HUNGARIAN STATISTICAL REVIEW

JOURNAL OF THE
HUNGARIAN CENTRAL
STATISTICAL OFFICE

VOLUME 90 2012. SPECIAL NUMBER 15

CONTENTS

ISSN 0039 0690

Published by the Hungarian Central Statistical Office
Responsible for publishing: dr. Gabriella Vukovich
Editor in Chief: dr. László Hunyadi
Printed by the Xerox Magyarország Kft.
2012.020 – Budapest, 2012

Managing Editor: Orsolya Dobokay-Szabó
Editor: dr. Cosette Kondora
Technical Editors: Ágnes Simon-Káli

Editorial Office: H-1024 Budapest II., Keleti Károly u. 5–7.
Mailing address: H-1525 Budapest, P.O. Box 51. Phone: +36-(1)345-6546,
Internet: www.ksh.hu/statszemle E-mail: statszemle@ksh.hu
Publishing Office: Central Statistical Office, H-1024 Budapest II., Keleti Károly u. 5–7.
Mailing address: H-1525 Budapest, P.O. Box 51. Phone: +36-(1)345-6000
The publication can be purchased at the Statistical Bookshop:
H-1024 Budapest II., Fényes Elek u. 14–18. Phone: +36-(1)345-6789

Inflation in Hungary After the Second World War*

Ádám Marton
Retired Head of Section
Hungarian Central Statistical Office
E-mail: Adam.Marton@ksh.hu

The currently used Hungarian currency (Forint) was introduced after the war in 1946. The paper presents the long-term developments of post-war inflation in Hungary. After the first hectic years, the market economy was replaced by the centralized „socialist" economy: the prices, parallel to a wage reform, were stabilized by 1952. The consumer market was also settled somehow at a very low level. Until 1968, due to strict central planning, consumer prices remained practically unchanged. After that the economic policy changed, the strict price control relaxed and the CPI slowly started to increase. The average price level in 1989 was 17 percent higher than in 1952. After 1990 the transition from the command to the market economy resulted in high inflation and a temporary decrease of the GDP, the real wages and the consumption.

KEYWORDS:
Inflation.
Economic development.
Price changes.

* Acknowledgments: The author wants to thank *László Hunyadi*, the Editor-in-Chief of the *Hungarian Statistical Review*, *Péter Bauer*, Economic Analyst of the Hungarian National Bank for many helpful contributions and suggestions which have greatly improved this article. The views expressed are those of the author and not necessarily reflect the policies of HCSO.

During the 20th century three legal currencies were used in Hungary: the Korona, the Pengő and the current Forint.

The first real Hungarian currency (Korona) was introduced in 1900. Earlier the official currency had been the silver-based Austrian Florin. On the 1st January 1900 the "gold" Korona (Crown) replaced the former currency. The exchange rate was: 1 Austrian Florin = 2 Hungarian Forints. During the First World War there was heavy inflation, which continued until the middle of the 1920's.

In June 1924 the Hungarian National Bank was established.

On the 1st October 1925 a new currency, the Pengő was introduced, the parity of which was: 1 Pengő = 12 500 Korona or 0.2632 g gold. It proved fairly stable till the beginning of the Second World War.

The inflation of the wartime currency Pengő started already during the war, in the middle of 1945 hyperinflation began and by the end of July 1946 the Pengő lost its value totally.

In January 1946 a "new" currency was introduced, called Tax Pengő.

The rate of exchange of the Tax Pengő:

1946 January 1st	1 Pengő
1946 April 1st	44 Pengő
1946 May 1st	630 Pengő
1946 June 1st	160 000 Pengő
1946 July 1st	7 500 000 000 ($7.5 \cdot 10^9$) Pengő
1946 July 31st	2 000 000 000 000 000 000 ($2 \cdot 10^{21}$) Pengő

(The Hungarian population practically lost all their money. Nevertheless this fact helped the stabilization process (*Grossman–Horváth* [2000]).)

In July 1946, due to the collapse of the monetary system, there were neither real prices nor real wages. The Pengő–Forint golden parity was the starting point of the new prices and wages.

At the end of 1945 a new coalition government was formed and soon it became obvious that a complete currency reform was needed. After consultations among the coalition parties and foreign experts the decision was made: the new Forint (Ft) currency to be introduced on the 1st August, 1946. At the same time there were some signs of slow recovery of the economy as well.

Parities: 1 Ft = 0.0757575 g gold; 1 US Dollar = 11.74 Ft; 1 Pengő (1938) = 3.47 Ft.

The transition was carried out through the Tax Pengő. (1 Ft = 200 000 Tax Pengő.) The new currency was not convertible.

By the end of the Second World War approximately 40 percent of the national wealth was destroyed, according to the estimates of the Hungarian Central Statistical Office (HCSO). The national income of the year 1945/1946 reached only less than half of that of 1938/1939 (*Ausch* [1958]).

Up to the present the current legal tender (Forint) has also been inflated, but to a reasonable extent. For the Hungarians it means that the depreciation of their money has been a permanent experience.

The paper, after this short historical review, presents the developments of consumer prices after the war and also of some basic indicators of the economy (GDP, real wages, consumption, living standard in general).

You can observe four characteristic periods in the last half century:

1945–1952: reconstruction, stabilization;
1952–1968: strong central planning;
1968–1989: economic reforms;
1989 – transition from the command to the market economy.

1. The new Forint consumer prices

The new Forint consumer prices were calculated from the 1938 Pengő prices, with the aim that the average price level be 3.7 of the one in 1938.

Table 1

Price multipliers, 1946
(1938 = 100.0)

Commodity group	Price multiplier
Average	3.7
Of which	
Clothing	6.6
Heating and lighting	3.8
Industrial goods	6.0
Rents, services, transportation	1.25

But the individual commodity prices were set on a different basis. Certain social and economic considerations distorted the former price ratios: some goods were taxed, some subsidized, that is, approximately 10 percent of the total consumption was priced on the average, 30 percent was highly taxed and all the other goods and services were subsidized.

As far as the wages were concerned, the bases were those of 1938, but the structure, the relation among the different occupations and the education levels greatly differed from the former relations. The real wages of 'workers' were set at 50 percent, while those of the employees at 25 percent of the 1938 level. It was supposed that this way equilibrium between the supply and the demand side of the consumer market could be established.

2. Price changes between 1946 and 1952

The reconstruction, the reparation payments meant a heavy burden on the economy. Due to the policy of the Soviet block, western loans were refused. The „Three-Year Plan" was started on 1st August 1947, with the aim to revitalize the country. In 1947–48 a serious political change took place in Hungary, copying the Soviet-type government planning and control of prices, wages, etc. In a short period of time more than 90 percent of the economy, including the means of production were nationalized.

The demand for foods was great, while the prices were 60 percent higher than in 1946. The increase of the prices of industrial products was only 12–14 percent. On average the 1948 price level was by 35 percent higher than the average level of the 1946 "starting prices". At that time huge amounts were spent on industrialization, mainly on heavy industry. The production of light industry and the import of consumer goods were far behind the demand, the budget showed a massive deficit, a shortage on the consumer market appeared, so in the next two years the increase of prices continued.

Nevertheless the Three-Year Plan was successful. In 1948–49 production and wages increased and the national income reached the 1938 level. Practically there was no unemployment; in 1950 the national income was approximately 25 percent higher than in 1938 and the living standard also attained the 1938 level. On 1st January 1950 the first Five-Year Plan was started. The agricultural and industrial production increased, the "socialist" central planning system somehow functioned. The living standard slightly increased, but on the consumer market shortage was observed soon. During two years' time the consumer prices increased by 27 percent and temporarily food rationing was effected. The government realized that measures were needed to stabilize the economy again.

To achieve the balance between supply and demand, a new price and wage reform was unavoidable. It was implemented on 2nd December 1951 and food ration-

ing was withdrawn. Under the strict state control (socialist planning system!) all prices were set and regulated by the government, except the so-called "seasonal goods" (vegetables, fruits, potatoes and eggs).

The retail prices were formed arbitrarily, regardless of the production costs. Social considerations shaped the price policy. Between the retail and producer prices a "turnover tax" or subsidy was applied. The result was that the retail price structure was very different from the cost relations. (Industrial products were taxed, while services, housing, public transport were highly subsidized.) The general consumer price level on 2nd December 1951 increased by more than 40 percent. (The average consumer price index (CPI) in 1952 based on 1946 was approximately 260 percent.)

Wages were also raised, but only by about 20 percent. Real wages decreased substantially.

Table 2

Consumer price indices
(1949 = 100)

Commodity group	1946	1949	1950	1952
Foods		100	112	210
Clothing	77	100	105	168
Industrial goods	77	100	96	145
Heating, lighting	77	100	101	109
Services	91	100	103	114
Total	66	100	107	174

The consumer prices were biased from the production costs. To make demand and supply meet, the idea was that a very detailed planning can manage it. Income allocation changed too. The centralized income increased, the personal income decreased. The position of the rural population deteriorated, while in the case of employees, the impact of price growth was compensated to a certain extent by wage increase. (In the socialist planning system the wages were also "planned", controlled by the government with the aim to "improve the living standard".[1])

On the role of prices in a planned economy *Béla Csikós-Nagy* [1958] wrote: "The role of prices in the topical economic system is very limited. In a socialist economy they serve only as an accounting unit for statistical and planning purposes." As far as the relation between producer and consumer prices is concerned, he says: "the producer and consumer prices are separated in order to achieve the different purposes of the producer and consumer price policy."

[1] In this paper the change of "living standard" means the change of real incomes.

The prices of the imported consumer goods were proportionally adjusted to the domestic price level. The foreign currency price was converted to Hungarian currency at the official exchange rate (1 USD = 11.74 Ft; 1 RUB =13.04 Ft; 1 USD = 0.9 RUB) and then a special "tax" served as a link between the calculated "price" and the much higher domestic price. (Since the official exchange rate was far from the purchasing power parity, in the tourist industry a special rate was applied. At that time it was 2.5 times the value of the "Devizaforint", that is, 30 Ft/USD.)

During these years there were several signs of economic recovery, the living standard was rising but the stabilization measures of December 1951 resulted in a necessary decline in real wages.

Table 3

Volume indices of consumption, accumulation and national income
(1949 = 100.0)

Year	Consumption	Accumulation	National income
1949	100.0	100.0	100.0
1950	110.9	134.3	120.6
1951	118.5	228.7	141.2
1952	127.7	187.4	138.5

Table 4

The indices of real income per capita
(1949 = 100.0)

Year	Wage earners	Rural population	Total
1949	100.0	100.0	100.0
1950	101.8	112.3	107.0
1951	90.8	116.8	103.8
1952	83.4	100.3	91.8

3. Change of consumer prices between 1952 and 1968

Between 1952 and 1968 consumer prices – except the prices of seasonal products – were fairly stable. Price changes were affected at longer intervals only, with no considerable effect on either the average price level or basic price relations. Up to

1957 official consumer price measures appeared exclusively in the form of price reductions.

In the first three years of the fifties, due to the arbitrary, faulty economic policy, living conditions deteriorated, the increase of national income stopped, by 1953 real wages decreased by almost 20 percent. In the middle of 1953 the new government changed the economic policy and relaxed the political suppression. In consequence, living conditions started to improve and by 1954 reached the 1950 level. Nevertheless, the political unrest led to the 1956 revolution. After that, in 1957, wages increased by more than 10 percent while the CPI increased only by less than 3 percent. Considering the whole decade, the prices slightly declined and the standard of living improved.[2]

In the fifties price indices were calculated on the basis of official price measures. It must be mentioned that at the same time so-called hidden price changes took place. The quality of a large number of consumer goods was changed without proper price adjustment, or prices of new goods were determined in no proportion to the already existing ones. It was not feasible to measure these "price changes", but according to certain estimates they moderately increased the general price level.

In the first half of the sixties there was again relative price stability. In 1966 there were comprehensive official price changes pushing up the price level by 1.2 percent. The aim was to adjust the consumer price relations: in the case of some food products and fuels, the high subsidy was reduced.

In 1966 preparation was going on to introduce the new economic management system. To a certain extent, price changes were carried out in accordance with this preparation.

Table 5

Consumer price indices by commodity groups

Year	Food	Alcoholic beverages, tobacco	Clothing	Electricity, gas and other fuels	Consumer durables	Other goods, including fuels and lubricants	Services	Total
	1952=100.0							
1952	100.0	100.0	100.0		100.0		100.0	100.0
1953	99.5	99.3	112.6		97.6		100.8	99.6
1954	95.2	95.7	116.6		94.7		100.6	94.7
1955	92.6	96.2	116.6		94.7		100.7	93.9
1956	92.5	92.3	116.7		93.2		100.7	93.0

(Continued on the next page.)

[2] In 1960 the real wages were nearly 50 percent higher than in 1955.

(Continuation.)

Year	Food	Alcoholic beverages, tobacco	Clothing	Electricity, gas and other fuels	Consumer durables	Other goods, including fuels and lubricants	Services	Total
1957	95.4	94.2	117.4		92.2		101.4	95.2
1958	95.2	95.2	117.7		94.4		102.8	95.5
1959	93.1	94.1	115.5		94.4		104.8	94.3
1960	94.3	93.6	114.7		94.4		105.5	94.9
	1960=100.0							
1961	100.9	102.9	100.0	99.8	99.8	100	102.0	100.9
1962	100.4	108.7	100.0	99.2	99.7	99.9	102.1	101.4
1963	99.7	108.7	99.5	98.8	98.4	98.8	102.0	100.8
1963	99.7	108.7	99.5	98.8	98.4	98.8	102	100.8
1965	102.9	110.9	98.3	97.2	97.3	98.9	102.6	101.9
1966	107.7	113.1	94.4	104.5	97.3	97.4	103.7	103.1
1967	108.5	114.1	93.8	106.9	97.3	97.3	104.9	103.5
1968	108.0	115.8	93.4	105.7	95.9	93.9	108.8	103.2
1968/1952	104.7	87.4				90.0	114.6	98.0

Special attention was paid to the seasonal goods which are very important for everyday life. The consumers were faced with their sometimes unpleasant price increases when all other prices were stately controlled.

Table 6

Price indices of seasonal goods
(1952=100)

Year	Live poultry	Eggs	Potatoes	Vegetables	Fruits	Total
1952	100	100	100	100	100	100
1956	104	112	89	88	91	95
1960	106	111	106	98	111	103
1965	106	120	162	130	136	131
1968	117	121	180	160	127	138

The planned management and control of the economy ensured economic development in the 1950's and at the same time a considerable rise in the standard of living. However, the planning did not function faultlessly, which appeared in the inefficiency of productivity and also in the functioning of the consumer market. An ad-

ministrative solution to this problem had been searched for in the fifties already, without positive results. A non-administrative answer was sought in the so-called great price debate of the early sixties, which finally led to a remarkable economic reform from the beginning of 1968.

4. Price movements and economic policy after 1968

In the New Economic Mechanism (NEM), introduced on 1st January 1968, the consumer price system substantially differed from the former rigid official price system.

According to the principles of the economic reform (*HSWP* [1966]), producer and consumer prices should have provided information for a rational basis of producer and consumer decisions. The prices were to improve the correspondence between supply and demand, promote technical progress and develop efficient consumption patterns, at the same time ensuring a relatively constant consumer price level even for the future.[3] Practically this means:

> – the consumer price structure should follow the structure of corresponding production costs, except certain luxury and agricultural products;
>
> – consumer prices should be flexible and the official price control should be very limited. The share of fixed and maximum priced products and services is less than half of the total consumption;
>
> – the changes of consumer prices should correspond to the steady increase in the living standard in such a way that as a result no population group should suffer a price increase.[4]

In addition there were specific methodological instructions for accounting principles to be used, calculating the prices of goods belonging to the non-fixed price category.

What has been said formerly can be put into four price categories: I. fixed, II. maximized, III. limited, IV. free prices. The Price Office ranged the goods and services into these groups. In 1969 the relevant allocation of goods was the following.

[3] If the CPI does not exceeds yearly 2 percent then we can talk about relative price stability. See also the Maastrich criteria.

[4] During the 40 years of command economy, it was not advised to speak about inflation, only about price corrections.

Table 7

The distribution of household consumption according to price categories in 1969
(percent)

Commodity group	Fixed	Maximum	Limited	Free	*Total*
	price category				
Foodstuffs	30	35	22	13	*100*
Clothing	–	6	55	39	*100*
Technical equipment, cultural articles	–	55	16	29	*100*
Fuels		100	–	–	*100*
Chemical and mineral oil product	20	44	–	36	*100*
Products of wood and paper industry	–	25	39	36	*100*
Construction materials	11	55	–	34	*100*
Glassware and china	–	15	–	85	*100*
Retail trade total	17	34	25	24	*100*
Services	17	21	22	40	*100*
Total consumption	16	31	24	29	*100*

In the new, fairly free price system, still remained some space for the government for price regulation. As regards limited prices, the Price Office first defined a basic price and retailers were allowed to vary price ±30 percent. The indirect state price control had also some other ways to influence the prices, for example, through indirect taxation. The crucial thing was that enterprises in this less rigid system had certain freedom of action.

The role of the turnover tax changed. Historically it had to ensure the desired preference between consumer and producer prices. From the late fifties the amount of individual tax rates was gradually decreased. Up to 1958 the turnover tax was 35-40 percent. In 1959 it was reduced to 13.5 percent, and by 1972 it decreased further by 2-3 percent. Divergences between producer and consumer price structures can be indicated by the turnover tax rates of the main commodity groups in 1969 (*Vincze* [1971]). (See Table 8.)

The Hungarian price system was unique in Europe. The tax burden was moved from the consumption side to the production sphere, which was unfavourable mainly for the foreign trade and the exchange rate policy. The long-term aim was to achieve a European-type (two-level) price system.

The most important target of price policy was to avoid inflation, but at the same time to be flexible enough to let change price relations among the main commodity groups. The price policy considered the impact of the change of productivity, of the world market prices. It was critically observed how the accidental price increases in-

fluenced the consumer prices and also the standard of living. In general it was supposed that the Price Office will be able to keep the annual CPI below 3.5 percent.

Table 8

Turnover tax rates of main commodity groups in 1969
(percent)

Commodity group	Turnover tax rate
Basic food	–25.3
Other unprocessed food	+17.6
Miscellaneous processed food products	+48.0
Food total	+9.5
Catering	–4.0
Clothing	+26.6
Miscellaneous goods other than food	+13.6
Products sold in retail trade, total	+12.9
Services purchased, total	–6.4
Total	+6.5
Total (with other allowances)	+3.5

5. Price changes between 1968 and 1989[5]

In 1968, in the first year of the NEM, the CPI did not change. Then, up to 1975, the yearly average increase was almost 2.8 percent. Service prices were steadily increasing. In the second half of the decade inflation accelerated up to 9 percent. In the eighties the annual CPI was about 8 percent, but in the last two years it was more than 15 percent. (See Table 9.)

The scope of the direct price regulation decreased, while the prices increased faster than earlier; it was an interesting question what was in the background: official and/or market induced effects. Estimates indicated that approximately two-thirds of the price increase was due to official decisions and one third to the "market" impetus, including seasonal goods in the 1970's. (Note: Retail firms were mainly in state ownership!) In 1986 the Yearly Economic Plan put great emphasis on reducing the

[5] The new market-type price policy needed direct representative price data collection. On the 1st January 1968 the HCSO collected consumer price statistics: 2 300 representative items were monthly observed by data collectors at least in 50 outlets. From 1968 onwards Laspeyres-type price indices have been published.

inflation. The result was 5.3 percent. But in the following three years the CPI accelerated to far above 10 percent again.

Table 9

Consumer price indices
(Previous year = 100.0)

Year	Food	Alcoholic beverages, tobacco	Clothing	Consumer durables	Energy	Other industrial products, fuels	Services	Total
1967	100.7	100.9	99.4	100.0	102.3	99.9	101.2	100.4
1968	99.5	101.5	99.6	98.6	98.9	96.5	103.7	99.7
1969	100.8	99.8	103.0	101.9	95.9	103.0	102.9	101.4
1970	100.9	100.5	102.3	99.9	98.2	103.3	101.9	101.3
1971	102.0	101.3	102.4	101.2	99.1	101.7	104.5	102.0
1972	101.1	107.0	104.0	102.4	97.9	102.3	103.5	102.9
1973	104.7	107.7	101.8	101.8	98.0	100.8	102.2	103.3
1974	100.5	102.3	102.0	102.4	105.6	102.1	101.8	101.8
1975	101.2	103.6	104.7	104.6	107.8	109.0	102.2	103.8
1976	110.2	100.5	105.2	102.9	99.1	106.7	102.4	105.0
1977	105.5	105.0	104.5	102.4	99.3	102.2	103.2	103.9
1978	103.6	109.2	105.0	103.1	100.3	103.3	103.9	104.6
1979	110.2	112.4	109.0	108.3	108.8	108.0	103.4	108.9
1980	113.4	101.7	105.0	113.2	121.4	109.5	107.0	109.1
1981	103.4	102.6	106.0	101.1	99.8	109.0	106.5	104.6
1982	104.8	113.1	105.5	102.1	110.0	106.4	107.2	106.9
1983	105.1	106.7	106.3	107.8	104.9	109.3	110.1	107.3
1984	112.1	105.0	111.0	105.5	104.9	107.3	108.0	108.3
1985	106.3	101.7	110.9	105.3	120.9	105.9	109.3	107.0
1986	102.0	105.2	109.4	106.1	103.5	105.0	108.9	105.3
1987	109.2	113.5	109.7	102.3	106.5	106.2	109.0	108.6
1988	115.8	114.3	120.0	108.5	112.8	116.3	117.5	115.5
1989	117.7	111.1	118.2	117.6	111.4	122.4	116.6	117.0

It is worth mentioning that price ratios among commodity groups did not change substantially up to 1989, but surprisingly clothing became more expensive and energy became cheaper. The tendency of durables already fit the long term tendency. (See Table 10.)

Table 11 gives an overview of the 40-year development of the consumer prices.[6]

[6] During 1946–1952 the general price level increased approximately 2.6-fold.

Table 10

Consumer price indices, 1968–1989[7]
(1968 = 100)

Year	Food	Alcoholic beverages., tobacco	Clothing	Consumer durables	Energy	Other industrial products, fuels	Services	Total
1968	100	100	100	100	100	100	100	100
1980	168	163	161	153	133	165	146	160
1989	346	328	401	262	269	378	353	343
Price ratio:	1.01	0.96	1.20	0.76	0.78	1.10	1.03	1.00
Weights (percent)	27	12	8	10	5	17	20	100

Table 11

Consumer price indices by main commodity groups

Year	Food	Alcoholic beverages., tobacco	Clothing	Consumer durables	Energy	Other industrial products, fuels	Services	Total
1938	14.5		10.8	20.4	13.3		40.6	16.6
1950	100.0		100.0	100.0	100.0		100.0	100.0
1952	195.1		156.8	108.9	143.2		111.9	169.4
	1952 = 100							
1952	100.0		100.0	100.0	100.0		100.0	100.0
1955	92.6		96.2	113.9	94.7		100.7	93.9
1960	94.3		93.4	114.7	94.4		105.5	94.9
1968	104.7		87.4	121.2	90.0		114.6	98.0
	1968 = 100							
1968	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1965	102.9	110.9	98.3	97.2	97.3	98.9	102.6	101.9
1970	109.9	116.2	98.4	99.6	97.6	99.9	114.1	106.0
1975	120.6	143.7	113.9	107.8	110.4	116.5	131.2	121.4
1980	181.4	189.2	150.5	140.6	147.0	155.2	159.3	164.6
1985	246.2	250.1	220.2	205.3	181.6	223.6	236.4	228.9
1989	373.7	379.3	374.7	284.3	251.5	354.8	384.2	353.9
1989/1952			350.7	326.1			405.3	335.9

[7] It is beyond the aim of this paper to analyze the causes of the change of the relative price structure. The yearly plans were responsible to regulate the production to meet the demand and supply on the consumer market (except the seasonal agricultural products).

In Hungary the foreign trade sector is relatively large. It is obvious that the change of world market prices influenced the domestic prices. In the socialist era there were many efforts to reduce this pressure, trying to make an independent (arbitrary) domestic price policy. After 1989 the changes of world market prices and the exchange rate of the Hungarian currency directly influenced the domestic prices.

After the devastating war, by 1949–50, the living standard and the national income reached the pre-war level. In the socialist economic planning system the main official economic goal was to increase the living standard. However, in the fifties, shortage prevailed practically in every sector of the economy. The emphasis was on increasing the volume of supply. The yearly plans managed the production, prices, consumption and there was some development in the economy. During the forty years of the socialist regime from the very low basis the economy developed substantially.

Table 12

Main indicators of the economic development, per capita
(1950 = 100.0)

Year	GDP	Real wages	Real incomes	Private consumption
1950	100.0	100	100	100
1960	177.0	154	154	115
1970	297.9	199	245	228
1980	483.3	243	333	314
1987	557.0	237	376	363
1989	560.3	227	385	358

In 1989 the GDP was approximately 460 percent higher than in 1950. During the same time real incomes increased to 385 percent and private consumption to 358 percent. There is no space here to analyse the details, but it must be mentioned, that the emphasis was on quantity and not on the good quality or a wide selection of products. Due to the lack of foreign exchange convertibility, the supply of imported consumer goods was also very poor.

Table 13 shows the structural change of consumption. The most striking data is that the volume of consumer durables increased 25-fold. Also the volume of energy and services increased high above the average, while beverages, tobacco only a little above that. This structural change was a good sign of the rising living standard.

Table 13

Volume indices of the per capita private consumption (estimates)
(1950 = 100)

Year	Foods	Beverages, tobacco	Clothing	Energy	Consumer durables	Services	Total
1938	97	99	104	99	115	78	93
1950	100	100	100	100	100	100	100
1960	134	158	161	138	369	155	152
1970	168	274	222	243	1008	235	228
1980	189	403	238	433	1532	376	314
1989	208	405	170	570	2544	478	358

6. Inflation, economic development after 1989

Two or three years before 1989 the former economic development slowed down and the yearly inflation increased to 15–17 percent.

By the end of the eighties economic difficulties got aggravated, social tensions increased and soon "the socialist planned economy" collapsed (parallel to the fall of the Soviet system). The transition from the command to the market economy started in 1989. The "socialist market" ceased to exist. The production of many uncompetitive, technically not up-to-date products stopped, the relatively huge foreign trade sector underwent a substantial structural change, the emphasis moved from the rouble to the dollar area. The privatization of the socialist agricultural and industrial sectors started. All these changes were carried out very quickly.

After 1989 price policy and the price system changed rapidly and substantially. Government and/or municipality price control did not cease to exist, but became limited (it covers approximately 20 percent of the total consumption). The introduction of the limited convertibility in 1996, then the full convertibility in 2001 was a great step toward capitalist market economy. The consumer prices (with few exceptions) depended on the producer and/or world market prices, taxes and import duties and on market competition. (In the case of agricultural products the weather influenced the prices as well.)

The rate of inflation accelerated. In the 1990's the yearly rate was above 20 percent. The CPI in 1991 was 135 percent, then till 1996 the yearly average was above 20 percent, until 1998 it was above 10 percent and only in the 2000's decreased below 10 percent.

In 2004 Hungary joined the European Union. A well-known fact is that the so-called Maastricht Criteria are very important part of the Convergence Program,

among others, to reduce the yearly inflation rate to 2-3 percent. Hungary is far from this condition.

Table 14

Consumer price indices 1989–2010
(percent, previous year = 100.0)

Year	Food	Beverages, tobacco	Clothing	Consumer durables	Energy	Other industrial goods, fuels	Services	Total
1990	135.2	130.7	123.3	120.8	127.6	128.9	125.6	128.9
1991	121.9	125.1	132.1	131.7	181.0	143.4	141.9	135.0
1992	119.4	119.6	123.0	114.3	143.2	127.2	126.0	123.0
1993	129.2	118.6	116.7	111.0	120.3	121.6	124.1	122.5
1994	123.4	116.4	116.1	111.8	111.7	119.0	120.3	118.8
1995	131.1	120.1	120.2	124.0	150.0	127.3	126.0	128.2
1996	117.3	126.6	125.6	119.2	132.5	125.7	126.4	123.6
1997	117.5	118.9	118.7	108.5	129.9	116.1	119.2	118.3
1998	114.4	115.3	114.1	108.1	117.9	110.7	116.2	114.3
1999	102.9	111.5	110.6	106.6	109.4	114.7	114.8	110.0
2000	109.2	111.0	105.8	101.7	109.1	115.0	109.7	109.8
2001	113.8	111.2	105.3	101.0	110.3	104.9	109.8	109.2
2002	105.4	109.7	104.0	98.4	105.5	104.1	106.4	105.3
2003	102.7	110.7	103.0	98.6	107.3	103.9	105.9	104.7
2004	106.5	111.3	103.4	99.4	114.1	103.9	107.6	106.8
2005	102.5	103.3	100.2	97.7	106.2	104.5	105.5	103.6
2006	107.7	104.3	99.3	96.0	106.4	102.7	104.1	103.9
2007	111.5	106.7	101.0	98.7	124.6	104.0	107.4	108.0
2008	110.2	105.6	100.0	99.3	112.7	104.6	105.0	106.1
2009	104.4	107.5	100.5	102.6	108.2	101.1	104.6	104.2
2010	103.2	108.2	99.6	100.2	106.3	108.8	104.3	104.9

During the 1990's the price increase was around 8-fold, while in the next decade it was only 70 percent. As Table 15 shows there was a great difference in the price dynamics of the groups of commodities.

The change of the CPI depends on the relative change of the weights and/or price indices of the individual goods. In practical terms this means that the current, perceived depreciation depends on the current market basket. That's why CPIs are calculated, as it will be seen in the following, with different commodity structures. You may put this idea in another way: what we want to know is how the price of foodstuffs, consumer durables or the energy changes, how much the CPI without tax-changes or calculated on the pensioner's basket comes to, etc.

One of the most important indicators is the "core" or "underlying" inflation rate (*Eckstein* [2008]). This is the so-called "except food and energy" approach to the CPI. The newer approaches exclude other items, for example tax changes, as well.

Table 15

Consumer price indices, 1989–2010
(percent, 1989=100.0)

Year	Food	Beverages, tobacco	Clothing	Consumer durables	Energy	Other industrial goods, fuels	Services	Total
1989	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1990	135.2	130.7	123.3	120.8	127.6	128.9	125.6	128.9
1991	164.8	163.5	162.9	159.1	231.0	184.8	178.2	174.0
1992	196.8	195.5	200.4	181.9	330.8	235.1	224.5	214.0
1993	254.3	231.9	233.9	201.9	398.0	285.9	278.6	262.2
1994	313.8	269.9	271.6	225.7	444.6	340.2	335.2	311.5
1995	411.4	324.1	326.5	279.9	666.9	433.1	422.4	399.3
1996	482.6	410.3	410.1	333.6	883.6	544.4	533.9	493.5
1997	567.1	487.8	486.8	362.0	1147.8	632.0	636.4	583.8
1998	648.8	562.4	555.4	391.3	1353.3	699.6	739.5	667.3
1999	667.6	627.1	614.3	417.1	1480.5	802.4	848.9	734.0
2000	729.0	696.1	649.9	424.2	1615.2	922.8	931.2	805.9
Ratio*	**0.90**	**0.86**	**0.81**	**0.53**	**1.99**	**1.14**	**1.15**	**1.00**
2001	829.6	774.1	684.3	428.4	1781.6	968.0	1022.5	880.0
2002	874.4	849.2	711.7	421.5	1879.6	1007.7	1087.9	926.6
2003	897.9	940.4	732.4	415.7	2016.1	1046.9	1152.6	970.2
2004	956.3	1046.7	757.3	413.2	2300.4	1087.7	1240.2	1036.2
2005	980.2	1081.2	758.8	403.7	2443.0	1136.6	1308.4	1073.5
2006	1055.7	1127.7	753.5	387.5	2599.4	1167.3	1362.1	1115.4
2007	1177.1	1203.3	761.0	382.5	3238.8	1214.0	1462.8	1204.6
2008	1297.1	1270.7	761.0	379.8	3650.2	1269.9	1536.0	1278.1
2009	1354.2	1367.2	764.8	390.5	3949.5	1283.8	1606.6	1331.8
2010	1397.5	1479.3	761.7	391.3	4198.3	1396.8	1675.7	1397.1
Ratio*	**1.00**	**1.06**	**0.55**	**0.28**	**3.01**	**1.00**	**1.20**	**1.00**
*Weights***	*23.2*	*9.5*	*4.4*	*6.9*	*9.9*	*17.2*	*28.9*	*100.0*

* 2010 commodity group indices divided by the total CPI.
** 2010 weights (percent).

The HCSO have calculated and published core inflation index since 1995. (See Table 16.)

Table 16

Consumer price index and core inflation
(percent, 1995 = 100)

Indicator	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
CPI*	124	146	167	184	202	220	232	243	260	269	279	301	319	332	348
Core inflation	123	144	163	178	193	218	223	234	248	253	259	273	287	299	308

* From CPI excluded fresh foods, household energy and fuels, services with stately controlled prices.

The core inflation in 2010 (308 percent) was a little bit lower than the CPI, due mainly to tax increases (*Mináry* [2010]).

The pensioners' market basket differs from the average family budget structure. It is justified to calculate separate indices to know how they are affected by the inflation. The prices of every day necessities increase generally above the average. But as Table 17 shows during the last decade there was no remarkable difference between the two inflation rates.

Table 17

Price indices of the pensioners and households, 2001–2010
(Previous year = 100.0)

Commodity group	Market basket	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total	Households	105.3	104.7	106.8	103.6	103.9	108	106.1	104.2	104.9
	Pensioners	105.3	104.6	107.3	103.9	105.0	110.7	106.9	104.9	104.5
Food	Households	105.4	102.7	106.5	102.5	107.7	111.5	110.2	104.4	103.2
	Pensioners	104.7	102.3	106.2	102.2	108.7	111.5	110.2	104.5	103.1
Beverages, tobacco	Households	109.7	110.7	111.3	103.3	104.3	106.7	105.6	107.5	108.2
	Pensioners	109.9	111.1	111.9	103.4	104.4	106.9	105.5	107.4	108
Clothing	Households	104.0	103.0	103.4	100.2	99.3	101.0	100.0	100.5	99.6
	Pensioners	104.0	102.9	103.5	99.9	99.2	100.8	99.8	100.5	99.4
Household durables	Households	98.4	98.6	99.4	97.7	96.0	98.7	99.3	102.6	100.2
	Pensioners	98.0	97.9	98.6	97.3	94.9	98.6	99.3	101.1	99.9
Energy	Households	105.5	107.3	114.1	106.2	106.4	124.6	112.7	108.2	106.3
	Pensioners	105.4	107.3	113.8	106.1	106.8	124.4	112.6	108.2	106.2
Other industrial goods, fuels	Households	104.1	103.9	103.9	104.5	102.7	104.0	104.6	101.1	108.8
	Pensioners	105.8	104.6	104.2	104.7	102.3	107.5	103.3	103.5	106.4
Services	Households	106.4	105.9	107.6	105.5	104.1	107.4	105.0	104.6	104.3
	Pensioners	106.2	105.9	107.8	105.9	104.0	108.5	104.3	104.6	104.0

After 1989 there was great restructuring in the economy, which was reflected in the main economic indicators. (See Table 18.) In the first four years the GDP declined by nearly 20 percent, the real wages by 15 percent, while private consumption by around 10 percent. The Hungarian economy reached the 1990 level again only by 2000. Then the increase continued and the GDP was 36 percent higher than in 1989 by the year 2008.

Unfortunately the worldwide recession has had a negative impact on the Hungarian economy as well.

Table 18

Economic indicators
(percent, 1949 = 100.0)

Year	GDP	Real wages	Private consumption
1989	100.0	100.0	100.0
1990	96.0	94.9	94.8
1991	84.9	90.0	86.1
1992	82.3	88.5	86.1
1993	81.7	85.0	87.6
1994	84.2	91.2	87.6
1995	85.5	80.2	81.7
1996	86.8	76.2	79.3*
1997	90.9	80.3	80.0
1998	95.3	83.0	82.6
1999	99.4	85.0	87.6
2000	104.6	86.4	91.2
2001	108.9	91.8	97.2
2002	113.4	104.1	107.2
2003	118.1	113.6	116.3
2004	123.8	112.9	119.9
2005	128.7	119.7	124.3
2006	134.0	123.8	127.1
2007	135.4	118.4	125.1
2008	136.3	119.0	126.3
2009	127.0	116.3	117.9
2010	128.8	119.0**	115.5**

* Change of methodology.
** Preliminary data.

7. Summary

The new currency (Forint) was introduced in 1946. At the same time "new" wages and prices were centrally set up. In 1948 there was a great socio-economic change: nationalization and almost total economic control and planning. The inflation between 1946 and 1952 was approximately 2.5-fold. In 1952 strict price control was introduced and real wages declined. After the 1956 revolution, prices remained under control and the living standard started to increase. The intention for a change of economic policy aroused which led to the introduction of NEM in 1968. In the following years the rigid price control and planning relaxed to some extent. The prices started to increase. The CPI in 1989 was 336 percent compared to 1952. In 1988 there was a large-scale tax reform.

In 1989 a great political and economic change (that is privatization) started. The GDP declined, the inflation accelerated. The recovery started in 1995–96 and then slow, but remarkable increase of the GDP and living standard could be observed up to 2008. During the last three years economic difficulties arose in Hungary as well.

References

AUSCH, S. [1958]: *Az 1945–1946. évi infláció és stabilizáció*. Kossuth Könyvkiadó. Budapest.

CSIKÓS-NAGY, B. [1958]: *Árpolitika az átmeneti gazdaságban*. Közgazdasági és Jogi Könyvkiadó. Budapest.

CSIKÓS-NAGY, B. [1974]: *Szocialista árelmélet és árpolitika*. Kossuth Könyvkiadó. Budapest.

ECKSTEIN, O. [1981]: *Core Inflation*. Prentice-Hall. Upple Saddle River.

GORDON, R. J. [1975]: Alternative Responses of Policy to External Supply Shocks. *Brookings Papers on Economic Activity*. Vol. 6. No. 1. pp. 183–206.

GROSSMAN, P. Z. – HORVÁTH, J. [2000]: The Dynamics of the Hungarian Hyperinflation, 1945–6: A New Perspective. *Journal of European Economic History*. Vol. 29. No. 2–3. pp. 405–427.

HCSO (HUNGARIAN CENTRAL STATISTICAL OFFICE) [1946– 2011]: Yearbooks. Budapest.

HSWP (HUNGARIAN SOCIALIST WORKING PARTY) [1966]: *Reform of Economic Mechanism*. Materials of the Session of the Central Committee of the Hungarian Socialist Working Party. 25–27 May.

JÁVORKA, E. [1967]: *A fogyasztói árak és dotációk*. Kossuth Könyvkiadó. Budapest.

MINÁRY, B. [2011]: Kormányzati intézkedések fogyasztói árszínvonalat befolyásoló szerepe és hatásának számszerűsítése. *Statisztikai Szemle*. Vol. 89. No. 7–8. pp. 743–767.

VINCZE, I. [1971]: *Árak, adók, támogatások a gazdaságirányítás reformja után*. Közgazdasági és Jogi Könyvkiadó. Budapest.

WYNNE, M. A. [2008]: Core Inflation: A Review of Some Conceptual Issues. *Federal Reserve Bank of St. Louis Review*. Vol. 90. No. 3. Part 2. pp. 205–228.

Preparations for the 2011 Population and Housing Census in Hungary and the Planned Implementation Method

Virág Erdei
Chief Counsellor
Hungarian Central Statistical Office
E-mail: Virag.Erdei@ksh.hu

This article summarizes all the preparatory work which was carried out at the Hungarian Central Statistical Office from 2007 to 2011, the census year. The reader can get a short overview of the census work from the start to the realization of the final execution.

KEYWORDS:
Census.
Survey.
Legislation.

The article presents the census preparatory work carried out at the Hungarian Central Statistical Office, from the starting ideas to the development of the census execution method. It contains a short description of the European and Hungarian census legislation and organization. As the paper was written before the census, it closes with a short summary on the realization.

1. The first steps of the census preparatory work

The preparatory work of Census 2011 started in 2007. What stood in the centre of our thinking is how to change from the traditional, interview-based census to a more modern method using partly or fully registers. At that time, a European Union project called "Preparation for the Population Census 2011" was running at the Population Statistics Department of the Hungarian Central Statistical Office (HCSO). It had two objectives:

– to discover which authorities in Hungary have such an administrative database that may contain census type information. Besides, as this information collection could have been the basis of a new type of Hungarian census using partly or fully registers, we wanted to find out where these data can be found, who the data owners are, how they can be used.

– to check the census addresses and block identifiers, that is, to correct the census address database (for example the related maps) of the Census Department of HCSO which serves as a basis for the census.

Simultaneously with this project, three census concepts were elaborated:

1. data are exclusively obtained from administrative registers, linking personal data for statistical purposes;

2. basic data are acquired from the Population and Address Register maintained by the Central Office for Administrative and Electronic Public Services and from the Address Register of HCSO, while supplementary data are collected from a representative sample survey;

3. data originates from full-scope enumeration.

The concepts indentified the way of implementation, (time) conditions as well as the material and human resources.

The first idea – relying exclusively on administrative registers – was soon rejected as it had no suitable legal background and there was not enough time for its implementation.

The second one – using basic data from the Population and Address Register – was a serious matter of consideration. Many analyses and comparative studies were conducted and their results were built in the methodology of the test censuses.

Besides these, the concept of the conventional census grounded on the latest technology (for example on using hand-held devices during the enumeration process or on the self-enumeration method (self-fulfilling paper or electronic questionnaires)) was also under deliberation.

2. Regulation of the European Parliament and of the Council on population and housing censuses

In recent years not only the preparations for Census 2011 have started in the countries but a substantial change has also occurred at European level. Around the millennium, the European censuses no longer met the quality criteria related to users' needs. There was a several-year difference among the reference dates of the censuses of various countries, and consistency errors were also found in data. As some countries forwarded their data to Eurostat with long delays, the European-level dissemination finished only in September 2005, 44 months after the end of the reference year. Instead of loose cooperation (gentlemen's agreement), a plan for stronger collaboration, that is, for creating a legally-binding European census regulation was outlined. The first draft of the European regulation on censuses was presented to the Directors of Social Statistics on 28 September 2005. Then intensive consultations started on the details with UNECE, the Eurostat and the representatives of national statistics, together with international experts. HCSO has also continuously participated in these consultations, followed and helped the process till its end. As a result of the multi-annual work, the Regulation of the European Parliament and of the Council on population and housing censuses (hereinafter referred to as the Regulation) was adopted on 9 July 2008.

Herewith the members of the European Union strengthened their cooperation related to censuses as the Regulation is compulsory for all member states. This legislative act stipulates that every country shall conduct a census in the same year, for the first time in 2011. It does not state the implementation method but declares the data to be collected. The member states have to meet both general and individual quality

requirements declared in the Regulation. The data collected compulsorily have to be provided to Eurostat within 27 months of the end of the reference year.

Three Commission regulations implementing the Regulation have also been adopted as regards the detailed rules (for example the technical way of data collection).

Even though the Regulation gives freedom to the member states to choose the implementation method, it also imposes a new obligation on them to prepare for the census.

3. Test censuses

Several test censuses and test processes were carried out in the years prior to the census in order to realise the planned, new methods. We conducted test censuses in 2008, 2009 and a pilot census in 2010. In doing so, we tested the usability of the Personal Data and Address Register which is the official state register and the citizens' reception of self-fulfilment of both paper and electronic questionnaires. We also analysed the main demographic features of those who chose these forms or the interview method and the time of answering via internet, regarding the whole enumeration period and the time of day. The advantages and disadvantages of the temporal separation or concurrence of the self-enumeration and conventional – enumeration with interviews – methods were also examined.

The questionnaire of the test censuses, the 'quality' of each question, the answering possibilities and the order of the questions were also tested. Before each test census, focus group researches were carried out for quality correction of the questionnaire during which a small number of participants were asked to fill it in and share their experience on how they understood the questions. The understanding of the questionnaire and the fact whether it allows an easy filling were analysed with this method.

While determining the census topics, special attention was paid that the compulsory data and their breakdowns provided for in the Regulation, could be collected and created.

Budget is always a fundamental point of censuses and this was so with us. It was the reason why we concentrated primarily on the compulsory information stipulated by EU legislation, followed by the domestic needs arising partly from the previous Hungarian censuses, partly from meeting the new demands of society.

Census topics have been negotiated with different organizations representing the needs of many professional and interest groups like the representatives of ethnic mi-

norities, disabled persons, churches and with the Commissioner for Data Protection and Freedom of Information, in order to meet all kinds of confidentiality rules.

4. Creation and content of the Hungarian act on population and housing census

The first version of the act was adopted in 2009 (Act CXXXIX of 2009 on the census of 2011, hereinafter referred to as Census Act), then it was modified two times in consequence of the change in the government which has taken place in the meantime. The final version was born on 30 November 2010 (Act CXXXIX of 2009 on the census of 2011 in a unified structure with Act XLIX of 2010 and Act CXXXVII of 2010 amending it). As a result of the modifications, the method of implementation was changed, the enumeration period was shortened, and the number of census topics was increased.

In accordance with the Census Act, a population and housing census shall be held in the territory of the Republic of Hungary, covering the natural persons living in the country and all dwellings, with the reference date of 1 October 2011. Data collection will take place between 1 and 31 October 2011. The supplementary enumeration of those persons who were omitted from the 'normal' enumeration for some reason has to be finished till 8 November 2011. Questionnaires can be answered by self-fulfilling a paper or an electronic questionnaire or by interview with an enumerator. Citizens can freely choose the way of answering.

The Census Act lists those data topics to be collected on natural persons, dwellings and institutions for which response is compulsory. Besides, there are questions on nationality, mother tongue, religion, long-lasting illness and disability, answering to which is not mandatory. The topics cover all variables prescribed in the Regulation, and, over and above, other subjects, for example, fertility and daily commute that are contained by the Census Act and were collected in the previous Hungarian censuses.

Respondents shall give true data on census topics. However, their names must not be recorded in the questionnaires. Dwelling addresses get an internal identification number which has to be handled separately from other collected census data after checking full coverage.

The Hungarian Central Statistical Office is in charge of the professional supervision of the (local) preparation and execution of the census; provides the technical conditions of the enumeration and performs data processing, dissemination and publication. The notary of the local government takes care of the local preparation of the

census and the execution of data collection in the administrative territory of the locality.

A separate government decree covers the detailed rules of the census (Government Decree 305/2010. (23 Dec) on the execution of Census 2011).

5. Features of the opportunities to respond

A major innovation of the Hungarian Census 2011 is that self-enumeration is provided as an option for inhabitants. This method was tested several times, and according to our experience, people do need self-enumeration through paper questionnaires or via internet, without enumerators, as some population groups are unwilling to answer to census takers. This option will be highlighted in communications to the population. Provision of data via internet is cost-effective for the statistical office because of data capturing and processing.

Before the reference date, the first task of the enumerators is to roam their enumeration areas in order to check the completeness of the list of addresses, to correct it, if necessary, and to distribute the respondent packages to every single address. Each respondent package contains an informative letter as well as a copy of a dwelling and a personal questionnaire.

Citizens may provide data in three ways:

– Questions can be answered via internet. For this action, respondents have to enter a website (www.enepszamlalas.hu) where they can start answering the questions after getting their unique internet access codes. These, namely the “Identifier” and the “Identification code”, can be found on the first page of the dwelling questionnaire included in the respondent package. In case of online fulfilment, all persons living in the same dwelling have to choose this answering method. Answering via internet is possible from 1 to 16 October 2011.

– Citizens may complete the paper questionnaires found in their respondent package on their own and then give it (being placed back in the envelope) to the enumerator at an agreed date. If more than one individual lives in the dwelling, the enumerator provides them the further copies of the personal questionnaire. Self-enumeration in this way is possible only in the first two weeks of data collection, from 1 to 16 October 2011, such as in the case of the internet method.

– Citizens can answer census questions with the help of an enumerator who visits and interviews them at an agreed date between 1 and 31 October 2011. It is necessary to keep the questionnaires found in the respondent package for this action.

6. The organization of implementation

The key to the success of the census is the implementation of data collection according to uniform principles throughout the country. The order of responsibility and the division of labour stipulated in the Census Act and the aforementioned government decree on the implementation of the census serve as a basis for the tasks of the actors taking part in it.

HCSO professionally manages the tasks concerning census preparatory work and directs the execution of the data collection. Its president has appointed a *county instructor* for each county and for the capital. The latter organizes the preparation for the local implementation of data collection, directs the territorial instructors' work in the given county or in the capital and maintains contact with the county administrative instructor of the census.

HCSO enforces the professional aspects through the territorial instructors' network. A *territorial instructor* represents the statistical office in the localities. He/she maintains strong professional relation with the *local administrative instructor of the census*. The tasks of a territorial instructor are the following: organizing the supervisors' network, supporting and controlling their job, providing professional supervision of the enumerators' work with the help of the supervisors and the local administrative instructor of the census.

Supervisors, who are at the lowest level of the managing and controlling hierarchy, directly manage, control and support the enumerators' work. Their task is to qualify and accept the fulfilled questionnaires and to record the information needed by the electronic monitoring system.

The leader of a government office (previously called public administration office) for the county or the capital appoints a *county administrative instructor of the census* for the period of the preparation and implementation of the data collection. He/she fulfils his/her task which is to check whether data collection is conducted according to the law, in close cooperation with the county instructor.

The notary of the locality or the district (in the capital) as the *local administrative instructor of the census* takes care of the local preparation and implementation of the data collection. In the case of larger settlements, a census officer and/or a census ad-

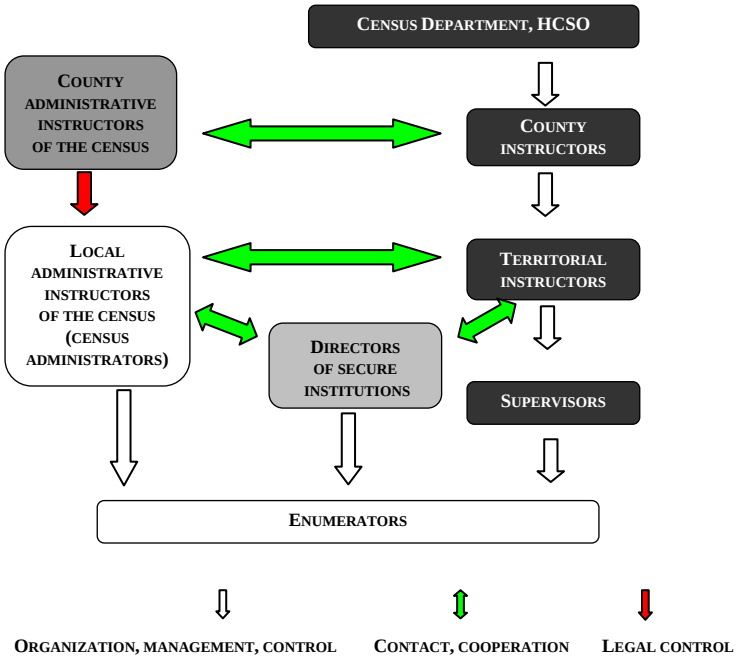
ministrator assists his/her work. The tasks of a local administrative instructor are to provide for the adequate number of enumerators, to ensure the technical conditions of implementation, to control the progress of data collection, to handle the data providers' problems and to organize supplementary data collection in the given locality, if necessary. The local administrative instructor of the census cooperates closely with the territorial instructor of HCSO.

The *enumerator* is in direct connection with the data providers. His/her task is to distribute respondent packages to the addresses, to conduct interviews, to collect paper questionnaires from those who completed them on their own and to correct the details of addresses in the address store. His/her work is directed and controlled by the supervisor.

The *director of a secure institution* is responsible for the implementation of the data collection process concerning persons living in that *institution* (maintained by the Hungarian Defence Forces or law enforcement agencies like police, national security services, etc.), in cooperation with the notary of the locality.

The relations of the actors of implementation are shown in the Figure.

The organization of enumeration



7. The dates, types and novelties of data dissemination

The users of census data including central and local administrative bodies, research institutions, individual researchers, interest groups and civil society organizations as well as the broad public rightly expect that the results of the census whose budget was significant should be accessible and usable as widely as possible, at the earliest opportunity, in a form and detail needed by them. When preparing the dissemination plan, our aim was to make such a user-friendly dissemination program which satisfies the formerly mentioned needs and also fulfils the reporting obligation of the country towards the European Union.

The most important change compared to the previous census is the reduction of paper publications and the better exploitation of the electronic dissemination potentials.

A stressed criterion of data dissemination is the earliest possible publication of the preliminary results. These data are planned to be released in March 2012. The final data and the relevant details will be published after completion of data processing, from the fourth quarter of 2012.

We will bring out the preliminary results, a national summary volume and its county versions in paper (printed) format as previously. Our aim is to reduce their content by combining print publications with CDs, decreasing the number of tables in the paper format but increasing it in the CD supplement.

Dissemination of the results will mainly rely on electronic publication forms. One part of them is static tables by territories and topics which can be used easily even by unprofessional users. Owing to the growing need for analytical studies, we plan to produce several downloadable publications for what the census database offers a great opportunity. We will continue the tradition of publishing our thematic volume set distributed previously in print but this time it will be disseminated in the form of downloadable internet publications.

The other part of the electronic dissemination channel is made up by databases that provide dynamic access for professionals and researchers having more interest in our data. A good example is the Dissemination Database of HCSO which would serve as a model for the Census Database.

There is always an increased interest in spatial data during census period. This is the reason why we plan to organize census data into regional/local databases besides the main census indexes of territorial (local, Budapest district, county, regional) level in line with the previous practice.

It is expedient to present the territorial data not only in the previously mentioned databases but also in interactive thematic maps such as the Atlas of Regions already available on the website of HCSO.

We are working on finding the appropriate way to make the census data accessible for researchers and hereby to provide certain micro data sets for scientific analysis.

The Regulation on population and housing censuses and the related implementing regulations define the dataset to be submitted to Eurostat, specifying exactly the topics and their breakdowns as well as the technical conditions, method and date of data submission. Data will be submitted in the so-called electronic hypercubes instead of Excel tables used in the previous European censuses. Sixty multidimensional hypercubes with defined content will be submitted by all member states of the EU. Among them, five hypercubes refer to the country level, 40 to the region level, 10 to the county level and 5 to the locality level.

The Eurostat recommends the Census Hub for data submission. This data dissemination and submission system is based on the concept of data sharing. Its main point is that the national statistical offices use their own IT system for storing their data to be supplied, which serves as a basis for query answering with the help of a uniform data query system provided by the Eurostat. The technical format of transmission shall be Statistical Data and Metadata eXchange (SDMX) developed for data submission of statistical data and metadata. Although the application of this system is not an obligation just a proposal, we would like to use it to offer an easier way to access data at international level, too.

*

Since the time the former part of the study was written, Hungary has successfully carried out its 15th census between 1 and 31 October, 2011.

The data collection period lasted from 1 to 31 of October and was followed by a supplementary data collection phase from 1 to 8 of November.

The 15th Hungarian census provided, for the first time, the opportunity for the inhabitants to choose from three methods of data provision: fulfilling the questionnaires via internet or on paper on their own, or answering to the enumerators. The electronic data collection was carried out in the first half of the month. 19 percent of the “addresses” completed the questionnaires through internet, while another 15 percent chose self-enumeration on paper. However, most of the households answered the questions about themselves and their dwellings to enumerators in traditional interviews.

Among the three methods, respondents chose that way of answering which was the most convenient for them. The electronic enumeration was practical, comfortable, fast and flexible for those who had only a short time for such an activity. The in-built instructions and checks as well as the answer dropdowns helped answering.

Self-enumeration on paper was preferred by those who did not want to agree on an appointment for enumeration or to let an "unknown enumerator" in their flat. This method proved to be the most difficult for the respondents as they could lean only on themselves in fulfilling the questionnaires, and thus, the accuracy of completion was determined considerably by their responsibility and attitude.

The personal interview was chosen by those who opened the door with confidence to the enumerators, trusted in their knowledge, competence and expertise and did not want to wrestle with the fulfilment.

The processing of the census data has already started. The preliminary data will be published in spring 2012, while the dissemination of the final data will begin from the end of 2012.[1]

[1] *Editorial comment.* The *Hungarian Statistical Review* will also report on the final results.

Characteristics of Private Farms and Family Farm Labour in Hungary by Settlement Size

Zsolt Andrási
Drafter, HCSO
E-mail: Zsolt.Andrasi@ksh.hu

Pál Bóday
Head of Section, HCSO
E-mail: Pal.Boday@ksh.hu

The Hungarian Central Statistical Office (HCSO) conducted a general agricultural census (ÁMÖ) in 2010, where enterprises and households performing agricultural activity were surveyed. According to the preliminary results, the number of private farms fell significantly in the last ten years. The present study concentrates on these farms and their labour force, drawing attention to a specific factor, the settlement size and the characteristics related to it.

The total population, the agricultural labour force as well as the private farms are not evenly distributed in Hungary by settlement size, specific disparities can be seen in this respect. The study aims to demonstrate not only these differences, but also several other specialities, for instance, the changing role of agricultural activity, the education level and age of the farm labour force on the basis of ÁMÖ 2000 and 2010 databases.

KEYWORDS:
Agricultural census.
Private farms.
Settlement size.

The size of settlements (described by the number of population) has a significant effect on the general situation, role, and demographic processes of the municipalities (*Beluszky* [1999] p. 293). In 2010, their 56 percent had a population of less than 1 000 inhabitants and nearly 9 percent of them reached the number of 5 000. Conversely, only 8 percent of the total population lived in places with less than 1 000 inhabitants, and nearly 69 percent in settlements with over 5 000 residents. (For the main characteristics of the Hungarian settlement structure, see *Perczel* [2003].)

Small villages are often associated with such concepts like disadvantageous, peripheral location, lack of supplies, ageing, etc. Although size is important, it is not the only factor which poses problems for small villages, other dimensions also should be taken into consideration (*Szabó* [2011]). They usually have small areas under cultivation and weak soil and are mostly located in areas surrounded and divided by hills (*Beluszky* [1999] p. 293.).

In Hungary the traditional and close connection between agriculture and villages radically changed in the last half century (after World War II) in consequence of the drastic reduction in the number of agricultural employees. Thus, nowadays, the formula that villages equal to the agricultural role is untenable without doubt (*Beluszky–Sikos* [2007]). It is also supported by (the dimension of) urbanization when differences are fading between rural and urban areas. This process does not only mean the improvement of the infrastructure and supplies, but also the transformation in the way of living and the spread of urban lifestyle. However, the changed, wider role of villages did not alter the fact that the main places of agricultural production are still the rural areas (smaller settlements) and their income from agriculture is significant. Nevertheless, beyond these obvious facts, it could be relevant to draw distinction by other features between agricultural activity of rural areas and that of urban regions. Our aim is to find these differences, paying attention only to the private farms and their labour force.

The study examines the preliminary data of the general agricultural census (ÁMÖ) 2010 compared with the previous census conducted in 2000 (where it is relevant). The eight categories of settlements, classified by total (midyear) population in 2010, and the database of the settlement-level data of private farms (part of the households exceeding the threshold of a certain size) were linked together.[1]

[1] ÁMÖ covered the farms which provided agricultural services over the previous 12 months or reached/ exceeded at least one of the following thresholds: they had at least 1 500 m^2 of productive crop land; 500 m^2 of orchards and vineyards together; 100 m^2 of area under glass or high (accessible) cover; or the following livestock: one head of larger livestock (cattle, pig, horse, sheep, goat, buffalo, ostrich), 50 heads of poultry (chicken, goose, duck, turkey, guinea-fowl), 25–25 rabbits, furry animals, or pigeons for slaughter; or 5 beehives.

1. Some preliminary results of ÁMÖ

The seventh general agricultural census of Hungary was conducted in June 2010 with the reference date of 1st June. Its targeted respondent units were divided into two groups: households and enterprises. While the latter were subject of the census on the basis of their agricultural activity, irrespectively of their size, households were only surveyed if exceeded the threshold of a certain economic size.

According to the preliminary data, 8 800 enterprises and 567 thousand private farms[2] were engaged in agricultural activity, however, almost further 1.1 million households had certain agricultural production in kitchen gardens and holiday homes under the threshold. The number of farms decreased by 41 percent compared with the previous census in 2000. In addition, the characteristics of farms have changed, too. (For details see *KSH* [2011].) Therefore attention shall be paid to some phenomena connected with private farms, which are also the objective of this study.

The analysis of the purpose of production shows that 60 percent of the farms produce only for own consumption which ratio roughly equals to the value of the previous census. However, the share of the market-oriented private farms in the private farming sector has increased from 8 to 20 percent over the last decade.

In 2010 1.1 million unpaid family labourers performed agricultural work on private farms. This means a 45 percent drop compared to the previous survey. Moreover, the share of family farm labour within working-age population has decreased notably.

The age structure of the labour force changed between 2000 and 2010, too: the average age rose by nearly 4 years. The share of young farmers (under 35) has decreased, while that of elder holders (over 54) has increased. Simultaneously with the improvement in the general educational level of family workers, the ratio of the labour force who graduated from agricultural colleges/universities has increased notably.

2. Private farms by settlement size category

As it was mentioned, the number of private farms has continuously decreased over the past decades in Hungary: in the last ten years the reduction was 41 percent. It was mainly caused by their disadvantageous farm structure, capital shortage and

[2] The private holdings providing only agricultural services are excluded from this study as their share within the total private holdings is under 1 percent. Therefore little more than 566 thousand households are included in the present study.

the lack of appropriate expertise (*Pintér* [2011]). The present study examines whether the different settlement sizes show coherence with the rate of this decline.

Table 1

Private farms and family farm labour force in Hungary by settlement size

Settlement size category (inhabitant)	Number of private farms		Share of the total in 2010 (%)	Family labour			Share of total population in 2010 (%)
	2000	2010		2000	2010	Share of the total in 2010 (%)	
under 500	71 655	40 241	7.1	151 983	78 273	7.3	2.8
500–999	107 255	63 036	11.1	232 324	123 267	11.5	4.9
1 000–1 999	179 264	108 138	19.1	384 795	209 177	19.5	9.1
2 000–4 999	247 350	152 766	27.0	512 520	290 702	27.0	14.4
5 000–9 999	120 441	69 179	12.2	239 364	129 032	12.0	9.0
10 000–49 999	170 549	99 120	17.5	334 399	180 264	16.8	24.2
50 000–99 999	21 062	13 417	2.4	44 684	26 192	2.4	6.6
100 000 and above	36 534	20 330	3.6	74 519	37 728	3.5	29.0
Hungary total	*954 110*	*566 227*	*100.0*	*1 974 588*	*1 074 635*	*100.0*	*100.0*

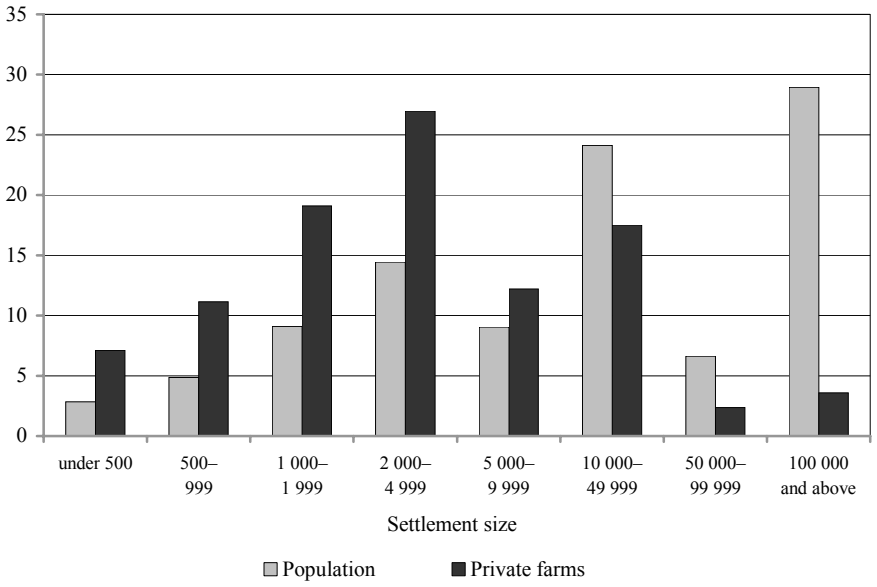
Between 2000 and 2010, the number of private farms fell by nearly 388 thousand, while their distribution by settlement size did not change significantly, and a spectacular difference could not be seen among categories. However, it shall be mentioned that the highest (44%) decrease of the number of private farms was in the two 'extreme' (the smallest and the biggest) settlement size groups, which could be partly explained by the formerly mentioned disadvantageous features of small villages and the naturally modest role of agriculture in large settlements. Still, due to the indefinite trend and the relatively small differences between categories, exact explanation cannot be given. Despite the fact that the situation of settlements with increasing number of inhabitants is improving in terms of agricultural production (see *Bóday–Kaposi–Konrád* [2001]), the result has shown that there is no direct link between the settlement sizes and the reduction in the number of private farms.

The role of agriculture has decreased significantly in economy over the past few decades. As mentioned formerly, the function of villages has also changed, their close connection with agriculture has slacked. This, of course, does not mean that the traditional structures broke up, agricultural activity, farms, and agricultural labour concentrate in smaller settlements even today.[3] (See Figure 1.) In 2010 more than

[3] To compare settlement categories appropriately, we should also take into consideration the large differences by the weight of farms and their labour.

three-quarters of the private farms and family farm labour were engaged in settlements with under 10 000 inhabitants, while only 40 percent of the total population lives there. Over a quarter of the private farms are located in settlements between 2 000 to 4 999 inhabitants but even places with 10 000 to 49 999 inhabitants have notable share. The previous census has shown approximately the same distribution.

Figure 1. The distribution of the population and private farms in Hungary by settlement size, 2010
(percent)



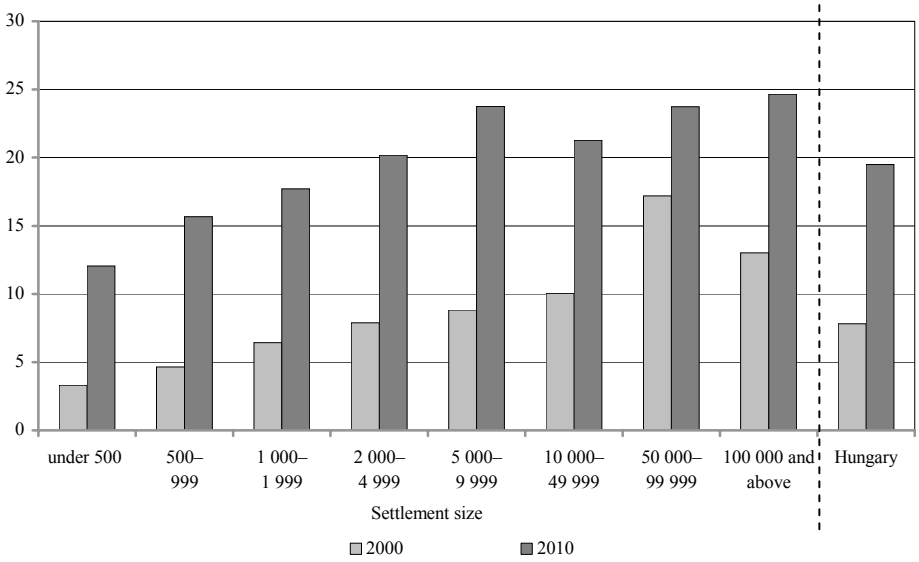
In contrast, other factors have changed considerably, which can illustrate the differing agricultural activities by settlement size. Among other, the census also examined the purpose(s) of production. Accordingly, farms vary depending on whether they

– produce only for own consumption;
– market the surplus over the production for own consumption;
– produce mainly for the market (more than 50 percent of the production is marketed).

Production only for own consumption is most typical of smaller settlements according to the data of the last two censuses. Approximately two-thirds of the farms in settlements having under 1 000 inhabitants have no marketing activity at all. This characteristic of the smallest villages seem to be stable over years, similarly to the

relevant Hungarian average (60%). Examination of the production for marketing is a more interesting task as certain changes can be experienced in this field. While the number of private farms decreased by 41 percent, that of the farms producing mainly for the market increased by 48 percent (more than 110 thousand farms), their total share within the private farming sector was 19.6 percent in 2010 (compared with 7.8 percent in 2000). In contrast, the number of those who produce only for household consumption has fallen by 41 percent and who sell only the surplus has dropped by 62 percent. These trends show that, simultaneously with the decrease in agricultural production only for own use, the profile of the farms has changed towards product selling.

Figure 2. The ratio of private farms producing mainly for the market within the private farming sector by settlement size
(percent)



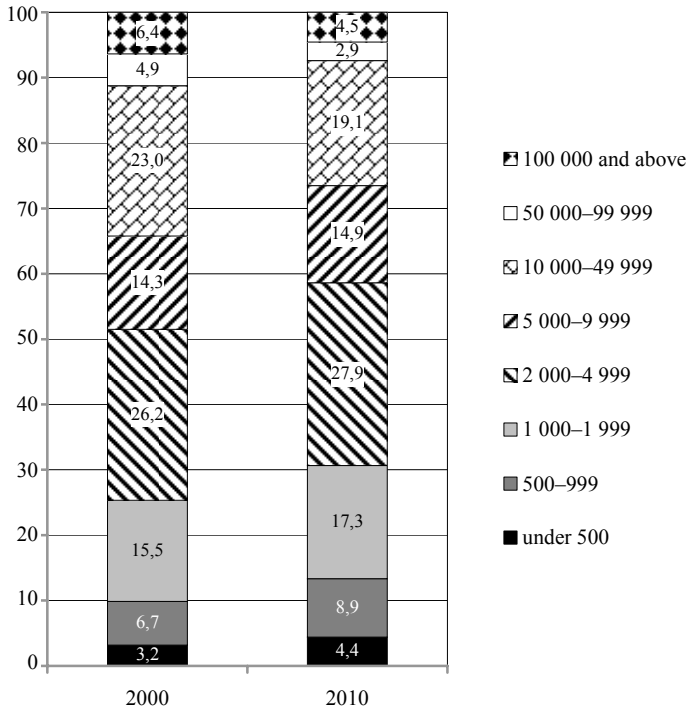
Examination of the production for sale by settlement size shows that the smallest villages have the lowest proportion of farmers producing mainly for the market, and the ratio increases in line with the number of inhabitants up to the category of the largest cities. (See Figure 2.) This could partly be explained by the previously mentioned disadvantageous location, negative agricultural capability and general characteristics of small settlements, for instance, by the education level (see later) or social situation of their inhabitants. The relatively high share of towns, large cities could be firstly the consequence of the smaller possibility of production in gardens for own consumption because of extensive building covered areas, secondly, of the difficul-

ties experienced during the survey in finding producers in these places, if they exist there.[4] Therefore, those who perform commercial, marketing activity, have a relatively high share within the small number of the total observed in these categories.

On the whole, the share of private farms producing mainly for the market was 2.5 times higher in 2010 than in 2000, however, it changed differently in the eight categories. For example, it was 3.5, 1.4, and almost 2 times higher in the smallest settlements, in the category of 50 000–99 999 inhabitants, and in the largest cities, respectively. These values can be explained, on the one hand, by the overall reduction in the production for household consumption and, on the other hand, by the fact that where the level of production for sale was previously high, the degree of growth was lower than in other categories.

In addition, marketing producers became more concentrated by size category.

Figure 3. The distribution of private farms producing mainly for the market by settlement size
(percent)



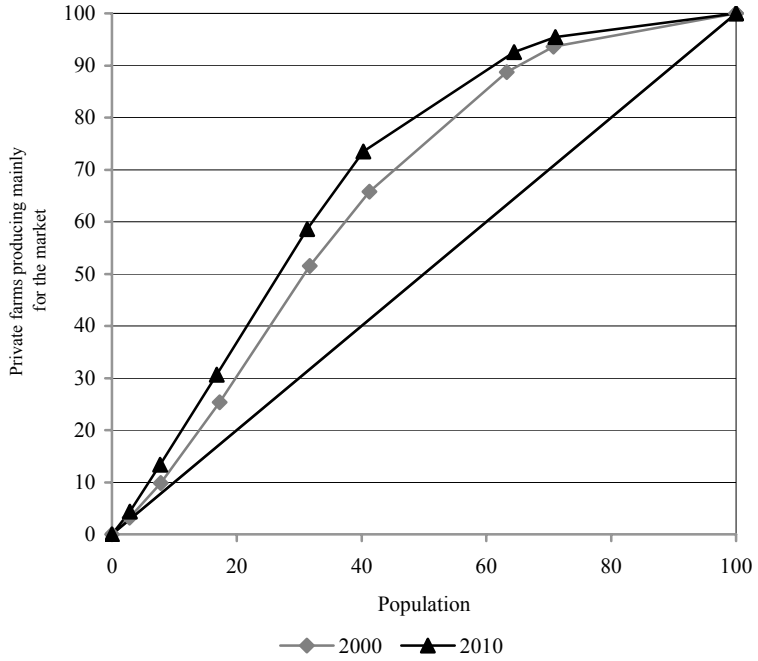
In categories under 10 000 inhabitants, the share of farms operating for the purpose of marketing has increased, above that, it has fallen. (See Figure 3.) In 2000,

[4] Farms in inner part of the cities were designated by address; in other areas every household was surveyed.

two-thirds, while in 2010, 73 percent of such private farms were located in settlements with less than 10 000 inhabitants. Presumably, it is in connection with the change of family farm labour experienced in the last ten years, with the lack of young holders in large settlements that is explained in Section 4.

The former trend can be also examined for the total population. (See Figure 4.) In the last ten years, the farms producing mainly for the market were more concentrated in certain settlement categories; and marketing activity was extending to smaller settlements.

Figure 4. The distribution of private farms producing mainly for the market in relation to the total population by settlement size
(percent)



3. The family labour force of private farms

According to the Labour Force Survey conducted in 2010, over 3.8 million people were employed throughout the country. In agriculture, hunting, forestry, and fish-

ing (by the renewed Standard Industrial Classification of All Economic Activities (TEÁOR 08 /NACE Rev. 2/)) 172 thousand persons, 4.5 percent of the workforce were employed. However, institutional labour statistics data shows that 77 thousand people, 2.8 percent of the employees worked in these areas. The labour force survey has the speciality that observes labour in private farms partially, furthermore, the institutional statistics not at all;[5] therefore the annual working unit (AWU) is a more precise indicator.[6] Accordingly, in 2010 agricultural activity in Hungary was as much as if 437 thousand employees would have worked in full-time job for enterprises and on private farms. This contains both salaried and non-salaried work. With reference to private farms, non-salaried work is equal to the work of family members. Therefore, the latter[7] could be expressed in 332 thousand AWU in 2010. This is in accordance with the approximately 1.1 million family members – surveyed by ÁMÖ – who did more or less agricultural work on farms. The nearly 2 million family labours fell nearly by 46 percent since the last survey in parallel to the decrease in the number of farms. However, their work – expressed in AWU – has not declined by such a large rate (38%), which means the concentration of agricultural work: less labour worked more in 2010.

The distribution of farms and their labour force by settlement size is very similar. (See Section 2.) The average number of the family labour force per a farm in each category has a value around 2, therefore, it is needless to repeat the data of Table 1. The share of the family labour force in the working-age population was 24 percent ten years ago and nearly 13 percent in 2010. Concerning settlement size, this indicator decreases in direct proportion to the growing number of inhabitants. (See Figure 5.) Consequently, a smaller settlement means a bigger role in agricultural activity.

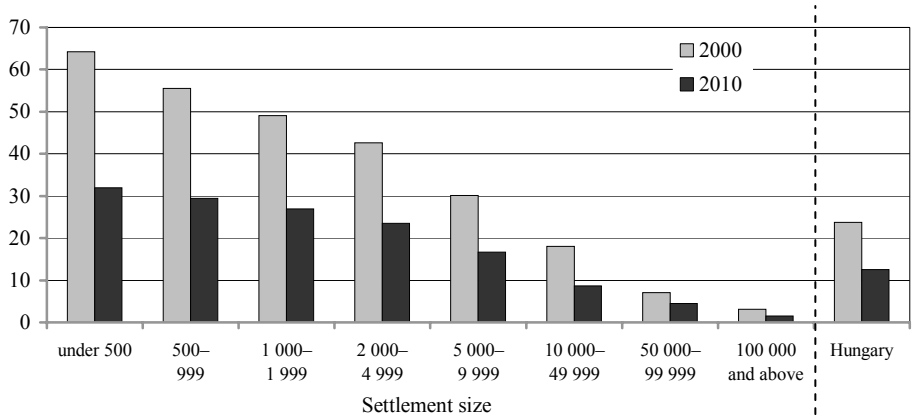
In municipalities with less than 1 000 inhabitants, the labour force engaged in agricultural activity was in majority within the total population in 2000, while in 2010 its share did not reach 50 percent (even in the smallest villages, where it was only 32 percent, half of the value in 2000). The categories with 2 000–4 999 and 5 000–10 000 inhabitants had the ratios of 24 and 17 percent in 2010, respectively. The level of decrease was almost the same in each category, regardless of the settlement size; thus, the weight of the categories has not changed significantly over the past ten years.

[5] The labour force survey observes employees of enterprises and workers of private farms whose main income is from activity in agriculture, forestry and fishing. Institutional labour statistics represents enterprises with more than 4 employees and the full scope of budgetary institutions, observes agriculture, forestry and fishing branches.

[6] One annual work unit corresponds to the work performed by one person who is occupied on an agricultural holding on a full-time basis. According to the Hungarian methodology, full-time means 1800 working hours, 225 working days of eight hours per day.

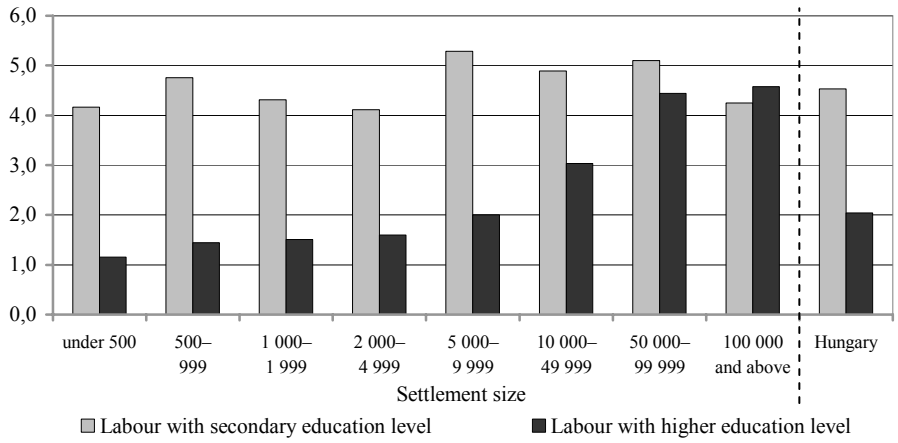
[7] Family labour force: a private holder (farmer) and his/her family members, who carried out agricultural work on a private farm during 12 months prior to the survey and did not get paid for it.

Figure 5. The share of family labour in total population (14+) by settlement size
(percent)



Beyond size and its change, the structural aspects of the labour force (for example education, age) are also important to study. Data only on the labour force with secondary and higher agricultural education were available for comparison as the methodology of data collection has changed. Risen from 1.2 percent of 2000, 2.0 percent of the 1.1 million family labourers have completed higher agricultural education by 2010. The ratio of farmers (holders of farms) who graduated from agricultural colleges/universities has increased from 1.8 to 2.8 percent. The bigger the settlements are, the greater these proportions.

Figure 6. The share of family farm labour with secondary and higher agricultural education by settlement size, 2010
(percent)



The data analyzed show a stable structure between 2000 and 2010, ten years ago the relations were the same, just the values were lower (in most cases). For example, the ratio of those family farm labourers who have completed secondary education was 4.5 percent in both years, no significant change occurred. In the case of secondary-level qualifications, no growing trend could be observed either at the time of the former or the latest census. Nevertheless, the comparison of the education levels and the purpose of production (see Section 2) confirms that the marketing activity presumes a higher level of qualifications and appropriate expertise, which is also proven by the settlement-level analysis.

Beyond education, the gainful activity and the source of income also can be analysed. The survey has the following categories: the family labour force *1*. has no income from gainful activity[8] outside the farm; *2*. has other gainful activity in full-time job; *3*. has other gainful activity in part-time job; *4*. mixed.[9]

In 2010, 57 percent of the family labourers did not have any other gainful activities outside the farm; however 40 percent had other full-time jobs. This has not changed in ten years notably. On the basis of settlement size, the complete equalization of the family labour that has no income from outside the farm can be seen. Previously, their proportion was higher in smaller villages and lower in bigger towns and cities. The changes were particularly due to the growth in upper categories.

Table 2

The share of family labour that has no income from outside of the farm
(percent)

Year	Settlement size category								Hungary
	under 500	500–999	1 000–1 999	2 000–4 999	5 000–9 999	10 000–49 999	50 000–99 999	100 000 and above	
2000	61.5	59.3	58.1	58.0	55.6	51.8	49.2	51.9	56.7
2010	58.1	58.4	57.6	57.2	56.0	55.8	58.1	56.6	57.1

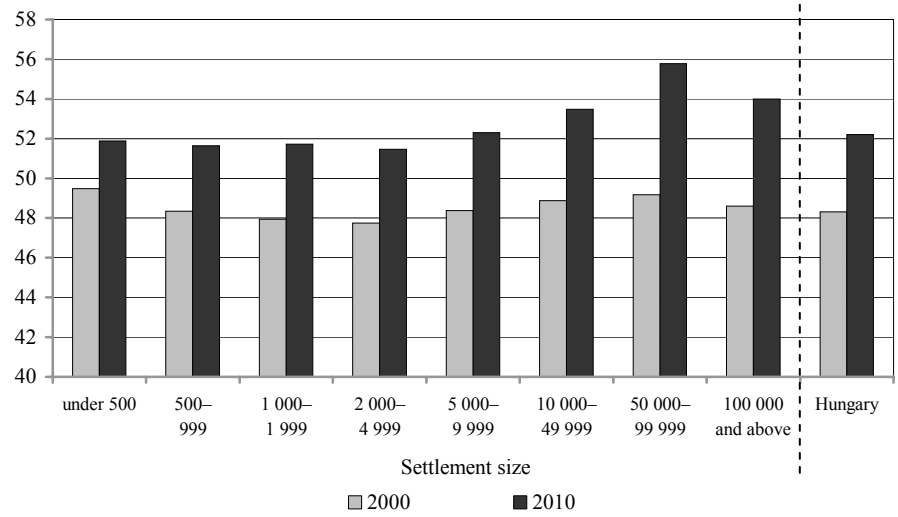
It can partly be explained by the ongoing flow of labour from agriculture to other sectors in villages, partly by the increased ageing of the family labour force in cities. The average age of family farm labourers and farmers was 52.2 and 56.2, respec-

[8] Retired persons' pension and other social allowances are not considered as income from gainful activities.

[9] The original categories of ÁMÖ in 2010 were as follows: the family labour force *1*. has no income from outside the farm; *2*. has other gainful activity in full-time job which has direct relation to the farm; *3*. has other gainful activity in full-time job which has no direct relation to the farm; *4*. has other gainful activity in part-time job which has direct relation to the farm; *5*. has other gainful activity in part-time job which has no direct relation to the farm, *6*. mixed.

tively in 2010.[10] The former figure has risen by 3.9 years, the latter by 1.6 years in ten years. In 2000, there were no significant differences among settlement sizes by either the average age of the family labour force or that of farmers; only the small villages had higher-than-average figures. By 2010, the dissimilarities have increased due to the higher level of ageing in larger cities.

Figure 7. The average age of family labour force by settlement size



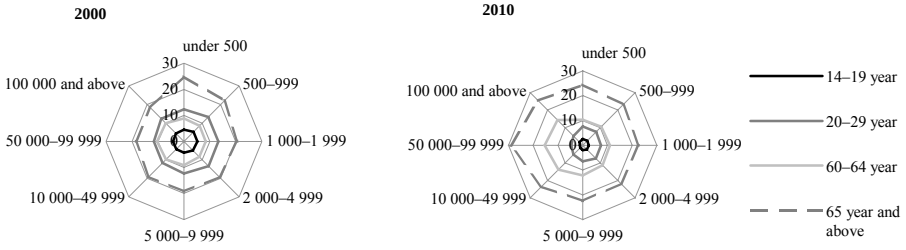
Significantly greater increase in the average age of labourers could be observed in settlements with more than 10 000 inhabitants than in other categories, and they have the highest values, too. This can be explained by the specific features of the Hungarian political and economic transformation of the 1990s, which played also part in the privatisation and land reform. In that period, plenty of the owners who had had land before nationalization were recompensed and got arable land(s) in rural areas and/or in cities. One part of them became agricultural entrepreneurs, while their other part rented their land(s) (*KSH* [2008]). In 2000 the average age had showed a balanced picture within settlement groups, which, however, in ten years later became different as generation change did not occur in cities, younger holders did not take the lead on farms which had started up by a kind of necessity.

This theory is confirmed by the examination of the concentration of certain age-groups in the eight settlement categories. Four age-groups (14–19, 20–29, 60–64, 65+) of the family labour force were chosen for the analysis. The shares of the two

[10] The mean age was counted by date, based on age-groups, as in the survey the exact age of persons was not questioned.

which contain young labourers became smaller within the family labour force in each settlement group from 2000 to 2010, while those of the groups of the elder increased significantly. In 2000, all four groups showed a balanced picture by settlement category; only in the smallest villages can be seen a higher rate of the elderly labour force. (See Figure 8.) In 2010, as it was mentioned, the proportions of the groups of people aged 14–19 and 20–29 decreased, especially in the bigger towns and cities. In contrast, those of the other two age-groups were significantly higher in large cities, urban areas, which confirm the phenomenon of the lack of young farmers in cities.

Figure 8. The share of the selected age-groups in family labour force by settlement size
(percent)



4. Conclusion

The present study demonstrates several disparities of the Hungarian private farms and family farm labour force, based on settlement size. The role of agricultural activity is continuously decreasing in both urban and rural areas, which is also shown by the number of farms and their labour force. The structures of production and the features of farm workforce have become different, too. Over the last ten years, the farms have changed their profile towards product selling, while their activity for own consumption has decreased. The smallest villages have the lowest share of private farms producing mainly for the market, and this ratio is increasing in line with number of inhabitants, in addition to the growing concentration of marketing producers. The proportion of the family farm labour force has decreased significantly in every group. However, a smaller settlement means a higher role in the agricultural activity of the population. The educational level of the farm labour force also shows disparities by settlement group: in settlements having greater number of inhabitants, the proportion of those who completed tertiary education is higher. Furthermore, stronger ageing can be experienced in large cities.

This analysis revealed several disparities, however, it does not mean that the settlement sizes and the observed differences are causes and effects at the same time. Other dimensions (for instance regional disparities) should be also integrated into the analysis to discover a wider and deeper coherence.

References

BELUSZKY, P. – SIKOS, T. T. [2007]: *Változó falvaink (Magyarország falutípusai az ezredfordulón)*. MTA Társadalomkutató Központ. Budapest.

BELUSZKY, P. [1999]: *Magyarország településföldrajza*. Dialóg Campus Kiadó. Budapest, Pécs.

BÓDAY, P. – KAPOSI, L. – KONRÁD, A. [2001]: *Az egyéni gazdaságok munkaerő felhasználása 2000*. Központi Statisztikai Hivatal. Budapest.

KSH (HUNGARIAN CENTRAL STATISTICAL OFFICE) [2008]: *A mezőgazdaság fejlettségének regionális különbségei*. Szeged.

KSH [2011]: *Agriculture in Hungary, 2010 (Agricultural Census)*. Budapest.

PERCZEL, GY. (ed.) [2003]: *Magyarország társadalmi-gazdasági földrajza*. ELTE Eötvös Kiadó. Budapest.

PINTÉR, L. [2011]: Magyarország mezőgazdasága a 2010. évi általános mezőgazdasági összeírás tükrében. *Statisztikai Szemle*. Vol. 89. No. 2. pp. 185–198.

SZABÓ, SZ. [2011]: *Vidéki térségek Magyarországon*. Trefort Kiadó. Budapest.

GÉSA: The Tool for Survey Control, Quality Assessment and Data Integration

Ildikó Györki
Senior Advisor
Hungarian Central Statistical Office
E-mail: Ildiko.Gyorki@ksh.hu

Supporting the flow of survey design and collection, the HCSO uses a standard metadata driven system, the so-called GÉSA system. This survey control system manages all economic and social statistical data collections of the office, observing the businesses and other institutions. The paper describes the aim and functions of the GÉSA system. It presents the unified tools of survey control for preparing the connection, integration of data collections through the way of the assignment of their survey frame. The paradata and other information connected to the survey frame and to the flow and result of the data collection are in the same structure and with the same code lists. It makes possible the unified monitoring, evaluation and quality assessment of the data collections.

KEYWORDS:
Statistical methodology.
Statistical survey.
Statistical system.

The transformation of the European Statistical System (ESS) is built on the Communication from the Commission to the European Parliament and the Council on the production methods of EU statistics (*Commission of the European Communities* [2009]). The so-called vision document (hereinafter referred to as vision) defines the objectives of reengineering to improve the coherence and comparability of data and to increase efficiency and cost effectiveness. Its two important elements are:

– replacement of the so-called "stove pipe model" in which statistics of different domains are produced independently, with an integrated model;

– standardization and integration of the formerly separated production processes.

The conception, methods and requirements of the vision fit well into the former developments and the strategy of HCSO.

The statistical office tried to create general systems for the phases of the statistical processing flow. These are built on metadata and their aim is to give a unified solution for as wide scope of data collections as possible. Besides, this unified design of the databases facilitates the integration of different statistical topics.

GÉSA (economic organisations and their data provision) is the survey control system for the observation of businesses and other institutions. It is the earliest metadata driven system in the HCSO which has been working since 1996. It is intended to give a general tool for supporting the tasks of the data collection phase of surveys, such as design, documentation, gathering of questionnaires, and evaluation of data collections. The solution is built on standard procedures.

At present, GÉSA manages and controls 132 data collections which account for 98 percent of all data collections where the data suppliers are institutions. In the case of more than 80, the questionnaires can be filled in and sent to HCSO via internet. Besides data collections, 50 administrative sources belong to GÉSA for the sake of unified processing. The number of modules, functions supporting the survey design, the data collection and its evaluation are more than 160. The system maintains a population with almost 2 million units, more than 380 000 data suppliers and 2-3 million pieces of questionnaires in one year.

As time went by, GÉSA has incorporated more and more functions and more and more data collections, its maintenance application has changed but its basic concept has remained stable.

Its main principles are the following.

– The terms and procedures connected to the data collections can be standardized. Giving a description of the unique characteristics of data collections in the metadatabase and using metadata, we can build metadata-driven procedures.

– The unique survey frames of data collections, their paradata[1] referring to the data suppliers and statistical units can be connected to a common master frame. Hereby,

– the standard management, monitoring, quality assessment of all data collections maintained by the system;

– common and unified communication with the data suppliers, and easy control of response burden become possible.

In the beginning, the GÉSA system dealt only with Business Register (GSZR)-based data collections, where data suppliers have a tax identification number. Later its scope was expanded to the data collections whose statistical units are institutions (dependent social institutions, non-profit organizations, etc.) without a tax identification number.

During this period, survey control became more precise as it was worked out for data collections with data suppliers which have to complete more than one questionnaire (for example by kind of activity units, local units, or settlements).

Concerning functionality, the most characteristic changes were brought by the widening scope of the proactive functions, thereby assisting data suppliers in responding in due time and in appropriate quality.

The support of work organization, the transparent task management and control are among the main functions of GÉSA. The system easily follows the changes in the organization of HCSO and in the responsibility of statisticians.

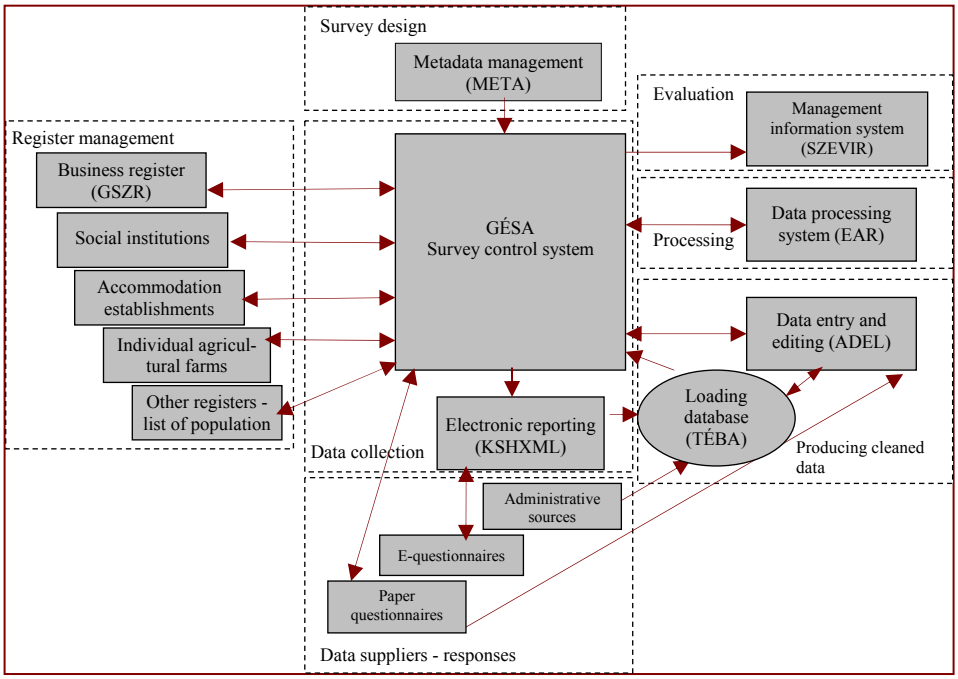
The main concepts, terms of the system and their relations are included in Appendix I.

1. The place and role of GÉSA in the statistical processing flow

GÉSA keeps track of data collections and questionnaires from survey design to loading and entering micro data into the database. Figure 1 shows the relation of GÉSA with the other phases and standard application systems of the processing flow.

[1] Paradata are data collected about the process of survey production to measure and improve quality and saving costs. Shortly, the difference between metadata and paradata is the following: metadata are data about data, paradata are data about process.

Figure 1. The environment of the GÉSA system – relation among the statistical processing elements



The wide scope of information, paradata which GÉSA gathers about data collections constitutes grounds for designing new data collections and for redesigning the existing ones after their evaluation. GÉSA provides a structured metadata description facility for documenting the information of the *survey design*.

Data collections involved in the system are built on several (business and other) registers. The snapshots of these registers connected to the reference time of data collections give input to the master frame. The *survey frame* and, in most cases, the sample frame of data collection are assigned from this common master frame, using the metadata description of the population and sampling.

GÉSA supports the personalized printing of the questionnaires and their mailing to the data suppliers, and forwards information to the electronic reporting system (KSHXML or the new system being under development) to inform the *electronic respondents* of their tasks. Several *proactive features* help data suppliers to send the questionnaires in time.

In the course of data collection, the work of statisticians is governed by the description of their tasks and that of the organizational units. All incoming questionnaires are registered. If some questionnaires are missing after expiry of the

deadline, *urging* letters of different degrees are sent automatically or individually to the data suppliers. *Response or non-response* is coded applying a unified nomenclature.

The progress made in data collection and in the processing of the questionnaires can be followed by both the statisticians and the management. Different statistics help the *evaluation* of the actual state and result of the process. From GÉSA information (on population, response, non-response, urging, etc., that is, from paradata), indicators on quality and response burden can be computed.

The GÉSA master frame, survey frames and paradata serve as a base for *other phases of the statistical production process*, for example, for loading micro data to the database (TÉBA – automatic registration and loading of questionnaire data), for entering, validating and correcting data (ADEL – frame system for data preparation), as well as for imputing, weighting, aggregating and making other procedures of statistical processing (EAR – unified statistical data processing system is under development).

The statistics about the result of the data collections are forwarded to the *management information system* every day where they can be analyzed in their tendency together with other information.

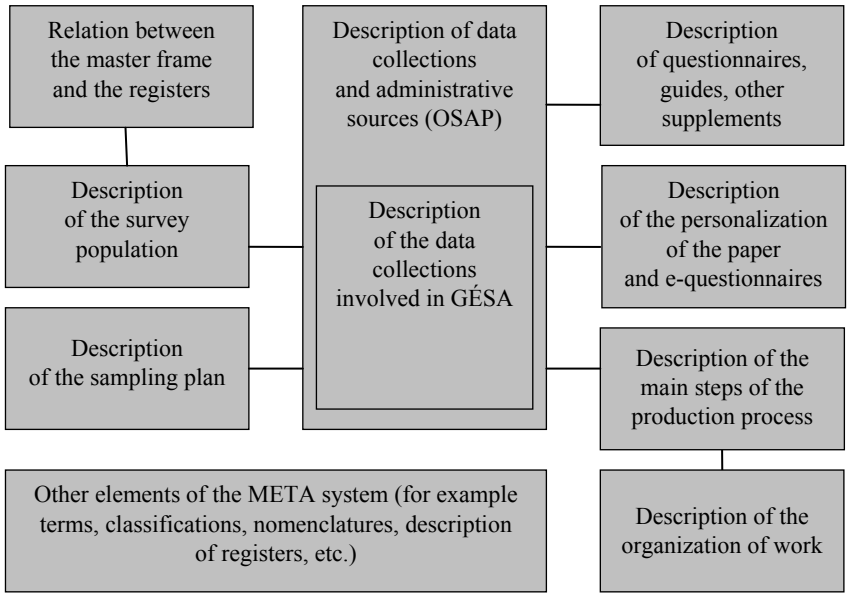
2. Description of data collections in the metadatabase

The metadata-driven operation of GÉSA is built on the description of the attributes of data collections, their supporting tools and steps of gathering data, as well as the division of work in the metadatabase. This metainformation is planned and entered in the survey design phase.

Figure 2 shows the overall diagram of the metadata necessary for the operation of GÉSA.

The primary aim of the *description of data collections* and administrative sources is to document the information needed to the yearly legislation about the National Data Collection Programme (OSAP). Besides legal commitments, it serves as a base for survey control, therefore this chapter as a part of the metadatabase contains all sources of statistics not only on mandatory but also on voluntary data collections and on the takeover of data from other (for example administrative) sources. The description renders account of the organizations that enacts and executes data collection, the topic, population, frequency, method (mail questionnaire, personal interview, etc.) and extension (full-scope, representative) of the observation, the size of the questionnaire, the response deadline, etc.

Figure 2. The main metadata groups describing data collections



As regards data collections included in GÉSA, these pieces of information are completed by more precise ones about the register such as the base of the survey frame, the type of data suppliers and statistical, reported units, the mode of printing and mailing the questionnaires, the possibility of electronic response, etc. The detailed information can be seen in Appendix II.

As it was already mentioned, assignment of a survey frame is built on such a snapshot of the given register that belongs to the reference time of data collection. To this function, the metadatabase has to describe the relation between the master frame (see Section 3) and the registers, as well as the exact *populatio*n, rule and algorithm of creating a survey frame.

The base of sampling is the *sampling plan* for the representative data collections. It contains the rule and method of sampling, the description of strata and the attributes framing a stratum. The system selects the sample from the sampling frame in accordance with these rules and the sample allocation for the given period.

The description of data collections also includes *that of the questionnaire*(*s*). If a subpopulation has to get another type of questionnaire that the others do, we have to identify and link it to that certain variant. Usually a guide and other supplements, code lists, a catalogue of terms, etc. belong to the forms, ordered to the data collections. Data suppliers get their questionnaires in a block whose composition (from different questionnaires, guides, etc.) is ruled also by the metadatabase.

Data suppliers get both paper and e-questionnaires in a personalized form. To standardize the operations related to, and the *personalization* of, the questionnaires, their design relies on *templates and rules*. The metadata-driven personalization is built on these templates and rules documented in the metadatabase.

A part of the metadatabase deals with the statistical processing flow. It describes the periods and deadlines of the various *phases of data collections*, for example, defines the deadline planned for response, or for the different urging types, according to the frequency of the given data collection.

To control and support the steps of data collections, it is necessary to describe the organization and staff of HCSO and the responsibilities connected to different functions and subpopulations. The task management built on this information defines the access to the data and functions of the survey control.

3. Creating a master frame

Most data collections are built on the Business Register (GSZR), in particular, on the description of legal units. Besides, there are other data collections where the survey frame comes from another register being only in loose connection with the Business Register and not integrated in it. These data collections include, for example, the personal interviews on individual agricultural farms that are listed in an independent register.

To provide a unified survey control system for different, independently maintained registers, GÉSA creates a unified survey frame, the so-called master frame. It is produced for each reference period from the given snapshots of the registers.

GÉSA differentiates two roles in the master frame. One of them is played by the data supplier units, for example, by legal units, institutions, individual farms, etc. HCSO has legal connection with them and expects their response. The other role is played by the statistical units which are either organizations as a whole or their parts defined by a given aspect, like activity, settlement, etc. Such a statistical unit can be:

– an economic unit, institution, nonprofit organization, individual farm, etc. (in their cases the data supplier and the statistical unit is the same);

– a part of the former category, engaged in an activity or in a group of activities (the so-called kind-of-activity unit – KAU);

– a settlement, a site of an organization (a local unit – LCU);

– an activity performed in a given site (a local kind-of-activity unit – LKAU), for example, accommodation, social activity, research activity, etc.;

– a special form of the previous category, the so-called specialized unit which aggregates the important activities at county level.

Both the data supplier and the statistical units are identified with eight characters. When they are legal units, the identifier is the first eight characters of their tax number. In other cases, the first digit of the identifier is a letter which characterizes the given type of units. Table 1 shows an example of data suppliers and statistical units in the GÉSA master frame.

The attributes characterizing the units include identification features, properties supporting the availability of units (name, seat address, postal address), demographic characteristics (date of establishment, beginning/ cessation of activities, operational status, etc.), economic, stratification aspects (principal activity, legal form, number of employees, county, settlement, composition of owners), links to other registers and to the owner, the maintenance organization.

They can be either administrative or statistical attributes. The first ones come from administrative sources, while the statistical attributes characterize the units according to their real activity. The value of the two attribute types may differ, for example, in the principal activity of the unit which can be either officially reported principal activity or statistical principal activity computed from value added. Similarly, the county where the headquarters is located can be other than the county where the unit performs its main activity. As regards operational status, the official status – legally 'alive' (active, under bankruptcy, liquidation or dissolution proceeding) or ceased (with or without successor) – does not always agree with the real status, as a lot of organizations do not wind up themselves.

For the purpose of statistical processing, the master frame makes difference between active and probably dead organizations, built on their official and non-official status. For the latter one the information is available from their announcements, data collections and tax returns. The population of active organizations is the so-called *base population* which consists of only a half (about 700 000) of the officially existing organizations.

Table 1

Data suppliers and statistical units in the GÉSA master frame

Data supplier identification number	Statistical unit identification number	Name of the statistical unit	Type of the statistical unit	Source
15329767	**15329767**	**Szent István University**	**Economic unit**	**Business register (GSZR)**
15329767	HA022187	Szent István University / Regional Knowledge Center	LKAU	Register of research and development units
15329767	HA025968	Szent István University / Pedagogic Faculty /	LKAU	Register of research and development units
15329767	K0139033	SZIE Dorm of Szent István University	LKAU	Register of accommodation establishments
15329767	K0169040	Zirzen Janka Dorm of Szent István University	LKAU	Register of accommodation establishments
16684801	**16684801**	**Family Help Center Szentes**	**Economic unit**	**Business register (GSZR)**
16684801	K0199087	Holiday Home of Szentes City	LKAU	Register of accommodation establishments
16684801	K0199088	Children's Camp of Szentes City	LKAU	Register of accommodation establishments
16684801	S0052898	Temporary Home of Families	LKAU	Register of social institutions
16684801	S0052902	Family Help Center – Temporary Home of Children	LKAU	Register of Social Institutions
12634048	**12634048**	**Szeged Water Joint Stock Company**	**Economic unit**	**Business register (GSZR)**
12634048	00019084	Szeged Water Joint Stock Company –Water collection, treatment and supply	KAU	Business register (GSZR)
12634048	00019091	Szeged Water Joint Stock Company – Waste collection and treatment	KAU	Business register (GSZR)
N0091559	N0091559	Golden Age 2004 Nursing Home Nonprofit Company	Non-profit institution	Register of non-profit organizations
M0010717	M0010717	Kiss Imre	Individual agricultural farm	Register of individual agricultural farms

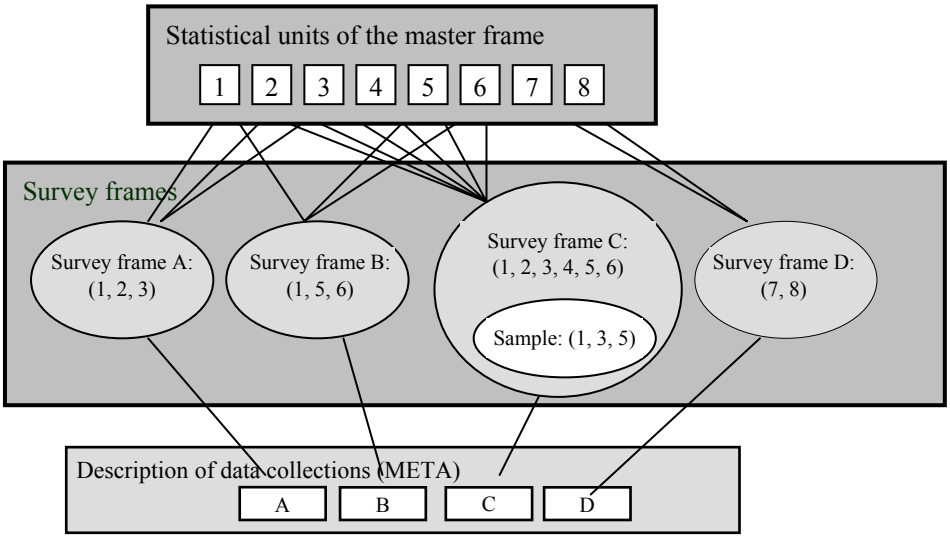
4. Assigning survey frames

The unified master frame serves as a base to select the survey frame of a given data collection by the algorithm, the definition of its population written in the meta-database. A unit of the master frame can be assigned to several individual survey frames.

To refine the coverage of the population, we can use the paradata of the given data collection for previous periods or those of other data collections for the same period available in the GÉSA system (such as the operating status of the data supplier, its response readiness, existence and operation of the observed activity, etc.)

Based on the former information, assignment of the survey frame is automatic. As regards annual data collections, it is built into the end-of-the-year snapshots of the registers. For data collections with short-term periodicity, survey frames are created for each reference period of the year to follow the changes in the organizations.

Figure 3. Assignment of the survey frames and samples in the GÉSA system



Concerning full-scope data collections, the survey frame, each unit of which has to supply data, defines the scope of the data suppliers. Representative data collections observe only a part of the survey frame: the units selected into the sample. In this case, the assignment of data suppliers depends on the method of sampling (which can be based on the survey frame) or another register may be used for this purpose. Where the master frame provides the sampling frame and the sampling plan describes random sampling, this process is a part of the GÉSA functions. In other

cases, sampling happens outside GÉSA, and the sample coming from other systems gives the scope of data suppliers of the representative data collection.

5. Conception of data integration in the GÉSA system

The provision of a base for the integration of statistical data begins with survey design. If we take integration into account during development of the questionnaires, definition and assignment of the survey frames and samples to data collections, data can be linked effectively. Otherwise success is not guaranteed.

Integration can be horizontal and vertical.

– The first means linking statistical measures from different sources for a given population. For example, sales data of a retail trade data collection can be linked to the data of a labour survey.

– In the case of vertical integration, we make a union of a given statistical measure for different, separately collected subsets of a particular population. For example, unifying data on the land usage of agricultural organizations, collected by self-enumeration with those of individual agricultural farms, collected by interviews creates data for the whole national economy.

Table 2

Features of data integration

Possibilities of data integration	Topics to be integrated horizontally				
	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Subpopulations to be integrated vertically					
Subpopulation 1	X	X		X	
Subpopulation 2	X	X	X		
Subpopulation 3	X	X			X
Subpopulation 4	X				X
Subpopulation 5	X				

GÉSA supports both types of integration as regards the description of the questionnaires and the specification of the survey frames.

5.1. Horizontal integration in the specification of the survey frames

A tool for realizing the horizontal integration is the specification, description of subpopulations in the master frame. The individual survey frames can be built up by several predefined and/or own subset definitions. The predefined, common subsets make possible that each data collection has the same data suppliers, statistical units as another, related one has. The data coming from data collections where the survey frames were created in this way can be directly matched to the common subsets. In GÉSA these predefined, identified subpopulations are called "segments". Their description – in addition to their identifier and name – contains information helping the definition and assignment of the population, for example, the algorithm of the selection of the subpopulation from the master frame, the type of observation (full-scope or representative), the relation between the data supplier and the statistical units, the handling of organizations with different operating status, etc.

Examples for segments:

Code	Segment name	Segment short name
4194	Big industrial enterprises (B, C, D economic branch having more than 50 employees) observed in full scope from 2008 for annual and from 2009 for short-term data collections	Big enterprises of industry
4195	Industrial small enterprises observed representatively (B, C, D economic branch, between 5 and 49 employees) from 2009 for short-term data collections	Small enterprises of industry – short term
3789	Big enterprises of construction (F economic branch having more than 50 employees) observed in full scope from 2008 for annual and short-term data collections.	Big enterprises of the construction

The survey frame of data collections can be compiled from the segments included in this example as follows:

Data collection	Name of data collection	Segment	Name of segment	Type of segment
OSAP 1042	Monthly integrated statistical survey of industry	4194	Big enterprises of industry	Predefined segment
OSAP 1043	Monthly integrated simplified statistical survey of industry	4195	Small enterprises of industry – short term	Predefined segment
OSAP 1874	Quarterly integrated statistical survey of industrial, construction and financial enterprises	4195	Small enterprises of industry – short term	Predefined segment
		4194	Big enterprises of industry	Predefined segment
		3789	Big enterprises of construction	Predefined segment
		…..	……	Predefined segment
		4409	Quarterly integrated statistical survey of industrial, construction and financial enterprises	Own segment of data collection

A predefined segment can be built not only on branches but also on any subsets of the master frame. In many data collections the municipalities are the data suppliers. Predefined segments belong to them as well.

Observation of municipalities:

Code	Segment name	Segment short name
3865	Municipalities from 2008 (without county and minority governments)	Municipalities

Some examples for data collections (among others) on municipalities:

Data collection	Name of data collection	Segment	Name of segment	Type of segment
OSAP 1206	Report on benefits provided in cash or in kind	3865	Municipalities	Predefined segment
OSAP 1832	Basic information on organizations providing social and child services	3865	Municipalities	Predefined segment
		1454	Child protection services	Predefined segment
OSAP 1761	Report on private accommodation establishments	3865	Municipalities	Predefined segment

5.2. Vertical integration in the specification of the survey frames

The formerly described segments contribute not only to horizontal but also to vertical integration. If the different subpopulations of a population are observed through different data collections (for example one data collection is conducted on industry, another one on construction and a last one on agriculture) which have common measures, indicators, and the subpopulations are specified with disjunctive segments of the population, measures for the total population can be computed from indicators characterizing the subpopulations.

If a significant proportion of the measures are identical, built on the same terms, and refer to the same period in the vertically connected data collections, we can connect them to a data collection group, to the so-called OSAP group. Within this, GÉSA supervises the disjunctive observation of subpopulations even when the characteristics of the units change.

Examples for the data collection groups are:

– the group of annual, integrated statistical surveys referring to different branches and group of branches, in the case of
– big enterprises of industry;

  - small enterprises of industry;
  - construction;
  - wholesale trade;
  - retail trade;
  - sale, maintenance and repair of motor vehicles and motorcycles;
  - agriculture and services branches;
  - financial intermediation;
  - government, social security and nonprofit institutions;
- land area and sown area on 31 May, based on
  - self-enumeration on the agricultural enterprises of the business register;
  - personal interviews on individual agricultural farms.

6. Proactive support to the data suppliers

Following the assignment of the survey frame and data suppliers of the data collections, the GÉSA system helps, informs and supports the data suppliers with different tools so that they know their obligation of supplying data and the deadline of submission.

At the end of every year they get personalized information about their duties for the next year.

– A calendar lists those data collections to what the data suppliers are assigned as well as the deadlines of response, and gives an address where the questionnaire has to be sent.

– The questionnaires and their annexes (guide, classifications, terms, and other aids) are sent to the data suppliers in personalized blocks. If they respond via internet, they get only a sample questionnaire, but if they have previously mailed such a report, they get the same number of copies as that of the reference periods of the given year.

The calendar with response deadlines is not only sent to the data suppliers in a paper form, but it can also be accessed on the Hungarian site (http://portal.ksh.hu/portal/page?_pageid=36,1&_dad=portal&_schema=PORTAL) of HCSO by their identification number.

Figure 4. The calendar of data supply

The name of data supplier:

XXXXXXXXXXX Trade and Service Ltd

Statistical identification number: XXXXXXXX 5610 113 15

CALENDAR
of the sending deadlines of the questionnaires in 2011

The data collection																
identifier	reference year	name	sending address	sending deadline												
				January	February	March	April	May	June	July	August	September	October	November	December
1045	2011	Monthly Survey of Retail Sale	See here!		21	21	20	20	20	20	22	20	20	21	20
1646	2010	Report on the sales of retail trade and catering by commodity groups	See here!	20											
1646	2011	Report on the sales of retail trade and catering by commodity group	See here!				20			20			20		
1872	2010	Monthly integrated statistical survey of agriculture, trade and services branches	See here!	12											
1872	2011	Monthly integrated statistical survey of agriculture, trade and services branches	See here!												
1878	2010	Quarterly integrated simplified statistical survey of agriculture, trade and services branches	See here!	20											
1878	2011	Quarterly integrated simplified statistical survey of agriculture, trade and services branches	See here!				20			20			20		
2009	2010	Report on the number of job vacancies	See here!	12											
2009	2011	Report on the number of job vacancies	See here!				12			12			12		

HCSO Regional Directorate, Pécs
Seat address: 7623 Pécs József Attila u. 10/A
Mailing address: 7602 Pécs Pf.: 371

The data suppliers, who already reported in the traditional way, get their questionnaires in a printed form. The personalized composition of the pack of questionnaires and the pre-printing of the known register data (identifier, addresses, name, some attributes) of data suppliers and statistical units facilitate the reduction of the data suppliers' response burden and the exact identification of the questionnaires.

In the questionnaire, the identification data (the OSAP number of the data collection, the reference period, the identifier of the data supplier and statistical unit) are printed not only with characters but also with a barcode. This, besides the formerly mentioned advantages, improves the effectiveness of the work of our colleagues responsible for data collection.

Personalization is built on the metadatabase chapter which deals with the description of printed materials (questionnaires, annexes, blocks, patterns, etc.). From the survey frames of various data collections, GÉSA packs and orders the questionnaires and their amendments belonging to the given data supplier. It selects the proper variant of the questionnaire and puts its copies into the pack based on the frequency of the particular data collection. It orders personal information to each page of the questionnaire. The unified personalization is helped by the templates. The printing office prints the blocks of questionnaires identified by a serial number, using the prepared information in the prescribed order. This method improves efficiency and quality and decreases the cost of mailing.

The data suppliers responding on the Web can see in the task list of the electronic reporting (data collection) system, which questionnaires they should complete in certain periods. The questionnaires of this system are also personalized just like the mailed ones. For the task list and personalization, information is provided by GÉSA.

Before the response deadline, the data suppliers and data providers (the agencies authorized to fill in and send the questionnaires in the name of the data suppliers) get an automatically generated reminder about the sending deadline by e-mail or − in absence of an e-mail address − by fax.

7. Supporting and evaluating data collection

The questionnaires sent to the office, the steps taken to ensure their submission, as well as other attributes and paradata of the data collection activity are stored, maintained by GÉSA in a unified data structure for all data collections and data suppliers, connected to the survey frames and the master frame. The collected information is detailed in Appendix III.

The data collection staff registers the questionnaires arriving by mail with the help of barcode readers or by entering their identifiers manually. The face of the questionnaires contains justification for a negative or a non-response if a blank questionnaire or a questionnaire containing a negative response is sent back. This information is managed by the so-called registration function of GÉSA.

Modification of the contact data of respondents, the time spent on filling in the questionnaires to weight response burden are maintained in a unified way in the phase of registration or entering the questionnaires.[2]

In the case of e-questionnaires, the registration data are the same as for the mailed ones: the arriving time, the reason for a negative answer, new, modified contact data, the time of questionnaire completion are all loaded to the survey frame in an automatic way.[3]

Submission of the missing questionnaires is urged. Similar to the formerly mentioned reminder sent after the deadline for arrival of the questionnaires, the persons responsible for completion and later their chiefs (response units) get an urging e-mail, fax, or letter. Urging has several degrees which is logged by GÉSA.

The reasons for non-responses or negative responses are coded according to a unified nomenclature. Their three types can be distinguished: reasons characterizing the problems/ errors of

– the units of the register describing the data suppliers and the statistical units:
  – false (107) or unknown (108) address;
  – not living or not active unit: ceased with successor (115), ceased without successor (101), under liquidation (102), under bankruptcy proceedings (103), not operating yet (104) or pausing activity (105);
  – active unit under liquidation (116) or under bankruptcy proceedings (117);
  – false classification: incorrect NACE (111), size by the number of employees (112)/ or settlement (county) codes (113);
– the register describing the observed activities of the units:
  – the data supplier has never performed the observed activity (201);
  – the unit has given up the observed activity (202);
  – the unit has paused the observed activity (203);
  – another reason that demands comment (204);
– readiness for data supplying
  – denial of response (801);
  – response is overdue (802) or will be overdue upon agreement (804);
  – connection with the data supplier has failed (803).

[2] See Data entry and validation of the ADEL system in Figure 1.
[3] See Registration and loading electronic questionnaires of the TÉBA system in Figure 1.

The reason for non-response serves as a base to prepare the missing questionnaires for processing, marking the units whose data have to be imputed. Later, in the processing phase, the description of the statistical units is updated with the type of imputation.

8. Feedback, support of data processing

The attributes of the survey frames and paradata of the data collection are also used in the related systems and in the following steps of statistical processing.

The errors of the master frame (and registers), a part of which can be corrected, appear during data collection. The register (e.g. GSZR) gets a direct feedback on the correction and its result.

Such kind of errors can be:

– The address errors are reported by the post office in the mailing procedure when it sends back the questionnaires to HCSO. The register is corrected after making contact with the units having a false address.

– The under coverage of the population can turn out if the classification of the unit is false or the activity of the unit we want to observe isn't described in the register. By correcting the register and adding the missing unit to the survey frame, data collection can be improved.

– The reason for non-response or negative response can show the error in the operating status or other attributes of the unit. GÉSA passes these codes to the business register (GSZR). In the next data collection period, the selection of the survey frames is built not only on the official but also on the non-official operating status codes from data collections. These can make the survey frames more precise.

The processing phases following data collection use the survey frames of GÉSA.

– The data entry and validation phase approves only the questionnaires on statistical units of the survey frame. The validation procedures varying by subset identify the subsets by the attribute of the survey frame.

– The imputation procedure finds the missing units assigned for imputation in the survey frame. Subsequently, the survey frame of GÉSA gets feedback on the applied method.

– The weighting, grossing procedures build the estimation on the survey frame of GÉSA using the non-response codes and the attributes for partitioning the survey frame.

– The aggregation and other statistical procedures rely also on the attributes of the population units described in the survey frame.

The tools and methods applied in a well-prepared assignment of the survey frames – as it was detailed in Section 5 – promotes the record linkage and integration of data from different data collections, decreases the differences between the observed populations and the questionnaires describing them and improves the comparison of statistical data.

9. Monitoring, evaluation, response burden, quality assessment

During data collection, every step and information available from the automatic or manual registration of the questionnaires can be *monitored* in any moment. The following can be queried:

– the rate of responses sent in time or after urging;

– the distribution of the different media types used for submission of responses (electronic, paper mail, etc.);

– the proportion of the mailed questionnaires that were entered or loaded into the database;

– the rate of the missing or empty questionnaires in the case of which the reason for the deficiency was found out;

– the fact whether any arrangements shall be made either by the data suppliers or the staff carrying out data collection.

The subsets of the data suppliers belong to the responsibility of different statisticians. This is described in the metadatabase. Thus, the formerly mentioned monitoring features can be applied not only to the whole data collection but also to the subsets by statistician.

Data collection statistics can be *queried and evaluated* by both the supervisor responsible for and the staff carrying out data collection. This allows the latter to time and to organize its tasks.

Data collection statistics can be queried in time series as well, which makes it possible to follow up the tendencies of the changes in data supply, drawing attention to their possible setbacks.

The GÉSA system sends daily statistics, reports to the management information system.

Response burden indicators can be computed automatically from GÉSA information. Presently four indicators are composed:

– the number of the types of the questionnaires to be completed by the data supplier in a year;

– the number of the questionnaires to be sent by the data supplier in a year (taking their frequency and the number of the statistical units for one data supplier into account);

– the number of the fields in the questionnaires that shall be completed by a data supplier in a year;

– the average time of completion of the questionnaires by data collection. The data supplier may give the filling time voluntarily on the face of the questionnaire. GÉSA stores and processes this information by data supplier, data collection and year.

The indicators of response burden can be analysed on the basis of the most important attributes of the population. This tool makes an in-depth analysis of the indicators possible along the attributes. Among data suppliers, the critical groups and those institutions can be selected whose response burden is the highest.

The information collected during the data collection period through the mailing procedure of the questionnaires and registration of the arrived or missing forms, gives a picture of the accuracy of the registers and the survey frames based on them. It enables the automatic creation of various *product and process quality indicators* for the accuracy of the registers and data collections.

The so-called "expected" number of data suppliers (and statistical units) serves as a base for quality indicators. The difference between the total number of the data suppliers of the survey frame and the expected number of data suppliers is the number of units under liquidation or bankruptcy proceedings. They are "possible data suppliers" which are probably not active.

Two types of the coverage error of the survey frame can be computed.

– *Over-coverage,* the number of units in the survey frame not belonging to the population means the number of those units, who got non-response code 101, 102, 103, 104, 105, 118, 201, 202, 203, or 204 according to the reasons for non-response and negative response mentioned in Section 7. *The rate of over-coverage* is the quotient of over-coverage and the expected number of data suppliers.

– The measure of *under-coverage* is a bit more difficult. It can be inferred from the missing units of the survey frame that came to our

knowledge not from the base register but from another source. One part of these deficiencies can be corrected during data collection when we expand our survey frame with these new units, while the other part is cleared up only after that. The number of these units is called under-coverage whose rate is the quotient of under-coverage and the expected number of the units.

The errors of contact attributes and misclassification can be measured as well.

– *Error of the contact attributes*: The measure has two sources. One is the number of the questionnaires returned by the postal service in the mailing procedure and sent to the data suppliers again after correcting their address. The other is the number and rate of the address errors remaining in the survey frame, shown by the non-response codes (107,108).

– *Misclassification error*: The error of the attributes of the register (master frame) units (NACE, size categories by the number of employees, county, settlement, sector code, etc.) can be measured by the number of data suppliers in the case of which an attribute was corrected during the data collection phase or the non-response code was 112 / 113.

The main indicators of accuracy connect to unit response. GÉSA can provide automatic indicators for response, non-response or imputation of the units.[4]

– *Response rate*: the quotient of the number of the questionnaires (with response or negative response) and the expected statistical units.

– *Response rate with data*: the quotient of the number of the questionnaires with data response and the expected statistical units.

– *Unweighted non-response rate*: the number of the missing questionnaires in relation to the number of the expected questionnaires.

– The *weighted non-response rate* could be composed from the data of sample allocation and the non-response data of different strata. In practice it is not computed yet.

– *Imputation rate of units:* the number of missing statistical units where data were created by one of the imputation methods. It is related to the number of the expected statistical units.

[4] Response, non-response or imputation indicators for the items are computed in the processing phase.

Some other indicators characterizing the data collections are the following.

– *The share of questionnaires according to the mode of submission* (mailed, electronic, e-mailed, etc.). 100 percent is the number of all arrived questionnaires.

– *The number and share of urges*: The number of the questionnaires sent prior urging or after 1, 2, 3, 4 urges (by phone, e-mail, letter, reminder, warning, etc.) correlating to the number of all arrived questionnaires.

10. Summary

The GÉSA system provides unified metadata-driven features for describing, maintaining all data collections observing institutions. The statisticians responsible for data collection can maintain all data collections belonging to them in one application system that makes their work easier. To collect a certain questionnaire, they can use paradata, reactions, remarks, contact information relating to other obligations of the given data supplier and the paradata of other data suppliers of the given data collection. These can help to improve the effectiveness of data collections. The standardized paradata, mailing modes, urging types, codes of response/ non-response and imputation provide the possibility of unified, efficient supervision and the evaluation of data collections.

The survey frames are managed in most of the statistical offices as part of single data collections. In some countries a need has also arisen to unify the data collection tasks and paradata, but we are unaware of a functioning system that gives such a standard solution that the GÉSA system does.

References

BÉRARD, H. – PURSEY, S. – RANCOURT, E. [2005]: *Re-thinking Statistics Canada's Business Register.* http://www.fcsm.gov/05papers/Rancourt_Berard_Pursey_IVB.pdf

CENTRAL STATISTICAL BUREAU OF LATVIA [2010]: *Metadata Case Studies.* http://www1.unece.org/stat/platform/display/metis/Central+Statistical+Bureau+of+Latvia

COLLEGE, M. [2004]: *Assessing and Improving Data Collection Programmes and Survey Design Principles. OECD/UNESCAP/ADB Workshop on Assessing and Improving Statistical Quality "Measuring the Non-observed Economy".* 11–14 May. Bangkok. http://www.unescap.org /stat/meet/wnoe/waisq_resource9p.pdf

COMMISSION OF THE EUROPEAN COMMUNITIES [2009]: C*ommunication from the Commission to the European Parliament and the Council on the production method of EU statistics: A vision for the next decade.* COM /2009/0404. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:HTML

COUNCIL OF THE EUROPEAN COMMUNITIES [1993]: Council Regulation (EEC) No 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community. *Official Journal of the European Union.* Vol. 36. 30 March. pp. 1–11.

EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION [2008]: Regulation (EC) No 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes and repealing Council Regulation (EEC) No 2186/93. *Official Journal of the European Union*. L 61. Vol 51. 5 March. pp. 6–16.

EUROSTAT [2009]: *ESS Standard for Quality Reports.* Methodologies and Working papers. Luxembourg.

GYÖRKI, I. – PAP, I. [2004]: *Metadata Driven Statistical Data Warehouse System at the Hungarian Statistical Office.* Work Session on Statistical Metadata. 9–11 February. Geneva. Working paper.

GYÖRKI, I. – RÓNAI, M. [1999]: *Metadata Management.* Conference of European Statisticians, UN/ECE Work Session on Statistical Metadata*.* 22–24 September. Geneva. Working paper.

GYÖRKI, I. [1996]: *Survey Control as a Subsystem of the Statistical Information System.* Seminar on Integrated Statistical Information Systems and Related Matters*.* 21–24 May. Bratislava. Working Paper.

GYÖRKI, I. [2000]: *Hungarian Solution for Metadata Model.* Conference of European Statisticians, UN/ECE Work Session on Statistical Metadata*.* 28–30 November. Washington, D.C. Working paper.

KREUTER, F. – COUPER, M. – LYBERG, L. [2010]: *The Use of Paradata to Monitor and Manage Survey Data Collection. Joint Statistical Meeting of the American Statistical Association.* 31 July –5 August. Vancouver. Invited paper.

MÉNARD, M. [2008]: *The Implementation of Tools to Support the Data Quality of the Business Register at Statistics Canada*. http://www.oecd.org/dataoecd/13/12/41691288.pdf

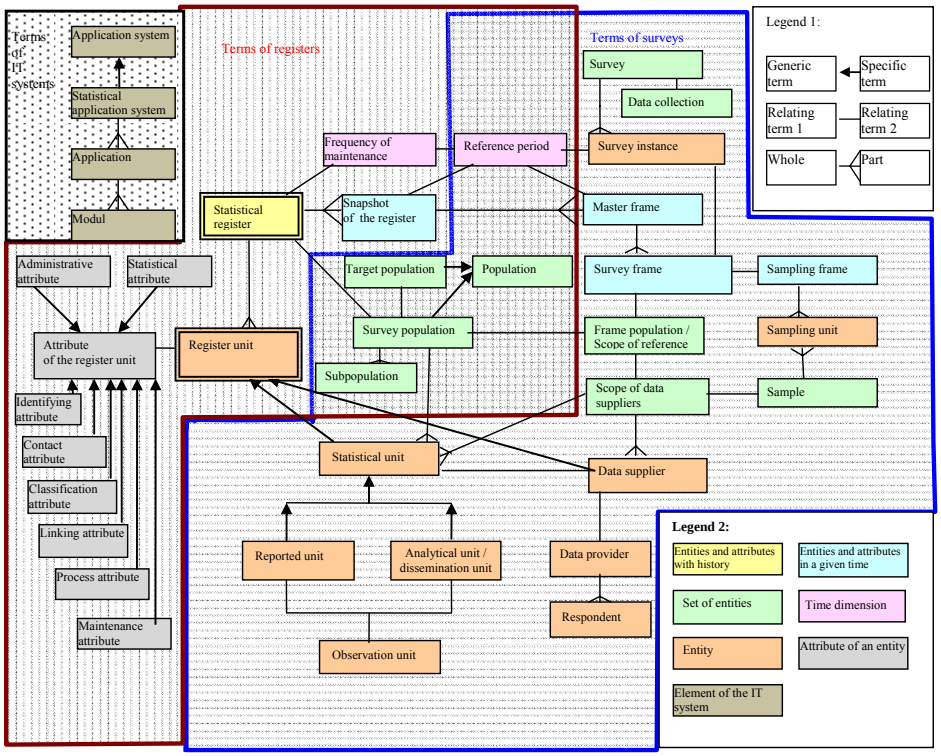
MÉSZÁROS, Z. [2005]: *Szociális regiszterek és a KSH monitorozás néhány kérdése.* http://www.allamreform.hu/letoltheto/szocialis_ugyek/hazai/Meszaros-Zoltan_Szocialis_regiszterek_es_a_KSH_monitorozas.pdf

STATISTICS CANADA [2011]: *Quality Guidelines – Survey Steps.* http://www.statcan.gc.ca/pub/12-539-x/steps-etapes/4147793-eng.htm

Appendix I

Terms of registers and survey control[5]

Relation of terms:



Administrative attribute: *Definition*: Administrative attribute of a register is a characteristic that can only be updated from administrative sources. Administrative attributes might change any time in accordance with the frequency of register maintenance. *Remark*: Administrative attributes cannot be modified, even if they are incorrect, however, reporting these errors toward the administrative sources is important. In some cases, correction of formal errors in the administrative data is allowed.

Analytical /dissemination unit: *Definition*: Analytical units represent real or artificially constructed units for which statistics are compiled. *Remark*: Analytical units are created by statisticians, often by splitting or combining observation units with the help of estimation and imputation.

[5] The aim of the relation of terms is to create a consistent set of concepts for the statistical data collection phase. The sources of terms are the Hungarian metadatabase, the RAMON Eurostat's Concepts and Definitions Database, the OECD Glossary of Statistical Terms and UN/ECE Terminology on Statistical Metadata. *Zoltán Vereczkei* and *Zsolt Kővári* helped the author edit the definitions.

The goal is to compile as detailed and homogeneous statistics as possible using data on observation units.

Application system: *Definition*: Application system is a logically related group of applications designed to perform a particular task.

Application: *Definition*: Application is a coherent group of functions supporting the maintenance, process or inquiry of data of a given phase of statistical processing. It is called on-line application if the functions are performed in real time and the navigation among functions is supported by a menu. The term of batch application is used if the application performs the series of functions in the background.

Attribute of register unit: *Definition*: Attribute of a register unit is a regularly updated characteristic of a register unit. *Remark*: Attributes of statistical register units can be arranged in groups. Accordingly, attributes referring to identification, contact, classification, demographic characteristics, relation to other register units, attributes supporting register maintenance and statistical processes (for example organization of data collection, sampling, etc.) can be defined. In respect of maintainability and changes of attributes over time, administrative and statistical attributes are distinguished.

Classification attribute: *Definition*: Classification attribute is an attribute supporting grouping of units by a given characteristic of the population. *Remark*: Typical classification attributes are NACE, classification of units by legal forms, size categories by the number of employees, attributes used for settlement description (county, region, resort area, etc.).

Contact attribute: *Definition*: Contact attributes are attributes supporting localization and accessibility of register units. Such attributes are the name, address, telephone number, e-mail address, etc. of a unit.

Data collection: *Definition*: Operation of statistical processing aimed at gathering statistical data and producing the input object data of a statistical survey.

Data provider: *Definition*: Data provider is the organ (for example the bookkeeper) or person authorized to report data in the name of the data supplier.

Data supplier: *Definition*: Data supplier is the unit of the frame population from which the data about the reporting and observation unit can be retrieved. The organ carrying out the statistical data collection is in legal relation with the data supplier. The data supplier is asked /obliged for providing data. *Remark*: In the majority of the surveys, the data supplier reports about itself, therefore the data supplier and the statistical units are the same. In other cases, the two terms are different, one data supplier accounts for one or more statistical units (for example an enterprise reports about its settlements, a local authority reports about its institutions).

Frame population / reference scope: *Definition*: Frame population (reference scope) is the set of population units described in the survey frame. *Remark*: The frame population (reference scope) is usually the same as the survey population. In absence of a direct register on the survey population, the register of units that are able to report about the object of the survey, serves as a base for reference scope.

Frequency of register maintenance: *Definition*: Frequency of register maintenance is the time interval of the register content alterations. *Remark*: Registers can be maintained from different sources with different frequencies. In such cases, the most frequently used source determines the frequency of the register maintenance.

Identifying attribute: *Definition*: Identifying attribute is a synonym of the unique identifier.

Linking attribute: *Definition*: Linking attribute is an identifier of another register unit that is in a sort of relation with the given register unit. *Remark*: The type of relation can be under or upper dependency, source of maintenance, etc.

Maintenance attribute: *Definition*: Maintenance attribute is an attribute of the register unit describing the date of registration, update, cause and source of maintenance, validity, etc.

Master survey frame: *Definition*: Master survey frame is a snapshot of a register (union of registers) to assign the survey frames based on the given register (registers). *Remark*: An example of the master survey frame is the snapshot of the business register to define the survey frames of different economic statistical data collections. Another example can be the snapshot of the address register to make a common frame for population surveys. The common master survey frame, the common reference period helps the integration and linking of statistical data coming from different surveys.

Module: *Definition*: Module is a logical unit of the application created to perform a function, a given part of a task.

Observation unit: *Definition*: Observation units are the entities for which information is received. *Remark*: During data collection, this is the unit for which data is recorded. It should be noted that this may or may not be the same as the reported unit (the reported unit is, for example, the settlement, while the observation unit is the product being produced there).

Population: *Definition*: Population is the total membership or "universe" of a defined class of people, objects or events. *Remark*: Specific population definitions are target population and survey population. Target population is also known as the scope of the survey and survey population is also called as the coverage of the survey.

Process attribute: *Definition*: Process attribute is an attribute of the register unit describing a function, characteristic of a population unit in the statistical working process. *Remark*: Typical examples of process attributes are the survey related characteristics of institutions, such as their role in the population, their willingness to provide data, etc.

Reference period: *Definition*: Reference period is a time interval or a date to which the observed attribute (variable, indicator, measure) refers. *Remark*: Not only variables but data collections and their survey frames have reference period. The reference period of data collections are in accordance with the reference period of the observed variables (for example monthly data collection usually refers to indicators that reflect monthly or end-of-the-month situations of a phenomenon). Reference period of the data collection determines the reference period of its survey frame as well. The reference period of a survey frame is related to the date of the register snapshot (for example the first day of the month for monthly data collections).

Register unit: *Definition*: Register unit is the unit, entity of the register population with related descriptive information on identification, accessibility and other attributes. *Remark*: Register unit type – that is the collection of a given type of individual units – and register unit instance – that is a concrete, individual register unit – are distinguished. In the surveying process, data processing and dissemination phases, register units might function as data supplier, data provider or statistical (reporting, observation, analytical, dissemination) units.

Reported unit: *Definition*: Reported unit – or with other name, accounting unit – is the statistical unit about which information is sought. The data supplier accounts for as many reported units

as were assigned to it from the survey frame or belong to it on the base of the observed activity, phenomenon. If the data supplier has more than one reported unit, it has to fill more questionnaires. If an enterprise is the data supplier and reports about its settlements, then the settlement is the reported unit. *Remark:* In some other terminologies the reported unit means the organisation who submits the data in the name of the data supplier. In our terminology that term is called data provider.

Respondent. *Definition*: Respondent is the person who completes or approves the questionnaire what provides information about the reported unit.

Sample: *Definition*: Sample is the set of sampling units selected from the sampling frame. *Remark*: The aim of the observation based on a sample is to estimate the attributes for the whole population without the observation of the whole population.

Sampling frame: *Definition*: Sampling frame is an information-set for the survey population or the attributes of stratification being used as a basis for sample selection and in subsequent estimation procedures. *Remark*: Sampling frame often corresponds to the survey frame. In multistage sampling, the sampling frame of the primary and secondary stages may be different. In that case the sampling units of the secondary stage belong to the selected units of the primary sampling frame. For example the first step of sampling selects the settlements involved in the sample and the second one chooses the addresses in the selected settlements.

Sampling unit: *Definition*: Sampling unit is the entity or a group of entities of the sampling frame from what selection is made at a given stage of sampling.

Scope of data suppliers: *Definition*: Scope of the data supplier is the set of entities of the frame population assigned for data reporting from which data can be retrieved for the investigated population (statistical and observation units). *Remark*: In full scope data collection, the scope of data suppliers corresponds to the frame population. In representative or combined data collection, the scope of data suppliers is only a part of the frame population. It doesn't contain the statistical units of the frame population not selected into the sample.

Snapshot of register: *Definition*: Snapshot of a register is its frozen state on a given date. Instead of a register, snapshots are used for statistical processing because, unlike register units (that can be updated frequently), population units and their attributes must be constant during data collection and statistical processing.

Statistical application system: *Definition*: Statistical application system is an integrated group of applications supporting the given phase of statistical processing. It performs the maintenance, process and inquiry of data of the particular phase and gives interfaces to the relating applications, application systems.

Statistical attribute: *Definition*: Statistical attributes of a register are all non-administrative attributes of the register unit used in the statistical working process. Statistical attributes can be updated, corrected any time, irrespective of the administrative sources. For comparability reasons, the classification attributes remain unchanged within a year.

Statistical Register: *Definition*: Register is a continuously updated set of objects for a given population containing information on identification, accessibility of population units and other attributes, supporting the surveying process of the population. The register contains the current and historical statuses of the population and the causes, effects and sources of alterations in the population. Register data of population units are stored in a structured database. *Remark:* Statistical regis-

try is a less complex version of statistical register. Logging, change and history management are not mandatory in statistical registries.

Statistical unit: *Definition*: Statistical units are units of observation for which data are collected or derived. *Remark*: Statistical units observed during data collection are called reported units. In the analytical and dissemination phases, the terms of analytical or dissemination units are used.

Subpopulation: *Definition*: Subpopulation is a subset of a population. *Remark*: Subpopulation refers to populations that require different handling in the statistical working process. Subpopulations are usually specified to understand the distinguishing characteristics of these populations.

Survey frame: *Definition*: Survey frame is the set of survey population units together with their attributes referring to a given reference period. *Remark*: Survey frame describes the identification, contact, classification attributes of the survey population units for a given reference period.

Survey instance: *Definition*: Survey instance is a particular survey and reference period in which data are collected from respondents.

Survey population: *Definition*: Survey population is the population for which information during the survey process can be obtained. *Remark*: Concurrence or difference of survey and target populations is measured by coverage. In HCSO terminology, the survey population is called reference scope.

Survey: *Definition*: Survey is an investigation on the characteristics of a given population by means of collecting data (including censuses, full scope and sample surveys, usage of administrative sources) and estimating their characteristics through the systematic use of statistical methodology.

Target population: *Definition*: Target population is the set of units about which information is wanted and estimates are required. *Remark*: Synonym of survey scope.

Appendix II

The topics of the survey control metadata

Chapter A – General description

1. Identification of the data collection/ administrative source:
- – code of the program elements of the planning system;
- – OSAP identification number of the data collection and data source;
- – its name;
- – year of the reference period, frequency;

2. Identifier of a particular subpopulation, the reason for creating subpopulations;

3. Type of the data supplier unit;

4. Type of the statistical unit;

5. Mode of questionnaire printing;

6. Mode of personalization;

7. Guides, amendments to the questionnaire, and the rule of their maintenance;

8. The way of mailing the questionnaires.

Chapter B – Description of the population (survey frame)

The description of the population (or a certain subpopulation) is built on and specified from elementary subsets. The whole population is the union of the elementary subsets. There is no overlap between the subsets.

An elementary subset can be described by

1. referring to an existing, predefined subset or
2. specifying a new subset by:
– the identifier of the subset;
– the observation mode of the subset (full-scope, representative);
– the way of maintaining the changes of organizations;
– the operating status of the units of the subset;
– the algorithm to assign the elements of the subset based on the attributes of the master frame (NACE, the size category of the number of employees, sector code);
– the condition referring to the observed activity;
– response/ non-response in the previous reference period;
– definition of the frame (or a part of the frame) from external source, the way of maintaining the source list of statistical units.

Chapter E – Mailing, registering the questionnaires

1. Periods and time of mailing;
2. Characteristics of the composition of questionnaire sets and blocks;
3. Attributes of data collection and validation:
a) identifier and contact person of the organization unit responsible for data collection;
b) organization unit responsible for data entry;
c) organization units responsible for different validation and editing steps;
d) the fact whether the questionnaire was involved in electronic data collection;
4. Response deadlines.

Chapter L, F – Assigning the survey frame from external source

This chapter is filled in when the assignment of the survey frame of a data collection is built on other sources different from the master frame. Here the following have to be described:

1. The fact whether the information from another source aims at
a) giving new units of the survey frame or
b) disabling the assigned units of survey frame;
2. Possibility of common use of the list by other data collections;
3. Way of maintaining the changes of the organizations;
4. Priorities among the different sources of the survey frame;
5. Frequency of the external list;

6. Specification of selection if the list comes from the processing phase of another data collection.

Chapter R – Sampling in the case of representative data collection

The chapter is filled in when data collection has a not full-scope subpopulation. In this case, the chapter describes:

1. whether sampling forms part of GÉSA or the sample comes from outside the system;

2. the relation between the master frame and sampling frame, the way of sampling;

3. in the case of a common sample, the list of data collections using it;

4. the operating status of units;

5. the condition and frequency of sample refreshment.

Appendix III

Information on the survey frame and data supply (paradata)

1. Information on the survey frame

a) Identifier of the data supplier;

b) Identifier of the statistical unit;

c) Identifier of the reference period of data collection;

d) Identifier of the particular subpopulation;

e) Identifier of the organization unit responsible for data collection;

f) Way of observation (full-scope or representative observation, deletion from the observation);

g) Strength of belonging to the population (voluntary or possible);

h) Belonging to the scope of data suppliers (whether the unit is part of the sample or only an element of the survey population);

i) Number of expected questionnaires (if the level of observation for the questionnaire is "deeper" than that for the statistical unit);

j) Source of getting into the survey frame;

k) Date of getting into the survey frame (in the case of follow-up recording);

l) Date of deletion, modification (in the case of follow-up modification);

m) Identifier of the user modifying the survey frame.

2. Information on the mailing procedure

a) State of mailing of the questionnaire;

b) Way of mailing (by post, web, e-mail, etc.).

3. Information on data supply

a) Arrival of the questionnaire (arrived, not arrived);
b) Way of data supply (post, web, e-mail, etc.);
c) Type of response (non-response, response with data, response without data (negative response), assignment for imputation, imputation by a given method);
d) Date of arrival;
e) Identifier of the user registering the arrival;
f) Reason of non-response or negative response;
g) Type of last urging (phone, reminder, warning, e-mail, fax, letter, urging e-mail, fax, letter, call-up letter, offence);
h) Date of last urging;
i) Identifier of the person urging data suppliers.

4. Comment on response or non-response
5. Filling time of the questionnaire
6. Contact data of the person

a) completing the questionnaire by chapters (name, phone, e-mail, identifier of the agency representing the respondent);
b) responsible for approval (name, phone, e-mail, post).

7. Log of the events of data collections

a) Identifier of the event (registry of the questionnaire, registry of response/ non-response, different phases of urging, etc.);
b) Date of the event;
c) User of the event.

8. Demand for exemption from data supply

Living Better, Living Longer? Is Ageing in Line with Economic Performance?*

Erzsébet Kovács
Professor
Corvinus University
of Budapest
E-mail: erzsebet.kovacs@uni-corvinus.hu

The goal of this paper is twofold. First, it is to describe and interpret the increase of life expectancy as a positive component of the human development index and the global competitiveness index. For developed countries, age-gain, defined as the difference between life expectancy at age 65 and at birth, is calculated and analyzed.

The second part of the paper is about Hungary which is a developed country according to the human development index and is in catching-up position according to the rankings of the World Economic Forum. Hungarian age-gain is one of the highest, it is above seven years and its old-age dependency ratio is fast increasing accompanied by a relatively low economic performance. To sustain the present Hungarian pay-as-you-go pension system, reliable projection for the ageing tendency is indispensable.

KEYWORDS:
Longevity.
Public Finance.
Pension.

* The research was supported by the Hungarian Scientific Research Grant (OTKA K 77420). An earlier version of this paper was given at the 71st International Atlantic Economic Conference, in Athens, on 17th March 2011.

Economic and mortality modelling have a very long but independent history. Measurement of economic performance was limited to the components of GDP, and GDP per capita was used as a standard statistical indicator of economic progress. Well-being is more than just income. It is a multi-dimensional concept with several aspects. Some of them are measurable, like income or education, and others are latent, for example protecting the environment Longevity in ageing societies is strongly associated with material well-being. Increasing life expectancy is a result of higher economic performance and but it also can moderate further economic growth.

Hundreds of economic and econometric studies tried to understand what determines the wealth of nations. The first attempt originates from *Adam Smith* who focused on specialization and the division of labour.[1] Neoclassic economists emphasize investment in physical capital and infrastructure. More recently, health, education, and training are listed as vital components of a country's competitiveness and productivity. Investments in education and health services are crucial for clear economic considerations. The *Commission of the European Communities* [2009] issued a paper entitled "GDP and Beyond" underlining the strategic importance of this topic.

In this paper, based on available data, the human development index and the global competitiveness index are compared with ageing in highly developed countries.

1. Three complex measures for similar purpose

There are different theories and opinions in the literature concerning the measures of economic development and well-being. According to several researchers, personal income is one of the key factors, and both low and high level of income inequalities may reduce the perception of well-being (*Cserháti–Takács* [2010]).

In the following, we concentrate only on indices including one way or another increasing life expectancy as a positive factor.

Human development index (HDI) has been used since 1990 and it was the first attempt to incorporate different aspects of quality of life. It was modified and redefined in 2010.[2]

[1] His one of the most influential work "An Inquiry into the Nature and Causes of the Wealth of Nations" was published in 1776. (See References.)

[2] Details can be found: http://hdr.undp.org/en/media/HDR_2010_EN_TechNotes_reprint.pdf

This composite index published by UN aggregates three dimensions: *1*. a long and healthy life, measured by life expectancy at birth; *2*. access to knowledge, combining mean and expected years of schooling; *3*. a decent standard of living, measured by GNI per capita (PPP[3] US$).

Higher value means better well being for nations. Norway has got the highest value (0.938) in 2010. Values above 0.788 are classified as very high human development. Next category is called high human development, whose lower limit is 0.67. This paper concentrates only on the analysis of the OECD member states with HDI above 0.67.

The World Economic Forum's annual competitiveness reports are examining 12 pillars to compute the global competitiveness index (GCI). This comprehensive index involves static and dynamic components as a weighted average of many different aspects. GCI combines survey data and hard data to capture microeconomic and macroeconomic foundations of national competitiveness through these pillars. Using all of the data, ranks and scores are published year by year.

The fourth pillar of GCI, as the measure of health and primary education is very important in our analysis. Higher education and training in the fifth pillar are strongly influencing life expectancy as well.[4] Sustainable economic growth cannot be achieved unless the other pillars are stable and efficient. Table 1 summarizes the 12 pillars.

Comparing the HDI as the geometric mean of its three components with the GCI compiling the formerly mentioned pillars, there are no statistically significant differences as it is shown in Figure 1. Higher HDI is followed by higher GCI. In spite of the non-linear relationship, the Pearson correlation coefficient is 0.69 ($p = 0.000$) and the Spearman rank correlation value is 0.658 ($p = 0.000$) for OECD countries ($n = 30$) in 2010. This is due to the overlapping content of the two indices.

Hungary is less developed and less competitive than the OECD's average, and is close to the virtual break point at GCI = 4.4 and HDI = 0.84 in Figure 1. The existence of the break point can be explained by the different content of the two measures. It is easier to gain higher HDI score with higher value in the three components, and more effort is needed to earn higher GCI point. *Zádor* and *Gáspár* [2010] came to a similar conclusion. Hungary is a highly developed country according to the HDI and is in a catching-up position according to the rankings of the World Economic Forum. Descriptive statistical measures of these two indices are presented in Table 2. The coefficient of variation of GCI is greater because of containing more information through the 12 pillars. HDI values have high kurtosis because OECD countries are concentrated around the mean. Both indices are skewed to left underlining the tendency that higher values are more frequent than low values in the OECD.

[3] Purchasing power parity.

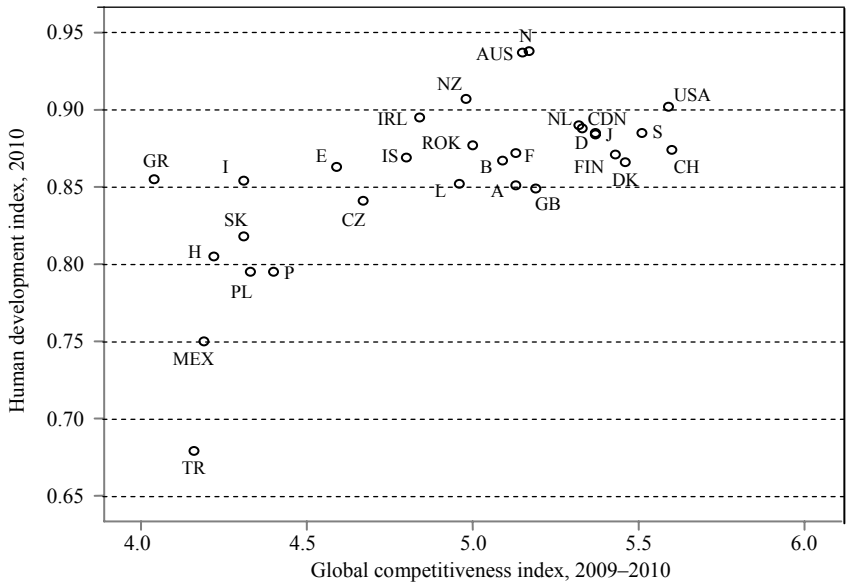
[4] List of variables combined in the fourth and fifth pillars are given in the Appendix.

Table 1

The 12 pillars of competitiveness

Name	Content
1. Institutions	Institutional environment, legal and administrative framework
2. Infrastructure	Infrastructure networks, transport, location of economic activity, telecommunications network
3. Macroeconomic stability	Fiscal deficit, inflation rate, public accounts managed by the government
4. Health and primary education	Healthy workforce and quantity and quality of basic education
5. Higher education and training	Well-educated workers, secondary and tertiary enrolment rates
6. Goods market efficiency	Supply and demand conditions, market competition, taxes
7. Labour market efficiency	Labour market flexibility, wage fluctuation, worker incentives
8. Financial market sophistication	Well-functioning financial sector, proper assessment of risk, regulation
9. Technological readiness	Access to information and communication technologies
10. Market size	Domestic and foreign market
11. Business sophistication	Quality of overall business networks and individual firms' operations and strategies
12. Innovation	Investment in research and development

Figure 1. HDI and GCI scores of OECD countries in 2010



Source: Author's calculation. In Figures 1–4 the international car codes are used.

Table 2

Statistics of HDI and GCI

Denomination	Human development index 2010	Global competitiveness index 2009–2010
Valid (sample size)	30	30
Mean	0.857	4.921
Median	0.868	5.045
Standard deviation	0.053	0.484
Skewness	–1.553	–0.403
Standard error of skewness	0.427	0.427
Kurtosis	3.754	–1.177
Standard error of kurtosis	0.833	0.833
Range	0.259	1.560
Minimum	0.679	4.040
Maximum	0.938	5.600

Source: Author's calculation.

Sustainable development in the ageing societies is investigated by several researchers and from different aspects in the European Union. A new complex measure in this field was proposed by *Stiglitz*, *Sen* and *Fitoussi* [2009]. Their paper, the so-called Stiglitz Report suggests shifting the emphasis from measuring economic performance to measuring people's well-being. The authors of this report propose eight dimensions to be taken into account simultaneously to represent the quality of life in different countries. Five from these eight dimensions (material living standard, health, education, personal activities (work), and economic insecurity (among them inadequate resources during retirement and volatility in pension payments)) are strongly connected to the problems of an ageing society.

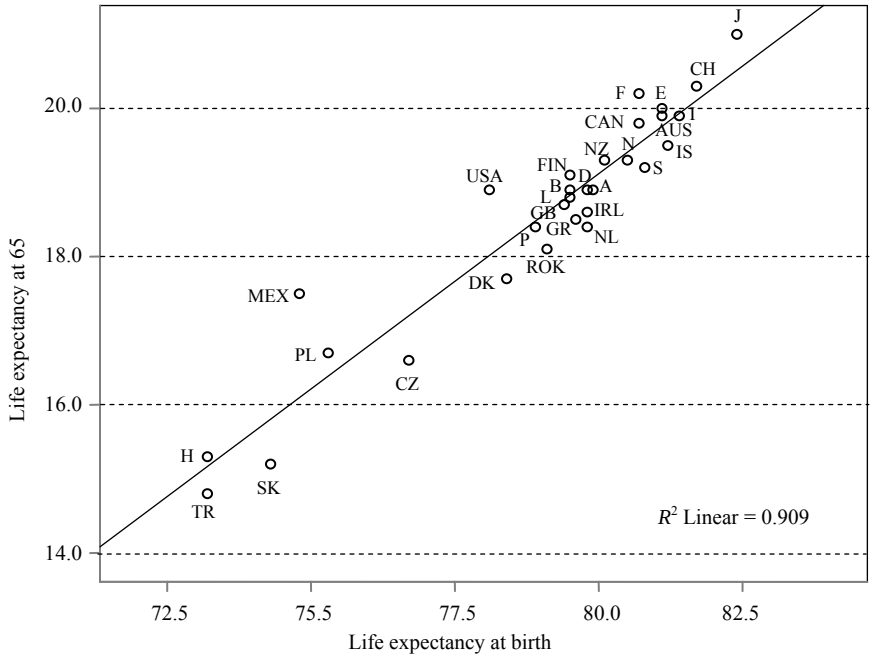
All three complex measures take increasing life expectancy as a positive result of the economic and social progress. However, consequences of ageing for the society have not been quantified in these measures. Longevity will have both positive and negative impact on developed nations. One strong evidence can be mentioned here as an example. Researchers predict very high proportion of health expenditure for individuals and for the society in the next decades. It is in line with the Pareto law: 80 percent of all health costs will be spent in the last 20 percent of our lifetime.

2. Measuring life expectancy at different ages

Life expectancy is a part of all complex measures. Forecasting life expectancy and modelling[5] mortality are key problems not only for economists but also for actuaries and demographers. Most methods of mortality forecasting are extrapolative and give aggregate measures, such as life expectancy at birth. It is only in the last 20–25 years that more sophisticated stochastic methods have been developed and applied. There are different mortality measures, they may refer to overall mortality or be decomposed by sex, socio-economic factors or cause of death. These components are strongly correlated over countries and time.

Life expectancy can be calculated not only at birth but also at different ages. It is higher and higher as age is increasing. Figure 2 shows this tendency in the OECD countries. Life expectancy at 65 has got crucial interest as 65 is the official retirement age in most of the developed countries.

Figure 2. Life expectancy at birth and at 65



Source: Author's calculation from OECD data.

[5] *Gompertz* published his law on mortality in 1825. For further details on mortality modelling see *Lee–Carter* [1992] or *Májer–Kovács* [2011].

Mortality improvement can be measured through increase in life expectancy. The median age at birth in the OECD member states is around 80 years. Under the condition that a person has reached the age 65, he/she can expect to live 19 more years. "Age-gain" is defined as the difference between life expectancy at 65 and life expectancy at birth for the same calendar year and population. This is a cross-sectional measure. The average age-gain is 4.5 year in the OECD countries. Hungary is the second on this list with 7.1 year. Table 3 shows the basic statistical results.

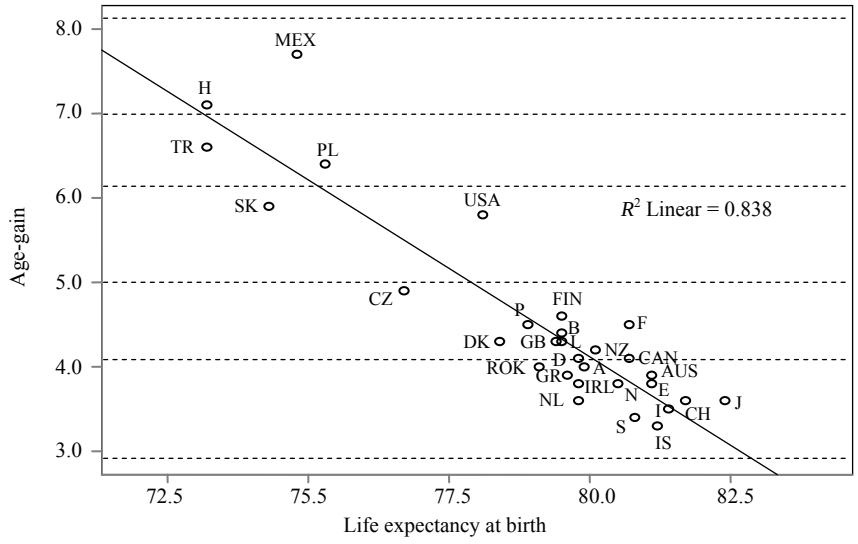
Table 3

Summary statistics of life expectancy

Denomination	Life expectancy at birth	Life expectancy at 65	Age-gain
Valid (sample size)	30	30	30
Mean	79.02	18.55	4.53
Median	79.70	18.90	4.15
Standard deviation	2.50	1.53	1.15
Minimum	73.20	14.8	3.30
Maximum	82.40	21.0	7.70

Source: Author's calculation from OECD data.

Figure 3. Determination of age-gain by life expectancy

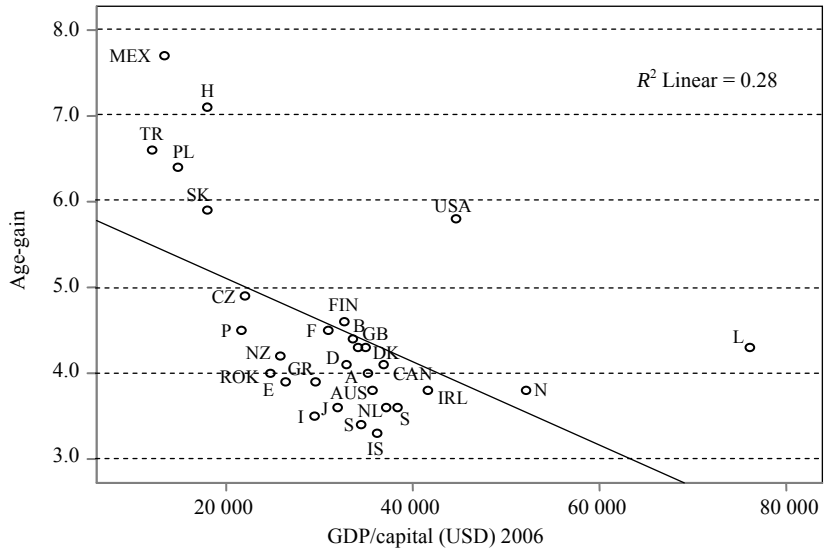


Source: Author's calculation from OECD data.

Higher life expectancy at birth is related to higher life expectancy at 65. But age-gain is decreasing with increasing life expectancy, as it can be seen in Figure 3. The coefficient of determination is 84 percent, so the unexplained variance is 16 percent. It could be interesting to analyze what else can explain the mortality improvement in highly developed countries.

Expected length of a person's life can be estimated in the process of underwriting by knowing individual health conditions and the level of education. These non-monetary measures can significantly increase people's life expectancy. GDP per capita as a proxy of economic performance explains only 28 percent of the age-gain variance. This is statistically significant but they are only weakly correlated. For details, see Figure 4. The level of education measured by the expected number of years of schooling and the sum of public and private health expenditure in absolute number or as a percent of GDP can be used to estimate age-gain more precisely from cross-sectional data. Age-gains in the case of Mexico and Hungary are the highest in the OECD.

Figure 4. Determination of age-gain by GDP per capita



Source: Author's calculation from OECD data.

Factor analysis, one of the multidimensional statistical methods is best suited for checking interactions and correlations among conventional economic indicators, like GDP and other complex indicators as it can be seen in Table 4.

Using six variables for 30 OECD countries, one factor can be identified with 69 percent of total information. All correlation coefficients are positive and significant in Table 4 (Component A). Instead of life expectancies, age-gain is taken into ac-

count in the second part (Component B). Here again one factor is calculated, but the sign of the correlation is negative, since increase of life expectancy becomes smaller at a higher level of education, health, innovation and GDP.

Table 4

*Two-component matrices**

Variables	Component (A)	Component (B)
Life expectancy at birth	0.892	–
Health and primary education	0.883	0.905
Life expectancy at 65	0.854	–
Innovation and sophistication factors	0.841	0.855
Higher education and training	0.829	0.888
GDP/capital (USD) 2006	0.674	0.679
Age-gain	–	–0.805

* Principal component analysis was used as an extraction method.
Source: Author's calculation from OECD data.

From the actuarial point of view, this factor describes the collective risk of ageing for countries. Slow economic development of the country can be associated with fast catch-up in life expectancy as health conditions are improving. Age-gain becomes more expensive at higher life expectancy. The situation cannot be summarized simply as "smaller result at higher cost". Additional effects are associated with this tendency. Long-term care system should be created and financed as old-age dependency ratio[6] is increasing. There is a worldwide problem with no best practice for benchmarking. Higher effective age of retirement or increasing contribution is needed to finance pension benefits in order to sustain pension systems.

3. Increasing life expectancy in Hungary

Hungarian life expectancy is among the lowest in Europe. Hungary will face the risk of the ageing society a bit later, mainly in the next decades. There are two reasons for this time lag. On the one hand, the baby boom started somewhat later, in the early 1950s and there was a second wave in the late 1970s. For details, see Figure 5.

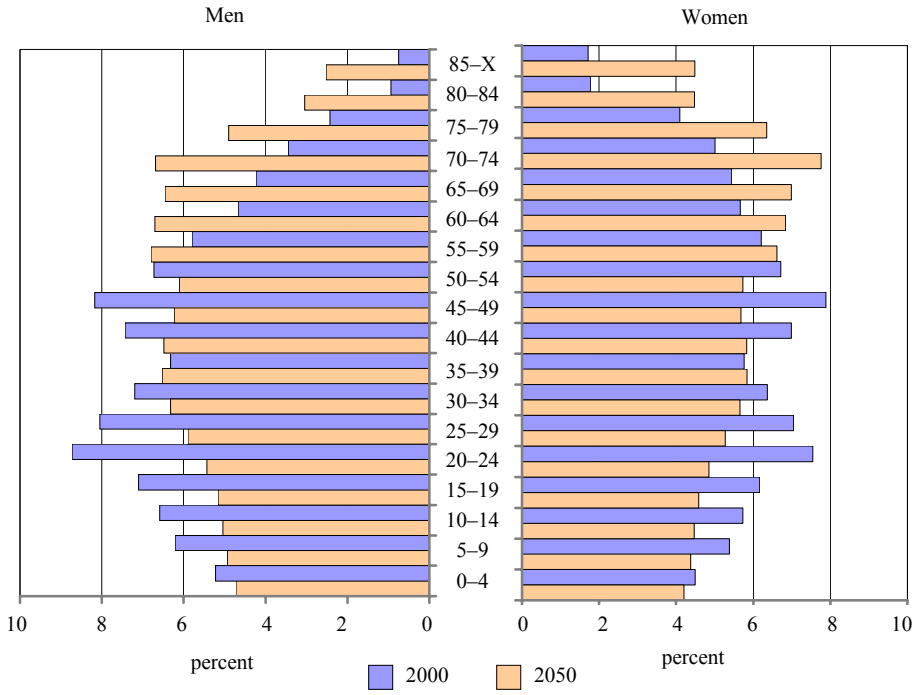
[6] The old-age dependency ratio is the ratio of the number of people aged 65 and over to the working-age population (those aged 15–64). Projected ratios are published in http://www.euphix.org/object_document/o5117n27112.html

On the other hand, the Hungarian male life expectancy is one of the lowest in the OECD. Life expectancy for 65 year-old men is 78.7, while 65 year-old women are expected to live four additional years compared to men.

Two tendencies meet in Hungary in this decade: *1*. fertility rate is below 1.3.; *2*. demographers expect a fast catch-up in life expectancy. Therefore the old-age dependency ratio is increasing at a high speed. Our pay-as-you-go pension system needs reliable projections for the ageing tendency.

Hungarian population by age and gender in 2000 and in 2050 (in percentage of the total population in each group) can be seen in Figure 5. The healthy triangle shape is not yet visible at present. But the picture of the future is more challenging. Dramatic change in old-age dependency is foreseen: it is expected to double in this 50-year-long period.[7]

Figure 5. Hungarian population in 2000 and 2050



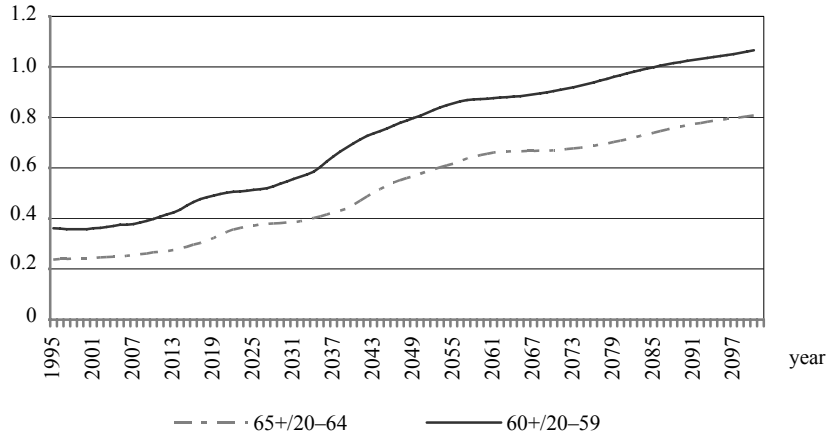
Note. The total population was 10.2 million in 2000 and is expected to be 8.7 million in 2050. The old-age dependency ratio was 0.24 in 2000 and will be around 0.5 in 2050.

Source: OECD Population Pyramids. (www.oecd.org/dataoecd/52/31/38123085.xls)

[7] The old-age dependency ratio is expected to be at least twice higher for most of the OECD member states in the investigated period. The increase comparing 2050/2000 for Japan will be 2.6, and the greatest growth rate is predicted for Mexico: 3.9.

Here is again a special Hungarian problem, which becomes clear when comparing Figure 6 and 7. The official age of retirement is 62 for both genders, but the effective age of retirement is less than 60 years. According to Figure 6, without any dramatic change in the economic activity, two active persons will finance one pension in 2020, that is, the solid line which shows the ratio of people above 60 to persons who are between 20 and 59 years, will reach 0.5 at that point. The lower dotted line shows the ratio of people above 65 to persons who are between 20 and 64 years.

Figure 6. Hungarian old-age dependency ratio



Source: Author's calculation from demographic prognosis elaborated by *László Hablicsek* [2010].

It has been decided to increase the official pension age to 65 within the next 6 years. Supposing real changes in this field, this problem can be postponed by 20 years.

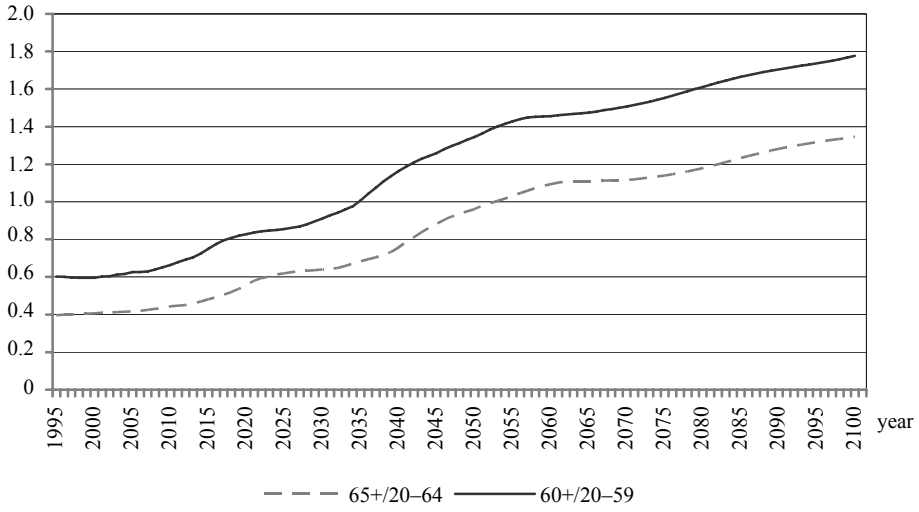
Old-age dependency ratio as a statistical measure should be adjusted for Hungary. Its numerator is higher because of early retirement. The denominator should be lower because of the grey economy: only 56 percent of the active population pays pension contributions regularly. Figure 7 presents a more realistic picture for the next period. One "contributor" will finance one pension in 2035 if the effective retirement age and the activity rate will not become significantly higher.

Report of the Pension and Old-Age Round Table (*Holtzer* [2010]) on its activities highlighted several problems and suggested different paradigms to support decision-making.[8] Appendix 6 of this report presented a demographic forecast with three scenarios. The basic version – based on low fertility rate and average migration – pro-

[8] Nineteen experts – among them the author of this paper – worked together between 2007 and 2009. The report covers projections for the present pension system and compares five possible pension schemes. The demographic prognosis was elaborated by László Hablicsek with the *component method,* taking into account childbirths (fertility), life expectancy (mortality), and international migration.

jected one-year increase in life expectancy in 5-6 years. Can we accept this for a constant growth rate?

Figure 7. Adjusted old-age dependency ratio



Source: Author's calculation from the demographic prognosis elaborated by László Hablicsek.

Reliable prediction of increasing life expectancy is very important for Hungary for several reasons: *1*. Health care, disability, and old-age pension reforms are all hot issues. *2*. Increase in life expectancy can be faster than the economic catch-up process. However, this tendency cannot be modelled by time series methods. *3*. New technology in medicine and genetics, innovations in pharmaceutics appear and cause revolutionary changes in life expectancy in the global world.

4. Longevity risk in the Hungarian pension system

The significance of life expectancy was usually underestimated in the past decades, while previously mortality probabilities were treated as deterministic input variables for life tables. Annuity providers, pension schemes must give annuities or pension benefits for a longer period than it was calculated earlier. As baby boomers retire, decumulation of wealth and longevity risk will become key issues. Instead of constant mortality, decreasing rates are preferred to be considered for the future.

The defined benefit (DB) pension system[9] introduced after the Second World War and extended in the 1960s is planned to change and will be transformed to a defined contribution (DC) scheme[10]. The DC system gives a chance for the government to balance the pension budget year by year. This idea is very similar to the German point system and the Swedish non-financial defined contribution (NDC) system. Both are pay-as-you go pension systems by their financial basis and are combined with individual accounts. To calculate monthly benefits, projected future life expectancy should be considered.

Based on experience, using a deterministic model, 15.4 years can be forecasted for a 65 year-old Hungarian in 2007. Unisex mortality data are used for pension calculations because of the EU principle of equal treatment between men and women.[11]

Estimation of increase in life expectancy requires age-specific mortality statistics, which are modelled by three factors: age, period, and cohort. Mortality reduction is of stochastic nature. More precisely, the improvements in mortality are viewed as stochastic processes, and cannot be projected by simple linear regression models.

The difficulty of projecting longevity risk lies in its complexity. It is composed of three underlying risks:

– modelling risk (due to limited availability of data, census is taken once in a decade);
– trend risk (changes in socio-economic environment or health care can significantly increase longevity); and
– idiosyncratic risk (random error).

Idiosyncratic and modelling risks can be managed in large populations. To quantify the trend risk, various mortality forecasting models have been proposed. One of the most popular extrapolative models is the Lee–Carter model introduced by *Lee* and *Carter* in 1992. It has several variations and extensions. In the original model: $\ln m_{x,t}^{(g)} = \alpha_x^{(g)} + \beta_x^{(g)}\kappa_t^{(g)} + \varepsilon_{x,t}^{(g)}$, where $m_{x,t}^{(g)}$ is the mortality rate of an x year old person at time t, $\alpha_x^{(g)}$, $\beta_x^{(g)}$, and $\kappa_t^{(g)}$ are estimated parameters, and $\varepsilon_{x,t}^{(g)}$ is the error term. Logarithm of the mortality rates can be efficiently projected by ARIMA models (*Baran et al*. [2007]).

[9] A pension scheme where the benefits accrued are linked to earnings and the employment career (the future pension benefit is pre-defined and promised to the member).

[10] A pension scheme where the level of contributions, and not the final benefit, is pre-defined: no final pension promise is made.

[11] For details on equal gender treatment visit the EU website or see Council Directive 2004/113/EC of 13 December, implementing the principle of equal treatment between men and women in the access to and supply of goods and services (http://curia.europa.eu/jcms/upload/docs/application/pdf/2011-03/cp110012en.pdf)

Based on the Lee-Carter model age-specific mortality statistics for a 65 year-old Hungarian was estimated by ARIMA(0,1,0) in the study of *Májer* and *Kovács* [2011]. This model gives 16.43 years as expected value and has a very high determination coefficient R^2= 92%. The lower bound is 15.35 years while the upper bound is 17.6 years at a 95 percent confidence level in 2007. The confidence interval keeps getting wider in the future as uncertainty grows.

This stochastic projection shows *one year extra life expectancy* for pensioners without taking into account any possible changes in the socio-economic environment or health care.

Annual budget of the state pension fund is around 12 percent of our GDP. This amount exceeds the total pension budget in most of the developed countries. The Hungarian pension system has no reserves to help finance a longer pension period. It got additional financial support from the budget in the last decade and this was one of the components in the high deficit. Stopping this process was among others a reason why the government weakened the second – mandatory private – pillar in 2010.[12]

5. Conclusion

Uncertainties in life expectancy estimations are significantly influencing the financial stability of pension systems. Pay-as-you-go pension systems strongly need reliable projections for the ageing tendency. Hungary will face the risk of an ageing society in the next decades. This period is characterized by lower GDP growth rate and moderate increase of productivity. Hungarian macro models and forecasts calculate with higher pension payments while politics keeps promising to maintain the present pension level. These projections underestimate future benefits without adding longevity to the model. It would be wise to link pension age to life expectancy as it is done in the Danish pension system.

Consequences of longer age are hidden using defined benefit pension rules. At retirement, calculation by the defined contribution principle reacts on increasing life expectancy year-by-year. The result should be lower pension or longer contribution period if we do not want to put this financial burden on the shoulders of the next generations.

We can observe two tendencies. Ageing is faster in highly developed countries, because more money is spent on education and health. At the same time, age-gain

[12] Only 100 000 people from the 3 million members of the second pillar insisted to remain in the private scheme knowing the new regulation introduced in 2010. The accumulated wealth from the second pillar is sold and used to reduce high deficit.

decreases as life expectancy grows. Poor economic performance and high unemployment rate after the recent financial crisis does not provide enough resources to support the pension system if it remains unchanged. This is also the lesson for Hungary. Our economic development has slowed down due to the crisis but life expectancy and the old-age dependency ratio keep increasing continuously. Longevity is going to be a dominating issue in the future development of the Hungarian society and poses several questions calling for further research.

References

BARAN, S. – GÁLL, J. – ISPÁNY, M. – PAP, G. [2007]: Forecasting Hungarian Mortality Rates Using the Lee–Carter Method. *Acta Oeconomica*. Vol. 57. No. 1. pp. 21–34.

BOOTH, L. – TICKLE, L. [2008]: Mortality Modelling and Forecasting: A Review of Methods. *Annals of Actuarial Science*. Vol. 3. No. 1–2. pp. 3–43.

COMMISSION OF THE EUROPEAN COMMUNITIES [2009]: *GDP and Beyond. Measuring Progress in a Changing World*. Brussels. http://www.beyond-gdp.eu/

CSERHÁTI, I. – TAKÁCS, T. [2010]: Analysis of Income Inequalities by Microsimulation. *Hungarian Statistical Review*. Vol. 88. Special No. 14. pp. 110–124.

GOMPERTZ, B. [1825]: On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of Royal Society,* (*Series A*). Vol. 115. pp. 513–583.

HABLICSEK, L. [2010]: Socio-demographic preliminary calculations to model the reconstruction of the pension system. Enclosure 6. In: *Holtzer, P*. (ed.): *Report of the Pension and Old-Age Round Table on Its Activities*. Prime Minister's Office. Budapest. http://econ.core.hu /file/download/nyika/jelentes_hu.pdf

HOLTZER, P. (ed.) [2010]: *Report of the Pension and Old-Age Round Table on Its Activities*. Prime Minister's Office. Budapest. http://econ.core.hu/file/download/nyika/jelentes_hu.pdf

KOVÁCS, E. [2010]: A nyugdíjreform demográfiai korlátai. *Hitelintézeti Szemle*. Vol. 9. No. 2. pp. 128–149. http://www.bankszovetseg.hu/anyag/feltoltott/HSz2_128_149ig.pdf

LEE, R. D. – CARTER, L. R. [1992]: Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association.* Vol. 87. No. 419. pp. 659–671.

MÁJER, I. – KOVÁCS, E. [2011]: Élettartam-kockázat – a nyugdíjrendszerre nehezedő egyik teher. *Statisztikai Szemle*. Vol. 89. No. 7–8. pp. 790–812.

OECD (ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT) [2006]: *Live Longer, Work Longer*: *A Synthesis Report*. OECD Publishing. Paris. www.oecd.org/els/ employment/olderworkers

OECD [2007]: *Society at a Glance 2006*: *OECD Social Indicators*. Demographic and Labour Force database. (www.oecd.org/els/social/indicators/SAG (See indicators GE2 Age dependency ratio); Ageing: Population Pyramids in 2000 and 2050. www.oecd.org/ dataoecd/52/31/38123085.xls)

OECD [2009a]: *Pensions at a Glance 2009. Retirement-Income System in OECD Countries*. OECD Publishing. Paris. http://www.oecd-ilibrary.org/docserver/download/fulltext/

8109081e.pdf?expires=1314019335&id=id&accname=guest&checksum=DF2CB11E52C9C0ECE1606E828294EBAA

OECD [2009b]: *Pensions and the Crisis*. 23 June. http://www.oecd.org/dataoecd/10/26/43060101.pdf

SMITH, A. [1776]: *An Inquiry into the Nature and Causes of the Wealth of Nations*. Strahan, W. and Cadell, T. London.

STIGLITZ, J. E. – SEN, A. – FITOUSSI, J.-P. [2009]: *Report by the Commission on the Measurement of Economic Performance and Social Progress*. http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf

UNDP (UNITED NATIONS DEVELOPMENT PROGRAM) [2010]: *Human Development Report*. http://hdr.undp.org/en/reports/global/hdr2010/
http://hdr.undp.org/en/media/HDR_2010_EN_TechNotes_reprint.pdf

UNIVERSITY OF CALIFORNIA DEPARTMENT OF DEMOGRAPHY – MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH [2011]: *Human Mortality Database – Hungary*. http://www.mortality.org/cgi-bin/hmd/country.php?cntr=HUN&level=1

WORLD ECONOMIC FORUM [2010]: *Global Competitiveness Report 2010–2011*. Geneva. http://www.weforum.org/issues/global-competitiveness

ZÁDOR, M. – GÁSPÁR, T. [2010]: Relative Total Performance Index. An International Comparison in a New Perspective. *Hungarian Statistical Review*. Vol. 88. Special No. 14. pp. 89–109.

Appendix

Components of pillars 4 and 5 of the global competitiveness index

4. Health and primary education
   - A. Health
     - 4.01 Medium-term business impact of malaria
     - 4.02 Medium-term business impact of tuberculosis
     - 4.03 Medium-term business impact of HIV/AIDS
     - 4.04 Infant mortality (hard data)
     - 4.05 Life expectancy (hard data)
     - 4.06 Tuberculosis prevalence (hard data)
     - 4.07 Malaria prevalence (hard data)
     - 4.08 HIV prevalence (hard data)
   - B. Primary education
     - 4.09 Primary enrolment (hard data)
5. Higher education and training
   - A. Quantity of education
     - 5.01 Secondary enrolment ratio (hard data)
     - 5.02 Tertiary enrolment ratio (hard data)
   - B. Quality of education
     - 5.03 Quality of the educational system
     - 5.04 Quality of math and science education

5.05 Quality of management schools

C. On-the-job training

5.06 Local availability of specialized research and training services

5.07 Extent of staff training

Characteristics of Serbian-Hungarian International Migration Before Schengen

Áron Kincses
PhD, Head of Section
HCSO
E-mail: Aron.Kincses@ksh.hu

Zoltán Takács
PhD student,
Research Fellow
Regional Science Association of Subotica (Serbia)
E-mail: takacszoli@stcable.rs

It is important to know what kind of features and spatial impacts characterise the international migration between Serbia and Hungary for two reasons. On the one hand, there has been significant population movement between these two countries since the Yugoslav wars, and from 19 December 2009 citizens of Serbia can travel to the Schengen area without a visa. On the other hand, one of the most considerable groups of foreign citizens in Hungary is the Serbs. In the study the authors examine the migration process from Serbia to Hungary. The analysis shows that there are two separate and considerable geographical motives, which mostly determine the spatial distribution of the location of foreigners in Hungary, namely, the effects of centres and borders. The first one means both a dynamic centre of immigration to Hungary and of emigration from Serbia. The second is proximity to the border, which is not a barrier but rather a contact zone. The authors describe migration between these two countries before Serbia joined the Schengen zone and analyse the period between 2001 and 2008 with the aim to show the effect Serbian citizens' emigration has on the Serbian and Hungarian societies and economies, as well as the role of Vojvodina as a region.

KEYWORDS:
International migration.
Regional disparities.

Serbian citizens may travel to the Schengen area without a visa since 19 December 2009. This fact may further increase the volume of migration from Serbia to Hungary which has been extensive anyway since the beginning of the Yugoslav civilian wars.

Serbian citizens make one of the largest foreign populations in Hungary. Therefore, it is important to examine the characteristics of these people, what regional effects their presence has in Hungary, as well as the impact of their absence from Serbia. We try to point out the specific relations between emigration and the target regions, for example the connection of the kinds of migrants from different parts of Serbia settling in various regions of Hungary. The study presents the spatial characteristics of this migration network.

During the analysis we primarily examine data from the period 2001–2008 but also consider the Serbian migration flows of the twentieth century as well as the current migration position of Vojvodina. We are convinced that it is important to understand the processes and the effects of the preceding events on the observed period, as well as to explore the socio-economic background of migration from Serbia towards Hungary.

1. Review of the international migration literature

According to the Hungarian statistical interpretation, international migration means leaving the country of the original place of living with the aim to find a new place of residence in another country for settling down or for performing earning activities (*HCSO* [2007]). This definition was completed by *Mária Rédei*, who introduced the term of educational migration (*Rédei* [2007]) for learning and by *Sándor Illés* with elderly migration (*Illés* [2006]) with the purpose of better utilising the purchasing power of pensions or searching for better living conditions. The Serbian definition of international migration is the permanent or temporary leaving of the territory of a nation-state for another by a great number of inhabitants. (Others like *Grečić* mean only the permanent settlement and movement of population by this term (*Grečić* [2001])). Legally it was considered as a neutral phenomenon: an activity that cannot be punished (*Szalma* [2001]). However, this Serbian approach has dramatically changed by 2009: in accordance with the European migration policy, illegal migration has become a criminal category that must be punished (*Đorđević* [2009]).

To summarise, international migrants change their place of residence at their own choice, which is influenced by external and internal factors, attractive and repulsive effects. This way migrants actively take part in the rapid transformation of the demographic and economic structure of the world, thus international migration is a visible form of numerous economic and demographic force fields.

In our opinion, besides the regional differences in wages (*Hatton–Williamson* [2005]), the geographical position of the target regions also has an important effect on the spatial distribution of migrants. Geographic and economic centres and peripheries play an important role in choosing the place of living. According to our hypothesis, an important factor in the spatial distribution of Serbian migrants in Hungary is the capital, Budapest, an economic hub which is a typical dynamic reception centre of international migration as well. Compared to the national average of settlements, Budapest is quite over-represented concerning the number of foreigners and their proportion per 1 000 inhabitants, being in tune with international trends, since the main target territories of migration are capitals. It offers numerous occupations and there are representatives of a wide range of sending countries in this place. Another significant factor is the proximity of borders, which presents additional values for many migrants. We can mention, for example, the better and easier contacts with family members who remained at home. These phenomena can be also recognised in the sending country, that is, migrants from Serbia who lived near the border would more probably settle down close to the Hungarian frontier than others. Serbian migrants, who mostly come from Vojvodina, prefer to live either in Budapest or not far from the common border.

Border regions were traditionally considered as disadvantageous territories according to location theories because of the barriers in international trade and the threats of military invasions (*Anderson–O'Down* [1999]). National borders negatively affected regional economies because they increased transactional costs. Taxes, different languages, cultures, and business practices made obstacles for cross-border trade in general, which reduced the willingness of national and foreign companies to locate in these regions (*Hansen* [1977]).

The alteration of this unfavourable image could generate "a new increase" in border regions through greater international economic integration – with lower trade barriers. Now these regions have several characteristics based on which they can be defined as active contact zones (*Nijkamp* [1988], *Van Geenhuizen–Ratti* [2001]).

Between 1988 and 1999, 3 982 Yugoslav investments were realised in Hungary (reaching the peak in 1993 with about one thousand enterprises), which made 16 percent of all foreign investments. Seventy percent of them were in the Southern Great Plain region, while only 20 percent in Budapest (*Szónoky–Ancsin* [2001]). *Nađ* [2006] characterises this period as the time of the leaving of intellectuals, the founding of companies and the escaping of capital. Concerning Hungarians in Vojvodina, *Gábrity Molnár* [2008] reports on migration losses of ethnic character.

Grečić [2001] considers Hungary and Poland as transit countries for migrants, resulting from the stricter migration policy of the EU in the 1990s (*Grečić* [2001]). Nađ considers these countries as a springboard (*Nađ* [2006]). However, the difficulty of entering the target country often made these people remain here (*Grečić* [2001]).

The study of migration in Serbia cannot be considered as continuous because of the lack of adequate data. The events of the period of 1991–2002 can hardly be backed up with exact numbers, and the Serbian literature after 2000 is based on estimates. We could find only few migration statistics in the 2002 census data compared to the previous ones, partly because of the modified methodology. For this reason, our main results are based on the statistical database of the countries of destination. However, even so, region-specific studies are difficult since receptive countries marked former Yugoslavia as the country of origin of migrants until 1998. The relevant Hungarian migration databases contain the settlements of the countries of origin (databases of the Hungarian Office of Immigration and Nationality and the Hungarian Central Statistical Office). This allowed us to identify the settlements one by one and to analyse the spatial distribution in detail.

In our study we consider foreign citizens living in Hungary who have valid living, immigration or settlement permits on 1st January of a given year (so tourists, diplomats and those who stay in Hungary for less than a year are not examined).

After presenting the migration situation in Serbia, we examine the migration processes in the 2001–2008 period based on the original (Serbian) and current (Hungarian) places of living, as well as the age group, occupation, qualification and proportion of taxpayers. The observed regional level is NUTS 3. Moreover, where it is relevant, we also give detailed data (at settlement level). We distinguished three (primary (ISCED 1–2), secondary (ISCED 3–4) and tertiary (ISCED 5–7)) educational levels, while the classification of occupations is based on the ISCO-88 structure.

2. Serbian citizens in Hungary

Hungary has had a surplus in international migration since the demolition of the Iron Curtain. Namely, more foreigners arrive in the country than the number of Hungarian citizens leaving it. On 1 January 2008 there were 174 697 foreign citizens staying permanently in Hungary, which was 1.7 percent of the population. This means that almost two in 100 people living in Hungary are foreigners. The proportion of immigrants has increased by 61 percent in eight years after the millennium.

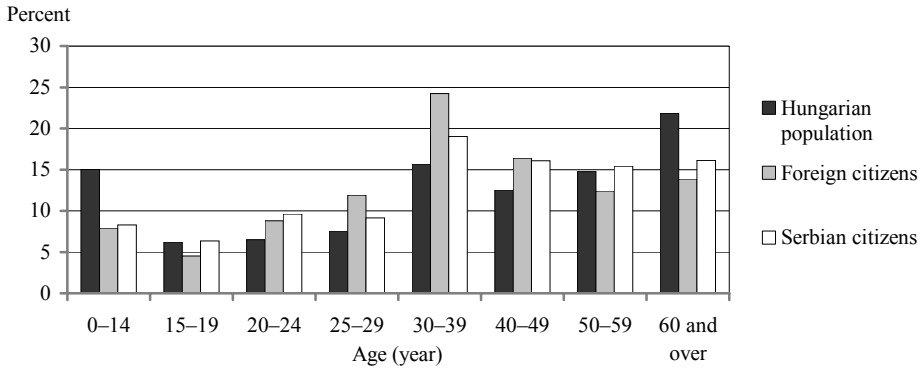
Direct and indirect effects of international migration are undoubtedly determining factors of the demographic situation in Hungary. This can be stated based on the

processes of the past two decades and those expected in the future. Direct effects are observed in the increase of the number of (active) population. Indirect effects together with the rejuvenating effect are those phenomena which are connected to the age structure and occur in the country of destination, for example births, deaths, marriages, studentification, and changes in real estate and labour markets.

Ten percent of foreigners staying in Hungary on 1 January 2008 were Serbian citizens. This rate has grown during the past three years but the proportion of foreigners in Hungary coming from other neighbouring countries is also on the increase. On 1 January 2008 about 17,186 Serbian citizens stayed in Hungary, while an additional 12 556 people have become Hungarian citizens since 1993 (97 percent of them have Hungarian as mother tongue). Therefore, according to official statistics, about *30 000 Serbian citizens have moved to and settled down in Hungary for 15 years*. This is about half of the natural decrease of population (the difference between live births and deaths) in Hungary per year.

As Figure 1 shows, the proportion of the active population of foreigners in Hungary is significantly higher than that of native Hungarians, who are over-represented in the 0–14 and 50–X age groups. The share of people aged 15–24 years is higher among Serbian citizens than among foreigners and in the whole population. Those aged 25–59 from Serbia also have higher proportion than Hungarians do, even though the high percentage of foreigners aged 25–49 years is not characteristic of them.

Figure 1. Native, foreign- and Serbian-born population in Hungary by age group, 1 January 2008



Source: Demographic Query System of the HCSO.

Due to the greater share of the active population, the proportion of taxpayers is also higher among foreigners. However, the distribution of foreign taxpayers across counties is not proportional; it is higher where taxable income per capita is higher, like in Budapest, where more than 4 percent of taxpayers are foreigners. This trend is less characteristic of Serbian citizens. Budapest and Pest County account for 62 per-

cent of foreign taxpayers, compared with less than 40 percent of Serbians. Nearly the same *number of them pays taxes in Csongrád County as in the capital.*

Taxpayers by citizenship and county of living, 2007

Counties	Number of taxpayers	Taxable income per capita (HUF)	Foreign taxpayers	
			Serbs	Total
Budapest	759 360	638 083	1 109	32 036
Baranya	169 041	289 332	131	1 334
Bács-Kiskun	227 188	263 798	339	2 015
Békés	161 601	237 811	79	1 151
Borsod-Abaúj-Zemplén	277 197	288 442	13	1 196
Csongrád	182 817	298 582	968	2 984
Fejér	207 272	367 202	142	2 118
Győr-Moson-Sopron	212 858	340 320	68	2 864
Hajdú	226 311	287 685	30	1 802
Heves	135 760	310 157	21	1 187
Komárom-Esztergom	151 416	358 457	31	2 146
Nógrád	87 407	264 147	6	521
Pest	514 948	421 507	255	10 735
Somogy	136 964	253 000	59	830
Szabolcs-Szatmár-Bereg	211 950	250 575	8	1 983
Jász-Nagykun-Szolnok	168 124	272 844	18	803
Tolna	104 360	291 682	53	708
Vas	130 716	300 813	23	705
Veszprém	174 841	289 702	30	1 063
Zala	138 475	279 140	76	862
Abroad*			220	18 034
Total	*4 378 606*	*373 033*	*3 679*	*87 077*

* Foreign citizens living abroad and pay tax in Hungary.
Source: Tax register of the Tax and Financial Control Administration.

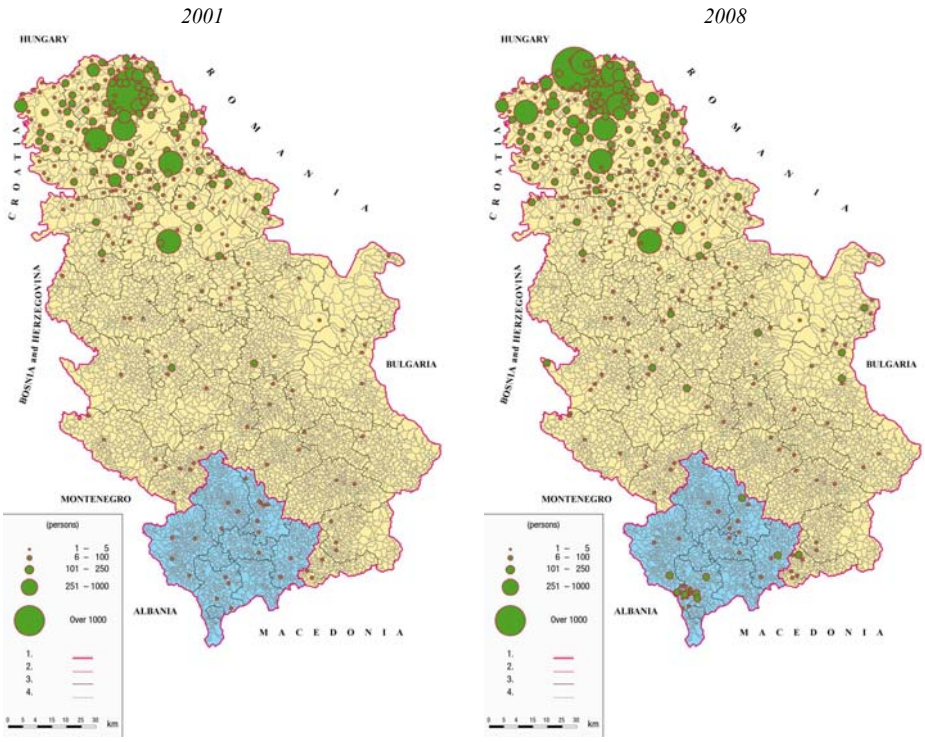
3. Regional study in Serbia and Hungary

There are traditionally strong migration relations between Serbia and Hungary. Migration concerns all Serbian and Hungarian regions, that is, there are migrants in

every Hungarian county from each part of Serbia. However, this process involves only 5.3 percent of Serbian and 24 percent of Hungarian settlements, thus strong regional effects can be observed.

Subotica (3 365 people), Senta (1 951), Novi Sad (1 020), Bačka Topola (705), Kanjiža (694), Bečej (524) and Zrenjanin (508) are the Serbian settlements, while North Bačka (5 869), North Banat (4 661), South Bačka (2 429) and West Bačka (1 005) are the regions that are particularly involved in immigration to Hungary. About 88 percent of emigrants from Serbia arrive from Vojvodina. On the other hand, among settlements in Hungary as a target area the most preferred ones are Szeged (4 481), Budapest (3 896), Kecskemét (499), and Baja (336); concerning counties these are Csongrád (6 270), Budapest (3 896), Bács-Kiskun (2 522), and Pest (945). Eighty percent of immigrants live in the foregoing regions. It is important to point out that Serbian citizens are characterised by convergence to their original and divergence to their current place of living.

Figure 2. Serbian citizens living in Hungary by settlement of origin, 1 January 2001 and 2008



Note. The maps are made by *Zsolt Bottlik*, Researcher of the Geographical Research Institute, Hungarian Academy of Sciences

Figure 3. Serbian citizens living in Hungary by settlement of destination, 1 January 2001

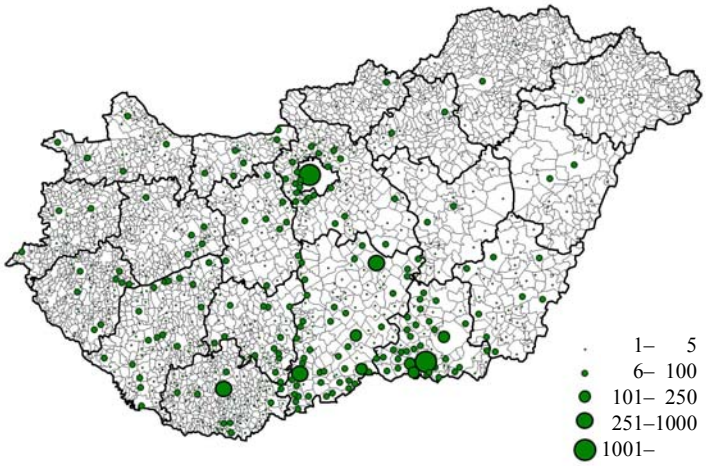
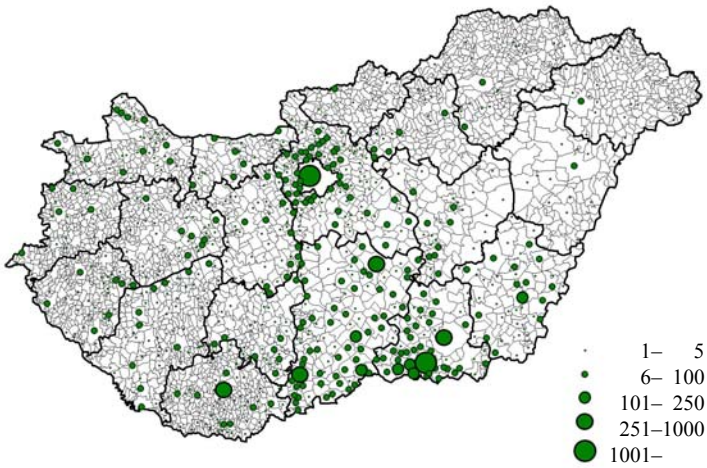


Figure 4. Serbian citizens living in Hungary by settlement of destination, 1 January 2008



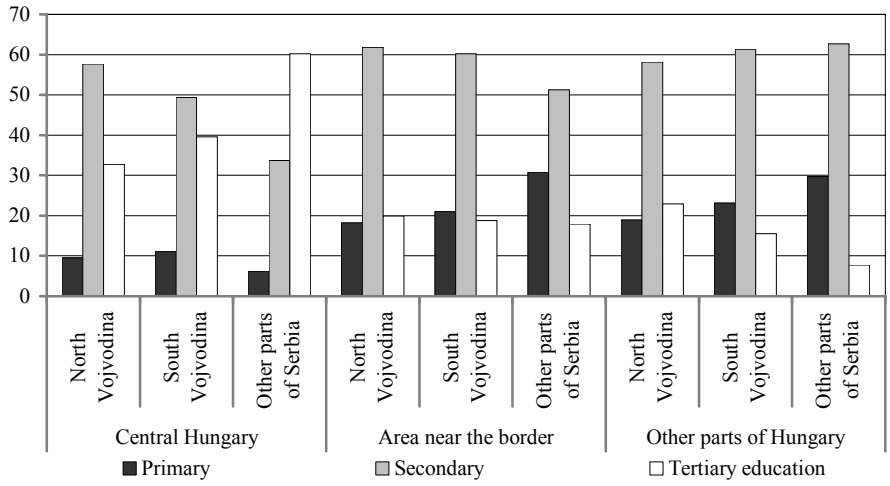
Our aim is to find a relationship between the original Serbian districts and the current Hungarian counties, that is, to analyse the Serbian migrants concerning this characteristic. According to a *"from where–to where" migration matrix* we can divide the regions of the original (Serbian) places of living into three groups. In the first group there are places near the border (North Bačka, North Banat, and West Bačka), in the second are South Bačka, South Banat, Central Banat, and Srem, while the third is the group of other regions. We distinguish three groups in Hungary, too,

for characterising the observed migration system. These are counties near the border (Csongrád, Bács-Kiskun), migration centres (Budapest and Pest County), and other counties.

It is important to note that there is a strong and concentrated flow of people from North Vojvodina to Csongrád or Bács-Kiskun counties, numbering 6 532, who represent 40 percent of all Serbian migrants. Besides a significant regional concentration, migration also shows a close relation between the proportion of Hungarian speakers (in an ethnic sense) and the willingness to emigrate from Serbian settlements. Serbian citizens who now live near the border (on the Hungarian side) mainly arrived from areas next to the border (78%), while many Serbian migrants from other parts of Serbia (49.4%) settled in the Hungarian migration centre. Migrants from other countries also prefer Budapest and Pest County irrespective of their original place of residence.

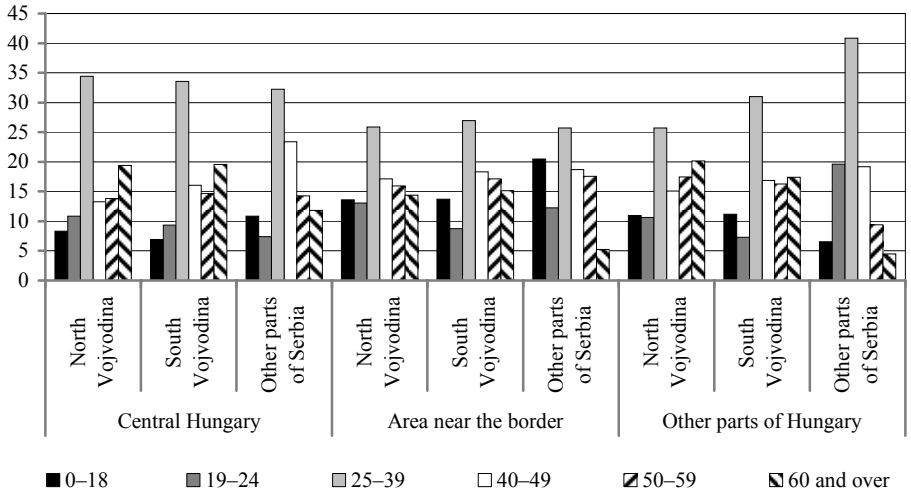
We further examined the Serbian migrants based on their qualifications, occupation, and age group. We found that people with higher educational attainment were attracted to the migration centre even from larger distance. Migrants who completed their secondary and tertiary education live principally in this area. Although the average educational level of Serbian citizens near the Hungarian border is lower than in other places, there are also a considerable number of migrants with higher education there, too. According to the data, qualifications depend more on the current place of living in Hungary than on the original one in Serbia.

Figure 5. Serbian citizens (aged 18 and over) staying in Hungary by spatial group and educational attainment, 1 January 2008
(percent)



The analysis of age groups (see Figure 5) shows a considerably different spatial distribution. The majority of Serbian citizens living in Central Hungary are in the active age group, but the proportion of older people is also relatively high. The share of older people who migrated from Vojvodina is much higher than of those from other parts of Serbia, irrespective of their current place of living in Hungary. More young people live near the border than in other parts of the country; there is also a high proportion of people aged between 19 and 24 there, so we could assume that they intend to obtain their higher qualifications in Hungarian colleges or universities. The distributions by age groups depend on both the original Serbian place of residence and the current settlement in Hungary.

Figure 6. Serbian citizens staying in Hungary by spatial group and age group, 1 January 2008
(percent)



Concerning the occupation of Serbian citizens aged above 18 years living in Hungary, we can see the following proportions: professionals (12%), technicians and associate professionals (10.7%), other jobs with higher or secondary qualifications (7.4%), service and craft workers (7.4%), clerks (5.4%), elementary occupations (5.1%), legislators, senior officials and managers (3.5%), plant and machine operators and assemblers (2.4%), and skilled agricultural and fishery workers (2%). Furthermore, the proportion of the unemployed, pensioners, housekeepers and personal care workers is 28.6 percent, while pupils and students account for 15.5 percent. The proportion of students among Serb migrants is the highest in Central Hungary (20%). As for the original place of residence, more than 16 percent of people above 18 years of age from North Vojvodina, 14 percent from South Vojvodina and 10.8 percent from other parts of Serbia are students. In Central Hungary the shares of profession-

als (16.5 percent of Serbian citizens aged above 18 living there) and legislators, senior officials and managers (5%) whereas near the border those of clerks (8.3%), skilled agricultural and fishery workers (2.5%) and elementary occupations (6.3%) are higher than the national average.

4. Conclusion

In our research we examined the migration of Serbian citizens to Hungary in the period between 2001 and 2008. The results show that the number of Serbian citizens staying in Hungary gradually increased after 2001. The ending of the ten-year period of Yugoslav wars does not mean decreased emigration willingness in Serbia. In reality, the consequences of events in Kosovo in 2008 may also have further political motives for emigration.

The analysis of spatial distribution of Serbian migrants in Hungary after the millennium shows that the Serbian and Hungarian (migration) centres and the (geographical) peripheries considerably influence the trends in international migration. Budapest and Pest County are the general dynamic migration centres of Hungary, while Vojvodina is the central territory of emigration. The proximity of the border is an important geographical motive, which is not a barrier but a contact zone from the aspect of migration flows.

In our opinion, the fundamental reasons for migration from Serbia/Vojvodina (to Hungary) after the millennium are the following:

– transformation of the economic structure, unsuccessful privatisation, unemployment, regional disparities, additional uncertainty (economic migration);

– dual enterprises, doing business both in Hungary and Serbia, capital investments (economic migration, commuting);

– (temporary or permanent) migration or commuting of new generations for educational purposes (educational migration);

– particular ethnocentric migration in the case of Vojvodina Hungarians (contact zone); and

– transit migration towards Western Europe (for political and economic motives).

We found that most of the Serbian citizens (78%) living near the border in Hungary had arrived from the frontier territory of Vojvodina. Therefore the border can be

considered as a contact zone rather than a barrier. Budapest and Pest County are also the favourite places for settling irrespective of the migrants' original place of living. Here we can find migrants with various occupations and qualifications, but the intensity of migration flow is lower than near the border.

We can assume that the structure of migration is defined by the location of the original Serbian and current Hungarian counties, the demographic and economic attributes of the migrants, and their acquaintances.

Migration benefits/losses on the two sides of the border are human and material. Summarising the results, there is a strong relation in the Serbian-Hungarian migration: about 17 186 Serbian citizens live in Hungary, and almost the same number (12 556) have become Hungarian citizens since 1993 (97 percent of them have Hungarian as mother tongue). So, in total about 30 000 Serbian citizens immigrated to Hungary during the past 15 years.

With 19 December 2009, a new era began in Serbian migration. The difficulties of the 1990s seem to lessen since Serbian citizens got international legitimacy for free movement. In the study we presented and analysed continuous and intensifying migration of Serbian citizens to Hungary in the past two decades until 2008. Instead of making assumptions on future trends, we want to call attention to the presented tendencies, data and statistics which would be greatly influenced by the new Schengen rules.

References

ANDERSON, J. – O'DOWN, L. [1999]: Borders, Border Regions and Territoriality: Contradictory Meaning, Changing Significance. *Regional Studies*. Vol. 33. No. 7. pp. 593–604.

ĐORĐEVIĆ, B. [2009]: Evropska politika upravljanja migracionim tokovima i Srbija kao njen deo. *Bezbednost zapadnog Balkana*. No. 12. pp. 78–91.

GÁBRITY MOLNÁR, I. [2008]: Vajdaság népességmozgása az utóbbi fél évszázadban. *Közép-Európai Közlemények*. Vol. 1. No. 2. pp. 74–85.

GREČIĆ, V. [2001]: Migracije sa prostora SR Jugoslavije od početka 90-tih godina XX veka. *Ekonomski anali.* Vol. 44. No. 153–154. pp. 57–84.

HCSO (HUNGARIAN STATISTICAL OFFICE) [2007]: *Demographic Yearbook, 2006.* Budapest.

HANSEN, N. [1977]: Border Regions: A Critique of Spatial Theory and a European Case Study. *Annals of Regional Science*. Vol. 11. pp. 1–12.

HATTON, T. J. – WILLIAMSON, J. G. [2005]: *Global Migration and the World Economy: Two Centuries of Policy and Performance*. MIT Press. Cambridge.

ILLÉS, S. [2006]: International Elderly Migration in Hungary. *Migracijske i Etnicke Teme*. Vol. 22. No. 1–2. pp. 53–77.

NAĐ, I. [2006]: Nekoliko obeležja migracije vojvođanske elite. *Zbornik Matice srpske za društvene nauke.* No. 121. pp. 445–456.

NIJKAMP, P. [1998]: *Moving Frontiers: A Local-global Perspective*. Research Memorandum No. 22. University of Amsterdam. Amsterdam.

RÉDEI, M. [2007]: *Mozgásban a világ – a nemzetközi migráció földrajza*. Eötvös Kiadó. Budapest.

SZALMA, J. [2001]: Migráció, kisebbségi jogok, nemzetközi jogszabályok. In: *Gábrity Molnár, I. – Mirnics, Zs*. (eds.): *Fészekhagyó vajdaságiak*. Magyarságkutató Tudományos Társaság. Szabadka. pp. 77–88.

SZÓNOKY, M.-NÉ – ANCSIN, G. [2001]: *A jugoszláv működő tőke a Dél-alföldön.* FÁROSZ Nyomda. Szeged. pp. 18–30.

VAN GEENHUIZEN, M. – RATTI, R [2001]: *Gaining Advantage from Open Borders. An Active Space for Regional Development*. Ashgate. Aldershot.

Mode Effects: Same Question, Different Answers – Theory and Experimental Assessment

Zoltán Vereczkei
Counsellor
Hungarian Central Statistical Office
E-mail: Zoltan.Vereczkei@ksh.hu

In an effort to minimize respondent burden and survey costs and to improve response rates, multi-mode data collection techniques are often used to collect information (*Pintér–Kátay* [2010]). With parallel data collection techniques, however, new methodological challenges are likely to emerge. One of them is the estimation of the so-called mode effects: respondents tend to answer differently to questions depending on the mode of data collection.

KEYWORDS:
Impact assessment.
Census.
Survey.

In 2011, the Hungarian Central Statistical Office (HCSO) collected census information from households and individuals using data collection modes of self-administered paper and online questionnaires and face-to-face interviews (*Waffenschmidt* [2011). This design involved the respondents' right to choose between these three modes. Choosing a mode to respond, however, is usually not random; it is related to the characteristics of the respondent. Differences in information gathered by each mode might be the result of the social desirability bias and the technical features of the different questionnaires, referred to as mode effects. In order to prepare for the methodological challenges raised by this technique, a preliminary experimental study was undertaken to evaluate the presence of mode effects, using data collected in the 2009 census rehearsal. The paper briefly reviews the Canadian and English practice regarding mode effect analysis and describes the key points of the Hungarian study – based on the Canadian practice –, the data reported using the self-administered internet and paper questionnaires and face-to-face interviews, the different characteristics of the respondents and gives a summary on measuring mode effects using the 2009 Hungarian census rehearsal data.

1. Multi-mode data collection and mode effects

With the development and exploitation of modern technology, online (internet-based) questionnaires offer real alternatives to traditional personal interviews and paper-based self-administered questionnaires. National statistical institutes use online data collection as it is considered to be cheaper and quicker than other techniques. Data entry errors can be neglected if data gets directly into a database, therefore data can be made available for analysis shortly after the collection process is complete. Along with hardly deniable advantages, the use of online questionnaires has its own difficulties. Internet response option is not yet available to all respondents due to lower internet coverage or the lack of broadband web access in some areas and respondents are expected to have the necessary computer skills to complete online questionnaires.

To overcome these difficulties, normally multi-mode designs are offered to respondents. The purpose of such design is usually to increase or maintain response rates in surveys while providing a chance to respondents to choose the mode they prefer the most. In multi-mode data collection, however, the so-called social desir-

ability bias might be a priority issue. This is the situation when other people (interviewers) are present during the data collection, some respondents might tend to provide answers they believe are socially acceptable or desirable. In such cases when no interviewer is present and the respondents need to use self-administered questionnaires, the answers for the same questions might be various. Even if the questionnaires and questions are quite similar in each mode, the climate in which the respondent needs to provide the answer can be quite different. In the presence of an interviewer, for example, the respondent can get support on how to interpret a question or which category to choose. Even though similar support is available with self-administered paper questionnaires (printed guide) or self-administered online questionnaires (online guide), the circumstances are technically quite different. In online questionnaires, automatic filtering or automatic adoption to already entered answers is possible. In case of paper questionnaires such automation is not available and usually instructions are provided if questions should be skipped based on previous answers.

All of these social desirability and technical effects arising from differences of data collection techniques are called mode effects. They can be observed when various data collection methods are used to collect survey information. The analysis of these effects emphasizes the importance of measurement. The results of the study could be used to modify questionnaires in order to minimise mode effects, to observe and quantify them as potential errors in the data, or to adjust the collected survey data.

2. Methodology

Mode effects have been brought into focus in the last few years; researchers and methodologists are searching for solutions to handle them. The related research can be divided into two areas: estimation and adjustment. The first one usually covers approaches and experiments with a purpose to find proof that this kind of effect is present in the data. The latter means that if it does exist, a solution is sought to modify or adjust data to remove it.

2.1. Estimation of mode effects

The literature on mode effect estimation is extensive compared to the topic of mode effect adjustment. Estimation usually starts with the concept of the multi-mode

design. Basically, there are two approaches. One suggests that questionnaires in different modes should correspond as closely to each other as possible, in terms of question wording, instructions and presentation of response choices, etc. A Canadian study (*Grondin–Sun* [2008]) uses this concept in order to minimize mode effects and to facilitate the integration of data received from different response modes. A different approach proposes that in the case of a multi-mode design, one should take advantage of all the potential each mode can offer (*de Leeuw* [2005]). This latter might result in different questionnaire designs. The idea is to use the features of web or paper questionnaires to maximum while give all the support during personal interviews to get the 'best' answer from the respondents.

The Hungarian mode effect investigation was based on the Canadian mode effect study as the circumstances of the analyses were quite similar. In the Canadian paper – where internet and paper self-administered questionnaires were compared – mode effect is defined as "any sign that the data reported on internet would have been different had it been reported on paper, whether it be due to the respondent or the internet application as such" (*Grondin–Sun* [2008]). This definition indicates that mode effects are related to both the specialities of data collection modes and the characteristics of the respondents choosing each response mode. Various research projects analyzed the latter. Based on their results, the characteristics of respondents using the internet for data transmission are observed to be different from those choosing other response modes. Internet users usually have more income, higher level of education and belong to younger age groups than other respondents.

The goal of the Canadian study is to detect if mode effect is present in the data collected. The methodology consists of four key parts:

1. Comparison of answers of each question by response mode (results are called *unadjusted differences*). Identifying the questions resulting in the biggest differences;

2. Compiling of variables explaining the preferences of choosing a response mode (for example highest level of educational attainment, age, place of residence, etc.). Using these factors as explanatory variables in a logistic regression model, a predicted probability value to each respondent can be calculated, representing the chance of choosing one response mode over another;

3. Classification of cases into subgroups, based on these probabilities. The probabilities within each subgroup are more resembling, therefore the comparison of the answers within subgroups is a possibility;

4. Matching the proportions by response mode by subgroups. The proportion of respondents by response mode among subgroups is expected to be different. In order to compare the final adjusted data of

the two groups, they must match. Using one of the groups as reference group, the proportion of the other group can be adjusted. Therefore adjustment factors are needed for each subgroup that can be used to standardize the subgroup proportions. As a result, the final *adjusted differences* can be compared to the unadjusted differences. This standardization is carried out by using "1" as a weight for the respondents by first mode (reference group), then by dividing the number of respondents in the reference group by the number of cases of the other group, an adjustment factor is calculated for the other respondents.

Any differences remaining after this adjustment cannot be explained by the explanatory variables used in the regression model. For questions, where the differences significantly decrease, the explanatory power of these factors can be considered high. Therefore any remaining difference can be a sign of the possible presence of mode effects in the data.

2.2. Adjustment of mode effects

Apart from estimation, adjustment of mode effects at the end of the data collection process comes with a purpose to adjust the data gathered by different modes. At a workshop on data collection for social surveys using multiple modes, held in September 2011 in Luxembourg,[1] *Gareth James* from the UK Office for National Statistics (ONS) presented a draft version of a mode effect study of ONS. According to the plans, mode effect estimation and expectedly an adjustment mechanism will be part of a labour force survey (LFS) web data collection project. This is going to be one of the first studies where mode effect adjustment is explicitly involved. In the study, a dual-frame analogy will be used. It means that respondents will be asked to register themselves at a website. As a result, registered and non-registered household subgroups will be available. With the registration process, contact information and basic details of the households can be collected, therefore the analysis of the non-respondents will also be available at the end of the data collection. In the first phase, face-to-face interviews will be carried out using a random sub-sample of both registrants and non-registrants; that will give the usual LFS data. In the second phase, registrants not participating in the first phase will be evaluated using web data collection. This second leg of the study is for the mode effect comparisons (as non-response will likely to occur, it can also be observed what kind of households typically reject to provide information on internet). The study highlights that

[1] Workshop presentations and papers are available at the ESSnet portal: http://www.essnet-portal.eu/essnet-data-collection-social-surveys-using-multiple-modes-documents-and-presentation-workshop-septe

difference in estimates based on information from registrants and non-registrants is attributable to sampling variation or mode effects (*Gareth* [2011]).

The Netherlands also indicated that they are working on their own mode effect study in order to separate selection and measurement effects in social surveys.

These studies are clearly experimental in nature and results are expected to be published in the next two or three years.

3. Experimental estimation of mode effects in the Hungarian census rehearsals

According to the regulation of the Hungarian Population and Housing Census Act (Act CXXXIX of 2009), Census 2011 was carried out using self-administered electronic and paper questionnaires, combined with personal interview technique. As part of the preparation, HCSO conducted its second census rehearsal in 2009, using a two-phase data collection technique. In the first phase, the inhabitants assigned to complete the census rehearsal questionnaire on a voluntary basis, could send their data using either paper or online self-administered questionnaires. In the second phase, interviewers visited the inhabitants not sending their data in the first phase and personal interviews were carried out. Using the data gathered during the rehearsal, HCSO carried out a study of mode effect estimation.[2]

3.1. Methodological background

The analysis is based on the methodology of the Canadian mode effect study. The model is designed around the assumption that respondents' decision to use a certain response mode is not random. As a consequence, the data gathered by each mode can be compared if the groups are as similar to each other as possible. If the gathered data, by mode, can be examined with the preferences of choosing response mode filtered out, then the filtered differences might be explained by these variables.

The first part of the study examined the relationship between data collected by paper and online self-administered questionnaires. The second focused on data obtained from face-to-face interviews and paper self-administered questionnaires.[3] Following

[2] *Editorial comment*. The study of *Virág Erdei* of this special number covers the Hungarian Census 2011.

[3] Three different questionnaires were used in the rehearsal. Questionnaire "A" focused on occupation, "B" on the highest level of educational attainment and "C" on family relations. In order to have a larger sample for the analysis, only the questions that were included in all three types were examined.

the approach laid down in the methodology, the first step in quantifying the mode effect is to compare the results by these three modes and see if there is any difference. In order to remove the preferences of choosing a response mode, explanatory variables for the regression model had to be identified. A detailed analysis of the 2009 census rehearsal, carried out by the HCSO, showed that respondents choosing each mode were different regarding their age, the highest level of educational attainment and economic activity. Based on the results of this analysis and other studies related to mode effects, the following variables were used to explain mode preferences: number of people living at the address, age group, highest level of educational attainment, age of the person responding to the questionnaire, the fact whether the respondent is habitually resident at the address, economic activity and the number of internet subscribers per capita at the settlement.[4] Using these explanatory variables and the mode as the dependent dummy variable (using "0" for mode "A" and "1" for mode "B"), logistic regression models were built using the appropriate module of SPSS to compare the results of online versus paper self-administered questionnaires and self-administered paper questionnaires versus face-to-face interviews.

3.2. Online versus paper self-administered questionnaires

In the first part of the study, altogether 20 119 self-administered questionnaires were analyzed. 16 894 questionnaires (84 percent) were submitted by post and 3 225 (16 percent) via internet.

Based on the regression variables, predicted probabilities are calculated for each respondent using the logistic regression model. By using "0" for paper and "1" for internet response mode, the higher values indicate higher probabilities of choosing internet. The summary of the model and included variables can be seen in Table 1.

With the adjusted differences available, the comparison with the initial unadjusted results is now possible. Table 3 highlights the variables, where the initial differences couldn't be reduced by the adjustment method. According to these results, the differences are reduced in most of the cases. The initial differences related to the tenure status and other residence questions remain basically unchanged. That means that the difference cannot be explained by the used variables. Similarly, the reduction in the difference of working activity status and the highest

[4] Data source: GKIeNET Internet Research and Consulting Ltd. – Regional Planning Information System (TEIR).

Even though the definition of mode effect used in this study indicates that characteristics of the respondents should be used only, the last explanatory variable is included to describe the conditions that have an overall effect on preferences (cannot expect high rate of internet respondents if internet coverage is at a low level).

level of educational attainment variables indicated that the whole initial difference is not entirely the result of the difference in the demographic variables used for the analysis. Since the unadjusted differences only slightly changed or haven't changed at all for these questions, it can be assumed that they can indicate the presence of mode effects.

Table 1

*Regression results for online versus paper self-complete questionnaires**

Variables in the equation	B	S.E.	Wald	df	Sig.	Exp(B)
Number of persons living at the address	0.419	0.018	533.938	1	0.000	1.520
Age group	–0.095	0.016	33.730	1	0.000	0.910
Highest level of educational attainment	0.231	0.022	111.831	1	0.000	1.260
Age of the person responding to the questionnaire	–0.019	0.002	98.531	1	0.000	0.981
Fact whether the respondent is habitually resident at the address	0.662	0.142	21.807	1	0.000	1.938
Economic activity	0.034	0.010	11.522	1	0.001	1.034
Number of internet subscribers per capita	2.424	0.231	110.468	1	0.000	11.294
Constant	–3.671	0.207	313.404	1	0.000	0.025

* Sex and marital status were not significant and were dropped from the model.

Note. Model summary: –2 Log likelihood: 15011.606, Cox & Snell R Square: 0.082, Nagelkerke R-Square: 0.139

3.3. Self-administered paper questionnaire versus face-to-face interview

With 46 991 questionnaires available for examination, a larger sample was used for the analysis, in contrast with the previous comparison. 30 097 questionnaires (64 percent) were collected using face-to-face interviews and 16 894 (36 percent) were submitted by post.

In this case "0" was assigned to the respondent using the personal interview mode and "1" was applied for paper respondents. Therefore higher probabilities indicate that respondents were willing to fill out the paper questionnaire by themselves. The summary of this model, and the variables included can be seen in Table 2.

Variables of sex and marital status proved to be significant this time, however, the R-square is lower, compared to the previous analysis. After classifying the respondents into 10 subgroups, the proportions of the personal interview mode were

considered as a reference group, since the proportions were more balanced among subgroups than in the paper reporters.

Table 2

Regression results for self-complete paper questionnaire versus personal interview

Variables in the equation	B	S.E.	Wald	df	Sig.	Exp(B)
Number of persons living at the address	–0.155	0.008	343.155	1	0.000	0.857
Age group	0.106	0.009	129.861	1	0.000	1.112
Highest level of educational attainment	0.205	0.011	380.976	1	0.000	1.228
Sex	0.059	0.021	7.926	1	0.005	1.061
Marital status	–0.072	0.014	26.528	1	0.000	.931
Age of the person responding to the questionnaire	0.015	0.001	244.395	1	0.000	1.015
Fact whether the respondent is habitually resident at the address	–0.384	0.063	36.727	1	0.000	.681
Economic activity	0.012	0.005	5.023	1	0.025	1.012
Number of internet subscribers per capita	–0.369	0.125	8.659	1	0.003	0.691
Constant	–1.397	0.101	190.793	1	0.000	0.247

Note. Model summary: –2 Log likelihood: 54917.847, Cox & Snell R Square: 0.064, Nagelkerke R-Square: 0.088.

Table 3 shows the adjusted and unadjusted differences for this comparison as well. According to the results, we can come to the same conclusion as for the previous analysis. Tenure status and other residence questions seem to be affected by other factors not covered in the logistic regression model. Also, the unadjusted differences for the marital status were not reduced as significantly as for most of the questions. The values for question relating to the economic activity are reduced. However, the unadjusted differences for this question were lower, compared to the online-paper analysis. This might indicate that mode effect related to this question is more characteristic for the online-paper relation or that it is a result of the relatively small sample ("only" 3 225 questionnaires were used, compared to the 16 894 in the second case).

3.4. Results

A summary of the results can be seen in Table 3. It contains all the variables examined in the mode effect analysis (variables resulting in the biggest unadjusted differences).

Table 3

Variables included in the analysis, unadjusted and adjusted differences by response mode
(percentage point)

Variable	Category	3.1. Online versus paper self-administered questionnaires		3.2. Self-administered paper questionnaire versus face-to-face interview	
		Unadjusted differences	Adjusted differences	Unadjusted differences	Adjusted differences
Age groups	–14	7.70	–2.00	–5.50	–0.20
	15–24	3.70	–2.40	–3.20	1.10
	25–34	6.90	2.30	–5.60	–3.00
	35–44	4.10	0.50	–2.00	0.60
	45–54	1.00	1.30	0.20	0.60
	55–64	–6.00	1.00	5.90	1.40
	65+	–17.40	–0.70	10.30	–0.50
Sex	Male	3.40	–0.50	–3.50	–0.50
Nationality	Hungarian only	0.00	–0.50	0.80	0.80
Marital status	Single	12.70	–1.70	–11.30	–3.40
	Married	–2.80	–1.70	7.60	4.70
	Widowed	–7.20	2.70	3.40	–0.90
	Divorced	–2.60	0.60	0.30	–0.50
Economic activity	Working	9.30	5.30	–4.50	–1.30
	Unemployed	–1.10	–1.00	–1.20	–0.50
	Child benefit, maternity leave payment	1.90	–0.10	–1.00	–0.30
	Retired	–21.30	–0.60	14.20	0.80
	Disabled	–1.40	–1.10	0.20	–0.20
	Family provider	0.30	0.10	–0.30	–0.20
	Student	8.10	–1.40	–4.40	3.20
	Child not attending school	3.50	1.70	–3.30	–1.50
	Other	0.70	–0.30	0.30	0.10
Highest level of educational attainment	Primary school or lower	–0.70	1.20	–10.30	–1.10
	Vocational school	–7.40	–3.50	0.90	2.00
	School-leaving certificate	–0.50	–2.90	5.70	1.50
	College, university diploma, PhD, DLA	8.60	5.10	3.80	–2.40

(Continued on the next page.)

(Continuation.)

Variable	Category	3.1. Online versus paper self-administered questionnaires		3.2. Self-administered paper questionnaire versus face-to-face interview	
		Unadjusted differences	Adjusted differences	Unadjusted differences	Adjusted differences
Tenure status	Registered (permanent address)	–2.90	–2.30	8.30	7.90
	Registered (temporary address)	0.60	0.40	–2.50	–2.40
	Unregistered	2.30	1.90	–5.80	–5.50
Have another address	No	–3.50	–2.50	6.60	6.80
	Yes, permanent address	2.50	1.80	–7.20	–6.90
	Yes, temporary address	0.20	0.10	0.50	0.30
	Yes, unregistered	0.60	0.40	0.50	0.20
	Yes, abroad	0.20	0.30	–0.30	–0.50
Is the respondent habitually resident at the address?	Yes	0.40	–0.60	–0.80	–0.10
	No, at other permanent address	–0.40	–0.20	0.10	–0.20
	No, at other temporary address	–0.20	0.00	0.30	0.10
	No, at other unregistered address	0.10	0.50	0.20	0.00
	No, abroad	0.00	0.10	0.30	0.10
Always lived at the address	Yes	6.00	–1.70	–2.40	1.80
Worked last week in December	No	–7.40	–4.90	4.80	0.20

The way the content of the table can be interpreted is the following: the results of tenure status "registered (permanent address)" for the online versus paper self-administered questionnaires, for example, show that respondents choosing internet mode are usually more likely to be registered at their address, compared to respondents using paper form. With the characteristics of the logistic regression model taken into account, this difference slightly decreases, indicating that the used explanatory variables can explain only a small portion of the difference. The decrease is not significant, meaning that possible explanatory variables need to be sought elsewhere.

4. Lessons learned

Even though the regression models were quite weak and the census rehearsals were not designed to serve as mode effect studies, experience was gathered on what is needed for a potential mode effect analysis in the future.

Methodology. In practice, various data collection modes are often used in a design where respondents have the chance to choose the mode they prefer (parallel data collection) or they are approached by different modes one after the other in case of no response (sequential design). As a consequence, the methodology to evaluate mode effects might be very affected by the approach used in the questionnaire design.

An article has been published recently in *Public Opinion Quarterly* (*Vannieuwenhuyze et al*. [2011]) dealing with the evaluation of mode effects using a mixed-mode dataset with a comparable single-mode dataset present. Methodology for mode effect estimation can differ if only a multi-mode dataset is available and there is no other comparable data or if a multi-mode dataset is available with comparable uni-mode data. There is no 'gold standard' methodology at this time for mode effect treatment.

A Hungarian research would also be useful to estimate the presence of mode effects. A study, designed solely for the purpose of mode effect estimation, could provide information about the likely effects of introducing a new data collection mode.

Purpose of the mode effect study. It is also different what researchers or national statistical institutes can use mode effect analysis for: estimation or adjustment. In case of estimation, it can be used as a 'questionnaire test' to identify questions being affected by mode effects. If so, the mode effect study can provide feedback and force the process to go back to the questionnaire design phase in order to minimize these effects at the very beginning. Then mode effect analysis can be carried out again until the effect is considered to be negligible enough to conduct a multi-mode survey. If the goal is to adjust data to remove mode effects, the methodology behind their estimation needs to be solid. There are no real examples or practices available on mode effect adjustment at this time.

Explanatory variables. Paradata. Staying with the current methodology (the Canadian approach), additional explanatory variables might be required to compute more exact adjusted differences, based on more sophisticated subgroups. The current analysis made an attempt to use wide range of variables in the regression model, however, most of them proved to be insignificant and therefore were dropped out. The use of larger samples can provide further information on understanding the differences between the results collected by different modes.

As census rehearsals were designed not for mode effect studies in the first place, variables were limited to the gathered data. The Canadian reference shows that additional variables can be entered into the analysis (for example, data about the survey process by which the data are collected). These so-called paradata are to provide information about the data collection process itself. As mode effects are directly linked to the data collection, useful results might be achieved if paradata were available for the analysis.

Sensitive questions. The current analysis was based on data collected from respondents in a voluntary census rehearsal. Since the analysis of mode effects can be the most useful if sensitive questions are present, it should be carried out using survey data containing real sensitive information.

Technical features. Differences in the responses can be the result of the technical implementation of different modes. The internet application used for data entry may have other features than traditional paper form. For this reason, the technical aspects should also be examined in detail. The application used for the 2009 census rehearsal sent messages to the user if no or incorrect answer was entered (format and range checks). Technical information whether the user changed the initial answer due to these warnings is required to measure the effects of such messages. The application, however, had no such feature; therefore the technical aspects could not have been examined. It is not known whether the respondents changed their initial answers as a result of these warnings or at which questions and how often such warning messages popped up. The Canadian study describes that analyzing the effects of pop-up menus and other features of the application can be useful to understand the initial differences in the responses.

5. Final conclusion

The Hungarian Central Statistical Office carried out an experimental mode effect analysis for the first time using the data collected during the 2009 census rehearsal. The mode effect estimation followed closely the methodology used in a Canadian mode effect study. Though the analysis of mode effects was not one of the declared goals of the rehearsal, the aim of the study was to measure if such effects were present in the collected data. According to the results of both the online versus self-administered questionnaires and self-administered paper versus face-to-face comparisons, tenure status and address questions might be affected by mode effects. Other questions, such as marital status or economic activity might be subject to the effect as well. However, based on these results, it is not certain if mode effects for

these questions appear only in certain mode relations or if they are the result of the small sample of internet respondents.

New data collection techniques will nevertheless become more popular in the future since they offer an alternative to traditional modes to motivate respondents to give information. Introducing new data collection modes, however, may require different questionnaire designs and technical implementation; therefore distortion effects might be present in the collected data, especially if different modes are used at the same time. Since mode effects can affect the quality of the collected information, planning such a multi-mode data collection technique requires thorough preparation.

The results of a mode effect study can provide feedback on data quality and can highlight questions that should be treated with attention during the preparations, including questionnaire design and technical implementation.

A twofold analysis can be carried out to test the presence of mode effects in the data. The current study aimed to measure the effects resulting from the different demographic characteristics of the respondent. Analyzing the technical aspect was not possible due to the application used to collect the census test data in 2009.

The data of this study is based on a relatively small sample and contained no real sensitive variables. Relying on the results, general conclusions cannot be drawn. For such inference, further testing and analysis is required, preferably based on different datasets and larger samples.

References

COUPER, M. P. – MILLER, P. V. [2008]: Web Survey Methods – Introduction. *Public Opinion Quarterly*. Vol. 72. No. 5. pp. 831–835.

GARETH, J. [2011]: *Introducing Web Data Collection in the UK*. Workshop on Data Collection for Social Surveys Using Multiple Modes. 22–23 September. Luxembourg.

GRONDIN, C. – SUN, L. [2008]: *2006 Census Internet Mode Effect Study*. American Statistical Association. Joint Statistical Meeting. Section on Survey Research Methods. 3–7 August. Denver. http://www.amstat.org/sections/srms/proceedings/y2008/Files/300977.pdf

HEERWEGH, D. – LOOSVELDT, D. [2008]: Face-to-Face versus Web Surveying in a High-Internet-Coverage Population. *Public Opinion Quarterly*. Vol. 72. No. 5. pp. 836–846.

HCSO (HUNGARIAN CENTRAL STATISTICAL OFFICE) [2008–2010]: Census working papers of the Hungarian Central Statistical Office. Budapest.

DE LEEUW, E. [2005]: To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*. Vol. 21. No. 2. pp. 1–23.

NOLAN, F. – WAGSTAFF, H. – WALLIS, R. [2009]: *Managing Multi Mode Collection Instruments in the 2011 UK Census*. Survey Methodology Bulletin. No. 65. September. Office for National Statistics.

PINTÉR, R. – KÁTAY, B. [2010]: A hibrid adatfelvétel módszertani kihívásai. *Statisztikai Szemle*. Vol. 88. No. 7–8. pp. 723–738.

SMITH, T. W. – DENNIS, J. M. [2008]: *Mode Effects on In-Person and Internet Surveys*: *A Comparison of the General Social Survey and Knowledge Network Surveys*. American Statistical Association. Joint Statistical Meeting. Section on Survey Research Methods. 3–7 August. Denver. http://www.amstat.org/sections/srms/proceedings/y2008/Files/300977.pdf

VANNIEUWENHUYZE, J. – LOOSVELDT, G. – MOLENBERGHS, G. [2011]: *A Method for Evaluating Mode Effects in Mixed-Mode Surveys*. http://poq.oxfordjournals.org/cgi/content/full/75/5/909 28th November 2011

WAFFENSCHMIDT, I. [2011]: A 2011. évi népszámlálást előkészítő próbafelvételek tapasztalatai. *Statisztikai Szemle*. Vol. 89. No. 7–8. pp. 729–742.