9.2	22.7	7.1	-3.9
2000	104.0	61.5	9.8	22.4	6.3	-4.0
2001	101.6	63.9	10.1	23.0	3.0	-1.6
2002	102.3	64.9	10.7	22.8	1.7	-2.3
2003	104.4	65.9	10.7	21.4	2.0	-4.4
2004	103.6	65.2	10.3	21.9	2.6	-3.6

Note. DU% is total domestic use, in percentage of GDP; FCH% is final consumption of households in percentage of total domestic use; FCG% is final consumption of government in percentage of total domestic use; GFCF% is gross fixed capital formation in percent of total domestic use; CI% is changes in inventories in percentage of total domestic use; EBGS% is external balance of goods and services in percent of GDP.

Source: Calculated from Table 1.

The data used until now are expressed in current prices, but the findings are even more confirmed by data expressed in constant prices and shown in Table 3.

As it can be seen from Table 3, data in comparable prices are available for the years from 1995 to 1998, from 1998 to 2000 and from 2000 to 2004. They appear in the first six rows of Table 3. The following three rows present rates of increase for these three time periods, and the fourth row the products of these three indices, i.e. the volume indices for the whole period of 1995–2004. These latter nine year rates of increase are obviously inconsistent with the assumption of state overspending on collective consumption. Final consumption of government at constant prices increased in these nine years by 21.57 percent what is less than half of 48.72, the percentage increase of the GDP.

The next three rows show the average annual rates of increases or volume indices for the previous three time periods, while the last row presents the average annual volume indices for the whole nine year time period. These latter volume indices clearly show that the average annual rate of increase of domestic use is more than two times and that of gross fixed capital formation more than three times higher than that of final consumption of the government. This is not government overspending on collective consumption but the opposite. If there is an overspending, it is the overspending of households, because the average rate of increase of final consumption of households exceeds that of the GDP. There is also an overspending on investments too, because their rate of increase exceeds that of the GDP. Considering that rapid increase of investments is highly desirable, this also means overspending of households, because the increase of investments is not accompanied by a corresponding decrease in the share of private consumption. Instead of this the share of private consumption increases.

Table 3

Gross domestic product, domestic use and external balance of goods and services, 1995–2004
(million HUF at constant prices)

Year	GDP	DU	FCH	FCG	GFCF	CI	EBGS
1995	5 614 042	5 616 747	3 730 258	617 700	1 125 389	143 400	– 2 705
1998	6 238 452	6 379 020	3 850 031	623 254	1 485 645	420 101	–140 567
1998	10 087 434	10 232 425	6 297 192	1 024 579	2 384 615	526 039	–144 991
2000	11 053 751	11 214 942	6 923 477	1 055 702	2 718 627	517 137	–161 191
2000	13 272 167	13 809 584	8 489 615	1 352 800	3 099 131	868 037	–537 410
2004	15 637 065	16 507 975	11 005 361	1 581 878	4 021 320	–100 585	–870 910
1998/1995T	111.11	113.51	103.21	100.91	132.01	..	..
2000/1998T	109.58	109.60	109.95	103.04	114.01	..	111.17
2004/2000T	117.82	119.54	129.63	116.93	129.76	..	162.06
2004/1995T	143.45	148.72	147.10	121.57	195.29	..	..
1998/1995Y	103.58	104.33	101.06	100.30	109.70	..	..
2000/1998Y	103.10	103.10	103.21	101.00	104.47	..	103.59
2004/2000Y	104.18	104.56	106.70	103.99	106.73	..	112.83
2004/1995Y	104.09	104.51	104.38	102.19	107.72	..	

Note. GDP is gross domestic product (at producers prices); DU is total domestic use; FCH is final consumption of households; FCG is final consumption of government; GFCF is gross fixed capital formation; CI is changes in inventories; EBGS is external balance of goods and services; 1998/1995T, 2000/1998T, 2004/2000T are volume indices for the given period; 2004/1995T is volume index for the given period, quotient of the previous three indices; 1998/1995Y, 2000/1998Y, 2004/2000Y are average annual volume indices for the given period; 2004/1995Y is average annual volume index for the given period.

Source: *National Accounts, Hungary, 2003–2004* [2006]. HCSO. Budapest. p. 18–19. and 22–23.

2. Foreign trade deficit

With respect to and in spite of what has been written previously it can be assumed that the root of the problem can be found in the field of foreign economic relationships, in the foreign trade deficit, i.e. in the inadequacy of exports and in overspending on imports. Table 4 presents the data of exports, imports and balance of trade for goods and services in order to see whether this assumption is valid.

Table 4

Exports, imports and trade balance of goods and services, 1995–2004
(million euro at current prices)

Year	XG	MG	BG	BG%	XS	MS	BS	BS%
1995	11 281	12 402	-1 122	-9.9	3 970	2 966	1 004	25.3
1996	12 743	14 080	-1 337	-10.5	4 683	3 177	1 506	32.2
1997	17 083	18 248	-1 165	-6.8	5 146	3 583	1 562	30.4
1998	21 057	22 742	-1 685	-8.0	4 811	3 736	1 075	22.3
1999	24 059	26 102	-2 044	-8.5	4 910	4 094	816	16.6
2000	31 278	34 457	-3 180	-10.2	6 429	5 195	1 234	19.2
2001	34 697	37 193	-2 496	-7.2	7 865	6 203	1 661	21.1
2002	36 821	39 024	-2 203	-6.0	7 820	7 233	587	7.5
2003	38 377	41 275	-2 898	-7.6	7 674	8 075	-401	-5.2
2004	45 083	47 536	-2 453	-5.4	8 660	8 533	127	1.5

Note. XG is exports of goods; MG is imports of goods; BG is trade balance of goods; BG% is trade balance of goods in percent of exports of goods; XS is exports of services; MS is imports of services; BS is trade balance of services; BS% is trade balance of services in percent of exports of services.

Source: Here and in the following table the National Bank of Hungary, Department of Statistics.

The trade balance of goods is negative in all of the ten years as it shown in Table 4, reaching its highest value in 2000, and decreasing later. The negative trade balance of goods was, however, almost completely counterbalanced by the positive balance of services in 1995, leading to the practical equality of GDP and domestic use in this year. The positive balance of services became higher than the negative balance of goods in the next two years leading to a positive external balance of goods and services appearing in the last columns of Table 1 and Table 2. In the last years of the period analysed here the negative trade balance of goods settled at about EUR2.5 billion per year while the positive balance of services practically disappeared. The trade balance of services was negative in 2003 and practically zero in 2004. This led to the high negative external balance of goods and services appearing in the last column of Table 1.

These findings confirm that the excess of domestic use over GDP is not resulting from state overspending as it is generally stated but it is the consequence of the stable negative balance of goods and the collapse of the positive balance of services. This is therefore a structural foreign trade problem and not a budget problem as it is generally asserted. The negative balance of goods is extensively analysed in *Szakolczai* [2005] where it has been shown to be the consequence of the dual character of the Hungarian economy. The greatest part of exports is produced by the foreign owned engineering firms while the greatest part of the Hungarian economy is dependent on imports but unable to produce the exports needed to pay for them. This problem is aggravated by the budget deficit, as part of this deficit is spent on imported goods. It would be almost nonsensical to say nevertheless that the dual character of the Hungarian economy or, particularly, the rapid disappearance of the positive trace balance of services is the consequence of government overspending.

The trade problem of services can be analysed by using data of Table 5.

Table 5

Exports, imports and trade balance of tourism and of services other than tourism, 1995–2004
(million euro at current prices)

Year	XT	MT	BT	BT%	XO	MO	BO	BO%
1995	2 258	1 158	1 100	48.7	1 712	1 808	–96	–5.6
1996	2 843	1 183	1 660	58.4	1 840	1 994	–154	–8.4
1997	3 384	1 325	2 060	60.9	1 762	2 258	–372	–21.1
1998	3 248	1 314	1 934	59.5	1 452	2 422	–859	–59.2
1999	3 359	1 450	1 909	56.8	1 551	2 644	–1 093	–70.5
2000	4 067	1 794	2 273	55.9	2 362	3 401	–1 039	–44.0
2001	4 654	2 022	2 632	56.6	3 211	2 992	–971	–30.2
2002	3 925	2 252	1 673	42.6	3 895	4 981	–1 086	–27.9
2003	3 577	2 289	1 288	36.0	4 097	5 786	–1 689	–41.2
2004	3 265	2 302	962	29.5	5 395	6 231	–835	–15.5

Note. XT is exports of tourism; MT is imports of tourism; BT is trade balance of tourism; BT% is trade balance of tourism in percent of exports of tourism; XO is exports of services other than tourism; MO is imports of services other than tourism; BO is trade balance of services other than tourism; BO% is trade balance of services other than tourism in percent of exports of services other than tourism.

The data of Table 5 show that the problem can be traced back to two causes. On the one hand, the earnings from tourism reached the highest level in 2001 and decreased in the following three years by about EUR1.4 billion while the domestic outlays on foreign tourism increased constantly in the whole period studied. On the other hand, the negative balance of services other than tourism increased to almost twenty fold of its original value between 1995 and 2003, and the negative balance remained high also in 2004. It can be stated therefore that origin of the problem of increasing excess of domestic use over the GDP lays in foreign trade problems and not in state overspending. This is a structural foreign trade problem that can be solved first of all by cutting back Hungarian expenses on foreign tourism and by attracting foreign tourists into Hungary, and also by developing Hungarian services other than tourism.

These results confirm the view that the problem of triple deficit has multiple causes and can be solved only by facing the structural problems of the Hungarian economy. It cannot be denied that one of these structural problems is the structural deficit of the budget, but it cannot be stated that it is the only or even the most important cause of the whole problem. It can, of course, be argued that the decrease of net earnings of tourism and the increase of expenses on services other than foreign tourism diminished the rate of increase of sources available for domestic use. The Hungarian economy failed to adapt itself to this change, and the domestic use increased as if the positive balance of services had not disappeared. This is obviously true, but this does not mean that the origin of the problem is not the disappearance of the positive trade balance of services, and that the reduction of government expenditures on collective consumption would have solved the problem or would solve it in the future.

3. Domestic product and national income

We must turn now from the analysis of the real side of the economy to the income side. Domestic use must not only be compared with the GDP, but first of all with the gross national income (GNI) and the gross national disposable income (GNDI), because no person, family, community or country can spend more for a longer time than his disposable income. The first relating data appear in Table 6.

Table 6

Gross domestic product, gross national income and gross national disposable income, 1995–2004
(million HUF at current prices)

Year	GDP	CE	PI	EUT	GNI	UT	GNDI
1995	5 614 042	1 401	–213 763	–	5 401 680	25 152	5 426 832
1996	6 893 934	11 355	–308 056	–	6 597 233	–938	6 596 295
1997	8 540 669	22 677	–517 469	–	8 045 877	38 228	8 084 105
1998	10 087 434	28 084	–647 939	–	9 467 579	52 224	9 519 803
1999	11 393 499	24 203	–692 616	–	10 725 086	103 309	10 828 395
2000	13 150 766	42 366	–753 725	–	12 439 407	99 984	12 539 391
2000	13 272 167	42 366	–723 281	–	12 591 252	99 984	12 691 236
2001	14 989 800	45 383	–833 727	–	14 201 456	115 864	14 317 320
2002	16 915 259	36 602	–956 328	–	15 995 533	127 780	16 123 313
2003	18 650 746	35 250	–955 398	–	17 730 598	149 639	17 880 237
2004	20 429 456	31 708	–1 243 523	60 575	19 278 216	63 887	19 342 103

Note. GDP is gross domestic product; CE is compensation of employees, net; PI is property income, net; EUT is European Union transfers, net; GNI is gross national income; UT is unrequested current transfers, net; GNDI is gross national disposable income.

Source: *National Accounts, Hungary, 2003–2004* [2006]. HCSO. Budapest. p. 13., the National Bank of Hungary, Department of Statistics.

HCSO publishes only data for gross national income that is GDP plus balance of labour and capital incomes. The first is, obviously, a small positive, and the latter an important negative item. Since the 2004 EU transfers, which are much less important as general public assumes, they appear separately. HCSO GNI data are presented in the fifth column of Table 6. We have obtained gross national disposable income by adding to the HCSO GNI figure the data of un-requested transfers published by the central bank. GNDI data calculated in this way appear in the last column of Table 6. Only GNI data are therefore directly compatible with the other HCSO data used here.

Domestic use must be compared with the GNI or rather with the GNDI, and we have therefore determined the excess of domestic use over the GNI and the GNDI. The first difference, is directly compatible with other HCSO data, the second is more correct in a theoretical sense. These data, together with data on financing this excess, are presented in Table 7.

Table 7

Excess of domestic use over gross national income and gross national disposable income and their financing, 1995–2004
(million HUF at current prices)

Year	EDU(GNI)	EDU(GNDI)	REFDI	OSF(GNI)	OSF(GNDI)
1995	215 067	189 924	–27 050	242 117	216 974
1996	264 830	265 768	78 210	186 620	187 558
1997	407 429	369 201	245 160	162 269	124 041
1998	764 846	712 622	252 470	512 376	460 152
1999	978 349	875 040	273 420	704 929	601 620
2000	1 236 860	1 139 876	280 760	956 100	859 116
2000	1 218 332	1 118 348	280 760	937 572	837 588
2001	1 025 980	910 116	385 626	640 354	524 490
2002	1 316 615	1 188 835	456 144	860 471	732 691
2003	1 731 821	1 582 182	445 095	1 286 726	1 137 087
2004	1 877 111	1 813 224	449 377	1 427 734	1 363 847

Note. EDU(GNI) is excess of domestic use over gross national income; EDU(GNDI) is excess of domestic use over gross national disposable income; REFDI is reinvested earnings of foreign direct investments, net; OSF(GNI) is other sources of financing the excess of domestic use over gross national income, EDU(GNI) = REFDI + OSF(GNI); OSF(GNDI) is other sources of financing the excess of domestic use over gross disposable national income, EDU(GNDI) = REFDI + OSF(GNDI).

Source: *National Accounts, Hungary, 2003–2004* [2006]. HCSO. Budapest. p. 13, and calculated from Tables 1 and 6.

The two left columns of Table 7 show the excess of domestic use over the GNI and the GNDI. This excess is considerable, it amounted to 10 percent of the GNI and 9.1 percent of the GNDI in 2000, decreased somewhat in 2001, increased again in the following two years, and attained practically the 2000 year-level in 2004. It is obviously impossible to maintain an excess of domestic use over the GNI or the GNDI approaching or even attaining 10 percent for a longer time period. Domestic use must therefore be decreased, as soon as possible, to the level of disposable income with other sources of financing that are available for a longer time. Though as it has been shown previously, the excess of domestic use is not the consequence of government overspending on collective consumption but of overspending in general.

It can be argued again that the Hungarian economy failed to face to the problem that a considerable part of the GDP is foreigners' property income, did not maintain domestic use within the limits of the GNDI. It behaved as if the whole GDP would have been at its disposal which is obviously false. Though this does not mean that the source of the problem is government overspending on collective consumption.

The excess of domestic use over the GNI or the GNDI must, of course, be financed. Part of it was financed by reinvested earnings of foreign direct investments, and it may be argued that this part does not lead to the increase of the burden of debt servicing. The greatest part of financing, more than three quarters of it, came, however, from other sources. This greater part, particularly if we consider the extremely high values of the last four and especially all of the last two years, increases and will also increase in the future the debt servicing burden of the country.

The shares and rates of increase of the main components of domestic use were presented in Tables 2 and 3, and it could be seen that the rate of increase of government outlays on collective consumption was much smaller than that of household consumption and of gross fixed capital formation. To show the income side of the problem, the percentage share of disposable income of the five sectors of the economy is shown in Table 8.

Table 8

Disposable income of the five sectors of the national economy, 1995–2004
(percentages)

Year	DIC%	DIF%	DIG%	DIH%	DIN%
1995	6.8	2.0	..	65.7	..
1996	6.2	1.6	..	66.2	..
1997	8.8	1.3	..	64.4	..
1998	9.5	1.6	..	64.2	..
1999	10.7	1.9	..	62.4	..
2000	10.2	2.0	..	61.4	..
2000	12.1	2.2	..	62.1	..
2001	11.1	2.2	20.3	63.1	..
2002	12.8	1.6	20.0	62.0	..
2003	13.7	2.1	19.8	61.0	..
2004	13.2	1.9	19.6	62.3	..
2000/1995	177.1	112.3	..	94.5	..
2004/2001	118.6	88.1	96.7	98.8	..

Note. DIC% is disposable income of non-financial corporations in percent of gross national disposable income; DIF% is disposable income of financial corporations in percent of gross national disposable income; DIG% is disposable income of government in percent of gross national disposable income; DIH% is disposable income of households in percentage of gross national disposable income; DIN% is disposable income of non-profit institutions in percent of gross national disposable income; 2000/1995, 2004/2001 are Indices.

Source: *National Accounts, Hungary, 2003–2004, 2002–2003, 2001–2002, 2000–2001, 1998–2000, 1998–1999, 1995–1997* [2006], [2005], [2004], [2003], [2002], [2001], [1999]. HCSO. Budapest.

Data of Table 8 lead to the same conclusions as those shown in Tables 2 and 3. The share of corporations is increasing, that of financial corporations is practically the same in the first and in the last years of the period studied while the share of households and of government decreases. This change is very articulate in the last three years. There is therefore no redistribution of income in favour of the government and not even in favour of households while the increasing share of corporations may further investments and growth. This structure can be considered as sound.

To turn back to the problem of triple deficit, these findings show the obvious. The trade problems are aggravated by the accumulated debt and by the possible increases of the debt servicing burden that may be caused by factors fully independent of this country and of its budget deficit. Although the budget deficit, if it cannot be financed by domestic savings, which is the present Hungarian case, obviously increases foreign indebtedness and debt servicing burden. It cannot be argued that the current account deficit is the direct consequence of the current budget deficit.

4. Budget deficit

Even if foreign trade and current account problems cannot be traced back to the problem of budget deficit, this latter is the second and obviously very important element of the triple deficit. Its analysis demands a rather detailed survey of the relevant data. The first of them is the distribution of income accounts of the government that were available for the author for the last four years and are presented in Table 9.

Table 9

Distribution of income accounts of government, 2001–2004
(million HUF at current prices)

Item	2001	2002	2003	2004
Operating surplus, net	–8 067	–13 861	–582	–13 497
Taxes on production and import	2 325 807	2 547 690	2 938 418	3 308 172
Subsidies	266 938	301 337	277 896	324 679
Interest, dividends and rents received	137 060	108 568	124 937	227 579
Interest, paid, consolidated	721 736	676 590	722 551	849 844
Balance of primary income	1 466 126	1 664 470	2 062 326	2 347 731
Current taxes in income and wealth	1 537 656	1 738 070	1 807 387	1 879 241
Social contributions, received	1 971 090	2 323 803	2 535 713	2 770 021
Social benefits other than in kind	1 916 602	2 284 145	2 617 404	2 885 518
Balance of other current transfers	–154 248	–222 306	–242 978	–317 634
Disposable income	2 904 022	3 219 892	3 545 044	3 793 841

Source: *National Accounts, Hungary, 2003–2004, 2002–2003, 2001–2002* [2006], [2005], [2004]. HCSO. Budapest. p. 88–89., 86., 86.

Data of Table 9 show that the structure of government incomes is relatively stable and do not point to any major problems. This stability may, however, be the root of the budget problems because the rates of increase of taxes on production and imports and particularly of current taxes on income and wealth are small which means that the increase of government income from taxes may not cover the increase of expenditures. Having seen the income side let us turn now to the side of expenditures. The relevant data are shown in Table 10.

Table 10

Use of disposable income of government, 2001–2004
(million HUF at current prices)

Item	2001	2002	2003	2004
Disposable income	2 904 022	3 219 892	3 545 044	3 793 841
Final consumption expenditure	3 231 080	3 931 661	4 588 886	4 866 355
Individual consumption expenditure	1 717 376	2 104 922	2 500 042	2 677 221
Collective consumption expenditure	1 513 704	1 826 739	2 088 844	2 189 134
Saving (+) or excess consumption (–)	–327 058	–711 769	–1 043 842	–1 072 514
Balance of capital transfers	–396 883	–673 234	–341 506	–230 826
Changes in net worth*	–723 941	–1 385 003	–1 385 348	–1 303 340
Capital accumulation, net**	–66 396	42 994	–195 736	–208 225
Net lending (+) or borrowing (–)	–657 545	–1 427 994	–1 189 612	–1 095 115

* Changes in net worth due to savings and capital transfers
** See Table 11.
Source: *National Accounts, Hungary, 2003–2004, 2002–2003, 2001–2002* [2006], [2005], [2004]. HCSO. Budapest. p. 90–91., 88., 88.

The first five rows of Table 10 present how savings or excess consumption are determined in the system of national accounts. Savings or excess consumption is the difference between disposable income and final consumption expenditures while the latter are the sum of government's outlays on individual and collective consumption. Data show that excess consumption doubled and trebled in 2002 and 2003, respectively, and stabilised in 2004. It can also be seen that individual consumption expenditures – transfers and social expenditures in the broad sense – increased more than outlays on collective consumption. The two increases are 55.9 and 45.6 of percent the 2001 data, respectively. The further rows of Table 10 show the balance of capital transfers and the changes in net worth. The detailed analysis of this data would be important but it is impossible without further background information.

The last two rows of Table 10 show capital accumulation and net borrowing. According to these rows that are based on the published HCSO data excess consumption is financed not only by net borrowing which is generally known but also by negative capital accumulation or capital destruction – the decrease of the capital stock of the government – which is not generally known and obviously absurd and unacceptable. Financing current consumption by capital destruction is contrary to all possible economic considerations and to the interests of the following generation. This side of the problem requires some further analysis for which Table 11 contains relevant data.

Table 11

Capital accumulation of government, 2001–2004
(million HUF at current prices)

Item	2001	2002	2003	2004
Gross fixed capital formation	563 251	815 684	652 995	730 680
Changes in inventories	1 550	2 697	283	4 002
Consumption of fixed capital*	614 308	759 340	813 665	872 506
Consumption of fixed capital**	576 146	593 451	631 298	657 761
Acquisition of non-produced assets***	–16 889	–16 047	–35 349	–70 401
Capital accumulation****	–66 396	42 994	–195 736	–208 225
Capital accumulatione*****	–28 234	208 883	–13 369	6 520

* Original data published in the first source indicated.
** Corrected preliminary data published in the second source indicated.
*** Acquisition less disposal of non-produced non-financial assets.
**** Calculated from the data published in the first source indicated.
***** Calculated from the data published in the second source indicated.
Source: *National Accounts, Hungary, 2003–2004, 2002–2003, 2001–2002* [2006], [2005], [2004]. HCSO. Budapest. p. 90–91., 88., 88. and www.ksh.hu.

Net capital accumulation – a notion is not appearing in this form in the national accounts – can be obtained by adding up the three elements of gross capital accumulation: *1.* gross fixed capital formation, *2.* changes in inventories and *3.* acquisition of less disposal of non-produced non-financial assets. Deducting consumption of fixed capital, i.e. amortisation (depreciation) can be also added. HCSO publishes data of for the government sector only, and two sets of these data are available. The first are the original data published in printed form on which capital accumulation appearing in the last but one row of Table 10 is based, and which can be seen in the third row of Table 11. The second are the corrected preliminary data published on the HSCO homepage (www.ksh.hu). These corrected preliminary data can be found in the fourth row of Table 11. Finally, capital accumulation based on these two sets of capital consumption data are shown in the last two rows of Table 11.

The first set of data lead to the conclusion that government financed current consumption by capital destruction in 2003–2004 because consumption of fixed capital (amortisation) exceeded gross capital accumulation in these years. The corrected preliminary data lead to the somewhat less startling conclusion that capital accumulation is these two years was practically zero. The other part of capital destruction is the result of selling government property, mainly real estates. This appears in the national accounts as acquisition (+) or sale (–) of non-produced non-financial assets. This element is negative in all four years for which data are available. These negative values are also unacceptable by all possible economic considerations, but their magnitude is not influenced by the difference between the various capital consumption data.

We must add as an explanation of what has been written previously that capital consumption (amortisation) data are rather "soft", as the are, necessarily, estimates based on assumptions, particularly in the government sector. The fact that the first set of these data led to the startling conclusion of capital destruction on a major scale might have led to the use of new, different assumptions and the resulting new set of amortisation data. It is therefore advisable to present here both sets of amortisation data and the results based on both sets because this seems to be the best way to describe the recent situation.

To turn back to the main line of analysis, saving or rather excess consumption, negative capital accumulation and negative net lending or rather net borrowing data appear in relative terms in Table 12.

Table 12

Main indices of capital accumulation and net borrowing of government, 2001–2004
(percent)

Item	2001	2002	2003	2004
Saving* in percent of DIG	–11.3	–22.1	–29.5	–28.3
Saving* in percent of GDP	–2.2	–4.2	–5.6	–5.3
Capital accumulation in percent of DIG**	–2.3	1.3	–5.5	–5.5
Capital accumulation in percent of GDP**	–0.4	0.3	–1.1	–1.0
Capital accumulation in percent of DIG***	–1.0	6.5	–0.4	–0.2
Capital accumulation in percent of GDP***	–0.2	1.3	–0.1	–0.0
Net lending**** in percent of DIG	–22.6	–44.4	–33.6	–28.9
Net lending**** in percent of GDP	–4.4	–8.4	–6.4	–5.4

* Saving (+) or excess consumption (–).
** Calculated from the data published in the first source indicated in Table 11.
*** Calculated from the data published in the second source indicated in Table 11.
**** Net lending (+) or borrowing(–).
Note. DIG is disposable income of the government; GDP is gross domestic product.
Source: Calculated from Table 11.

Negative saving i.e. excess consumption of the government surpassed government disposable income by more than 10, by more that 20 and by nearly 30 percent in 2001, 2002 and 2003–2004, respectively. The same excess amounted to more than 2 percent of the GDP in 2001 and to more than 5 percent in 2003–2004. The negative capital accumulation obtained using the first set of amortisation data was 5.5 percent of government disposable income and about 1 percent of the GDP in the same two years, while the negative capital accumulation obtained using the second set of amortisation data was 0.4 and 0.2 percent, respectively, of government disposable income and about 0.1 and 0.0 percent of the GDP in the same two years. Finally, the negative net lending i.e. borrowing of government attained its peak at almost 45 percent of government disposable income and 8.4 percent of the GDP in 2002. In the years of 2003–2004 values stabilised at about 30 percent of the government disposable income and 6 percent of the GDP.

To turn finally to the economic implications, it is obvious that the situation described cannot be maintained. Current consumption may not surpass disposable income for a longer time and particularly not in the mentioned degree, and the excess cannot be financed from credits. Capital accumulation cannot be negative or even zero. According to some estimates capital accumulation of government ought to attain at least 3 percent of the GDP but even this seems to be a low figure. Current consumption may not be financed by selling off property. The conclusions are therefore obvious and the difference between the economic interpretation of capital destruction or zero capital accumulation is marginal. The whole situation and the policies leading to this point suggest serious reconsideration.

Table 13

Disposable income, saving, capital accumulation and net lending in subsectors of the government sector, 2001-2004
(million HUF at current prices)

Item	Year	CG	LG	SS	G
Disposable income	2001	2 026 775	492 839	384 408	2 904 022
	2002	2 246 162	548 288	425 442	3 219 892
	2003	2 421 512	675 601	447 931	3 545 044
	2004	2 572 489	769 584	451 768	3 793 841
Savings (+) or excess consumption (–)	2001	438 913	–838 224	72 253	–327 058
	2002	278 788	–1 028 849	38 292	–711 769
	2003	154 315	–1 187 651	–10 506	–1 043 842
	2004	203 783	–1 221 130	–55 167	–1 072 514
Capital accumulation	2001	–107 025	36 963	3 666	–66 396
	2002	–110 251	150 950	2 295	42 994
	2003	–317 056	116 611	4 709	–195 736
	2004	–289 696	83 340	–1 869	–208 225
Net lending (+) or borrowing (–)	2001	171 145	–895 415	66 725	–657 545
	2002	–278 883	–1 183 360	34 246	–1 427 997
	2003	154 922	–1 328 061	–16 473	–1 189 622
	2004	258 324	–1 300 139	–53 300	–1 095 115

Note. CG is central government; LG is local government; SS is social securities; G is government, total.
Source: *National Accounts, Hungary, 2003-2004, 2002-2003, 2001-2002* [2006], [2005], [2004]. HCSO. Budapest. p. 89–91., 88., 88.

This conclusion is reinforced by the subdivision of government into its three subsectors, *1.* the central government, *2.* the local government and *3.* the social securities. This analysis is based on published data which are consistent with the published data shown previously and which are presented in Table 13.

Local government data are the most striking. According to HCSO data local government excess consumption was close to HUF1 billion in 2001 and above HUF1 billion in the following three years. Local government negative let lending i.e. borrowing was even somewhat higher because of positive capital accumulation. According to the same data central government net lending was positive in 2001 and in the years of 2003–2004, and central government resorted to borrowing only in 2002.

These conclusions are, however, consequences of the methodology used to consolidate data. According to this methodology central government income transfers to local governments do not appear as a part of the disposable income of local government but as a part of their borrowing. Credits taken up by central government in order to finance income transfers to local governments appear therefore as credits taken up by local governments. This is, on the one hand, reasonable, because these credits finance the expenses of local governments, though on the other hand, misleading, because the debtor is the central and not the local government. Data should be interpreted in this way.

The capital accumulation data presented in Table 13 are not influenced by the method of consolidation. It is therefore true that local governments had a small positive capital accumulation even if we use the old amortisation values. To turn to the social securities, their position deteriorated continuously in these years, but the saving, excess consumption and net lending or borrowing of this subsector was almost negligible as compared with those of central and local government.

5. Income of households

Considering that more detailed analysis of budget problems cannot be attempted here we must turn to the third element of triple deficit, the problem of income and saving of households. The first relating data are shown in Table 14.

Table 14 builds up the disposable income of households starting from their primary income.[1] In order to obtain disposable income of households shown in the last column social benefits in cash and other transfers received by households are added, while current taxes on income and wealth and other current taxes, social contribu-

[1] There is a slight difference due to methodological reasons between the data of disposable income of households used in Tables 14, 15 and 16, on the one hand, and Tables 17 and 18, on the other. The former are published in the subsequent publications quoted previously, and the latter the in last edition of the National Accounts. The latter is the latest version of the disposable income of households (DIH), but it cannot be compared with the other data shown in Tables 14, 15 and 16.

tions of households and other transfers paid by households are deducted. Further analysis is built upon these data but the stability or instability of the distribution of households' disposable income can only be analysed using the distribution data. (See Table 15.)

Table 14

Distribution of households' disposable income, 1995–2004
(million HUF at current prices)

Year	PIH	SBM	TrORH	TH	SCH	TrOPH	DIH
1995	3 831 410	910 459	485 354	411 965	894 523	361 326	3 559 409
1996	4 744 483	992 604	439 101	520 171	1 013 284	259 747	4 382 986
1997	5 732 626	1 158 386	544 595	599 410	1 260 235	356 404	5 219 558
1998	6 661 130	1 405 781	545 595	700 579	1 499 300	314 380	6 098 247
1999	7 370 429	1 583 399	570 139	818 496	1 599 084	352 580	6 753 807
2000	8 571 615	1 748 658	623 153	998 707	1 898 529	371 887	7 674 303
2000	8 552 039	1 755 180	605 554	999 708	1 891 917	196 031	7 825 117
2001	9 721 559	1 982 249	829 986	1 185 800	2 206 805	227 654	8 913 535
2002	10 669 256	2 356 536	843 335	1 341 514	2 473 817	204 769	9 849 077
2003	11 668 853	2 693 533	847 092	1 389 906	2 680 953	229 287	10 909 332
2004	12 847 691	2 985 769	904 681	1 441 203	2 953 022	286 695	12 057 221

Note. PIH is primary income of households, net; SBM is social benefits other than social transfers in kind; TrORH is other current transfers, received by households; TH is current taxes on income, wealth etc. of households; SCH is social contributions of households, total; TrOPH is other current transfers paid by households; DIH is disposable income of households.

Source: *National Accounts, Hungary, 2003–2004, 2002–2003, 2001–2002, 2000–2001, 1998–2000, 1998–1999, 1995–1997* [2006], [2005], [2004], [2003], [2002], [2001], [1999]. HCSO. Budapest. p. 104–105., 102–103., 102–103., 94–95., 116–117., 120–121.

The data of Table 15 point to a remarkable stability of the distribution, but three important changes can already be ascertained. Social benefits in cash attained their lowest share in 2002 and their share increased in the following two years while taxes and social contributions of households attained their highest share in 2002 and this share decreased in the two following years. These data point therefore to a redistribution of income on behalf of households and to the detriment of government which is in conformity with the findings presented previously. The inference is obvious. If there is no state overspending on collective consumption as it has been shown, and if there is a decrease in the share of taxes and social contributions paid by and an increase in the share of social benefits in cash obtained by households within households' disposable income as it has been presented here, the resulting budget deficit is obviously not the consequence of state overspending on collective consumption but redistribution on behalf of households in the form of decrease of taxes and increase of social benefits in cash.

Table 15

Distribution of households' disposable income, 1995–2004
(percent)

Year	PIH%	SBM%	TrORH%	TH%	SCH%	TrOPH%
1995	107.6	25.6	13.6	11.6	25.1	10.2
1996	108.2	22.6	10.0	11.9	23.1	5.9
1997	109.8	22.2	10.4	11.5	24.1	6.8
1998	109.2	23.1	8.9	11.5	24.6	5.2
1999	109.1	23.4	8.4	12.1	23.7	5.2
2000	116.9	22.8	8.1	13.0	24.7	4.8
2000	109.3	22.4	7.7	12.8	24.2	2.5
2001	109.1	22.2	9.3	13.3	24.8	2.6
2002	108.3	23.9	8.6	13.6	25.1	2.1
2003	106.9	24.7	7.8	12.7	24.6	2.1
2004	106.6	24.8	7.5	12.0	24.5	2.4

Note. PIH% is primary income of households, net, in percent of disposable income of households; SBM% is social benefits other than social transfers in kind in percent of disposable income of households; TrORH% is other current transfers, received by households in percent of disposable income of households; TH% is current taxes on income, wealth etc. of households in percent of disposable income of households; SCH% is social contributions of households, total, in percent of disposable income of households; TrOPH% is other current transfers paid by households in percent of disposable income of households.
Source: Calculated from Table 14.

The same conclusions are confirmed and completed by the same data presented in a rearranged form in Table 16.

The first four columns of Table16 present the time series of social benefits in cash, disposable income of households, social transfers in kind and adjusted disposable income of households. As it has already been shown, social benefits in cash are, obviously, part of disposable income, while adjusted disposable income is obtained by adding social transfers in kind to disposable income. The last two columns present social benefits in cash and social transfers in kind as the percentage of disposable income of households. Both ratios decrease until 2000 but show a remarkable increase after this year. These increases and particularly the increase of social benefits in cash represent a redistribution of income in favour of households, and along with the decrease of the share of taxes and social contributions paid, account for at least the greatest part of the budget problems of the last years of the time period studied.

To analyse our findings we can state the following. On the one hand there is no redistribution in favour of government but rather a redistribution in favour of households in the form of decreasing taxes and social contributions and increasing social benefits and social transfers. It is obvious that the decrease of taxes and social contributions and the increase of social benefits and social transfers cannot go hand in hand as this leads to budget deficit what actually happened. There are two ways to solve this problem: to raise taxes and social contributions or to decrease social benefits and social transfers. The common view is that the only way out is the decrease of social benefits and social transfers, i.e. the retreat of the welfare state. This view is contrary to the symmetry of the problem and has no theoretical foundation. Further considerations are needed to determine which of the two possibilities or what a combination of them is preferable.

Table 16

Disposable income of households, social transfers and social benefits, 1995–2004
(percentages)

Year	SBM	DIH[a]	STrK	DIHA	SBM%	STrK%
1995	910 459	3 559 409	763 186	4 322 595	23.8	21.4
1996	992 604	4 382 986	890 760	5 273 746	20.9	20.3
1997	1 158 386	5 219 558	1 063 885	6 283 443	20.2	20.4
1998	1 405 781	6 098 247	1 288 607	7 386 854	21.1	21.1
1999	1 583 399	6 753 807	1 445 511	8 199 318	21.5	21.4
2000	1 748 658	7 674 303	1 607 786	9 282 089	20.4	21.0
2000	1 755 180	7 825 117	1 626 130	9 451 247	20.5	20.8
2001	1 982 249	8 913 535	1 903 416	10 816 951	20.4	21.4
2002	2 356 536	9 849 077	2 321 511	12 170 588	22.1	23.5
2003	2 693 533	10 909 332	2 749 726	13 659 058	23,1	25.2
2004	2 985 769	12 057 221	2 970 654	15 027 875	23.2	24.6

Note. DIH is disposable income of households; SBM is social benefits other than social transfers in kind; DIHA is adjusted disposable income of households; STrK is social transfers in kind; SBM% is social benefits other than social transfers in kind in percent of disposable income of households; STrK% is social transfers in kind in percent of disposable income of households.

Source: *National Accounts, Hungary, 2003–2004, 2002–2003, 2001–2002, 2000–2001, 1998–2000, 1998–1999, 1995–1997* [2006], [2005], [2004], [2003], [2002], [2001], [1999]. HCSO. Budapest. p. 104–105., 102–103., 102–103., 94–95., 116–117. 120–121.

The most important considerations are the following. First, it is generally accepted that future development can be promoted first of all by better health and education, and this requests a high level of social benefits in kind and the maintenance of the welfare state. Second, the argument in favour of maintaining or increasing household money incomes is politically motivated. Third, it is obvious that the decrease of the share of both taxes and social contributions and also of social welfare expenditures in the broader sense increases the inequality of income distribution. The lowering of taxes benefits the social strata with higher incomes while the reduction of social benefits and transfers is against to the interests of the lower income social groups. It is generally accepted, however, that greater inequality is disadvantageous for development. The supporters of increasing income inequality are therefore working against growth even if they pretend that their policy recommendation are growth promoting. Finally, this is, obviously, not a technical problem which can be solved by experts but a problem connected with value judgements. Consequently the author, even if he holds the view that expenditures on health and education must be maintained or even increased, cannot propose anything else than to inform the public in a correct way about the problems, the possible solutions and their consequences.

6. Savings and investments of households

Some space must be devoted to the third element of the triple deficit, the inadequacy of domestic savings. We have already seen that government is dissaving, the analysis of corporate savings is difficult to carry out for several reasons (e.g. unavailability of data) so household savings will be analysed here not in all detail. The most important data are presented in Table 17.

Table 17

Disposable income, saving, capital transfers, gross capital formation and net lending of households, 1995–2004
(million HUF at current prices)

Year	GDP	DIH	SH	CTrH	GCFH	LH
1995	5 614 042	3 565 712	598 640	41 701	277 441	362 900
1996	6 893 934	4 367 488	857 889	56 748	347 537	567 100
1997	8 540 669	5 205 091	1 019 289	42 614	406 803	655 100
1998	10 087 434	6 112 666	1 176 007	24 110	405 917	794 200
1999	11 393 499	6 755 823	1 056 369	22 033	435 502	642 900
2000	13 150 766	7 693 926	1 137 710	36 563	505 673	668 600
2000	13 272 167	7 881 646	1 121 138	78 787	677 498	522 127
2001	14 989 800	9 037 644	1 290 591	105 409	867 507	525 493
2002	16 915 259	9 995 602	1 175 432	98 756	996 948	277 240
2003	18 650 746	10 909 332	923 596	121 565	1 131 111	–85 960
2004	20 429 456	12 057 221	1 323 719	102 908	1 317 421	109 206

Note. GDP is gross domestic product; DIH is disposable income of households; SH is saving of households; CTrH is capital transfers of households, net; GCFH is gross fixed capital formation of households; LH is lending of households, net.

Source: *National Accounts, Hungary, 2003–2004* [2006]. HCSO. Budapest. p. 13. 106–107.

Data of the third column of Table 17 show that savings of households at current prices increased marginally between 1997 and 2004 while, as data of the second column show, disposable income of households almost doubled in the same period. This must mean that the rate of saving decreased heavily in this time. This development is highly disadvantageous as the high rate of private savings is the obvious prerequisite of dynamic economic growth. The problem becomes even more pronounced if household savings are compared with gross fixed capital formation i.e. with residential building of households. The last column of Table 18 shows that net lending of households – the difference of household savings and households' gross fixed capital formation presented in this column – practically disappeared by the end of the period studied. Net lending of households can also be compared with capital transfers of households appearing in the fifth column. It can be seen that in 2004 capital transfers obtained by households were practically equal to net lending of households which means that households obtained as much from the other sectors of the economy as they allotted to these other sectors in the form of net lending.

The interpretation of these finding is obvious. On the one hand, net lending of households ought to finance both budget deficit and the greatest part of private sector investments. Zero net lending of households is a situation which cannot be maintained. On the other hand, it is obvious that zero net lending of households is not the consequence of state overspending or of budget deficit but an independent problem that cannot be solved by reducing the budget deficit and cutting state expenditures.

The results are even more striking if data are shown in percent of the GDP as it is done in Table 18.

Table 18

Disposable income, saving, capital transfers, gross capital formation and net lending of households, 1995–2004
(percent)

Year	DIH%	SH%	CTrH%	GCFH%	LH%
1995	63.5	10.7	0.7	4.9	6.5
1996	63.4	12.4	0.8	5.0	8.2
1997	60.9	12.1	0.5	4.8	7.7
1998	60.6	11.7	0.2	4.0	7.9
1999	59.3	9.3	0.2	3.8	5.6
2000	58.5	8.7	0.3	3.9	5.0
2000	59.4	8.5	0.6	5.1	3.9
2001	60.3	8.6	0.7	5.8	3.5
2002	59.1	7.0	0.6	5.9	1.6
2003	58.5	5.0	0.7	6.1	-0.5
2004	59.0	6.5	0.5	6.5	0.5

Note. DIH% is disposable income of households in percent of GDP; SH% is saving of households in percent of GDP; CTrH% is capital transfers of households, net, in percent of GDP; GCFH% is gross fixed capital formation of households in percent of GDP; LH%: is lending of households, net, in percent of GDP.

Source: Calculated from Table 17.

It can be seen that household savings decreased from a little more than 10 percent of the GDP to 6.5 percent, and that they became practically equal to residential construction of households leading to the disappearance of net lending of households.

These results add much to what has been told previously. It has been shown that general domestic overspending and overspending of the households must be stopped. There are two ways to attain this: to reduce households' incomes and to increase households' savings. It is obvious that the latter way is more advantageous for the households. The increase of household savings and lending is therefore not only a necessity but also advantageous for the households in the longer run.

References

DORNBUSCH, R.[1988]: Balance of payments issues. In: *Dorbusch, R. – Helmers, F. R. L.* (ed.): *The open economy*. Oxford University Press. Oxford.

SZAKOLCZAI GY. [2005a]: A magyar gazdasági növekedés és felzárkózás kulcsa: az exportorientált gépipari fejlesztés. *Statisztikai Szemle*. Vol. 83. No. 1. p. 5–23. old.

SZAKOLCZAI GY. [2005b]: A folyó fizetési mérleg kumulálódó hiánya és a hiány finanszírozásának lehetőségei. *Statisztikai Szemle*. Vol. 83. No. 3. p. 238–357.

Structure of agricultural holdings in Hungary in years 2000, 2003 and 2005

László Csorba,
head of Census Section of the HCSO
E-mail: laszlo.csorba@ksh.hu

In order to provide the necessary information for conducting Common Agricultural Policy (CAP), Member States are obliged to carry out regularly Farm Structure Surveys (FSS). FSS serves as a backbone of the whole agricultural statistics and also indispensable for the domestic agricultural policy-makers. Nowadays the information collected through FSS is essential for preparing the Rural Development Policy (RDP), which accompanies and complements CAP. The appearance of RDP can be explained by the significance of the complex relation between agricultural activity and the environmental and social aspects, as well as of problems coming from industrial agriculture.

Hungary carried out Agricultural Census in 2000 (AC 2000) and a Farm Structure Survey in 2003 and 2005. This paper presents the main characteristics of the agricultural holdings in years 2000, 2003 and 2005, the prevailing trends in the period of 2000 to 2005 and the conclusions of the analysis.

KEYWORDS:
Agricultural statistics.

Natural conditions, the climate, the location, the water supply and the soil provide excellent opportunity for agricultural production in Hungary. In the history of the past hundred years the agriculture was the locomotive of the economic development in certain times in Hungary, and it provided a level of food supply exceeding the demand of population, thus a considerable surplus of agricultural products was sold on export markets every year.

The proportion of agricultural area in Hungary is the second highest among European countries after Denmark. In 2005 the area under agricultural cultivation was nearly two thirds of the total area and considering the per capita agricultural land area, so Hungary belongs to the group of European countries with the highest rate. Coinciding with the international tendencies, the rate of people engaged in agriculture compared to the total economically active population and the share of agriculture in the total GDP decreased continuously in the past years, meantime the policy makers have become aware of the complex relation between the agricultural activity and the environmental and social aspects, as well as of the problems associated with industrial agriculture.

Farm Structure Survey (FSS) is a coherent series of statistical surveys on the structure of agricultural holdings carried out to meet the information needs of the Community institutions conducting the Common Agriculture Policy (CAP). Every member state is obliged to implement at regular intervals farm structure surveys: at least every ten years a census (full-scope survey) and between the censuses at 2-3 year intervals structural surveys based on the census. The majority of the statistical surveys on agriculture is based on the farm structure surveys. (FSS is regarded as the backbone of agricultural statistics.) Hungary joined the system of FSS and carried out Agricultural Census in 2000 (AC 2000) and a Farm Structure Survey in 2003 (FSS 2003).

Far reaching and considerable social and economic changes have been taking place in the Hungarian agriculture since 1989. To follow and analyse the rapid structural changes, the decision-makers, scientific researchers, farmers, international organisations and other interest groups demand reliable and detailed statistics about this sector. The information collected through FSS is indispensable for preparing the different measures to be taken on Hungarian agriculture and supports to elaborate the rural development policy[1] focusing on such problems as ensuring the sustainable development of rural areas. As the "second pillar" of CAP the rural development policy accompanies and complements it.

[1] Rural development policy is managed through legal instruments and absorbs approximately 15 percent of the total CAP budget. The main headlines of the new regulation are the following: enhancing competitiveness of farming and forestry, environment and countryside, improving quality of life and diversification of the rural economy.

Regarding the Council Regulation No 2467/1996 the Hungarian Central Statistical Office carried out a Farm Structure Survey (FSS 2005) between 14 and 28 November 2005. The implementation of FSS 2005 was stipulated by Governmental Decree No. 303/2004 (XI.2.), relating to the National Program of Statistical Data Collection (NPSDC).[2] This survey was the first and of high priority after the EU accession to provide a credible and accurate picture about the Hungarian agriculture. The relevant EU regulation specifies the agricultural holding constituting the target population of FSS. This definition was adapted in Hungary and in practice two main groups of farming can be distinguished as follows: the private holding and the agricultural enterprise.

An enterprise engaged in any kind of agricultural activity regardless of its size is the agricultural enterprise. The selection of agricultural enterprises is based on the information from Business Register. Household engaged in any agricultural activity reaching a physical threshold at the reference time of the survey is the private holding. The physical threshold[3] applied in FSS 2005 fits the strict coverage criterion of 571/88 EC which says: fixing the threshold at a level excluding only the smallest holdings which together contribute 1 percent or less to the total Standard Gross Margin (SGM).[4]

The agricultural enterprises were surveyed on full-scope basis by mail, meanwhile in case of private holdings a sample survey was designed in FSS 2005 as follows: within the randomly selected survey districts all the active private holdings on 1st December 2005 were observed by the enumerators completing the questionnaires during a face-to-face interview. 8 400 agricultural enterprises and 960 thousand private holdings were observed in AC 2000. The same figures were 7 800 and 766 thousand in FSS 2003 and 7 900 and 707 thousand in FSS 2005. The sharp decline in the number of private holdings also reflects the rapid structural changes still characterizing the Hungarian agriculture. Besides the agricultural enterprises and private holdings a number of other organisations and institutions such as national parks, water management authorities, airports, etc. – passive users of land not for production – were not included in the observation, whose agricultural activity was negligible and the agricultural activity in kitchen and home gardens was also noticeable.

This paper presents the main characteristics of the agricultural holdings in 2000, 2003 and 2005 and the prevailing tendencies in the following topics:

[2] All the surveys of a given year are included in NPSDC approved each year in a Government Decree.

[3] According to the physical threshold of FSS 2005 the private holding on 1st December 2005 uses at least 1500 m^2 of productive land area (including jointly or severally arable land, kitchen garden, orchard, vineyard, meadow, pasture, forest, fish-pond, reed); 500 m^2 of orchards or vineyards, jointly or severally or; 100 m^2 of land area under cover; 50 m^2 of mushroom area; has at least one head of bigger animals, such as cattle, pig, horse, sheep, goat, buffalo, emu, ostrich, donkey; 50 heads of poultry jointly or severally, such as hens, geese, ducks, turkeys, guineafowls; 25-25 heads of rabbits, furry animals, pigeons for slaughter, or 5 bee colonies. This is the same definition used in FSS 2003 and similarly to the one applied in AC 2000, the difference is only that the agricultural service were included in the threshold definition of AC 2000.

[4] Roughly defined SGM = sales – variable costs; this characteristic is very near to Gross Value Added and the core of the EU typology of agricultural holdings.

– number of agricultural holdings,
– type of farming and objective of production,
– agricultural land use,
– livestock,
– farm size by Gross Production Value (GPV),
– farm labour force,
– tractors, cultivators, machinery and equipment,
– non-agricultural activity (gainful activities other than agriculture).

1. Number of agricultural holdings

In the past three decades the number of agricultural enterprises manifested hectic changes. Following the mergers of enterprises for large-scale production in the 1970s and 1980s, the appearance of organisations of new types in the 1990s resulted in an expansion in the number of agricultural enterprises. In the years after the turn of the millennium the difficulties in farming caused again a moderate reduction. On the 1st of December 2005 there were around 7 900 agricultural enterprises active in Hungary and it means that the number of agricultural enterprises has hardly changed since 2003.

Table 1

Number of agricultural holdings

Year	Private holdings		Agricultural enterprises		Total		Number of people occupied in agriculture* (thousand)
	number (thousand)	index (percent) 1972=100	number (thousand)	index (percent) 1972=100	number (thousand)	index (percent) 1972=100	
1972	1842	100.0	6.1	100.0	1848	100.0	1167
1981	1530	83.1	1.4	23.0	1531	82.9	984
1989	1435	77.9	1.5	24.6	1437	77.8	840
1991	1396	75.8	2.6	42.6	1398	75.7	710
2000	959	52.1	8.4	137.7	967	52.3	252
2003	766	41.6	7.8	127.9	773	41.8	215
2005	707	38.4	7.9	129.5	715	38.7	194

* Labour Force Survey.

In the past three years the number of private holdings shrank continually. Between 1991 and 2000 the fall was due to the cessation of farm gardening.[5] The 26 percent

[5] Agricultural activity, which was carried out by the members of agricultural co-operatives for their own benefit.

drop in the number of holdings in the years following the turn of the millennium can be explained by the lack of capital, the insufficient agricultural qualification and aging of private farmers as well as the unfavourable structure of land use. The tendency between 2000 and 2003 has continued, although the decline slowed down in the two past years: both the number of holdings engaged in animal husbandry and the livestock reduced, many households still stopped or diminished their agricultural activity. Simultaneously, the concentration of land use decelerated, fewer holdings cultivated hardly larger agricultural land area, while in the meantime the number of private farmers using agricultural land area over 300 hectares increased slightly. (See Table 1.)

2. Type of farming and objective of production

Classifying the agricultural holdings by activities, three main types of farming can be distinguished as follows: crop, livestock and mixed farming. The crop farming means that the agricultural activity of the holding based on land use solely, the livestock farming includes only animal husbandry, while the mixed farming covers both of them.

It is evident, when the agricultural holding makes decision on which type of activity should be preferred, it asses all the factors influencing the profitability. In the period of 2000 to 2005 the livestock farming faced more difficulties than the crop farming and this situation has not changed since the EU accession, the difference even has widened further: on the one hand SAPS (Simplified Area Payment Scheme) and the additional domestic subsidies have improved the profitability of crop farming, on the other the imports from EU member states have put the agricultural holdings engaged in animal husbandry under price pressure. After the vigorous shift to crop farming characterised agricultural enterprises between 2000 and 2003, the structure has not changed on the merits since 2003. Approximately three-quarters of agricultural enterprises has continued to specialize for crop farming, whilst the proportion of those engaged solely in animal husbandry has reduced slightly to 9 percent.

Table 2

Distribution of the number of agricultural holdings by type of farming, 2000 2003 and 2005

(percent)

Type of farming	Private holdings			Agricultural enterprises			Total		
	2000	2003	2005	2000	2003	2005	2000	2003	2005
Crop farming	39.7	37.2	46.7	42.0	71.6	73.6	39.7	37.6	47.0
Livestock	22.0	24.6	21.6	9.4	11.0	9.4	21.9	24.4	21.5
Mixed	38.3	38.2	31.7	48.6	17.4	17.0	38.4	38.0	31.5
Total	*100.0*	*100.0*	*100.00*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*

In case of private holdings the specialization towards crop farming also emerged in the past two years, however its intensity was much smaller than in the agricultural enterprises between 2000 and 2003. Compared to 2003 the proportion of crop farming increased by 9 percentage point, at the same time the proportion of livestock and mixed farming lessened by 3 and 6 percent, respectively. The private holdings could be classified in 2005 as follows: 47 percent carried out crop farming, around one fifth was engaged solely in livestock farming and the rest (32 percent) implemented mixed farming. (See Table 2.)

Focused on the distribution of private holdings at regional level, the change in type of farming showed a various picture in the past two years. Considering crop farming, the largest widening in Northern Great Plain (13 percentage point), the smallest in Central Hungary (2 percentage point) were measured. The proportion of livestock farming expanded by 5 percent in Central Hungary, in the meantime the reduction was the same in Northern Great Plain. The lowering of proportion relating to mixed farming by regions was between maximum of 8 percent in Northern Great Plain and minimum of 4 percent in Western Transdanubia.

The question referring to the objective of production was answered by the private farmers during the face-to-face interviews, thus the private holdings also could be grouped in this respect as subsistence holding, where the agricultural products are produced only for own consumption, semi-subsistence holding, which sells a certain part of the agricultural products on the market beyond the own consumption and market-oriented holding, which runs as a business aiming at profit in the first place.

In 2000 and 2003 nearly 60 percent was the rate of subsistence holdings, however it has declined by 9 percent by 2005. The proportion of market oriented holdings was a mere 8 percent in 2000, but it has risen continuously since 2000 and attained 15 percent in 2005. The rate of subsistence holdings was the highest in livestock farming (78 percent) and the smallest in mixed farming (38 percent) in 2005. Nearly one fifth of crop farming holdings was market oriented, whereas mere a few percent of holdings engaged in livestock farming produced primarily for the market.

Within the framework of National Rural Development Plan (NRDP) financial support is granted to private farmers who can submit a solid business plan targeting at converting their subsistence or semi-subsistence holding to market-oriented.

Table 3

Distribution of the number of private holdings by the objective of production, 2000, 2003, 2005
(percent)

Objective of production	2000	2003	2005
Subsistence	60.5	59.4	51.4
Semi-subsistence	31.5	29.0	33.1
Market oriented	8.0	11.6	15.5
Total	*100.0*	*100.0*	*100.0*

The share of subsistence holding differed from each other significantly by regions in 2005. It was over 60 percent in Central Hungary and in Central and Western Transdanubia, whilst these figures just exceeded 40 percent in Northern and Southern Great Plain. The market oriented holdings worked in the greatest proportion (20 percent) in Northern and Southern Great Plain, but their proportion was below 10 percent in Central Transdanubia.

3. Agricultural land use

Before dealing with the detailed statistics, some phenomena are needed to mention: the ownership and use of land separated sharply from each other, whilst the number of farmers living in urban areas increased in the early 1990's in Hungary and the land cultivated by agricultural enterprises very often divided into several parcels.

In 2005 the agricultural holdings cultivated 5 710 thousand hectares productive land area[6] of which 4 268 thousand hectares were agricultural land. The structure of the agricultural land area by land use categories in 2005 was the following: 84 percent of arable land, 11 percent of grassland, 2 percent of vineyard and orchard, respectively and 1 percent of kitchen garden. The land area is divided between the two main groups of farming in 2005 as follows: 40 percent of productive land and 50 percent of agricultural land area was cultivated by the private holdings. 93 percent of the private holdings and nearly 90 percent of the agricultural enterprises used productive land. Considering agricultural land, the same figures are 93 and 73 percent.

The number of agricultural holdings using productive land area has reduced by 7 percent since 2003. If this reduction is divided in two main groups of farming, the number of agricultural enterprises has grown by 2 percent, while in case of private holdings a 7 percent fall has happened since 2003. At a modest pace, but the tendency evolved between 2000 and 2003 has continued: the average productive land area of agricultural enterprises lessened by 3 percent, and that of private holdings increased by 3 percent. Accordingly, the average size of the productive land area reached 487 hectares in the agricultural enterprises and around 3.5 hectares in the private holdings in 2005.

In the past two years contradictory trends appeared in the number of agricultural holdings using agricultural land and arable land area, respectively. In both main groups of farming the number of holdings using arable land increased slightly. Simultaneously, the arable land area cultivated by agricultural enterprises enlarged by nearly 3 percent, meantime approximately 2 percent lowering was measured in case of private holdings.

[6] Productive land = Agricultural land (arable land, grassland, vineyard, orchard, kitchen garden) + Forest + Reed + Fish-pond ; FSS aggregates of different land use categories cover only the area that can be connected to the agricultural holdings as users, at the same time the published aggregates of current statistics also contain and reflect some additional expert estimations.

Some changes in the use of plantations have taken place in both main groups of farming since 2003. On the one hand the number of agricultural enterprises using orchard has increased by 7 percent, on the other this figure showed an 8 percent fall in the private holdings. The orchard area per holding lessened by 0.5 percent in the agricultural enterprises and by 18 percent in the private holdings. The number of holdings using vineyard and the average size of vineyard area by two main groups of farming moved reversely in the past two years. The number of agricultural enterprises increased by 14 percent, at the same time the drop reached nearly one-quarter of private holdings. The average plantation area in this highly labour intensive branch increased by 4 percent in the agricultural enterprises and reduced by 5 percent in case of private holdings.

The grassland area lessened in both main groups of farming in the last two years. Around one fifth of agricultural enterprises and merely 6 percent of private holdings cultivated such areas in 2005. The size of grassland used by enterprises fell significantly by 20 percent compared to 2003, meanwhile the reduction was much smaller in case of private holdings (3 percent). (See Table 4.)

Table 4

Size of land area per agricultural holding by land use categories, 2000, 2003 and 2005

(hectare/holding)

Denomination	Private holding			Agricultural enterprise		
	2000	2003	2005	2000	2003	2005
Arable land	3.11	4.42	4.22	506.93	384.88	385.78
Grassland	2.86	4.12	5.93	161.20	146.72	167.31
Agricultural land area	2.51	3.10	3.21	533.49	384.11	373.55
Productive land area	2.74	3.33	3.44	663.00	503.09	486.82

Significant differences between the two main groups of farming have remained in the structure of land use and the most important characteristics have continued to hold since 2003 as follows. 87 percent of the agricultural enterprises used their productive land exceeding one hectare, which altogether amounted to 99 percent of the total productive land area. 22 percent of agricultural enterprises used productive land area over 300 hectares, which accounted for around 88 percent of the total area. The significant part of agricultural activity was implemented in the group of enterprises using productive land over 1 000 hectares (merely 8.9 percent of all the enterprises, but they use the 72 percent of the total area).

73 percent of the private holdings have continued to use productive land area below 1 hectare, which amounted to 5 percent of the total productive land. Whilst the proportion of private holdings using productive land area over 50 hectares hardly was larger than 1 percent, their productive land area was nearly 40 percent of the to-

tal. The cultivated productive land area exceeded 300 hectares occurred in the case of a very few private holdings.(See Figure 1.)

Figure 1. Concentration of productive land area per agricultural holding, 2005

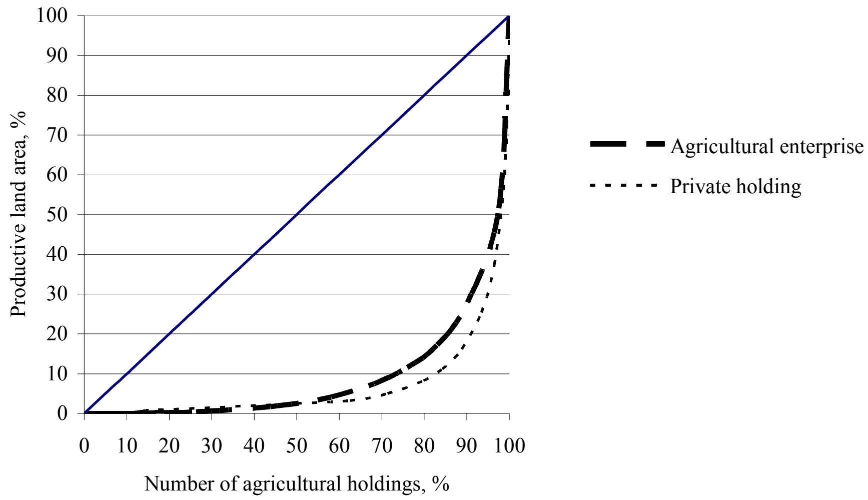


Table 5

Distribution of agricultural land area involved in organic farming by land use categories, 2005
(percent)

Denomination	Arable land	Orchard	Vineyard	Grassland	Other	Total
Organic farming	43.12	1.20	0.45	50.70	4.53	100.00
Under conversion to organic farming	34.69	2.26	0.55	59.27	3.23	100.00
Total	*40.24*	*1.56*	*0.48*	*53.62*	*4.10*	*100.00*

In compliance with the EU regulation, the agricultural land area on which organic farming production methods were applied and the agricultural area which were under conversion to organic farming amounted together to 123 thousand hectares in Hungary The largest part of this land was utilised as arable land (40 percent) or grassland (54 percent). Considerable funds are allocated in the framework of EU rural development policy to support the European farming sector to move towards more environmentally sustainable practices. The rural funds also play a key role in helping to maintain farming systems, which, although they may have limited economic viability, play an essential role in land management. Much of Europe's biodiversity and natural resources management depend on such farming systems. (See Table 5.)

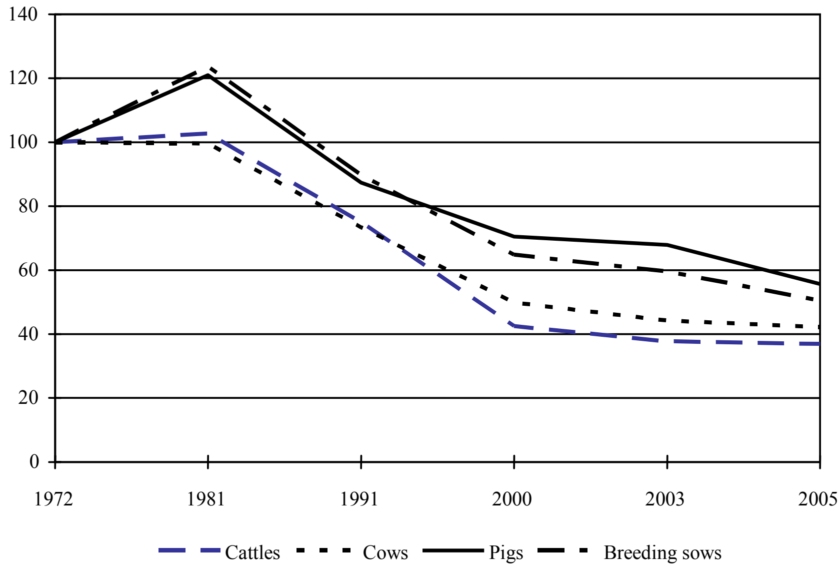
4. Livestock

2 078 agricultural enterprises (26 percent) and 377 thousand private holdings (53 percent) were engaged in animal husbandry reaching the physical threshold, of which 156 agricultural holdings applied organic production methods to the animal production in 2005. In this sector, compared to 2003, the number of enterprises and private holdings shrank by 6 and nearly 20 percent, respectively. In the latter case this drop resulted mostly from stopping (giving up) the production for own consumption (subsistence holdings).

The cattle stock has kept on reducing since the 1980s. In 2003 714 thousand, in 2005 699 thousand cattle were in Hungary. In 2005 more than two thirds of cattle and cow stock belonged to agricultural enterprises. In 2005 around 60 percent of the 3.8 million pig stock was kept by the agricultural enterprises, including the 65 percent of the breeding sow stock as well. The main reason for a 24 percent decline compared to 2000 was that the private farmers has reduced their stock by more than a million heads. (See Figure 2.)

The sheep stock has grew by 13 percent since 2003 and exceeded 1.4 million heads. Nearly 90 percent of stock was kept in private holdings. Regarding poultry stock in 2005, hardly more than 40 percent of miscellaneous birds (chicken, hen, cocks) and nearly a half of gooses and ducks belonged to the private holdings, while one third of turkeys was kept in the agricultural enterprises.

Figure 2. Cattle and pig stock
(Index: 1972 = 100.0)



The structure of livestock in 2005 by species[7] were dominated by cattle (53 percent) and pig (37 percent) in the agricultural enterprises. In the private holdings the shares of three species were overwhelming: cattle (36 percent), pig (34 percent) and sheep (18 percent). Organic production was characterised by cattle and sheep heeding. In the past two years the number of cattle breeding agricultural enterprises changed hardly anything, whilst the fall in the number of pig keeping was 15 percent. In case of private holdings the drop in the number of cattle and pig keeping farms was intensive: 23 and 27 percent, respectively. After a 27 percent fall in the number of the enterprises keeping sheep in 2003, the increase was 12 percent in the past two years. The developments relating to private holdings are similar: a 16 percent reduction followed by a 4 percent increase. Traditionally, goat breeding has continued to be typical for private holdings. In 2005 the number of enterprises was unchanged relative to 2003. In the past two years the number of holdings keeping gallinaceous birds (hen, chicken etc.) sank by 12 percent in both main groups of farming.

Table 6

Number of agricultural holdings keeping major species and livestock by species, 2000, 2003 and 2005

Major species	Number of holdings			Livestock (thousand heads)		
	2000	2003	2005	2000	2003	2005
Cattle	52 182	32 273	25 108	850.4	714.3	698.8
Pig	484 527	434 731	316 486	5 050.5	4 658.1	3 815.5
Horse	38 079	28 168	28 971	80.6	69.4	72.4
Goat	27 249	19 246	20 648	106.0	80.1	94.1
Sheep	25 094	20 994	21 862	1 287.3	1 280.9	1 442.7
Chicken	596 988	436 784	383 897	42 419.0	34 757.6	27 435.8

Significant differences in the size of livestock emerged in 2005 by species and the form of farming (agricultural enterprise or private holding).

– 40 percent of the cattle breeding agricultural enterprises kept more than 500 heads, at the same time the private holdings mostly kept 1 or 2 heads (35 percent) or 3-9 heads (46 percent).

– Approximately one-quarter of pig breeding agricultural enterprises kept a livestock exceeding 5 000, on the meantime in 16 percent of them the livestock was in the range between 2000 and 4999 heads. Two thirds of pig breeding private holdings kept 1-2 heads, and 27 percent of them had a livestock in the range of 3-9 heads.

At last it is worth of dealing with the problem of agricultural holdings keeping livestock, but having no agricultural area at all. These holdings have to take more

[7] Calculated in livestock-unit.

risk in course of their business, because they are exposed to the possible rise in price and the poor quality of forage. Some of them jeopardize the environment, because their manure storage facilities are not sufficient. The shares of these holdings keeping cattle or pig was 6 and 25 percent, respectively in 2005.

5. Farm size by Gross Production Value (GPV)

In 2005 the agricultural enterprises produced HUF433 billion of GPV[8] calculated at constant price of the year 2000. The share of the agricultural enterprises in the total GPV was 54 percent. The volume index of GPV on fixed basis of 2000 (2000=100.00) to 2005 amounted to 98 percent. The GPV per agricultural enterprise was around HUF55 million.

Regarding the number of enterprises, where the GPV was not more than HUF9.5 million in 2005 the highest proportion (55 percent) was measured, however their contribution to the total GPV was merely 3 percent. The number of the enterprises with GPV above HUF150 million was 731 (9 percent), meanwhile they produced nearly the 70 percent of the total GPV in 2005. (See the different groups in Table 7.)

Table 7

Distribution of number of agricultural enterprises and their GPV by size groups of GPV, 2000, 2003 and 2005
(percent)

Size classes of GPV (1000 HUF)	Number of agricultural enterprises			Gross Production Value*		
	percent					
	2000	2003	2005	2000	2003	2005
≤ 600	24.63	17.04	16.02	0.05	0.07	0.07
601-9500	32.48	39.01	38.91	1.65	2.36	2.44
9501-150000	30.99	34.16	35.85	23.51	26.51	28.37
150001-500000	9.84	7.90	7.25	41.29	36.99	34.54
500001≤	2.06	1.89	1.97	33.50	34.07	34.58
Total	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*

* Calculated at constant price of year 2000.

Examining the period of 2000 to 2005, the number of enterprises under 600 thousand of GPV has dropped by 25 percent and simultaneously their proportion to the total number of enterprises has also decreased from 25 percent to 16 percent, meanwhile their contribution to the total GPV has continued to be very small. Measured

[8] The estimated value of the crop and animal products. It is calculated by multiplying the physical characteristics of land use and livestock (harvested area, livestock by species) and the specific coefficients per area unit or animal unit (these coefficients based on the appropriate procurement and producer prices).

the change in number and in value, the group of enterprises between HUF9.5 and 150 million of GPV has strengthened their position since 2000 (5 percentage point in number and in value as well). This expansion was also justified by 8 percentage point increase in agricultural land area.

In 2005 the private holdings produced HUF365 billion of GPV calculated at constant price of the year 2000. The share of the private holdings in the total GPV was 46 percent. The volume index of GPV on fixed basis of 2000 (100.00=2000) to 2005 amounted to 79 percent, which is the result of the sharp decline in the number of private holdings has been occurred since 2000. The GPV per private holding was around HUF517 thousand.

In 2005 the number of private holdings was the highest in the group of holdings with GPV not more than HUF200 thousand, meanwhile their share in the total GPV was 12 percent. It can be assumed that they were all subsistence holdings, their production did not appear on the market. The next group of holdings producing GPV between HUF201 and 600 thousand can be classified as small holdings. The majority of them sell only the surplus of production left over the own consumption (semi-subsistence holdings). The highest contribution (52 percent) to the total GPV was manifested in the group of holdings with GPV of HUF601–9500 thousand, however their proportion to the total number of holdings reached barely 13.7 percent. The holdings of GPV between HUF9.5–70 million and over HUF70 million can be regarded as large and very large holdings. They have a key role on the market and – mainly whose GPV is over HUF70 million – the abilities to react to the marketing tendencies flexibly. These holdings make their business plan on long term. Unfortunately, the number of holdings with more than HUF70 million of GPV is rather small, they were merely 47 in 2005. (See the different groups in Table 8.)

Table 8

Distribution of number of private holdings and their GPV by size groups of GPV, 2000, 2003 and 2005
(percent)

Size classes of GPV (1000 HUF)	Number of private holdings			Gross Production Value*		
	2000	2003	2005	2000	2003	2005
≤ 200	61.42	62.14	65.97	10.95	10.88	11.54
201–600	23.29	22.45	19.63	16.42	13.99	12.68
601–9500	14.86	14.77	13.74	52.52	50.00	51.54
9501–70000	0.42	0.63	0.66	16.39	20.75	22.15
70001≤	0.01	0.01	0.00	3.72	4.38	2.09
Total	100.00	100.00	100.00	100.00	100.00	100.00

* Calculated at constant price of year 2000.

Focusing on the structural changes over the period of 2000 to 2005, small holdings had a considerable reduction in number and the value of GPV as well (4 percentage points). At the same time, a contradictory tendency could be revealed in the group of large holdings and they have extended their proportion to the total number and to the total GPV of holdings by 0.3 and 6 percentage point, respectively. This increment was also reflected by the physical characteristics, thus a 9 and a 16 percentage point of growth have occurred in their productive land area and in their livestock of cattle.

If the farm size is examined by type of farming, the result is not surprising: nearly one third of mixed farming holdings were above 600 thousand of GPV, whilst the appropriate proportion was not more than 10 and merely 3 percent in case of crop and livestock farming. It corresponds to the assumption that mixed farming is a more complex activity than crop or livestock farming, severally and rather market oriented.

6. Farm labour force

The main characteristics of labour force input are very different in the agricultural enterprises and in the private holdings. Usually, the employees work eight hours per day in the enterprises, whilst just a few hours daily work occurs in many private holdings. For this reason it was needed to apply such a unit of measure (it is called Annual Work Unit, AWU=1800 hours), which makes possible to convert the part-time work into full time to compare the farm work doing in the enterprises and in the private holdings. In 2005 the agricultural output was produced by farm work which equals to the work implemented by 512 thousand full-time farm workers. This is an 11 percent decrease compared to that of 2003.

Applying AWU is a favourable approach to analyse the structure of farm labour force, meantime it is not applicable to compare the agricultural work with work done in the other sections of national economy. To examine the labour force characteristics at the level of national economy, the statistics from Labour Force Survey (LFS) and Survey on Employment (SE) can be used.[9] The problem with the LFS and SE is that the former covers the agricultural activity carried out in private holdings partly, whilst the latter excludes it totally.

Unlike the mentioned LFS and SE, FSS covers the agricultural activity (regardless primarily or secondary) carried out in the enterprises, which can be classified[10] in various industries of the national economy other than agriculture. The principle is similar in case of private holdings and each type of agricultural activity is accounted

[9] The LFS is a sample survey and collecting labour force data on natural persons aged between 15 and 74 years belonging to the selected households. LFS covers only people carrying out agricultural activity in the agricultural enterprises and private holdings as his/her major occupation. LFS statistics showed that the occupied population in 2005 amounted to 3.9 million people in the national economy, of which 194 thousand were occupied in the agriculture, forestry and fishery. This is 5 percent of the total. The SE collects labour force data on people employed by businesses and organisations in the governmental sector, where the number of employees is at least 5. According to SE the number of people employed in the agriculture, forestry and fishery was 98.2 thousand in 2005 which is around 4 percent of the total.

[10] According to the Hungarian Uniform System of Activity Classification by Branches.

in FSS which is implemented within the household exceeding a physical threshold.

In the agricultural enterprises 85 thousand employees on regular, 15 thousand employees on non-regular basis did the work in 2005. The appropriate data of private holdings are 4 and 34 thousand, which means a 36 and 38 percent decline relative to 2003. Parallel with the employees, around 1.3 million people as non-paid family labour were involved to some extent in the agricultural work in the private holdings.

86 percent of managers and 73 percent of employees were male in the agricultural enterprises. About three-quarters of farmers were also male in the private holdings and this proportion has not changed since 2000.

In the period of 2000 to 2003 the private farmers were getting older and older justified by the proportion of private farmers aged over 54 years increased, whilst the proportion of those below 34 years lessened. Since 2003 on the one hand the rate of private farmers aged over 54 years has continued to widen (52 percent), on the other the proportion of the youngest ones has moved to a favourable direction and grown from 6 to 8 percent. If the distribution is examined by gender as well, contradictory tendencies could be revealed in the past two years.

Table 9

Distribution of private farmers by age groups and gender, 2000, 2003 and 2005
(percent)

Age groups	Male	Female	Together	Male	Female	Together	Male	Female	Together
	2000			2003			2005		
14–34 years	10.2	5.9	9.1	6.5	5.4	6.3	8.5	5.3	7.8
34–54 years	45.0	28.4	41.0	44.7	34.6	42.7	44.3	28.4	40.5
55 years ≤	44.8	65.7	49.9	48.8	60.0	51.0	47.2	66.3	51.7
Total	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*

The number of people working as non-paid family labour in private holdings has reduced slightly by 1 percent (11 thousand people) since 2003. The average number of family labour force per holding has grown from 1.8 person to 1.9 person.

Parallel with the reduction of the number of private holdings and a moderate increase of their average productive land area, the number of days worked per private farmer lowered from 82 to 75 days in the past two years. The male farmers worked by average 18 percent (78 day) more than the female. Of 1 000 private farmers 610 carried out the agricultural activity as his/her major occupation in 2005 and this figure was nearly the same in 2003. Simultaneously, the proportion of holders having other gainful activity grew from 36 to 38 percent. 76 percent of female and 57 percent of male of the labour force considered agriculture as their major occupation.

Unfortunately, the situation on agricultural qualification of private farmers has not improved since 2000. Merely 2 percent had graduated from a college/university specialized for agriculture, 6 percent had finished an agricultural vocational school, so most of them could apply only the many years of working experience during the agricultural activity. Generally, the agricultural qualification of male farmers was higher than that of females. The share of farmers completed at least secondary school amounted to 9 percent in case of males, on the meantime it did not exceeded the 3 percent in case of females. In the framework of rural development policy the human potential is targeted by a set of measures: organising training courses, supporting young farmers parallel with encouraging early retirement of farmers, setting up farm advisory services etc.

Table 10

Some figures of private farmers by gender, 2000, 2003 and 2005

Denomination	Male farmers			Female farmers			Total		
	2000	2003	2005	2000	2003	2005	2000	2003	2005
	Agricultural qualification, percent								
Secondary school	7,1	6,7	6,7	2,3	1,9	1,9	5,9	5,5	5,6
Collage/university degree	2,3	2,6	2,2	0,6	0,7	0,6	1,9	2,1	1,8
	Form of activity, percent								
Unemployed	5,9	3,3	3,7	3,3	2,3	2,9	5,3	3,0	3,5
Retired	42,9	42,5	44,3	66,0	61,7	65,5	48,4	47,1	49,3
	No other gainful activity, percent								
	52,7	57,7	56,5	72,0	75,0	76,3	57,3	61,9	61,2
	Average annual days worked, days								
	79	86	78	65	70	66	75	82	75

7. Tractors, cultivators, machinery, and equipment[11]

In the period of 2000 to 2005 significant support from Special Accession Programme for Agriculture and Rural Development (SAPARD), Agricultural and Rural Development Operational Programme (ARDOP) and National Rural Development Plan (NRDP) reached the agricultural holdings in Hungary to enlarge or modernise their machinery. In 2005 the numbers of tractors used in the agriculture exceeded that of the year 2000 by 6 percent. The figures on power rating reflected that the tractors with relatively low performance have been replaced with heavy-duty ones in both main groups of farming. The number of combine-harvesters has not changed since 2000. Regarding the number of tractors and combines in the enterprises and in

[11] This group of characteristics was not observed in FSS 2003.

the private holdings, respectively, contradictory tendencies have prevailed the period of 2000 to 2005. The number of lorries have dropped by 39 percent since 2000.

Table 11

Number of tractors, combines and lorries by forms of farming
(number)

Denomination	Private holdings		Agricultural enterprises		Total	
	2000	2005	2000	2005	2000	2005
Tractor	87 107	96 922	26 199	23 557	113 306	120 479
Combine	6 453	8 157	5 660	3 980	12 113	12 137
Lorry	15 320	9 462	10 505	6 411	25 825	15 873

8. Non-agricultural activities

According to rural development policy there is a need to promote changes in rural areas by helping agricultural holdings to diversify farming activities towards non-agricultural activities, develop non-agricultural sectors, improve employment and basic services and carry out investments making rural areas more attractive in order to reverse trends towards economic and social decline and desertion of the countryside.

Table 12

Propotion of agricultural holdings involved in non-agricultural activities, 2000, 2003 and 2005
(percent)

Activity	Private holdings			Agricultural enterprises		
	2000	2003	2005	2000	2003	2005
Meat-processing	0.41	0.40	0.81	1.15	0.76	1.31
Milk-processing	0.13	0.62	0.49	0.39	0.38	0.39
Fruit and vegetable processing	0.21	0.50	1.25	1.32	0.78	0.97
Wine-making, wine-bottling	2.9	1.3	0.6	2.14	2.46	3.43
Fodder-mixing	0.06	0.01	0.06	7.02	1.92	1.58
Tourism, catering	0.13	0.06	0.14	3.15	1.93	2.44
Transportation, delivery	0.49	4.53	0.41	11.52	5.77	6.91

Unfortunately, the number of agricultural holdings involved in non-agricultural activities (gainful activities other than agriculture) has continued to be very low. The developments showed that in general a moderate growth occurred in the proportion of the agricultural enterprise except for the activities of fodder-mixing and milk-processing since 2000. The trends relating to the private holdings have been various since 2000: a slight, however the greatest expansion has taken place in fruit and vegetable processing as well as in meat-processing, while a sharp decline could be observed in transportation and delivery. (See Table 12.)

9. Conclusions

Generally descending tendency characterised the period of 2000 to 2003 has continued in the Hungarian agriculture, although it slowed down in the two past years: both the number of holdings engaged in animal husbandry and the livestock reduced, many households still stopped or diminished their agricultural activity.

The structure of land use and the farm size are unfavourable, they are economically not optimal, agricultural holdings can utilize the advantages of the well-mechanized farming only in a limited way. The concentration process of land use decelerated between 2003 and 2005. The sensitive balance has broken between the crop and livestock farming, thus the production of crop farming exceeds the demands. The activity of agricultural holdings keeping livestock, but having no agricultural land area is risky biologically and environmentally as well. Agriculture suffers from the lack of capital and the insufficient professional agricultural qualification and aging of private farmers.

Ecological agriculture including organic farming is a reaction and alternative to evade some of the strategies associated with industrial agriculture, which have had a damaging effect on rural society and agricultural ecosystems. Rural development policy supports the European farming sector to move towards more environmentally sustainable practices and to maintain farming systems, which, although they may have limited economic viability, play an essential role in land management. Such farming systems can be the base of biodiversity and natural resources management of Europe.

The share of agriculture in the total GDP (3.3 percent in 2004) seems to be not significant at the first glance, however, when one would like to make a thorough analysis about this section of the national economy, the complex relation between the agricultural activity and the environmental and social aspects cannot be ignored. Accordingly, such aims in the new EU rural development policy can be found as ”ensuring that environmental issues are taken into account, developing complementary and alternative activities that generate employment, with a view to slowing the depopulation of the countryside and strengthening the economic and social fabric of rural areas, improving living and working conditions, promoting equal opportunities guarantying the safety and quality of foodstuffs etc.”[12] The conclusion is that the optimal structure of the agriculture based on purely economic considerations does not agree with that optimum including also environmental and social aspects, thus being desirable for the society.

[12] Enlargement strengthens the case for more support for rural development in the EU [2000] EurActiv 2000–2005. (www.euractiv.com)

Is there causal relationship between the value of the news and stock returns?

Zsolt Bedő

PhD candidate, teaching assistant of the University of Pécs

E-mail: zsoltbedo@ktk.pte.hu

Gábor Rappai

CSc., Habil. associate professor of the University of Pécs

E-mail: rappai@ktk.pte.hu;

We introduce the application of event study methodology in conjunction with the theory of information entropy. This new mean of time series analysis creates connection between the value of the news announcement and the behaviour of stock prices. As we make distinction between the rates of "surprise" of a particular news announcement the corresponding return behaviour can be assessed more efficiently. We find in our empirical application that the announcements of profitability ratios at different entropy values explain the behaviour of cumulated abnormal returns.

KEYWORDS:
Financial applications, financial and stock market.

In the case of modelling time series (of share prices) the researcher is facing two contradictory statistical characteristics: on one hand time series are long with high frequency (advantageous), but on the other their volatility (variance) is rather high (disadvantageous). In order to overcome the problem of high volatility countless models have been proposed in the literature. Methods concerned with the variance of time series can be divided into two groups: in the *first* improvement of model specifications or estimation of parameters can be found, while in the *second* the transformation of time series into a more homogenous set is proposed. Event study methodology belongs to the second group, which grants the foundation of behavioural finance.

The foundation of event study methodology was introduced in our prior study (*Bedő–Rappai* [2004]), when we demonstrated the determination of abnormal and cumulative abnormal returns and also the means of event analysis. Our sole objective in that study was to scrutinize the impact of "good" and "bad" events, whether they effect share prices differently. We based our hypothesis on the findings of *Hong, Lim* and *Stein* [2000], which highlighted the differing behaviour of share prices at the occurrence of positive and negative news. At the end of that study we took an attempt to build trading strategy based on our finding, with little success.

In this study we take the analysis a step further by connecting the behaviour of the share price and the impact of a particular event, by assessing the relationship between the rate of surprise at the release of company data and the price reaction. To assess the behaviour of share price we calculate cumulative abnormal returns – spread between the realized and estimated returns – around the release of return on equity (ROE) and earnings per share (EPS) figures. The flow of the cumulative abnormal return determines whether investors are able to correctly "price" the particular company share. Using past ROE and EPS data we calculate the probability of occurrence of the last ROE and EPS figures, which allow us to specify the corresponding entropy value. This entropy value enables us to state whether the announcement of ROE and EPS ratios were considered a "surprise" or not. By observing entropy values in conjunction with cumulative abnormal returns conclusions can be drawn concerning the effect of news' "surprise" on price disturbances.

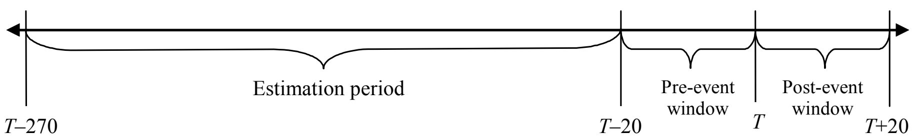
The remainder of the paper is as follows. In the first section we review the basics of event study methodology to calculate abnormal and cumulative abnormal returns. In the second section we introduce the method of event "valuation" to use the distribution of return on equity and earnings per share data in order to establish the entropy indicator. Section 3 introduces the data under analysis, data transformation and indicators. In section 4 results are presented and analysed. We summarize our conclusions in section 5.

1. Review of event study methodology

Assessing the value of an instrument – traded on an open market – around the appearance of particular information allows the researcher to identify the economic meaning of the underlying event. In financial research event study methodology has been used both to determine the value of certain information affecting economic performance and to specify the efficiency of markets in order to draw conclusion regarding the behaviour of market participants.

By applying the method one is able to compare the "normal" return to the one actually is occurring on the open market. The gap between the two returns will reveal the meaning of the information, whether market participants have correctly assessed the upcoming event or have been able to adjust to the new circumstances. The release of the information concerning the event in question is put in the centre of the window at time *T.* (See Figure 1.) In the pre-event period the precision of market anticipation appears which also provides a signal about the channels through which the information spreads out into the market. In case of gradual increase of the spread between normal and actual returns one can state that the news penetrated the market through informal channels are available only to a few investors. In the post period the speed of investor reaction or in another word their ability to adjust to the new circumstances will be assessable.

Figure 1. The design of time window around the event and the position of the estimation period



To determine the normal rate of return actual returns in the *estimation period* are inputted into a simple model. The length of the estimation period, in our view should at least be 250 units long, which is in the current case measured in day. In order to determine the normal rate of return the researcher can choose from three types of models typically applied in the literature: *1.* market adjusted model, *2.* mean adjusted model, and *3.* market model.

The market adjusted model assumes that the best predictor of the normal rate of return for a given security is the current return on the market. The market return is predominantly equivalent to a market wide index.

$$E\left(r_{it}\right) = \hat{r}_{it} = r_{mt}, \qquad /1/$$

where the expected (normal) return on security *i* in period *t* is equal to the actual rate of return on market index *m* in period *t*.

The normal return generated by the mean adjusted model is the average of actual returns of security i for the time period t = 1, 2, 3, ..., T.

$$E(r_{it}) = \hat{r}_{it} = \overline{r_i} = (1/T)\sum_{t=1}^{T} r_{it} . \qquad /2/$$

The market model[1] is equivalent to an ordinary least square (OLS) method, wherein returns on a given security i are regressed against the concurrent return on the market m.

$$E(r_{it}) = \hat{r}_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it} . \qquad /3/$$

The dependent variable is the expected return on security i in period t. On the right hand side α is the intercept, β is the slope coefficient, ε is the error term, and the independent variable is the market index in period t.

In case of the previously described three models estimation biases can arise as the result of unsynchronized trading or low liquidity, which is caused by low trading volume of the particular security. In order to overcome these obstacles literature offers several solutions, which are detailed in *Bedő–Rappai* [2005]. These solutions either propose modifications in the estimation methods or take an attempt to alter the specification of the model.

Meanwhile *Varga–Rappai* [2002] take another approach in the estimation of the slope parameter β by suggesting the application of GARCH specification for the random variable.[2] By using GARCH specification the residual term in the usual OLS model /3/ is not treated as a white noise term but its variance is estimated as a series of past innovations, where $\rho(L)$ and $\theta(L)$ are the standard lagged polynomial terms. This is pursued in order to eliminate the bias generated by the autocorrelation in the volatility of the time series. The proposed model is the following:

$$Var(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, ...)\sigma_t^2 = \omega + \rho(L)\ \sigma_{t-1}^2 + \theta(L)\ \xi_t^2 . \qquad /4/$$

It can empirically be proven that the GARCH (1.1) specification is sufficiently able to correct the problem, which implies that our model is reduced to the following form:

$$\begin{aligned} r_{it} &= \alpha + \beta r_{mt} + \varepsilon_t , \\ \sigma_t^2 &= \omega + \rho\, \sigma_{t-1}^2 + \theta\, \varepsilon_{t-1}^2 . \end{aligned} \qquad /5/$$

As it was stated before all the previously described models are to determine the normal return, which is compared against the actual rate of return of the securities.

[1] First proposed by *Fama et al.* [1969].
[2] See also *Engle* [1982].

The difference between the normal return ($\hat{r}_{it}$), and the realized rate of return (r_{it}) is called abnormal return, which is calculated from the first day of the pre-event window (*T–20*) until the last trading day of the post-event window (*T+20*). The formula of the abnormal return is as follows:

$$AR_{it} = r_{it} - \hat{r}_{it} \ . \qquad /6/$$

In order to eliminate all idiosyncratic effects that arise on the individual security level one has to calculate abnormal return for several securities with the same event allocated at time *T* (announcement of EPS or ROE ratios). Abnormal returns generated for securities under analysis have to be averaged cross sectional for each trading day. This is depicted in equation /7/.

$$\overline{AR_t} = \frac{1}{N}\sum_{i=1}^{N} AR_{it} \ , \qquad /7/$$

where *N* is the number of securities with calculated abnormal return for the given date *T*.

To be able to assess the tendency of change of abnormal returns throughout the event window *cumulative abnormal return* is calculated. This allows us to study the pre and the post event window, specifically to understand the precision of anticipation of market actors and the diffusion of information after the release of the news. The beginning of accumulation is at day *K* until trading day *L*.[3] The formula of cumulative abnormal return calculated for security *i* is as follows:

$$CAR_i^{K,L} = \sum_{t=K}^{L} AR_{it} \ . \qquad /8/$$

By cross sectional averaging individual cumulative returns we obtain the global cumulative return for the total portfolio, which takes the following form:

$$CAR^{K,L} = \frac{1}{N}\sum_{i=1}^{N} CAR_i^{K,L} \ . \qquad /9/$$

CAR values in the negative dimension imply “pessimism” in the market as the realized returns are lower than the normal ones (this is valid if we assume that the model which generates normal returns is correct). When CAR values are positive the market is said to have too “optimistic” expectations in relation to the event. If the market correctly values an event than the CAR value is 0.

[3] *K* is the first day of the event window (*T*–20), while *L* is the last day of the event window (*T*+20).

In the following sections when referring to announcement of company data we mean the release of ROE and EPS news, which will be put in the focus of our empirical analysis. The meaning and descriptive statistics of the two fundamental ratios are discussed in the data description section of the paper.

2. Value of a news, entropy of an event – methodology, hypotheses

One of the most well known axiom of information theory states that as the probability of occurrence of an event increases the level of surprise decreases, due to the diminishing rate of information content of the news.[4] In case of particular news let us call the measure of information content *discrete information*, which is indirectly proportional to the probability of occurrence of the event that the news describes. To formalize discrete information, in compliance with all properties, the most appropriate functional form is the following:

$$h(p) = c \log \frac{1}{p}, \qquad /10/$$

where $h(p)$ a is the measure of discrete information, p is the a priory probability of occurrence of the event, c and the base of the logarithm are constant.

In practice the constant c usually takes the value of one, while the base of the logarithm takes the value of 2 or e. In case the base of the logarithm is 2 the unit of information is referred to as *bit*, while if base value is e the unit of information is called *nit*. As the probability of occurrence of an event takes the value of 0.5 the information content of the news describing the underlying event is one:

$$h(0.5) = \log_2 \frac{1}{0.5} = 1\, bit, \qquad /11/$$

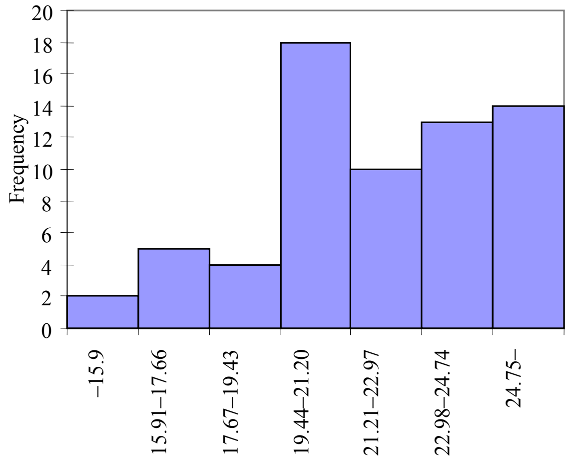
or

$$h(0.5) = \ln \frac{1}{0.5} = 0.693\, nit. \qquad /12/$$

To present how entropy value is determined let us use the following example. As it appears in the quarterly financial statements of General Electric Ltd., a company well known in Hungary, the return on equity figures in the past 16 years fluctuated between 14.12 and 26.52 percent. The distribution of the 66 data occurring in the past is presented in Figure 2.

[4] The concept of entropy is known not only in information theory but in thermodynamics and in statistical mathematics as well. Information entropy is also referred to as Shannon's entropy in honour of *Claude E. Shannon*, who in his 1948 paper introduced the concept (*Shannon* [1948]).

Figure 2. The distribution of quarterly ROE ratios of General Electric Ltd. between 1990 and 2006



When creating the histogram we converted continuous ROE values into a frequency table by applying the usual conventions to determine the number of frequency classes, and the length of frequency classes (*Hajdu et al*. [1994]). To proceed with our example consider General Electrics' ROE value as of the first quarter of 2006, was 9.49. This implies that this value belongs to the first frequency class resulting in a 0.03 estimated probability of occurrence, which is based on our previous sample. From this estimated probability of occurrence the value of the news is the following:

$$h(0.03) = \log_2 \frac{1}{0.03} = 5.04 \; bit \,, \qquad /13/$$

or

$$h(0.03) = \ln \frac{1}{0.03} = 3.50 \; nit \,. \qquad /14/$$

To interpret these values by themselves is a rather difficult task, while the fact that the relationship between the probability of occurrence of ROE ratio and the level of the entropy value is inverse, implies that as uncertainty increases the level of surprise as well.

Based on the previously introduced definitions we will present the concept of information entropy, which is the expected value of the distribution of individual news values. Information entropy measures the *uncertainty* of an empirical distribution, which equals to the average of the "surprises" that the occurring news generate. In another words it is equivalent to the weighted average of the surprises associated with the news that arrive:

$$H(p)=\sum_{i=1}^{k} p_i \log\frac{1}{p_i} = -\sum_{i=1}^{k} p_i \log p_i \ . \qquad /15/$$

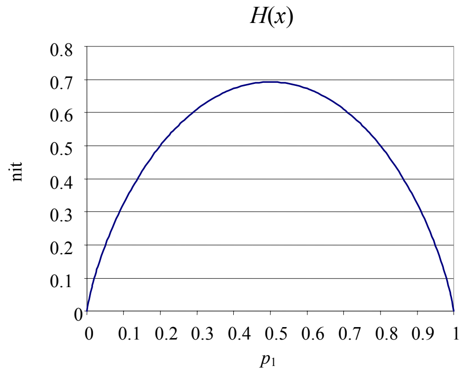
In light of the fact that the probability of occurrence of an event can take the value of 0, and inasmuch as

$$\lim_{x\to 0} x\log\frac{1}{x} = 0 \ , \qquad /16/$$

at elements of $H(p)$ with the property of $p_i = 0$ the value will be zero. As the value of the probability of occurrence have no dimension information entropy is measured either with bit or nit just as discrete information.

With regards to the previous definitions the entropy function, in case of *k*=2 takes the following form (implicitly $p_2 = 1 - p_1$):

Figure 3. The entropy function plotted against the probability of occurrence of the event



As it appears from Figure 3 the minimum and maximum points of the functions are 0 and ln2 (log*k* in general). The maximum of the function is at "total equality", in another word where

$$p_1 = p_2 = 0.5 \, , \qquad /17/$$

than

$$H(p) = -2\times 0.5\times \ln 0.5 = -\ln 0.5 = 0.693\, nit \ , \qquad /18/$$

or equivalently

$$H(p) = -2\times 0.5\times \log_2 0.5 = -\log_2 0.5 = 1\, bit \ . \qquad /19/$$

Note that while the maximum of the function is always at "total equality" as the number of events increase the information content of the value increases as well. To illustrate it consider the case when

$$p_1 = \ldots = p_4 = 0.25 \,, \qquad /20/$$

which implies

$$H(p) = -4 \times 0.25 \times \log_2 0.25 = -\log_2 0.25 = 2 \; bit \,. \qquad /21/$$

Such a case entails that the occurrence of events with equal probabilities generates higher rate of surprise, due to, in statistical terms the lack of concentration.

In order to compare entropy values based on different frequency tables we introduce *relative entropy*, which takes the following form:

$$H_r(p) = \frac{H(p)}{\log k} \,. \qquad /22/$$

The measure of relative entropy lies between 0 and 1 and as its relative characteristic implies it is without dimension.

Continuing our previous example, the frequency table based on the ROE ratios of General Electrics Ltd. is presented in Table 1.

Table 1

Frequency table of General Electrics' ROE ratios

ROE	Frequency	p_i
– 15.90	2	0.03
15.91 – 17.66	5	0.08
17.67 – 19.43	4	0.06
19.44 – 21.20	18	0.27
21.21 – 22.97	10	0.15
22.98 – 24.74	13	0.20
24.75 –	14	0.21
Total	*66*	*1.00*

As a consequence the entropy value, using natural based logarithm is as follows:

$$H(p) = \sum_{i=1}^{k} p_i \ln \frac{1}{p_i} = 0.03 \times \ln \frac{1}{0.03} + \ldots + 0.21 \times \ln \frac{1}{0.21} = 1.76 \; nit \,. \qquad /23/$$

The relative entropy is:

$$H_r\left(p\right)=\frac{H\left(p\right)}{\ln k}=\frac{1.76}{\ln 7}=0.904\ . \qquad /24/$$

Given that the maximum of relative entropy is 1, this result shows that the future announcements of ROE ratios can generate significant surprise. If the relative entropy value is one, than the probability distribution of ROE values in the past are even.

As we have introduced how news is valuated and price behaviour is scrutinized let us define our hypotheses:

Hypothesis 1: As the probability of occurrence of certain ROE and EPS figures decrease the absolute values of abnormal and cumulative abnormal returns at $t = T$, $t = T–20$, $t = T+20$ increase.

Hypothesis 2: As the value of the relative entropy tends to unity the absolute value of abnormal and cumulative abnormal returns at $t = T$, $t = T–20$, $t = T+20$ increases.

3. Data description

Our sample consists of 65 shares of publicly traded companies on the New York Stock Exchange (NYSE) and National Association of Security Dealers Automated Quotation (NASDAQ), which are constituents of Dow Jones Composite Index (DJCI) as of the first quarter of 2006. Share price and company data were acquired from Reuters database, which is a global vendor of financial data, news and analysis. To calculate abnormal and cumulative abnormal returns daily share prices were used, while return on equity (ROE) and earnings per share (EPS) ratios are based on quarterly financial data. The time series of ROE and EPS ratios reach back to the second quarter of 1992, AR and CAR are determined for the four quarters of 2005.

To determine market response to company event ROE and EPS announcements were put into the zero time period of the event window (see Figure 1). When conducting event study calculations we designed the event window with a –20, +20 bound, which implies that the window starts twenty days before and ends 20 days after the announcement of ROE and EPS ratios. The length of the estimation period is 250 trading days, which implies that the estimation period starts 270 days before the announcement of the financial ratios.

Return on equity is defined as the ratio of net income to shareholders' equity. Earning per share (EPS), which is the portion of the company's profit allocated to each outstanding share of common stock, is defined as the net income minus dividend on preferred stock to average outstanding share. These measures have been frequently used as performance indicators in the financial literature.

Table 2

Return on equity and earnings per share ratios in the sample at 2006Q1

Indicator	Mean	SD	Minimum	Median	Maximum
ROE (percent)	15.78	76.02	-2671.74	14.16	2931.37
EPS (USD)	1.58	2.28	-41.94	1.56	14.27

Note. ROE is the net income/shareholders' equity, EPS is the net income dividend on preferred stock/average shares outstanding. SD is the standard deviation. Number of firms in the sample is 65, the length of the time series of financial ratios is 65 quarters from 1990 Q1 until 2006 Q1.

Source: In this and in the following table Reuters database and the authors' computation.

Table 2 reveals summary statistics for the two performance measures. The ROE indicator exhibits substantial variation on its level as depicted by the standard deviation. The minimum and maximum figures in case of the ROE indicator are the result of the extremity of AMR Corporation, which is positioned in the air transportation industry (see the industry distribution of companies in the Appendix). The maximum appeared in the first quarter of 2003, while the minimum at the end of the same financial year. Although it does not imply a significant downturn in the performance of the company, since the positive figure in the first quarter was the result of a negative net income and shareholders' equity. By the end of the financial year the denominator (shareholders' equity) became positive, while the net income remained negative yielding a negative ROE indicator.

Table 3

Entropy indicators for the ROE and EPS ratios, AR and CAR values at t = T, t = T–20, t = T+20.

Indicator	Entropy indicator	Mean	SD	Minimum	Median	Maximum
ROE	*H*(bit)	2.1363	0.5587	0.2262	2.3323	2.7725
	H(nit)	1.4808	0.3872	0.1568	1.6166	1.9218
EPS	*H*(bit)	2.2336	0.4558	0.1371	2.3522	2.7599
	H(nit)	1.5482	0.3159	0.0950	1.6305	1.9130
CAR	$t = (-20)$	0.0025	0.0112	–0.0177	0.0004	0.0506
	$T = 0$	0.0058	0.0394	–0.0948	0.0004	0.1478
	$t = 20$	0.0099	0.0611	–0.1185	–0.0005	0.2572
AR	$t = (-20)$	0.0018	0.0078	–0.0135	0.0006	0.0352
	$T = 0$	0.0004	0.0153	–0.0633	0.0004	0.0419
	$t = 20$	0.0012	0.0066	–0.0073	0.0004	0.0344

Note. SD is standard deviation, *H*(bit) is the entropy index, the base of the logarithm is 2, *H*(nit) is the entropy index with natural base logarithm. CAR is the cumulative abnormal return, AR is the abnormal return $t = T–20$ and $t = T+20$ are the boundaries of the event window, $t = T$ is the announcement date. *N* for the number of firms is 65.

Using the return on equity and earnings per share data sets the value of announcements of the two indicators are calculated in nit and bit as well. The descriptive statistics of the entropy indices for the two types of news are indicated in Table 3, which also contains the descriptive statistics of abnormal and cumulative abnormal returns generated with formulae /7/ and /9/. The variation of the two entropy indices is low, which implies that the values of the news (ROE and EPS announcements) between the 65 companies are not too distinct. The maximum entropy index reveals the value with the largest surprise, in another word the company that this value belongs to has the highest uncertainty in its performance. It further entails that the value of information is the highest in case of this particular security as the foresee ability is the lowest in this instance. The minimum entropy values were generated by companies that have rather constant, predictable performance. The value (or uncertainty) of the ROE and EPS news are quite similar meaning that investors are able to utilize both performance indicators by bearing similar risk in the analysis.

Interestingly AR figures in the three points of the event window show that the precision of investors, implied by the actual return is the highest at the announcement day (mean of 0.04 percent). 20 days before and after the information release realized returns are 0.18 and 0.12 percentage points away from the normal returns respectively. While the mean of AR figures is the lowest at $t = 0$ the variation is the highest at the same time, which entails increasing uncertainty. The higher standard deviation naturally is the result of the higher maximum and minimum values. An investor, who goes long at the beginning of the day and shorts the same share at the end of the trading session can end up with either –0.633 percent or 0.419 percent of abnormal rerun implying loss or gain respectively.

By observing cumulative abnormal returns the tendency of AR development can be understood. The discrepancy between the normal and the actual returns is the lowest before the release of the information. In this period the variation of this spread is also the lowest compared to the announcement date and the period that follows. As the announcement takes place the uncertainty enhances ($\mathrm{SD}(t_{-20}) = 1.12\%$; $\mathrm{SD}(t_0) = 3.94\%$), that results in the decrease of the precision of anticipation ($\mathrm{mean}(t_{-20}) = 0.25\%$; $\mathrm{mean}(t_0) = 0.58\%$). This trend continues after the announcement by almost reaching the mean of 1 percent of cumulative abnormal return with 6.11 percent of standard deviation. A person investing into the share that produces the maximum (minimum) CAR value is able to earn 26 percent more (12 percent less) return than the normal, estimated by our model.

4. Results

Following the previous line of argument we determine certain dependent variables that are to be estimated with high level of precision in order to be able to build an efficient trading strategy. Based on this requirement we specified four variables to be estimated. The dependent variables and the reasoning for their selection are the following.

1. Abnormal return at the time of the news announcement (*AR0*); we argue that the uncertainty around the occurrence of the news must have an effect on prompt reactions of share prices.

2. The maximum deviation from the estimated daily returns during the 20 days time period subsequent to the announcement. In another word the maximum absolute value of daily abnormal returns (*maxAR*). In our view there is a deterministic relationship between the value of the news (rate of surprise) and the deviations of returns from the estimated ones following the announcement.

3. The maximum value of the cumulated abnormal returns during the 20 days following the announcement: this implies the maximum of the absolute value of the cumulative abnormal return in the post event window (*maxCAR*).

4. Finally the cumulative abnormal return measured at the last day of the post event window (20^{th} day of the post event window – *CAR20*). Based on our findings discovered in our previous study we can state that by the 20^{th} day following the announcement share prices return to their "normal" tendencies. Following this argument we are interested, whether we can generalize on this in the current case as well.

The descriptive statistics of the previously described dependent variables appears in Table 4.

Table 4

Descriptive statistics of dependent variables

Variables	Mean	Standard deviation	Minimum	Maximum
AR0	0.000418	0.0152941	–0.0633	0.0419
maxAR	0.019393	0.0114395	0.0083	0.0633
maxCAR	0.053439	0.0456879	0.0114	0.2572
CAR20	0.009946	0.0610757	–0.1185	0.2572

Note. AR0 is the abnormal return on the day of announcement of ROE and EPS ratios, *maxAR* is the maximum absolute value of abnormal return values in the 20 days time period subsequent to the announcement, *maxCAR* is the maximum absolute value of cumulated abnormal return values in the 20 days time period subsequent to the announcement, and *CAR20* is the cumulative abnormal return at the 20^{th} day subsequent to the announcement.

Note that values in Table 4 are based on time series of daily returns, which implies that to transform these values to an annual scale they have to be multiplied by

365. The fact that standard deviation values equal or greater to the average values should not be surprising as in time series analysis this characteristic is a common one.

4.1. The effect of the value of the news on returns

In the first phase of our empirical analysis we investigate the consequences of the value of news (level of EPS, ROE ratios) release on the dependent variables. Interestingly enough the average of the news values (in case of *EPS_NEWS* – 2.26 nit, and for *ROE_NEWS* 2.31 nit) implies that the average of the probability of occurrence is around 10 percent. One can predict from the histograms of independent variables that the distributions of news values are close to normal, which implies results similar to the Bera–Jarque test results. (See Table 5.)

Table 5

Bera–Jarque test, and p-values of independent variables

Variables	Bera–Jarque	*p*-value
EPS_NEWS	2.057	0.357
ROE_NEWS	0.515	0.773

This entails that the assumption of normality in case of both independent variables is feasible.

In light of our hypotheses we can state that value of the news has a positive relationship with the rate of surprise. In another word there is a strong correlation between the independent and dependent variables.

Table 6

*Correlation coefficients of dependent and independent variables**

Dependent variables	Independent variable	
	EPS_NEWS	*ROE_NEWS*
AR0	–0.141	0.030
maxAR	0.066	0.016
maxCAR	–0.220	0.314
CAR20	–0.135	0.090

* Table 6 is a segment of a correlation matrix. The remaining parts of this matrix are irrelevant to the current argument.

In case of 65 observations the hypothesis of independence can be rejected at the 5 percent significance level if the absolute value of the correlation coefficient is greater than 0.278. In our particular case this implies that the relationship between the *ROE_NEWS* and the cumulated abnormal returns by the 20^{th} day of the post event window is statistically significant. This result is in line with our hypothesis as the high level of uncertainty around the event generates a long lasting disturbance in stock prices, which results in a continuous deviation of daily returns from the estimated ones.

Meanwhile our further assumptions of causality between the rate of surprise and return variables are not supported by the correlation coefficients. This implies that a model based on the previous dependent and independent variables is unable to describe the mechanism of stock prices around the release of the assessed events.

4.2. The predictability of announcements and the abnormal returns

As we were unable to detect significant relationship between the value of the announced news and the abnormal returns subsequent to the announcement we argue that market participants had already factored the volatility of the particular news into their expectations. We have presented the calculation of the value of the news announcement with the application of entropy and relative entropy. The relative entropy value tends to 0 (H_r) entails that the news is rather predictable, while H_r closer to 100 percent is likely in case of uncertainty of future announcement.

The mean of relative entropy values, in case of EPS and ROE announcements, are 76 and 79 percent respectively, while for individual stocks the deviation is on a much wider scale (for EPS announcements individual stocks deviate between 5 and 98 percent, for ROE announcements between 8 and 98 percent).

We assumed that the value of the news announcement is not particularly able to explain the behaviour of abnormal returns subsequent to the event because the historical volatility of entropy values for stocks are also considered by investors. This implies that for a share with hectic announcements in the past newly occurring salient news will not be as credible as for a share with constant performance. As credibility alters abnormal returns differ as well. To assess this assumption we calculated the correlation matrix of the relative entropy values and abnormal return variables (introduced in the previous section).

However, the absolute values of correlation coefficients in Table 7 exceed values in Table 6 statistically significant result is solely detectable in the case of ROE and *maxCAR* variables. Consistent with our previous line of argument we examine whether individual news announcement values and the predictability of the news together have an influence on the behaviour of time series of returns. We selected linear regression models to assess this relationship, which is justified by the presented normality of the datasets.

Table 7

Correlation coefficients of dependent and independent variables (relative entropy values)

Dependent variables	Independent variables	
	EPS_H$_r$	*ROE_H$_r$*
AR0	0.151	0.145
maxAR	–0.122	–0.212
maxCAR	–0.070	–0.403
CAR20	0.218	–0.154

In order to generate more informative results, beyond the determination of explanatory power of the outcomes we applied stepwise method (*Mundruczó* [1981]). Table 8 shows the output of the stepwise regression. Table 8 contains only those estimated parameters of independent variables and their *t*-statistics (*t*-statistics with value of one is included, with 0.5 is excluded) that according to the stepwise regression method exceeded the average level of statistical significance. Coefficients with explanatory power were adjusted to the number of independent variables (see adjusted R^2 in Table 8).

Based on the output of the regression analysis we can conclude that the explanatory power of the optimal models is rather low. The only variable with acceptable explanatory power was the one that was constructed from the cumulated abnormal returns. It is not surprising that the correlation matrix reveals that the predictability of maximal cumulated abnormal return is the highest; while it is promising that the same is true for the cumulative abnormal return by the 20th day subsequent to the event. (See Table 8.)

Table 8

Results of the step wise regression analysis

Independent variables	Dependent variable							
	AR0		*maxAR*		*maxCAR*		*CAR20*	
	$\hat{\beta}$	*t*-statistics	$\hat{\beta}$	*t*-statistics	$\hat{\beta}$	*t*-statistics	$\hat{\beta}$	*t*-statistics
Constant	–0.0012	–1.218	–0.0026	–4.407	0.0088	3.216	-0.0013	-0.348
EPS_NEWS	–0.0004	–1.871	0.0002	1.045			-0.0021	-2.458
ROE_NEWS					0.0011	1.987		
EPS_H$_r$	0.0018	1.376					0,1530	3.150
ROE_H$_r$	0.0011	1.092	–0.0014	–1.941	-0.0080	-2.991	-0.0066	-1.723
R^2	*0.040*		*0.031*		*0.187*		*0.150*	

Though we did not find any model that contains all independent variables at a high enough level of explanatory power, nevertheless the variable that is to represent the impact of the announcement of ROE is present in each model. Furthermore the fact that most of the coefficients are of negative sign indicates that unexpected announcements and hectic past announcement structures make investors "resistant" implying low level of effect on stock prices.

5. Conclusion

It has been presented that the event study methodology can be sufficiently extended as independent variables instead of just being binary variables (occurrence of news indicate the value of 1, no news = 0) represent the value of the news that had been announced in the middle of the event window ($t = 0$).

We were able to prove that from among the conventional dependent variables the aggregate of the deviations from the actual return series (cumulated abnormal return) is the only one, which can be predicted with certainty while individual deviations (daily abnormal returns) and daily returns are still unpredictable in light of the news announcement.

Regression analysis reveals that the announcement of the earnings per share data generates a higher impact on share prices as opposed to the return on equity ratio. This implies that investors are rather using EPS to specify the performance of their company. In light of entropy values ROE data seems to be more relevant as investors become resistant to extremely turbulent announcement history, which results in lower impact of announcements on share prices.

Finally we suggest that from our dependent variables investor public should utilize the maximal deviation from actual return time-series, which is represented by the maximal cumulated abnormal return. As this variable was predicted with the highest precision we recommend its application in hedging strategies, which might generate extra return around the date of the announcement of financial indicators.

Appendix

Distribution by industry of sample firms as of 31 December 2005.

Industry	Number of companies
Aerospace and military technology	2
Automobiles	1
Banking	1
Beverages and tobacco	1
Broadcasting and publishing	1
Business and public services	5
Chemicals	1
Data processing and reproduction	2
Electrical and electronics	1
Electronic components and instruments	1
Energy sources	2
Financial services	2
Food and household products	2
Health and personal care	3
Industrial components	1
Insurance	1
Leisure and tourism	2
Machinery and engineering	1
Merchandising	2
Metals – non ferrous	1
Telecommunications	2
Transportation – airlines	5
Transportation – road and rail	8
Transportation – shipping	2
Utilities – electrical and gas	14
Multi-industry	1

Note. N for number of companies in the sample is 65.
Source: Reuters database and authors' computation.

References

BEDŐ, ZS. – RAPPAI, G. [2004]: The application of event study methodology to shares listed on the Budapest Stock Exchange: Is news, influencing stock prices, valuable or not? *Szigma*. Vol. 35. No. 3–4. p. 113–127.

ENGLE, R. [1982]: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica.* Vol. 50. No. 4. p. 987–1007.

FAMA, E. ET AL. [1969]: The adjustment of stock prices to new information. *International Economic Review.* Vol 10. No. 1. p. 1–21.

HAJDU O. ET AL [1994]: *Statisztika I*. Pécsi Tudományegyetem. Pécs.

HONG, H. – LIM, L. – STEIN, C.J. [2000]: Bad news travel slowly: Size analyst coverage, and the profitability of momentum strategies. *Journal of Finance.* Vol. 40. No. 1. p. 265–295.

MUNDRUCZÓ, GY. [1981]: *Alkalmazott regressziószámítás*. Akadémiai Kiadó. Budapest.

SHANNON, C. E. [1948]: Mathematical theory of communication. *The Bell System Technological Journal*.Vol 27. No. 2. p. 379–426.

VARGA, J. – RAPPAI, G. [2002]: Heteroszkedaszticitás és a szisztematikus kockázat hatékony becslése GARCH modell alapján – A magyar részvénypiac elemzése. *Szigma*. Vol. 33. No. 3–4. p. 108–116.

The examination of the factors of growth in the Hungarian small- and medium size business sector*

László Szerb,
Associate professor
University of Pécs
E-mail: szerb@ktk.pte.hu

József Ulbert,
Associate professor
University of Pécs
E-mail: ulbert@ktk.pte.hu

Despite increasing interest, it is still a lack of rigorous model building and econometric method application in the area of small business growth in Hungary. This paper constructs an integrated conceptual model of growth. The model incorporates four major categories of variables such as personal demographic, personal behavioural, business demographic and business behavioural variables. To test the empirical validity of the model, different measures of growth, including employment, sales, equity and composite factors, are used in five empirical models. The stepwise regression method is proved to be a proper tool to identify the significant factors of business growth. However, the different dependent variables of growth are mainly influenced by different independent factors. Most observed outcomes corresponding to previous empirical results and the alterations can be explained by the limited market economy experience and the transitional nature of Hungary. Business behavioural factors of investment, technology development, exports, organizational change and strategic orientation are found to be the major determinants of business growth. Personal behavioural features like ownership experience in other businesses as well as business size, age, legal form, the number of founders and foreign owners are significant but less important determinants of growth.

KEYWORDS: Enterprise statistics.

* *László Szerb* acknowledges the financial support from the Bolyai János Research Foundation, and *József Ulbert* is grateful for the support from OTKA Research Fund Nr. T035105. The authors would also like to thank for the referee *Iván Major* and for *Gábor Rappai* for useful comments and suggestions. *Judit Bedőné Károly* and *Orsolya Kisantal* provided assistance in the data collection. Thanks to *Hugh Spall, Katherine Spall*, and *Siri Terjesen* for language correction.

Despite considerable statistical and interpretation problems, the performance and the future prospect of a country is most frequently measured and judged by economic growth, unemployment and inflation. Out of these measures, probably the examination of growth has received the widest interest among economic scholars. After *Solow* [1956] and *Lucas* [1988], *Romer's* [1990] so-called endogenous model has became the major theoretical framework in examining economic growth since the late 1990s. Besides that all of these macro models mention the importance of the underlying economy – Lucas builds on that – both the theory and the empirics lack to demonstrate and prove the connection between macro and micro economy, or in other words between macroeconomic and firm level growth.

The examination of firm level growth raises several difficulties. A major problem emerges during the aggregation of firm data, when we loose important information about the personal performance, characteristics and behaviour of the individual businesses that are associated with the growth of a business. For example, just recent researches have presented evidences that only a tiny portion of the young and small businesses, called gazelles, are responsible for most the creation of new employment in an economy (*Birch* [1987], *Autio* [2005], *Parker–Storey–Witteloostuijn* [2005]). Aggregation also hides the dynamics behind growth, i.e. even in the most prosperous sector there are stagnating and disappearing firms, or, on the contrary, in the cases of declining segments, some firms still grow. To understand the reason of this different performance specific, firm level data collection and examination are required.

Since the second half of the 1990s, a new wave of research has emerged aiming to investigate small business growth (*Davidsson* [2003], *Davidsson–Wiklund* [1999], *Delmar–Davidsson–Gartner* [2003], *Davidsson–Achtenhagen–Naldi* [2005], *Reid* [1993], *Storey* [1994], *Wiklund–Stepherd* [2004], *Weinzimmer–Nystrom–Freeman* [1998]). Instead of rigorous modelling, these researchers have focused on empirical testing. The mentioned authors identified and explained several factors of growth. Moreover, they presented evidences about the problems associated with the different measures of growth. A major shortcoming of these researches is the reliance on business registry data that does not make possible to analyze important behavioural characteristics of business growth. These caveats call for specific firm based data collection as well.

Besides numerous publications in the area of business growth, Hungarian researchers have been focused more on studying large, mainly foreign-owned businesses that determined and lead macroeconomic growth in Hungary since the mid 1990s (*Szerb–Ulbert* [2002]). However, it is no doubt that the interest toward small businesses have been increasing.

Regarding the growth of the newly established businesses and their effect on employment are analyzed by the Global Entrepreneurship Monitor (GEM) studies where mainly the connection among individual demographic characteristics, entrepreneurial traits and growth are in the spotlight (*Ács et al.* [2002], *Szerb et al.* [2004], *Szerb* [2005]). The investigation of the growth potential of the newly established and of privatized businesses is the main issue of *Laki* [1998], [2001].

In a comprehensive empirical study, *Czakó et al*. [1995] analyzed the small business sector in the early years of transition. The authors distinguished six groups of small businesses according to the several features like business versus household orientation, full time or part-time focus, and demographic characteristics. They also noted that most recently founded businesses did not fulfil the classical definition of entrepreneurship.

The limited growth capacity of the small businesses is also recognized by *Jávor–Rozgonyi* [1995] and *Laky* [1998]. By studying the efficiency of the Hungarian business sector, *Halpern* and *Kőrösi* [2001] observe the relative backwardness of the smaller size businesses as well as the improved performance of the corporate sector in the second half of the 1990s. Another notable study is *Major* [2002], which analyzed the performance of the Hungarian small and medium sized enterprises (SME) and recognized the efficiency problems not only in the micro but also in the medium size businesses sector.

The influential factors and the consequences of growth have been less frequently investigated with the exception of the financial (capital) problems. *Kuczi* and *Makó* [2000] pointed out how important the effect of social capital can be on business growth. The employment of micro businesses is the main topic of *Vajda* [1999].

The structure and the structural change of the domestic small business sector have been examined probably by the most researchers.

Notable papers can be found from this field including *Kőhegyi* [1998], *Mészáros–Pitti* [2003], and *Román* [2002], [2005]. Recently, *Kőhegyi* [2001] has provided the most comprehensive study on firm growth by presenting the regional, sectoral, financial features and differences of the growing and shrinking businesses in the 1996–1999 time period. However, Kőhegyi could not analyze the behavioural factors of firm growth. *Szirmai* [2002] has probably the only one that deals with the limited number of the fastest growing Hungarian businesses, the gazelles. By relying on the life-cycle model, *Salamonné* [2006] presented evidences about the development phases of Hungarian businesses. Based on a sample of 50 carefully selected small businesses she suggested a minor modification of the original life-cycle model to fit to the local conditions and the transitional nature of the Hungarian economy.

Despite considerable domestic theoretical and empirical development in the area of small businesses, there is still a lack of rigorous modelling and the application of econometric methods in Hungary.[1] Therefore the aim of this paper is twofold. First, we would like to build a conceptual model of small business growth that is different from the popular life-cycle models, recently used in the Hungarian literature (see

[1] In this area of research only *Halpern–Kőrösi* [2003] and *Major* [2002] applied econometric technique. However both of these studies focus on efficiency rather than growth.

Szirmai [2002], *Salamonné* [2006]). Second, to test the validity of the conceptual model we intent to rely on econometric regression methods. By relying on a unique, but non-representative individual small business data set, we aspire to identify the most important determinants of business growth and evaluate the sometimes contradictory results.

In the following part of the paper we built a conceptual model of business growth that contains four major groups of influential factors of growth. The theory building is followed by the description of the data set and the methodology. The expected outcomes and the regression results are evaluated in the third section. Section four contains the discussion of the overall findings.

1. The conceptual model of firm growth

There is probably one thing in which the different authors and experts agree: examining growth and its influential factors on firm level is a difficult and multidimensional issue. Over years, more or less researchers have found that besides demographic features, certain behavioural factors, like clear growth strategy orientation, product and process innovation, the application of specific financing resources like angel money, entrepreneurial skills as risk taking, opportunity recognition, managerial knowledge and expertise have significant influence on firm growth (*Acs* [1996], *Baumol* [2002], *Dodgson–Rothwell* [1995], *Kirzner* [1979], *Moran–Ghoshal* [1996], *Porter* [1985], *Reid* [1993], *Storey* [1994], *Wennekers–Thurik* [1999]).

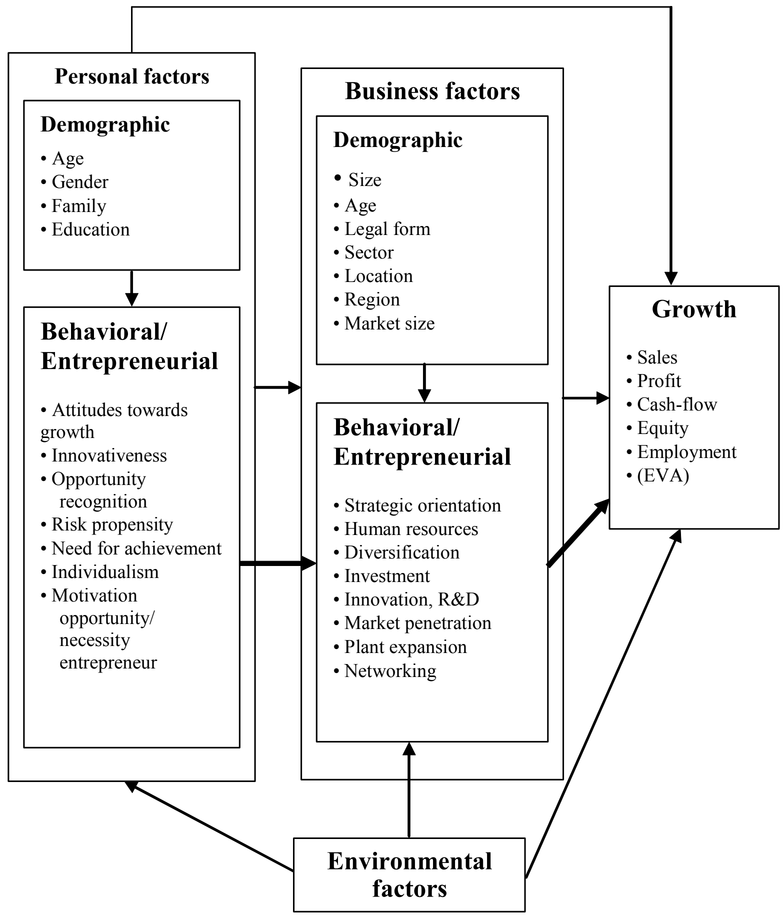
In a comprehensive conceptual framework model, Figure 1 shows the influential factors of business growth.

Figure 1 groups the prominent elements of firm growth in three major categories as personal, business and environmental factors. While the importance of the environment is acknowledged, we are focusing on the personal and business factors.

Both of these categories consist of two other subcategories, one is a demographic that includes basically given features, and the other is behavioural that is associated with learnt characteristics.

Both personal and business demographic characteristics can have an effect on behavioural factors. The existence of different measures of growth is also incorporated in the model. According to the model, personal factors influence business characteristics that determine business growth. At the same time, personal factors can directly influence growth. However, as noted by the bold arrow, the major direction of the impact on growth is assumed to derive from the behavioural features.

Figure 1. The conceptual model of firm growth



We also label the behavioural factors as entrepreneurial, highlighting that there is a direct connection between entrepreneurship and growth. This relationship is particularly salient because it provides both the theoretical and the practical application of the concept of entrepreneurship in cases of already existing business (*Davidsson–Achtenhagen–Naldi* [2005]). The connection between economic growth, new employment and entrepreneurship are important issues of *Wennekers–Thurik* [1999] and *Acs et al.* [2003]. Some authors, including *Sexton* [1997] claim, that "growth is the very essence of entrepreneurship" (*Sexton* [1997] p. 97. also cited by *Davidsson–Achtenhagen–Naldi* [2005] p. 4.).[2] Despite *Davidsson's* [2003] argument

[2] Just examining journal papers and leading textbooks between 1988 and 1992, *Morris et al.* [1994] found 77 definitions. *Wennekers* and *Thurik* [1999] claim that "entrepreneurship is an ill-defined, at best multidimensional, concept" (p. 29.).

that we have the necessary blocks to build a strong paradigm for entrepreneurship research, the full and generally agreed definition of entrepreneurship is still missing (*Gartner* [1990], *Wennekers–Thurik* [1999], *Shane–Venkatamaran* [2000]).

However, it is agreed that entrepreneurship includes two, many times interrelated factors, traits and behaviour. Several researchers separate entrepreneurial and managerial roles, where entrepreneurial style is associated with specific traits like high risk taking propensity (*Knight* [1990]), need for achievement (*McClelland* [1961]), creativity and innovativeness (*Schumpeter* [1934]). The recognition and pursuit or exploitation of opportunity is considered to be the heart of entrepreneurship by *Kirzner* [1979], *Timmons* [1999], *Venkataraman* [1997] and *Shane–Venkataraman* [2000]. These features are shown on Figure 1. Longer or shorter lists of other entrepreneurial traits can be found in many publications including *Herbert–Link* [1989], *Timmons* [1999], *Chell et al*. [1991] or *Hisrisch et al*. [2005]. Personal-demographic traits include gender, age, education, and family background have been widely investigated and connected to business growth. According to recent GEM research results, a typical entrepreneur is middle aged (35-45 year old) male, with higher education degree, and having an entrepreneur in the family or knowing an entrepreneur personally. In terms of these features, Hungarian entrepreneurs do not differ from their colleagues from abroad (*Ács et al* [2002], *Szerb* [2005]).

Entrepreneurial businesses are characterized by distinct behaviour including processes, sets of actions, and functions. Several aspects of strategic orientation, like strategy formulation in general (*Stevenson–Roberts–Grousbeck* [1989], *Sadler-Smith et al*. [2003]), development of human and intellectual capital (*Becker et al.* [1997], *Glancey* [1998], *Ucbasaran–Westhead–Wright* [2005]), diversification (*Berger–Ofek* [1995]), networking (*Holm– Erikson* [1999], *Tsai–Ghoshal* [1998], *Wedin* [2003]) and investment propensity (*Iyigun–Owen* [1997]) are examined and associated with ultimate growth. Following *Schumpeter's* [1934] famous notion about the creation of something new, many researchers have identified innovation as the distinctive attribute of entrepreneurship and therefore a source of growth (*Drucker* [1985], *Dodgson–Rothwell* [1995], *Baumol* [2002], *Teece* [1998]). Out of the firm demographic characteristics, the examination of firm size, age and growth has received the widest interest. According to *Jovanivic* [1982] *Evans* [1987], and *Hall* [1987] smaller and younger firm grow at a faster rate than older and larger businesses, denying Gibrat's law. The examination of the legal form, business sector, location, region, and market size can be found mainly in empirical studies (e.g. *Davidsson et al.* [2000], *Reid* [1993], *Storey* [1994]). All of these factors are present on Figure 1.

While the measurement of macroeconomic growth is relatively well developed it cannot be said about firm level growth. In order to arrive at a balanced view at diverse determinants and measures of growth, multiple indicators are required. Different indicators result in very different outcomes. The correlations between different growth measures are generally weak (*Weinzimmer–Nystrom–Freeman*

[1998]). Useful empirical applications of the growth concept are provided by *Davidsson–Wiklund* [1999], *Davidsson et al.* [2002], *Delma–Davidsson–Gartner* [2003], *Davidsson–Achthagen–Naldi* [2005], *Wiklund–Stepherd* [2004]), who highlight the multidimensional characteristics of growth. Following the Swedish researchers and other well recognized empirically oriented studies (*Reid* [1993], *Storey* [1994], *Weinzimmer–Nystrom–Freeman* [1998]), we intent to apply three measures of growth as changes in sales, employment and own capital. We demonstrate that applying or favouring only a single growth factor may be misleading. It should be noted that neglecting of opportunity costs, liquidity premium and compensation of risk weakens the application of this approach (*Ventakamaran* [1997], *Shane–Venkataraman* [2000]). However, our data set does not make possible to apply some other measures like economic value added (EVA). Moving to the following section, an additional comment should be added. The mentioned model does not contain all the factors; probably any category can be expanded. However, we focus on the entrepreneurship related features; therefore, factors that are associated with the firms' behaviour, such as pricing, marketing, and finance, are only implicitly included in the model.

2. Data description, variables and methodological considerations

Our data set consists of 320 individual small and medium size business that had at least two years of operation (were established before 1st. January 2002). First, it should be emphasized that this research was experimental; therefore it lacked the representativeness of the data set. Originally, the questionnaire aimed to collect small firm growth data, but some individual characteristics were also recorded. Respondents were the main owners of the businesses and also had a position in the leadership: in most cases they were the executive managers. Data were collected between March 2004 and March 2005, and the harmonization of the data set and the collection of the missing data took place in November 2004 and May 2005. The basic characteristics of the sample and the applied (significant) variables can be found in Tables 1 and 2.

Out of several possible growth measures, we use employment, sales, and own equity data.[3] Both sales and equity data are in real terms. We do acknowledge the availability of a large number of growth measures as described in many studies (*Davidsson* [2005], *Davidsson–Wiklund* [1999], *Delmar–Davidsson–Gartner* [2003], *Weinzimmer–Nystrom–Freeman* [1998], *Wiklund–Stepherd* [2005]).[4]

[3] Own equity includes the subscribed capital, the annual profit and the reserved profit from previous years.

[4] Probably many experts, readers miss the growth of profit from the growth measures, however we had a lot of missing data that would have decreased the sample size. Moreover, it is well-known that most data are not really correct in the case of small businesses, but we think that the unreliability of profit data is the most prevailed.

Table 1

Sample description

Characteristics	Size category based on number of employees in 2002						
	0–1	2–5	6–9	10–19	20–49	50–249	Sum/average
Number of firms	66	102	35	49	44	24	320
Age of firm in 2004 (year)	7	8.2	9.7	10	11.2	11.8	9.1
Male ownership (percent)	56	57	69	80	82	67	66
Age of entrepreneur in 2004 (year)	43	45	46	49	49	51	46
Opportunity orientation (percent)	35	27	46	29	30	8,3	30
Independent existence orientation (percent)	42	53	57	65	39	25	49
Necessity orientation (percent)	40	52	54	55	52	71	52
Other business start-up (percent)	29	41	42	55	41	21	39
Number of founders	2	2.4	2.8	3.4	4.6	6.7	3.1
Family business (percent)	64	61	40	47	36	25	51
Foreign owners (percent)	4.5	5	5.7	6.1	4.5	8.3	5.3
Number of employees in 2002	0,8	3	7.2	13.8	31	102	16.1
Net sales in 2002 (million HUF)	5	42	91	246	383	882	183
Own capital in 2002 (million HUF)	1.7	7	24	34	104	376	56.5
Plant expansion 1998–2004 (percent)	9	13	11	24	20	17	15
Change of ownership 1998–2004 (percent)	23	14	9	39	25	17	21
Organizational change 1998–2004 (percent)	3	5	17	24	23	25	13
New product introduction 1998-2004 (percent)	14	8	11	33	23	33	17
New technology introduction 1998-2004 (percent)	14	16	20	31	27	46	22
New investment 1998-2004 (percent)	47	61	54	71	89	67	63
Has a business plan (percent)	56	60	74	61	55	58	60

Table 2

The list of applied significant variables

Variable	Use	Type	Description
Growth measures			
EMPCREG	D	Continuous	Change of employment between 1998–2004 or from start-up to 2004, the slope of regression
SALESCHREG	D	Continuous	Change of real sales between 1998–2004 or from start-up to 2004, the slope of regression
CAPITCHREG	D	Continuous	Change of real own equity between 1998–2004 or from start-up to 2004, the slope of regression
CLUSTERREG3	D	Continuous	Cluster of employment, of real sales and of real own equity changes calculated from regression slopes
CLUSTERREG2	D	Continuous	Cluster of employment, and of real sales changes calculated from regression slopes
Personal demographic			
AGEENTR	I	Categorical	The age of entrepreneur, 1: 18–35 year age, 2: 36–45 year age, 3: 46–55 year age, 4: 56 and up age
AGEENTRSQ	I	Categorical	The square of the AGEENTR
Personal behavioral			
STARTEXP	I	Categorical	The entrepreneur has participated in other start-up: 0: no start-up experience, 1: start-up experience 2: start-up+ ownership experience
Business demographic			
SIZE	I	Categorical	The size of business based on the number of employees in 2002. 1: 0–1 employees, 2: 2–5 employees, 3: 6–9 employees, 4: 10–19 employees, 5: 20–49 employees, 6: 50–249 employees
AGEBUS	I	Categorical	Age of business. 1: 1–3 year, 2: 4–6 year, 3: 7–9 year, 4: 10–12 year, 5: 12–14 year, 6: 15 and up year
LEGFORM	I	Categorical	The legal form of the business. 1: sole proprietorship, 2: unlimited liability, 3: limited liability
FOUNDER	I	Categorical	Number of founders in categories: 1: 1 founder 2: 2 founders, 3: 3–5 founders, 4: 6 or more founders
FORIGNOWN	I	Dummy	Foreign ownership: 0: no foreign ownership, 1: foreign ownership
Business behaviour			
INVEST	I.	Dummy	Investment in the business between 1998–2002, 0: no investment 1: investment
EXPORT	I	Dummy	Export of sales in 2004: 0: no export, 1: export
INNOVTECH	I	Dummy	New technology introduction between 1998–2004, 0: no technology innovation 1: technology innovation
STRATORIENT	I	Dummy	Strategic orientation of the business, 0: no business plan, written strategy: 1: written business plan, strategy 2: written business plan and strategy

Note. I: Independent variable, D: dependent variable.

Following *Weinzimmer–Nystrom–Freeman* [1998], we use the regression based measure of growth where the slope of the regression curve over the examined time period serves to identify the rate of growth.[5] A second measure of growth was calculated by the factor analysis technique where employment, real sales and real equity growth rates constituted one factor, and employment and real sales growth rates (as the two most important growth measures) constituted another factor.[6] The examined time period ranges from 1998 to 2004, or if the firm was established after 1998, then from the first year of start-up to 2004.

As it happens frequently in experimental data collection, not all the necessary factors described on Figure 1 are available. Of the individual factors, most entrepreneurial traits are missing except the opportunity/necessity variable. Moreover, all the business characteristics data are available except the human resource and networking variables. However, there is an important limitation of the model testing: the small sample size does not make possible to apply simultaneous equation system in order to identify the major influential factors of growth originated from personal features then affecting business characteristics as it is implied by Figure 1. Therefore, we are testing the effect of personal and of business factors within the framework of one model. Since the instrumental variable method of two stage least squares (2SLS) is also proved to produce inconclusive results, we selected the stepwise regression technique.

A stepwise regression procedure is able to find the best predictors of the dependent variables. The forward method starts with the most significant variable and adds the most statistically significant term (the lowest p-value) at each step, until there are no more significant variables left. The backward method begins with including all of the variables, then the independent variable with the smallest partial correlation with the dependent variable is removed first if it meets with the selection criteria. Here, the ten percent F value is chosen as a selection criterion. In the following, the same method is applied, and the removal of the variables continues until all the independent variables are significant with the dependent variable. Since it is not guaranteed that the stepwise regression provides the best results, both the forward and the backward methods are tried and the best results are reported.[7] The results are provided by the SPSS Version 11 statistical program package. Data points with missing values were omitted from the analysis.

By the application of the stepwise regression procedure, we could select only the significant determinants of growth, therefore the relative importance of personal demographic, personal behavioural, business demographic and business entrepreneurial should be tested. We assume that, in general, business factors would be more important determinants of growth than personal features. Moreover, we believe that business entrepreneurial characteristics would be even more significant

[5] We also used other relative and absolute growth rates; however, the results were about the same.
[6] Note, that the application of absolute instead of relative measures gives partially different results.
[7] For details regarding the stpewise regression method see *Mundruczó* [1981] or *Rappai* [2001].

than business demographic attributes. As a supplement, the correlation coefficients will serve to support the idea about the connections between the personal and the business characteristics. In the way of result evaluation we should also consider those Hungarian specialties that basically derive from the limited market economy experience of domestic entrepreneurs and businesses. Though, it is mostly agreed that Hungary, along with other similar countries like the Czech Republic, Poland and Slovenia, have successfully laid down the foundations of the market system, entrepreneurs and businessmen make their decision according to market rules and price signals (*Transtion economies* [2002], *The first ten years* [2002], *Szerb–Ulbert* [2002]). Therefore we do not expect too much different influential factors of the growth of Hungarian businesses as compared to other matured market economy enterprises.

3. Model testing: expectations and results

In the following, the five stepwise regression results are reported in Table 3. Note that we included only those variables that proved to be significant at most at ten percent.

According to the *F* test, all regressions are significant. However, the explanatory powers are not too high: the adjusted R^2 are below 0.15. There was no need to test for potential multicollinearity, since the stepwise regression method excluded this possibility. Moreover, the Glejser test did not imply heteroscedasticity problems. As it has been expected, the five different growth variables are affected by different independent variables underlying the notion about the application of various growth measures. Relatively, the growth measures that contain sales provide the worst results. During the evaluation, if one independent variable has not been proven to be significant at least two times, then we consider the effect of variable as not important (not robust).

Table 4, containing the correlation coefficients among the applied variables serves as an addition instrument to our analysis.

During our evaluation we would like to compare our outcomes whith the expectations based on the literature as well as on other empirical results.

Table 3

Stepwise regression analysis results of the influential factors of growth indicators

	Employment growth		Sales growth		Own capital growth		Cluster three factors		Cluster two factors	
	Coefficient	Significant	Coefficient	Significant	Coefficient	Significant	Coefficient	Significant	Coefficient	Significant
Constant	–0.934		0.514		–11.74		0.367	**	0.121	
Personal demographic										
Age of entrepreneur	1.759	*			7.313	***				
Age of entrepreneur sq.	–0.387	**			–1.994	*				
Personal behavioural										
Other business experience			12.771	**	–3.079	*			0.158	**
Business demographic										
Size			6.664	**			0.100	**	0.091	**
Business age	–0.341	***	–6.097	**	1.672	*	–0.115	***	–0.126	***
Legal form	0.765	***								
Number of founders.	–0.976	***					–0.214	***	–0.193	***
Foreign ownership	–1.873	**			15.717	***	–0.499	**		
Business behaviour										
Organization change					11.367	***				
Investment	0.312	***			2.780	***	0.072	**	0.074	**
Export			44.096	***			0.318	*	0.366	**
Technology innovation	1.096	**					0.240	*	0.305	**
Strategic orientation	0.400	*							0.136	**
Number of cases	313		313		286		286		313	
Adjusted R^2/*Wilks' Lambda*	0.137		0.106		0.144		0.109		0.133	
F-test	6.519	***	8.791	***	8.056	***	5.983	***	6.990	***

Note.* report a significance level of 10 percent, ** of 5 percent, and *** of 1 percent.

Table 4

Pearson correlation coefficients between growth, personal and business characteristics

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	EMPCREG	1.00																												
2	SALESCHREG	**0.50**	1.00																											
3	CAPITCHREG	0.04	0.02	1.00																										
4	CLUSTERREG3	**0.88**	**0.88**	*0.10*	1.00																									
5	CLUSTERREG2	**0.87**	**0.87**	0.04	1.00	1.00																								
6	AGEENTR	*-0.09*	0.02	-0.05	-0.06	-0.04	1.00																							
7	AGEENTRSQ	***-0.11***	0.02	-0.07	-0.07	-0.05	**0.98**	1.00																						
8	GENDER	0.06	0.04	-0.07	0.07	0.06	0.07	0.08	1.00																					
9	FAMILY	0.07	-0.02	-0.07	0.05	0.03	**-0.14**	**-0.14**	**-0.11**	1.00																				
10	OPPORTUNITY	0.01	0.06	0.02	0.05	0.04	-0.02	-0.03	0.02	***-0.12***	1.00																			
11	NECESSITY	0.01	0.02	0.02	-0.01	0.02	0.01	0.01	-0.08	0.06	-0.30	1.00																		
12	INDEPENDENT	0.02	0.00	-0.04	0.01	0.01	-0.06	-0.05	0.03	*0.11*	0.02	**-0.16**	1.00																	
13	STARTEXP	0.08	**0.14**	-0.06	0.08	***0.13***	-0.03	-0.03	0.05	0.01	0.08	***-0.11***	0.06	1.00																
14	SIZE	*0.10*	**0.19**	**0.22**	**0.16**	**0.17**	**0.24**	**0.22**	**0.18**	**-0.24**	**0.12**	-0.09	-0.05	0.07	1.00															
15	AGEBUS	***-0.11***	-0.05	**0.16**	-0.08	*-0.09*	**0.31**	**0.30**	0.03	-0.08	0.05	-0.01	0.00	0.00	**0.40**	1.00														
16	LEGFORM	*0.09*	***0.13***	0.02	*0.11*	***0.13***	**0.25**	**0.24**	**0.24**	**-0.29**	***0.13***	**-0.14**	0.04	**0.15**	**0.43**	**0.17**	1.00													
17	FOUNDER	**-0.14**	0.00	0.04	*-0.11*	-0.08	**0.25**	**0.25**	**0.13**	**-0.50**	**0.14**	-0.05	-0.09	0.00	**0.37**	**0.18**	**0.52**	1.00												
18	FORIGNOWN	***-0.11***	0.07	**0.14**	*-0.10*	-0.02	0.04	0.04	0.02	**-0.24**	***0.12***	-0.06	**-0.15**	0.08	0.03	-0.04	0.08	0.03	1.00											
19	DIVERSIF	*-0.10*	-0.04	-0.01	***-0.13***	-0.08	***0.13***	***0.11***	-0.03	0.00	0.00	-0.06	-0.02	0.06	0.06	***0.14***	**0.16**	*0.10*	*0.10*	1.00										
20	INVEST	**0.17**	**0.16**	**0.30**	**0.18**	**0.19**	0.08	0.07	**0.16**	**-0.24**	***0.11***	-0.07	0.02	0.05	**0.52**	**0.26**	**0.33**	**0.25**	0.05	0.02	1.00									
21	INNOVPROD	**0.14**	0.03	0.05	***0.12***	*0.10*	0.02	0.02	**0.21**	0.06	-0.05	0.00	*0.09*	0.01	**0.20**	**0.22**	0.06	0.00	-0.04	-0.07	**0.21**	1.00								
22	INNOVTECH	**0.18**	***0.13***	0.01	**0.16**	**0.18**	0.03	0.03	*0.09*	-0.01	0.00	0.03	0.00	0.04	**0.21**	0.08	**0.15**	0.05	-0.02	*-0.10*	**0.20**	**0.29**	1.00							
23	LEGFORMCH	0.00	-0.01	-0.05	-0.01	-0.01	-0.03	-0.03	0.05	0.07	0.06	-0.03	-0.04	0.04	**0.00**	*0.10*	0.14	***0.04***	-0.06	-0.02	0.05	0.07	0.08	1.00						
24	OWNERCH	-0.03	0.08	0.08	-0.01	0.03	0.02	0.02	*0.11*	**-0.19**	-0.03	-0.08	0.03	***0.13***	*0.07*	*0.09*	0.19	0.19	*0.09*	***0.04***	*0.09*	*0.07*	**-0.01**	*0.09*	1.00					
25	ORGCH	***0.13***	**0.15**	**0.21**	***0.12***	**0.16**	-0.02	0.00	**0.14**	-0.04	0.01	-0.07	0.08	**0.15**	**0.26**	***0.11***	***0.13***	*0.10*	0.08	-0.01	**0.22**	**0.19**	**0.23**	-0.01	**0.20**	1.00				
26	PLANTEXP	**0.22**	***0.12***	*0.10*	**0.23**	**0.20**	-0.01	-0.02	0.03	0.07	***0.13***	0.01	0.06	0.00	0.09	-0.01	0.06	0.02	-0.05	0.01	0.07	0.07	*0.09*	-0.08	-0.03	*0.10*	1.00			
27	GEOGREXP	0.08	***0.12***	0.04	0.08	*0.11*	0.01	-0.01	0.08	-0.06	0.02	-0.01	0.01	0.09	**0.17**	-0.03	***0.13***	*0.09*	0.08	-0.02	**0.20**	0.08	***0.11***	-0.03	0.09	**0.17**	***0.11***	1.00		
28	EXPORT	*0.09*	**0.27**	0.08	**0.18**	**0.21**	0.07	0.07	**0.14**	-0.08	0.02	-0.04	0.03	0.01	**0.33**	0.07	**0.18**	*0.09*	*0.10*	***-0.12***	**0.25**	**0.15**	**0.14**	-0.06	0.09	**0.21**	0.06	**0.27**	1.00	
29	STRATORIENT	0.07	0.07	0.04	0.08	0.08	-0.03	0.00	0.03	-0.05	-0.02	0.02	0.03	-0.07	0.03	0.04	*0.09*	0.09	-0.06	-0.04	-0.02	0.08	0.00	-0.01	0.00	0.04	0.04	-0.06	0.05	1.00

Note. Bold numbers mean a significance level of 1 percent, bold and italic of 5 percent, and italic of 10 percent.

3.1. The effect of personal demographic factors

Out of the *personal demographic* variables, the effects of age and gender on business growth have been the most widely examined. One the one hand, a younger age of the entrepreneur may imply higher risk aversion and a stronger attitude towards growth as compared to older entrepreneurs. On the other hand, as the entrepreneur gets older, he/she is learning more and more about business growth and has more and more managerial experience, implying a positive impact of age on growth. Following *Storey* [1994] we consider not only the linear but also the quadratic effect of age.

Our finding show mixed results. The parameters of the age and the square of the age of the entrepreneur seem to support the quadratic effect of age on growth in the cases of employment and own capital growth. Therefore the business of a middle age person who possesses enough resources and experiences on the one hand and has a young spirit to drive achievement on the other hand, grows more than younger or older person's firm. This age effect is not significant in terms of sales growth. Based on Table 4, young entrepreneurs' business grows faster, so the sign of the parameter is such as expected, but the overall effect is insignificant. In the cases of the composite growth measures neither the age nor the square of age is significant.

Gender is also a frequently examined factor of growth. In general, female entrepreneurs are believed to be more "growth averse" than males. However, as found in several empirical studies after controlling for demographic characteristics and for industry, the differences between male and female businesses disappear (*Carter–Williams–Reynolds* [1997], *Chell–Baines* [1998], *Davidsson–Achtenhagen–Naldi* [2005], *Johnsen–McMahon* [2005], *Du Rietz–Henrekson* [2000], *Storey* [1994]). Since we found gender to be insignificant, our result support the idea that gender is not an important demographic determinant of growth.

Some studies have found that family owned businesses are more averse in hiring new employees and this averseness limits the growth of these types of businesses (*Ward* [1997], *Cromie et al*. [1999]). However, recent studies find no differences between family and non-family owned business growth orientation (*Bogaert et al* [1999], *Malinen–Stenholm* [2004]). Since we found that the growth in family and non-family businesses do not differ from each other; therefore, we reject the notion about the importance of family ownership. The correlation coefficients also show mixed sign and a very weak correlation between family orientation and growth.

3.2. The effect of personal behavioural factors

The opportunity/necessity motives of business establishment on growth and survival are amongst the most frequently analyzed connections. In general, opportunity oriented business owners are better prepared, more skilled, and have a stronger desire to grow than necessity driven entrepreneurs who start the business mostly because of unemployment or the fear of it (*Autio* [2005], *Reynolds et al.* [2001], *Reynolds–*

Bygrave–Autio [2004], *Storey* [1994]). However, this may not be true in Hungary where entrepreneurs, independent of start-up motivation, had no previous market experience, proper business skills or education. This is particularly true for businesses started in the early 1990s. According to our findings, neither opportunity nor necessity motives seem to be significant factors of growth. The correlation coefficients present generally surprising results: both opportunity and necessity business establishment motives have a weak but positive influence on growth. This can be seen as a Hungarian special feature resulting from the transition to market economy.

Experience in managing other businesses or prior ownership can decrease the probability of failure (*Storey* [1994]). We also expect that multiple business owners possess more of the managerial skills required for successful business growth. This expectation is supported by our outcome in the sales and the two factor cluster cases. A surprise is that experience negatively influence the growth of own capital, however, the level of significance is marginal only at 10 percent. For this phenomenon we could not find a reasonable explanation.

3.3. The effect of business demographic factors

Denying Gibrat's law, most empirical studies found that business size is not a stochastic determinant of growth: smaller businesses grow at a faster rate (*Davidsson et al.* [2000], *Dunne–Roberts–Samuelson* [1988], *Evans* [1987], *Hall* [1987], *Storey* [1994]). A popular interpretation of this finding is that smaller firms grow more quickly in order to achieve the minimum efficient size (*Jovanovic* [1982]). A few empirical studies have found that larger businesses grow faster (*Glancey* [1998], *Storey* [1994]). However, these authors did not provide a clear justification for this phenomenon. Since Hungarian business growth in the smaller size sector is considered to be constrained it may happen that size has a positive influence on it (*Laky* [1998], *Kőhegyi* [2001], *Szerb–Ulbert* [2002]). Based on our finding, i.e. the size of the business positively influence growth, this later argument seems to be supported.

The examination of the effect of age on business growth is similarly interpreted as in the case of the age of the entrepreneur: younger businesses are expected to grow faster than old firms. Our findings support this statement except in the case of own capital. The implication of this outcome is that younger Hungarian businesses face serious capital limitations therefore in the early years of growth they select a non-capital intensive growth path. Note, that out of the business demographic characteristics, age seems to be the most important explanatory variable.

The legal form, especially the difference between limited and unlimited liability forms, can make a difference in terms of growth. More rapid growth rates are experienced in the case of limited liability businesses as compared to partnerships or sole proprietorship legal forms (*Davidsson et al.* [2000], *Storey* [1994]). The correlation coefficients of our data set imply a positive but weak relationship between growth and the legal form of business. According to the stepwise regression result, establishing a limited liability form of business is associated with more rapid employment growth rates, but the overall effect is negligible.

The number and the composition of business owners can also be an important determinant of growth. Since the business requires different managerial skills it is expected that firms with more numerous owners can grow faster (*Storey* [1994]). By surprise, our result is the opposite what most other studies found, but fits well to our common knowledge about Hungarian ownership problems. So, there is a potential counterbalancing effect: if the number of owners is increasing then the potential disagreement about business strategy and growth can be higher, implying a negative growth rate effect. The effect is the strongest in the case of employment growth, but highly significant also in the cases of composite growth measures. The composition of the owners can also be an important influential growth factor. In this regard, the presence of foreign owners who posses more relevant market experience and managerial skills than Hungarian ones is particularly vital. It is expected that foreign owners bring fresh blood to the business and therefore these businesses grow at a faster rate. Surprisingly, we found that the presence of foreign owners has a negative significant effect on employment growth and positive significant effect on the growth of own capital. The overall effect of foreign ownership in the case of composite growth rate is negative.

3.4. The effect of business behavioural factors

Among *business behavioural* factors, the degree of diversification has not been empirically researched. More diversified businesses are less exposed to industry specific changes and, according to the portfolio theory, can expect a smoother growth rate with less frequent sudden changes. However, increased diversification can lead to the lost of strategy focus. Since small businesses posses less resources than larger firms, the inefficient use of the limited resources can negatively influence growth rates. As it shown in Table 3, diversification alone seems to negatively influence growth rates, but according to the stepwise regression results, the strength of this effect seems be insignificant. The impact of investment on business growth is probably the most straightforward effect. It is expected that investment positively affects business growth, and this statement is supported by our regression results. The coefficient of investment is insignificant only in the case of sales growth. In general, investment propensity seems to be the most important behavioural influential factor of growth.

In general, any kind of innovation is expected to positively influence business growth (*Brouwer–Kleinkleht–Reijen* [1993], *Roper* [1997], *Storey* [1994]). However, innovation activity in the Hungarian small business sector is generally weak. Most small firms introduce only marginally new products or technology (*Török–Papanek* [2005], *Inzelt* [2003], *Inzelt–Szerb* [2006], *Pakucs–Papanek* [2003]). The new products are marginal improvements in existing products and are mainly introduced at the declining stage of the life-cycle of existing products. They are barely enough to maintain competitive position (*Inzelt–Szerb* [2006]) and not sufficient to induce substantial growth. Therefore, the effect of innovation effort on Hungarian business growth is insignificant. Our findings present mixed evidences: based on Table 4, both product and process (technology) innovations positively and significantly influence any kinds of growth measures. However, the stepwise regression outcome implies that product innovation is less important, insignificant factor of growth while technology innovation is vital for employment growth as well as when growth is measured as a composite factor.

A change of ownership, legal form or organization structure in the business can bring new owners, new expertise, and new impulsion. Moreover, all of these changes can be considered as innovation in the sense of *Schumpeter* [1934]. Therefore it is expected to have a positive influence on growth. However, our results imply a modest, non-significant influence of these factors expect organizational change that has a positive influence on capital change. (See Table 3.) The expansion of a business can be interpreted in different ways. An increase of the number of plants or the acquisition of customers in different places is expected to influence business growth positively. Additionally, it is also anticipated that internationalized businesses exporting their product to foreign counties grow at a faster rate than businesses selling only in domestic markets (*Davidsson et al.* [2000], *Storey* [1994]). Stepwise regression results underline the initial expectation in terms of the export but not in the case of plant expansion. However, Table 4 shows strong, significant correlation among plant expansion and different growth measures. It seems that this significance disappeared due to cross-correlation among different independent variables. Exports are found to be the single most important factor of sales growth. A strong influence of exports are also noticed in the composite growth measure cases.

Planned business strategy can be vital for successful growth. If the business has a business plan or other written document on strategy, then it is expected to deal with growth problems more successfully than those businesses with no such plan. Moreover, written plans can be interpreted as a sign of growth orientation of the business owner/manager. According to our results, strategy formulation is only significant in the cases of two factor composite growth measures and employment change. However the level of significance, as well as the strength of influence, seem to be modest. Moreover, strategic orientation seems to have a limited influence on growth, according to Table 4. Overall, the results are ambiguous: out of 11, four behavioural features have proved to be significant in business growth, in at least two cases.

4. Interpretation of findings

While data limitations does not make possible to test our model fully, together with OLS regressions and Person's correlation methods, an overall picture of the growth factors emerges. It is clear, that different measures of growth are influenced by different variables and different magnitudes, so it does matter what kind(s) of growth measures are applied. Our experience suggests on the use of not only one but more growth measures. An example for supporting this statement can be the comparison of capital change and of sales change: ownership experience in other business has a positive effect on sales growth but a negative one to own capital changes. A possible explanation is that business owners with multiple ownerships can spear their money amongst many firms. A more realistic explanation can be made in terms of foreign ownership. By surprise, foreign ownership has a negative effect on employment change but a positive one to own capital growth. Even before completing this study, we knew that foreign owned businesses were better capitalized, though recently we have evidences that foreign businesses are rather fire than hire new employees, therefore they prefer a capital intensive growth. Without multiple growth measures these phenomena would have been hidden.

In general, personal demographic characteristics seem to have the least influential direct effect on growth with one notable exception: the age of the owner. In this respect, a quadratic age effect on business growth is supported. It implies that a business with a median (average) aged entrepreneur (around 40) who has some business management experience, and a need for achievement has the highest rate of business growth. In older age, stability over growth is probably more appreciated. Neither being a female nor family ownership implies lower growth rates. However, there are high correlations between personal and business demographics factors. It means, that larger and older businesses are owned by older, most of the times male entrepreneurs. While family owned businesses do not grow slower than other type of firms, these ventures are generally smaller, have smaller number of owners and select the simpler unlimited liability sole proprietorship or general partnership legal forms.

Business demographic factors are found to be important determinants of growth. Out of this, size is definitely the most significant feature of a business growth. Larger businesses are older, operate in limited liability forms, and have multiple owners. Moreover, larger firms invest and export more, renew their products and technology (innovate) more frequently, attract more customers from different geographic location and change their inside organizational structure more often than smaller businesses do. Since most developing countries empirical studies found negative relationship between firm size and growth, this outcome may reinforce that there is a growth problem in the Hungarian micro and small size business sector. Most empirical studies have found that smaller and younger businesses grow at a higher rate. In our analysis, the age effect is found to correspond to these previous results. There is only one exception: the growth of the capital is higher at larger businesses, reinforcing the limited capital raising capability of the young and risky businesses. In addition, younger firms invest and export less and renew their technology rarely. All of these findings correspond well to other empirical outcomes of the problems of young businesses.

The legal form of the business makes a difference in growth: limited liability firms grow faster than unlimited liability or sole proprietorship firms. The number of founders is another element where our outcome differs from other empirical results. It is believed that a larger number of owners have a positive influence on growth since the experience as well as the responsibility can be shared. However, if the number of owners is too high, then owners' conflicts and potential disagreements can demolish business growth. These businesses are established because of opportunity reasons with relatively large size, diversified product portfolio, and they invest more as compared to other businesses owned by less numerous members. At first, the negative effect of the number of founders on employment growth seems surprising. In the case of Hungary the large number of founders, who probably not just own but work at the business, could limit further employment. Potential disagreement amongst owners may cause frequent organizational changes and over time the number of owners decreases. A further possible reason of the limited growth is the lack of strategy focus.

Of the personal behavioural factors, opportunity or necessity start-up motives do not have an influence on business growth. This is different from other empirical studies, but it is consistent with the short history of the Hungarian market economy. In the transitional period, there is not much difference between opportunity and necessity entrepreneurs in terms of managerial experiences and attitudes so similar growth rates are not surprising. However, it is also prevailed that opportunity oriented businesses are larger, generally established in limited liability forms, have more founders, invest more and open new production plants more frequently than necessity oriented businesses.

Only one individual behavioural factor has proved to influence growth significantly. Previous or present business ownership and managerial experience in other business is an important individual factor in the cases of sales and capital growth. However, the negative sign of the parameter is unpredictable in the case of capital growth.

Overall, regardless of how growth is measured business behavioural variables are the most important factors of business growth. Investment, the introduction of new technology, exports, organizational changes and strategic orientation are found to be the major characteristics that influence growth. Out of these factors, investment, organizational changes and strategic orientation effects do not require further explanation, their impact is obvious. It is a surprise that product innovation is an insignificant factor of growth. This phenomenon can be explained by the late introduction of new products when the previous products are already in the declining stage of their life cycle or possibly it is the result of limited product improvements that are not valued by customers. Technology changes (innovation) influence growth rates positively, and this effect is significant in the cases of employment growth and of the two composite factors. The geographic expansion of sales is found to have an insignificant effect on growth rates unlike exports which is a major factor of sales growth.

5. Conclusion

Based mainly on entrepreneurship and small business growth literature, we constructed a conceptual model of growth applicable to small businesses in the first part of this paper. The influential factors of growth are grouped into four categories: *1*. personal demographic, *2.* personal behavioural, *3.* business demographic and *4.* business behavioural factors. After reviewing the potential application of the growth model in the Hungarian SME sector, we formulated our expectations about the factors of growth. We tested the model by relying on stepwise regression analysis and Pearson's correlation methods. For measuring growth, different factors are identified: the change in the number of employees, the change in real growth of sales, and the change in real growth of equity. In addition, two composite factors of growth are also incorporated in the model. Growth rates in each of the five cases were calculated as the slope of the regression over the examined time period. Out of the four major factors, business behavioural features are the most important while business demographic variables rank second. Personal characteristics show a limited influence on growth rates with the exception of the age of the business owner and the experience in other business start-ups. However, personal characteristics show high correlation with certain business demographic and behavioural variables, which calls for further research.

Examining growth on a micro level is a multidimensional problem focusing on either the measurement or the determinants of growth. An important aim of this study was to show an overall picture how to examine business growth according to a conceptual model and testing its validation by econometric method. This paper has probably raised some new questions while failed to answer all the old ones in a satisfying way. However, we are convinced that the analytical and conceptual methods developed to investigate advanced economy businesses have proved to be also applicable in the case of Hungary. Our results are sometimes contradictory which calls for further research. Since behavioural characteristics play the major role in explaining business growth, usual business registry data are insufficient and a careful and representative survey is necessary for further examination.

References

ACS, Z. J. (ed.) [1996]: *Small firms and economic growth*. Edward Elgar. Cheltenham.

ACS, Z. ET AL. [2003]: *The missing link: The knowledge filter, entrepreneurship and endogenous growth.* Working paper.

ÁCS, J. Z. ET AL. [2002]: *GEM, Vállalkozások Magyarországon 2001.* PTE. Pécs.

AUTIO, E. [2005]: *Report on high-expectation entrepreneurship*. www.gemconsortium.org

BAUMOL, W. J. [2002]: *The free-market innovation machine: Analyzing the growth miracle of capitalism*. Princeton University Press. Princeton.

BECKER, B. E. ET AL. [1997]: HR as a source of shareholder value: Research and recommendation. *Human Resource Management.* Vol. 36. No. 1 p. 39–48.

BERGER, P. G. – OFEK, E. [1995]: Diversification's effect on firm value. *Journal of Financial Economics.* Vol. 37 No. 1. p. 39–65.

BIRCH, D. [1987]: *Job generation in America*. Free Press. New York.

BROUWER, E. – KLEINKNEHT, A. – REIJNEN, J. [1993]: Employment growth and innovation at the firm level. *Journal of Evolutionary Economics*. Vol 3. No. 2. 153–159.

BOGAERT, I. ET AL. [1999]: *Sales growth in family businesses: The case of Belgian wholesale sector.* http://cref.u-bordeaux4.fr.

CARTER, N. M. – WILLIAMS, M. – REYNOLDS, P. [1997]: Discontinuance among firms in retail: The influence of initial resources, strategy and gender. *Journal of Business Venturing*. Vol. 12. No. 2. p. 125–145.

CZAKÓ, E. ET AL. [1995]: A kisvállalkozások néhány jellemzője a kilencvenes évek elején. *Közgazdasági Szemle.* Vol. 42. No. 4. p. 399–419.

CHELL, E. – BAINES, S. [1998]: Does gender affect business performance? A study of microbusinesses in the business services in the United Kingdom *Entrepreneurship and Regional Development.* Vol. 10. No. 2. 117–135.

CHEEL, E. – HAWORTH, J. – BREARLEY, S. [1991]: *Entrepreneurial personality*. Routledge. London. New York.

CROMIE, S ET AL. [1999]: Family firms in Scotland and Northern Ireland: An empirical investigation. *Journal of Small Business and Enterprise Development.* Vol. 6. No. 3. p. 253–266.

DAVIDSSON, P. [2003]: The domain of entrepreneurship research: Some suggestions. In: *Katz, J. – Stepherd, S.* (ed.): *Advances in entrepreneurship firm emergence and growth.* Elsevier, JAI Press. Oxford.

DAVIDSSON, P. – ACHTENHAGEN, L. – NALDI, L. [2005]: *Research on small business growth: A review*. EFMD 35th. EISB Conference on sustaining the entrepreneurial spirit over time. Barcelona. Working paper.

DAVIDSSON, P. – DELMAR, F. – WIKLUND, J. [2002]: Entrepreneurship as growth. In: *Sexton, D. et al.* (ed.): *Growth as entrepreneurship, in strategic entrepreneurship*. Blackwell. Oxford.

DAVIDSSON, P. ET AL. [2000]: *Factors underlying business growth in Sweden*. ICSB World Conference. Brisbane. Working paper.

DAVIDSSON, P. – WIKLUND, J. [1999]: *Theoretical and methodological issues in the study of firm growth.* Jonköping Business School. Working paper.

DELMAR, F. – DAVIDSSON, P. – GARTNER, W. B. [2003]: Arriving at the high growth firm. *Journal of Business Venturing*. Vol. 18. No. 2. p. 189–216.

DODGSON, M. – ROTHWELL, R. (ed.) [1995]: *The handbook of industrial innovation*. Edward Elgar. London.

DRUCKER, P. F. [1985]: *Innovation and entrepreneurship: practice and principles*. Butterworth-Heinemann. Oxford.

DU RIETZ, A – HENREKSON, M. [2000]: Testing the female underperformance hypothesis. *Small Business Economics.* Vol. 14. No. 1. p. 1–10.

DUNNE, T. – M. ROBERTS. M. – SAMUELSON, L. [1988]: Pattern of firm entry and exit in U.S. manufacturing industries. *Rand Journal of Economics.* Vol. 104. No. 4. p. 671–698.

EVANS, D. [1987]: The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *Journal of Industrial Economics.* Vol. 35. No. 2. p. 567–581.

GARTNER, W. B. [1990]: What are we talking about when we talk about entrepreneurship? *Journal of Business Venturing*. Vol. 5. No. 1. p. 15–28.

GLANCEY, K. [1998]: Determinants of growth and profitability in small entrepreneurial firms. *International Journal of Entrepreneurial Behaviour and Research.* Vol. 4. No. 1. p 18–27.

HALL, B. H. [1987]: The relationship between size and firm growth in the U.S. manufacturing sector. *Journal of Industrial Economics.* Vol. 35. No. 2. p. 583–606.

HALPERN, L. – KŐRÖSI, G. [2001]: Efficiency and market share in Hungarian corporate sector. *The Economics of Transition*. Vol. 9. No. 3. p. 559–592.

HÉBERT, R. F. – LINK, A. N. [1989]: In search of the meaning of entrepreneurship. *Small Business Economics*. Vol. 1. No. 1. p. 39–49.

HISRICH, R. D. – PETERS, M. P. – SHEPHERD, D. A. [2005]: *Entrepreneurship*. McGraw-Hill, Irwin. New York.

HOLM, D. B. – ERKISON, K. [1999]: Creating value through mutual commitment to business network relationship. *Strategic Management Journal.* Vol. 20. No. 5 p. 467–486.

INZELT A. [2003]: A kicsik K+F és innovációs tevékenysége. *Külgazdaság.* Vol. 47. No. 11. p. 24–42.

INZELT A. – SZERB L. [2006]: *The innovation activity in a stagnating county of Hungary*. Working paper.

IYIGUN, M. – OWEN, A. [1997]: *Risk, entrepreneurship and human capital accumulation.* www.federalreserve.gov

JÁVOR I. – ROZGONYI T. [1995]: Vállalkozás, vállalkozásfejlődés, vállalkozástörténetek. *Társadalomkutatás.* Vol. 13. No. 3–4. p. 222–239.

JOHNSEN, G. – MCMAHON, R. [2005]: Owner-manager gender, financial performance and business growth amongst SMEs from Australia's Business Longitudinal Survey. *International Small Business Journal*. Vol. 23. No. 2. p. 115–142.

JOVANOVIC, B. [1982]: Selection and the evolution of industry. *Econometrica*. Vol. 50. No. 3. p. 649–670.

KIRZNER, I. M. [1979]: *Perception, opportunity and profit: studies in the theory of entrepreneurship*. University of Chicago Press. Chicago.

KNIGHT, F. H. [1990]: Risk, uncertainty and profit. In *Casson, M*. (ed.): *Entrepreneurship*. Edward Elgar Publishing Ltd. Hampshire.

KŐHEGYI K. [2001]: Növekvő és zsugorodó vállalkozások. *Közgazdasági Szemle*. Vol. 48. No. 4. p. 320–337.

KŐHEGYI K. [2005]: Vállalkozói hálózatok, *Közgazdasági Szemle.* Vol. 52. No. 5. p. 480–493.

KUZCI T. – MAKÓ CS. [2001]: A vállalkozók társadalmi erőforrásai és a vállalkozások növekedési képessége. *Vezetéstudomány* Vol. 31. No. 1. p. 46–55.

LAKI M. [1998]: A magánvállalkozások növekedési esélyei Magyarországon. *Külgazdaság.* Vol. 42 No. 4. p. 41–54.

LAKI M. [2001]: Az újonnan alapított magánvállalatok teljesítménye. *Közgazdasági Szemle*. Vol. 48. No. 11. p. 965–979.

LAKY T. [1998]: A kisvállalkozások növekedésének korlátai. *Szociológiai Szemle.* Vol. 8. No. 1. p. 23–39.

LUCAS, R. E. [1988]: On the mechanics of economic development. *Journal of Monetary Economics*. Vol. 22. No. 1. p. 3–42.

MAJOR I. [2002]: Miért (nem) sikeresek a magyar középvállalatok? *Közgazdasági Szemle.* Vol. 49. No. 10. p. 993–1014.

MALINEN, P. – STENHOLM, P. [2004]: *Who dares wins: Modeling of Finnish family business growth orientation*. ICSB World Conference. Working paper.

MCCLELLAND, D. [1961]: *The achieving society*. Van Nostrand. Princeton, NJ.

MÉSZÁROS T.– PITTI Z. [2003] A vállalkozási szerkezet módosulása különös tekintettel a középvállalkozásokra és az EU-sajátosságokra. *Európai Tükör.* Vol. 8. No. 1. p. 47– 69.

MORAN, P. – GHOSHAL, S. [1996]: Value creation by firms. In: *Keys, J. B. – Dosier, L. N.* (ed): *Academy of management best paper proceedings*. George Southern University. Statesboro.

MORRIS, M. H. – LEWIS, P. – SEXTON, D. [1994]: Reconceptualizing entrepreneurship: an input-output perspective. *SAM Advanced Management Journal*. Vol. 59. No. 1. p. 21–31.

MUNDRUCZÓ, GY. [1981]: *Alkalmazott regressziószámítás*. Akadémiai kiadó. Budapest.

PAKUCS, J. – PAPANEK, G. [2003]: Innovációs esélyek és lehetőségek. *Harvard Business Manager.* Vol. 4. No. 3. p. 46–57.

PARKER, S.C. – STOREY, D. J. – VAN WITTELOOSTUIJN, A. [2005]: *What happens to gazelles? The importance of dynamic management strategy*. www.dur.ac.uk. Working paper.

PENROSE, E. [1959]: *The theory of growth of the firm*. John Wiley and Sons. New York.

PORTER, M. E. [1985]*: Competitive advantage: Creating and sustaining superior performance*. Free Press. New York.

REID, G. [1993]: *Small business enterprise*. Routledge. London.

REYNOLDS, P. ET AL. [2001]: *Global entrepreneurship monitor 2001. Executive Report*. London Business School.London.

REYNOLDS, P. – BYGRAVE, W. – AUTIO, E. [2004]: *GEM 2003 Global Report*. Business School. London. Working paper.

ROMER, P. [1990]: Endogenous technological change. *Journal of Political Economy.* Vol. 98. No. 5. p. 71–102.

ROPER, S. [1997]: Product innovation and small business growth: A comparison of the strategies of German, UK and Irish companies. *Small Business Economics*. Vol. 9. No. 6. p. 523–537.

RAPPAI G. [2001]: *Üzleti statisztika Excellel*. Központi Statisztikai Hivatal. Budapest.

ROMÁN Z. [2002]: A kis- és középvállalatok a magyar gazdaságban. *Statisztikai Szemle.* Vol. 80. No. 8. p. 752–768.

ROMÁN Z. [2005]: A kis- és középvállalatok és a vállalkozás-statisztika helyzete. *Gazdaság és Statisztika*. Vol. 17(56) No. 4 p. 37–53.

SADLER-SMITH, E. ET AL. [2003]: Managerial behavior, entrepreneurial style and small firm performance. *Journal of Small Business Management*. Vol. 41. Iss. 1. p. 47–67.

SALAMONNÉ HUSZTY A. [2006]: Magyarországi kis- és középvállalkozások életútjának modellezése. *Competitio*. Vol. 5. No. 1. p. 51–68.

SCHUMPETER, J. [1934]: *The theory of economic development*. Harvard University Press. Cambridge.

SEXTON, D. L. [1997]: Entrepreneurship research needs and issues. In: *Sexton, D. L. – Smilor, R. W.* (ed.): *Entrepreneurship 2000*. Upstart Publishing Co. Chicago.

SHANE, S. – VENTAKARAMAN, S. [2000]: The promise of entrepreneurship as a field of research. *Academy of Management Review*. Vol. 25. No. 1. p. 217–226.

SOLOW, R. M. [1956]: A contribution to the theory of economic growth. *Quarterly Journal of Economics*. Vol. 70. No. 1. p. 65–94.

STEVENSON, H. H. – ROBERTS, M. J. – GROUSBECK, H. I. [1989]: *New business ventures and the entrepreneur*. IL: Irwin. Homewood.

STOREY, D. [1994]: *Understanding the small business sector*. Routledge. London.

SZERB L. – ULBERT J. [2002]: Entrepreneurial growth and the role of venture capital. In: *Varga A. – Szerb L.* (ed.): *Innnovation, entrepreneurship and regional development.* University of Pécs. Pécs.

SZERB L. ET AL. [2004]: Az új vállalkozások hatásai nemzetközi összehasonlításban. *Közgazdasági Szemle*. Vol. 51. No. 7–8 p. 679–698.

SZERB, L. [2005]: *Vállalkozásindítás, vállalkozói hajlandóság, és a vállalkozási környezeti tényezők alakulása Magyarországon a 2000-es évek első felében*. Pécsi Tudományegyetem. Pécs.

SZIRMAI, P. [2002]: *A kisvállalkozások fejlődési szakaszai és a szakaszváltások konfliktusai.* Kisvállalkozás –Fejlesztés Központ. Budapest.

TEECE, D. [1998]: Capturing value from knowledge assets: The new economy, markets for know-how, and intangible assets. *California Management Review.* Vol. 40. No. 3. p. 55–79.

TIMMONS, J. A. [1999]: *New venture creation: Entrepreneurship for the 21st century*. Irwin/McGraw-Hill. Boston.

TÖRÖK, Á. – PAPANEK, G. [2005]: *Az EU-tagországok innováció- és kkv politikájának kapcsolódása.* Magyar Vállalatgazdasági Kutatásokért Alapítvány. GKI Gazdaságkutató Rt. Budapest.

Transition economies: How much progress? [2002]. IMF Staff Papers. Vol. 48. Special Issue.

TSAI, W. – GHOSHAL, S. [1998]: Social capital and value creation: The role of Intrafirm networks. *Academy of Management Journal*. Vol. 41. No. 4. p. 464–476.

Transition: The first ten years [2002]. The World Bank. Washington, D.C.

UCBASARAN, D. – WESTHEAD, P. – WRIGHT, M. [2005]: *The focus of entrepreneurial research: Contextual and process issues*. www. vsb.nott.ac.uk

VAJDA Á. [1999]: Munkahelyteremtés a mikrovállalkozásokban. *Közgazdasági Szemle.* Vol. 46. No. 6. p. 530–547.

VENKATARAMAN, S. [1997]: The distinctive domain of entrepreneurship research: An editor's perspective. In: *J. Katz, J. – Brockhaus,* R. (ed.): *Advances in entrepreneurship, firm emergence and growth*. JAI Press. Greenwich.

WARD, J. [1997]: Growing the family business: Special challenges and best practices. *Family Business Review.* Vol. 10. No. 4. p. 323–338.

WEDIN, T. [2003]: *Value creation in industrial networks*. www.bath.ac.uk

WEINZIMMER, L. G. – NYSTROM, P. C. – FREEMAN, S. J. [1998]: Measuring organizational growth: Issues, consequences and guidelines. *Journal of Management.* Vol. 24. No. 2. p. 235–262.

WENNEKERS, S. – THURIK, R. [1999]: Linking entrepreneurship to economic growth. *Journal of Small Business Economics.* Vol. 13. No. 1. 27–55.

WIKLUND, J. – STEPHERD, D. [2005]: *Knowledge accumulation in growth studies: The consequences of methodological choices.* http://econ.geog.uu.nl

The quasi-regression form of calibration estimates

László Mihályffy,
Retired senior statistical adviser of the HCSO
E-mail: laszlo.mihalyffy@ksh.hu

For an arbitrary calibration estimator, an alternative expression called quasi-regression form is can be used in variance computations. In the case of simple random sampling it yields an explicit expression for the difference between the estimated variance of the arbitrary calibrated estimate and that of the generalized regression estimate.

KEYWORDS:
Estimations.

In the literature on calibration methods, the *Deville–Särndal* paper [1992] is a key reference. It is shown in that paper that under some mild conditions any calibration estimator is asymptotically equivalent to the generalized regression estimator (called henceforth GREG), and therefore the variance and the estimated variance of the latter may be used for the former. A small Monte Carlo study with simple random samples of size $n = 200$ from a population consisting of $N = 2000$ units has yielded practically the same variance for the most common calibration estimators in use.

In this paper a method is given to assign approximate variance and sample estimate of the variance – different from those of the GREG – to an arbitrary calibration estimator. By the Deville–Särndal principle, these variances will be quite close to their counterparts corresponding to the GREG, yet in some cases the difference may be interesting, and the extra computing needed is not substantial. The idea of our method is to re-write a given calibration estimator in a form similar to that of the GREG, and then the variance and the variance estimate can be determined in a similar way as in the case of the latter. The GREG in this paper plays the role of the baseline, therefore we begin with a brief review on that estimator.

Provided we are given a sample $\{1,\ 2,\ ...,\ n\}$ from a finite universe of size *N*, and the design enables the use of the Horvitz–Thompson estimator, consider the following problem referred to as (P1) in the subsequent considerations. Find the calibrated weights $w_1,\ w_2,\ ...,\ w_n$ by minimising the distance function

$$\sum_{j=1}^{n}\left(w_j - d_j\right)^2 / d_j \ , \qquad /1/$$

subject to the calibration constraints

$$\sum_{j=1}^{n} x_{ji} w_j = X_i \ , \qquad i = 1,\ 2,\ ...,\ m \ . \qquad /2/$$

In equations /1/ and /2/, $d_1,\ d_2,\ ...,\ d_n$ stand for the design weights, $x_{j1},\ x_{j2},\ ...,\ x_{jm}$ are the values of the auxiliary variables observed on sample unit *j*, and $X_1,\ X_2,\ ...,\ X_m$ are the population totals of the auxiliary variables. The unique solution of the problem (P1) for w_j can be given explicitly, and the calibrated total of some study variable y_j can be written as

$$\hat{Y}^{\text{reg}} = \hat{Y} + \sum_{i=1}^{m} b_i \left(X_i - \hat{X}_i\right). \qquad /3/$$

$\hat{Y}^{\text{reg}}$ is called generalized regression estimate of the population total Y; $\hat{Y}$, $\hat{X}_1$, ... and $\hat{X}_m$ are Horvitz–Thompson estimates based on the design weights d_j, and b_1, b_2, ..., b_m are generalized regression coefficients estimated from the sample.

To emphasize the baseline function of the GREG in this paper, the results coming from the problem (P1) will be denoted with symbols having a superscript $(.)^{\text{o}}$; thus e.g. w_1^{o}, w_2^{o}, ..., w_n^{o} will stand for the calibrated weights and /3/ will be re-written as

$$\hat{Y}^{\text{reg}} = \hat{Y} + \sum_{i=1}^{m} b_i^{\text{o}} \left(X_i - \hat{X}_i \right). \qquad /3a/$$

Matrix algebra will often be used in this paper hence we need matrix-vector notations, too. Some of the most important of those are as follows. The superscript $(.)^T$ denotes transpose of matrices or vectors;

$$\mathbf{d} = \left(d_1,\ d_2,\ ...,\ d_n\right)^T,$$

$$\mathbf{w}^{\text{o}} = \left(w_1^{\text{o}},\ w_2^{\text{o}},\ ...,\ w_n^{\text{o}}\right)^T,$$

$$\mathbf{y} = \left(y_1,\ y_2,\ ...,\ y_n\right)^T,$$

$$\mathbf{x} = \left(x_{ji}\right), \quad j = 1,\ 2,\ ...,\ n, \quad i = 1,\ 2,\ ...,\ m,$$

$$\mathbf{b}^{\text{o}} = \left(b_1^{\text{o}},\ b_2^{\text{o}},\ ...,\ b_m^{\text{o}}\right)^T,$$

$\mathbf{\Omega}$ is the diagonal matrix with entries d_1, d_2, ..., d_n in the main diagonal.

Note that

$$\sum_{j=1}^{n} d_j y_j = \mathbf{d}^T \mathbf{y} = \hat{Y};$$

by analogy we have

$$\mathbf{d}^T \mathbf{x} = \left(\hat{X}_1,\ \hat{X}_2,\ ...,\ \hat{X}_m\right).$$

Further notations:

$$X = \left(X_1,\ X_2,\ ...,\ X_n\right)^T,$$

$$\hat{X} = \left(\hat{X}_1,\ \hat{X}_2,\ ...,\ \hat{X}_m\right)^T.$$

Except for the last two symbols, matrices and vectors are denoted by bold-face letters that may be capital, lower case or even Greek characters. Note also that with these notations the vector $\mathbf{b}^{\mathrm{o}}$ of regression coefficients can be written as follows:

$$\mathbf{b}^{\mathrm{o}} = \left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1} \mathbf{x}^T \mathbf{\Omega} \mathbf{y} \ .$$

In some cases a generalized version of the problem (P1) is considered where the distance function /1/ has the following form:

$$\sum_{j=1}^{n} \left(w_j - d_j\right)^2 / q_j d_j \ , \qquad /1a/$$

and $q_1, q_2, \ ..., q_n$ are positive weights chosen properly. For any unit *j* in the sample or in the population, q_j can always be identified with the reciprocal of the variance σ_j^2 of the random variable Y_j in the super-population model, $j = 1, 2, ..., N$; see e.g. *Särndal*, *Swensson* and *Wretman* ([1992] p. 225–229.). However, the option of using weights q_j other than unity would have no impact on our conclusions therefore we assume throughout that $q_j = 1$ for all *j*. In any case, it is interesting to note that the estimator /3/ – or /3a/ – can be derived in two different ways: either by solving the calibration problem (P1) or by means of the super-population principle.

1. The general calibration estimator and its quasi-regression form

With the same assumptions on sample and universe as in the introductory section, consider the following calibration problem (P2). Find the calibrated weights $w_1, w_2, ..., w_n$ by minimising the distance function

$$F = F\left(w_1, w_2, ..., w_n, d_1, d_2, ..., d_n\right) \ , \qquad /4/$$

subject to the calibration constraints

$$\sum_{j=1}^{n} x_{ji} w_j = X_i \ , \qquad i = 1, 2, ..., m \ . \qquad /2/$$

and the individual bounds on the calibrated weights

$$L \leq w_j / d_j \leq U \ . \qquad /5/$$

The distance function F is supposed to be strictly convex and continuously differentiable at least twice. In the majority of cases it is also assumed that F is separable which means that it is of the form

$$F = \sum_{j=1}^{n} G\left(w_j, d_j\right),$$

where G is strictly convex and continuously differentiable at least twice; term j in this representation depends only on w_j and d_j.

Denote $\mathbf{w} = \left(w_1, w_2, ..., w_n\right)^T$ the unique solution of (P2) – distinguishing it in this way from the solution of (P1) – and denote $\hat{Y}^{\text{cal}}$ the calibrated estimate of Y with these weights. We point out the following.

Result 1. $\hat{Y}^{\text{cal}}$ can be written in form as follows:

$$\hat{Y}^{\text{cal}} = \hat{Y} + \sum_{i=1}^{m} b_i \left(X_i - \hat{X}_i\right) = \hat{Y} + \left(X - \hat{X}\right)^T \mathbf{b} \ , \qquad /6/$$

where $\mathbf{b} = \mathbf{b}^{\text{o}} + \mathbf{b}'$, and

$$\begin{aligned} \mathbf{b}' = \frac{\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}}{\left(X - \hat{X}\right)^T \left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1} \left(X - \hat{X}\right)} \left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1} \left(X - \hat{X}\right) \overset{\text{def}}{=} \\ \overset{\text{def}}{=} C_{\text{o}} \left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1} \left(X - \hat{X}\right) \end{aligned} .$$

Note that $\mathbf{b}$ depends on the problem (P2) only through the expression $\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}$, and that $\hat{X}$ depends on the sample and the design weights d_j.

Proof. Starting with the right-hand side of /6/, we have

$$\begin{aligned} \hat{Y} + \left(X - \hat{X}\right)^T \mathbf{b} = \hat{Y} + \left(X - \hat{X}\right)^T \mathbf{b}^{\text{o}} + \left(X - \hat{X}\right)^T \mathbf{b}' = \\ = \hat{Y}^{\text{reg}} + \left(X - \hat{X}\right)^T \mathbf{b}' = \hat{Y}^{\text{reg}} + \left(\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}\right), \end{aligned}$$

as was to be shown.

While Result 1 is almost trivial, expression /6/ is useful in examining the estimated variance of $\hat{Y}^{\text{cal}}$. It is easy to see that the existence of $\left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1}$ is sufficient for that of $\hat{Y}^{\text{reg}}$ and also for the "quasi-regression" representation /6/, thus the term "quasi-regression form of calibrated estimates" is justified.

2. Linearization and variance expressions

With the quasi-regression forms introduced in the preceding section, one should proceed in the same way as in the case of "ordinary" regression estimates.
To this end:

– first the quasi-regression estimate should be linearized, then
– the linearized expression can be treated as the Horvitz–Thompson estimate of a total, and
– expressions for the variance and the sample estimate of the variance should be identified, and finally,
– the unknown population values in the variance estimate from the sample should be replaced by the corresponding sample estimates.

Before starting this procedure, the population value of the quasi-regression coefficients **b** should be found. This will be done for the two terms of $\mathbf{b} = \mathbf{b}^{\mathrm{o}} + \mathbf{b}'$ separately. By the principle of the super-population model, the population value of $\mathbf{b}^{\mathrm{o}}$ is $\mathbf{B}^{\mathrm{o}}$, the vector of regression coefficients in the population ($\mathbf{B}^{\mathrm{o}} \neq E\left(\mathbf{b}^{\mathrm{o}}\right)$). As for **b'**, it is straightforward to take the expectation **B'** of **b'** over all samples in the design in consideration as population value. In cases where $(\mathbf{x}^T \boldsymbol{\Omega} \mathbf{x})^{-1}$ does not exist we take $\mathbf{b}^{\mathrm{o}} = \mathbf{b}' = \mathbf{b} = 0$. The population value of **b** is then defined as $\mathbf{B} = \mathbf{B}^{\mathrm{o}} + \mathbf{B}'$, its components will be denoted by $B_1, B_2, \ldots, B_m$.

Now we have to linearize $\hat{Y}^{\mathrm{cal}}$ given by /6/. This estimated total depends on $\hat{Y}$, $\hat{X}_1, \hat{X}_2, \ldots \hat{X}_m$, and a certain number of other sample-depending values determined basically by the distance function F in /4/. Denote $\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_h$ these arguments of $\hat{Y}^{\mathrm{cal}}$; we shall see soon that we need not to have much information on them. Differentiating yields

$$\partial \hat{Y}^{\mathrm{cal}} / \partial \hat{Y} \equiv 1;$$

$$\partial \hat{Y}^{\mathrm{cal}} / \partial \hat{X}_i = \sum_{\mathrm{k}=1}^{\mathrm{m}} \frac{\partial b_k}{\partial \hat{X}_i} \left(X_k - \hat{X}_k \right) - b_i, \qquad i = 1, 2, \ldots, m;$$

$$\partial \hat{Y}^{\mathrm{cal}} / \partial \hat{z}_i = \sum_{k=1}^{m} \frac{\partial b_k}{\partial \hat{z}_i} \left(X_k - \hat{X}_k \right), \qquad i = 1, 2, \ldots, h.$$

Setting the arguments in the last two relations equal to the corresponding population values implies

$$\partial\hat{Y}^{\mathrm{cal}} / \partial\hat{X}_i|_{\hat{X}_i = X_i} = -B_i\,, \quad i = 1,\ 2,\ \ldots,\ m\,, \text{ and } \partial\hat{Y}^{\mathrm{cal}} / \partial\hat{z}_i|_{\hat{z}_i = z_i} = 0\,.$$[1]

This suggests that $\hat{Y}^{\mathrm{lin}}$, the linearized version of $\hat{Y}^{\mathrm{cal}}$ can be written as follows:

$$\hat{Y}^{\mathrm{lin}} = Y + \left(\hat{Y} - Y\right) + \sum_{\mathrm{k}=1}^{\mathrm{m}} -B_k\left(\hat{X}_k - X_k\right) = \hat{Y} + \sum_{k=1}^{m} B_k\left(X_k - \hat{X}_k\right), \qquad /7/$$

i.e. the linearization yields that the quasi-regression coefficients b_i are replaced by the corresponding population values. From now on, variance expressions for $\hat{Y}^{\mathrm{cal}}$ are derived in the same way as in the case of the ordinary regression estimator. The approximate variance of $\hat{Y}^{\mathrm{cal}}$ is the variance of $\hat{Y}^{\mathrm{lin}}$, and since $\sum_k B_k X_k$ is constant over all samples, we have

$$AV\left(\hat{Y}^{\mathrm{cal}}\right) = Var\left(\hat{Y}^{\mathrm{lin}}\right) = Var\left(\hat{Y} - \sum_{k=1}^{m} B_k \hat{X}_k\right) = Var\left(\hat{Z}\right),$$

where $\hat{Z}$ is the total of the residuals $z_j = y_j - \sum_{k=1}^{m} B_k x_{jk}$ weighted with the design weights d_j, and $Var\left(\hat{Z}\right)$ is computed with the variance formula of the Horvitz–Thompson estimator. The sample estimate of the variance is also based on the residuals z_j, but the unknown population values B_k should be replaced by the corresponding sample values b_k; moreover, Deville and Särndal advocate the use of calibrated weights w_j in variance estimates rather than that of d_j. It should be emphasized that in this way the estimated variance of $\hat{Y}^{\mathrm{cal}}$ – and not that of $\hat{Y}^{\mathrm{reg}}$ – is determined; and in practice presumably not the Yates–Grundy formula

$$var\left(\hat{Y}^{\mathrm{cal}}\right) \approx var\left(\hat{Z}\right) = \sum_i \sum_{j>i} \frac{\left(\pi_i \pi_j - \pi_{ij}\right)}{\pi_{ij}} \left(z_i / \pi_i - z_j / \pi_j\right)^2$$

will be used, but e.g. the jackknife method.

In the particular case of simple random sampling an explicit expression can be given for $var\left(\hat{Z}\right)$. We have the following.

Result 2. Assume that the design is simple random sampling without replacement and one of the auxiliary variables assumes the value 1 for each unit of the popula-

[1] The notation is simplified; all arguments in the partial derivatives should set equal to the corresponding population values.

tion.[2] In this case the following relation holds for the sample estimate of the variance of $\hat{Y}^{\text{cal}}$:

$$var\left(\hat{Y}^{\text{cal}}\right) \approx var\left(\hat{Z}\right) = var\left(\hat{Y}^{\text{reg}}\right) + var\left(\hat{X}b'\right) \qquad /8/$$

where $z_j = y_j - \sum_{k=1}^{m} b_k x_{jk}$ and $\hat{Z} = {N}/{n} \sum_{j=1}^{n} z_j$. Furthermore,

$$var\left(\hat{X}\mathbf{b}'\right) < \frac{(1\text{-}f)N^2}{n(n-1)} \frac{\left(\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}\right)^2}{\left(X - \hat{X}\right)^T \left(\mathbf{x}^T\mathbf{x}\right)^{-1} \left(X - \hat{X}\right)}, \qquad /9/$$

where $f = n/N$.

Proof. It is easy to see that the well-known estimated variance for an estimated total under simple random sampling (see *Cochran* [1977] p. 26.) can be re-written in matrix-vector form as follows;

$$var\left(\hat{Z}\right) = \frac{(1-f)N^2}{n(n-1)} \mathbf{z}^T \left(\mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}^T\right)\mathbf{z},$$

where $\mathbf{z} = \left(z_1, z_2, ..., z_n\right)^T$, $\mathbf{I}$ is unit matrix of order n and $\mathbf{e}$ is a vector with each component being equal to 1. Thus we have

$$var\left(\hat{Z}\right) = C_1 \left(\mathbf{y}^T - \mathbf{b}^T\mathbf{x}^T\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}^T\right)\left(\mathbf{y} - \mathbf{x}\mathbf{b}\right) \qquad /10/$$

where $C_1 = \dfrac{(1-f)N^2}{n(n-1)}$. Now $\mathbf{b}^{\text{o}}$+$\mathbf{b}'$ should be substituted for **b.** We have to take into account that, owing to simple random sampling, the matrix $\mathbf{\Omega}$ in the expressions of $\mathbf{b}^{\text{o}}$ and $\mathbf{b}'$ is now *N*/*n* times the unit matrix. However, the factor *N*/*n* will not occur in the formulae, since it always appears simultaneously in the numerator and in the denominator. Consequently, the factor $\mathbf{y} - \mathbf{x}\mathbf{b}$ becomes

$$\mathbf{y} - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y} - C_{\text{o}}\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\left(X - \hat{X}\right) =$$
$$= \mathbf{y} - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y} - C_{\text{o}}\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\left(\mathbf{w}^{\text{o}} - \mathbf{d}\right),$$

or denoting the matrix $\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T$ by **P**,

[2] From the viewpoint of regression this means that there is an intercept.

$$\mathbf{y}-\mathbf{x}\mathbf{b}=\mathbf{y}-\mathbf{P}\mathbf{y}-C_{\mathrm{o}}\mathbf{P}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)=\left(\mathbf{I}-\mathbf{P}\right)\mathbf{y}-C_{\mathrm{o}}\mathbf{P}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right), \qquad /11/$$

where $C_{\mathrm{o}}=\dfrac{\hat{Y}^{\mathrm{cal}}-\hat{Y}^{\mathrm{reg}}}{\left(X-\hat{X}\right)^{T}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\left(X-\hat{X}\right)}$;

note that $\mathbf{\Omega}$ has disappeared from here, too. The matrix **P** is a symmetric projection and, because of the assumption on the auxiliary variable having the value 1 for any unit, the vector **e** is an eigenvector of $\mathbf{P}$: $\mathbf{Pe}=\mathbf{e}$. Substituting the right-hand side of /11/ for **y – xb** in /10/ implies

$$var\left(\hat{Z}\right)=C_{1}\left(\mathbf{y}^{T}\left(\mathbf{I}-\mathbf{P}\right)-C_{\mathrm{o}}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)^{T}\mathbf{P}\right)\left(\mathbf{I}-\frac{1}{n}\mathbf{e}\mathbf{e}^{T}\right)\left(\left(\mathbf{I}-\mathbf{P}\right)\mathbf{y}-C_{\mathrm{o}}\mathbf{P}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)\right)=$$
$$=C_{1}\mathbf{y}^{T}\left(\mathbf{I}-\mathbf{P}\right)\mathbf{y}+C_{1}C_{\mathrm{o}}^{2}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)^{T}\left(\mathbf{P}-\frac{1}{n}\mathbf{e}\mathbf{e}^{T}\right)\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right).$$

Substituting here $\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}$ for **P** and making use of the expression for $\mathbf{b}^{\mathrm{o}}$ and the relation $\mathbf{x}^{T}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)=X-\hat{X}$ one obtains

$$var\left(\hat{Z}\right)C_{1}\left(\mathbf{y}-\mathbf{x}\mathbf{b}^{\mathrm{o}}\right)^{T}\left(\mathbf{I}-\frac{1}{n}\mathbf{e}\mathbf{e}^{T}\right)\left(\mathbf{y}-\mathbf{x}\mathbf{b}^{\mathrm{o}}\right)+$$
$$+C_{1}C_{\mathrm{o}}^{2}\left(\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\left(X-\hat{X}\right)\right)^{T}\left(\mathbf{I}-\frac{1}{n}\mathbf{e}\mathbf{e}^{T}\right)\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\left(X-\hat{X}\right).$$

Using again the argument that an additive constant of the form $\sum_{k}b_{k}X_{k}$ has no impact on the variance, it is easy to see that the right-hand side of the last equality equals $var\left(\hat{Y}^{\mathrm{reg}}\right)+var\left(\hat{X}\mathbf{b}'\right)$ which verifies /8/. Inequality /9/ follows by omitting the matrix $\mathbf{I}-\mathbf{e}\mathbf{e}^{T}/n$ from the second term and making use of the fact that its norm equals /1/. The proof is thereby complete.

3. A numerical example

We have considered a universe consisting of $N=2899$ households. In those households, there were $X_{1}=1076$ individuals aged 15-24 years, $X_{2}=4239$ individuals aged 25-54 years, $X_{3}=1382$ individuals aged 55-74 years, $X_{4}=3193$ males aged 15-74 years, $X_5 = 3504$ females aged 15-74 years, and, finally $Y = 3656$ individuals aged 15-74 who participated in the labour market.

From this universe simple random samples consisting of 25 units were selected, thus the design weight was 116.96 for each unit in the samples. Using X_1, X_2, X_3, X_4, X_5 and $X_6 = N$ as controls,[3] two calibration estimates of Y were computed for each sample. One of them was $\hat{Y}^{\text{reg}}$, the baseline estimate, the other was $\hat{Y}^{\text{cal}}$ obtained with raking, obeying also the individual bounds $40 \le w_j \le 600$ for the final weights.

The following table shows $\hat{Y}^{\text{reg}}$, $\hat{Y}^{\text{cal}}$, $\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}} = \left(X - \hat{X}\right)^T \mathbf{b}'$ and the corresponding standard errors based on /8/ and /9/ for the first six samples.

Estimates and standard errors obtained with two calibration estimators for samples from an artificial population

Number of Sample	$\hat{Y}^{\text{reg}}$		$\hat{Y}^{\text{cal}}$		$\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}$	
	Estimate	S. E.	Estimate	S. E.	Estimate	S. E.
1	2878	308.9	2933	310.5	55	30.9
2	4815	331.4	4797	331.5	–18	7.2
3	3306	393.4	3346	394.1	40	24.1
4	3773	343.1	3739	344.0	–34	19.6
5	2884	253.7	2959	254.7	75	22.8
6	3494	409.4	3575	412.6	81	50.1

It might be surprising that the asymptotic equivalence of calibration estimators is manifest even at such moderate sizes of sample and population as $n = 25$, $N = 2899$.

References

COCHRAN, W. G. [1977]: *Sampling techniques*. John Wiley & Sons. New York.

DEVILLE, J-C. – SÄRNDAL, C-E. [1992]: Calibration estimators in survey sampling. *Journal of the American Statistical Association.* Vol. 37. p. 376–382.

SÄRNDAL, C-E. – SWENSSON, B. – WRETMAN, J. [1992]: *Model assisted survey sampling*. Springer Verlag. New York.

[3] Note that $X_1+X_2+X_3=X_4+X_5$.

Exact inference on poverty predictors based on logistic regression approach

Ottó Hajdu,
Associate professor
Budapest University of Technology
E-mail: hajdu@finance.bme.hu

The paper deals with the exact inference on poverty indicators included in a predictive logistic regression model as predictor variables. Based on a multiple stratification applied in a household survey, small size or unbalanced subgroups are likely to occur in practice with regard to the number of poor and hence the standard unconditional maximum likelihood estimation of a regression parameter may fail to exist. Focus is brought on exact inference which is still possible to make even at this case. The paper gives a brief overview of problems of exact *p*-value and confidence interval calculation in small samples for the case when the unconditional maximum likelihood estimate does not exist or the large sample asymptotic properties are violated. Besides, some empirical examples are presented based on a survey of Hungarian households.

KEYWORDS:
Logistic and loglinear models.
Estimation. Poverty and social deprivation.

Logistic regression is a straightforward method to evaluate the risk of being poor based on an appropriate poverty line.[1] This paper focuses mainly on the problem of the relevancy of poverty indicators when the so-called *p*-value criterion is applied to select a set of predictor variables. Since several types of homogenous households can be defined by their socio-economic and demographic characteristics, small size or unbalanced subgroups are likely to be formed for measuring poverty in a stratified survey. Calculating *p*-values for inference accurately as possible is clearly a key stage of the model building process but the standard unconditional maximum likelihood approach exhibits appealing asymptotic properties only in the large sample case. Otherwise, considering small or unbalanced samples the exact approach of inference gives the correct *p*-values and confidence intervals. The main purpose of this paper is to give a brief guide to apply exact methods and to interpret their results correctly.

1. The predictive model

Consider a set of independent binary random variables, $\mathbf{Y} = \left(Y_1, Y_2, ..., Y_n\right)$ where $i = 1,\ 2,\ ....,\ n$ stands for an individual household. The response variable Y_i takes the value of 1 in the case of households falling below the poverty line and 0 otherwise. Corresponding to each response variable, Y_i, there is a $\left(p \times 1\right)$ covariate vector $\mathbf{x}_i = \left(x_{i1}, x_{i2}, ..., x_{ip}\right)'$ of predictor variables. Let π_x denote the conditional probability $Pr\left(Y = 1 \middle| \mathbf{x}\right)$ i.e. the probability that $Y = 1$ conditioned on the levels of the covariates. Using the so-called odds-measure defined as $\pi_x / \left(1 - \pi_x\right)$ the conditional probability of the event „1" can be written as follows:

$$\pi_x = \frac{\pi_x / \left(1 - \pi_x\right)}{1 + \pi_x / \left(1 - \pi_x\right)} = \frac{\text{odds}_x}{1 + \text{odds}_x}\ . \qquad /1/$$

When π_x exceeds a critical value of C the prediction is $\hat{Y} = 1$ otherwise the predicted value is 0. In the case of logistic regression models the dependency of π_x on the predictor variables is expressed through the relationship

[1] For a recent work in the field see for instance *Havasi* [2005].

$$\log\left(\text{odds}_x\right)=\boldsymbol{\beta}\mathbf{x}\,, \qquad /2/$$

where $\boldsymbol{\beta}$ is a $(1\times p)$ vector of unknown parameters. An *unstratified* model is defined when there is only one single constant term included among the parameters. This distinguishes it from the *stratified* model when data come from several strata and a different stratum specific constant term is introduced to each stratum provided common slope parameters.

Based on a random sample of size n included in the $(n\times 1)$ vector $\mathbf{y}=\left(y_1,y_2,...,y_n\right)'$ the unconditional likelihood of the sample to be maximized with respect to $\boldsymbol{\beta}$ is

$$L(\boldsymbol{\beta})=\prod_{i=1}^{n}\frac{\left(\text{odds}_i\right)^{y_i}}{1+\text{odds}_i}=\frac{e^{\sum_{i=1}^{n}y_i\sum_{j=1}^{p}\beta_j x_{ij}}}{\prod_{i=1}^{n}\left(1+e^{\sum_{j=1}^{p}\beta_j x_{ij}}\right)}=\frac{e^{\sum_{j=1}^{p}\beta_j t_j}}{\prod_{i=1}^{n}\left(1+e^{\sum_{j=1}^{p}\beta_j x_{ij}}\right)}\to\max\,, \qquad /3/$$

where the

$$t_j=\sum_{i=1}^{n}y_i x_{ij}\quad\left(j=1,2,...,p\right) \qquad /4/$$

sample statistic is termed as *sufficient statistic* for the parameter β_j in the literature of the exact logistic regression.[2] Recalling now that $y_i=\{0,1\}$, t_j is simply the sum of the predictor variable x_j for those who fall below the poverty line. Hence, if a constant term is included in the model, its sufficient statistic means the number of poor households. Obviously, t_j is the sample outcome of the random variable T_j.

In order to make inference on parameters of our interest consisting of the vector $\boldsymbol{\beta}_1$, partition the parameter vector $\boldsymbol{\beta}$ into two components $\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\right)$. Suppose we are primarily interested in inference about $\boldsymbol{\beta}_1$, and regard $\boldsymbol{\beta}_2$ as „nuisance" parameter(s). The partitioned vector of the sufficient statistics corresponding to $\boldsymbol{\beta}=\left(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\right)$ is $\mathbf{t}=\left(\mathbf{t}_1,\mathbf{t}_2\right)$ or in matrix notation

$$\mathbf{X}_1'\mathbf{y}=\mathbf{t}_1,\quad \mathbf{X}_2'\mathbf{y}=\mathbf{t}_2\text{, where} \qquad /5/$$

$$\mathbf{X}=\left[\mathbf{X}_1,\mathbf{X}_2\right] \qquad /6/$$

[2] For the definition of a sufficient statistic see *Garthwaite–Jolliffe–Jones* [1995].

is the partitioned form of the $(n \times p)$ covariate matrix **X**.

The unconditional likelihood can be written in an equivalent form of

$$f_{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2}\left(\mathbf{t}_1,\mathbf{t}_2\right)=\frac{c\left(\mathbf{t}_1,\mathbf{t}_2\right)e^{\boldsymbol{\beta}_1\mathbf{t}_1+\boldsymbol{\beta}_2\mathbf{t}_2}}{\sum_{\mathbf{u}}c\left(\mathbf{u}_1,\mathbf{u}_2\right)e^{\boldsymbol{\beta}_1\mathbf{u}_1+\boldsymbol{\beta}_2\mathbf{u}_2}}, \qquad /7/$$

where $c(\mathbf{t})$ is the number of ways of selecting the binary sequence **y** so as to satisfy the $\mathbf{X}'\mathbf{y}=\mathbf{t}$ sample condition. The summation in the denominator is over all $\mathbf{u}=\mathbf{X}'\mathbf{Y}$ generated by any $(n\times 1)$ binary sequence **Y**. Maximizing /7/ by choice of $\boldsymbol{\beta}$ is equivalent to maximizing /3/ by choice of $\boldsymbol{\beta}$ (see *LogXact 7 User Manual* [2005] p. 503).

2. Hypothesis testing

Three methods of inference are available for testing: *1.* unconditional likelihood inference, *2.* conditional likelihood inference, and *3.* conditional exact inference. Unconditional likelihood inference is based on estimating the entire parameter vector $\boldsymbol{\beta}$ by maximizing the unconditional likelihood function. This approach assumes the asymptotic normality of the maximum likelihood estimates to make inferences about $\boldsymbol{\beta}_1$. Conditional likelihood inference, in contrast, is based on maximizing the likelihood function derived by conditioning on the sufficient statistics for $\boldsymbol{\beta}_2$ in /7/. By definition this likelihood function is *free* of the nuisance parameters $\boldsymbol{\beta}_2$. Finally, conditional exact inference is based on deriving the exact permutational distribution $c(\mathbf{u}_1,\mathbf{t}_2)$ of the sufficient statistics for $\boldsymbol{\beta}_1$ conditional on the sufficient statistics for $\boldsymbol{\beta}_2$.

It is apparent, that conditioning on $\mathbf{t}_2$ plays a key role in inference because thereby parameter $\boldsymbol{\beta}_2$ can entirely be eliminated from the likelihood:

$$f_{\boldsymbol{\beta}_1}\left(\mathbf{t}_1\,|\,\mathbf{t}_2\right)=L\left(\boldsymbol{\beta}_1\,|\,\mathbf{t}_2\right)=\frac{c\left(\mathbf{t}_1,\mathbf{t}_2\right)e^{\boldsymbol{\beta}_1\mathbf{t}_1}}{\sum_{\mathbf{u}_1}c\left(\mathbf{u}_1,\mathbf{t}_2\right)e^{\boldsymbol{\beta}_1\mathbf{u}_1}}\rightarrow \max. \qquad /8/$$

Maximizing /8/ with respect to $\boldsymbol{\beta}_1$ yields conditional maximum likelihood estimates (CMLE). The fundamental difference between unconditional maximum likelihood estimates (MLE) and conditional CMLE inference is that MLE needs to estimate $\boldsymbol{\beta}_2$ even if it were only a nuisance parameter, interest being focused mainly on $\boldsymbol{\beta}_1$ while based on the CMLE approach $\boldsymbol{\beta}_2$ is conditioned out from the estimation.

To test the null hypothesis

$$H_0 : \boldsymbol{\beta}_1 = \mathbf{0}, \qquad /9/$$

based on the MLE approach three basic methods are available: *1.* the *scores* statistic (also known as Lagrange-Multiplier test statistic), *2.* the *likelihood ratio* statistic and *3.* the *Wald* statistic. All of these statistics are asymptotically chi-squared distributed on *d* degrees of freedom under H_0 where *d* is the number of restrictions imposed. It must be emphasized that the scores statistic does not depend on the full model MLE. It is derived based on the MLE of the restricted model only. This means that the scores statistic may exist even when the MLE of the full model does not exist.

When CMLE is applied the corresponding conditional versions of these test statistics are in hand to be applied. Further, when data are from several strata and number of the stratum specific "nuisance constants" is large relative to the number of observations the MLE estimates for the slope parameters may be inconsistent. In this case the CMLE estimates for the slope parameters $\boldsymbol{\beta}_1$ (eliminating the stratum specific constant terms included in $\boldsymbol{\beta}_2$ from the computation) are consistent. Since we have eliminated the stratum specific constant terms from the likelihood function it is no longer possible to test hypothesis about these parameters.

For small or highly imbalanced data sets the asymptotic chi-squared distributions for the scores, likelihood ratio and Wald tests might not hold. This is the situation when it is appropriate to generate the true permutation distribution of $\mathbf{T}_1$ given $\mathbf{t}_2$. Armed with this permutation distribution one can perform exact hypothesis tests and generate exact confidence intervals for the parameters of interest. Based on /8/ under $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ the exact conditional probability that $\mathbf{T}_1 = \mathbf{t}_1$ given $\mathbf{t}_2$ that is the null-distribution of $f_{\boldsymbol{\beta}_1}(\mathbf{t}_1 \mid \mathbf{t}_2)$ reduces to

$$f_0(\mathbf{t}_1 \mid \mathbf{t}_2) = \frac{c(\mathbf{t}_1, \mathbf{t}_2)}{\sum_{\mathbf{u}_1} c(\mathbf{u}_1, \mathbf{t}_2)}. \qquad /10/$$

An exact *p*-value for testing H_0 is then defined as follows:

$$p = \sum_{\mathbf{u}_1 \in R} f_0(\mathbf{u}_1 \mid \mathbf{t}_2), \qquad /11/$$

where *R* is the region of the conditional sample space of $\mathbf{T}_1$ given $\mathbf{t}_2$ in which the values $\mathbf{T}_1 = \mathbf{u}_1$ are all considered to be more extreme under H_0 than the observed value $\mathbf{T}_1 = \mathbf{t}_1$.

An adjusted version of p, the so-called *mid-p-value* can be applied to correct the discrete test without compromising on its significance level:

$$p_{mid} = p - 0.5 f_0\left(\mathbf{t}_1 \mid \mathbf{t}_2\right). \qquad /12/$$

The choice of R depends on the type of exact test selected. We provide three methods for selecting an exact test: *1.* exact conditional scores test (based on either asymptotic or exact variance); *2.* exact conditional probability test; and *3.* exact likelihood ratio test.

For the *exact conditional scores test*, R, the extreme region of the sample space over which the p-value is calculated is defined to be all values of the test statistic having a value greater than or equal to the *observed* test statistic:

$$q = \left(\mathbf{t}_1 - E\left(\mathbf{T}_1\right)\right)' Cov_{\mathbf{T}_1}^{-1} \left(\mathbf{t}_1 - E\left(\mathbf{T}_1\right)\right). \qquad /13/$$

For the *exact conditional probability test*, R, the extreme region of the sample space over which the p-value is calculated is defined to be all values of the test statistic having a probability smaller than or equal to the probability of the observed test statistic. Finally, for the *exact likelihood ratio test*, R, the extreme region of the sample space over which the p-value is calculated is defined to be all values of the test statistic having a likelihood ratio value greater than or equal to that of the observed data.

Although the exact tests are guaranteed to protect from type-1 error at any specified level, the conditional scores test may be the preferred one. The main reason is that characterizing the rejection region R in terms of larger conditional scores rather than smaller conditional probabilities is intuitively appealing. Considering the exact conditional probability approach in the univariate case (when $\mathbf{t}_1$ is a scalar t_1) it may happen that the conditional probability distribution $f_0\left(t_1 \middle| \mathbf{t}_2\right)$ has multiple modes so that the rejection region R is not a contiguous interval. This problem does not occur when exact conditional scores test is applied. Moreover, the conditional maximum likelihood estimate fails to exist in certain situations hence it is impossible to carry out a likelihood ratio test. Notice, that for the conditional scores test, parameter estimates are not required.

3. Exact parameter estimation

Estimation of the parameters by maximizing /8/ with respect to $\boldsymbol{\beta}_1$ yields conditional CMLE point estimates. However, since the marginal interpretation of the slope parameters is meaningful only in a partial sense it is reasonable to perform exact inference on each individual scalar parameter β_j separately, step-by-step. This is carried out by taking the special partitioning $\boldsymbol{\beta}_1 = \beta_1$ which is done successively for each parameter of interest:

$$f_{\beta_1}\left(t_1 \mid \mathbf{t}_2\right) = \frac{c\left(t_1, \mathbf{t}_2\right) e^{\beta_1 t_1}}{\sum_{u_1 = t_{1\min}}^{t_{1\max}} c\left(u_1, \mathbf{t}_2\right) e^{\beta_1 u_1}} \to \max \,, \qquad /14/$$

where $t_{1\min}$ and $t_{1\max}$ are the minimum and maximum values respectively in the range of the random variable T_1 conditional on $\mathbf{t}_2$. Notice that /14/ depends only on β_1.

If the observed value of the sufficient statistic t_1 is at one extreme of its range (when either $t_1 = t_{1\min}$ or $t_1 = t_{1\max}$) it is no longer possible to maximize /14/ with respect to β_1. This is the case when the sample is *separated* with respect to the predictor x_1 (see *Christmann–Rousseeuw* [2001]). In this case the likelihood function increases strictly monotonically as $|\beta_1|$ goes towards ∞. When the CMLE fails to exist the so-called "median unbiased estimate" (MUE) is available for exact conditional estimation by solving the following equation:

$$f_{\hat{\beta}_1}\left(t_1 \mid \mathbf{t}_2\right) = 0.5 \,. \qquad /15/$$

Considering the two-sided α level confidence interval (CI) the upper and lower bounds β_+ and β_- are defined respectively based on the left and right tails of the distribution of $T_1 \middle| \mathbf{t}_2$ at the observed value $T_1 = t_1$ in the following manner:

$$F_{\beta+}\left(t_1\right) = \sum_{u_1 = t_{1\min}}^{t_1} f_{\beta_+}\left(u_1 \mid \mathbf{t}_2\right) = \alpha / 2 \,, \qquad /16/$$

$$G_{\beta_-}\left(t_1\right) = \sum_{u_1 = t_1}^{t_{1\max}} f_{\beta_-}\left(u_1 \mid \mathbf{t}_2\right) = \alpha / 2 \,. \qquad /17/$$

Solving /16/ and /17/ gives interval as desired. A one-sided CI can also be obtained spending the entire α error using only the left tail distribution $F_{\beta_+}\left(t_1\right) = \alpha$, when $F_0\left(t_1\right) \le G_0\left(t_1\right)$ and the right tail distribution $G_{\beta_-}\left(t_1\right) = \alpha$, when $F_0\left(t_1\right) > G_0\left(t_1\right)$. The resulted intervals are $\left(-\infty, \beta_+\right)$ and $\left(\beta_-, \infty\right)$ respectively.

Also, regardless the type of CI requested only an open interval is available when either $t_1 = t_{1\min}$ or $t_1 = t_{1\max}$ because the cumulative probability on the entire range of the random variable $T_1 \left| \mathbf{t}_2 \right.$ always equals 1 and hence it is independent at the value of β_1. Then the CI bounds are constructed as follows:

$$F_{\beta+}\left(t_1\right) = \sum_{u_1=t_{1\min}}^{t_{1\min}} f_{\beta_+}\left(u_1 \mid \mathbf{t}_2\right) = \alpha / 2, \quad \beta_{\_} = -\infty \,, \qquad /18/$$

$$G_{\beta_{\_}}\left(t_1\right) = \sum_{u_1=t_{1\max}}^{t_{1\max}} f_{\beta_{\_}}\left(u_1 \mid \mathbf{t}_2\right) = \alpha / 2, \quad \beta_{+} = +\infty \,. \qquad /19/$$

Finally, whenever a CI reported at α level it is necessary to compute a *p*-value that preserves the consistency between the conclusions derived from the CI and from the *p*-value. Based on this particular method it is ensured that the exact *p*-value is less than α if and only if the exact $(1-\alpha)$ CI excludes the corresponding model parameter. According to this alternative definition the exact two-sided *p*-value is double the one-sided *p*-value: $p_2 = 2p_1$, where the one-sided *p*-value is the minimum of the left and right tail probabilities:

$$p_1 = \min\left\{F_0\left(t_1\right), G_0\left(t_1\right)\right\}. \qquad /20/$$

4. Some empirical examples

Let us consider the subgroup of *households* in *Budapest* with *6 persons or more*, in the year of 2003. The source of data is the HCSO, Household Expenditure Survey 2003. Those households with per capita income below the poverty line (0.6 median income) are considered to be poor.[3] Their status is recorded in the binary poverty variable by the value of Poverty = 1 which plays the role of the response variable. The examples of our interest are as follows.

4.1. The unconditional, asymptotic maximum likelihood estimation fails to exist

The estimation results are shown in Table 1 and the frequency distributions of the predictor variables considered are included in Table 2. The computations are carried out by the program LogXact 7 (www.cytel.com).

[3] The per capita median income level in the year of 2003 is HUF754.000 and per *capita* here stands for per *consumption unit* as defined by the OECD scheme using 1, .7 and .5 units representing first and further adults and children respectively.

Table 1

Parameter estimates when MLE does not exist

Model	Type	Beta	SE(Beta)	Type	95 percent CI lower	95 percent CI upper	2*1-sided=p_2
Model term 1							
Constant	MLE	?	?	asymptotic	?	?	?
Gender	MLE	?	?	asymptotic	?	?	?
	MUE	4.481	NA	exact	2.804	+INF	1.094e-024
Model term 2							
Constant	MLE	?	?	asymptotic	?	?	?
Permanently sick	MLE	?	?	asymptotic	?	?	?
	MUE	–5.29	NA	exact	–INF	–3.614	5.809e-052
Model term 3							
Constant	MLE	–8.522	0.3566	asymptotic	–9.221	–7.823	2.493e-051
Education score	MLE	0.5927	0.03053	asymptotic	0.5328	0.6525	2.327e-043
	CMLE	0.5926	0.03053	exact	0.534	0.6547	3.763e-202
Model term 4							
Constant	MLE	?	?	asymptotic	?	?	?
Education level	MLE	?	?	asymptotic	?	?	?
	MUE	7.092	NA	exact	5.418	+INF	6.977e-257

Note. NA for not applicable, ? means does not exist, INF is infinite, e is exponent.

Table 2

Frequency distributions of the predictor variables

Denomination	Poverty = 0	Poverty = 1	Total
Gender			
0: Female	601	0	601
1: Male	6294	642	6936
Permanently sick persons			
0: not present	5678	642	6320
1: present	1217	0	1217
Education score			
3	1009	0	1009
5	1573	0	1573
7	370	0	370
8	1383	0	1383
11	545	355	900
12	1809	126	1935
13	206	161	367

(Continued on the next page.)

(Continuation.)

Denomination	Poverty = 0	Poverty = 1	Total
Education level			
1	1009	0	1009
2	3326	0	3326
3	2560	642	3202
Unemployed persons			
0	6459	516	6975
1	436	0	436
2	0	126	126
Economic activity			
111 type	681	0	681
112 type	986	161	1147
113 type	410	0	410
114 type	656	0	656
115 type	1520	0	1520
117 type	996	0	996
121 type	609	481	1090
122 type	601	0	601
232 type	436	0	436
Total	*6895*	*642*	*7537*

Firstly, taking the gender of the head of household as a single predictor variable (Model term 1) "Female" is a perfect predictor hence the MLE does not exist (this is indicated by a ? mark) while the MUE point estimate and a one-sided confidence interval are available. The upper bound of the CI is +INF because the zero frequency occurs at the lower extreme value on the range of gender i.e. at female when Gender = 0 in Table 2.

In contrast, let us consider an another single binary predictor variable namely whether a permanently sick person is present in the household "1" or not "0" (Model term 2). The conclusions are similar to those made earlier with the only exception that the lower bound of the CI is -INF since the zero frequency in Table 2 occurs at the upper extreme value of 1 in the range i.e. when a permanently sick person is present in the household.

Merging categories can also influence the existence of the MLE. Considering the education score of the head of household as a single predictor variable (see Model term 3)[4] it is apparent that both MLE and CMLE exist *despite* the fact that zero

[4] The entire ordinal range of the level of education: [1,2,...,13] with 13 indicating a PhD degree.

frequencies appear *only* at the lower *tail* of the distribution published in Table 2. However, merging the scores into only three levels as shown in Table 2 the MLE does not exist any longer as it can be seen in Table 1 under Model Term 4.

4.2. Differences between exact and asymptotic *p*-values

Despite the relatively large sample size – given that the sample is unbalanced (i.e. the poor/non-poor ratio is 642/6895) – one can expect that the asymptotic and exact *p*-values differ significantly. Let us take the number of unemployed persons in the household as a single predictor variable (Model term 5). Results of Table 3 show that this is not the case. The number of unemployed persons is significant at any level and the point and interval estimations are quite similar in magnitude.

Selecting now the number of the dependent persons in the household as predictor variable (Model term 6), Table 3 shows that the exact *p*-value can considerably differ from the unconditional one. Apparently, the number of dependent persons is not significant at usual error levels but if the extreme 40 percent level were applied as a critical cut-off-value the different methods would yield different conclusions. It might be the case of several predictors in several strata not investigated in this paper.

Table 3

Parameter estimates when the MLE does exist

Model	Type	Beta	SE(Beta)	Type	95 percent CI lower	95 percent CI upper	2*1-sided=p_2
Model term 5							
Constant	MLE	–2.6420	0.04741	asymptotic	–2.7350	–2.5490	3.92e-085
Unemployed	MLE	1.4910	0.07773	asymptotic	1.3390	1.6440	6.443e-043
	CMLE	1.4910	0.07772	exact	1.3360	1.6470	3.333e-073
Model term 6							
Constant	MLE	–1.4590	0.41440	asymptotic	–2.2710	–0.6464	0.000432
Dependents	MLE	–0.0877	0.10120	asymptotic	–0.2860	0.1106	0.386
	CMLE	–0.0876	0.10110	exact	–0.2912	0.1158	0.4143

4.3. Strata specific constant terms are conditioned out

Let us discuss again the impact of the number of unemployed persons in the household as the single predictor but controlling on the type of households regarding the economic activity of the households as a stratum variable (Model term 7 in Table 4). Several strata can be defined based on the classification whether an active person is present in the household or not and besides regarding the activity of the head of the household as well. The codes of these types are presented in Table 2 but their exact meaning is not relevant from our methodological point of view. The fact must be highlighted that after this stratification the MLE is not available but exact MUE exists and exact p-value in Table 4 tells us that variable “Unemployed” is still significant at any level. Notice, that both the intercept term and the nuisance stratum specific constant terms are *cancelled out* from the estimation.

Table 4

Stratified estimates by the type of household’s economic activity

Model	Type	Beta	SE(Beta)	Type	95 percent CI lower	95 percent CI upper	2*1-sided=p_2
Model term 7							
Unemployed	MLE	?	?	asymptotic	?	?	?
	MUE	2.8680	NA	exact	2.0230	+INF	9.471e-050
Model term 8							
Education score	MLE	–0.2139	0.09588	asymptotic	–0.4018	–0.02596	0.0257
	CMLE	–0.2139	0.09588	exact	–0.4065	–0.02154	0.02889

Finally, Table 4 reconsiders the model with the education score of the household’s head as the predictor variable but using the stratified sample described previously. Although both the MLE and the CMLE exist the predictor “score of education” at 2 percent level is no longer significant, moreover, the *signs* of both parameters have changed. Notice again that both the intercept term and the nuisance stratum specific constant terms are cancelled out from the estimation.

4.4. Conflicting test results

So far, in our study only the 2*1-sided type p_2-value was applied to ensure the consistency with the 95 percent CI published. However, the exact p-values may vary depending on the special test statistic type applied, such as *scores*, *likelihood ratio* and *Wald*, especially when the sample size is extremely small. This problem is illustrated as follows.

Table 5 gives the exact test results for two predictors. Firstly, for the age of the head of the household (Age) then subsequently for the response (yes or no) whether the household has suffered poverty *ever before* (Suffered). The sample has been restricted to the *types* of the households in *Budapest* those with *more than 6 persons* exhibiting gender of the household's head. Table 5 shows that in the case of Age only the score test exists among the asymptotic tests but its *p*-values yield different decisions at 5 percent error level depending on the type of test considered. Moreover, exact tests give the same $p = 0.07143$ value in this case (this is not necessary in general) but the *p-mid* value for the exact likelihood ratio test accepts the null-hypothesis at 5 percent level while the other exact tests reject it.

Finally, considering the question of "Suffering poverty ever before" based on the *p*-values at 5 percent error level the conclusion of the exact probability test differs from that of the other exact types and both the *p*-values and the *p-mid* values vary substantially.

Table 5

Exact test results

Type of test	Statistics	DF	*p*-value	*p*-mid
	H_0 :Beta_Age = 0			
Score	4.317	1	0.03774	NA
Likelihood ratio	?	?	?	?
Wald	?	?	?	?
Exact score_asy	4.317	NA	0.07143	0.05357
Exact score	3.777	NA	0.07143	0.05357
Exact probability	0.03571	NA	0.07143	0.05357
Exact likelihood ratio	8.997	NA	0.07143	0.03571
	H_0 : Beta_Suffered = 0			
Score	6.107	1	0.01347	NA
Likelihood ratio	?	?	?	?
Wald	?	?	?	?
Exact score	5.343	NA	0.03571	0.01786
Exact probability	0.03571	NA	0.07143	0.05357
Exact likelihood ratio	8.997	NA	0.03571	0

Finally, we can conclude that when a stratification of the data is required in many dimensions in order to control some factors, other covariates of the response variable are likely to become perfect predictors. In this case the exact approach of inference is appropriate providing reliable results if they exist.

References

AGRESTI, A. [2002]: *Categorical data analysis*. Wiley. New York.

ALBERT, A. – ANDERSON, J. A.[1984]: On the existence of maximum likelihood estimates in logistic models. *Biometrica.* 71. p. 1–10.

BULL, S B. – MAK, C. – GREENWOOD, C. M. T. [2002]: A modified score function estimator for multinomial logistic regression in small samples. *Computational Statistics and Data Analysis*. 39. p. 57–74.

CHRISTMANN, A. – ROUSSEEUW, P. J. [2001]: Measuring overlap in logistic regression. *Computational Statistics and Data Analysis*. 37. p. 65–75.

COX, D. R. – SNELL, E. J. [1989]: *Analysis of binary data*. Chapman and Hall. London.

CRAMER, J. S. [1999]: Predictive performance of the binary logit model in unbalanced samples. *The Statistician*. 48. p. 85–94.

GARTHWAITE, P. H. – JOLLIFFE, I. T. – JONES, B. [1995]: *Statistical inference*. Prentice Hall. New York**.**

HAVASI, É. [2005]: *A jövedelem mint az anyagi jólét és a szegénység mérőszáma*. KSH. Budapest.

HIRJI, K. F. [1992]: Exact distributions for polytomous data. *JASA*. 87. p. 487–492.

HIRJI, K. F. – MEHTA, C. R. – PATEL, N. R. [1987]: Computing distributions for exact logistic regression. *JASA*. 82. p. 1110–1117.

HIRJI, K. F. – TSIATIS, A. A. – MEHTA, C. R. [1989]: Median unbiased estimation for binary data. *The American Statistician.* 43. p. 7–11.

HUNYADI, L. [2001]: *Statisztikai következtetéselmélet közgazdászoknak* . KSH. Budapest.

KING, G. – ZENG, L. [2001]: Logistic regression in rare events data. *Political Analysis*. 9. p. 137–163.

KING, E. N. – RYAN, T. P. [2002]: A preliminary investigation of maximum likelihood logistic regression versus exact logistic regression. *The American Statistician*. Vol. 56. No. 3. p. 163–170.

LogXact 7 User Manual, Cytel Statistical Software & Services [2005]. Cytel. Cambridge.

MEHTA, C.R. – PATEL, N. R. [1995]: Exact logistic regression: Theory and examples. *Statistics in Medicine*. Vol. 14**.** p. 2143–2160.

MEHTA, C. R. – PATEL, N. R. – SENCHAUDHURI, P. [2000]: Efficient Monte Carlo methods for conditional logistic regression. *Journal of the American Statistical Association*. Vol. 95. No. 449. p. 99–108.

SANTNER, T. J.– DUFFY, D. E. [1989]: The statistical analysis of discrete data. Springer-Verlag. New York.

TRICHLER, D. [1984]: An algorithm for exact logistic regression. *Journal of the American Statistical Association*. 79. p. 709–711.

Digital analysis: Theory and applications in auditing

Tamás Lolbert,
auditor, State Audit Office of Hungary
E-mail: lolbertt@asz.hu

Newcomb in 1881 and *Benford* in 1938 independently introduced a phenomenon, that in randomly collected numbers certain digits are more often leading digits than others. They stated that the frequency of starting digits follows the logarithmic distribution, so that for any starting digit d=1 … 9, Prob(first significant digit=d)=$\log_{10}(1+1/d)$. This empirical law was recognized by many mathematicians so that several possible explanations have been derived. Meanwhile the phenomenon has not only theoretical aspects, since it can be applied in detecting fraud, (deliberate) misstatements or fabrication of data, or in several other fields, but still most notably in auditing of financial statements. It has other applications as well, ranging even to the design of future computer processors. This study gives an overview on Benford's law and its history, lists the main mathematical results and last but not least introduces the most important application, digital analysis.

KEYWORDS:
Theory of probability.
Financial applications, financial and stock market.

For boring novels in a library it is not unusual, that the first 10-20 pages are much dirtier than the last ones. One would however not presume that it would be true for dictionaries or encyclopaedias. Nor for collections of tables of real science. But sometimes reality contradicts intuition. In 1881, 125 years ago, an astronomer-mathematician named *Simon Newcomb* wrote a two page article where he stated: "That the ten digits do not occur with equal frequency must be evident to anyone making use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones. The first significant digit is oftener 1 than any other digit, and the frequency diminishes up to 9" (*Newcomb* [1881] p. 39–40). Without further explanations or examples he concluded that the law of this frequency satisfies the following equation (see values in Table 1.):

$$\text{Prob (first significant digit} = d) = \log_{10}(1 + 1/d), \text{ where } d = 1, 2,\ldots 9.$$

Table 1

Frequencies of decimal leading digits

First digit	Frequency (percent)
1	30.10
2	17.61
3	12.49
4	9.69
5	7.92
6	6.69
7	5.80
8	5.12
9	4.58

Although this observation was old that time, it became unremembered and was rediscovered some 60 years later by the physicist *Frank Benford* who had a paper (*Benford* [1938]) on "The law of anomalous numbers". He begins with the same observation on logarithmic tables as Newcomb, and continues with the analysis of different data sets, e.g. the area of 335 rivers, populations of 3259 locations, mol weights of elements, powers of natural numbers, street addresses of famous people from a magazine, and so on, examining a total of 20229 data. He derived the same logarithmic law as Newcomb, and noted that this law will prevail even if we took reciprocals of the data or mutatis mutandis if we write the numbers in any other base.

Since the 1938 paper of Benford this phenomenon is called "Benford's law" by many, however some other names are also in use today, such as "Newcomb-Benford law", "First digit law", or "Significant digit law". The expression "Digital analysis", found in the title of this paper as well, refers to some techniques based on the phenomenon.[1]

Since the 1970s, several statistical applications have been developed in order to make use of Benford's law. The field of possible applications is wide-ranging: starting from finance/accounting/auditing, via quality control of socio-economic models to fields of computer development. This study has the following structure. In section 1, we will make a short overview of the possible heuristic, philosophical and all other explanations of this phenomenon. Section 2 lists the important mathematical results, while audit applications and the descriptions of the most commonly applied statistical tests can be found in section 3.

1. Explanations of the first digit phenomenon

As Benford's law is a widespread empirical phenomenon, several authors tried to explain it. Obviously, some data series do not obey this law at all: look at for example the body heights of adults in meters: more than 90 percent begins with 1; or telephone numbers of a city: some digits are not available as starting digits. Despite these simple counterexamples, some authors stated that this phenomenon is the inherent result of our numeral system (the way we currently write numbers, the positional system), so Benford's law roots in our numeral system. There are some more complex explanations as well, and in this part we will show some genuine ideas. But first let us investigate empirical observations that cannot be accounted to simple "coincidence". An obvious example for the digit 1's excess frequency is the basic multiplication table.

The ratio of leading 1's (18/81=2/9) in the basic multiplication table differs from the ratio 30.1 percent, but it is still the double of the expected one-ninth. (See Table 2.) In the continuous analogy, consider the product of n real numbers, each between 1 and 10 (10 not included). It can be shown that the frequency of first digits tends to obey Benford's law as n goes to infinity.

The most straightforward and real life example is the one of bank account balances. Suppose that we have EUR100 in our bank account with 10 percent interest. It takes more than 7 years to reach EUR200. To reach EUR300, more than additional 4 years are needed. One can easily see that the least time passes when we wait between EUR900 and EUR1000. Later, however, between EUR1000 and EUR2000, we have to wait once again more than 7 years. So, one can see that in the largest part of times, our account balance starts with 1, and in the smallest part

[1] Nevertheless the word "digital" can be confusing for those not acquainted with the topic, maybe a much clearer name would be "analysis of digits".

of times, it starts with 9. Obviously, it remains true if we use any other starting amount or interest rate, so if we suppose that new bank deposits accumulate in a constant rate, the first digits of all bank balances in a given moment must approximately obey Benford's law. Thus the law is true for any geometrical sequence in the form of $a \cdot q^n$ except the case where q is a power of 10 (for the function $f(x) = a \cdot q^x$, no such restrictions apply).

Table 2

Basic multiplication table

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	1	2	3	4	5	6	7	8	9
2	2	4	6	8	10	12	14	16	18
3	3	6	9	12	15	18	21	24	27
4	4	8	12	16	20	24	28	32	36
5	5	10	15	20	25	30	35	40	45
6	6	12	18	24	30	36	42	48	54
7	7	14	21	28	35	42	49	56	63
8	8	16	24	32	40	48	56	64	72
9	9	18	27	36	45	54	63	72	81

Several socio-economic data series and even data on natural phenomena follow lognormal-like distribution. It is quite intuitive, that in case of such underlying distribution, the first digits from a random sample will not follow uniform distribution and there will be a bias towards the smaller digits.

One could summarize Benford's law as 'if we pick a number at random, the probability that it starts with digit d is $\log_{10}(1+1/d)$'. The main fallacy of this sentence is that in fact we have no natural method to "pick a number at random" from the set of all positive (real or integer) numbers. The set of numbers starting with digit d has no natural density, neither among the reals nor among the integers, unlike for example even or odd numbers have (1/2).

The first intuition to calculate a density is as natural as simple: suppose that our set of positive integers has an upper limit, as real sets mostly have (for example areas of rivers from Benford's paper must not be larger than the area of Earth itself). If all upper limits were equally likely in each magnitude, the following logic could be used to calculate the probability of choosing a number with starting digit 1 (in a certain magnitude); in a given magnitude, let us say numbers between 10 and 99, we have:

$$\text{Prob (first digit} = 1) = \frac{1}{90}\left(\frac{1}{1}+\frac{2}{2}+\ldots+\frac{10}{10}+\frac{10}{11}+\frac{10}{12}+\frac{10}{13}+\ldots+\frac{10}{90}\right) \approx 0.35 .$$

The reader is called to try the same for other magnitudes, e.g. 100–999, or 1000–9999, etc. It is worth the simple spreadsheet exercise that will keep showing similar ratios among all magnitudes. But do not forget: we assumed something not at all innocent about the distribution of upper limits. If we had used other upper limit distribution, we had had other results.

A possible way to correctly treat the first digit problem mathematically is to calculate the frequencies of starting digits between 0 and n, and then evaluate its limit as n goes to infinity. These sequences are however not convergent, neither on real numbers nor on natural numbers so in order to calculate them, we should apply a summation method. For the discrete case with integers only, the first result known to me is the one described in *Flehinger* [1966] that makes use of a reiterated Cesaro method.[2] Let $p_d^1(n)$ denote the rate of positive integers $<n$ so that the first digit equals or less than d. As we saw before, $p_d^1(n)$ does not converge, so we define $p_d^t(n) = \frac{1}{n}\sum_{i=1}^{n} p_d^{t-1}(i)$ as a re-iterated "frequency". Flehinger proves that these series have diminishing oscillation and that the process converges to $\log_{10}(d + 1)$, in the following sense:

$$\lim_t \liminf_n p_d^t(n) = \lim_t \limsup_n p_d^t(n) = \log_{10}(d+1).$$

This result includes the original law, since the rate of occurrence, the "probability" of digit d equals $\log_{10}(d+1) - \log_{10}((d-1)+1) = \log_{10}(1+1/d)$.

It can be shown that the Flehinger method is stronger than (it means its equivalent to but can be applied on more series than) any other iterations of Cesaro methods, and that it is equivalent to any matrix method whenever it applies. There are also continuous versions of this approach, with integration schemes or Fourier analysis. One interesting result is the so called Stigler's law, where the probability of the digits differs from Benford's law, but the tendency remains. The inconsistency between Stigler's law and Benford's law indicates that summation methods are ad hoc, and they constitute no real explanation to the phenomenon.

Besides summation methods, invariance principles are another interesting interpretation of the problem. The most basic version of invariance is scale-invariance (*Pinkham* [1961]), or the independence of measurement units. Empirical tests have shown indeed that any set that obeys Benford's law keeps to obey it when multiplied by any constant, for example when exchanging metric units to imperial ones, or dollars to euros. It was conjectured that if a general law of significant digits exists at all, than it must be scale-invariant, i.e. that the frequency of leading digits should remain the same if the underlying set is multiplied by any constant. Similar argument leads

[2] If a series, like for example $\Sigma(-1)^n=1+(-1)+1+(-1)+\dots$ is divergent, it is still possible to evaluate it by a summation method. Such evaluations should have the usual value for convergent series; therefore they are generalizations of the standard definition. Cesaro method is one of the basic summation methods.

to another invariance property, translation invariance (*Raimi* [1969]). Some changes of measurement units come with translation, like for example between Kelvin and Celsius grades.

A more general invariance property is base invariance, or the independence of numeral system bases. For base invariance arguments, a more general form of Benford's law is used:

Prob (first significant digit base $b = d$) = $\log_b(1 + 1/d)$, where $d = 1, 2,\ldots b–1$.

It can be proved that some invariance concepts are (nearly) equivalent to Benford's law. By invariance arguments it is possible either to characterize (a set of) probability measures on positive reals (or integers) so that the measure itself is invariant to the given concept, or simply to characterize sets where the natural measure for significant digit distribution obeys the invariance concept. After the characterization, the explanation of first digit phenomenon is simply based on the fact that most real-world distributions match or almost match the characteristics defined by the appropriate invariance concept.

The last group of explanations is concerned with mixture of distributions, or sampling from random distributions. Individual data series in Benford's paper did not closely follow Benford's law, but for the union of all data series there was a good fit. Even the most fundamental paper on the early-stage history of Benford's law (*Raimi* [1976]) devotes a separate section to this approach, listing several important results, including the Furry–Hurwitz convolution which roughly states that the product of i.i.d. random variables converges to Benford's law. The most important result in this field can be found in *Hill* [1996], and it roughly states that if the data are taken from diverse-enough sources, the combined data series will tend to obey Benford's law.

2. Mathematical aspects of Benford's law

In this section we provide a short insight to the mathematics of Benford's law. In order to facilitate the later results, we need some basic definitions.

Definition 1. (mantissa-function). For each $b > 1$, the base *b* mantissa function is the function $M_b : \mathbb{R}^+ \to [1, b)$, such that $M_b(x) = x \cdot b^{-[\log_b x]}$ for any $x \in \mathbb{R}^+$, where $[r]$ means the integer part of *r*, the largest integer not larger than *r*. $M_b(x)$ will be called the mantissa of *x*.

Remark. From here on the numeral base is always a positive real number, larger than 1.

Example. The mantissa-function does nothing else than shifts the integer-separator (decimal point) so that the result is between 1 and b: $M_{10}(e) = M_{10}(100 \cdot e) = M_{10}(e/100) = e$.

Definition 2. (significant digit-k-string). The i^{th} significant digit-k-string of a positive real number x is denoted by $S_{i,k}^{(b)}(x)$ and it means the string of digits in $M_b(x)$ from position b^{-i+1} to position b^{-i-k+2}. In the default case of $k = 1$, the k subscript will be omitted, just like the (b) superscript, if we do not want to stress the base we are working with. In case of equivalent b-adic expansions, their lexicographic maximum will be considered.

Remark. By definition of $M_b(x)$, the first significant digit $S_1(x) \neq 0$.

Example. $S_1(\pi) = S_1(\pi/100000) = 3$, $S_2(1999.\dot{9}) = S_2(0.02) = 0$, $S_{1,3}(\pi) = 314$, $S_1^{(3)}(\pi) = 1$.

The main concern in the literature was the definition of "picking a number at random", since both Newcomb's and Benford's paper lacked the proper definition of the probability space. Now we will define the "natural" measure space for the first digit phenomenon.

Definition 3. (base-b mantissa sigma-algebra). For any $E \in \mathcal{B}([1,b))$ put $\mathcal{M}_b(E) = \bigcup_{n \in Z} b^n E$, and put $\mathcal{M}_b = \{\mathcal{M}_b(E)\colon E \in \mathcal{B}([1,b))\}$. Hereafter $\mathcal{M}_b$ will be called base-b mantissa sigma-algebra. ($\mathcal{B}(A)$ denotes the Borel sets on A).

Remarks. It is easy to see that $\mathcal{M}_b$ is a σ-algebra and that it is exactly the σ-algebra generated by M_b, the mantissa-function. One can check that bounded sets are never measurable on the measure-space ($\mathbb{R}^+$, $\mathcal{M}_b$) so the problem cited e.g. by *Raimi* [1969] of finding the median of all positive reals does not arise. In fact, all nonempty sets in a mantissa σ-algebra are log-periodic, and therefore unbounded with accumulation points at zero and infinity.

Example. $\mathcal{M}_b(\{1\}) = \{b^n\colon n \in \mathbb{Z}\}$, $\mathcal{M}_b([1,b)) = \mathbb{R}^+$.

Lemma 4. (properties of $\mathcal{M}_b$). For all bases, the following properties hold for $\mathcal{M}_b$:

1. $\mathcal{M}_b$ is closed under scalar multiplication, i.e., if $A \in \mathcal{M}_b$ and $\alpha > 0$, then $\alpha A \in \mathcal{M}_b$.

2. $\mathcal{M}_b$ is closed under integer roots, i.e., if $A \in \mathcal{M}_b$ and $a \in \mathbb{N}^+$, then $A^{1/a} \in \mathcal{M}_b$.
3. $\mathcal{M}_b(E) = \mathcal{M}_{b^n}\left(\bigcup_{k=0}^{n-1} b^k E\right) = \bigcup_{k=0}^{n-1}\left(\mathcal{M}_{b^n}\left(b^k E\right)\right)$ for all $n \in \mathbb{N}$ and $E \in \mathcal{B}([1,b))$.
4. $\mathcal{M}_b \subset \mathcal{M}_{b^n} \subset \mathcal{B}\left(\mathbb{R}^+\right)$ for all $n \in \mathbb{N}$.

Remarks. Inclusions in *4* are strict. The set $\mathcal{M}_{100}(2,20)$ for example is not in $\mathcal{M}_{10}$, but all elements of $\mathcal{M}_{10}$ are also elements of $\mathcal{M}_{100}$, since $\mathcal{B}([1,10)) \cup \mathcal{B}([10,100)) \subset \mathcal{B}([1,100))$. Note that contrary to scalar multiplication, $\mathcal{M}_b$ is not closed under scalar addition. For example translating the set $\mathcal{M}_{10}(\{1\})$ by 1, $\mathcal{M}_{10}(\{1\})+1 = \{\ldots, 1.01, 1.1, 2, 11, 101, \ldots\}$ is obviously not in $\mathcal{M}_{10}$. (Remember that unlike this set, all sets in a mantissa σ-algebra have an accumulation point at 0.)

Proof. All conclusions follow directly from definitions.

Definition 5. (base independent mantissa sigma-algebra). The base independent mantissa σ-algebra is defined as the σ-closure of all $\mathcal{M}_b$-s, formally $\mathcal{M}_\infty = \sigma(\cup \mathcal{M}_b)$.

Remarks. By the usual arguments, $\mathcal{M}_\infty$ (along with $\cup \mathcal{M}_b$) keeps being closed under scalar multiplication and integer roots. Additionally, they are also closed under powers.

The motivation to introduce $\mathcal{M}_\infty$ is that it is the tightest σ-algebra where all events like "(first digit base 10 is 1) and (third digit base 7 is 3)" can be described. By the next lemma, the tightest such σ-algebra is the set of Borels.

Lemma 6. (characterization of $\mathcal{M}_\infty$). $\mathcal{M}_\infty = \mathcal{B}\left(\mathbb{R}^+\right)$.

Proof. Since $\mathcal{M}_b \subset \mathcal{B}\left(\mathbb{R}^+\right)$ for all b and $\mathcal{M}_\infty = \sigma(\cup \mathcal{M}_b)$, we only have to prove that $\mathcal{M}_\infty$ includes all intervals on $\mathbb{R}^+$. Now consider any interval I on $\mathbb{R}^+$, and define $I_b = \bigcup_{i \in Z} M_b(I) \cdot b^i$ for all b= 2, 3, 4… . Obviously, $I_b \in \mathcal{M}_b$ for all base b. Since $\mathcal{M}_\infty$ is a σ-algebra, it will include $I = \cap I_b$ for all I.

Definition 7. (base-b mantissa distribution). For any probability measure $\mathcal{P}$ on $(\mathbb{R}^+, \mathcal{M}_\infty)$ the base-b mantissa distribution $\mathcal{P}_b$ is defined on $([1,b), \mathcal{B}([1,10)))$ in the following way: $\mathcal{P}_b(E) = \mathcal{P}(\mathcal{M}_b(E))$.

Definition 8. (Benford's law). A probability measure $\mathcal{P}$ on $(\mathbb{R}^+, \mathcal{M}_\infty)$ is said to obey Benford's law base b, if for any $x \in [1,b)$ $\mathcal{P}_b([1,x)) = \log_b x$. $\mathcal{P}$ is said to obey Benford's law if it obeys Benford's law for any base. From here on $\mathcal{P}_{BL}$ will be used to represent any probability measure that obeys Benford's law.

Remarks. Benford's law for binary (base 2) numbers means that the probability of a starting 1 is $\log_2 2=1$, as in fact it is. Given the base, we can easily derive specific laws for n^{th} digits, or for groups of digits as we will do in the examples. Another interesting feature of Benford distribution is that the individual probabilities of n^{th} digits are not independent; this dependence however decreases as n goes to infinity.

Example. In base 10, the Benford probability of a number to start with 102 (first digit 1, second digit 0, third digit 2) is the measure of mantissa-interval [1.02, 1.03): $\log_{10}1.03-\log_{10}1.02=\log_{10}(1+1/102)\approx 0{,}42\%$.

Example. The Benford probability of a second digit d can be calculated by summing the probabilities of starting strings $1d$, $2d$, $3d$,… $9d$, $\mathcal{P}_{BL}\left(S_2(x)=d\right)=$ $=\sum_{i=1}^{9}\log_{10}\left(1+1/(10i+d)\right)$.

Example. To show that individual digit probabilities are not independent, we simply have to calculate the marginal probabilities and the joint probability: $\mathcal{P}_{BL}(S_1(x)=1)=\log_{10}(1+1/1)$, $\mathcal{P}_{BL}(S_2(x)=1)=\sum_{i=1}^{9}\log_{10}\left(1+1/(10i+1)\right)$ and $\mathcal{P}_{BL}(S_1(x)=1\,\&\,S_2(x)=1)=\log_{10}(1+1/11)$. The products of marginal probabilities will differ from joint probability; hence individual digit probabilities are not independent.

The following theorem is a kind of by-pass in our logic, but since it states the "converse" of Benford's law, we focus to it at this point. It is important, because it is the theoretical justification for the last-digit tests in digital analysis.

Proposition 9. (least significant digit law). If a random variable X has a density, then for all bases, $S_i^{(b)}(x)$ converges to uniformly distribution on $\{0, 1, \ldots, b-1\}$ as i goes to infinity.

Proof. See *Hill* and *Schürger* [2005], Corollary 4.4.

Benford's law has been regarded almost from the beginning as a "universal law". As such, it was expected to be independent of measurement units, of "the scale", which mostly means the invariance of digit distribution if the whole data set is multiplied by any positive constant. (Translation invariance will not be considered here.) The next paragraphs clarify the connection between scale-invariance and Benford's law.

Definition 10. (scale-invariant). Let $\mathcal{F}$ be a collection of sets over $\mathbb{R}^+$ that is closed under scalar multiplication, with $\mathbb{R}^+\in\mathcal{F}$ and $\mathcal{G}=\sigma(\mathcal{F})$. A probability measure

$\mathcal{P}$ on ($\mathbb{R}^+$, $\mathcal{G}$) is called scale-invariant on $\mathcal{F}$ if for any $S \in \mathcal{F}$ and for any $\alpha > 0$, $\mathcal{P}(S) = \mathcal{P}(\alpha S)$.

Remarks. Note, that in this sense no scale-invariant probability (or even nontrivial finite) measure can exist on any $\mathcal{F}$ that includes bounded sets, like e.g. $\mathcal{B}\left(\mathbb{R}^+\right)$ does. If there are bounded sets, the measure of $\mathbb{R}^+$ must either be 0 or ∞, because of scale-invariance.

In the light of this remark, the result of the next corollary is somewhat disappointing.

Corollary 11. There is no scale-invariant probability measure on $\mathcal{M}_\infty$.

Proof. Obvious from the remark and the lemma 6.

The base-independent σ-algebra was introduced in order to measure joint significant digit statements from different bases. It turned out, however, that it equals the Borels, and any significant law of this generality cannot be totally scale invariant. The following theorem therefore shall suffice with a tighter characterization.

Theorem 12. (characterization of scale-invariant probability measures). A probability measure $\mathcal{P}$ on ($\mathbb{R}^+$, $\mathcal{M}_\infty$) is scale-invariant on $\cup\mathcal{M}_b$ if and only if it obeys Benford's law for any base.

Remarks. Note, that this theorem precludes the possibility of such scale-invariant probability measure on $\cup\mathcal{M}_b$ that obeys Benford's law for some base b, but not for some other bases. In case of a partly scale-invariant measure (only on a subset of $\cup\mathcal{M}_b$), however, it is possible to create such base-dependent behaviour.

Proof. Obviously, if $\mathcal{P}$ is scale-invariant then it is also scale-invariant on $\mathcal{M}_b$ for any base. By Theorem 8 in *Hill* [1995a], the scale-invariant measure on ($\mathbb{R}^+$, $\mathcal{M}_b$) is unique and it is exactly Benford's law. It is an easy calculation that for different bases b and c, the two Benford measures of any $A \in \mathcal{M}_b \cap \mathcal{M}_c$ are the same, and the logarithmic mantissa measure itself is scale invariant on any $\mathcal{M}_b$, thus on $\cup\mathcal{M}_b$ also.

Now let us have another by-pass, partly back to the definition of $\mathcal{M}_\infty$. Consider the natural logarithm as an isomorphism between $\mathbb{R}^+$ and $\mathbb{R}$. All the (previously log-periodic) sets in $\mathcal{M}_b$ are transformed into periodic sets, with periodicity of ln b. The image of $\mathcal{M}_b$ is simply consists of all sets that are periodic with ln b, hence the isomorphic image of $\cup\mathcal{M}_b$ equals the collection of all periodic subsets of the real line. Because of the transformation, multiplication is now translation and powers are multiplication. If we multiply a set in the image, it remains periodic, but with other periodicity. In the original, it means the change of base which is the intuition behind our next definition.

Definition 13. (base-invariant). Let $\mathcal{F}$ be a collection of sets over $\mathbb{R}^+$ that is closed under powers, with $\mathbb{R}^+ \in \mathcal{F}$ and $\mathcal{G}=\sigma(\mathcal{F})$. A probability measure $\mathcal{P}$ on $(\mathbb{R}^+, \mathcal{G})$ is called base-invariant on $\mathcal{F}$, if for any $S \in \mathcal{F}$ and for any $a \in \mathbb{R} \setminus \{0\}$, $\mathcal{P}(S) = \mathcal{P}(S^a)$.

Remarks. In his original definition, Hill worked on $\mathcal{F}=\mathcal{G}=\mathcal{M}_b$, and therefore he could only prescribe the invariance for integer roots ($\mathcal{M}_b$ is closed under integer roots). In our case, however, $\mathcal{F} = \cup\mathcal{M}_b$ is closed under any powers, not only integer roots. Note, that this definition includes inversion-invariance, which is in fact the power -1.

Example. By simple calculation, $\mathcal{P}_{BL}$ is base-invariant on $\cup\mathcal{M}_b$.

Example. The most trivial example for base-invariance on $\cup\mathcal{M}_b$ is the Dirac measure concentrated on 1, denoted by $\delta_1(\cdot)$.

Remarks. Note, that $\delta_1(\cdot)$ keeps being base-invariant even on $\mathcal{B}(\mathbb{R}^+)$. Hence, contrary to scale-invariance, there exists base-invariant probability measure on $\mathcal{M}_\infty$.

If there are some base-invariant probability measures on $\mathcal{F}$, it is easy to show that all their convex combinations are base-invariant also. Therefore it is maybe surprising, that the class of all base-invariant measures on $\cup\mathcal{M}_b$ is restricted to the examples given previously (and of course to their convex combinations.)

Theorem 14. (characterization of base-invariant probability measures). If a probability measure $\mathcal{P}$ on $(\mathbb{R}^+, \mathcal{M}_\infty)$ is base-invariant on $\cup\mathcal{M}_b$, then there exists a unique scalar $q \in [0,1]$ such that $\mathcal{P}(A)=q\cdot\mathcal{P}_{BL}(A)+(1-q)\cdot\delta_1(A)$ for any set $A \in \cup\mathcal{M}_b$.

Proof. If $\mathcal{P}$ is base-invariant on $\cup\mathcal{M}_b$, then it is also base-invariant on $\mathcal{M}_b$ in the sense of Hill's definition. For such measures on $(\mathbb{R}^+, \mathcal{M}_b)$ the proof can be found under Theorem 3.5 in *Hill* [1995a]. Now suppose indirectly that for two different bases b_1 and b_2, we have two different scalars, q_1 and q_2 respectively, where $q_1 \neq q_2$. Consider a "good-behaving" set A from the base b_1. By the power closeness of $\cup\mathcal{M}_b$, we always have an appropriate power (for example ln b_2/ln b_1) so that the transformed image of A is in the base b_2 mantissa σ-algebra. But $\mathcal{P}$ is base-invariant, and q for a given base is unique, which is a contradiction.

The next concept shows another feature of Benford's law.

Definition 15. (sum-invariant). A probability measure $\mathcal{P}$ on $(\mathbb{R}^+, \mathcal{M}_\infty)$ will be called sum-invariant, if with any fixed i, k and base b, the expected value of $M_b(x)\cdot 1_{\left\{x:S_{i,k}^{(b)}(x)=d_1 d_2 \ldots d_k\right\}}(x)$ is constant for any k-string $d_1 d_2 \ldots d_k$.

Remarks. In words, sum-invariance means that the "expected-mantissa-sum" of any k-string set is independent of the digits that form the k-string. Specifically, summing the mantissae of all numbers starting with 1 and starting with 9 will yield the same sum in expectation.

Theorem 16. (characterization of sum-invariant probability measures). A probability measure $\mathcal{P}$ on ($\mathbb{R}^+$, $\mathcal{M}_\infty$) is sum-invariant if and only if it obeys Benford's law.

Proof. See for example Theorem 4.1. in *Allaart* [1997].

Corollary 17. If X_i, i=1…n are random variables with distribution $\mathcal{P}_{BL}$, then the expected value of $\sum\left\{M_b(X_i) : S_{i,k}^{(b)}\left(X_i\right) = d_1 d_2 ... d_k\right\}$ is constant for any k-string $d_1 d_2 ... d_k$.

Remarks. Note the difference between 15 and 17: the corollary states that *a randomly generated set of numbers* from a Benford distribution has equal expected mantissa sums for all possible k-strings.

Proof. See the proof of Corollary 4.2. in *Allaart* [1997]. (This corollary is somewhat weaker than the corollary proved by Allaart.)

3. Applications

The evolution of digital analysis has been long and slow. In 1972, the now famous economist, *Hal R. Varian* suggested in a reader's letter to the editor this law to test the reasonableness of economic models. His reasoning was the following:

"After all, Benford's law is just a curious empirical, almost numerological, phenomena; why should it have anything to do with economic forecasting? However, one must admit that if the input data did obey Benford's law while the output didn't… well, that would make one feel a trifle uneasy about actually using that output." (*Varian* [1972] p. 65.)

This little suggestion did not attract too much attention and the first real applications of Benford's law date from the second half of the 1980s. In *Carslaw* [1988] Benford's law was used on second digits to show that New Zealand companies are systematically rounding up their income figures: second digit 0 appeared in excess frequency among the examined income statements, while there were fewer second digit 9s than expected. This study was repeated and expanded on United States data in *Thomas* [1989] where the opposite effect (rounding down) was also found true for reported losses, and additionally, reported earnings per share (EPS) in the United States were multiples of 5 cents more often and had an ending digit 9 fewer times than expected.

The phenomenon has found its most important appliers among (forensic) auditors, mostly tax auditors of internal revenue services and internal auditors of larger firms, since they mostly have to analyze huge data series and find the signs of manipulation. They can apply Benford's law because, as *Hill* [1988] showed, if people are "guessing" numbers, these sequences will not conform Benford's law. Since 1992, *Nigrini* has the most pioneering results in developing new applications in auditing.

There are still some other applications besides "hardcore auditing" but they are mostly somehow border-line applications. One interesting application is for example the one proposed by *Schraepler* and *Wagner* [2003], the detection of faked interviews in statistical surveys. But in the broader sense, it is auditing as well. Maybe a good example for real non-audit applications is optimizing floating-point calculations. Recent results of computer science show that floating-point calculations are based on inputs that follow Benford's law rather than uniform distribution. In this case, it is possible to take advantage of this prior knowledge when designing future computer architectures, just like in the case of typewriter (and computer) keyboard layouts which were based on the known differences of character usage frequencies (which even earlier were used by printers-typesetters to optimally allocate lead characters in cases). Audit applications are themselves wide-ranging; here we list only the most common areas of analysis. After listing the areas, we give an overview on the statistical test that can be applied during the analysis of these areas.

The basic process of auditing can be divided from our point of view into two important parts: analysis and testing. Analysis is needed to mark risky areas in order to optimally allocate audit resources (the precious working time of auditors) during the (substantive) testing procedures. Note, that auditing is full scope only in exceptional cases, in most cases audit opinions are based on sample testing. In audit applications – as a rule of thumb – the riskier the area, the larger the sample size is. The first main task is to see whether the assumption of Benford-like behaviour of the audited population is reasonable. The main methods in this field are the 1st, 2nd, 3rd and 4th digit analyses.

– *First digits analysis* is an initial test of reasonableness which is not intended to select audit samples. Generally it is the first test to be performed during digital analysis of accounts.

– *Second/third/fourth digits analysis* is used as an additional test of reasonableness. These tests are also not appropriate for sample selection because of the large resulting sample size. Along with the first digits test, they are high level tests to establish whether the assumption of Benford-like distribution is reasonable. After performing the reasonability tests, the selection of risky areas[3] can start.

– *First two digits analysis* (sometimes along with first three or even first four digits analysis) can be used for sample selection but mostly in an indirect way. Those combinations of digit groups (or specifically the transactions underlying these digit groups) that have significantly different frequency from the postulated one are ideal candidates for closer investigation. Since real data sets are only approximately Benfordian, the first two/three/four digits analysis often results in false alarms which in

[3] "Risky areas" means here only those risky areas that have abnormal digital behavior.

turn can mean obsolete audit efforts. So these tests mostly play an indirect role in sample selection rather than a direct one. They indicate which digit groups are "overused" and where the other analytic/substantive procedures shall focus.

– The *number duplication test* is a natural extension and a natural sequel of first two digits analysis. The common belief is that excess frequency of some leading digits is caused by duplications which can signal management inefficiencies, errors, (deliberate) misstatements and ultimately fraud. An example of management inefficiency in accounting is the procedure of one-by-one recording of repeated procurements instead of using a summary bill. Misstatement is for example the systematic rounding up of figures and fraud if someone in the firm sets up fictitious invoices for services that never happened. Special attention shall also be devoted to numbers just below psychological thresholds or internal authorization limits, since they are most prone to misstatements and fraud. When performing a number duplication test, a so-called Number Frequency Factor (NFF) is calculated for all data subsets defined by the auditor. In any subset, let c_i mean the number of items with the i^{th} value (but: instead of c_i=1, c_i=0 for single items, without duplication). If n is the cardinality of the subset, NFF is defined by:

$$NFF = 1 - \frac{\sum c_i^2}{n^2}.$$

If there are no duplicated items in the subset, then NFF=1, in the other extreme case when 1 item is duplicated n times NFF=0. The subsets chosen to further analytic/substantive testing are those with low NFF.

– *Rounded numbers test* is used in areas where estimation is in no way acceptable. Excess frequencies of numbers multiples of 5, 10, 25, 100, 1000 can signal unacceptable negligence or even possible fraud. Rounded numbers tests are especially appropriate for example when testing the inventories of a takeover target and the target itself provided the figures for the potential acquirer.

– *Last two digits test* is a more focused version of rounded numbers test and it is relevant when auditors presume that systematic "number invention" is occurring. As we saw before in the mathematics section, the bias towards smaller digits diminishes as one move to the right. Although that result is an asymptotic one, simulations show that it is reasonable to suppose for real-life data sets that the last two digits are uniformly distributed with relative frequency of 1/100.

Modern auditing softwares (like CaseWare Idea, ACL etc.) mostly feature a so-called Benford module that can be used to perform the analyses described previously. Later in this paper, we will bring some demonstrative screenshots with the results of some performed test. Deeper insight to the analyzed data sets and results is not possible, due to the confidentiality of the source.

In most audit applications it is conjectured that the underlying distribution obeys (or almost obeys) Benford's law, except in "continuous auditing", where peculiar distributions different from Benford distribution are allowed; the reason is that some firm-specific factors can divert empirical frequencies and can cause a constant bias, peculiar to the given firm. In other auditing applications the firm is generally not yet known to the auditor, and examining the biases can reveal inefficient procedures (and of course efficient, but still peculiar procedures also).

Such a bias factor is for example if a hospital buys hypodermic syringes in standard packages and each buy is registered one-by-one. If the price of a package begins e.g. with five, the frequency of leading fives will be significantly more than their Benford frequency because of the volume hospitals are buying hypodermic syringes. In this case the excess frequency of leading fives is considered normal. The only change it inflicts in the analysis methods is that the previously audited peculiar distribution is taken as baseline instead of Benford distribution.

Any deviation from the postulated digit distribution can be attributed to one of the following two factors:

– sampling error,
– data manipulation, fraud, or errors, inefficiencies.

To distinguish between the two possible sources of deviation, several classical and non-classical tests can be applied. We will consider the following tests in this paper:

1. visual inspection,
2. Kolmogorov–Smirnov goodness-of-fit test,
3. mean absolute deviation test,
4. chi-square goodness-of-fit test,
5. z-statistic test,
6. a test based upon regression of theoretical values on empirical values,
7. summation test.

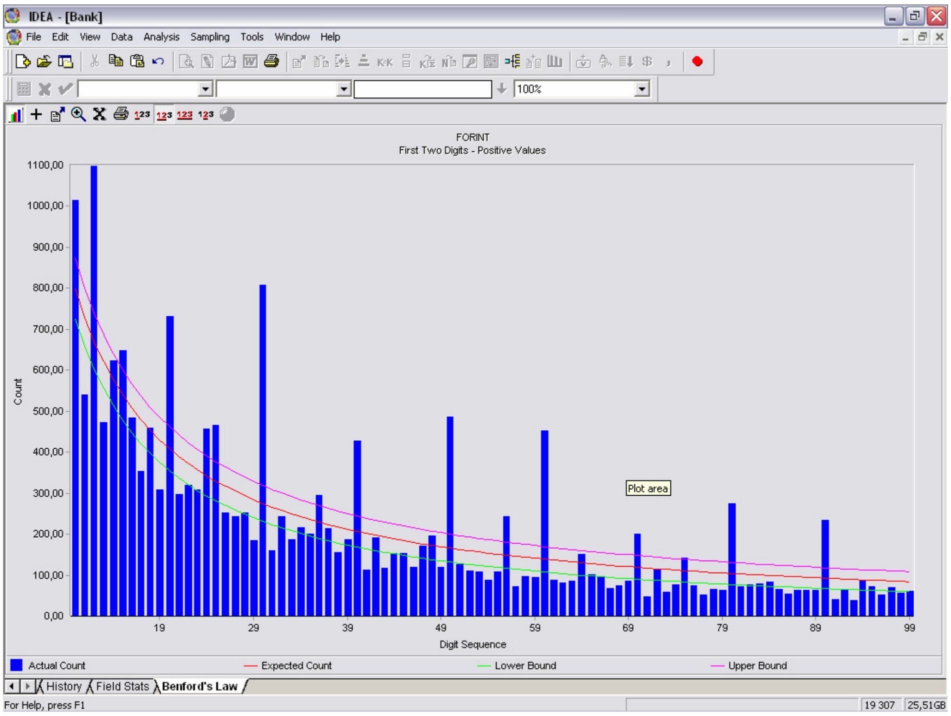
Tests from 1 to 6 are all to "measure" how well actual data fit the Benford distribution. Test 7 has other aims; it uses the sum invariance property of Benford distributions.

Visual inspection is an important first step in digital analysis, just like in many statistical applications. Plotting values together can provide a fast hint on the next step of application, therefore auditing software mostly support visual inspection. See for example Figure 1. for the visual inspection of a first two digits test.

The *Kolmogorov–Smirnov test* is based on the difference of empirical and theoretical cumulative distribution functions: in fact it measures their "distance" in a metric generated by the maximum norm on functions. The one-sided Kolmogorov–Smirnov test statistics are defined by

$$D_n(x) = \max\left|F_n(x) - F(x)\right|.$$

Figure 1. Visual inspection of the first two digit distribution in CaseWare Idea



For the first digits case the empirical distribution function F_n for sample size n is defined by

$$F_n\left(x\right)=\frac{\#(y:M_b\left(y\right)\leq x)}{n},$$

while the theoretical distribution $F\left(x\right)$ is given by 0, $\log_{10}x$, and 1 before 1, between 1 and 10 and after 10 respectively. (For other cases the functions are given analogously.) The probability distribution of this statistics, given that the null hypothesis of equality of distributions is true, does not depend on what the hypothesized distribution is, as long as it is continuous. The Kolmogorov–Smirnov test has however several important limitations:

– It only applies to continuous distributions (generally met in digital analysis).
– It tends to be more sensitive near the centre of the distribution than at the tails.
– Perhaps the most serious limitation is that the distribution must be fully specified. That is, if location, scale, and shape parameters are

estimated from the data, the critical region of the Kolmogorov–Smirnov test is no longer valid (generally does not apply in digital analysis).

Critical values for Kolmogorov–Smirnov test can be taken from statistical tables or statistical software packages.

The *mean absolute deviation test* (MAD) is based on the mean absolute deviation between empirical and theoretical relative frequencies. Formally the test statistics are given by

$$MAD = \sum_{i=1}^{9} \frac{\left| p_s^i - p_b^i \right|}{9} \text{ for first digit test,}$$

and by appropriate changes for other tests. Here p means relative frequency, i superscript denotes an appropriate first digit (or other category, e.g. digit group, last digit etc.) and the lower-case letters denote "sample" and "benford" ("baseline" in continuous auditing) respectively.

Instead of MAD some authors prefer MSE, the squared average error, but in fact it does not help solving the most serious limitation of this test: no objective critical values exist. As long as we know, there are currently two, somewhat subjective, methods to determine critical values. The first and oftener used method is given by *Nigrini*. His critical values are based on practical test experience and therefore must be used with caution. Table 3 lists these critical values for 3 types of tests.

Table 3

Critical values for three MAD tests suggested by Nigrini

Test type / decision	First digits only	Second digits only	First two digits
Close conformity	< 0.004	< 0.008	< 0.006
Acceptable conformity	0.004–0.008	0.008–0.012	0.006–0.012
Marginally acceptable conformity	0.008–0.012	0.012–0.016	0.012–0.018
Nonconformity	>0.012	>0.016	>0.018

Source: *Nigrini* [2000].

Other possible approach to get critical values is Monte Carlo simulation (*Posch* [2004]). In this case a random Benford set is generated first, and it is gradually contaminated with appropriate non-Benford random numbers. Evaluating the expected MAD at all stages gives a rough estimation how contaminated the original data set can be. This approach must also be used with caution since its results *depend* on the way non-Benford numbers are generated, and in addition, the higher the level of contamination the higher the variance of simulated MAD is.

Pearson's chi-square test is the standard textbook method in statistics to test how well an empirical population fits to a given theoretical distribution. It has however some drawbacks that should be considered before use:

– it is only an approximation so with small samples exact tests (e.g. multinomial test) should be preferred;
– chi-square distribution needs independent summands.

Testing for Benford distribution is performed on large data sets (1000+ or even 10000+) almost in all cases, so condition 1 is not a problem. In some audit applications the largest concern arises with the second condition, because data in some data set are not independent. The standard formula used to calculate the test statistic is:

$$\chi^2 = n \cdot \sum_{i=1}^{k} \frac{\left(p_s^i - p_b^i\right)^2}{p_b^i},$$

where n is the sample size, p means relative frequency, the i upper-case letter denotes an appropriate first digit (or other category, e.g. digit group, last digit etc.) and the lower-case letters denote "sample" and "benford" (or other baseline) respectively. Critical values for appropriate level of significance and degrees of freedom ($k – 1$) can be taken from statistical tables, but most software packages include them as well.

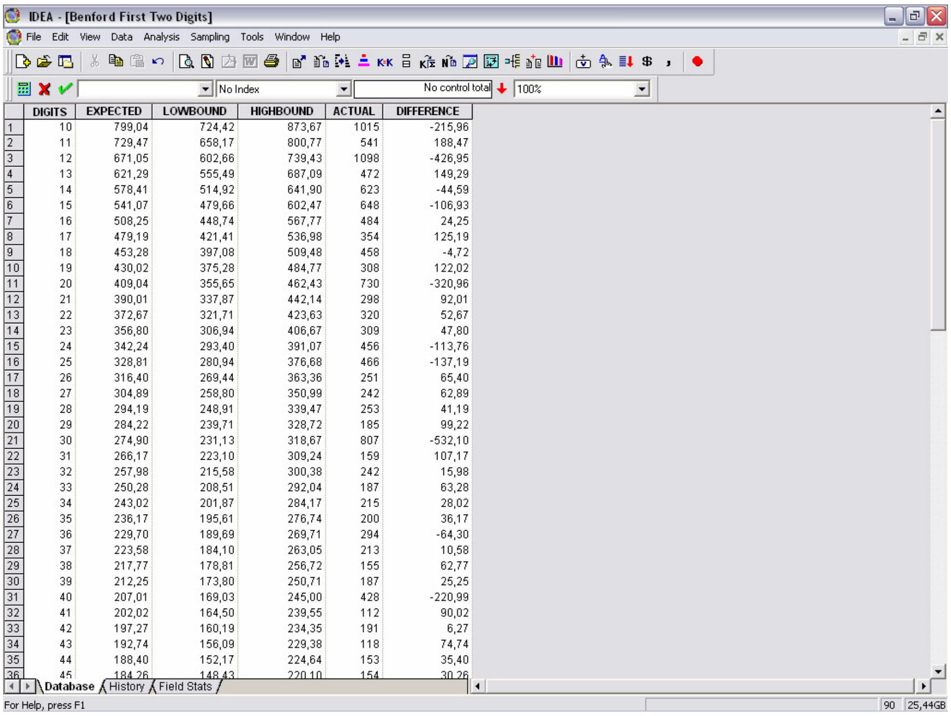
The *z-statistic tests* are best used to detect which category has an excess frequency in the sample, relative to theoretical values. While a chi-square test evaluates the data set as a whole, they are partial tests for individual categories. The z-statistic is calculated with the standard formula

$$z = \frac{p_s - p_b \pm \frac{1}{2n}}{\sqrt{\frac{p_b\left(1 - p_b\right)}{n}}},$$

where $1/2n$ is the correction term for continuity. This test is two-sided with critical values from standard normal distribution.

Figure 2 is the source data table for the graph of Figure 1 (the values given by Idea are absolute values). The software can automatically generate reports on the digital frequencies, and uses a high (99%) built-in level of significance for computing upper and lower bounds. The level of significance is set high in order to avoid false alarms (note that digital analysis is a tool for marking risky areas: false alarms mean larger samples than needed, therefore unnecessary audit efforts, thus waste of precious resources).

Figure 2. Raw data table of Figure 1. in CaseWare Idea (screenshot)



IDEA - [Benford First Two Digits]

	DIGITS	EXPECTED	LOWBOUND	HIGHBOUND	ACTUAL	DIFFERENCE
1	10	799,04	724,42	873,67	1015	-215,96
2	11	729,47	658,17	800,77	541	188,47
3	12	671,05	602,66	739,43	1098	-426,95
4	13	621,29	555,49	687,09	472	149,29
5	14	578,41	514,92	641,90	623	-44,59
6	15	541,07	479,66	602,47	648	-106,93
7	16	508,25	448,74	567,77	484	24,25
8	17	479,19	421,41	536,98	354	125,19
9	18	453,28	397,08	509,48	458	-4,72
10	19	430,02	375,28	484,77	308	122,02
11	20	409,04	355,65	462,43	730	-320,96
12	21	390,01	337,87	442,14	298	92,01
13	22	372,67	321,71	423,63	320	52,67
14	23	356,80	306,94	406,67	309	47,80
15	24	342,24	293,40	391,07	456	-113,76
16	25	328,81	280,94	376,68	466	-137,19
17	26	316,40	269,44	363,36	251	65,40
18	27	304,89	258,80	350,99	242	62,89
19	28	294,19	248,91	339,47	253	41,19
20	29	284,22	239,71	328,72	185	99,22
21	30	274,90	231,13	318,67	807	-532,10
22	31	266,17	223,10	309,24	159	107,17
23	32	257,98	215,58	300,38	242	15,98
24	33	250,28	208,51	292,04	187	63,28
25	34	243,02	201,87	284,17	215	28,02
26	35	236,17	195,61	276,74	200	36,17
27	36	229,70	189,69	269,71	294	-64,30
28	37	223,58	184,10	263,05	213	10,58
29	38	217,77	178,81	256,72	155	62,77
30	39	212,25	173,80	250,71	187	25,25
31	40	207,01	169,03	245,00	428	-220,99
32	41	202,02	164,50	239,55	112	90,02
33	42	197,27	160,19	234,35	191	6,27
34	43	192,74	156,09	229,38	118	74,74
35	44	188,40	152,17	224,64	153	35,40
36	45	184,26	148,43	220,10	154	30,26

Database History Field Stats

For Help, press F1

Consider for example the first two digit group "10" in Figure 2. It was found in 1015 cases of total 19307 (5.2%) while Benford's law postulated 799 cases (4.1%). With the given formula that difference is significant at any levels in use, which is not surprising looking at Figure 1. Auditing applications that support digital analysis therefore automatically draw upper and lower confidence limits to facilitate visual inspection.

An interesting solution is to *regress theoretical frequencies on empirical frequencies*. The following equation is estimated:

$$p_s = \alpha + \beta \cdot p_b \,.$$

The null hypothesis (perfect fit) means $\alpha=0$ and $\beta=1$. We can test it either partially or jointly with the common tests. Note that in case of first (second, third etc.) digit frequencies we have no more than 10 observations so this type of testing is only appropriate if we test 2 or more digits jointly (90 or more observations).

Summation test is an alternative approach which has been developed by *Nigrini* [1992], but its mathematical foundations were laid down later by *Allart* [1997] who showed that summation invariance is an exclusive property of Benford distribution.

As we could see in Part 2, summation invariance is defined on mantissae, but this test is based on absolute numbers which makes it invalid in theory. Some empirical results by *Nigrini* [2000] show, however, that it is still valid in practice. It is because, in practice, magnitudes in a Benford set are independent of mantissae, while, in theory, it does not necessarily follow from the definition.

The main use of this test is to detect magnitude errors by summing values with the same starting digits. Significantly different sums may signal a possible error. *Nigrini* [2000] cites a true story from the Wall Street Journal where a man who expected a tax refund of USD513 received a letter from the US Internal Revenue Service informing him that he still owes USD300 000 007.57, and this happened that year to 3000 additional people. Such errors are not detectable with standard digital analysis tests since 3000 errors are no way significant among the whole Unites States population.

Table 4

When to apply and not to apply digital analysis

Uses of Benford analysis	Examples
When Benford analysis is likely useful	
Sets of numbers that result from mathematical combination of numbers – Result comes from two distributions	Accounts receivable (number sold × price) Accounts payable (number bought × price)
Transaction-level data – No need to sample	Disbursements, sales, expenses
On large data sets – The more observations, the better	Full year's transactions
Accounts that appear to conform – When the mean of a set of numbers is greater than the median and the skewness is positive	Most sets of accounting numbers
When Benford analysis is not likely useful	
Data set is comprised of assigned numbers	Check numbers, invoice numbers, zip codes
Numbers that are influenced by human thought	Prices set at psychological thresholds (USD1.99), ATM withdrawals
Accounts with a large number of firm-specific numbers	An account specifically set up to record USD100 refunds
Accounts with a built in minimum or maximum	Set of assets that must meet a threshold to be recorded
Where no transaction is recorded	Thefts, kickbacks, contract rigging

Source: *Durtschi–Hillison–Pacini* [2004].

So far in this part we had a short overview on the evolution of digital analysis, the main areas of application, and the statistical tests used in the analyses. What still remains to give some guidelines on when and how to apply digital analysis. In the field of auditing, most accounting-related data can be expected to obey Benford's law. It does not mean, however, that if a data set differs significantly from the expected then it is fraudulent. As we earlier mentioned, some firm-specific peculiarities can significantly divert digit distribution, as well as some universal ones can. Marketing activity often results in psychological pricing, and no one will be surprised to see firms rounding their figures. ATM withdrawals are predefined round numbers so extensive use of ATM-s can result in biased bank ledger. Administrative fees (for example the highway tax collected for the use of highways almost everywhere) are also predefined numbers and can have a similar effect. Table 4 summarizes when it is appropriate to use digital analysis.

There are some limitations also, based on the type of fraud. Significant deviations from Benford's law occur if the person committing fraud changes, removes or adds items so that he meanwhile hurts the original first digit distribution. To hurt a distribution significantly, the subset modified must be large enough relative to the whole set. If no registered items (the fraud is not even registered, like theft) or very few items are affected, digital analysis cannot be expected to detect anything. Other type of frauds, which affect only data sets that are not suitable for digital analysis, are also not to be detected by means of digital analysis: a good example is the same address of a vendor and of the person endorsing payments. Some possible fraud cannot be detected by digital analysis uses the invariance principle: multiplying all items with the same number remains undetected.

Durtschi–Hillison–Pacini [2004] had a good indicative example on how effective digital analysis can be. If we suppose that 3 percent of all transactions is fraudulent, and when checked, digital analysis identifies fraudulent accounts in 75 percent of times, the probability of a real fraud when digital analysis signals fraud can be calculated using Bayes' theorem:

$$P\left(F \mid S\right) = \frac{0.75 \cdot 0.03}{0.75 \cdot 0.03 + 0.25 \cdot 0.97} = 0.085 \approx 9\% \,.$$

This 9 percent is three times more than with simple random sampling, but it is still less than 10 percent. The result shows clearly the value of digital analysis: it is not a silver bullet; it is only a useful compliment to the arsenal of an auditor.

Reference

ALLAART, P. C. [1997]: An invariant-sum characterization of Benford's law. *Journal of Applied Probability*. Vol. 34. No. 1. p. 288–291.

BENFORD, F. [1938]: The law of anomalous numbers. *Proceedings of the American Philosophical Society.* Vol. 78. No. 4. p. 551–572.

BERGER, A. – BUNIMOVICH, L. A. – HILL, T.P. [2005]: One-dimensional dynamical systems and Benford's law. *Transactions of the American Mathematical Society.* Vol. 357. No. 1. p. 197–219.

BERGER, A. [2004]: *Dynamics and digits: on the ubiquity of Benford's law*. Vienna University of Technology. Working paper.

CARSLAW, C. [1988]: Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review.* Vol. 63. No. 2. p. 321–327.

DIACONIS, P. [1977]: The distribution of leading digits and uniform distribution mod 1. *Annals of Probability.* Vol. 5. No. 1. p. 72–81.

DRAKE, P.D. – NIGRINI, M. J. [2000]: Computer assisted analytical procedures using Benford's law. *Journal of Accounting Education.* Vol. 18. No. 2. p. 127–146.

DURTSCHI, C. – HILLISON, W. – PACINI, C. [2004]: The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting.* Vol. 5. No. 1524–5586. p. 17–34.

FELLER, W. [1971]: *An introduction to probability theory and its applications*. J. Wiley. New York.

FLEHINGER, B.-J. [1966]: On the probability that a random integer has initial digit A. *American Mathematical Monthly.* Vol. 73. No. 10. p. 1056–1061.

HILL, T. P. [1988]: Random-number guessing and the first digit phenomenon. *Psychological Reports.* Vol. 62. p. 967–971.

HILL, T. P. [1995a]: Base-invariance implies Benford's law. *Proceedings of the American Mathematical Society.* Vol. 123. No. 3. p. 887–895.

HILL, T. P. [1995b]: The significant-digit phenomenon. *American Mathematical Monthly.* Vol. 102. No. 4. p. 322–326.

HILL, T. P. [1996]: A statistical derivation of the significant-digit law. *Statistical Science.* Vol. 10. No. 4. p. 354–363.

HILL, T. P. [1999]: The difficulty of faking data. *Chance.* Vol. 26. No. 1. p. 8–13.

HILL, T. P.– SCHÜRGER, K. [2005]: Regularity of digits and significant digits of random variables. *Journal of Stochastic Processes and their Applications.* Vol. 115. No. 10. p. 1723–1743.

MILLER, S. J. – NIGRINI, M. J. [2006]: *Differences between independent variables and almost Benford behaviour.* Arxiv preprint math.PR/0601344. www.arxiv.org.

NEWCOMB, S. [1881]: Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics.* Vol. 4. No. 1. p. 39–40.

NIGRINI, M. J. [1992]: *The detection of income tax evasion through an analysis of digital frequencies.* University of Cincinnati. Cincinnati, OH. PhD. thesis. Working paper.

NIGRINI, M. J. [1996]: A taxpayer compliance application of Benford's law. *Journal of the American Taxation Association*. Vol. 18. No. 1. p. 72–91.

NIGRINI, M. J. – MITTERMAIER, L. [1997]: The use of Benford's law as an aid in analytical procedures. *Auditing: A Journal of Practice and Theory.* Vol. 16. No. 2. p. 52–67.

NIGRINI, M. J. [2000]: *Continuous Auditing*. Working paper.

PINKHAM, R. S. [1961]: On the distribution of first significant digits. *Annals of Mathematical Statistics.* Vol. 32. No. 4. p. 1223–1230.

POSCH, P. N. [2004]: *Ziffernanalyse in der Fälschungsaufspürung. Benford's Gesetz und Steuererklärungen in Theorie und Praxis.* Working paper.

RAIMI, R. A. [1969]: On the distribution of first significant figures. *American Mathematical Monthly.* Vol. 76. No. 4. p. 342–348.

RAIMI, R. A. [1976]: The first digit problem. *American Mathematical Monthly.* Vol. 83. No. 7. p. 521–538.

RAIMI, R. A. [1985]: The first digit phenomenon again. *Proceedings of the American Philosophical Society.* Vol. 129. No. 2. p. 211–219.

SCHRAEPLER, J.-P. – WAGNER, G.G. [2003]: *Identification, characteristics and impact of faked interviews in surveys*. IZA Discussion paper. No. 969. www.iza.org.

THOMAS, J. K. [1989]: Unusual patterns in reported earnings. *The Accounting Review.* Vol. 64. No. 4. p. 773–787.

VARIAN, H. R. [1972]: Benford's law. *The American Statistician.* Vol. 26. No. 3. p. 65–66.

On estimation of generalized logarithmic series distribution*

Anwar Hassan,
associate professor of
PG. Department of Statistics
University of Kashmir,
Srinagar, India.
E-mail: anwar_husan@yahoo.com

Khurshid Ahmad Mir,
senior lecturer of
Department of Statistics
Amar Singh College,
Srinagar, India
E-mail: khrshdmir@yahoo.com

In this paper we have studied the estimation of generalized logarithmic series distribution (GLSD) by the method of weighted discrepancies between observed and expected frequencies. The maximum likelihood, minimum chi-square and the discrimination information methods are special cases of the weighted discrepancies method. A new weighted technique, the empirical weighted rates of change (EWRC) for estimating the GLSD parameters has been obtained. We have fitted the GLSD to several zero-truncated biological data by different methods and observed that in most of the cases the GLSD provided a better fit than the usual logarithmic series distribution (LSD) by using EWRC method.

KEYWORDS:
Probability distributions. Estimations.

* The authors are highly thankful to the editor-in-chief *László Hunyadi* and the refree *József Varga* for their helpful comments and suggestions.

The generalized logarithmic series distribution (GLSD) characterized by two parameters α and β was defined by *Jain* and *Gupta* [1973]. The probability function of the GLSD model is given by

$$P(X=x)=\begin{cases}\dfrac{\theta\ \Gamma(\beta x)\alpha^{x}(1-\alpha)^{\beta x-x}}{x!\ \Gamma(\beta x-x+1)} & ;\ x=1,\ 2\ \dots,\\ 0 & \text{otherwise}\end{cases} \qquad /1/$$

where $\beta \geq 1$, $0<\alpha<\beta^{-1}$ and $\theta=-\dfrac{1}{\log(1-\alpha)}$.

The GLSD model /1/ is also a limiting form of the zero-truncated generalized negative binomial distribution of *Jain* and *Consul* [1971]. *Patel* [1981] also defined GLSD and obtained the estimates of the parameters by the method of moments. The model /1/ reduces to the simple logarithmic series distribution (LSD) when $\beta=1$. *Patil* [1962] studied the estimation of LSD. The GLSD model is a member of *Gupta* [1974] modified power series distribution and also can be found in Lagrangian probability distributions of *Consul* and *Shenton* [1972]. *Famoye* [1987] showed that the GLSD is unimodal and the mode is at the point $x=1$. Some methods of sampling from GLSD /1/ are provided by *Famoye* [1997]. *Mishra* [1979], and *Mishra–Tiwary* [1985] showed that the GLSD provides a very close fit to the observations coming from various fields such as medicine, engineering etc. *Tripathi-Gupta* [1988] gave the another generalization of the logarithmic series and geometric distributions.

Jani [1977] obtained the minimum variance estimators, *Famoye* [1995] acquired the moment estimators, *Mishra* [1979] and *Jani–Shah* [1979] discussed the use of maximum likelihood and moment method of estimation for the two parameter GLSD /1/. *Mishra–Tiwary* [1985] suggested an alternative method of estimation based on the first three moments. *Mishra–Hassan* [1996], [1997] recommended a quick and simple method for the estimation and they also obtained the Bayesian estimate of GLSD.

In this paper we study the estimation of the parameters of GLSD /1/ using the maximum likelihood (ML), minimum chi-square (MC), weighted discrepancy (WD) and empirical weighted rate of change (EWRC) methods in the same line as has been performed by *Famoye–Lee* [1992] in case of generalized Poisson distribution (GPD).

1. Maximum likelihood method

Let a random sample of size N be taken from the GLSD /1/ and let the observed frequencies be f_x; $x = 1, 2 \dots k$ so that $\sum_{x=1}^{k} f_x = N$, where k is the largest of the observed values having non–zero frequencies. The likelihood equation of the GLSD /1/ can be written as

$$L = \frac{\theta^N \alpha^{N\bar{x}} \prod_{\chi=1}^{K} \prod_{j=1}^{x-1} (\beta x - j)^{f_x} (1-\alpha)^{(\beta-1)N\bar{x}}}{\prod_{i=1}^{K} (X_i !)^{f_x}} . \qquad /2/$$

The log likelihood function is given as

$$\begin{aligned} \log L &= N \log\theta + N \bar{x} \log\alpha + \sum_{x=1}^{k} \sum_{j=1}^{x-1} f_x \log(\beta x - j) + \\ &+ (\beta - 1) N \bar{x} \log(1-\alpha) - \sum f_i \log X_2 . \end{aligned} \qquad /3/$$

The two likelihood equations can be obtained as

$$\frac{\partial \log L}{\partial \alpha} = -\frac{N\theta}{(1-\alpha)} + \frac{N\bar{x}}{\alpha} - \frac{(\beta-1)N\bar{x}}{(1-\alpha)} = 0 , \qquad /4/$$

$$\frac{\partial \log \mathrm{L}}{\partial \beta} = \frac{-N\bar{x}}{\theta} + \sum_{x=1}^{k} \sum_{j=1}^{x-1} \frac{x f_x}{\beta x - j} = 0 , \qquad /5/$$

where $\bar{x}$ is the sample mean. From equation /4/, we get

$$\beta = \frac{1}{\alpha} - \frac{\theta}{\bar{x}} . \qquad /6/$$

Putting this in equation /5/, we get

$$\varphi(\alpha) = -\frac{N\bar{x}}{\theta} + \sum_{x=1}^{k} \sum_{j=1}^{x-1} \frac{x f_x}{\left(\frac{1}{\alpha} - \frac{\theta}{\bar{x}}\right) x - j} = 0 . \qquad /7/$$

The previous equation does not seem to be directly solvable and hence some iteration method can be used to solve it. For this we find second derivatives of log L as

$$\frac{\partial^2 \log L}{\partial \alpha^2} = -\frac{N\bar{x}}{\alpha^2} - \frac{(\beta-1)\ N\bar{x}}{(1-\alpha)^2} - \frac{N\ \theta}{(1-\alpha)^2} = \frac{N\bar{x}}{\alpha^2} - \frac{N\left[\theta + (\beta-1)\bar{x}\right]}{\left(1-\alpha^2\right)}, \qquad /8/$$

$$\frac{\partial^2 \log L}{\partial \beta^2} = -\sum_{x-1}^{k} \sum_{j=1}^{x-1} \frac{x^2 f_x}{(\beta x - j)^2}, \qquad /9/$$

$$\frac{\partial^2 \log L}{\partial \beta\ \partial \alpha} = -\frac{N\ \bar{x}}{1-\alpha}, \qquad /10/$$

$$\frac{\partial^2 \log L}{\partial \alpha\ \partial \beta} = -\frac{N\ \bar{x}}{1-\alpha}. \qquad /11/$$

The values of these second derivatives can be put in the following equation in the matrix form as

$$\begin{bmatrix} \dfrac{\partial^2 \log L}{\partial \alpha^2} & \dfrac{\partial^2 \log L}{\partial \beta \partial \alpha} \\ \dfrac{\partial^2 \log L}{\partial \alpha \partial \beta} & \dfrac{\partial^2 \log L}{\partial \beta^2} \end{bmatrix}_{\alpha_0, \beta 0} \begin{bmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \dfrac{-\partial \log L}{\partial\ \alpha} \\ \dfrac{-\partial \log L}{\partial \beta} \end{bmatrix}_{\alpha_0, \beta 0}, \qquad /12/$$

where $\hat{\alpha}$, and $\hat{\beta}$ are the ML estimators of α and β respectively and α_0, β_0 are the initial values of the parameters. For initial values the moment estimators can be used or it can be obtained by equating the first three observed relative frequencies to the corresponding theoretical probabilities. The system of two equations may be used repeatedly till a good approximation of α and β are obtained.

2. Weighted discrepancies (WD) method

Let f_x denote the observed frequencies $x = 0, 1, 2, ... k$. Obviously, k is the largest of the observations. Let $N = \sum_{k=1}^{k} f_x$.

The corresponding relative frequencies are given by

$$n_x = \frac{f_x}{N}; \qquad x = 0,\ 1,\ \dots .\ k\ . \tag{/13/}$$

The log likelihood function can be written as

$$\log L = \sum_{x=1}^{k} N n_x \log p_x\left(\alpha,\beta\right) \ . \tag{/14/}$$

The likelihood equations are

$$\sum_{x=1}^{k} n_x \frac{\partial}{\partial\alpha}\log p_x = 0\,, \tag{/15/}$$

$$\sum_{x=1}^{k} n_x \frac{\partial}{\partial\beta}\log p_x = 0 \ . \tag{/16/}$$

Again as $\sum_{x=1}^{k} p_x = 1$, we have

$$\sum_{x=1}^{k} p_x \frac{\partial}{\partial\alpha}\log p_x = 0 \ \text{ and} \tag{/17/}$$

$$\sum_{x=1}^{k} p_x \frac{\partial}{\partial\beta}\log p_x = 0\ . \tag{/18/}$$

Subtracting /17/ from /15/ and /18/ from /16/, we get

$$\sum_{x=1}^{k}\left(n_x - p_x\right) \frac{\partial}{\partial\alpha}\log\ p_x = 0\,, \tag{/19/}$$

$$\sum_{x=1}^{k}\left(n_x - p_x\right) \frac{\partial}{\partial\beta}\log p_x = 0\ . \tag{/20/}$$

Substituting the corresponding expressions of the derivatives to /19/ and /20/, we get

$$\sum_{x=1}^{k}\left(n_x - p_x\right)\left[\frac{N}{\left(1-\alpha\right)\log\left(1-\alpha\right)} + \frac{N\overline{x}}{\alpha} - \frac{\left(\beta-1\right)N\overline{x}}{\left(1-\alpha\right)}\right] = 0 \ \text{ and} \tag{/21/}$$

$$\sum_{x=1}^{k}\left(n_x - p_x\right)\left[N\bar{x}\log\left(1-\alpha\right)+\sum_{x=1}^{k}\sum_{J=1}^{x-1}\frac{x\,f_x}{\left(\beta x - j\right)}\right]=0 \ , \qquad /22/$$

which has referred to *Kemp* [1986] as an equation from minimum discrimination information and ML estimation and called as weighted discrepancies estimation method.

3. Minimum chi-square (MC) method

We know that

$$\chi^2 = \sum_{X=1}^{K}\frac{\left(n_x - p_x\right)^2}{p_x} \qquad /23/$$

is approximately distributed as chi-square. Differentiating it with respect to α and β, we obtain

$$\sum_{x=1}^{k}\left(n_x - p_x\right)\left(1+\frac{n_x}{p_x}\right)\frac{\partial}{\partial\alpha}\log p_x = 0 \ , \qquad /24/$$

$$\sum_{x=1}^{k}\left(n_x - p_x\right)\left(1+\frac{n_x}{p_x}\right)\frac{\partial}{\partial\beta}\log p_x = 0 \ . \qquad /25/$$

Substituting the corresponding expressions of the derivatives to /24/ and /25/ we get

$$\sum_{x=1}^{k}\left(n_x - p_x\right)\left(1+\frac{n_x}{p_x}\right)\left[\frac{N}{\left(1-\alpha\right)\log\left(1-\alpha\right)}+\frac{N\bar{x}}{\alpha}-\frac{\left(\beta-1\right)N\bar{x}}{\left(1-\alpha\right)}\right]=0 \text{ and} \qquad /26/$$

$$\sum_{x=1}^{k}\left(n_x - p_x\right)\left(1+\frac{n_x}{p_x}\right)\left[N\bar{x}\log\left(1-\alpha\right)+\sum_{x=1}^{k}\sum_{j=1}^{x-1}\frac{x\,n_x}{\left(\beta x - j\right)}\right]=0 \ . \qquad /27/$$

By a similar argument in section 3 the resulting equations /26/ and /27/ are known as minimum chi-square equations.

The weights in equations /15/ and /16/ for ML method depend only on the observed frequencies while the weighted discrepancies method equations /21/ and /22/ including the minimum chi-square method equations /26/ and /27/, both have weights depending on the parameters as well as observed frequencies.

4. Empirical weight rates of change (EWRC) method

The expression

$$\frac{\partial}{\partial \theta_j} \log p_x \ ; \quad j = 1, 2 \ , \qquad /28/$$

where $\theta_1 = \alpha, \ \theta_2 = \beta$, is common to /15/, /16/ and /19/, /20/ and /26/, /27/ which are, the ML equations, the weighted discrepancies equations and minimum chi-square equations respectively. The common term /28/ can be seen as the relative rates of change in the probabilities on the parameters as α and β change. We refer to /28/ as the score function and it is being weighted by the relative frequencies in case of ML estimation method as in equations /15/ and /16/ and weighted by the discrepancy between observed relative frequency and estimated probability in case of WD estimation method as in equations /26/ and /27/. In order to obtain an estimation which is closer to the actual parameter value, it is quite natural to consider the combination of these two methods of estimation. Thus, we will use a weighting factor which is the product of the weights of ML and WD methods. This leads to equations

$$\sum_{x=1}^{k} n_x (n_x - p_x) \frac{\partial}{\partial \theta_j} \log p_x = 0, \qquad j = 1, 2 \ , \qquad /29/$$

where $\theta_1 = \alpha, \ \theta_2 = \beta$.

Estimators obtained from /29/ will be referred as empirical weighted rates of change estimators (EWRC). This method weights the scoring function by $n_x \left(n_x - p_x \right)$ which weights the discrepancy by factor n_x. If large discrepancies occur on the extreme *x* values, then small weights are applied. Meanwhile, if large discrepancies occur on the more frequent *x* values, large weights are applied. Therefore this method can be viewed as a generalization of WD method.

5. Fitting to the GLSD

In this section the method will be presented on a biological example. In our paper we have fitted the logarithmic series distribution (LSD) and GLSD /1/ to the same zero truncated biological data which was used by *Jani* and *Shah* [1979] though they only used the method of moment estimation. Here we have operated with the different methods of estimation like ML, WD, MC and EWRC to find the estimators for fitting the LSD and GLSD.

Data provided in Tables 1 and 2 are the zero-truncated data of *P. Garman (Jani–Shah* [1979]) on counts of the number of European red mites on apple leaves.

Table 1

Estimation of LSD model

Number of mites per leaf	Leaves observed	Expected frequency of LSD methods of estimation ML	 Moments	 MC	 WD	 EWRC
1	38	44.05	43.46	42.53	40.58	41.26
2	17	15.67	6.24	0.09	18.09	17.07
3	10	9.03	8.09	9.01	9.06	9.03
4	9	3.97	4.53	4.25	3.56	4.56
5	3	2.49	2.71	2.79	2.72	3.89
6	2	1.34	1.69	1.57	1.34	1.56
7	1	1.56	1.08	1.79	2.09	2.01
≥8	0	1.89	2.20	2.97	2.56	0.62
Total	80	80.00	80.00	80.00	80.00	80.00
Mean	2.1500	.	.	.	.	.
S.D	1.4504	.	.	.	.	.
χ^2	.	2.30	1.81	1.038	0.935	0.80508
D.f	.	2	2	2	2	2
$P(\chi^2)$	.	0.3166	0.40	0.595	0.6265	0.6686
	Estimates					
α		0.7578	0.7473	0.7398	0.7296	0.7216

Table 2

Estimation of GLSD model

Number of mites per leaf	Leaves observed	Expected frequency of LSD methods of estimation ML	 Moments	 MC	 WD	 EWRC
1	38	40.56	39.10	38.46	38.16	38.89
2	17	18.56	17.40	16.03	15.98	16.05
3	10	9.34	9.73	10.63	10.21	9.81
4	9	4.79	5.83	5.26	5.81	5.34
5	3	2.59	3.55	3.29	3.75	3.89
6	2	2.06	2.17	1.38	1.56	1.34

(Continued on the next page.)

(Continuation.)

Number of mites per leaf	Leaves observed	Expected frequency of LSD methods of estimation ML	Moments	MC	WD	EWRC
7	1	0.89	1.27	2.97	3.01	2.79
≥8	0	1.21	0.95	1.98	1.52	1.89
Total	80	80.00	80.00	80.00	80.00	80.00
Mean	2.1500	.	.	.	.	.
S.D	1.4504	.	.	.	.	.
χ^2	.	1.38	0.16	0.102	0.097	0.084
D.f	.	1	1	1	1	1
$P(\chi^2)$	.	0.2401	0.69	0.75	0.76	0.772
	Estimates					
α		0.8904	0.8898	0.8878	0.8823	0.8789
β		0.9526	0.9129	0.9103	0.91001	0.9001

It is evident from Tables 1 and 2 that in all the cases GLSD /1/ provides a better fit than the usual logarithmic series distribution. We notice that all the estimation techniques, the ML, the WD, the moment and the MC method do not perform well in comparison to the EWRC method in estimating the LSD and GLSD parameters. Also, we observed that EWRC method seems to be better than any of the other methods for fitting the data either to LSD or GLSD.

References

CONSUL, P. C. – SHENTON, L. R. [1972]: Use of Lagrange expansion for generating discrete generalized probability distributions. *SIAM Journal of Applied Mathematics.* Vol. 23. No. 2. p. 239–248.

FAMOYE, F. [1987]: A short note on generalized logarithmic series distribution. *Statistics and Probability Letters.* Vol. 5 p. 315–316.

FAMOYE, F. [1995]: On certain methods of estimation for the generalized logarithmic series distribution. *Journal of Applied Statistical Sciences.* Vol. 2. p. 103–117.

FAMOYE, F. [1997]: Sampling from the generalized logarithmic series distribution. *Computing.* Vol. 61. p. 365–375.

FAMOYE, F. – LEE, C. M. S. [1992]: Estimation of generalized Poison distribution. *Communications in Statistics–Simulation and Computations.* Vol. 21. No. 1. p. 173–188.

GUPTA, R. C. [1974] : Modified power series distribution and some of its applications. *Sankhya.* Series. B. 36. p. 288–298.

JANI, P. N. [1977]: Minimum variance unbiased estimation for some Left-Truncated Modified Power Series Distribtions. *Sankhya.* Series B. 39. p. 258–278.

JAIN, G. C. – CONSUL, P. C. [1971]: A generalized negative binomial distribution. *SIAM Journal of Applied Mathematics*. Vol. 21. No. 4. p. 501–513.

JAIN, G. C. – GUPTA, R. C. [1973]: A logarithmic type distribution. *Trabjos Estadist*. Vol. 24. p. 99–105.

JANI, P. N – SHAH, S. M. [1979]: On fitting of the generalized logarithmic series distribution. *Journal of the Indian Society for Agricultural Statistics.* Vol. 30. No. 3. p. 1–10.

KEMP, A. W.[1986]: Weighted discrepancies and maximum likelihood estimation for discrete distributions. *Communication in Statistics – Theory and Methods.* Vol. 15. No. 3. p. 783–803.

MISHRA, A (1979): *Generalization of some discrete distributions*. PhD thesis Patna University. Patna. Working paper.

MISHRA, A – TIWARY, D. [1985]: On generalized logarithmic series distribution. *Journal of the Indian Society for Agricultural Statistics.* Vol. 37. No. 3. p. 219–222.

MISHRA, A. – HASSAN, A. [1996]: On Bayesian estimation of generalized logarithmic series distribution. *Research Journal of Statistics*. *Assam Statistical Review*. Vol. 10. No. 2. p. 120–124.

MISHRA, A. – HASSAN, A. [1997]: A simple method of estimating generalized logarithmic series distribution. *Journal of Statistical Research*. Vol. 31. No. 2. p. 63–69.

PATEL, I. D. [1981]: A generalization of logarithmic series distribution. *Journal of Indian Statistical Association.* Vol. 19. No. 1. p. 29–132.

PATIL G. P. [1962]: Some methods of estimation for the logarithmic series distribution. *Biometrics.* Vol. 18. p. 68–75.

TRIPATHI, R. C. – GUPTA, R. C: [1988]: Another generalization of the logarithmic series and the geometric distributions. *Communication in Statistics – Theory and Methods.* Vol. 17. No. 5. p. 1541–1547.