HUNGARIAN STATISTICAL REVIEW

JOURNAL OF THE
HUNGARIAN CENTRAL
STATISTICAL OFFICE

79. VOLUME 2001. SPECIAL NUMBER 6

CONTENTS

MONEY SUPPLY, GDP AND INFLATION: THE DYNAMIC ECONOMETRIC ANALYSIS OF MACRO-EQUILIBRIUM*

TAMÁS MELLÁR[1] – GÁBOR RAPPAI[2]

The paper investigates some theoretical problems of the Hungarian economy. The fundamental question is, whether the Hungarian economy meets the requirements of a modern, developed market economy at the end of the transition period. In order to answer the underlying question the authors present several macroeconometric results, based on quarterly analysis of macroeconomic time series concerning the interrelations of GDP, money supply, velocity and inflation. No unanimous answer to the basic question is given, nontheless interesting findings for the relevance of some theoretical macromodels, the equilibrium–disequlibrium situation of the economy and the causal relations of some macrovariables are revealed.

KEYWORDS: Quantity theory of money; Disequilibrium analysis; Econometrics.

The study of the economic relations in the transition period is of both theoretical and practical importance. The two theoretical problems which require clarification are the following: *a*) has the transformation of the country into a market economy been completed, i.e. is the transitory stage over for the Hungarian economy; and *b*) is the emerging new system identical with the highly developed market economies, or should it be considered as something special, a 'third way' formation? The two questions are, of course, not independent from each other. It would hardly be possible to give comprehensive, and in many respects unquestionable answers, therefore it seems more useful to focus research on certain special areas and use the findings to facilitate the understanding of the nature of the new system. The present study follows this approach by attempting to analyse the equilibrium–disequilibrium of the macro level monetary processes with the help of empirical tools.

The determinant significance of the monetary processes did not become self-evident until after the transformation into the market economy. Attention was, however, focused mainly on the microeconomic processes in the finances and the money market rather than

* The study was written within the framework of the research programme No. T 019925, of the Hungarian Scientific Research Fund (OTKA) entitled. 'The econometric analysis of the macro-economic relations of transition.

[1] President of the Hungarian Central Statistical Office.
[2] Associate professor of statistics, at the University of Sciences Pécs.

on the macro-level interrelationship between the monetary and real processes or the changes in the equilibrium. The authors of this study attempt to apply econometric tools to the analysis of this area, which has so far been less broadly investigated either from theoretical or practical (economic policy) aspects. The following three issues will be dealt with.

1. The relationship between the money supply and the GDP, with special regard to the fact that money can be active, i.e. used for the purposes of monetary policy, or can be passive, playing a neutral role in the shaping of macroeconomic processes.

2. The interrelation between the supply of money and the changes in the price level, the impact of the changes on the money supply on inflation, and the influence of inflation on the money demand and supply.

3. The changes in the money demand and supply, the study of the equilibrium on the money market, periods of disequilibrium on the money market: their changes in time and the macro-economic problems they refer to.

There are several factors making it quite difficult to perform the investigations proposed by the authors. The first and most important of these is the lack of a firm theoretical foundation which could serve as a basis for the analysis. This is the consequence of the transitional condition. The second difficulty is the lack of adequate, homogeneous data sets, and reliable, high quality macro-data, that could be the unquestionable starting point for the empirical study. Thirdly, the application of the econometric methods including testing raises various methodological problems. (Not to mention only the most obvious of these: our data-base contains not more than 40 observations, which makes the findings of the stochastic time series modelling less convincing. However, it can hardly be accepted that the 'optimal' size of the data-base does not allow macroeconomic modelling for at least twenty years to come.) In spite of all these difficulties we do believe that certain statements will prove to be true and some of the analyses will be useful in the future as well.

TESTING OF THE BASIC INTERRELATIONSHIPS AND THE SURVEY OF THE DATA

In this section we introduce the basic variables of the study, give a short description of their change in time in the period of transition and analyse their causal interrelations.

Quantity theory of money and the velocity

As a starting point for the analysis, the basic interrelationship of the quantity theory of money was used:

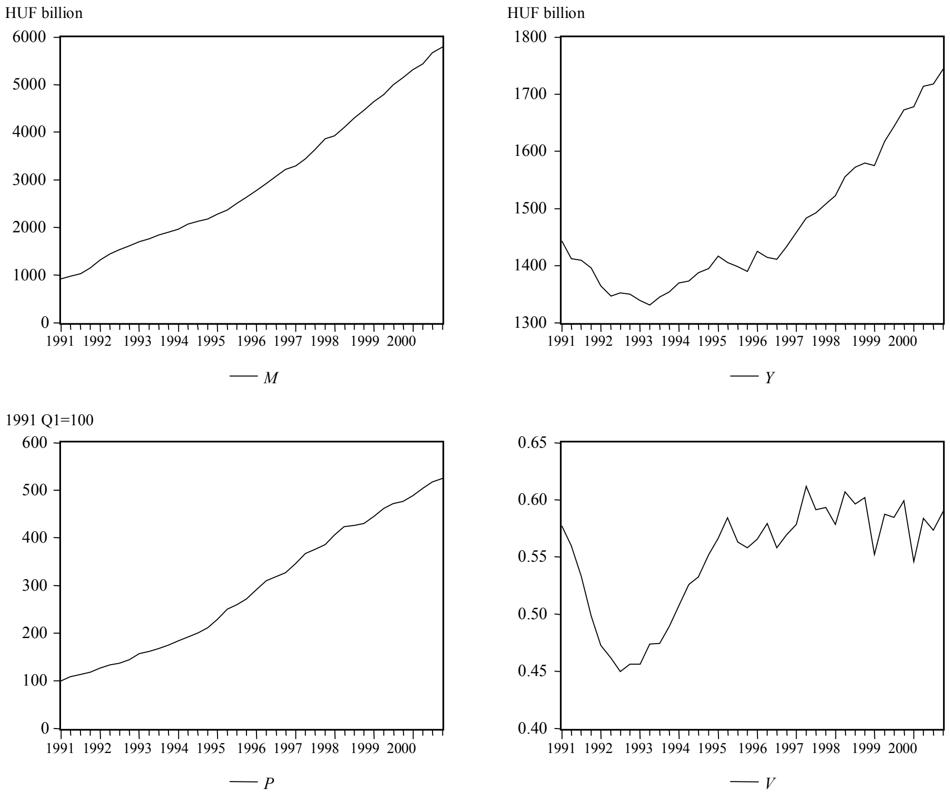
$$M \cdot V = P \cdot Y \qquad /1/$$

where M means the quantity of money, V is the velocity of money, P is the price level and Y represents real GDP. Following the conventions, the logarithms of the variables is given in lowercase letters, consequently the previous relationship /1/ can be transcribed as follows:

$$m - p = y - v \qquad /1a/$$

Several theoretical assumptions have been put forward to interpret the changes in the four variables. Monetarists say that the velocity of circulation remains constant – especially in the long run – and real GDP will change according to its potential growth path, consequently changes in money supply and price level are closely related. Keynesians state that the velocity cannot be considered as constant, and the adjustment of the price level is slow and comes with some delay, but the quantitative adjustment is much faster and more vigorous, therefore significant interrelation actually develops between money supply and the GDP. Which of the two theoretical assumptions can more appropriately be applied on the Hungarian conditions is hard to tell a priori, on a theoretical basis. The answer for this question can be attempted only after the analysis of macroeconomic data. For the period between 1991–2000, a relatively homogeneous quarterly data-set serves as a basis for the empirical studies.[3]

Figure 1. Time paths of the basic variables
(1991–2000, seasonally adjusted series)



[3] All these data available for the whole period can be found in the official publications of the Hungarian Statistical Office (except for the GDP, because quarterly GDP data are available only for the period 1996-2000). The GDP quarterly data for the period preceding 1996 were estimated from the annual data by *Viktor Várpalotai*. The detailed methodology of this estimation is given in *Várpalotai* (2000). With regard to the sensonality the variables are seasonally adjusted (except for the interest rate). The definitions and the observed time series of the variables can be found in the Appendix.

The first step was the exploration of the data generating processes (DGP) of the four basic variables. Three of them are observable (see the Appendix), while velocity of money is derived from /1/ as

$$V = \frac{P^d Y}{M},$$

where P^d is the quarterly GDP deflator. Since this GDP deflator was partly estimated from annual data, we mainly use quarterly consumer price index for *P* in the rest of this paper. The only exception is the real money, were the GDP deflator will be applied as M/P^d. The changes of the four variables over time are shown in Figure 1.

As the graphic representation of the basic variables suggests, these processes cannot be considered as stationary, so the level of integration was determined by means of hypothesis testing. The DGP of the basic variables was explored by the augmented *Dickey-Fuller* test (ADF). Table 1 shows the findings of the tests. The data-generating processes of the individual variables were specified separately (to find out whether they contained constant terms, or trends). The model was chosen on the basis of partial *t*-tests. Table 1 contains the test values calculated for the optimal model.

Table 1

ADF tests for the basic variables

Variable	Level	First difference	Second difference
M	-2.882	-2.674*	-6.506***
Y	-1.949	-4.631***	-6.338***
P	-2.760	-6.344***	-14.278***
V	-2.746	-3.246*	-5.537***

Note: The random walk null-hypothesis is to be rejected * at 10, ** at 5, and *** at 1 percent level.

Table 2

ADF tests for the logarithms of the basic variables

Variable	Level	First difference	Second difference
m	-2.438	-3.493*	-6.010***
y	-2.333	-4.464***	-6.428***
p	-2.323	-4.198***	-12.523***
v	-2.720	-3.121**	-5.552***

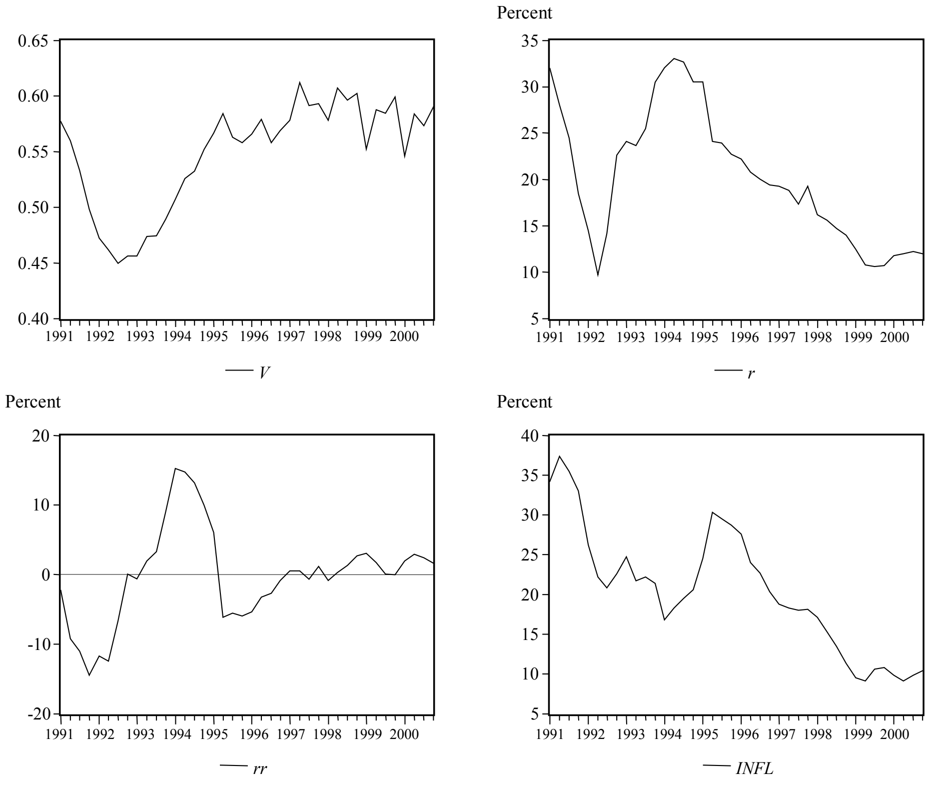
Note: The random walk null-hypothesis is to be rejected * at 10, ** at 5, and *** at 1 percent.

It can be seen that of the basic variables, the GDP and the consumer price index measured as the percentage of the first quarter of 1991 and representing price level, are most probably integrated of order 1. In the case of money supply and velocity, the null can be rejected only at a 10 percent level in the case of the first differences, i.e. the variables can be interpreted as integrated of order 2. This ‘uncertainty’ in the quantity of money is probably due to the structural break (the ‘*Bokros* package’) experienced in the first quarter of 1995. It will be ‘transferred’ to the other variables as well, since the velocity was simply transformed using the other variables and equation /1/.

Since later, in the course of analysis, the original variables will be replaced with their logarithms (see, for example equation /1a/), the mentioned tests were performed for the transformed variables as well. Table 2 shows that the conclusions to be drawn from the tests are in perfect agreement with the statements made about the original variables.

Let us examine now in more details the changes in the velocity (see Figure 2) The time sequences lead us to two conclusions (in addition to the fact that its expected value is not constant, as can be seen in the analysis of the order of integration): on the one hand, there was a significant decrease after the change of regime, due to the transformation crisis, and later, in the period of consolidation, it returned to the former level; and, on the other hand, its value got stabilised in the second part of the period under consideration (there were only small deviations).

Figure 2. Time paths of the velocity and its factors



What kind of factors can be related to the change in the velocity? First of all, the interest rate (*r*), because the trends in the costs of keeping money must have a decisive impact on the duration of keeping the money. On the other hand, the co-integration test has shown that although there is a relationship between the velocity of circulation and the nominal interest rate, or real interest rate (*rr*, nominal interest rate – rate of inflation), the direction of the relationship is not positive, but negative – in contrast with the expectations – and the error correction mechanism does not function properly either.

A more thorough look at Figure 2 will help us to find an acceptable explanation of this phenomenon. The figure shows quite clearly that both the nominal and the real interest rates decreased significantly in the early period of the transition; the decrease of the former came as a consequence of the economic and financial crisis, while the latter for the same reasons and because of the increasing inflation. Later, after 1992, both showed an increase, while inflation slowed down and this process continued until the end of 1994. Actually, up to this point the trend was the same as in the case of the velocity of money. After 1994, however, inflation highly increased again, while both interest rates fell significantly. This accelerated inflation was not caused by a setback in the performance of the economy, but by the change in the economic policy, therefore the velocity did not decrease again, but stabilised around the given level. This short historic overview leads to two conclusions: on the one hand, a distinction between the period before 1995 and the period after it will make the close relationship between velocity and interest rates even more obvious. On the other hand, considering that the changes in the real interest rate were not caused by the inflation, but by the nominal interest rate, one can suppose that the velocity is more strongly related to inflation than to the real interest rate.

Let us examine now, whether these two hypotheses can be supported by the estimates. The relationship between the velocity of circulation and the real interest rate was tested again using a dummy variable, and the following co-integrating equation was set up.

$$\hat{V} = 0.5799 + 0.00214 \cdot rr - 0.08364 \cdot DUM \quad ,$$

where *rr* is the real interest rate, *DUM* stands for dummy (the *t*-statistics were 4,05 and -16.47). Dividing the period into two parts the following equations were estimated

$$\hat{V} = 0.5799 - 0.08154 \cdot rr \qquad (1991-1995)$$

$$\hat{V} = 0.5799 + 0.00214 \cdot rr \qquad (1995-2000)$$

It can be seen that the real interest rate has different impacts in the two periods. This role, however, is not equally strong; the negative role between 1990–1995 was significantly stronger than the positive one played in the second period, that of stabilization.

The co-integrating relationship between the velocity and quarterly inflation is quite convincing, and the error correcting mechanism also operates adequately. As it was expected, the relationship is negative ($\hat{V} = 6849 - 3.2833 \cdot \Delta p$) and there was no need to divide the period into two parts.

The relationship of the three basic variables

After the analysis of the velocity the three further variables of the starting equation of the money supply /1/, i.e. the changes in the nominal money supply, the price level and the GDP will be examined. Since the velocity changes in time, it is not reasonable to assume a stationary combination of these variables. It may, therefore, be more useful to continue the analysis by setting up a VAR model, instead of co-integration.

The VAR estimates of the three variables are considered to be quite adequate and reliable. There is, however, a fundamental problem, which jeopardises further analysis, namely, that for each variable only its own lagged value was significant, the other two were not, with two unimportant exceptions. These were: the GDP lagged by one and two periods was significant for the money supply and the price level. On the whole, however, this does not alter the previous conclusion: the changes in the three variables do not give a coherent picture, consequently it may be more helpful to examine the pairvise relationships.

If we examine the relationship between the supply of money and the price level, it seems to be true that changes in money supply is an I(2) process, while those in price level is an I(1), but since the former was brought about by a structural break in the middle of the period under consideration, we can reasonably expect to discover a co-integrating relationship between them, even if certain restrictions are necessary. The estimated equation (m = 3.447 + 0.799p, t = 23.11) shows that the relation is positive and quite steady, the latter being also proved by the significant error correction equation. After the testing of the causal relation it can be established that the change in money supply may be more caused by the change in price level than by the opposite relation.

Table 3

Significance values of Granger causality test

X is not the cause of Z	Lags in the VAR model				
	1 period	2 period	3 period	4 period	5 period
$p \Rightarrow m$	0.0189	0.008	0.008	0.047	0.020
$m \Rightarrow p$	0.5829	0.7570	0.1141	0.3522	0.5286
$\Delta p \Rightarrow \Delta m$	0.2873	0.1668	0.5117	0.0504	0.0124
$\Delta m \Rightarrow \Delta p$	0.6217	0.6086	0.6266	0.5343	0.2452
$m \Rightarrow y$	0.0001	0.0001	0.0001	0.0001	0.2980
$y \Rightarrow m$	0.0632	0.0633	0.0379	0.4683	0.7174
$\Delta m \Rightarrow \Delta y$	0.0018	0.0116	0.0297	0.6207	0.5672
$\Delta y \Rightarrow \Delta m$	0.7209	0.3575	0.5136	0.4283	0.4996

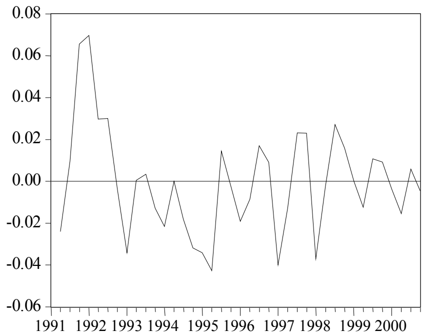
In the course of further analysis the dynamics of changes, i.e. the relationship between the increase in money supply and inflation is examined. The co-integrating regression is estimated as

$$\Delta\hat{m} = 0.01431 + 0.7837 \cdot \Delta p .$$

The estimated regression parameter can be interpreted in such a way that a 1 percent increase in the quarterly inflation brought about a 0.78 percent (quarterly) increase in money supply. In good agreement with the former statements (see Table 3) the direction of the causal relation remained the same, i.e. it is the inflation that causes increase in money supply and not the other way round (the increase in money supply causes inflation). This conclusion is drawn from the widely, but not generally accepted view that the inflation experienced in the transition period was basically generated not by the demand, but by the collapse of the supply side, and is to be considered as cost inflation originating in the structural transformation.

In spite of the revealed causality relations, we cannot be convinced that inflation is the only factor influencing the increase in money supply. Table 3 drawn up on the basis of co-integrating regression makes is quite clear that the high growth rate of the early 1990s cannot be attributed to the acceleration of inflation alone. The decrease in the velocity could also play a role in the increase of money supply, among many other things. A more detailed analysis of this issue would require a study of further money market factors, but this would be beyond the scope of the present study.

Figure 3. Residuals of the co-integrating regression between the increase in money supply and increase in quarterly inflation



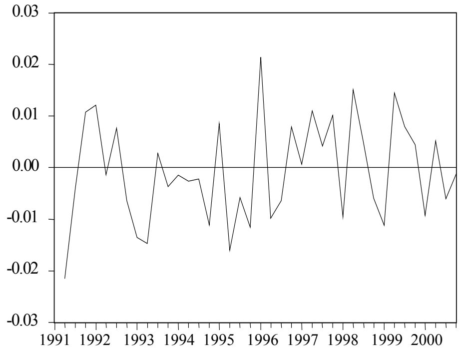
Let us turn our attention to the relation between money supply and real GDP. Testing the co-integration can be performed here as well, but with regard to our statement made earlier about money supply and price level. The test found only a weak co-integrating relation between the two variables, and the error correction mechanism does not work either. In principle these findings do not contradict any theoretical hypotheses, because the latter are usually concerned with the existence or non-existence of a relation between real money supply (instead of nominal money supply) and real output. This is also supported by the causality analysis, which has failed to produce any significant result (see Table 3).

We continue with the analysis of the relation between the increments. The co-integration test allows us to establish the following interrelationship:

$$\Delta\hat{y} = 0.0260 - 0.4352 \cdot \Delta m$$

and the relevant error correction mechanism is also correct. Figure 4 presents this relationship and shows satisfactory adjustment, i.e. well behaving residuals.

Figure 4. Residuals of the co-integrating regression between the increment in GDP and that of quarterly inflation



The negative relation between the two variables can be attributed to the fact that at the beginning of the decade production fell dramatically, leading to a very high inflation, and consequently to an increase in money supply. In the second half of the decade, however, consolidation resulted in increasing production, and a decrease in the growth of money supply and inflation. In the case of 1-3 lags the analysis of the causal relation between the rates of increase has shown beyond doubt that the rate of increase in money supply is actually the cause of the GDP's growth rate. Since there is a negative relation between the two variables, it can by no means be considered as a manifestation of the Keynesian stimulating monetary policy or of active money. What we have got here is rather the negative impact of inflation: high or increasing inflation enforces an increase in the growth of money supply, and, at the same time, impairs the conditions in real economy or makes them more uncertain. As a result, the increase in GDP is also slowed down.

The co-integrating relation is in agreement with equation /1/ and also with the relation established formerly between the velocity and inflation. Since $\Delta y - \Delta m$ is stationary, $\Delta v - \Delta p$ should be stationary as well, in terms of /1/. This requirement is not in contradiction with the co-integrating equation between the velocity and inflation given previously.

2. ANALYSIS OF THE DEMAND SIDE

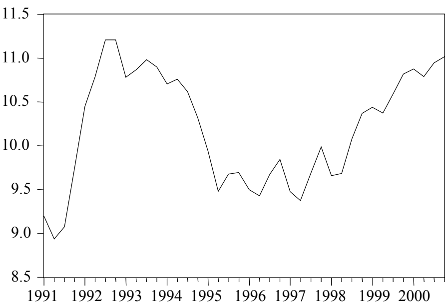
In this section analysing the demand side of two models, a traditional and the *Cagan*-type money demand model will be treated. An interesting question of this investigation is whether the adaptive or the rational expectations of inflation seem to be more plausible in describing the Hungarian economy.

The traditional money demand equation

We have so far examined the changes in nominal money supply as a basic variable and its relationships. From an economic point of view, however, the time series analysis of the real money supply is more important (*M/*P, or its logarithm *m-p*).[4] Figure 5 shows the changes in real money supply during the decade under examination.

[4] In this case – as a second exception – the quarterly deflator index is used as variable *P*.

Figure 5. Changes in the real money supply



Based upon the unit root test it can be stated that the changes in real money follow an I(1) process, (the ADF-test for the level of the variable being -2.740; the test value of the first difference is -4.509.) Therefore the changes in real money can be described by an ARMA (2, 1) model. Because the estimated roots of the dynamic process are less than one, the process is stationary. The complex roots, on the other hand, indicate a cyclical movement, though a cycle with a decreasing amplitude.

In terms of the traditional, simple Keynesian money demand function, real money demand depends on the nominal interest rates and real GDP, on the former in a negative way, on the latter positively, with respect to the speculative and transactional demand. In a general equation form:

$$\left(\frac{M}{P}\right)^D = f(Y, R) \qquad f_Y > 0, f_R < 0 \qquad /2/$$

If R is defined as a vector, which designates the yields of the alternative money holdings, then we might use the portfolio-theory approach. Fundamentally, R means short-term interest rates. It is a common assumption, that the money market is one of the most flexible ones, and therefore the market always clears, i.e. demand equals supply. It is to be stressed that in this approach this is a one-sided market adjustment, since the money demand is the only one which adjusts to the money supply. Moreover, there is another often used assumption, that money supply equals the quantity of money in circulation. Using these simplifications we get the following simple, and empirically testable formula:

$$\frac{M}{P} = f(Y, r) \qquad f_Y > 0, f_r < 0 \qquad /2a/$$

In equation /2a/ there is another hidden assumption, which states that the demand for money (the attitude of the actors of the market) remains unchanged even if the money supply changes. (This is the essence of the *Lucas*-critique, which is neglected by this assumption, and is totally avoided by the relation /2a/).

Testing empirically formula /2a/ we used a dummy variable to separate the periods before 1995 from the period after, the same way as we did previously. The results of the

OLS estimates of the model including real money, real GDP and interest rate variables are the following:

$$(m-p^d) = 5.497 + 0.3384 \cdot y - 0.0065 \cdot r + 0.0743 \cdot DUM \ ,$$

where p^d is the quarterly GDP deflator. Though all the parameters are significant, it is difficult, however, to interpret the coefficient of y, because according to the conventional interpretation the value of the transactional money demand coefficient should be somewhere between 0.5–0.8 (depending on the changes in the velocity). The discouraging results (e.g. $R^2 = 0.44$, $D-W = 0.24$) indicate, that the traditional concept of the demand function cannot really be considered relevant to the Hungarian economy in the transition period.

Consequently, it might be expedient to release the basic assumptions used so far. First, we remove the assumption, that the behavioural rule determining the demand for money is independent from the supply of money. Accordingly, let us take a look at the following five equations:

$$\begin{aligned}
&(m-p)_t^D = a_0 + a_1 y_t - a_2 r_t + \varepsilon_t^D, \\
&y_t = \beta_1 y_{t-1} + \beta_2 \left[(m-p)_{t-1} - (m-p)_t^e\right], \\
&(m-p)_t^e = \alpha_1 (m-p)_{t-1} + \alpha_2 y_{t-1}, \\
&(m-p)_t^S = \alpha_1 (m-p)_{t-1} + \alpha_2 y_{t-1} + \varepsilon_t^S, \\
&(m-p)_t^D = (m-p)_t^S = (m-p)_t.
\end{aligned}$$

The first equation defines the traditional behaviour of money demand, the second one defines a *Lucas*-type supply function, the third – based on the rational expectations – assumes that the actors of the economy know the rule of money supply, which is formulated by the fourth equation. And finally, the fifth equation further maintains the assumption, that the equality of demand and supply automatically exists without time lag on the money market. From the five equations we can get the following reduced form:

$$(m-p)_t = a_0 + a_1\beta_2(1-\alpha_1)(m-p)_{t-1} + a_1(\beta_1 - \beta_2\alpha_2) y_{t-1} - a_2 r_t \, . \qquad /3/$$

We can take the following three relations as an alternative to the previous five equation model:

$$\begin{aligned}
&(m-p)_t^D = a_0 + a_1 y_t - a_2 r_t + \varepsilon_t^D, \\
&(m-p)_t^S = \alpha_1 (m-p)_{t-1} + \alpha_2 y_{t-1} + \varepsilon_t^S, \\
&\Delta(m-p)_t = \gamma \left[(m-p)_t^D - (m-p)_t^S\right].
\end{aligned}$$

In these equations, however, the condition of the immediate adjustment of demand and supply was ruled out, the parameter γ shows the velocity of adjustment.

Further reducing these three equations we can get the following formula:

$$(m-p)_t = \gamma a_0 + (1-\gamma\alpha_1)(m-p)_{t-1} + \gamma(a_1-\alpha_2)y_{t-1} - \gamma a_2 r_t \qquad /4/$$

The two reduced forms show very similar structures, though their parameters are obviously different. The common estimation of the two reduced forms is the following:

$$\left(m-p^d\right)_t = 1.9562 + 0.8293(m-p^d)_{t-1} - 0.07517 y_{t-1} - 0.0029 r_t + u_t \ .$$

The parameters are all significant, $R^2 = 0.92$ and $D-W = 1.83$. Unfortunately, the structural parameters cannot be obtained 'back' from the estimated reduced form (under-identified model). At the same time, however, with the testing of the stability of the parameters, we can answer, even if only indirectly, the question: weather the rule of money supply has changed over time. In our case the *Chow*-test – using the most probable 1995 Q1 breakpoint – shows instability, moreover, the distribution of the residual variable of the model is not normal. As a result, we cannot make any decisive statements about the rule of money supply.

Testing the Cagan-model

In Hungary in the period of transition there was a significant inflation, which had different levels and was very varied in nature, yet was continuously present. This in itself justifies, that we use the *Cagan*-type money demand model to determine the money demand function. The model is defined by the following very simple relation:

$$(m-p)_t^D = b - a\Delta p_{t+1}^e + u_t \ . \qquad /5/$$

The demand equation expresses the empirically observed behaviour, that as the expectations about inflation increase the demand for holding money decreases. Here *b is* constant, *u* is the random variable, indicating the shocks, caused by the changes in demand and the velocity, and the parameter *a* (since we work with log-variables) expresses the percentage decrease of money demand derived by 1 percent increase of the expected inflation.

We take two cases for defining the expectations: one is the adaptive, the other is the rational expectations. First, let us take a look at the *adaptive expectation*, which according to the widely accepted definition is:

$$\Delta p_{t+1}^e = \theta\Delta p_t + (1-\theta)\Delta p_t^e \ . \qquad /6/$$

For the sake of simplicity let us assume that money supply equals the quantity of money, and the adjustment of demand-supply can be defined by the simple form presented in the previous part:

$$(m-p)_t^S = (m-p)_t$$
$$\Delta p_t = \gamma\left[(m-p)_{t-1}^D - (m-p)_{t-1}^S\right] .$$

If we assume that the adjustment is immediate and perfect, that is $\gamma \to \infty$, then based on the equations /5/ and /6/ we get the following reduced form:

$$(m-p)_t = \theta b + (1-\theta)(m-p)_{t-1} - a\theta\Delta p_t \ . \qquad /7/$$

The OLS estimates yielded relatively good results:

$$(m-p^d)_t = 1.1986 + 0.8505(m-p^d)_{t-1} - 0.3780\Delta p_t + u_t \,.$$

Because the model is just identified, the structural parameters can be determined exactly: $b = 8.0174$; $\theta = 0.1495$; $a = -2.5284$. Based on this, we can conclude on the one hand, that the last pieces of information play a relatively small role in the formation of adaptive expectations ('slow forgetting'), as opposed to the earlier expectations, on the other hand a 1 percent increase of the expectations about inflation will bring about a two and half percent decrease in the demand for money.

If the adjustment of demand supply is not perfect, then the following reduced form occurs:

$$\Delta p_t = \frac{\gamma\theta}{1+\gamma\theta a} b + \frac{1-\theta}{1+\gamma\theta a}\Delta p_{t-1} - \frac{\gamma}{1+\gamma\theta a}(m-p)_t + \frac{(1-\theta)\gamma}{1+\gamma\theta a}(m-p)_{t-1} \,. \qquad /8/$$

Based on the results of the corresponding OLS estimate the estimated reduced form is:

$$\Delta p_t = 0.6353 + 0.2200\Delta p_{t-1} - 0.1986(m-p^d)_t + 0.1221(m-p^d)_{t-1} + u_t \,.$$

Similarly to the previous case, this model is just identified, so we can determine the structural parameters, which are the following: $b = 8.30$; $\theta = 0.3852$; $a = -8.394$; $\gamma = 0.555$. Practically, only parameter *a* deviates significantly from the earlier estimates, which indicates, that if there is no immediate adjustment between demand and supply, just a gradual one (the adjustment parameter is less than 1), then as a compensation, the elasticity of demand of the inflationary expectations will be much higher.

Considering the other type of expectations, the *rational expectations*, estimation of /5/ becomes very simple. Taking into account the definition of rational expectations:

$$\Delta p_{t+1} - \Delta p^e_{t+1} = \varepsilon_{t+1}$$

where ε is a white noise variable, inserting this into equation /5/, and assuming, that supply equals the real money supply and the adjustment is perfect, then we get the following form:

$$(m-p^d)_t + a\Delta p_t = b - a\Delta^2 p_{t+1} + u_t + a\varepsilon_{t+1} \,. \qquad /9/$$

Because inflation is an I(1) process, the growth rate of inflation is stationary, and if there is no demand shock, then the whole right hand side becomes stationary, conse quently the two elements of the left hand side must be co-integrated, with co-integrating coefficient *a*.

Estimating this relation we get the following co-integrating equation:

$$(m - p^d)_t = 8.1431 - 6.493\Delta p_t + u_t \ .$$

The parameters are very close to those, estimated in the model of the adaptive expectations. The error correction mechanism works well. Testing of the residuals showed, that they may follow a normal distribution with expectation close to 0, which means that the assumption of the rational expectations proved to be acceptable.

3. EQUILIBRIUM ANALYSIS OF THE MONEY MARKET

After examining the demand side of the Hungarian money market, let us take a look at the whole of the market, and the equilibrium relations dominating the market. This obviously requires that the money supply function should be defined, which – opposed to the demand function – possesses a significantly less theoretical base. After the specification of the supply function, for the analysis of the equilibrium we must choose one of the following two principles.

– According to the *equilibrium approach*, the money demand and supply move together in the long-run (do not necessarily coincide, but the difference between them is nearly constant), when either of these variables diverts from this equilibrium path – due to an incidental shock – the mechanisms (perhaps automatisms) which return the market into the state of equilibrium immediately switch on (error correction takes place). When some influence (e.g. price increase) evokes a change leading to an increase of inflation, then with the increase of supply the equilibrium will come about, and vice versa.

– The *disequilibrium approach* on the other hand assumes that the difference between demand and supply is not steady, not stationary (not constant), there are no forces on the market that might stimulate demand or supply to ‘follow’ each other. A certain adjustment variable (e.g. the real interest rate) continuously provides information about the earlier relations of demand and supply, it decreases when the supply is prevalent, and increases when the supply decreases.

From modelling point of view this means that:

a) we specify the money demand and money supply function (when specifying the demand function we obviously rely on our earlier argumentation);

b) we estimate – independently – the value of the target variable based on the demand and the supply function, then by testing the co-integration of the two estimated variables we analyse the equilibrium hypothesis;

c) after we have a satisfactory adjustment equation, we estimate the models in a disequilibrium approach and carry out a specification analysis;

d) Based on the fit of the model estimated at two different approaches and on the specification analysis, an assessment on the equilibrium–disequilibrium situation of the economy can be set.

The specification of the market model

Based on our earlier findings it becomes obvious that the study of the equilibrium situation requires the re-specification of the money demand and money supply functions. Our study with regard to the money demand function made it clear, that the demand for real money supply depends on the GDP, on the inflation as well as on the interest rate. Consequently, we specified the following demand function:

$$D_t^M = \alpha_0 + \alpha_1 y_t + \alpha_2 \Delta p_t + \alpha_3 r_t + \alpha_4 DUM + u_t^D \ ,$$

where D_t^M means the demand for money supply, DUM denotes the dummy variable, (introduced earlier for the quantification of the impacts of the 'Bokros-package'). From the point of view of equilibrium the intended and the actual demand coincide in the long-run, its value equals that of the supply, and the actual value of the real money supply (in our earlier models we used the notation $(m-p)_t$). It is to be mentioned, that we did not want to use a lagged endogenous variable in our model, because it could result in an 'over-adjustment' especially in a later disequilibrium model.

During the specification of the supply function we could not rely on theoretical considerations as we did before. With regard to the whole of the period it would not seem feasible to have a rule of money supply or a monetary policy with exclusive effect, so we considered the two most obvious factors as basic variables: the domestic interest rate influencing the money supply of the commercial banks, and the interest premium influencing the flow of foreign currency.[5] Based on all these findings the final form of the supply function is as follows:

$$S_t^M = \beta_0 + \beta_1 r_t + \beta_2 rp_t + \beta_3 DUM_t + u_t^S$$

where S_t^M corresponds to the quantity of supply, and rp_t represents interest premium.

In both the demand and the supply function we took the logarithms of real money supply, real-GDP, and the price level variables.

The estimation and the analysis of the equilibrium model

According to the equilibrium hypothesis, the demand for money and the supply of money was in dynamic equilibrium during the past decade on the Hungarian money market, the two sides of the market practically equalled the actual money supply, that is:

$$D_t^M = S_t^M = (m-p)_t$$

The results of the OLS estimates are the following (*t*-statistics are in brackets):

$$\begin{aligned}(m-p^d)_t &= 53190 + 0.379 y_t - 0.711 \Delta p_t - 0.005 r_t + 0.083 DUM + u_t^D \\ &\quad (3.90) \quad (2.12) \quad (-1.49) \quad (-2.64) \quad (3.39)\end{aligned} \ ,$$

[5] The definition of interest premium and the referring data are displayed in the Appendix.

$$\left(m-p^{d}\right)_{t}=\underset{(315.6)}{8.026}-\underset{(-4.46)}{0.008}r_{t}-\underset{(-5.86)}{0.441}rp_{t}+\underset{(2.76)}{0.054}DUM_{t}+u_{t}^{S}.$$

The coefficient of determination of the demand function is 0.498; while the corresponding indicator of the supply function is 0.607. (It must be mentioned, that the values of the *Durbin-Watson d*-statistic – a generally accepted diagnostic test – are very low for both functions.)

The parameters of the models basically meet the theoretical expectations, with only one exception, this is the negative sign of the interest premium. According to the logic of the uncovered interest parity, the higher the interest premium is the more foreign currency will flow into the country, and after the exchange it will increase the quantity of money, that is the money supply. If, however, the monetary authorities follow a strict and consistent sterilising policy, that is they compensate for the foreign currency flow by withdrawing corresponding amounts of money from the internal market, then we can easily get a negative coefficient, because the withdrawal of money usually affects the 'active' money, while the foreign currency represents 'passive' money after the exchange (see for example the privatisation purchases). Based on the estimation of the money supply function we can conclude that during the period examined, the Hungarian National Bank followed a consequent sterilising policy.

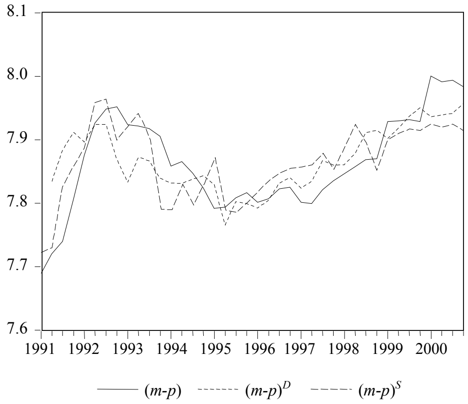
Based on OLS estimates we carried out the ex-post forecasts of the variables of the two equations, which seem to be co-integrated at a 5 percent significance-level (the value of the *Johansen* LR-test function is 17.546).

According to the co-integrating regression

$$(m-p)_{t}^{D}=-1.878+1.238(m-p)_{t}^{S}$$

where the superscripts indicate, which function the forecasts derive from. Figure 6 shows the changes in estimated demand and supply, and the actual real value of money supply.

Figure 6. The time sequence of money demand and money supply



The figure supports that demand and supply change together, so based on our findings we can assume, that the market was in equilibrium in the long-term, the error correcting mechanism prevailed itself.

The analysis of the disequilibrium model with regard to money markets

The essence of the disequilibrium concept is, that the market does not ‘clear itself’, that is – as we called it earlier there is no error correction – the actual transactional quantity (in our case the real money supply) is determined by the short side of the market at all times. The specification of the disequilibrium model resembles that of the equilibrium model described earlier, but differs from it in some – conceptual – questions. The model used by us is:

$$D_t^{\hat{M}} = \alpha_0 + \alpha_1 y_t + \alpha_2 \Delta p_t + \alpha_3 r_t + \alpha_4 DUM + u_t^D$$
$$S_t^{\hat{M}} = \beta_0 + \beta_1 r_t + \beta_2 rp_t + \beta_3 DUM_t + u_t^S$$
$$(m-p)_t = \min\left(D_t^{\hat{M}}; S_t^{\hat{M}}\right)$$
$$\Delta rr_t = \gamma\left(D_{t-1}^{\hat{M}} - S_{t-1}^{\hat{M}}\right)$$

where $D_t^{\hat{M}}$ and $S_t^{\hat{M}}$ mean the intended demand for money supply and the supply of money respectively. The latter – in case of the disequilibrium model – might not necessarily correspond to the actual supply. The model is completed by the minimum condition and a so-called adjustment equation, in which we assume that changes in the real interest rates are functions of the excess demand of the previous period. (Depending whether this equation contains a random variable we have to use different methods for estimating the parameters of the model).[6] The maximum likelihood estimates are as follows:

Parameter	Estimate	Estimated standard error
α_0	6.980	3.712
α_1	0.143	0.503
α_2	-1.373	1.246
α_3	0.005	0.003
α_4	-0.104	0.078
σ_{u^D}	0.077	0.011
β_0	7.987	0.073
β_1	-0.149	0.415
β_2	-0.005	0.004
β_3	-0.047	0.056
σ_{u^S}	0.063	0.020
γ	39.852	13.907
log *L*	-46.569	

[6] Given that both equations contain the same explanatory variables, construction of the likelihood function is not trivial. The details of this problem are beyond the scope of this study (see *Quandt*; 1988 and *Rappai*; 1989).

It can be seen, that the parameters show only a slight deviation from the values estimated earlier by the OLS method. Some of the striking differences are the result of the large standard errors which make even the signs uncertain. All this indicates that the empirical verification of the disequilibrium hypothesis is based on very precarious grounds.

Figure 7. The intended demand for and supply of money, and the changes in money supply

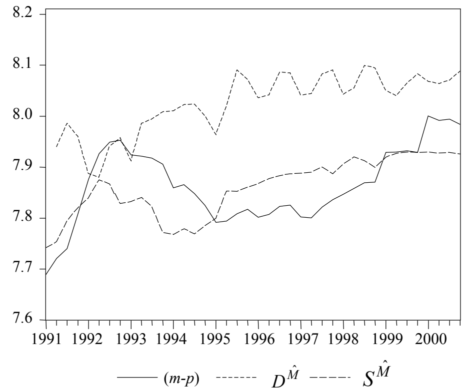


Figure 7 shows the changes in estimated intended demand, and supply having a disequilibrium conception, as well as the actual real money supply. We can see, that in the past decade under examination the intended demand continuously and significantly exceeded the intended supply, and – due to the effectiveness of the short side of the market – the actual real money supply as well. Nontheless it is true, that the 'excess demand' shows a decreasing tendency over time, but even then it cannot justify sufficiently the dynamic changes and the interrelation of money demand and money supply. Based on these findings, the disequilibrium hypothesis with regard to the period between 1991–2000 cannot be considered statistically verified.

4. CONCLUSION

Since the observation period is too short, the results allow us to draw conclusions of only limited validity. Nevertheless, the results of a many-sided model analysis provide the possibility to make a few careful and not too rigorous conclusions.

1. From the point of view of monetary processes the period between 1991–2000 cannot be regarded homogeneous. The Bokros-package of March 1995 has a relevance here from two aspects: on the one hand, the artificial acceleration of inflation, and on the other hand, the changing of the exchange rate mechanism. Its impacts can be well experienced during the modelling of the different monetary processes.

2. According to the evidence gained from our empirical analyses, the changes in the velocity of money in the transition period can be explained by the inflation rather than the changes in interest rates. This specific situation could arise, because the change in real interest rate (despite its high value) was basically determined not by the inflation, but by the change in the nominal interest rate.

3. Our research rendered it possible that in the Hungarian economy, money did not play an active, but rather a passive role in the period of transition. Our model-estimates confirm that the increase and at a later stage the decrease of inflation was caused by the decrease and the stabilization of real output (GDP) (with the exception of 1995, when inflation was artificially accelerated). The nominal money supply (with a short time-lag) basically adjusted to the changes of inflation.

4. Effectiveness of the two-factor, containing GDP and interest rate money demand function – when it is assumed that the demand of money adjusts immediately and perfectly to the quantity of money (which is equal to the money supply as well) – cannot be proven by empirical investigations. We can get far better results if we overrule the assumption that quantity of money = money supply, and the immediate adjustment of demand supply (money supply = money demand), and we transform the money demand function accordingly.

5. The *Cagan*-type money demand function provided surprisingly good results, with both adaptive and rational expectations. This might come as a surprise, because though the level of inflation in the Hungarian economy was quite significant throughout the period under consideration, yet there was no real danger of a hyperinflationary situation (that is why the *Cagan*-model was originally developed for). One explanation of this fairly good result could be, that the expectations about inflation played an important role in the changes of the domestic inflation process.

6. With respect to the mutual relations between the demand for money and the supply of money and to the balanced situation of the money market, the findings of our tests firmly suggest that we accept the equilibrium hypothesis against the disequilibrium hypothesis. Our estimates suggest, that the demand for money and the supply of money adjusted to each other relatively fast and well through the determining factors, and accordingly, significant disequilibrium did not occur during the whole period.

APPENDIX

Observed time series

Year/quarter	Variables						
	M2	GDPC	GDF	CPI 4	INFL	INTEREST	IPREM
1991 Q1	914.7	592.0	1363.0	100.00	34.2	31.94	23.06
1991 Q2	963.5	614.2	1398.4	109.17	37.3	28.05	28.47
1991 Q3	1018.1	623.6	1418.6	113.64	35.5	24.51	13.05
1991 Q4	1183.0	674.0	1481.9	117.66	33	18.49	16.78
1992 Q1	1311.8	666.6	1289.7	126.23	26.2	14.49	17.68
1992 Q2	1422.2	703.3	1333.3	133.46	22.2	9.72	10.11
1992 Q3	1519.8	724.8	1361.7	137.33	20.8	14.14	0.72
1992 Q4	1666.5	795.2	1433.9	144.26	22.6	22.61	-0.15
1993 Q1	1687.6	802.0	1265.3	157.38	24.7	24.10	-7.63
1993 Q2	1743.1	856.6	1318.2	162.37	21.7	23.65	-11.54
1993 Q3	1819.8	890.0	1353.5	167.81	22.2	25.46	-5.27
1993 Q4	1967.1	986.7	1438.2	175.11	21.4	30.44	10.41
1994 Q1	1957.5	1015.2	1294.3	183.83	16.8	32.03	7.68

(Continued on the next page.)

(Continuation.)

Year/quarter	Variables						
	M2	GDPC	GDF	CPI 4	INFL	INTEREST	IPREM
1994 Q2	2041.3	1093.6	1359.2	192.10	18.3	33.07	-3.30
1994 Q3	2102.4	1138.2	1396.8	200.52	19.5	32.66	4.71
1994 Q4	2246.0	1258.1	1480.7	211.18	20.6	30.54	1.07
1995 Q1	2262.8	1298.4	1338.8	228.84	24.5	30.54	-8.24
1995 Q2	2343.1	1380.5	1391.5	250.30	30.3	24.13	9.97
1995 Q3	2481.1	1407.8	1407.7	259.69	29.5	23.94	11.07
1995 Q4	2715.4	1527.3	1476.0	271.76	28.7	22.73	9.64
1996 Q1	2759.3	1577.1	1346.7	292.03	27.6	22.24	6.67
1996 Q2	2891.1	1692.8	1401.2	310.43	24	20.76	5.44
1996 Q3	3043.0	1717.0	1420.3	318.63	22.7	20.02	4.16
1996 Q4	3318.0	1907.1	1521.0	326.97	20.3	19.41	3.53
1997 Q1	3268.8	1910.5	1377.8	346.87	18.8	19.28	3.34
1997 Q2	3399.0	2101.2	1468.5	367.21	18.3	18.80	3.43
1997 Q3	3593.4	2147.5	1501.6	376.01	18	17.34	2.05
1997 Q4	3975.3	2381.5	1601.5	386.14	18.1	19.27	4.17
1998 Q1	3901.2	2276.6	1438.4	406.18	17.1	16.21	1.99
1998 Q2	4050.1	2479.0	1540.5	423.40	15.3	15.62	-5.26
1998 Q3	4242.8	2548.1	1582.7	426.38	13.4	14.71	2.98
1998 Q4	4593.4	2783.7	1676.9	429.81	11.3	14.00	14.07
1999 Q1	4618.2	2567.4	1487.8	444.78	9.5	12.51	6.01
1999 Q2	4733.1	2802.7	1601.4	461.92	9.1	10.78	6.87
1999 Q3	4930.7	2906.3	1654.4	471.54	10.6	10.64	5.58
1999 Q4	5310.2	3212.1	1775.1	476.23	10.8	10.73	5.85
2000 Q1	5282.6	2913.3	1585.4	488.38	9.8	11.78	1.69
2000 Q2	5370.5	3166.2	1696.6	503.95	9.1	12.00	2.47
2000 Q3	5595.1	3237.5	1728.2	517.73	9.8	12.20	0.89
2000 Q4	5968.3	3563.6	1851.4	525.71	10.4	12.00	3.51

Note. Signs and abbreviations used:
M2 – M2 money supply (HUF billion).
GDPC – Quarterly GDP in current prices (HUF billion).
GDF – Quarterly GDP in 1995 constant prices (HUF billion).
CPI 4 – Quarterly consumer price index (1991Q1=100).
INFL – 12 month consumer price index – CPI.
INTEREST – Average yield of three month discount treasury bills, computed for annual level.
IPREM – Interest premium, computed from the relation of the so-called non-covered interest parity, $rp = r - r^f - \Delta e / e$,
where r^f means the interest rate of the world market, $\Delta e / e$ means the rate of devaluation of the domestic currency.

REFERENCES

BÓTA, L. – BUDAVÁRI, P. – HOÓS, J. – MUSZÉLY, GY. – PEHARTZ, F. (2001): Ökonometriai modell a fiskális politika szolgálatában. *Statisztikai Szemle*, Vol. 79. No. 6. p. 469–489.

DUCK, N. W. (1993): Some international evidence on the quantity theory of money. *Journal of Money, Credit and Banking*, Vol. 25. February, p. 1–12.

ENGSTED, T. (1993): Cointegration and Cagan's model of hyperinflation under rational expectations. *Journal of Money, Credit and Banking*, Vol. 25. August, p. 350–360.

FISCHER, A. M. (1993): Is money really exogenous? Testing for weak exogenity in Swiss money demand. *Journal of Money, Credit and Banking*, Vol. 25. May, p. 248–258.

HENDRY, D. F. – ERICSSON, N. R. (1991): Modelling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, Vol. 35. May, p. 833–866.

HENDRY, D. F. – ERICSSON, N. R. (1991): An econometric analysis of UK, money demand in monetary trends in the United States and the United Kingdom by Friedman and Schwartz. *The American Economic Review*, Vol. 81. March p. 8–38.

HULYÁK, K. (2000): A pénzkeresleti függvényről (On the money demand function), *Statisztikai Szemle*, Vol. 78. No. 6. p. 435–443.
KIGUEL, M. A. (1989): Budget deficits, stability, and the monetary dynamics of hyperinflation. *Journal of Money, Credit and Banking*, Vol. 21. May, p. 148–157.
PHYLAKTIS, K. – TAYLOR, M. P. (1993): Money demand, the Cagan model and the inflation tax: Some Latin American evidence. *The Review of Economics and Statistics*, Vol. 75. February, p. 32–37.
QUANDT, R. E. (1988): *The econometrics of disequilibrium*. Basil Blackwell, Oxford, p. 296.
RAPPAI, G. (1989): A fogyasztási javak piacának nem egyensúlyi modellje. *Statisztikai Szemle*, Vol. 67. No. 7. p. 663–678.
TAYLOR, J. B. (1989): Monetary policy and the stability of macroeconomic relationships. *Journal of Applied Econometrics*, Vol. 4. December p. 161–178.
TAYLOR, M. P. (1991): The hyperinflation model of money demand revisited. *Journal of Money, Credit and Banking*, Vol. 23. August. p. 327–351.
VÁRPALOTAI, V. (2000): *A negyedéves felhasználás oldali GDP adatok visszabecslése 1991-ig.* Mimeo, MNB Közgazdasági és Kutatási Főosztály, Modellezési osztály, Budapest.
YASHIV, E. (1994): Money demand in a high inflation economy: The case of Israel. *The Review of Economics and Statistics*, Vol. 76. February, p. 186–191.

EMPLOYMENT RESPONSE TO REAL EXCHANGE RATE MOVEMENTS: EVIDENCE FROM HUNGARIAN EXPORTING FIRMS*

MIKLÓS KOREN[1]

This paper estimates the labor demand response of Hungarian exporting firms to real exchange rate movements. The use of firm level export–import data enables the separation of two channels through which the exchange rate affects labor demand. First, a real depreciation raises the forint-equivalent price of foreign competitors, thereby boosting demand for the firm's export and, hence, the firm's demand for labor. Second, by raising the cost of imported inputs, a depreciation has an adverse effect on employment through the cost channel. A higher marginal cost induces a decrease in production and thus shrinks labor demand. Since firms with higher export share tend to import more, this latter negative effect might offset the former positive one. The cost effect may be dampened if labor and imported inputs are substitutes.

The paper shows that the relative importance of the demand and cost effects is industry specific. The short-run exchange rate and employment elasticity stemming from the demand effect is around 0.04. This channel is most pronounced in the case of the Food and tobacco industry. Machinery, on the other hand, exhibits a cost effect of roughly –0.04. Surprisingly, there is no evidence that export share affects exchange rate exposure.

KEYWORDS: Labor demand; Exchange rate; Panel data.

The present paper addresses the question of how workers of Hungarian exporting firms are affected by movements in the Hungarian real exchange rate. In particular, it examines the changes in labor demand attributable to exchange rate movements. To the extent a weaker forint implies an expansion in Hungarian exports, firms may be willing to expand their labor force. Besides the magnitude and speed of this demand-driven adjustment, we are interested in the cost side of the exchange rate. It is a common observation that Hungarian exporting companies use a substantial amount of imported inputs. This means that a depreciation of the forint raises the marginal cost of production, thereby affecting the demand for labor non-trivially. This paper attempts to separate the demand- and cost-side effects.

* This paper is a revised and updated version of my master's thesis at Central European University. I wish to thank *Gábor Kőrösi* for his extensive help as my thesis supervisor. I also thank *János Vincze* and two referees for their comments. Any remaining errors are mine.

[1] PhD Student, Harvard University, Department of Economics.

Using a panel of large Hungarian exporting firms from 1992–1996, the paper quantifies the effect of real exchange rate movements on labor demand. Although this effect is well documented for U.S. industries, we are unaware of any such empirical study concerning transition countries. As these economies have opened up to international trade, it is important to quantify how much their labor force is exposed to external market conditions. The immediate observation that most Central and Eastern European countries have limited the fluctuation of their exchange rate in one fashion or another does not render the question irrelevant. Even if the nominal exchange rate is stabilized, the real exchange rate faced by firms can vary remarkably depending on foreign and home market conditions. As for the case of the HUF, we document firm-specific real exchange rates based on product- and firm-level data on Hungarian exports and imports and show that there is much more fluctuation in these rates than in the nominal HUF rate. Also, the recent tendency of more flexible exchange rate regimes further justifies the analysis of exchange rate movements.

The rest of the paper is organized as follows. Section 1 gives a brief survey on the existing literature on U.S. estimations. Section 2 derives the estimable equations from a standard dynamic model of labor demand. Estimation methodology and results are reported in section 3. Finally, section 4 concludes. There is an appendix describing the data used and some of the technical derivations of the estimation technique.

1. PREVIOUS RESULTS

In this section we present some of the empirical literature that have addressed a wide range of questions concerning labor market adjustment to exchange rate shocks. Besides net changes in employment (either in number of employed or in hours worked) and wages, gross job flows are also analyzed.

Employment and wage response

In one of the earliest papers, *Branson* and *Love* (1988) examine the effects of the real appreciation of the U.S. dollar in the early eighties. Using industry-level data from 1970 to 1986, they estimate how a labor-cost based real exchange rate can explain changes in U.S. manufacturing employment. They find a sizeable exchange rate effect, an average elasticity of 0.11 for non-durable goods and 0.29 for durable goods.[2] That is, a 10 percent real appreciation of the dollar results in a 1.1 percent, respectively 2.9 percent decline in industry employment. They also document that employment reacts very slowly, so the long-run coefficients may well be three to ten times higher than the short-run elasticities.

Burgess and *Knetter* (1998) estimate the *reduced form* of a simple dynamic labor demand model for G-7 countries. They examine the elasticity of employment with respect to exchange rates, as well as the speed of adjustment. They find that both of these relationships heavily differ among industries and countries. Their estimation for employment-exchange rate elasticity are rather mixed, ranging from −1.5 to +1.2.[3] They conclude that in roughly 30 percent of the country–industry pairs there is a significant em

[2] *Branson–Love* (1988) p. 249, significant coefficients only.
[3] *Burgness–Knetter* (1998), Table 3, significant coefficients only.

ployment response to exchange rates: an appreciation of the home currency results in loss of jobs. Germany and Japan seem somewhat less responsive to exchange rate fluctuations than the other G-7 countries. The stylized fact that the labor markets of Anglo-Saxon countries (United Kingdom, United States, Canada) adjust faster is also confirmed.

Dekle (1998) applies a partial equilibrium model of monopolistic competition on Japanese manufacturing firms. He calculates industry specific real exchange rates as demand shifters, by using industry specific GDP deflators and trade shares. Dekle founds that there is a sizeable impact of real exchange rate changes to employment. He rejects, however, that this impact is linked with the industries' exposure to foreign markets (either export or import shares). The estimated employment/exchange rate elasticity ranges from 0.71 to 1.26, but the speed of adjustment is rather slow.[4]

In contrast to the result of *Burgess* and *Knetter* (1998), *Campa* and *Goldberg* (1998) find very weak relationship between employment and exchange rates in U.S. manufacturing industries. They investigate both the *demand* and the *supply* of labor, measuring job, working hours and wage responses to fluctuations of the U.S. dollar between 1972 and 1995. Using industry-level data, they find that wages respond significantly to exchange rate changes (with an average elasticity of 0.04), whereas the number of jobs and working hours is virtually unaffected (with an elasticity of 0.01).[5] In explaining the differences in labor market adjustment among sectors, the competitive structure (measured by the price over costs markup) and the external orientation of the industry (a higher export share yields larger labor demand increase in response to a real depreciation, while a higher import share results in lower, or even in negative demand change) and the skill-level of the labor force have proven to be the most important factors.

Lebow (1993) criticizes those studying the behavior of tradable goods sectors alone. He argues that the cross sectoral adjustment of labor supply (although mobility is imperfect in his model) affects the wage level of the nontradable goods sector, as well. The overall effect of exchange rate on relative and aggregate real wage is thus ambiguous, its sign depends on the degree of labor mobility and the share of the tradable sector in production and employment. He also finds that wages are more responsive to export and import *price*s than the real exchange rate.

Job flows

The use of industry-level employment *stock* data can yield misleading predictions. In the last few years it has become fashionable to address the question with the use of firm-level *flow* data. Instead of number of jobs or total working hours, research is directed towards job creation and job destruction of firms and entry–exit decision of workers. The simultaneous creation and destruction of jobs (resulting from heterogeneity across firms) may cover a substantial part of labor market adjustment, even if net employment is unchanged. These gross job flows may be correlated with the firm's external exposure.

At the individual worker's level, employment stability has got little to do with net employment flows. Newly created jobs may require very different skills from those that a freshly laid-off worker has. Even in absence of these structural differences, higher gross

[4] *Dekle* (1998), p. 797.
[5] *Campa–Goldberg* (1998), p. 24.

job flows give rise to higher search unemployment. This is why the analysis of gross flows is essential in understanding the labor market consequences of real exchange rate movements.

Analyzing U.S. job flows, *Gourinchas* (1998) finds that although net industry employment responds very little to dollar exchange rate movements, there is a substantial response of job destruction and creation rate within the industries. A weaker dollar implies both less job creation and less job destruction, and, conversely, a strong dollar means more job adjustment. *Goldberg, Tracy* and *Aaronson* (1999) incorporate another important margin of labor market adjustment, the worker's switching of industry. They document that roughly half of the job-changers change their two-digit industry as well.[6] The authors use matched samples from the Current Population Survey, which contains individual-level data on employment, too. They estimate a probit model to explain the job or industry switching of the worker by industry specific export and import real exchange rates, and also examine the asymmetry of these effects. They find that manufacturing workers' likelihood to switch industry is largely effected by the import real exchange rate, that is, an appreciation reduces the likelihood of switching. The relationship is most pronounced in the non-durable goods sector and in sectors with high import shares. For non-manufacturing workers, both export and import exchange rates influence the probability of industry switching, while the probability of job changing remains unaffected. This means that sectoral composition of the labor market is responsive to exchange rate movements.

Similarly to *Gourinchas* (1998), *Klein, Schuh* and *Triest* (2000) turn to gross job flows in U.S. manufacturing industries during 1973 and 1993. They calculate job creation and destruction rates from firm-level employment data, showing substantial heterogeneity across industries in this respect. They find that the differences are connected with the sector's external exposure. Interestingly, this connection is asymmetric: whereas job destruction increases with an appreciation of the dollar (a 1 percent appreciation causing a 0.47 percentage point higher destruction rate), job creation seems to be unaffected.[7] This may be due to asymmetric adjustment costs: creating new jobs is likely to be more costly than laying off workers. The authors also find evidence that the responsiveness of job destruction is higher in industries more open to international competition.

Goldberg and *Tracy* (2001) improve upon *Goldberg et al.* (1999) by incorporating a wage equation in their estimation. This enables them to address the puzzling result that wages are much more responsive to exchange rates than employment is (see *Campa-Goldberg*, 1998, for instance). They show that these large wage changes are mostly associated with job transitions (though they are not captured by industry-level data), that is, workers remaining on the same job face little wage change.

2. ANALYTICAL FRAMEWORK

This section introduces the analytical framework used to derive the estimable equations.

[6] *Goldberg et al.* (1999), p. 206, Table 1.
[7] *Klein et al.* (2000).

Product Demand

Consider a Hungarian exporting firm, competing in two markets, the domestic (henceforth indexed by *H*) and the foreign (hence *F*). Demand for the firm's products Q^D depends on its price relative to its competitors,

$$Q^D = f\left(P^F / P^F_{comp}, P^H / P^H_{comp}\right), \qquad /1/$$

and both relative prices have negative effect on the demand.[8] If the demand elasticities are constant in both markets, then markups are constant, too. This implies that the prices charged depend only (positively) on the marginal cost of production, hence one can write total production as a function of competitor's prices and marginal cost shifters.

$$Q = g(\underset{+}{P^F_{comp}}, \underset{+}{P^H_{comp}}, \underset{-}{W}, \underset{-}{R}, \underset{-}{P_M}), \qquad /2/$$

where W is the wage rate, R is the rental cost of capital, and P_M is the price of imported inputs. All prices are expressed in domestic currency (henceforth referred to as HUF). Adopting a log-linear functional form,

$$q = \alpha_0 + \alpha_1 p^F_{comp} + \alpha_2 p^H_{comp} + \alpha_3 w + \alpha_4 r + \alpha_5 p_M \qquad /3/$$

where lowercase letters denote logarithmized variables. We expect α_1 and α_2 to be positive while α_3, α_4, and α_5 to be negative.

Labor demand

With any static production function, labor demand of the firm is a function of the production level, the wage rate, and the price of other inputs that either complement or substitute labor in production. Two such factors will be examined: capital and imported materials.

$$L = h(\underset{+}{Q}, \underset{-}{W}, \underset{?}{R}, \underset{?}{P_M}), \qquad /4/$$

where L is the amount of labor demanded (measured in the number of workers) and Q is the level of production. We will assume a log-linear form,

$$l = \beta_0 + \beta_1 q + \beta_2 w + \beta_3 r + \beta_4 p_M \quad . \qquad /5/$$

Here β_1 is expected to be positive, β_2 to be negative, while the signs of β_3 and β_4 are indeterminate. They depend on whether capital and imported inputs substitute or comple

[8] Insofar as Hungary can be considered a small open economy, the foreign disposable income does not affect demand for Hungarian products. Specifications with foreign and Hungarian spending have been tested but none of them entered significantly.

ment labor. If the factors are gross substitutes then the coefficient is positive, otherwise, if they are gross complements, it is negative. Substituting in for q from /3/ to avoid endogeneity bias in the estimation yields,

$$\begin{aligned} l = (\beta_0 + \alpha_0\beta_1) + \alpha_1\beta_1 p^F_{comp} + \alpha_2\beta_1 p^H_{comp} + \\ + (\beta_2 + \alpha_3\beta_1)w + (\beta_3 + \alpha_4\beta_1)r + (\beta_4 + \alpha_5\beta_1)p_M , \end{aligned} \qquad /6/$$

or, after redefining the parameters,

$$l = \gamma_0 + \gamma_1 p^F_{comp} + \gamma_2 p^H_{comp} + \gamma_3 w + \gamma_4 r + \gamma_5 p_M \ . \qquad /7/$$

This is the static version of the labor demand equation that we wish to estimate. We will also refer to the previous equation as $l = \boldsymbol{\gamma}^T \mathbf{x}$ with $\boldsymbol{\gamma}$ being the vector of coefficients and **x** being the vector of demand and cost shifters.

There are three channels through which exchange rate affects labor demand. First, a real depreciation raises the forint-equivalent price of foreign competitors (p^F_{comp}), thereby boosting demand for the firm's export.[9] This is what we call the *demand channel*. The exchange rate elasticity of labor demand, γ_1 is expected to be positive and increasing in the company's export exposure. Second, by raising the cost of imported inputs (p_M), a depreciation has an adverse effect on employment through the *cost channel*. A higher marginal cost induces a decrease in production and thus shrinks labor demand. Third, depending on the production function, an increase in imported material prices may cause substitution towards labor to the extent that these two factors are substitutable in production. This *substitution channel* dampens the effect of the cost channel. Their gross impact is summarized in coefficient γ_5. The overall employment-exchange rate elasticity will be $\gamma_1+\gamma_5$.[10]

Let us now turn to a dynamic version of the previous model.

Sluggish Adjustment

The static model of labor demand presented in the previous subsection is only valid in absence of adjustment costs. However, there are significant costs of *hiring* and *firing* workers. This means that the present employment decision is influenced by past employment and also by the expectation to future market conditions. Let us briefly examine the dynamic considerations arising from adjustment costs. We will only consider net changes in the labor force, leaving working hours unchanged.

In general, the firm would solve the following dynamic optimization problem:

$$\max_{\{l_t\}} \mathrm{E}_0 \sum_{t=0}^{\infty} \left[\delta^t \pi(l_t, \mathbf{x}_t) - C(\Delta l_t) \right] , \qquad /8/$$

[9] 'Real' deprecation means that domestic prices are unchanged.

[10] It has been assumed that the wage rate and the rental cost of capital are unaffected by the real exchange rate. Although wages can play a significant role in labor market adjustment to external shocks (see *Campa–Goldberg*; 1998), we anticipate that this occurs over a longer horizon due to some rigidity in the nominal wages. As for the cost of capital, a liberalized capital market ensures that it is relatively fixed.

where $\delta = 1/(1+r)$ is the firm's discount factor, π_t is profit, l_t is (log of) employment, $\mathbf{x}_t$ is a vector of exogenous variables (demand and cost shifters, among others, most importantly, foreign product prices and prices of other factors) in period t, $C(\cdot)$ is the adjustment cost depending on the one period net percentage change in employment, $\Delta l_t = l_t - l_{t-1}$.

The simplest specification of adjustment costs is the most widely used *quadratic form*:

$$C(\Delta l) = c\Delta l^2 ,$$

where c is a parameter representing the size of adjustment costs. Together with a quadratic profit function, this assumption ensures that labor demand evolves according to the following dynamics:[11]

$$l_t = \mu l_{t-1} + (1-\mu)(1-\delta\mu)\sum_{s=0}^{\infty}(\delta\mu)^s \mathrm{E}_t\, l^*(\mathbf{x}_{t+s}) , \qquad /9/$$

where l^* is the *optimal* level of labor demand that would prevail itself in absence of adjustment costs, and μ is a positive parameter depending positively on the magnitude of adjustment costs. Observe that optimal employment is a weighted average of previous period employment and present and all future optimal employments.

Now the question remains how expectation on future market conditions, $\mathbf{x}_{t+s}$'s can be specified. In the estimations we will assume that the exogenous variables follow a first-order vector-autoregression, that is,

$$E_t \mathbf{x}_{t+s} = \mathbf{A}^s \mathbf{x}_t ,$$

where $\mathbf{A}$ is the matrix of the VAR coefficients. Recall that $l^* = \boldsymbol{\gamma}^T \mathbf{x}$, implying that $E_t l^*_{t+s} = \boldsymbol{\gamma}^T \mathbf{A}^s \mathbf{x}_t$. If the Leontief inverse of $\delta\mu\mathbf{A}$ exists then /9/ simplifies to

$$l_t = \mu l_{t-1} + \underbrace{(1-\mu)(1-\delta\mu)\gamma(\mathbf{I}-\delta\mu\mathbf{A})^{-1}}_{\boldsymbol{\theta}} x_t . \qquad /10/$$

Here $\boldsymbol{\theta}$ denotes the vector of directly estimable short-run (or contemporaneous) parameters. However, we might also be interested in the long-run parameters, γ. If we know $\mathbf{A}$ and δ then we can calculate

$$\gamma = \frac{\boldsymbol{\theta}(\mathbf{I}-\delta\mu\mathbf{A})}{(1-\mu)(1-\delta\mu)} .$$

We will apply the previous dynamic framework to estimate the production equation, /3/, too. Since adjusting the production level can indeed be a costly decision (We have al

[11] See *Nickell* (1986), p. 502–504 for the details of the algebra.

ready discussed the role of hiring and firing costs, but adjustment costs also occur in investment and disinvestment), it is right to estimate /3/ in a dynamic setting. In this case, present production will also depend on last-period production, so lagged production will enter the labor demand equation, too.

$$l_t = \theta_0 + \mu l_{t-1} + \eta q_{t-1} + \theta_1 p^F_{comp,\,t} + \theta_2 p^H_{comp,\,t} + \theta_3 w_t + \theta_4 r_t + \theta_5 p_{M,t}\,, \qquad /11/$$

$$q_t = \alpha_0 + \lambda q_{t-1} + \alpha_1 p^F_{comp,\,t} + \alpha_2 p^H_{comp,\,t} + \alpha_3 w_t + \alpha_4 r_t + \alpha_5 p_{M,t}\ . \qquad /12/$$

These are the equations to be estimated. We are interested in θ_1, the demand-side exchange rate elasticity of employment, θ_5, the cost-side elasticity, and their production-equation counterpart, α_1 and α_5.

These are only the short-run elasticities. To see the magnitude of the long-run elasticities, we must take the speed of labor demand adjustment and the expectation for future market conditions into account. The sample is too short to forecast exogenous variables so we are not calculating long-run elasticities based on the VAR approach outlined former. A crude measure can be obtained if we assume that all the exogenous variables follow a random walk (i.e., **A**=**I**). The respective demand- and cost-side exchange rate elasticities of employment are $\theta_1/(1-\mu)$ and $\theta_5/(1-\mu)$.

3. EMPIRICAL ANALYSIS

In the following section I first describe the data and the econometric methodology used, then discuss the main empirical findings.

Data

The dataset consists of a panel of Hungarian exporting companies from 1992 to 1996. Data were matched from three different sources, the Customs Statistics, the firms' balance sheet and earnings statement data, and Eurostat's Extra-EU Trade Statistics. 'Foreign market' of the firms is identified with the European Union because this is the largest market segment that we have data on. This approximation is valid up to the extent that Hungarian export is oriented towards the EU. The median firm in the sample collects 65 percent of its export revenues from the EU, and this number is above 99 percent for the top decile, so we regard the use of EU data as a good approximation.

The Customs Statistics dataset contains the annual export and import traffic of Hungarian firms, both in value (HUFs and U.S. dollars) and in tons, so we are able to calculate unit value measures. The traffic is divided into product categories broken down to HS6 (the Harmonized System) level. The use of the Harmonized System makes prices and quantities comparable to European external trade statistics. Annual EU exports and imports are given in ecus and tons for each HS6 category. This enables us to calculate import and export unit values as a proxy for the average foreign price for each product. These prices are then converted to HUF and are averaged for each firm as described in Appendix 1. The companies were then matched with their balance sheets and profit and loss accounts, to obtain data on employment, sales and costs. Appendix 2 describes how the variables were constructed.

Those large companies were selected, whose performance depend largely on export markets. A firm was chosen if it exported at least 100 million HUFs and possessed at least 100 million HUFs book equity in 1994. The dynamic nature of the model restricts us to use only those companies who have data from consecutive years. Out of the resulting 356 firms we selected those in one of the four most export-oriented industries: Food and tobacco (SIC 15 and 16), Chemical industry (SIC 23 through 26), Metallurgy (SIC 27 and 28) and Machinery (SIC 29 through 35). This limited the number of firms in the unbalanced panel to 266 with an average span of 2.7 years, which means 707 observations.

Altogether, the companies in the sample represent a substantial fraction of Hungarian exports (see Table 1). Their share is between one fourth and one third, although in 1992 there are remarkably few firms in the sample (31 firms with an export share of 8.4 percent).[12] Although the selection by size can introduce selection bias, the large degree of representativity may justify this choice.

Table 1

Sample allocation

Industry	Number of firms						Average span (years)
	1992	1993	1994	1995	1996	Total	
Food and tobacco	8	28	43	44	51	67	2.6
Chemical	8	38	46	37	55	63	2.9
Metallurgy	5	23	28	29	29	50	2.6
Machinery	10	43	57	48	77	90	2.3
Total	*31*	*132*	*174*	*158*	*212*	*266*	*2.7*
Share in total Hungarian exports* (percent)	8.4	24.3	26.2	28.2	35.2	27.4	

* *Source*: Statistical Yearbook of External Trade (1997–1998). Külkereskedelmi Statisztikai Évkönyvek 1997, 1998. KSH. Budapest.

Since we have an unbalanced panel, we need to check whether falling out of the panel is endogenous, in which case we would encounter serious selection bias in the estimation. Fortunately, the variables of interest do not explain fall-out from the panel (we have estimated a probit equation to examine this problem), meaning that survival is indeed random.

Table 2 summarizes the external exposure of firms within each industry. We report the share of exports in total revenues as a measure of exposure to demand shocks, the share of import costs in total costs to capture the cost-effect of the exchange rate, and the share of foreign owned firms in the sample. This latter ratio may be relevant to the extent that foreign owned firms respond differently to external conditions than domestic firms. Sources of this difference may include greater market power in external markets and

[12] This is most probably because of the accounting and bankruptcy reforms of 1992; a lot of firms have gone out of business or changed their status and thus their tax registration number. Hence they cannot be linked into a panel and drop out of the sample.

transfer pricing between the multinational company and its Hungarian subsidy. After experimenting with several specifications (not reported in this paper), we were unable to identify significant differences between foreign and domestically owned firms.

Table 2

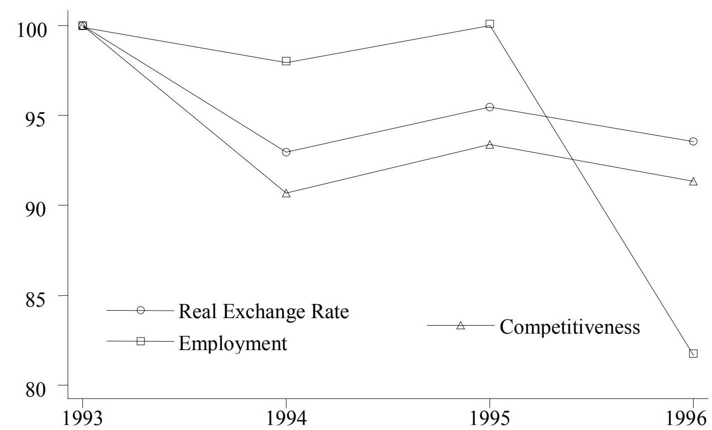
External exposure
(percent)

Industry	Export share in revenues	Import share in costs	Sample share of foreign owned firms
Food and tobacco	26.1	13.7	58.1
Chemical	47.6	29.5	59.8
Metallurgy	45.0	34.3	39.5
Machinery	65.8	40.3	58.3
Total	*47.2*	*29.8*	*55.6*

It is a surprising observation how much export and import exposure correlate. The industry with the lowest export share, Food industry (26.1%) has also the lowest share of imports (13.7%), and, conversely, that with the highest export share (Machinery, 65.8%) has the highest import share, too (40.3%). This amplifies the importance of looking at the cost-channel of exchange rate: it may well be the case that a weakening HUF has a negative impact on employment because the price increase of imports outweighs the expansion of demand.

To get a sense of how the exchange rate and employment co-move, let us have a look at Figures 1 through 4. They display cumulated change in the real exchange rate, the firm's external competitiveness, and employment averaged across each of the industries (1993=100).[13]

Figure 1. Mean employment and the real exchange rate: Food and tobacco industry
(1993=100)



[13] Year 1992 has been omitted because there are too few observations to calculate meaningful averages.

Figure 2. Mean employment and the real exchange rate: Chemical industry
(1993=100)

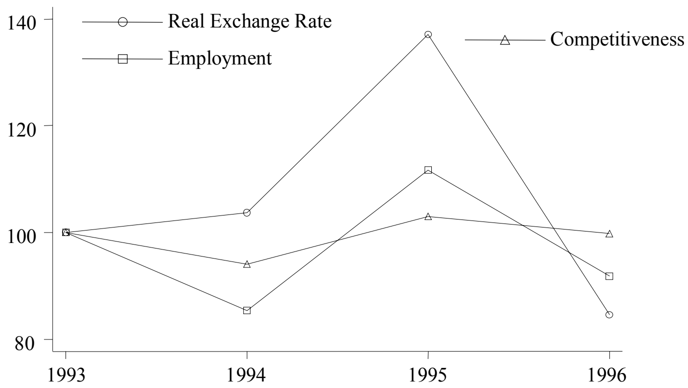


Figure 3. Mean employment and the real exchange rate: Metallurgy
(1993=100)

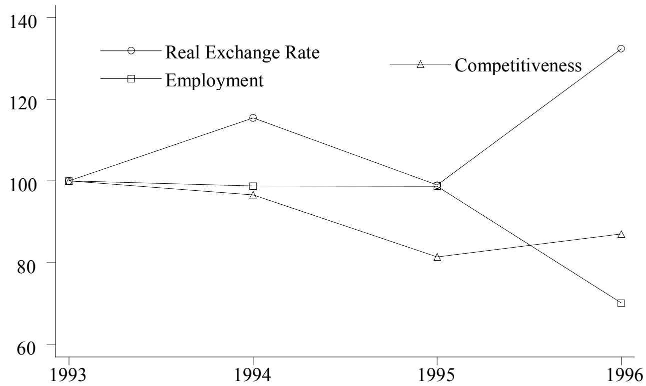
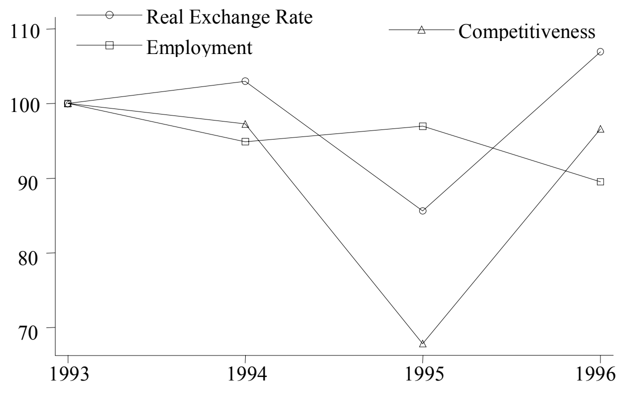


Figure 4. Mean employment and the real exchange rate: Machinery
(1993=100)



The real exchange rate is average foreign price relative to domestic price (in common currency), while competitiveness is measured as the foreign competitors' price relative to the firm's export price (see Appendix 1 for more details on variable definitions). We can see that the real exchange rate moves along a wide range during the years with the patterns differing remarkably across industries. For example, Machinery has experienced almost the opposite real exchange rate movements than the Chemical industry did. There is also a large within-industry variation in the real exchange rate not reported in the figures. These findings call for the use of industry-specific, or even firm-specific exchange rate instead of an aggregate macroeconomic measure.

Also observe that in the case of Chemical industry and Metallurgy competitiveness is less volatile than the real exchange rate although they move in the same direction. This means that the firm's export price reacts less to the exchange rate than the domestic price does. This may be due to either nominal price rigidities, or local currency price stability stemming from pricing-to-market behavior.[14]

We see significant co-movement of employment and real exchange rate for the Food and tobacco and the Chemical industry. In particular, a weaker forint is associated with a labor expansion. No pattern is visible for Metallurgy and even a reversed response can be seen in the case of Machinery. This may be due to the high import share of this industry: a weaker forint raises the price of inputs, thereby lowering the demand for labor. Let us now turn to an econometric evaluation of these findings.

Methodology

Recall the estimable equation from /11/. We have subtracted $l_{i,t-1}$ from both sides.

$$\Delta l_{i,t} = \theta_0 + (\mu - 1) l_{i,t-1} + \eta q_{i,t-1} + \theta_1 p^F_{comp,it} + \\ + \theta_2 p^H_{comp,it} + \theta_3 w_{it} + \theta_4 r_{it} + \theta_5 p_{M,it} + u_{it} \qquad /11a/$$

where i indexes firms, t indexes time, and u_{it} is the error term. Let us discuss some of the methodological problems of estimating /11a/.

– *Endogeneity of factor prices*. If the firm's size is not negligible relative to its factor markets, then factor prices may be correlated with the error term. Consider a firm-specific shock that raises the firm's demand for labor. (Macroeconomic shocks are controlled for by using time dummies.) This may well increase the equilibrium wage rate, and may also alter the other factor prices. Since the cost of imported materials is measured as a European average price (see Appendix 1), it is not likely to be affected by a small-country firm. In the case of capital, we argue that the large degree of capital mobility equates capital costs across regions, so the individual firm has little effect on its capital cost. This argument does not hold for labor, since it is rather immobile within Hungary. Hence we used an Instrumental Variable (IV) approach to correct for the endogeneity of wages. Current wage of the firm was instrumented by lagged wage and regional supply shifters, as unemployment rate and labor activity rate.

[14] See *Goldberg–Knetter* (1997) for an overview of the pricing-to-market literature.

– *Firm-specific error*. If there are omitted firm-specific factors that affect labor demand (e.g., firm-specific assets, managerial skills), then the error term is correlated with lagged employment, thus making the OLS estimation inconsistent. We can incorporate firm-specific error terms if we estimate the model with one of the panel methods. A fixed effect model is immediately ruled out, since the lagged dependent variable renders the estimation inconsistent.[15] A random effect specification can be employed, and the maximum likelihood estimator remains consistent in this case, as described in Appendix 2.[16]

To tackle the previous problems, the following estimation procedure was adopted. First, we fitted the wage rate on all the exogenous labor demand shifters herein, as well as supply shifters, such as lagged wage and regional measures of unemployment and labor force activity. We used the predicted wage instead of the actual wage in all of the model specifications later. This two-step procedure is equivalent to the standard IV-method if the second step is a linear estimation. If it is not, then the estimation is still consistent by a method-of-moments argument.[17] However, the asymptotic covariance matrix of the estimator may be different if we treat the predicted wage as given and do not take into account that it was estimated in the auxiliary regression. It can be argued that if the second estimator is close to linear, this bias of the covariance matrix should be negligible. Otherwise, we should bear this caveat in mind when testing the significance of coefficients.[18]

To test the specification of the labor demand model, we also estimated an equation for the production of the firm, using equation /12/ (and subtracting $q_{i,t-1}$ from both sides):

$$\Delta q_{it} = \alpha_0 + (\lambda - 1)\, q_{i,t-1} + \alpha_1\, p^F_{\text{comp},\, it} + \\ + \alpha_2\, p^H_{\text{comp},it} + \alpha_3\, w_{it} + \alpha_4\, r_{it} + \alpha_5\, p_{M,\, it} + w_{it}. \qquad /12a/$$

If we have omitted variables important to both production and labor demand decisions then the two error terms, u_{it} and w_{it} are likely to be correlated. This is why the seemingly unrelated regressions (SUR) technique is to be applied.

To test whether export exposure affects the demand channel of the exchange rate, we split the sample in two parts: those firms whose share of exports in total sales is above 50 percent are termed ‘High export share’ firms, the others have ‘Low export share.’ A dummy controls for high export share. The coefficient of foreign prices and domestic prices are allowed to vary with export share. We would expect that a high export share increases the effect of foreign prices and reduces that of domestic prices.

Results

Table 3 displays the results of the random effect estimation. Since industries differ with respect to their trade exposure and production function, we estimate the labor demand equation for each of them separately.

[15] This is known as the *incidental parameters* problem, see *Chamberlain* (1984) for a discussion.

[16] For a similar panel-problem while estimating investment response to exchange rates, *Nucci* and *Pozzolo* (2001) use GMM instead of the ML procedure described here.

[17] The first part of the moment function is the orthogonality condition of the auxiliary regression and the second part is derived from the second step, e.g., a maximum likelihood estimation.

[18] We have also tried the error component two-stage least squares procedure suggested by *Baltagi* (1995) and it did not change the results qualitatively.

Table 3

Random effect model, maximum likelihood estimates

Employment, difference	Pooled sample	Food and tobacco	Chemical	Metallurgy	Machinery
Lagged employment	-0.1642***	-0.2235***	-0.1123**	-0.1728***	-0.2433***
	(0.0250)	(0.0372)	(0.0536)	(0.0650)	(0.0446)
Lagged production	0.0696***	0.1034***	0.0146	0.0928*	0.1310***
	(0.0222)	(0.0348)	(0.0485)	(0.0514)	(0.0408)
Foreign prices	0.0186*	0.0363**	-0.0030	0.0268	0.0057
	(0.0099)	(0.0146)	(0.0186)	(0.0223)	(0.0300)
Domestic prices	0.0836	-0.0089	0.0788	0.0237	-0.1121
	(0.0512)	(0.1329)	(0.1294)	(0.1188)	(0.2056)
Wage rate	-0.1377*	-0.2726***	-0.0161	-0.2418*	-0.1753
	(0.0720)	(0.0963)	(0.1669)	(0.1399)	(0.1445)
Rental cost	-0.0217	0.0095	-0.0444	-0.0042	-0.0597
	(0.0219)	(0.0421)	(0.0456)	(0.0394)	(0.0413)
Import cost	-0.0038	0.0030	-0.0034	0.0149	-0.0405**
	(0.0068)	(0.0101)	(0.0120)	(0.0111)	(0.0186)
Foreign prices × High export share	-0.0187	0.0676	0.0098	-0.0977***	-0.0138
	(0.0136)	(0.0629)	(0.0247)	(0.0251)	(0.0380)
Domestic prices × High export share	-0.0582	-0.1706	-0.0390	0.1385	0.2150
	(0.0640)	(0.1809)	(0.0901)	(0.1416)	(0.2115)
	Short-run exchange rate elasticity				
Low export share	0.0148	0.0393**	-0.0064	0.0417*	-0.0348
High export share	-0.0039	0.1070*	0.0034	-0.0560***	-0.0487*
	Approximate long-run exchange rate elasticity				
Low export share	0.090	0.176**	-0.057	0.241	-0.143
High export share	-0.024	0.479	0.030	-0.324	-0.200
	Restrictions				
H_0: demand = cost = 0 [a)]	3.53	6.47**	0.13	4.00	4.85*
	Descriptives				
Number of observations	707	174	184	114	235
$\chi^2(16)$ [a)]	171.40***	65.44***	66.67***	60.57***	69.34***
$\sigma(v_i)$	0.1161***	0.0000	0.1245***	0.1549***	0.1284***
$\sigma(u_{it} - v_i)$	0.2144***	0.1831***	0.1520***	0.1152***	0.2735***
$\sigma(\Delta l_{it})$	0.2690	0.2217	0.2187	0.2115	0.3455
R^2	0.1785	0.3179	0.1929	0.1669	0.2352

Notes: All variables are in logs. The significance of coefficient or test: * 10 percent, ** 5 percent, *** 1 percent.
[a)] Likelihood ratio test, χ^2 distribution.

For all the industries except for the Food and tobacco industry we find that the firm-specific error term is highly significant. It explains a large fraction of the variation in annual change of employment. This validates the use of the random effect model.

The short-run wage elasticity of labor demand is of the expected negative sign and significantly different from zero in the case of Food and tobacco industry and Metallurgy. As for the magnitude, it is somewhat smaller than that was found by *Kőrösi* (1998), who uses a slightly different dynamic formulation for labor demand estimation on a larger Hungarian dataset.[19] The coefficient on lagged production, $\mu-1$, is rather high, that is, significantly greater than -1. This means that the speed of employment adjustment is fairly slow, just as in *Kőrösi*'s analysis (1998). Also note that lagged production enters significantly into three of the industries, suggesting that production adjusts sluggishly, too. Let us now turn to the parameters of key interest.

Although foreign prices enter with a significant positive coefficient in the pooled sample, it is only the Food and tobacco industry that has a significantly positive demand-side exchange rate elasticity (0.0363 for the low export share firms). This means that a 10 percent real depreciation of the HUF causes labor demand to rise by 0.36 percent in the same year. This number is higher for high export share firms, although not significantly. We do not find support for the hypothesis that export share affects exchange rate exposure.[20] It affects neither the coefficient on foreign prices, nor that on domestic prices. This may be due to several reasons. Firstly, as shown in Table 2, export and import shares are highly correlated across industries, meaning that a higher export share also means more pronounced cost effect thereby offsetting the increase in the demand effect. However, we expect the export and import shares to be less correlated within industries (we do not test this because we only have a crude measure of import share for separate firms) making this explanation unreasonable. Secondly, firms exporting more to the EU may have more market power in their foreign market. Then they absorb exchange rate fluctuations more in their markup than in their level of production and employment.

Import cost only affects labor demand of Machinery significantly. Here a 10 percent HUF depreciation implies that employment is cut by 0.41 percent the same year. It is important to note that this number may potentially be dampened by an incomplete exchange rate pass-through. As documented by *Goldberg* and *Knetter* (1997), it may well be the case that the price of imported inputs reacts less than one to one to exchange rate movements.

The overall effect of exchange rate on labor demand remains ambiguous. The null hypothesis that there is no effect whatsoever is only rejected in Food and tobacco industry and Machinery (the former exhibiting a positive, the latter a negative effect). Low export share firms in the Food and tobacco industry and Metallurgy have a total short-run exchange rate elasticity around 0.04. This number goes up to 0.11 for high export share Food and tobacco industry firms and falls down to −0.05 for firms in Machinery. On the other hand, the speed of labor adjustment is very slow (the reported $\mu-1$ coefficient and hence μ is large), especially in the case the Chemical industry and Metallurgy. This means that the *long-run* effect of exchange rate may be 4 to 9 times higher than the short-run. The approximate long-run elasticities are also reported in Table 3.

Table 4 and 5 report results from the estimation of the employment and the production equations. We use the SUR method to estimate the two equations jointly. That is, we allow the two error terms to be correlated but we do not allow them to be firm specific.

[19] *Slaughter* (1997) may also serve as a basis for comparison.
[20] This negative result is in line with the findings of *Dekle* (1998) but contradicts those of *Klein et al.* (2000).

This may introduce a bias in the estimates since, as we saw in the previous model, there is a significant firm-specific component of the error term. Nonetheless, the parameters in the employment equation are remarkably stable when compared to the random effect estimates, suggesting that this bias may be small. Short-run overall exchange rate elasticity varies between 0.04 and 0.12, which increase to 0.13 and 0.42 in the long run.

Table 4

SUR estimates: Employment equation

Employment, difference	Pooled sample	Food and tobacco	Chemical	Metallurgy	Machinery
Lagged employment	-0.1489***	-0.2755***	-0.0794*	-0.0601	-0.2166***
	(0.0206)	(0.0353)	(0.0441)	(0.0630)	(0.0388)
Lagged production	0.0694***	0.1434***	0.0063	0.0135	0.1248***
	(0.0190)	(0.0337)	(0.0400)	(0.0535)	(0.0365)
Foreign prices	0.0193**	0.0336**	-0.0040	0.0317	0.0087
	(0.0093)	(0.0154)	(0.0172)	(0.0222)	(0.0307)
Domestic prices	0.0891*	0.0174	0.0571	-0.1600	-0.1103
	(0.0481)	(0.1412)	(0.1301)	(0.1652)	(0.2063)
Wage rate	-0.1348**	-0.3622***	-0.0181	-0.0729	-0.1546
	(0.0623)	(0.0956)	(0.1411)	(0.1672)	(0.1334)
Rental cost	-0.0151	0.0148	-0.0164	0.0564	-0.0537
	(0.0201)	(0.0441)	(0.0403)	(0.0413)	(0.0395)
Import cost	-0.0026	0.0021	0.0133	0.0204	-0.0348*
	(0.0069)	(0.0108)	(0.0127)	(0.0161)	(0.0190)
Foreign prices × High export share	-0.0149	0.0807	-0.0048	-0.0679**	-0.0088
	(0.0131)	(0.0661)	(0.0251)	(0.0272)	(0.0384)
Domestic prices × High export share	-0.0880	-0.2396	-0.0454	0.3106	0.1736
	(0.0642)	(0.1824)	(0.1007)	(0.2040)	(0.2242)
	Short-run exchange rate elasticity				
Low export share	0.0167*	0.0358**	0.0093	0.0521**	-0.0261
High export share	0.0018	0.1165*	0.0045	-0.0159	-0.0349
	Approximate long-run exchange rate elasticity				
Low export share	0.112*	0.130*	0.117	0.867	-0.120
High export share	0.012	0.423*	0.057	-0.265	-0.161
	Restrictions				
H_0: demand = cost = 0 [a]	2.22	2.47*	0.56	2.90*	1.69
	Descriptives				
Number of observations	704	172	184	114	234
$F(K, N-K-1)$	11.02***	7.23***	3.43***	1.73*	4.54***
$\sigma(u_{iti})$	0.2468	0.1927	0.2002	0.2032	0.3125
$\sigma(\Delta l_{it})$	0.2690	0.2217	0.2187	0.2115	0.3455
R^2	0.1779	0.3142	0.2259	0.1912	0.2344

Notes: All variables are in logs. The significance of coefficient or text: * 10 percent, ** 5 percent, *** 1 percent.
[a)] *Wald* test, *F* distribution.

The production equation fits much better than the employment equation. Also, we find that production adjusts slower than employment does. As expected, the error terms of the two equations are positively correlated. This may be due to the omission of variables affecting labor demand and the scale of production in the same direction. The coefficients on domestic prices and rental cost are often significant with the wrong sign, most probably because of the poor proxies that we use.

Table 5

SUR estimates: Production equation

Employment, difference	Pooled sample	Food and tobacco	Chemical	Metallurgy	Machinery
Lagged production	-0.0943***	-0.1694***	-0.1268***	-0.0742**	-0.0724**
	(0.0139)	(0.0374)	(0.0226)	(0.0314)	(0.0291)
Foreign prices	0.0369**	0.0426	0.0065	0.1164***	0.0817*
	(0.0152)	(0.0306)	(0.0317)	(0.0391)	(0.0468)
Domestic prices	-0.3436***	-1.0569***	-0.0030	-0.7347**	-0.8773***
	(0.0789)	(0.2788)	(0.1706)	(0.2879)	(0.3204)
Wage rate	-0.3958***	-0.3792**	-0.5055***	-0.2412*	-0.3689**
	(0.0669)	(0.1466)	(0.1228)	(0.1290)	(0.1581)
Rental cost	0.0865**	0.1124	0.2260***	0.1987***	0.0143
	(0.0339)	(0.0877)	(0.0719)	(0.0727)	(0.0613)
Import cost	-0.0098	-0.0212	0.0517**	-0.0114	-0.0476*
	(0.0117)	(0.0215)	(0.0234)	(0.0283)	(0.0287)
Foreign prices × High export share	-0.0403*	0.3039**	-0.0462	-0.1252***	-0.0617
	(0.0222)	(0.1316)	(0.0462)	(0.0477)	(0.0598)
Domestic prices × High export share	-0.0311	0.3292	-0.1104	0.2329	0.2093
	(0.1090)	(0.3611)	(0.1851)	(0.3584)	(0.3484)
	Short-run exchange rate elasticity				
Low export share	0.0271*	0.0214	0.0582*	0.1050***	0.0341
High export share	-0.0132	0.3254**	0.0120	-0.0202	-0.0276
	Approximate long-run exchange rate elasticity				
Low export share	0.287	0.126	0.459	1.415	0.471
High export share	-0.140	1.921**	0.095	-0.272	-0.381
	Restrictions				
H_0: demand = cost = 0 [a]	2.96*	1.34	2.92*	4.75***	2.43*
	Descriptives				
Number of observations	704	172	184	114	234
F (K, N–K–1)	18.14***	6.84***	9.12***	6.23*	10.29***
σ (u_{iti})	0.4188	0.3827	0.3695	0.3558	0.4860
σ (Δl_{it})	0.4806	0.4526	0.4521	0.4174	0.5334
R^2	0.2548	0.3392	0.3796	0.3570	0.2161
	Cross-equation tests				
Correlation of error terms [b)]	0.4154***	0.3946***	0.4219***	0.4509***	0.4161***
H_0: No exchange rate effect in either equation [a)]	1.89	1.62	1.49	3.17**	1.78
H_0: Constant returns and no substitution effect [a)]	0.44	1.37	3.21*	1.55	0.22

Notes: All variables are in logs. The significance of coefficient or test: * 10 percent, ** 5 percent, *** 1 percent.
[a)] *Wald* test, *F* distribution.
[b)] *Breusch – Pagan* test, χ^2 distribution.

Foreign demand affects production significantly in the Food an tobacco industry, Metallurgy and Machinery. Demand effect ranges from 0.04 to 0.30. Import cost has a significant impact on Machinery only. Embarrassingly, export share does not explain exchange rate exposure; only in the Food and tobacco industry do we find that firms with higher export share react more to foreign demand. The overall exchange rate elasticity of production ranges from 0.03 to 0.33 in the short run. The cross equation hypothesis that exchange rate affects neither employment, nor production can only be rejected in the case of Metallurgy, an industry, which shows no significant effects of labor response to exchange rate.

We have also tested whether the coefficients on import cost are significantly different in the two equations. If we assume constant returns to scale (β_1=1), this would mean that there is some substitution between labor and imported inputs. In Table 5, we report the test of the joint hypothesis of constant returns and no substitution. Only the Chemical industry is significant, in which the coefficient of import cost is of the wrong sign.

4. CONCLUSION

The paper estimates labor demand of Hungarian exporting firms in response to real exchange rate movements. The use of firm-level export–import data enables us to separate two channels through which the exchange rate affects labor demand. First, a real depreciation raises the forint-equivalent price of foreign competitors, thereby boosting demand for the firm's export and, hence, the firm's demand for labor. Second, by raising the cost of imported inputs, a depreciation has an adverse effect on employment through the *cost channel*. A higher marginal cost induces a decrease in production and thus shrinks labor demand. Since firms with higher export share tend to import more, this latter negative effect might offset the former positive one.

We find that the short-run elasticity stemming from the demand effect is around 0.04. That is, a 10 percent real depreciation causes labor demand to rise by around 0.4 percent the same year. This effect is most pronounced in the case of the Food and tobacco industry. Machinery, on the other hand, exhibits a cost effect of roughly the same magnitude but of opposite sign. Since labor demand adjustment is sluggish, the long-run effect of the exchange rate can be an order of magnitude higher than the short-run.

Surprisingly, we do not find support for the hypothesis that export share affects exchange rate exposure. This may be either because a higher export share also means a higher import share and cost effect, thereby offsetting the increase in the demand effect, or because firms exporting more have more market power in their foreign market.

The results suggest that the analysis of market power deserves more attention. This could be accomplished by a more thorough investigation of the pricing behavior of exporting firms and its interconnection with employment decisions.

APPENDIX

Here we provide definitions of the primary and the constructed variables. The datasource is given in brackets.

1. VARIABLE DEFINITIONS

– *Industries*. Industries are identified by their two-digit classification. Food and tobacco is 15 and 16, Chemical is 23 through 26, Metallurgy is 27 and 28, and Machinery is 29 through 35. (Earnings statement.)

– *Products*. A product is defined as a HS6 category. Though some of these categories are rather broad, this is the deepest possible, internationally comparable classification. Prices of different products are then averaged geometrically for each firm and year pair. (Customs Statistics.)

– *Production*. The level of production is defined as the value of sales at 1991 HUF prices. Export and domestic sales are deflated by the appropriate industry producer price index. Since material inputs are incorporated into the model, total sales, not value added is used to measure production. (Earnings statement, Statistical Yearbook of Hungary, 1993, 1995, 1997.)

– *Foreign price*. Foreign competitors' price is calculated as the CIF (Cost Insurance Free) value of total import shipments to the European Union divided by the net weight of the shipments (unit value). The values are originally expressed in ecus, and are converted to HUF with a firm-specific exchange rate using the following procedure.

Because sales are unevenly distributed throughout the year, different firms face different average nominal exchange rates within a year. Since the value of export shipments is reported in both dollars and HUFs, we have a firm-specific HUF–USD exchange rate for each product. From this we can calculate the *estimated month of shipment* within the year (as if the shipment were made in a single month), and use that month's average HUF–ecu exchange rate to convert EU-prices to HUF. This procedure ensures that we do not impose a common, average exchange rate on the firms thereby losing information. The EU prices of products are then averaged for each firm and year to get firm-specific foreign price indexes. We use geometric average with the net weight of exports as weights. (EUROSTAT, National Bank of Hungary.)

– *Domestic price*. The domestic producers' price index in the 4-digit SIC industry of the firm. (Statistical Yearbook of Hungary, 1993, 1995, 1997.)

– *Real exchange rate*. Real exchange rate is measured as the price of foreign competitors relative to domestic prices. If relative foreign prices go up, the real exchange rate rises, that is, we have a real depreciation. (EUROSTAT, Customs Statistics.)

– *Competitiveness*. With data on the actual unit value of each shipment, we are able to construct a better measure of external competitiveness. This is calculated as the price of foreign competitors relative to the export price charged by the firm. (EUROSTAT, Customs Statistics.)

– *Wages*. Total labor cost over average annual number of workers. (Earnings statement.)

– *Capital cost*. Capital cost is proxied by depreciation cost divided by the stock of fixed assets. Here we follow the lead of *Kőrösi* (1998), who uses the same proxy in his labor demand estimations. (Earnings statement, Balance sheet.)

– *Price of imported inputs*. The ecu price of an imported product is calculated as the FOB value of exports of the product from the EU divided by the total net weight. It is converted to HUF as outlined former. A price index is calculated as a weighted geometric average of individual prices. (EUROSTAT, Customs Statistics.)

2. ESTIMATION PROCEDURE

Let v_i denote the firm-specific error term of the equation. Assume that the error terms are jointly normally distributed and are independent of the explanatory variables. In particular,

$$v_i \,|x_{i1},\ldots x_{iT}, l_{i0} \overset{i.i.d.}{\sim} N\left(0,\tau^2\right),$$
$$u_{it} \,|\, x_{i1}, x_{i2},\ldots,x_{it}, l_{i0},\ldots l_{i,t-1} \overset{i.i.d.}{\sim} N\left(v_i,\sigma^2\right),$$

that is, the error term (u_{it}) is independent of the exogenous variables and past employments, *conditional* on knowing the firm-specific omitted variable, v_i. In the estimation we treat l_{i0}, i.e., the first realization of employment as given, and express all the distributions conditional on it.[21]

[21] Here we have also assumed that v_i is independent of l_{i0}. For a more general error structure and an endogenous treatment of l_{i0}, see *Chamberlain* (1984, 1999).

The joint density of the error terms for firm i conditional on the firm-specific omitted variable is the product of the per-period conditional density functions.

$$f(u_{i1},u_{i2},\ldots u_{iT} \mid x,l_{i0},\nu_i)=$$
$$= f(u_{i1} \mid x,l_{i0},\nu_i) f(u_{i2} \mid u_{i1},x,l_{i0},\nu_i)\ldots f(u_{iT} \mid u_{i1},u_{i2},\ldots,u_{iT},x,l_{i0},\nu_i)=$$
$$= \prod_{t=1}^{T} f(u_{it} \mid u_{i1},u_{i2,}\ldots,u_{i,t-1},x,l_{i0},\nu_i) \qquad \text{/App. 1/}$$

The marginal probability density function of (the vector) u_i is obtained from the joint density of v_i and u_{it} (which is the product of the previous two densities since v_i and u_{it} are independent) by integrating out v_i.

$$f(u_{i1},u_{i2},\ldots,u_{iT} \mid x,l_{i0}) = \int_{-\infty}^{\infty} f(u_{i1},u_{i2},\ldots,u_{iT} \mid x,l_{i0},\nu_i)\, dF(\nu_i) =$$
$$= (2\pi)^{-T/2} \det(\mathbf{\Sigma})^{-1/2} \exp\left[-\frac{1}{2}\mathbf{u}_i'\mathbf{\Sigma}^{-1}\mathbf{u}_i\right], \qquad \text{/App. 2/}$$

where $\mathbf{u}_i=(u_{i1},u_{i2},...,u_{iT})^T$ is the vector if error terms for firm i and

$$\mathbf{\Sigma} = \begin{bmatrix} \tau^2+\sigma^2 & \tau^2 & \tau^2 & \ldots \\ \tau^2 & \tau^2+\sigma^2 & \tau^2 & \ldots \\ \tau^2 & \tau^2 & \tau^2+\sigma^2 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

is its covariance matrix.

One may observe that this likelihood contribution for firm i is the same as in a model without a lagged dependent variable with the minor difference that the standard panel likelihood is a product of independent marginal densities while here we have a product of conditional densities. Since the functional form is the same, the standard maximum likelihood estimation procedure can be used for this dynamic version of the model.

REFERENCES

BALTAGI, B. H. (1995): *Econometric analysis of panel data*. John Wiley and Sons, New York.

BRANSON, W. H. – LOVE, J. P. (1988): *The real exchange rate, employment, and output in manufacturing in the U.S. and Japan*. In: R. C. MARSTON (ed.), Misalignment of Exchange Rates: Effects on Trade and Industry, University of Chicago Press, Chicago. Chapter 9.

BURGESS, S. – KNETTER, M. M. (1998). An international comparison of employment adjustment to exchange rate fluctuations. *Review of International Economics*, Vol. 6. No. 1. p. 151–163.

CAMPA, J. M. – GOLDBERG, L. S. (1998): Employment versus wage adjustment and the U.S. dollar. *NBER Working Paper*, No. 6749.

CHAMBERLAIN, G. (1984). *Panel data*. In: Z. GRILICHES – M. D. INTRILIGATOR (eds), Handbook of Econometrics, Vol. 1. North Holland. Chapter 22.

CHAMBERLAIN, G. (1999): *Econometric methods*. Lecture Notes. Harvard University.

DEKLE, R. (1998). The yen and Japanese manufacturing employment. *Journal of International Money and Finance*, Vol. 17. p. 785–801.

GOLDBERG, L. – TRACY, J. (2001): Exchange rates and wages. *NBER Working Paper*, No. 8137.

GOLDBERG, L. – TRACY, J. – AARONSON, S. (1999): Exchange rates and employment instability: Evidence from matched CPS data. *American Economic Review. Papers and Proceedings*, Vol. 89. No. 2. p. 204–210.

GOLDBERG, P. K. – KNETTER, M. M. (1997): Goods prices and exchange rates: What have we learned? *Journal of Economic Literature*, Vol. 35. p. 1243–1272.

GOURINCHAS, P. O. (1998): Exchange rates and jobs: What do we learn from job flows? In: B. BERNANKE – J. ROTEMBERG (eds), *NBER Macroeconomics Annual*, NBER, Cambridge MA.

HAMERMESH, D. S. (1986): The demand for labor in the long run. In: O. ASHENFELTER – R. LAYARD (eds), *Handbook of Labor Economics*, North-Holland, chapter 8.

KLEIN, M. W. – SCHUH, S. – TRIEST, R. K. (2000): Job creation, job destruction, and the real exchange rate, *NBER Working Paper*. No. 7466.

KŐRÖSI G. (1998). Labour demand during transition in Hungary: Econometric analysis of Hungarian exporting firms. *Budapest Working Papers on the Labour Market*, No. 5.

LEBOW, D. E. (1993): Import competition and wages: The role of the nontradable goods sector. *Review of Economics and Statistics*, Vol. 75. No. 3. p. 552–558.

MARSTON, R. C. (1990): Pricing to market in Japanese manufacturing. *Journal of International Economics*, Vol. 29. p. 217–236.

NICKELL, S. J. (1986): Dynamic models of labour demand. In: O. ASHENFELTER – R. LAYARD (eds), *Handbook of Labor Economics*, North-Holland. Chapter 9.

NUCCI, F. – POZZOLO, A. F. (2001): Investment and the exchange rate: An analysis with firm-level data. *European Economic Review*, Vol. 45. No. 2. p.259–283.

SLAUGHTER, M. J. (1997): International trade and labor demand elasticities. *NBER Working Paper*, No. 6262.

PRICING INITIAL PUBLIC OFFERINGS IN PREMATURE CAPITAL MARKETS: THE CASE OF HUNGARY*

IBOLYA SCHINDELE[1] – ENRICO C. PEROTTI[2]

This paper investigates the determinants of underpricing at initial public offerings in the Hungarian Initial Public Offerings (IPO) market in 1990–1998, a period of transition from socialist to market economy and immaturity of the domestic capital market. The evidence suggests that political issues played a significant role in the process**:** we have found greater discount at privatization IPOs than at private issues, and a positive relation between underpricing and the proportion of shares offered for compensation coupons. These findings reinforce the hypothesis that governments in transition may pursue political objectives by selling shares at discount. Besides, the results show larger initial returns at early IPOs compared to later issues, which implies a negative relation between the discount and the maturing of the capital market. Most of the asymmetric information theories, empirically justified for well-developed stock markets, receive no support. Some results suggest that the transition related determinants of underpricing disappear as the securities market becomes more mature.

KEYWORDS: Initial public offerings; Underpricing; Privatization.

A common feature of Initial Public Offerings (IPO) in capital markets all around the world is that firms systematically sell equity at significant discount compared to the true value of the company. Empirical evidence suggests that investors' initial return, the percentage price change between the offer price and the first day closing price at the stock exchange, is usually higher than 10 percent, and in many cases it reaches 45-50 percent.

This paper investigates the structure of Initial Public Offerings of shares in Hungary. As in other capital markets IPOs in Hungary appear to be underpriced, so that buying shares at the offer price at issue and selling them on the first day of trading post flotation provides investors a positive return on average. In the period 1990–1998, approximately 60 companies offered shares publicly in Hungary. These issues were underpriced by 22 percent on aver

* We would like to thank *Gábor Kőrösi* for his encouragement and valuable suggestions during the realization of this project, as well as *György Martin Hajdu* for his useful comments. Discussions with seminar participants at the Summer School in Financial Markets, at the Central European University, helped also to improve the paper. Financial support from PHARE '97 is greatly acknowledged.

[1] PhD candidate at the University of Amsterdam and Tinbergen Institute Amsterdam.

[2] Professor at the University of Amsterdam and Fellow at Centre for Economic Policy Research, London.

age, which is not high by international standards, particularly for an emerging market (*Ibboston–Ritter*; 1994).

Table 1

Average underpricing

Country	Sample size	Time period	Average return (percent)
Belgium	28	1976–1989	10.2
Chile	19	1982–1990	16.3
France	187	1983–1992	4.2
Korea	347	1980–1990	78.1
Malaysia	132	1980–1991	80.3
Portugal	62	1986–1987	54.4
Spain	71	1985–1990	35.0
United Kingdom	2 133	1959–1990	12.0
United States	10 626	1960–1992	15.3

Source: *Ibboston–Ritter*; 1994.

The paper seeks to establish whether this significant level of underpricing in the early period of development of the Hungarian market reflects the usual factors suggested by classical IPO theories or rather some specific features of the transition from socialist to market economy. In the period 1990–1998, economic transition is a distinguishing characteristics of the capital market in Hungary: a large proportion of IPOs is related to privatization and the market is at a premature stage of development.

The next section reviews classical theories of IPO pricing and reveals the specific characteristics of the Hungarian IPO market, which provide a background for our hypotheses. The second part of the paper describes the characteristics of issues in Hungary and details our empirical results.

1. THEORETICAL ISSUES

In this section we introduce theories of underpricing based on asymmetric information. These theories allow us to built several hypotheses. Further on the privatization and transition-related factors of underpricing are reviewed with a particular attention to the Hungarian case. At the end of this section the specific privatization-related hypotheses are given.

Theories based on asymmetric information

Several theoretical models conclude that underpricing results from asymmetric information among groups of agents taking different roles in the IPO process. Underpricing is then an incentive used to stimulate the uninformed group to act in the interest of the informed one.

Asymmetric information may exist between a firm and its investment banker. *Baron* (1982) considers a principal-agent model for new issue underpricing based on this type of information asymmetry. The investment bank being an agent of the firm has superior information concerning its value. His compensation is a function of the proceeds from issue and the post-flotation price. The price discount, therefore, serves to induce the investment banker to put enough effort in advising and selling the firm's shares. Evidence found by *Muscarella* and *Vetsuypens* (1989) on investment banks going public, however, refuted this model. Underpricing proved to be significant at IPOs by investment banks as well, even though no asymmetric information existed since issuers acted as their own agents in the going public process.

Another approach to underpricing has focused on differential information of investors participating in the IPO market. *Rock's model* (1985) is based on information asymmetry between two groups of investors: informed and uninformed. The informed group knows well the prospects of firms and therefore is able to avoid buying low value IPO shares. Uninformed people have no information on firms' value, which results in a bias in their purchases towards less profitable equity issues. Anticipating this rationing bias, the uninformed group has no incentive to participate in buying shares. Underpricing is necessary to make this group enter the IPO market.

Rock's explanation was extended by *Beatty* and *Ritter* (1986) who studied the degree of underpricing as a function of ex-ante uncertainty concerning firm value. An increasing in uncertainty should be associated with a raise in information asymmetry, and thus with higher underpricing. As a proxy for this uncertainty, Beatty and Ritter use the post floatation variability of the share price. A positive relationship is found between this proxy and the discount involved in the offer price. As small firms have more volatile prices, this model may imply a negative relationship between firm's size and the level of discount. This hypothesis is supported by empirical studies on equity issues in the United States where large firms' offerings seem to be less underpriced. Thus measures of informational asymmetry such as size and post flotation price volatility should be regarded as explanatory factors for underpricing.

Signaling theories of underpricing focus on another type of asymmetry: differential information between firms and investors. These models assume that it is the selling entrepreneur who has superior information about the value of the IPO firm. To overcome adverse selection, companies with favourable prospects are interested to signal their value and thereby convince potential investors to buy shares. Signaling models for IPOs differ in the mechanism through which signaling occurs.

The first mechanism is signaling by equity retention. *Leland* and *Pyle* (1977) claim that firm value is positively related to equity retained at initial share issues. Managers of profitable companies intend to convey information about their quality to outsiders: retaining equity might be a signal of high quality. Since managers of firms with less attractive prospects have incentives to mimic good quality by issuing the signal, misleading signals have to be costly. Retaining equity can serve as a credible signal, since it is costly for firms that produce lower than average results. The higher the level of equity retained by the owners of a low value firm, the lower is the value of their residual investment and the higher the undiversified risk of their portfolio. Signaling good quality by ownership retention is then only in the interest of high value firms. This result is strongly supported by several empirical tests (*Downes–Heinkel*; 1982, *Ritter*; 1984, *Keasey–McGuinnes*; 1992).

Underpricing is another type of signalling mechanism. Signaling by equity retention implicitly assumes that the firm has only one opportunity to sell equity in the short term. In practice, most entrepreneurs sell off their equity holdings in several stages.

Models suggesting that firms underprice initially to let investors realize larger proceeds from secondary issues were proposed by *Welch* (1989), *Allen* and *Faulhaber* (1989) and *Grinblatt* and *Hwang* (1989). Pricing initial offerings at a discount is a credible signal of firm quality: only good firms are expected to recoup the loss due to initial underpricing.

A firm that retains a large proportion of its equity may not need to discount substantially the offer price. Observing equity retention, investors rationally think the firm makes a good investment since managers have information on the firm's true quality. Alternatively, when the issuer wants to sell a large proportion of its ownership, underpricing is necessary to convince investors that the firm is of high value. Equity retention and underpricing might thus be substitute signals, which ceteris paribus, implies a negative relationship between them.

Besides underpricing and quantity retained, firms have *other signals* available. In the context of an IPO, original shareholders can signal firm's quality by the name and reputation of advising agents they hire. Advising agents of high reputation may reinforce the public's confidence in high firm value and bring about a reduction in asymmetric information and thus required underpricing. Several papers in the literature discuss the relation between pricing and quality of the advising agents. *Titman* and *Trueman* (1986) consider the choice of an auditor as an important signal, while *Booth* and *Smith* (1986) emphasize the quality of the underwriter. The correlation between these variables and firm's value is well supported by empirical findings (*Keasey–Short*; 1997).

Indebtedness may be a further factor referring to the future prospects of firms, since a high level of debt discourages equity investment (*Myers*; 1977, *Jensen–Meckling*; 1976). Thus indebted firms may face hurdles to signal high value at an IPO. As a consequence, it could be expected that a more indebted firm should underprice its share issue to a greater extent.

Hypotheses built on IPO theory

The previously reviewed theories imply a number of hypotheses that serve as a basis for our empirical investigation:

1. Underpricing negatively depends on firm's size. Large firms price their issues more accurately. We consider firm size (ASSET)[3] as the balance sheet value of total assets of the firm preceding the issue.

2. The level of underpricing is a positive function of risk (RISK), which is to be measured by the ex-post variability of share price returns.

3. Underpricing and equity retained by the issuer (RETAINED) are negatively related, since they are substitutable methods of signaling high firm value.

[3] Further reference to each explanatory factor is under the name indicated in brackets. See the Appendix for a more detailed description of the explanatory variables.

4. Underpricing is lower when the advising agents of the issuing firm have good reputation. Thus we hypothesize a negative relationship between the quality of underwriter (DBROKER) and underpricing.

5. The more indebted the firm, the more its shares are underpriced at an IPO. High level of indebtedness (DEBT), potentially a bad quality signal, must be outweighed by other means, like underpricing. A positive relationship is hypothesized therefore.

Privatization IPOs

In our sample, many IPOs represent privatization sales. This raises the question whether other factors than those cited in the classical literature contribute to explain the underpricing phenomenon in Hungary.

A theoretical model of IPO pricing at privatization is proposed by *Perotti* (1995). This explanation suggests that underpricing and equity retention serve the goal of signaling commitment of the selling government to a privatization policy without future redistribution of asset value. The provision of informative signals for the public is necessary because at the time of privatization, significant uncertainty exists about the government' s real intentions towards the firm to be sold. For political and social reasons, governments in transition might engage in value redistribution, which may take place via taxation, indirect interference, through regulation or entry-deregulation.

A committed government is more willing to bear the costs of a delay in the final sale by retaining a part of the shares, as well as the costs of underpricing the issue. Gradual sales imply that the government is willing to bear the risk of a falling price after the sale, suggesting that it does not expect any change in its policy. When fast privatization is a prime objective, however, the government needs to sell quickly a large proportion of the shares, underpricing is needed as a costly signal of commitment. Thus ownership retention and underpricing are substitutable signals for the government to convey its political commitment. If the privatization program is not reversed, stock prices should rise over time and the government may earn substantial amounts when selling shares at subsequent issues at a smaller discount (*Laeven–Perotti*; 2001).

In the absence of large political uncertainty, privatization sales may not be underpriced to a higher degree than private issues. For several countries, however, there is much in evidence that privatization offers are more underpriced than their private issue counterparts (*Jenkinson–Mayer*; 1988, *Perotti–Guney*; 1993). *Dewenter* and *Malatesta* (1996) compare returns on initial privatization offers and on private IPOs in several countries with different levels of capital market development. Although their evidence indicates that in the United Kingdom privatization IPOs are underpriced more than private issues, they have opposite findings for other countries. While they cannot identify a general tendency for privatization sales to be more underpriced than private IPOs, they do find that issues subject to larger potential political risk are characterized by a greater discount.

Besides the goal of credible privatization, underpricing may result from other political objectives of the government. The most important of these considerations are buying political support, targeting dispersed ownership, giving ownership to employees, and promoting capital market development. A hypothesis advanced in the literature is that underpricing may serve these goals (*Schmidt*; 1997, *Biais–Perotti*; 2001).

Privatization IPOs in Hungary

Apart from explanations discussed in the corporate finance and privatization literature, specific characteristics of the privatization process in Hungary could also have an impact on the pricing of issues.

The privatization experience in Hungary can be characterized as following a variety of approaches. The methods applied in a particular stage of transition depended on the preferences of the ruling government. The transformation of ownership rights started in 1988. Selling companies through initial public offers became common only in the later stages of privatization. In 1990 and 1991, most of the initial offers were share sales or capital increases by private firms. Only 3 out of 13 issues were privatization IPOs, while in the years 1992–1994, 14 issues out of 18 were sales by the State Property Agency (SPA), the organization responsible for privatization of the state property. From 1995, ownership transformation through IPOs became more widespread, partly as a consequence of the implementation of the Law on Privatization (1995). Table 2 shows the distribution of IPOs in our sample, which is a good reference for the number of issues in different stages of the privatization process.[4]

Table 2

Distribution of IPOs in the sample, 1990–1998

Year	Privatization	Private issue	*Total*
1990	1	3	*4*
1991	2	7	*9*
1992	2	0	*2*
1993	5	2	*7*
1994	7	2	*9*
1995	5	1	*6*
1996	2	3	*5*
1997	4	4	*8*
1998	1	2	*3*
Total	*29*	*24*	*53*

An important feature of the privatization process in Hungary is its fundamental relation to restitution. Compensation of families, which suffered expropriation under the socialist regime, was a primary goal of the governments during the transition era. The main form of compensation was free distribution of the so-called 'compensation coupons' which could be exchanged for ownership in firms under privatization at initial share issues. These coupons were introduced to the Budapest Stock Exchange and were traded at prices usually significantly below face value. In spite of the fact that they were interest paying securities, in some periods their market value did not exceed 40 percent of the face value.

At most of the initial issues, compensation coupons were counted in the share price at full nominal value plus accrued interests. Buying coupons at the market price (or receiv

[4] Table 2 is based on a group of 53 initial offers, which is somewhat less than the total number of IPOs in the period 1990–1998, but representative for the distribution of issues. Section 2 gives an explanation on the difference.

ing them for free) and exchanging them for shares at IPOs, brought substantial additional returns to investors above the proceeds from classical underpricing. Thus, selling shares for coupons resulted in a special form of targeted underpricing at privatization IPOs. The proportion of stocks offered in exchange for compensation coupons provides therefore reference whether compensation policy and its related political objective had some role in underpricing the issues.

Aiming at maximizing privatization revenues, the SPA organized its sales in several rounds.[5] To improve the market's sentiment, at initial offers a large underpricing was applied, while substantial revenues were generated by decreasing the discount at later share issues. When aftermarket share performance is an important consideration, equity retention and underpricing may be thus positively correlated at the initial sale. When uncertainty concerning the government's future policy is high, a large initial underpricing is needed to improve investors' confidence. At the same time, most of the shares should be retained for later offers at which the discount may be set smaller. If over time political risk is resolved, gradual sales generate higher proceeds (*Laeven–Perotti*; 2001). This confronts the implication of *Perotti* (1995) that after controlling over political risk, underpricing and equity retention are substitutable signals of commitment.

A core characteristic of privatization in Hungary was the choice of selling companies to (usually foreign) strategic investors through private placements. The presence of strategic investors may refer to the possibility of more efficient restructuring and may therefore create more confidence from the public and a higher demand for shares. If the presence of strategic investors increases the public's confidence in the firm, their equity share and underpricing might be substitutable signals of high value, which generates a negative dependence between the two. The proportion of shares held by strategic investors after the issue might thus be an interesting explanatory variable.

Another goal of the privatization process was the allocation of shares to employees. Selling underpriced shares to employees helps to create support from insiders, which is an essential condition for the success of privatization. Without employees' acceptance, a change in ownership might be impossible to realize. In Hungary, at several IPOs a significant fraction of shares was separated and offered exclusively to the firm's employees. Such a practice creates strong labor support for privatization and stimulates employees to contribute to an efficient company operation following the ownership change, which were necessary conditions for the success of transition. The proportion of shares offered to employees should therefore be included as a possible explanatory factor in the investigation on IPO pricing.

Another distinguishing feature of the privatization process in Hungary may be the role it played in the creation of the country's stock market. Governments during transition regarded the goal of capital market development as of primary importance, since a liquid stock market may provide funds for corporate investments and attract foreign capital. The substantial number of privatization IPOs (see Table 2) and their large proceeds (see Table 3)[6] compared to the revenues from private issues indicate that privatization indeed contributed to the creation of the stock market in Hungary.

[5] Maximizing proceeds from privatization was an important objective of the governments in transition, since Hungary needed to pay back all of its debt accumulated during the socialist era, for which the financial means came primarily from privatization.

[6] Table 3 is based on our sample. See description of the sample in Section 2.

Table 3

Proceeds from IPO sales, 1990–1998

Year	Privatization	Private	Total
	1000 USD		
1990	36 750	27 786	64 536
1991	8 613	35 346	43 959
1992	44 237	0	44 237
1993	37 837	15 775	53 613
1994	181 408	1 336	182 744
1995	244 399	6 316	250 714
1996	251 324	106 508	357 832
1997	1 136 788	64 869	1 201 657
1998	6 691	106 172	112 863
Total	*1 948 047*	*364 108*	*2 312 155*

We claim that on markets in a premature state, one needs to underprice more than on markets in a more progressed stage of development. This hypothesis is supported by evidence that underpricing is much larger in developing countries than in efficient capital markets (see Table 1).

This phenomenon may imply that IPOs occurred in the early 90-ies in Hungary, soon after the reopening of the stock exchange and thus in a primitive capital market, were priced at a greater discount than later issues.

'Privatization-related' hypotheses

Privatization and the process of transition certainly influenced the development of the IPO market in Hungary. Thus we need to consider further hypotheses that relate underpricing to the former discussed 'transition- and privatization-related factors'.

6. Privatization IPOs are more underpriced than private issues. In our sample, we distinguish privatization offers from private issues by a dummy variable (DPRIV).[7]

7. The proportion of shares sold in exchange for compensation coupons (COMPENS) is positively related to underpricing.

8. Underpricing and equity retained at the initial issue are correlated. When revenue maximization is the main objective, a positive relationship may exist. These variables are however negatively related when the principal consideration of the government is signaling its commitment to a sustained privatization policy.[8]

9. The greater the proportion of strategic investors in the ownership structure, the lower the discount. Our measure is the post-IPO ownership share of strategic investors in the firm (STRATEGIC).

[7] This hypothesis does not unambiguously originate from the transition state of the economy. Greater underpricing at privatization IPOs is also a characteristic of several developed stock markets.

[8] Hypothesis 8 is a modification of the original statement about the relation between underpricing and retained equity (hypothesis 3).

10. The greater the proportion of shares that the firm's employees own following an IPO (EMPLOYEE), the higher is the discount at the issue.

11. The more developed the capital market, the less underpriced the issue. The later the IPO occurred in the sequence of issues at the Budapest Stock Exchange (BSE), the lower is the discount at sale. We identify each issue by a counting variable (DSEQUENCE), which is to be increased by one at each subsequent IPO.

2. EMPIRICAL RESULTS

In this section, first the methods of the investigation are briefly summarized. This is followed by the description of the data used and the remaining part is devoted to a detailed analysis of the estimations.

Methodology

To estimate the relationship between underpricing and the explanatory variables, we apply simple OLS (Ordinary Least Squares) methodology. We justify our choice by several test statistics on residual series, which we include in the tables (see Table 5 and 6) showing our results. The definition of the dependent variable, underpricing, is an important methodological issue in IPO studies. Underpricing is usually defined as the percentage change between the offer price and the first day closing price of the trade. This measure is called raw underpricing.

$$UP_i = \frac{P_{i,1} - P_{i,0}}{P_{i,0}},$$

where

$P_{i,0}$ is the offer price at issue i,

$P_{i,1}$ is the closing price of share i on the first day of trading on the Stock Exchange.

However, actual stock market tendencies may influence prices to a large extent. Therefore, an adjustment is necessary to make in order to see the real value of the discount. As it is customary in the literature, we estimate regressions for adjusted underpricing, which is the difference between raw underpricing and the percentage change in the market index between the date of subscription and the first day of trading.

$$AUP_i = UP_i - \frac{INDEX_{i,1} - INDEX_{i,0}}{INDEX_{i,0}},$$

where $INDEX_{i,0}$ is the closing value of the market index on the last day of subscription of IPO*i* and $INDEX_{i,1}$ is the closing value of the market index on the first day of quotation of the shares issued at IPO*i*.

Description of Data[9]

Our dataset is built on the entire population of IPOs occurred in Hungary before January 1999. 53 share issues are considered, which is somewhat less than the total number of IPOs in this period. Altogether 67 companies offered shares publicly in Hungary after the reopening of the Budapest Stock Exchange in 1990, until the end of 1998. We disregard 14 issues for the following reasons:

– 6 out of the 14 were offerings of shares of investment funds or specialized asset management companies. Considering that portfolio management companies bear different characteristics than companies in the usual production and services industries, their valuation may require a different approach. Therefore, characteristics of the underpricing phenomenon in relation to these companies might also substantially differ.

– Another 8 issues (most of these firms were soon de-listed) are disregarded, because data were not available either on the issue or the firm itself. For most of these firms, the number of stock exchange transactions was insignificant for a period of at least six months following the issue (in a few cases even underpricing could not be calculated).

In conclusion, the dataset includes almost all IPOs in Hungary. Therefore, although the sample size is somewhat small, significant variables identified in regressions must have explanatory power as the sample represents almost the entire population.

Among the 53 initial issues, 29 are privatization IPOs and 24 are private share issues (capital increase or private sale). Adjusted underpricing is 22 percent on average in the entire population. Privatization IPOs are underpriced by 31 percent on average; while for private issues the average discount is only 10.7 percent. Thus the hypothesis that private issues tend to be priced more accurately seems to hold in the IPO market of Hungary.

Table 4

Descriptive statistics on underpricing

Denomination	Privatization IPOs		Private IPOs		Whole sample	
	underpriced	adjusted underpricing	underpriced	adjusted underpricing	underpriced	adjusted underpricing
Sample size	29	29	24	24	53	53
Mean (percent)	30.4	31	6.8	10.7	19.7	21.8
Standard deviation (percent)	34.3	38	25.3	35.3	32.5	37.9
Minimum (percent)	-20.4	-20.4	-52.2	-58.4	-52.2	-58.4
Maximum (percent)	102.4	110	63.5	110.4	102.4	110.4

Taking adjusted underpricing into account, the maximum level of the discount in the population is 110.4 percent. At the same time, 15 initial issues were overpriced (negative underpricing). The lowest value is –58,4 percent. Among the overpriced issues 7 were privatization IPOs and 8 were private sales.

[9] The authors collected data from issued prospectus of IPO firms, annual financial statements of companies, and time series of share prices and the BUX index, all of which were provided by the Information Center of the Budapest Stock Exchange. The Company Fact Book, which contains information on companies floated in Hungary, was also a useful source of information.

Among the 53 Initial Public Offers, 11 implied pure capital increases, while in case of the other 42 issues an (at least partial) sale of already existing shares occurred.

Estimation

Tables 5 and 6 in this section present our empirical results. They show regression estimations for adjusted underpricing as a dependent variable. Each column represents a different regression. Values for the estimated parameters are in rows; *t*-statistics are in brackets. Stars indicate significant explanatory variables.

Table 5

Estimation results

Variables	Model			
	REGR1	REGR2	REGR3	REGR4
Constant	0.09 (0.56)	0.10 (0.70)	0.14 (0.87)	0.22 (1.34)
log(ASSET)				
DPRIV				
COMPENS	0.66*** (4.25)	0.87*** (5.88)	0.60*** (3.56)	0.82*** (4.50)
DEBT	0.26 (1.27)	0.40** (2.21)		
DSEQUENCE	-0.003 (-1.12)	-0.01*** (-3.30)	-0.002 (-0.62)	-0.01*** (-2.72)
DBROKER			0.07 (0.53)	0.17 (1.34)
STRATEGIC	0.02 (0.09)	0.16 (0.89)		
RISK	-0.03 (-0.12)	-0.08 (-0.30)		
RETAINED			-0.04 (-0.24)	-0.05 (-0.31)
EMPLOYEE				
Statistics				
Sample size	53	44	42	34
F-statistics	3.74***	6.77***	3.41***	6.45***
Adjusted R^2	0.21	0.40	0.20	0.40
White heteroscedasticity	1.31	0.62	1.71	0.93
Durbin–Watson statistics	1.81	1.76	2.06	2.37
Residuals' skewness	0.40	0.26	0.53	0.86
Residuals' kurtosis	3.26	3.21	2.85	3.18
Chow forecast test at 44	2.86***		2.64**	

* 10 percent significance level.
** 5 percent significance level.
*** 1 percent significance level.
Note. *t*-statistics in brackets.

The sample size may differ in the regressions presented. Table 5 shows four regressions that constitute two pairs. Both regressions in a pair are based on the same explana tory variables but they are built on different samples. REGR 1 considers the entire sample (53 observations), while REGR 2 concerns share issues exclusively in the period 1990-1996, and thus builds on a smaller sample (of 44 observations). The same holds for the other pair, REGR 3 and 4 in Table 5, and also for the two pairs of regressions in Table 6.

Table 6

Estimation results

Variables	Model			
	REGR5	REGR6	REGR7	REGR8
Constant	0.18	0.28***	0.15	0.28**
	(1.92)	(2.97)	(1.40)	(2.46)
log(ASSET)	0.02	0.02		
	(0.65)	(0.67)		
DPRIV			0.21**	0.21**
			(2.04)	(1.92)
COMPENS	0.58***	0.70***		
	(3.99)	(4.49)		
DEBT				
DSEQUENCE	-0.004	-0.01***	-0.002	-0.01***
	(-1.06)	(-2.81)	(-0.62)	(-2.28)
DBROKER				
STRATEGIC				
RISK				
RETAINED				
EMPLOYEE	-0.10	-0.04		
	(-0.24)	(-0.08)		
Statistics				
Sample size	53	44	53	44
F-statistics	4.38***	6.63***	2.16	3.78**
Adjusted R^2	0.20	0.34	0.04	0.11
White heteroscedasticity	0.69	0.38	1.19	1.37
Durbin–Watson statistics	1.83	1.86	1.48	1.43
Residuals' skewness	0.35	0.26	0.54	0.78
Residuals' kurtosis	3.11	3.10	2.98	3.84
Chow forecast test at 44	2.24**		1.59	

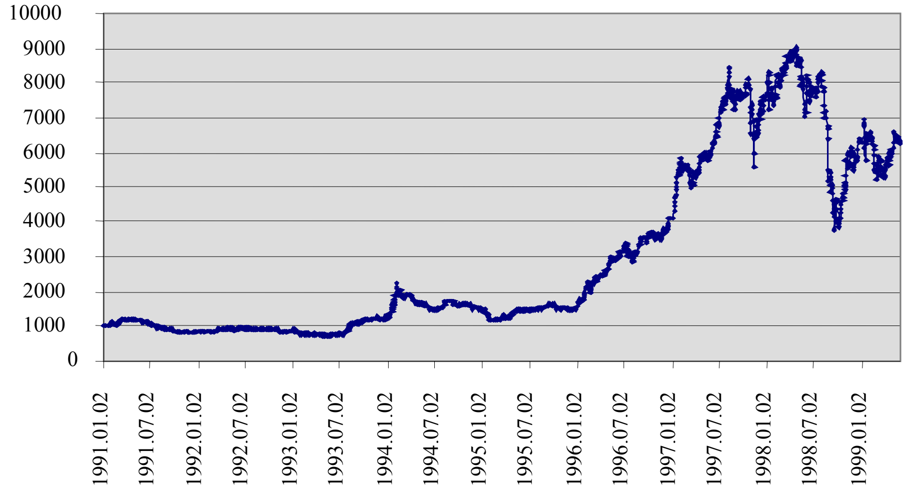
* 10 percent significance level.
** 5 percent significance level.
*** 1 percent significance level.
Note. *t*-statistics in brackets.

The reason for using different samples to estimate underpricing is the following. We claim that in Hungary, the primary characteristics of capital market development differ before and after January 1997, which may influence IPO pricing to a great extent. The early period of transition, 1990–1996, saw a balanced expansion of the stock market, characterized with a progressive upward movement of the market index (BUX) and low volatility. On the contrary, in the more mature phase of market development after 1997, the index exhibited extremely high volatility and market crashes (see Figure 1). On the one hand this is a consequence of international economic and financial market develop ments: the Russian and Asian financial crises seriously affected trading at the Budapest Stock Exchange during 1997–1998. On the other hand, the relatively large number of shares quoted (see Table 1) in this period, high liquidity compared to earlier stages, and the high values of the index indicate a more mature phase of development of the capital market for 1997–1998. We find justification by stability tests that features of IPO pricing indeed differ in the two periods and that the determinants of underpricing can be more easily explained for the earlier, rather smooth phase of market evolution.

The sample size may differ for another reason. When retained equity (RETAINED) is chosen as an explanatory variable, only a small number of observations (42 in the entire period) can be used for estimation purposes (REGR 3 and 4).[10] The reason for this is that equity retention applies only to sales of existing shares. For the group of capital increase IPOs, the relation between equity retention and underpricing cannot be tested.

Concerning the structure of our estimations, we are limited by correlation among the explanatory variables. Offers for compensation coupons occurred only at privatization IPOs, thus we need to estimate the relation of COMPENS and DPRIV to underpricing in separate regressions. Firm size is also correlated with the privatization dummy, since privatization IPO firms are typically of greater size than privately owned issuing companies.

Figure 1. Index values at the Budapest Stock Exchange, 1991–1998



In most of the regressions, explanatory variables have a strong joint effect on adjusted underpricing (significant *F*-statistics). The best explaining variables are COMPENS and DPRIV, which are significantly independent of the sample size. Both have the positive hypothesized sign. This supports our hypotheses that compensation for expropriation under socialism substantially affected IPO pricing and that privatization IPOs were underpriced to a greater extent than private issues in the Hungarian stock market.

[10] In regression 3, the sample size is 42 instead of 53. 11 observations are excluded from the estimation because equity retention is irrelevant (capital increase IPOs).

By far the most significant and highest value coefficient estimate concerns the proportion of shares offered for compensation coupons (COMPENS). This indicates that underpricing occurred particularly in case of those privatization IPOs that allowed for buying shares in exchange for coupons. At sales for coupons underpricing remains hidden in a sense, since its targeted value is implied in the low coupon price instead of a low announced issue price. This is especially true if we take into account (as discussed earlier) that at certain issues coupons could be purchased at the BSE at a price substantially below face value and that they were counted at their nominal value increased with interests at the exchange for shares.[11] This may reinforce our hypothesis on governments in transition pursuing political objectives by the means of underpricing. Especially targeted underpricing may encourage more people to obtain share ownership at IPOs, since it allows for a purchase of shares at an extremely low cash price. Support by a new middle class of citizens might bring substantial political benefits for the privatizing government. Higher discount at privatization IPOs may thus be further evidence on the government's incentives to gain popularity.

The large initial underpricing at privatization deals may be related to the government's consideration for the IPO shares' post-flotation performance. In case of many companies, the SPA applied the strategy of gradual sales on the stock market accompanied with substantial underpricing at the initial public offer. The objective was to achieve a smaller discount and generate large revenues at subsequent share issues. This strategy to privatization is theoretically justified in *Laeven* and *Perotti* (2001): when the company is sold gradually, later sales are priced more accurately reflecting the resolution of political risk.

Governments in transition may have an incentive to sell shares at discount to signal their commitment to a privatization policy without future interference (*Perotti*; 1995). Only a committed government can bear the costs of the initial underpricing of privatization shares. Equity retention can be another means of signaling credibility, thus equity retained by the government at issues should be negatively correlated with the discount (*Perotti–Guney*; 1993). In our estimations the coefficient of retained equity has a negative sign, but underpricing is not significantly affected by this variable.

Greater discount at privatization issues may also refer to the desire of achieving dispersed ownership structures. Governments in transition may aim at creating widely dispersed ownership of public firms in order to bring liquidity to the capital market and thus to give strength to its development. We claim that the substantial underpricing at privatization IPOs in Hungary was to a large extent to stimulate the expansion of the stock market. This is supported by the finding that underpricing was smaller at later issues than at early IPOs, especially in the more premature phase of stock market development (before 1997), as the negative coefficient of DSEQUENCE shows.

Estimations based on the entire population of IPOs in 1990–1998 (REGR 1, 3, 5) suggest that the relationship between underpricing and the independent variables is unstable. This is shown by significant Chow forecast test statistics: in May 1997 (at observation 44), there may be a structural change in the estimated models. Re-estimation of each

[11] Our data on underpricing involve this 'hidden underpricing' as well. See a detailed explanation on calculation of underpricing in the Appendix.

equation for the sub-sample of issues in 1990–1996, (REGR 2, 4, 6, 8) gives better results (higher coefficient values and R^2, higher F-statistics). This change in the explanatory characteristics of underpricing may provide evidence that the features of stock market development differ before and after March 1997. Underpricing can be better explained for the period 1990-1996, when a smooth, progressive expansion and small volatility describes the stock market. When the entire period of 1990–1998 is considered, our models can justify only a smaller part of the variation in the discount.

We underline that the balanced evolution of the stock market in the period 1990–1996 occurred at a low level of market development. From 1997 onwards the volume of trading achieved much higher levels than ever before and volatility also increased to a large extent. Uncertainty implied by a more mature but still inefficient stock market may prevent the establishment of well fitting regression models that could explain the pricing phenomenon. At such a 'low maturity' phase, transition and privatization related explanations might become inappropriate. At the same time, classical asymmetric theory models may not apply yet.

Apart from higher explanatory power and better fit for the period 1990–1996, we find justification for the relation between underpricing and the variable representing the level of stock market development (DSEQUENCE). This implies that early IPOs were priced at a greater discount than late ones, but this holds only for the period 1990–1996. This finding justifies the idea that transition related underpricing disappears as the stock market enters a more progressed phase.

We emphasize that almost all significant explanatory variables (COMPENS, DPRIV, DSEQUENCE) relate to privatization, transition, and the immaturity of the capital market. The finding that privatization IPOs are underpriced to a greater extent than private sales holds for some more matured markets as well. Our estimations reject most of the classical asymmetric information theories on IPO pricing. The coefficient of firm size has a positive sign, which is opposite to our expectations. This may be due to the fact that in our sample offers by large firms relate to privatization of state holdings, which may result in higher underpricing. However, the coefficient is not significantly different from zero. Neither the relation between underpricing and risk, nor the dependence of the discount on the quality of underwriter can be justified. Equity retention has no statistical significance either.

The only asymmetric information theory that is reinforced by our estimations concerns the leverage of issuing firms. Classical corporate finance theory claims that a high level of debt may prevent refinancing projects and therefore implies a negative signal on the firm's future prospects. Indebted IPO firms should therefore use underpricing as a counter-signal to convince investors of their high value. For the period 1990–1996, we indeed find a significant positive relationship between underpricing and debt level at firms issuing shares in Hungary.

Since in the early transition period, bank debt was hardly available and therefore rarely used by Hungarian firms, companies capable to obtain debt financing were better than average. Thus we claim that for the Hungarian sample debt financing and underpricing should be complementary rather than substitutable signals of high value: only good firms, able to raise debt, could recoup the costs of underpricing. This may explain the positive relationship in our estimation (regression 2).

3. CONCLUSION

The most important finding of this paper is that initial public offerings are underpriced in Hungary just as in other countries. The level of this underpricing was approximately 22 percent in the period 1990–1998. This is higher than the average discount in countries with well functioning stock markets but substantially lower than initial returns on several premature capital markets (*Ibboston–Ritter*; 1994).

The underpricing phenomenon in the period considered strongly relates to the transition state of the economy and the low maturity of the capital market. Particularly, privatization of state owned companies, restitution, and thus the privatizing governments' political objectives play a central role in the determination of the discount. Asymmetric information theories, many empirically justified for well-developed markets, do not receive support for IPOs in Hungary.

The most important explanatory factor for the discount concerns compensation. The higher the proportion of shares offered in exchange for compensation coupons, the greater the underpricing at IPO. The low coupon price that characterizes most of the period allows for a relatively high offer price and a low cash price of shares, at the same time. Thus the discount is involved in the coupon price instead of the offer price itself. We call this phenomenon hidden underpricing.

We argue that hidden underpricing in a period of transition may be a policy by the government to pursue objectives such as gaining political support, signaling commitment, and stimulating the capital market. This result reinforces some privatization theories for transition economies, such as *Perotti* and *Guney* (1993), *Perotti* (1995), *Schmidt* (1996), and *Biais* and *Perotti* (2001), and provides further justification for the empirical findings by *Jones et al*. (1999).

Another main result of this paper is that privatization IPOs are underpriced to a significantly greater extent than private sales. The average discount at issues by the State Property Agency (the government institution responsible for privatization) was 31 percent (adjusted for market variation), while private issues were underpriced on average only at 11 percent. This finding is in accordance with empirical results by *Paudyal et al*. (1998) concerning IPOs in an emerging economy, Malaysia.

An interesting finding of our paper is that underpricing can be much better explained by factors advanced by the privatization literature when one considers offers exclusively in the period 1990–1996. Estimations for the entire population of issues in 1990–1998 show less significance and smaller explanatory power. We claim that the reason is that transition related explanations better describe the underpricing phenomenon in the early transition stage, a period of smooth and balanced expansion of the capital market, than in a more mature but still inefficient state of the stock market in the years 1997–1998. At the same time in the immature stage, asymmetric information theories are not yet appropriate. The evidence that IPO underpricing decreases with market development and the fact that this holds only for the early transition period, reinforces our idea that transition-related underpricing slowly disappears as the securities market becomes more mature.

These results suggest that in a transition period, when privatization IPOs take place within a primitive capital market, explanatory variables, which relate to the particular circumstances of the privatization process and the country itself are needed. The main conclusion of this research is the strong significance of these privatization- and country-related explanations.

APPENDIX

DEFINITION OF EXPLANATORY VARIABLES

Underpricing (UP): percentage return of investors if they buy at the offer price and sell at the first day closing price at the initial issue.

Remarks: *1*. When a significant proportion of shares was offered for compensation coupons (CC) and investors could buy coupons at the market price, the offer price is a weighed average of the compensation coupon price and the cash price. The weight is the proportion offered in exchange for compensation coupons.

2. At several IPOs, only certified compensation coupons were accepted in exchange for shares. Certified CCs could be used only by their original owners and were not traded on the Budapest Stock Exchange. However, an unofficial secondary market existed for these certified coupons so that private investors could buy them from their owners at roughly 40 percent premium above the prevailing price at the BSE, and use them to purchase shares at IPOs. Our raw underpricing measure does not reflect this premium in the coupon price and may thus, to a small extent, overestimate the discount at privatization IPOs. Since prices at this secondary market are hypothetical, we use the coupon price as quoted on the BSE in our calculations of the underpricing variable.

3. If the introduction of the shares to the Budapest Stock Exchange occurred more than a month following the issue, we considered the first day closing price deflated to its value at the time of subscription (for certain times annual inflation was above 30 percent during the period under investigation).

Adjusted underpricing (AUP): difference between raw underpricing and the percentage change in the market index between the date of subscription and the first day of trading (see a more detailed explanation in Section 2).

Assets (ASSET): total assets of the IPO firm according to the balance sheet closing the year preceding the issue.

Risk (RISK): variability of excess returns on the share (post-floatation), expressed by the variance of the differences between daily returns on buying and selling the particular share and daily returns on buying and selling the market portfolio.

Remark: As the expected return on a share depends on the performance of the entire market, and such market risk is unavoidable in a premature market, the risk of an investment is better described by the difference between the behaviour of a share and the whole market. We thus calculate the variability of excess returns on shares compared to the market portfolio, represented by the BUX index, a weighed average of price changes of the most often traded securities in the Budapest Stock Exchange.

Retained equity (RETAINED): ratio of the number of shares retained to the total number of shares owned by the issuer.

Reputation of the firm's underwriter (DBROKER): a dummy variable taking the value 1 if the issue is underwritten by a foreign investment bank or brokerage house, or by a syndicate of at least three financial service companies, otherwise its value is 0.

Indebtedness (DEBT): ratio of total debt to total assets as stated in the balance sheet of the year preceding the IPO.

Type of the issue (DPRIV): a quality variable taking a value of 1 if the issue is a privatization IPO, and 0 otherwise.

Proportion of shares offered for coupons (COMPENS): the number of shares offered in exchange for compensation coupons as a percentage of the total number of shares offered for sale at the IPO.

Employee ownership (EMPLOYEE): proportion of shares owned by employees following the IPO.

Ownership of strategic investors (STRATEGIC): the percentage share of the firm's equity strategic investors own following the issue.

Level of stock market development (DSEQUENCE): a counting variable, which has a value 1 for the first IPO in the sequence of issues, and the value of which is increased by 1 at each subsequent issue.

REFERENCES

ALLEN, F. – FAULHABER, G. R. (1989): Signaling by underpricing in the IPO Market. *Journal of Financial Economics,* Vol. 23. No. 1. p. 303–323.

BARON, D. (1982): A model of demand for investments banking advising and distribution services for new issues. *Journal of Finance,* Vol. 37. No. 4. p. 955–976.

BEATTY, R. P. – RITTER, J. R. (1986): Investment banking, reputation and the underpricing of initial public offerings. *Journal of Financial Economics,* Vol. 15. No.1. p. 213–232.

BIAIS, B. – PEROTTI, E. C. (2001): *Machiavellian privatization*. (Manuscript.)

BOOTH, J. R. – CHUA, L. (1996): Ownership dispersion, costly information, and underpricing. *Journal of Financial Economics,* Vol. 41. No. 2. p. 291–310.

BOOTH, J. R. – SMITH, R. L. (1986): Capital raising, underwriting and the certification hypothesis. *Journal of Financial Economics,* Vol. 15. No. 1. p. 261–281.

DEWENTER, K. L. – MALATESTA, P. H. (1997): Public offerings of state-owned and privately-owned enterprises: An international comparison. *Journal of Finance*, Vol. 52. No. 4. p. 1659–1679.

DOWNES, D. – HEINKEL, R. (1982): Signaling and the valuation of unseasoned new issues. *Journal of Finance,* Vol. 37. No. 1. p. 1–10.

GRINBLATT, M. – HWANG, C. Y. (1989): Signaling and the pricing of new issues. *Journal of Finance,* Vol. 44. No. 2. p. 393–420.

IBBOSTON, R. G. – RITTER, J. R. (1994): Initial public offerings. In: *Handbook of operations research and management science: Finance.* North Holland, Amsterdam.

IBBOSTON, R. G. – SINDELAR, J. L. – RITTER, J. R. (1994): The markets' problems with the pricing of initial public offerings. *Journal of Applied Corporate Finance,* Vol. 7. No. 1. p. 66–74.

JENKINSON, T. – MAYER, C. (1988): The privatization process in France and the UK. *European Economic Review,* Vol. 32. No. 1. p. 482-490.

JENSEN, M. – MECKLING, W. (1976): The theory of the firm: Managerial behaviour, agency costs, and capital structure. *Journal of Financial Economics,* Vol. 3. No. 4. p. 305–360.

JONES, S. L. – MEGGINSON, W. L. – NASH, R. C. – NETTER, J. M. (1999): Share issue privatizations as financial means to political and economic ends. *Journal of Financial Economics,* Vol. 53. No.2. p. 217–253.

KEASEY, K. – SHORT, H. (1997): Equity retention and initial public offerings: the influence of signaling and entrenchment effects. *Applied Financial Economics,* Vol. 7. No. 1. p. 75-85.

KEASEY, K. – MCGUINNESS, P. (1991): Prospectus earnings forecasts and the pricing of new issues on the United States market. *Accounting and Business Research,* Vol. 21. No. 88. p. 133–145.

LELAND, H. – PYLE, D. (1977): Information asymmetries, financial structure, and financial intermediation. *Journal of Finance,* Vol. 32. No. 2. p. 371–387.

MENYAH, K. – PAUDYAL, K. N. (1996): Share issue privatizations: the UK experience. In: *Empirical issues in raising equity capital.* Elsevier, Amsterdam.

MUSCARELLA, C. J. – VETSUYPENS, M. R. (1989): A simple tests of Baron's model of IPO underpricing. *Journal of Financial Economics,* Vol. 24 No. 1. p. 125–136.

MYERS, S. C. (1977): Determinants of corporate borrowing. *Journal of Financial Economics,* Vol. 5. No. 2. p. 147–175.

PAUDYAL, K. – SAADOUNI B. – BRISTON, R. J. (1998): Privatization initial public offerings in Malaysia: Initial premium and long-term performance. *Pacific-Basin Finance Journal,* Vol. 6. No. 5. p. 427–451.

PEROTTI, E. (1995): Credible privatization. *American Economic Review,* Vol. 85. No. 4. p. 847–859.

PEROTTI, E. C. – LAEVEN, L. (2001): *Confidence building and emerging stock markets.* Working paper. The William Davidson Institute. Ann Arbor.

PEROTTI, E. C. – GUNEY, S. E. (1993): The structure of privatization plans. *Financial Management*, Vol. 22. No.1. p. 84–98.

RITTER, J. R. (1984): The 'hot' issue market of 1980. *Journal of Business,* Vol. 57. No. 2. p. 215–240.

RITTER, J. R. (1987): The costs of going public. *Journal of Financial Economics,* Vol. 19. No.1. p. 269–282.

ROCK, K. (1985): Why new issues are underpriced? *Journal of Financial Economics,* Vol. 15. No.1. p. 187–212.

ROLAND, G. (1993): The political economy of restructuring and privatization in Eastern Europe. *European Economic Review*, Vol. 37. No. 1. p. 533–554.

SCHMIDT, K. M. (2000): The political economy of mass privatization and the risk of expropriation. *European Economic Review,* Vol. 44. No. 2. p. 393–421.

TITMAN, S. – TRUEMAN, B. (1986): Information quality and the valuation of new issues. *Journal of Accounting and Economics,* Vol. 8. No. 2. p. 159–172.

WELCH, I. (1989): Seasoned offerings, imitation costs, and the underpricing of initial public offerings. *Journal of Finance,* Vol. 44. No. 2. p. 421–449.

INFLUENCE OF FOREIGN CAPITAL ON CORPORATE PERFORMANCE OF THE HUNGARIAN FOOD PROCESSORS

CSABA JANSIK[1]

The ultimate objective of the study is to measure the impacts of foreign capital in the Hungarian food processing. The analysis was designed and carried out at meso-level, which means that the individual food processing industries are in the focus of investigations. Yet, calculations rely on corporate data, which were used to uncover the impact of foreign ownership on company performance.

The analysis sets forth with grouping the Hungarian food processing companies by their ownership structure and comparing the performance of the two major groups of owners, foreign investors and domestic private capital.

The investigation of the dynamic performance gap concludes that foreign owned companies surpass the domestically owned private companies in all the important efficiency and performance categories. The major tendency of performance gap between the two groups of companies has been of opening nature between 1995 and 1998.

KEYWORDS: Foreign investments; Food processing; Corporate performance.

The Hungarian food processing suffered from severe crisis in the beginning of the 1990s. Restructuring and privatisation of companies, and fundamental changes in the operational environment exacerbated and deepened the recession, which culminated in the beginning of the decade. The aggregate performance indicators of food processing have improved since 1993. Although some of the figures stagnated or even slightly declined in 1998, a definite recovery characterised the overall performance of the industry in the second half of the 1990s. Table 1 presents value figures in dollars by using average annual exchange rates.

Foreign investors played and active role in reshaping the ownership structure of food processing. By 1998, they acquired over 60 percent of the aggregate company capital in the industry (see Table 1). This study intends to detect the influence of foreign capital in the Hungarian food processing by searching what impact corporate ownership has made on company performance.

[1] Researcher at MTT Agrifood Research Finland, Economic Research (Maa- ja elintarviketalouden tutkimuskeskus, Taloustutkimus MTTL), Helsinki.

Table 1

Indicators of Hungarian food processing between 1992 and 1998

Indicators	1992	1993	1994	1995	1996	1997	1998
Sales revenues (billion USD)	6.58	6.14	7.27	7.85	7.85	7.59	7.42
Export (billion USD)	1.01	0.83	1.05	1.40	1.52	1.65	1.55
Profitability (percent)	-2.31	0.24	0.73	0.70	1.55	2.77	2.72
Labour productivity (thousand USD/capita)	36.61	40.68	49.25	56.76	59.83	58.65	55.91
Share of foreign ownership in registered capital (percent)	31.82	43.09	47.95	52.66	52.92	60.44	62.62

Source: own calculations based on data of AKII (1997; 1998a).

The opening question of the analysis rises pursuing the previous research objective: is the performance of either foreign owned or domestically owned food processors superior to that of the other group? The hypothetical response states that ownership does have an impact on company performance in the food processing and that foreign affiliates have better performance indicators than their domestically owned rivals.

This hypothesis is built upon the finding that comparative advantages of the foreign investors are soon after investing abroad internalised into their production subsidiaries (*Dunning*; 1997). This results in an overall performance and efficiency growth at the foreign owned companies, which – in the case of considerable foreign participation – would have a measurable impact on the performance of the entire food processing.

Rare examples can be found in the literature that would feature empirical or accounting based impact analyses of multinational or other foreign subsidiaries in the Hungarian food processing. Most studies consider food processing as one of the processing industries in manufacturing comparisons.

Hamar (1995, p. 115) also viewed food processing as one in the set of all processing industries. She attributed the rapid growth of food processing to the participation of foreign investments. Her study also documented a uniform cost structure in the industry.

In one of the most recent studies, *Szabó* (2000) introduced an interesting and novel approach. He traced down the history of 38 large food processors between 1990 and 1998, which were transferred from state control to foreign ownership. The sample included only those companies that had distinct equivalents in both years of observations. This criterion eliminated several multinational enterprises from the investigations. The restricted size of the sample may be the explanation for the surprising fact that the aggregate growth of sales revenues and profit earnings at the sample companies stayed below the industrial average. On the other hand, the growth of equity and total assets in the sample surpassed the industrial average; it indicates quick development and the intensity of investments. Besides, the companies in the sample were also among the best ones in terms of productivity growth, which is a result of their internal rationalisation and intensive labour lay-off (*Szabó*; 2000. p. 47).

The direction in the current article differs from the mentioned two studies in terms of

1. the applied database includes the accounting information – excerpts from balance sheets and income statements – of the complete set of Hungarian food processing companies that is over two and half thousand enterprises between 1995 and 1998;

2. the analytical approach, namely the dynamic measurement of performance difference between foreign and domestically owned companies that gives novelty to the current analysis.

1. IMPACT OF OWNERSHIP ON CORPORATE PERFORMANCE FIGURES

In the first part of the calculations, performance of the companies is measured by sales revenues, profit earnings, export sales, and investment activities. Sales revenues is an essential indicator of corporate performance, since it embodies market power. In an earlier analysis, market power was proved to significantly motivate the inflows and industry choices of foreign capital in the food sector (*Jansik*; 2000a, p. 83).

1.1. Indicators of corporate performance and data set

In the post-socialist economic environment, profit earnings have provided rather distorted information on the performance of food processing companies. In the second half of the decade, partly owing to the halved corporate tax rates, the profit performance of the companies improved.[2]

Although experience indicates that food industrial foreign investors settle themselves primarily to supply domestic markets, the traditional export performance also strengthened in the Hungarian food processing sector. The amount of investments is an indicator that signifies the future objectives and anticipations of the companies.

The impact of ownership on corporate performance is first identified by segmenting the companies. The database used in the study embraces all food processing companies for four years from 1995 to 1998. The most recent data were used in the initial calculations; total number of observations in 1998 amounted to 2977, which narrowed to 2961 after eliminating the companies with zero registered capital. Two major groups were identified according to the ownership structure of the companies: *1*. majority-foreign owned and *2*. majority-domestically owned companies.

The segmentation was done on the basis of foreign versus domestic ownership share in the registered company capital of each enterprise.

1.2. Results of the company segmentation

Table 2 demonstrates the corporate figures of the companies in the two groups. Beside aggregate figures, it also displays group averages. The data in the table confirms the ‘size-superiority’ of foreign owned companies. The difference in sales revenues refers to the fact that foreign owned companies are typically among the largest ones in food processing.

In terms of average profit and export sales, the superiority originating from foreign ownership seems to be even more pressing than in the case of sales revenues, while average investments of foreign owned companies are nine times more than average investments of the domestic companies.

[2] Nevertheless, many domestically owned companies are presumed to ‘hide’ their profit, while foreign owned companies may tend to exercise hidden profit repatriation. Yet, the only available profit figures are official reported ones, which were also used in the calculations.

Table 2

Aggregate figures of Hungarian food processing companies by majority owners in 1998

Ownership	Net sales revenues	Profit	Export sales	Value of investments
	billion HUF			
Total majority foreign (*N*=398)	764.96	45.14	188.62	14.35
Total majority domestic (*N*=2563)	826.08	5.78	143.18	10.17
Average of foreign owned	1.9220	0.1134	0.4739	0.0361
Average of domestically owned	0.3223	0.0023	0.0559	0.0040

The mere data of Table 2 may well raise a doubt: since all large companies are in foreign ownership, the real reason for different performance is not the ownership type, but rather the magnitude of companies. A two-step segmentation was to address this apprehension. The initial grouping attribute is the size of the companies by using their sales revenues: the first group includes all companies, whose revenues exceeded HUF 100 million, the second group included all companies below that limit. The segmentation follows the previously applied method afterwards. The objective of double segmentation is to compare the foreign and domestically owned companies within their own size groups. Table 3 shows the results of the analysis.

Table 3

Corporate figures of Hungarian food processors by size and ownership type in 1998

Size and ownership	Sales revenues	Profit	Export sales	Value of investments
	million HUF			
Sales revenues above HUF 100 million				
Total majority foreign (*N*=167)	760 953.8	43 886.9	187 560.2	13 720.7
Total majority domestic (*N*=718)	788 898.2	7 552.1	141 789.1	8 439.6
Average of foreign owned	4 556.6	262.8	1 123.1	82.2
Average of domestically owned	1 098.7	10.5	197.5	11.8
Sales revenues below HUF 100 million				
Total majority foreign (*N*=231)	4 005.6	1 250.0	1 062.9	632.8
Total majority domestic (*N*=1845)	37 182.1	-1 770.5	1 393.1	1 734.9
Average of foreign owned	17.3	5.4	4.6	2.7
Average of domestically owned	20.2	-1.0	0.8	0.9

A little more than 40 percent of the foreign owned companies are situated in the group of the large companies, their power continues to be evident (see Table 3). Although they make up less than one-fourth of the number of the large companies, their combined sales revenues still amount to nearly half of the entire large-scale group. Consequently, the difference in average sales revenues between the two ownership types in the group of large companies is more than fourfold. The difference in terms of average profit is even considerably wider than in the case of sales revenues. It indicates that foreign owned companies are much more profitable than domestically owned ones. The differences between group averages in the field of profit, exports and investments surpass the magnitude of difference in sales revenues.

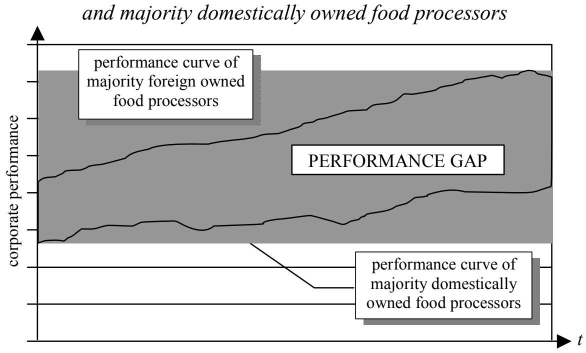
The group of the small companies results in some unexpected findings as opposed to the group of large companies. The average sales revenues of both ownership groups are around the same magnitude; in fact, the figure of domestically owned companies is even slightly higher than that of the foreign owned group. Due to the similar size among the small companies, any difference in corporate performance can purely be attributed to the differences in ownership structure. Further data inform about the superiority of foreign ownership: foreign subsidiaries earned considerable profits compared to the losses of domestically owned companies, their exports are nearly sixfold and the value of investments are threefold in comparison to the respective average figures of domestically owned processors.

Investment is an extremely important performance category in terms of future corporate growth. Investments of the foreign owned companies surpass the respective figures of domestically owned enterprises in both company size groups. The difference is not that surprising in the group of large companies, where capital strength of the large multinational enterprises is overwhelming. In the other group, small and medium-scale ventures could be the ones that are mobile, flexible and perceptive to emulate the modern techniques of foreign owned companies. They are believed to be the germ of a strong Hungarian owned food processing segment and they are expected to constitute a competitive group against foreign subsidiaries. Since these small Hungarian enterprises can not easily compete with the large foreign ones in terms of innovation, product differentiation or know-how, the only relevant way of corporate growth appears to be physical investments. Therefore, it is regrettable to recognise that Hungarian owned small-scale processors take such a low-key approach in investments.

2. DEFINITION OF THE CORPORATE PERFORMANCE GAP

A difference in corporate performance has been verified to prevail between the foreign and domestically owned food processors in Hungary; the corporate operational data in the previous analysis confirm the existence of the gap. Therefore, the analysis can be continued one step further and incorporate the dynamic aspect. The opening question of the study now takes a modified form as follows: is the performance gap between foreign and domestically owned food processors widening or narrowing over time?

Figure 1. The concept of performance curves of majority foreign and majority domestically owned food processors



The dynamic concept of the performance gap is illustrated in Figure 1, where shaded area signifies the gap. The widening or narrowing nature of the gap can be revealed only through a care study of dynamic indicators. Such calculations require comparable data for many years. Data availability limited the surveyed period to four years, from 1995 to 1998. Again, enterprises with zero registered company capital were eliminated from the data set.

3. COMPARISON OF FOREIGN AND DOMESTICALLY OWNED COMPANIES

In order to quantify the changes of the performance gap, a wider group of indicators of corporate performance is computed.

3.1. Relative performance indicators used in the comparison

Calculations contain four profitability, one productivity and three other accounting indicators:[3]

Profit rate (profitability margin):[4] $ROS_i = \frac{P(bt)_i}{R_i}$ /1/

Return on equity:[5] $ROE_i = \frac{P(bt)_i}{E_i}$ /2/

Profit to assets: $p_i^a = \frac{P(bt)_i}{TA_i}$ /3/

Profit per capita: $p_i^{emp} = \frac{P(bt)_i}{EMP_i}$ /4/

Labour productivity:[6] $prod_i^{emp} = \frac{R_i}{EMP_i}$ /5/

Own capital intensity: $e_i = \frac{E_i}{TA_i}$ /6/

Share of export sales: $exp_i = \frac{EXP_i}{R_i}$ /7/

Asset efficiency: $r_i^a = \frac{R_i}{TA_i}$ /8/

[3] The group of indicators were sorted out based on the traditions of the corresponding international literature; see *Jansik* (2000b, p. 248–258.) for a detailed literature overview.

[4] The indicator is identical to the category of Return of Sales – ROS.

[5] The measure is not fully identical to Return on Equity – ROE indicator, since the figure of so called ‘own capital’ is used in the denominator as a substitute for equity.

[6] Labour productivity is usually calculated with value added produced by the company. However, when value added figures are unavailable, net sales/capita is an internationally applied proxy. The indicator does not directly denote labour productivity, but it expresses corresponding trends of the particular corporate performance aspect *(Frydman et al.*; 1999).

where

$P(bt)_i$ is the i^{th} company's profit before taxation,
R_i is the i^{th} company's sales revenues,
EXP_i is the i^{th} company's export sales,
E_i is the i^{th} company's own capital,
TA_i is the i^{th} company's total assets,
EMP_i is the i^{th} company's labour force.

3.2. Results of the dynamic performance gap analysis

Pursuant to the definition explained in Figure 1, data in Table 4 show a snapshot of the dynamic performance gap for *t*=1998. The numbers prove a significant disparity between the performance of foreign and domestically owned food processors. Indicators signify a notable advantage of foreign owned companies: their labour productivity was twice, while their profit rate was seven times as high as the respective figures of domestically owned enterprises.

The advantage of domestic ownership was disclosed in the case of one indicator, asset efficiency. This indicator, however, carries rather illusive than real competitive advantages. A logical explanation lies behind the domestic superiority in the case of asset efficiency. Many domestically owned enterprises operate with almost fully or fully depreciated assets. Hence, the lower value of denominator results in higher measures of asset efficiency than in the case of foreign owned companies with their typically more valuable or recently installed assets. The indicator of capital intensity already gives a better understanding on real power relations.

Table 4

Relative corporate indicators of majority foreign owned and majority domestically owned food processors, 1998

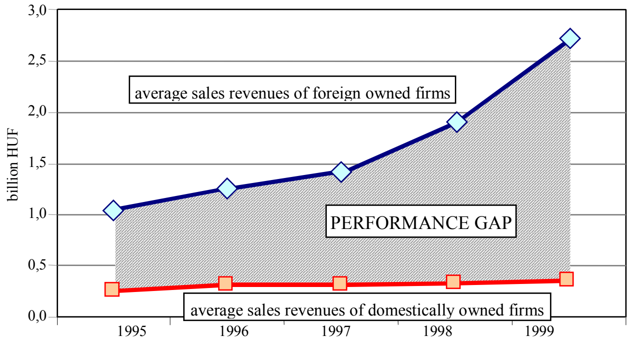
Performance indicator	Majority foreign owned (*N*=398)	Majority domestically owned (*N*=2563)	Total food processing (*N*=2961)
1. Profit rate – ROS (percent)	4.94	0.70	2.74
2. Return on equity – ROE (percent)	12.56	2.8	8.59
3. Profit to assets (percent)	6.08	1.12	3.83
4. Profit per capita (thousand HUF /capita)	868.4	64.8	328.5
5. Labour productivity (million HUF /capita)	17,569	9,264	11,989
6. Own capital intensity (percent)	48.4	40.1	44.6
7. Share of export sales (percent)	24.65	17.33	20.85
8. Sales to assets (percent)	122.95	160.27	139.86

Source: the author's calculations based on the data of AKII.

The dynamic approach of the performance gap is evidently more essential than a static type of comparison. Data in Table 4 get full sense by placing them into the dynamic context of a longer period.

In order to illustrate the development of the gap in practice, Figure 2 shows the curves of average sales revenues for foreign and domestically owned food processors in Hungary. The figure confirms that the difference in average sales revenues between the two ownership groups has been growing. Although such a chart of performance curves is illustrative, it has to be admitted that the mere comparison of annual values may imply the danger of inaccurate calculation of the gap.

Figure 2. Average sales revenues of foreign and domestically owned food processors in Hungary between 1995 and 1999



Since indicators of both groups of companies change over time, the comparison of absolute values would really not capture the opening or closing nature of the gap.[7]

In order to resolve the issue, a similar methodology is applied that *Pilat* (1996) developed at OECD to measure international productivity gaps. The performance of the most productive country was fixed at unity (or at 100 percent), and the productivity of other countries was expressed as its proportion. Productivity gap is defined by the difference between the most productive country and the other ones. This approach allows a reliable and accurate dynamic interpretation, even if the denominator changes in the meantime, or the title of the most productive country is taken over by a new one.

Data in Table 5 were calculated by applying a similar concept to the productivity gaps suggested by *Pilat*. In order to quantify the performance gap among the Hungarian food processors, the figure of the better group was fixed at unity. The denominating base of the proportion was the performance of the foreign owned companies in the majority of cases. The following formula expresses the performance gap (PGAP) for a particular indicator:

$$PGAP = 1 - \frac{\sum_{i=1}^{n} P_i^w}{\sum_{j=1}^{m} P_j^b},$$

[7] A simple arithmetic example enlightens the problem. Let us suppose two economic actors – enterprises or nations – whose performance figures are compared for three subsequent years. Let the values of the better performer be 10, 12 and 15, while the values of the weaker performer 2, 3, and 5, respectively. Then, the absolute difference between them will be 8, 9 and 10, which would suggest a widening gap. The performance gap given by the PGAP formula, however, will be 0.8, 0.75 and 0.66, or in percentage form 80, 75 and 66 percent. The values calculated on the basis of the PGAP formula confirms exactly the opposite trend that absolute figures suggested, namely a narrowing performance gap.

where:

P_i^w is the indicator of the i^{th} company in the ownership group with weaker performance,

P_j^b is the indicator of the j^{th} company in the ownership group with better performance.

The formula of PGAP determines that the figures of performance gap can take values between zero and unity. Multiplying the values by 100, PGAP may also be expressed in percentage form. Then, the arbitrary value of a PGAP=0.72 can be interpreted in two different ways: the performance gap is 72 percent of the performance of the better group, or the performance figure of the weaker group reaches 28 percent of the performance of the better group. This latter one is an intermediate step in the calculation process. Due to its demonstrative power, however, it is also included among the results in Table 5. P^W is expressed by the following two formulas:

$$PGAP = 1 - P^w \quad \text{and} \quad P^w = \frac{\sum_{i=1}^{n} P_i^w}{\sum_{j=1}^{m} P_j^b}$$

The values of PGAP and P^W should be interpreted based on how they change over time: descending values of P^W or ascending values of PGAP would indicate the widening of the performance gap and vice versa, ascending P^W and descending PGAP values would signify the narrowing of the gap.

Although access to data limited the time span of the analysis to four years, the amount of computed indicators is sufficient to draw pertinent conclusions. The changes in the performance gap demonstrated by eight corporate indicators disclose a dramatic shift in the Hungarian food processing sector. All the profitability, productivity and export indicators notify a distinctly widening performance gap: domestically owned companies constantly keep falling behind foreign owned processors. In terms of capital intensity, the lag has been stagnating since 1996, the only advantage of the domestically owned companies is a stable superiority shown in the case of asset efficiency.

It would be early to alert on the basis of four year development of the performance gap, although future prospects do not promise spectacular improvement for the domestically owned processors. The figures in Table 5 reveal a notable disparity among investment activities between the two ownership groups, which may continue to open the performance gap also in the coming years. This danger equally shades the future of both small and large domestically owned firms in the Hungarian food processing.

Results shown in Table 5 should be interpreted with caution also due to another fact. Since the performance gap is calculated from group averages, the numbers conceal the heterogeneity in the performance of group members. In the multitude of food processors, there might be laggards amongst the foreign owned companies just as well as rapidly developing domestically owned enterprises.

Table 5

Performance gap between foreign and domestically owned food processors between 1995 and 1998

Indicators/Elements of performance gap	Formulas	1995		1996		1997		1998	
		Foreign	Domestic	Foreign	Domestic	Foreign	Domestic	Foreign	Domestic
		majority owned companies							
Number of firms	n, m	399	2 153	397	2 289	400	2 468	398	2 563
Sales revenues/ enterprise (billion HUF)	$\frac{\sum_{i=1}^{n} R_i}{n}, \frac{\sum_{j=1}^{m} R_j}{m}$	1.034	0.262	1.246	0.307	1.544	0.324	1.917	0.322
SL*	P^W		0.253		0.246		0.210		0.168
PG**	***PGAP***	***0.747***		***0.754***		***0.790***		***0.832***	
1. Profit rate (Profit to sales, percent)	$\sum_{i=1}^{n} ROS_i, \sum_{j=1}^{m} ROS_j$	0.07	1.06	2.23	1.07	5.03	1.04	4.94	0.7
SL*	P^W	0.065			0.478		0.206		0.142
PG**	***PGAP***	***(0.935)***		***0.522***		***0.794***		***0.858***	
2. Return on equity (percent)	$\sum_{i=1}^{n} ROE_i, \sum_{j=1}^{m} ROE_j$	0.16	3.56	6.32	4.35	11.68	4.26	12.56	2.80
SL*	P^W	0.045			0.689		0.364		0.223
PG**	***PGAP***	***(0.955)***		***0.311***		***0.636***		***0.777***	
3. Profit to assets (percent)	$\sum_{i=1}^{n} p_i^a, \sum_{j=1}^{m} p_j^a$	0.08	1.60	2.94	1.66	6.22	1.68	6.08	1.12
SL*	P^W	0.048			0.562		0.271		0.185
PG**	***PGAP***	***(0.952)***		***0.438***		***0.729***		***0.815***	
4. Profit per capita (in 1000 HUF/capita)	$\sum_{i=1}^{n} p_i^{emp}, \sum_{j=1}^{m} p_j^{emp}$	6.5	63.9	277.8	83.0	777.2	92.8	868.4	64.8
SL*	P^W	0.102			0.299		0.119		0.075
PG**	***PGAP***	***(0.898)***		***0.701***		***0.881***		***0.925***	
5. Labour productivity (in million HUF/capita)	$\sum_{i=1}^{n} prod_i^{emp}, \sum_{j=1}^{m} prod_j^{emp}$	9 431.6	6 021.5	12 434.1	7 772.1	15 441.8	8 949.5	17 569.5	9 264.5
SL*	P^W		0.638		0.625		0.580		0.527
PG**	***PGAP***	***0.362***		***0.375***		***0.420***		***0.473***	
6. Own capital intensity (percent)	$\sum_{i=1}^{n} e_i, \sum_{j=1}^{m} e_j$	47.34	44.98	46.62	38.07	53.26	39.55	48.39	40.07
SL*	P^W		0.950		0.816		0.743		0.828
PG**	***PGAP***	***0.050***		***0.184***		***0.257***		***0.172***	
7. Share of export sales (percent)	$\sum_{i=1}^{n} exp_i, \sum_{j=1}^{m} exp_j$	18.28	17.12	22.56	17.35	25.36	18.92	24.66	17.33
SL*	P^W		0.936		0.769		0.746		0.703
PG**	***PGAP***	***0.064***		***0.231***		***0.254***		***0.297***	

(Continued on the next page.)

(Continuation.)

Indicators/Elements of performance gap	Formulas	1995		1996		1997		1998	
		Foreign	Domestic	Foreign	Domestic	Foreign	Domestic	Foreign	Domestic
		majority owned companies							
8. Asset efficiency (percent)	$\sum_{i=1}^{n} r_i^a,\ \sum_{j=1}^{m} r_j^a$	110.5	150.9	131.8	155.0	123.6	162.2	123.0	160.3
SL*	P^W	0.733		0.850		0.762		0.767	
PG**	***PGAP***	***(0.267)***		***(0.150)***		***(0.238)***		***(0.233)***	

* SL – Standardised level of weaker group.
** PG – Performance gap.
Note: the values of PGAP in parenthesis indicate the superiority of domestically owned group.

Domestically owned food processors face three main options as far as their future is concerned:

1. Catching up; in other words it is the alternative of survival. It is a function of initial level of development but to a certain extent it also depends on industrial affiliation. Those domestic companies may enter the group of up-comers that will narrow the performance gap with the application of state-of-the-art technology and modern business techniques. Capital for the development will hardly be available from own resources, it will need to be mobilised in the domestic or international capital markets. Domestically owned companies may develop competitive advantages by two ways: one is product differentiation, and meeting uncommon or specific consumer needs; the second way is special market strategy in a geographic sense, exploring the white spots of Hungarian market, or specialising in export markets.

2. Falling behind. Companies utilising outdated equipment and management methods, and restrained marketing strategies will inevitably be forced on the way of dropping back. These companies would have typically similar product mix to that of their large foreign competitors; they lack resources for development and access to financing channels.

3. Middle alternative. The middle alternative would involve co-operation or collusion with large competitors. It is possible only for some companies with special status in given industries. Large competitors may need co-operating companies because of market distribution or other regional reasons. Middle alternative has a very limited applicability. Its outcome is risky, since company acquisitions may soon cease the independence of the smaller partner in any strategic alliance.

4. INDUSTRIAL IMPACTS OF OWNERSHIP STRUCTURE

The impact analysis of foreign ownership ends with unveiling industry-specific tendencies. The foreign ownership share in the registered company capital represents ownership structure (OSTR). The value of variable OSTR ranges from 0 to 1, this way it expresses the division between foreign and domestic ownership. The impact of ownership structure is estimated in the case of six corporate performance indicators with descriptive regression analysis:

$$PER_i^j = a_0 + a_1 OSTR_i + \varepsilon ,$$

where

PER_i^j is the j^{th} corporate performance indicator in the i^{th} industry,

$OSTR_i$ is the variable of corporate ownership structure in the i^{th} industry.

The following six corporate performance indicators were used to measure industry-specific impacts: the categories of sales revenues, profit, export sales, investments, labour productivity and asset efficiency. Estimations were run in all 12 industries, where the number of observations allowed such calculations. Owing to the nature of OSTR variable, the positive sign of estimated parameters indicate a larger impact – or in other words the advantage – of majority foreign owned companies, while negative values signify a bigger effect – or advantage – of majority domestically owned companies in the case of the specific indicator.

Based on the calculations so far in the article, the superiority of foreign owned companies is anticipated in the first five performance categories, while the advantage of domestically owned companies is more probable in the field of asset efficiency. Table 6 shows the sign of estimated parameters of OSTR variable in the cases of various indicators and industries. The findings do verify the anticipations, the superiority of foreign owned companies can be declared overwhelming with the exception of asset efficiency.

Table 6

The impact of ownership structure on selected corporate indicators in the food processing sub-sectors

Performance indicators \ Industries	Meat	Poultry	Fruit and vegetable	Vegetable oil	Dairy	Milling	Feed	Bakery	Confectionery	Distilling	Beer	Soft drinks
Sales revenues	+	+	+	+	+	+	+	+	+	+	+	+
Profit	+	+	+	+	-	+	+	+	+	+	+	+
Export	+	+	+	+	+	+	+	+	+	+	+	+
Value of investments	+	+	+	+	+	-	+	+	+	+	+	+
Labour productivity	+	-	+	+	+	+	-	-	+	+	+	+
Asset efficiency	-	-	-	-	-	-	-	-	-	-	-	-

The detailed explication of industry-specific disparities in the corporate indicators is based on the results of regression analysis in the Appendix.[8] The influence of foreign

[8] Results also reveal the fact that the superiority in the case of several industries and indicators is statistically not significant. In certain instances, low R^2 are attributable to the fact that the relations measure the impact of one single variable, the ownership structure. The current research did not strive for the expansion of the model. The purpose of the estimates was to map the influence of the two major ownership categories.

ownership on *sales revenues* is significant in all industries with the exception of milling and baking, where foreign ownership is initially low.

The *profit* category shows a less pronounced impact of the foreign ownership, real significant influence can only be detected in the vegetable oil and feed industry as well as in the manufacturing of beverages. Results again verify the fact that foreign companies soon adapted to the Hungarian conditions and they do not necessarily display the full amount of profits. Foreign ownership determines the development of profits in a positive way, the only exception is dairy industry. Most probably industry concentration and fierce competition for market shares deteriorated the profits of foreign owned companies; this is why the fact of domestic ownership itself may result in higher profits than foreign ownership.

Export sales inform about the unequivocal superiority of foreign owned companies. The fact of foreign ownership presumes more intense exporting activities in every industry, most significantly in the poultry, vegetable oil, dairy, grain processing, confectionery and brewing industries.

The result in the case of *value of investments* also refers to the strong activity of foreign owned enterprises. Among the individual industries, the investments of foreign owned companies are worth highlighting in the vegetable oil and feed industries as well as in the manufacturing of beverages.

The explanatory power of ownership structure in the field of *labour productivity* and asset efficiency is scarcely significant. Foreign owned companies tend to be more productive than their Hungarian rivals in the entire food processing, but principally in the dairy and beer industries. Exceptions are the poultry-, feed-, and bakery industries, where domestic ownership is the labour productivity increasing factor.

The influence on asset efficiency is mostly insignificant, although the sign of the parameters indicate the positive impact of Hungarian ownership in every industry. The result is not at all surprising, what is more, it is in compliance with the respective calculations of earlier studies about the entire set of Hungarian enterprises. *Major* (1996) reported the advantage of domestic ownership for the return on fixed assets. *Szanyi* (1998) explained the phenomenon with the age and composition of assets, which differ considerably at the foreign- and domestically owned companies. The results of the current calculations lead to the conclusions that the same patterns prevail also in food processing; the aggregate data of asset efficiency proved the tendency for the entire food processing industry (see Table 5).

The regression analysis in the individual food processing industries verified that foreign ownership has an influence on corporate performance. This impact is insignificant in the majority of the cases, but owing to the nature of the OSTR variable, the positive or negative sign of its estimated parameter means the better performance of either one of the ownership types. The results of calculations unveiled the advantages of foreign ownership in the case of most industries. The superiority of foreign owned companies is especially strong in the vegetable oil, dairy and beverages manufacturing sub-sectors. Hungarian owned companies appear to be competitive in the grain and poultry processing sub-sectors.

5. CONCLUSION

The objective of the article was to analyse the impacts of foreign investments in the Hungarian food processing. Subsequent to corporate restructuring and privatisation in the food industry, a pronounced disparity emerged between the foreign owned and domestically owned company groups in the second half of the 1990s. The majority-foreign owned enterprises enjoy unequivocal superiority over the majority-domestically owned ones in most indicators of corporate operation and performance.

The dynamic analysis of performance gap enlightened the opening tendency of corporate performance gap. Between 1995 and 1998 the Hungarian owned food processors could not reduce their lag. The dominance of foreign owned companies prevails also on the level of individual industries with the exception of poultry industry and the entire grain processing chain.

The superior corporate figures of foreign owned companies translate into constantly improving performance of Hungarian food processing. The calculations suggest a definite overall positive impact of foreign capital in the industry.

Hungarian ownership appears to play an overwhelming effect on asset efficiency, although it is a sign of worn out fixed assets. The phenomenon coupled with low investment activities casts an ominous shade on the future of Hungarian owned food processors. Nevertheless, catching up is in principle still a relevant option in the present conditions and framework. After Hungary's potential accession to the EU, the current opportunities will most probably diminish.

It is the responsibility of domestically owned food processors, how much they utilise and take advantage of the spillover effects of FDI. Economic policy may alleviate the size-based disadvantages of domestically owned enterprises with dedicated supportive policy to small- and medium-scale enterprises and permanent requirement and incentives for efficiency. The banking sector and capital markets can provide a solid ground for more investments.

These can contribute to the successful development of domestically owned processors. In order to catch up to the high-performers, they have to be committed to state-of-the-art technology, perceptive to modern management techniques and risk-takers to implement new investments. Developments should be made on the markets of particular product groups, which are driven by stable or growing consumer demand, and where there is reasonable room for expanding processing capacities.

APPENDIX

The impact of ownership structure on corporate performance in the Hungarian food processing

Variables	Meat	Poultry	Fruit and vegetable	Vegetable oil	Dairy	Milling	Feed	Bakery	Confectionery	Distilling	Beer	Soft drinks
	Sales revenues											
Constant (a_0)	0.6947	1.4713	0.4854	0.0018	1.3360	0.7192	0.6811	0.1132	0.0794	0.1496	-0.0110	0.0340
OSTR (a_1)	2.8660	11.843	1.2173	7.8600	3.7918	0.1560	2.9530	0.3281	2.8666	2.8291	7.0963	4.5054
R^2	0.063	0.127	0.083	0.131	0.132	0.001	0.126	0.032	0.139	0.285	0.414	0.237

(Continued on the next page.)

(Continuation.)

Variables	Meat	Poultry	Fruit and vegetable	Vegetable oil	Dairy	Milling	Feed	Bakery	Confectionery	Distilling	Beer	Soft drinks
						Profit						
Constant (a_0)	0.1939	1.2609	0.0045	-0.0061	1.1464	0.0394	0.0238	0.0017	-0.0519	0.0126	-0.0831	-0.1888
OSTR (a_1)	5.1056	5.2609	0.0407	0.9730	-1.2735	0.5737	0.3943	0.0215	4.6303	1.1273	8.8279	6.6215
R^2	0.044	0.015	0.007	0.121	0.005	0.001	0.196	0.018	0.059	0.414	0.128	0.102
						Export						
Constant (a_0)	0.1502	0.4684	0.1948	-0.0136	0.0702	0.0394	0.0212	0.0004	-0.0037	0.0234	-0.0029	0.0277
OSTR (a_1)	1.1358	5.3645	0.7117	2.7982	0.5758	0.5737	0.5433	0.0723	0.8066	0.1842	0.5623	0.4044
R^2	0.057	0.133	0.079	0.126	0.135	0.104	0.178	0.053	0.149	0.045	0.267	0.099
						Value of investments						
Constant (a_0)	0.0052	0.0116	0.0098	0.0017	0.0125	0.0057	0.0008	0.0009	0.0036	-0.0012	0.0003	-0.0031
OSTR (a_1)	0.0242	0.0717	0.0135	0.0204	0.0779	-0.0046	0.1260	0.0205	0.0096	0.3082	0.0953	0.2169
R^2	0.047	0.073	0.015	0.166	0.072	0.003	0.351	0.030	0.033	0.390	0.207	0.124
						Labour productivity						
Constant (a_0)	14.970	16.637	11.486	10.140	13.341	11.081	25.297	5.395	4.967	7.578	2.282	6.650
OSTR (a_1)	4.327	-5.547	1.1293	14.948	20.127	58.095	-2.025	-712	6.189	26.156	26.783	7.006
R^2	0.002	0.003	0.030	0.035	0.119	0.098	0.000	0.000	0.064	0.037	0.192	0.073
						Asset efficiency						
Constant (a_0)	5.1	4.8	1.9	2.4	3.4	2.1	3.4	3.3	2.5	1.3	1.5	2.5
OSTR (a_1)	-3.8	-4.2	-0.3	-1.1	-1.4	-0.9	-2.2	-0.5	-0.8	-0.7	-1.0	-1.4
R^2	0.054	0.019	0.003	0.008	0.062	0.024	0.038	0.001	0.012	0.007	0.018	0.053

Note. Estimated parameters and coefficients of determination.

REFERENCES

AGARWAL, J. (1980): Determinants of foreign direct investment: A survey. *Weltwirtschaftliches Archiv*, Vol. 116. p. 739–773.

AKII, Agrárgazdasági Kutató és Informatikai Intézet (1995): *Az élelmiszereket, italokat és dohánytermékeket gyártó szervezetek gazdálkodásának főbb adatai 1984–1993*. Budapest.

AKII (1997): *A mezőgazdasági és élelmiszeripari szervezetek gazdálkodásának főbb adatai 1992–1996*. Budapest.

AKII (1998a): *A mezőgazdasági és élelmiszeripari szervezetek gazdálkodásának főbb adatai 1993–1997*. Budapest.

AKII (1998b): *Élelmiszeripari szervezetek név- és címjegyzéke 1997*. Budapest.

AKII (1999): *Élelmiszeripari szervezetek név- és címjegyzéke 1999. május 31-én*. Budapest.

BÉHM, I. (1998): *Vállalkozások megítélése*. Perfekt Kiadó, Budapest.

BEVAN, A. – ESTRIN, S. – SHAFFER, M. (1999): *Determinants of enterprise performance during transition*. Working Paper, Centre for Economic Reform and Transformation (CERT), Edinburgh.

BOARDMAN, A. – VINING, A. (1989): Ownership and Performance in Competitive Environments. A Comparison of the Performance of Private, Mixed, and State-owned Enterprises. *Journal of Law and Economics*, Vol. 32. p. 1–33.

DUNNING, J. (1997): *Alliance capitalism and global business.* Routledge, New York.

ÉLTETŐ, A. – SASS, M. (1997): Motivations of foreign investors – FDI and foreign trade, The Hungarian case. In: WITKOWSKA, J. – WYSOKIŃSKA, Z. (eds): *Motivations of foreign direct investors and their propensity to eports in the context of European integration process – Empirical studies with special reference to Eastern and Central European countries.* University of Łódź, p. 77–105.

FRYDMAN, R. – GRAY, C. – HESSEL, M. – RAPACZYNSKI, A. (1999): When does privatization work? The impact of private ownership on corporate performance in the transition economies. *The Quarterly Journal of Economics*, No. 4. p. 1153–1191.

HAMAR, J. (1995): *A külföldi működő tőkebeáramlás hatásai a magyar gazdaság átalakulására (1989–1995) – Mérhető makro és mikrogazdasági hatások a nemzetközi elméleti és gyakorlati tapasztalatok, valamint a hazai tényleges folyamatok elemzése alapján*. PhD. Thesis, Budapest.

JANSIK, C. (2000a): Foreign direct investment in the Hungarian food sector. *Hungarian Statistical Review*, Vol. 78. Special Number 4. p. 78–104.

JANSIK, C. (2000b): *A magyar élelmiszeriparba áramló külföldi működő tőkebefektetések determinánsa és hatása közép- és kelet-európai kontextusban*. PhD Thesis, Budapest University of Economics and Public Administration.

LIZONDO, S. (1993): Foreign direct investment. In: ALIBER, R. – CLICK, R. (eds.): *Readings in international business – A decision approach.* Massachusetts Institute of Technology, The MIT Press, Cambridge.

PILAT, D. (1996): Competition, productivity and efficiency. *OECD Economic Studies*, Vol. 2. No. 27. p. 107–146.

PORTER, M. (ed.) (1986): *Competition in global industries*. Harvard Business School Press, Boston.

SZABÓ, M. (2000): A multinacionális vállalatok hatása a magyar élelmiszergazdaságra és az EU csatlakozásunk folyamatára. Agrárgazdasági Kutató és Informatikai Intézet, *Agrárgazdasági Tanulmányok,* No. 12. Budapest.

MOBILITY OF HUMAN RESOURCES IN HUNGARY: AN ANALYSIS AND A PROPOSAL FOR REGULAR COLLECTION OF STATISTICS*

ERZSÉBET VISZT[1] – ZSUZSA PLANK[2] – BALÁZS BORSI[3]

This article is one of the first results of the work undertaken in the OECD working group. It targets to explore the measurement possibilities of labour force mobility and to lay the foundations for future cross-country comparisons. Highly qualified labour force, which has substantial impact on a country's innovative performance is focused in the OECD work. The authors' aim was twofold: to study the available Hungarian Labour Force Survey database (1993–1999) to reveal the possibilities of these statistics used for analysing the mobility processes of the transition economy and to show development alternatives of the Labour Force Surveys. This way the mobility of the Hungarian highly qualified labour force is becoming reliably measurable and internationally comparable.

KEYWORDS: Human resources; Innovation; Skilled labour mobility.

Although the human resource factor is often referred to in studies on innovation, due to the lack of availability of data the discussion usually goes little beyond general statements about its importance. The paper presents initial results based on the Hungarian Labour Force Survey (LFS), as an introduction to a discussion of possible ways of collecting statistical data on the mobility of the human resources devoted to science and technology (HRST) in Hungary. First, we review the conceptual background and present the key results of our 1998 pilot study, before going on to propose possible methods for measuring HRST. Finally, in the tables of the Appendix, we summarise the advantages and disadvantages of the methods proposed.

1. CONCEPTS AND CURRENT STATE OF RESEARCHES

Studying human mobility from basic statistics is never an easy task. In Hungary, human resource statistics of science and technology development are far from complete; unlike the Nordic countries, Hungary does not collect register data. The Hungarian Central Statistical

* The original version of this article was published in OECD (2001): *Innovative people*, Paris.
[1] Research manager of the GKI Co.
[2] Senior adviser of the HCSO.
[3] PhD student of the Budapest University of Technology and Economics.

Office (HCSO) collects statistical data on the number of employees with high qualifications, as well as on their sectoral, professional and other structural characteristics, but contains only limited information on changes of workplace and migration. However, mobility of highly qualified workers – not only among but also within companies and institutions (including changes in positions and individual career paths) – is an important catalyst for research and technical development; it is thus reasonable to aim its measurability by statistical methods. The NIS (National Innovation System) Focus Group on Human Mobility was initiated by the OECD with two aims: to map how the various OECD countries measure knowledge and the flows of knowledge which are important for innovation; and to obtain statistical measurements which would be comparable across countries.

The internationally recommended measurement methods – including the definition of mobility of human resources – are provided in the so-called 'Canberra Manual', compiled by the OECD (1995). According to the Manual, both those people with higher education whose qualifications are considered to be important from the aspect of technical development and those who work in this field despite having a different educational background, should be considered as the target group for research on the human resource aspects of science and technology development. Statistical analysis of workers with high qualifications is a relatively new topic, not only in Hungary but also in many other countries, and the research projects launched so far have only been able to examine some parts of the field.

2. USING THE LFS DATABASE TO MEASURE LABOUR MOBILITY IN HUNGARY

When the EU-harmonised quarterly surveys of the labour force were first launched in Hungary, the main objective was to study the economic activity of households. Thus, mobility was not, and is still not, the focus of these surveys. In 1992, however, a number of amendments were made in the survey structure in line with international recommendations. Accordingly, we can now analyse to what extent the LFS databases are suitable for studying mobility in Hungary. For this study, we have followed the methodological recommendations of the OECD, paying special attention to the study by *Mikael Åkerblom* on mobility of the highly qualified labour force (*Åkerblom*; 1999). The findings are presented in the following.

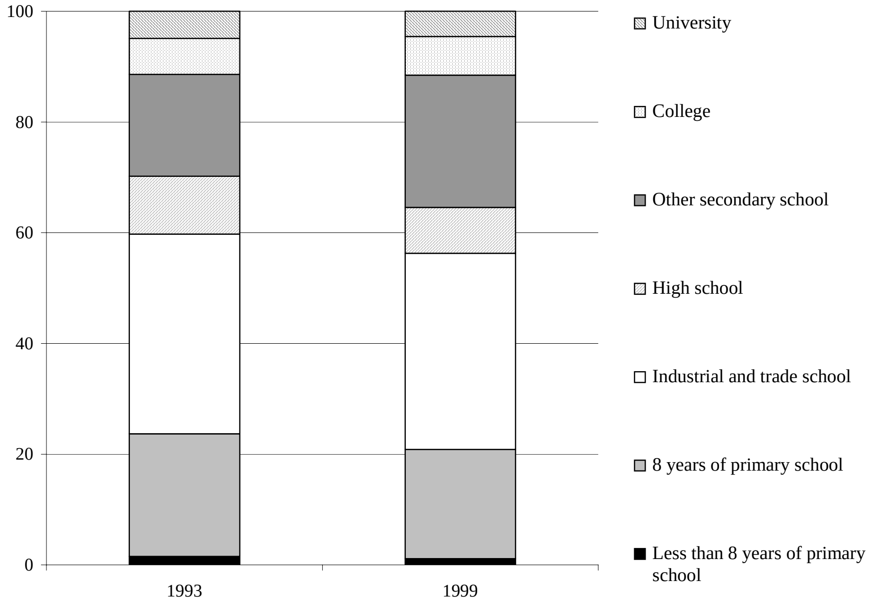
In 1993–1999, employment in Hungary shifted in favour of the young, with an increase in the share of 20-29 year olds entering the labour market. On the supply side, this rise in the most qualified and flexible manpower is a good sign. Nonetheless, the fall in the number of employees in the youngest age bracket indicates that the younger generation increasingly prefers to remain in training and further education. At the same time, expectations for older employees have deteriorated significantly.

In terms of labour force inflow by qualification, most of those entering the labour market had completed industrial and trade school and, by 1999, the second largest group comprised those with a specialised secondary qualification ('other secondary school' see in Figure 1). Thus, a substantial change in employment patterns is currently taking place: the low skilled – regardless of whether or not they have completed primary or secondary education (final examination at a high school) – are increasingly unable to find jobs. In

1993–1999, the share of highly qualified employees remained stable among those entering the labour market; this trend contrasts with that experienced in developed economies in the same period.

Figure 1. Distribution of those entering the labour market

By highest qualification (percentages)

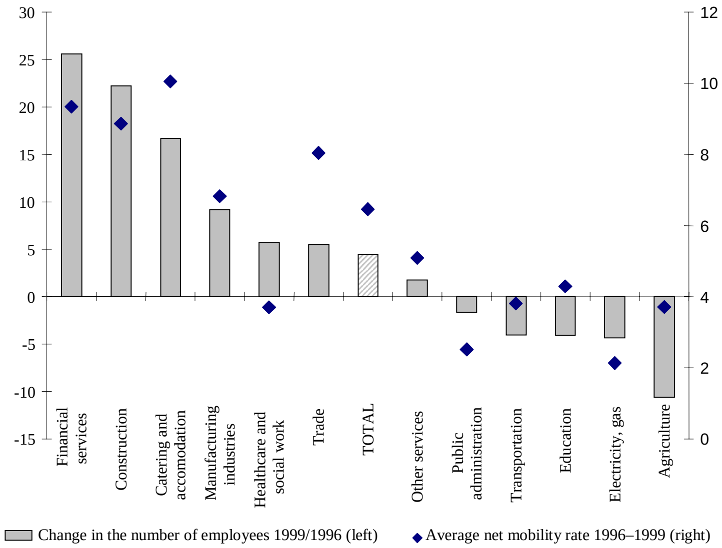


In 1993–1999, important sectoral changes took place, the results of which can be seen in the labour force outflow figures. The share of employees leaving the labour market decreased in manufacturing and agriculture, but it rose sharply in the sectors financed mostly by the state (education, healthcare, some cultural activities, etc.). There were two main reasons for this: public administration employees were laid off in an effort to achieve cost-savings; and wages in the public sector lag far behind those in the private (business) sector.

The share of those changing their place of work was high in industries where employment was on the rise. One possible reason for this is that the most mobile labour force – the young and skilled – is seeking jobs in the developing industries; these industries also offer greater opportunities for trying out several workplaces. In industries where employment has fallen (education, transport, agriculture, public administration), mobility tends to be fairly low. Low mobility is obviously a drawback with respect to innovative capacities – knowledge flows – since it signals that shrinking sectors also lag behind in terms of human resources, leading to further disadvantages.

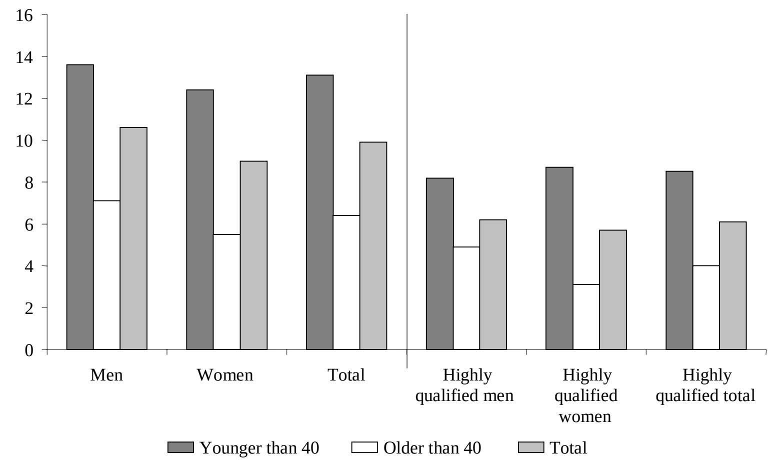
The 1998 database was used to examine the suitability of the LFS for constructing internationally comparable mobility rates. It was concluded that both the ‘broad’ and ‘narrow’ – or ‘gross’ and ‘net’ as termed in the literature – interpretations of the mobility rate could be computed. Broad mobility rates are expressed as the total inflows and outflows for the total number of employees, while narrow rates show the ratio of 'job-changers' only (as a percentage of total employees; net mobility rate concerns job changing among employees).

Figure 2. Change in the number of employees and the net mobility rate
(Percentages)



In 1998, 10 percent of employees changed their place of work; this share is approximately in line with the narrow rates of the EU-countries. The share for highly qualified workers, at 6.1 percent, is also similar to that in the developed countries.

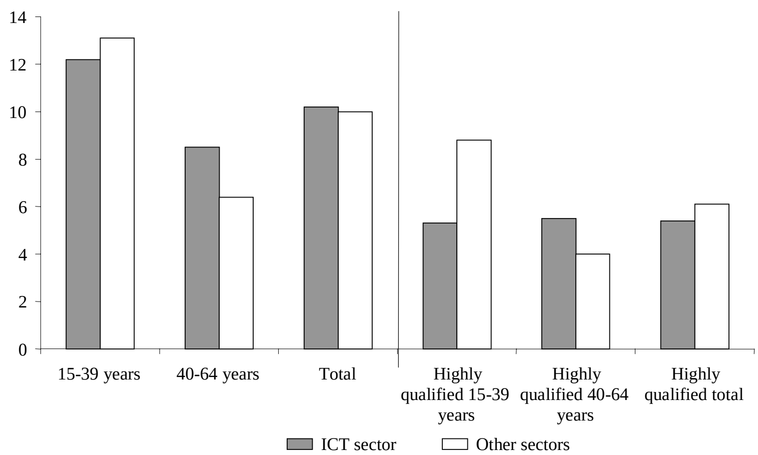
Figure 3. Mobility rates of the total and the highly qualified employees, 1998
(Percentages)



The breakdown by gender – a basic distinction in international practice – is also available. The figures show that men are more mobile than women, both in the sample and among employees with higher education degrees. The Hungarian figures correspond to known international patterns in other respects as well: the younger generation is more mobile, with no significant variations between genders, while among the older generation, men are more mobile than women. In both age brackets, the highly qualified are less mobile than the unskilled.

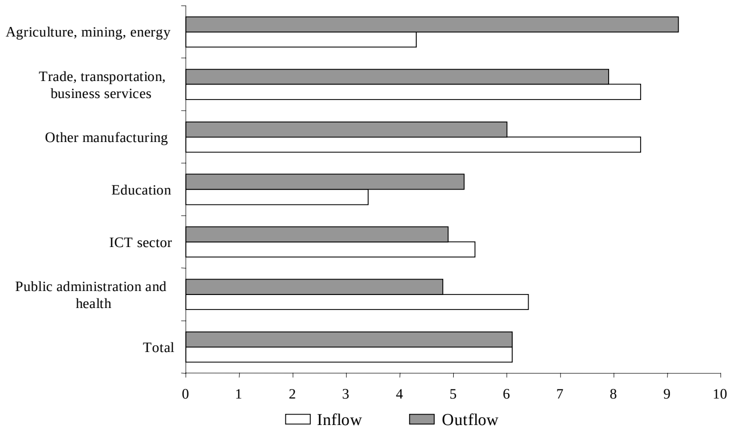
In the ICT (Information and Communication Technology) industry (mostly high-technology sectors belong here, see the Appendix), mobility of the older (more experienced) generation is higher than in other industries, which may mean that the outflow of 'knowledge' via mobility is more intensive in this sector. Differences in the inflows and outflows of highly qualified workers reflect the attractiveness and perspectives of a given industry.

Figure 4. Mobility rates of employee total and the highly qualified labour force, 1998*



* For an explanation of 'highly qualified' and the ICT sectors, see the Glossary.

Figure 5. Rates of inflow and outflow of the highly qualified, 1998



The higher the rate of outflow, the lower the attractiveness of an industry; higher inflows reflect greater opportunities for the highly qualified. In 1998, education and agriculture had the lowest rates of inflow, while manufacturing and some service sectors (business services – real estate and financial services) had the highest inflow. Agriculture was not attractive to job-starters and experienced a substantial outflow. A similar situation prevails in education, while high rates of both inflow and outflow in services highlight the dynamic changes taking place in this very fast-growing sector.

The share of highly qualified workers reflects knowledge intensity and its growing economic importance; the LFS therefore enables us to track the weight and dynamics of employment among this segment of the labour force.

Figure 6. Share of employees with higher education degrees, by branch
(Percentages)

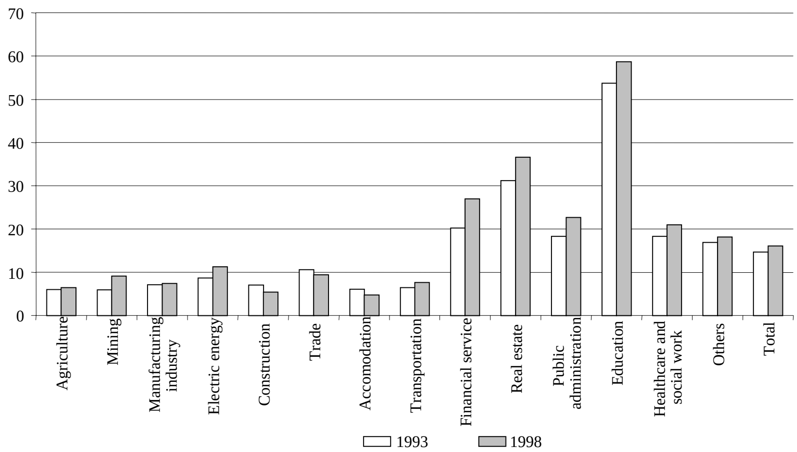
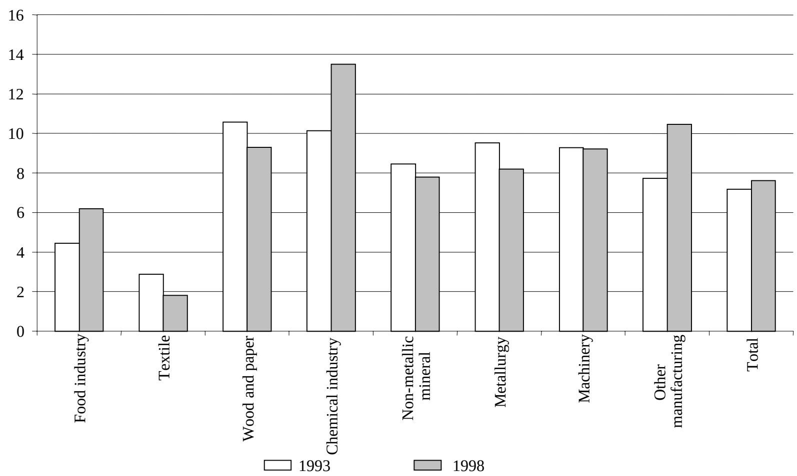


Figure 7. Share of employees with higher education degrees in manufacturing
(Percentages)

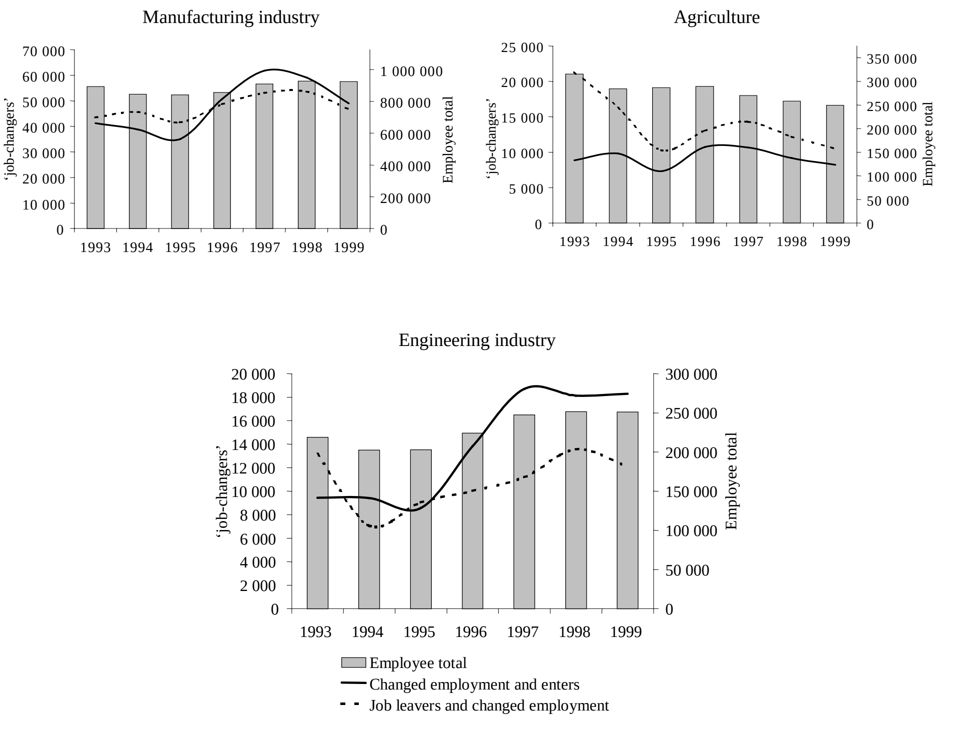


In the production sectors and some services (trade, catering and accommodation, transport), the share of employees with higher education degrees is low (10 percent and less). The public administration and healthcare sectors are more knowledge-intensive, with a share of between 20 percent and 30 percent; the business service sector boasts between 30 percent and 40 percent. The qualifications of employees in education are outstanding: more than half of those employed in this sector have a higher education degree.

Among the manufacturing industries, the chemical industry leads in terms of the share of employees with higher education degrees, although the food industry has caught up in recent years. However, in engineering – the sector with the highest performance in terms of output, productivity and exports – the share of highly qualified workers remains low at under 10 percent. These figures also refer to the R&D activity of the respective sectors.

The figures shown were constructed from the Labour Force Surveys of the Central Statistical Office of Hungary. Our experimental analysis enabled us to show sectoral differences between the broad and narrow (gross and net) mobility rates of employees. Due to the shortcomings of representativity, however, more detailed analysis using the current methodology of the HCSO was not possible. Misleading results would be obtained if, for example, the engineering industry was broken down into its sub-branches and employment categories, although this is clearly a very exciting field for future analysis.

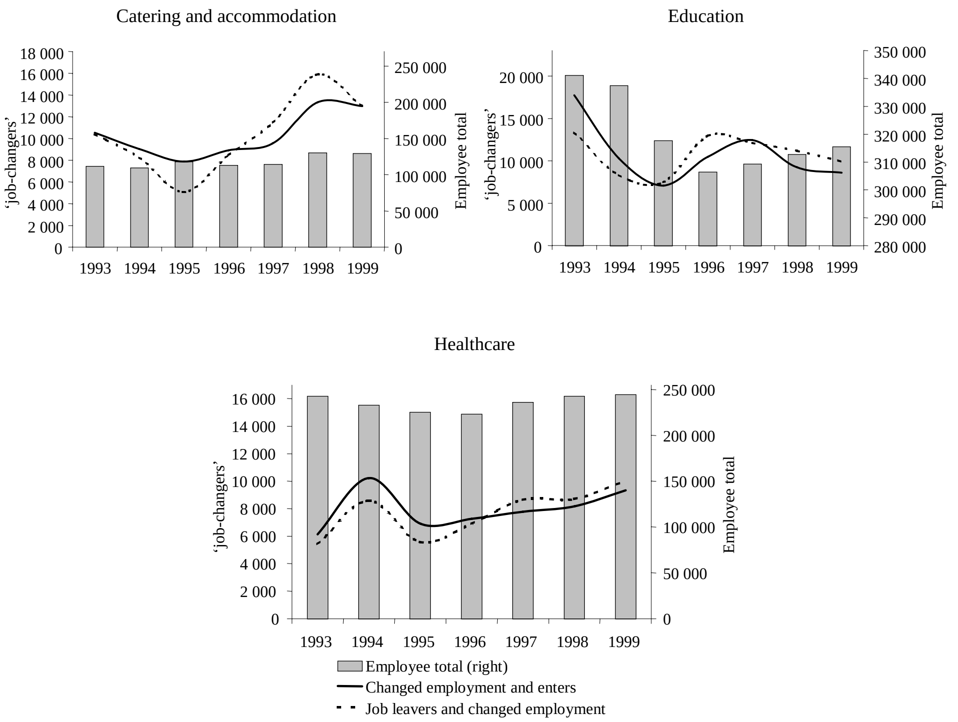
Figure 8. Number of employees and 'job-changers'



The overall net mobility rate in the 1990s was not high in Hungary, with substantial differences across sectors. The major structural changes which took place in the Hungarian economy explain some of these differences: manufacturing – especially engineering – gained momentum as a result of foreign direct investment and the activities of multinational firms. The most substantial depression took place in agriculture, which has a low mobility rate (4–5%). Structural changes are still underway in the engineering industry: the broad (or gross) rate of inflows and outflows stabilised at a high of 15 percent and 10 percent respectively; more young employees are being attracted to this sector today than was the case in the past.

Interesting changes are also taking place in some service sectors. In catering and accommodation, the number of employees remained stable. However, as the number of market actors – i.e. firms – is rising, mobility has picked up since 1996. In the state-owned, or government-dominated service sectors – public administration, education, healthcare – outflow is higher than inflow. Nonetheless, the narrow (or net) mobility rate is low due to the fact that these branches comprise many large organisations (especially in healthcare), and intersectoral mobility of teachers and medical professionals is poor. This is the case even though wages in these sectors remain below those in the competitive sphere.

Figure 9. Number of employees and 'job-changers'



In summary, in order to study human mobility – which is an important factor in explaining the restructuring and innovation processes – there is a need for more detailed statistics. If two- or three-digit sectoral data were available (General Industrial Classification of Economic Activities within the European Communities – NACE or Standard International Trade Classification – SITC) by profession (International Standard Classification of Occupations – ISCO) and education and skills (International Standard Classification of Education – ISCED), very interesting research projects could be undertaken despite the relatively short time series available. To achieve this, the current practice and methodology of Labour Force Surveys need to be revised and developed.

3. HOW TO GENERATE THE HRST DATABASE IN HUNGARY?

In principle, the human resources devoted to science and technology (HRST) database should include the highly qualified. Sources of relevant information available in Hungary are the HCSO censuses and micro-censuses which contain comprehensive data regarding education, orientation of graduate study and occupation. Censuses are held every ten years. In contrast with some European countries, the other possible registers (social insurance register, tax register, population register) do not contain complex information relevant to this subject of research and thus cannot be considered as a possible frame for sampling (in any event, current Hungarian regulations would not allow their utilisation). Only the census can provide the basis for the continuous sampling of individuals and households.

In the light of the previous, we investigated three different implementation methods for developing the Hungarian HRST database:

Method 1. Organising a panel from a representative sample of highly educated workers based on census data on the population with appropriate qualifications or occupation.

Method 2. Supplementing the questionnaire of the Labour Force Circulation surveys.

Method 3. Supplementary data collection of persons with higher education degrees in conjunction with the Labour Force Survey.

Data collection from a panel of people with higher education degrees

The census enables the compilation of a comprehensive database from which an appropriate sample may be selected and basic information gathered. However, time series could only be generated in one of the following two cases. An extended questionnaire containing questions on the various stages of the individual's career could be added to the census questionnaire. Accordingly, data could be analysed through a couple of years after the first questionnaire which would have to be filled in annually by the same sample of population. (This would, of course, imply legal and technical considerations, e.g. according to the current legal regulation, personal data must not be identifiable after the survey etc.)

Experts say that the questionnaire respondents would not have difficulties relating their position and workplace changes over the previous ten years of their careers, since in international comparison Hungarians do not change their workplace or position very frequently. The survey should involve at least 10 000 people, representing 1.7 percent of the total population with higher education degrees. A final decision on the size of the sample, however, would only be possible after testing.

A major drawback to this solution is its high cost: data collection from panels tends to be very expensive. The sample has to be relatively large since a certain number of people are expected to drop out from one year to another. Complicated, time-consuming questionnaires require trained interviewers, which increases the wages to be paid to the interviewers. More importantly, the selection and maintenance of the sample, and the constant development of methods for reconstructing the sample according to the original population properties, increase costs considerably. Furthermore, processing by individual programmes or organisational procedures is more costly than in the case of a simple survey or continuous population survey.

On the other hand, large amounts of data could be collected during the census, and this is one of the great advantages of this method. A relatively short supplementary questionnaire would therefore be sufficient, and is more likely to be filled in by the people interviewed. Another advantage is that it is possible to include in the sample those people who were most co-operative during the census, especially during the first interviews.

Supplementing the questionnaire of the Labour Force Circulation surveys

Until 1993, the Hungarian Central Statistical Office collected data on Labour Force Circulation (LFC) from the institutional labour statistics. Since the late 1980s, however, organisational changes – liquidations, start-ups – have become so frequent that this method can no longer be applied. The reason for the introduction of a supplementary survey on Labour Force Circulation in the first quarters since 1994 – in conjunction with the Labour Force Survey (LFS) data collection introduced in 1992 – was to make up for this deficiency (since then, data collection refers to individuals; the Appendix refers to this kind of data collection). Through interviews, the Hungarian Central Statistical Office collects data on those who quitted their employment, liquidated their private enterprise, entered new employment or started a private enterprise during the previous year.

From the basic sample of the LFS, those who have changed employment constitute the sample for the Labour Force Circulation survey; sample sizes of the Labour Force Circulation supplementary surveys ranged from 4 200 to 7 200 between 1994–2000. The sample covered 327 people with higher education degrees. This sample is too small to enable the collection of reliable information on the mobility of people with higher education, only a new survey can provide sufficient data to allow the more subtle characteristics of the mobility of the highly educated to be analysed.

Supplementary questions in the Labour Force Survey

The Labour Force Survey is the largest continuous statistical survey of households in Hungary. The core questionnaire does not include questions on mobility. The related supplementary survey would gather information on the career paths of the highly qualified workers included in the sample and on the motives underlying any change in employment.

The supplementary questionnaires could be distributed together with the core questionnaires, this is the a main advantage of this method. As a consequence, collecting basic information on the selected people or households and on the general characteristics of the labour market would not burden the supplementary survey. Furthermore, the sample of highly educated people belonging to HRST could provide retrospective information on changes in employment and other parameters (sector, position etc.).

The address register of the Labour Force Survey consists of 12 775 districts of the census performed in 1990, including 751 settlements and 626 000 addresses. Due to its size (approximately 10 percent of the registered addresses), the sample is representative at regional level NUTS2 as well.

The current sample of the LFS would enable some 6 000 people with higher education degrees to be interviewed. The LFS contains sufficient data on education to allow groupings according to both the Hungarian classification and the ISCED '97 (International Standard Classification of Education – ISCED) international classification. The occupations of people aged 15-74 are coded on the basis of a detailed description. A three-digit classification of occupation, corresponding to the ISCO '88 (International Standard Classification of Occupations – ISCO) international classification system, was introduced in 2000. In 2001, the Hungarian system will be enlarged to four digits to bring it in line with EU regulations.

Any drawbacks caused by the rotational character of the survey could be eliminated by the introduction of characteristics: 1/6 of the 6 000 people – i.e. 1 000 people – would remain in the sample for one and a half years due to the rotating method mentioned.

In the panel survey, interviewees would answer questions relating to

– Qualifications and skills.
– Changes in employment.
– Reasons for changing employment: motives of choosing the new employment; changes in the activities performed at work.

The panel survey would enable researchers to obtain information relating not only to the year in question but also to the professional careers of the respondents, thus facilitating analysis of the direction of knowledge flows over a longer period of time.

This method is more cost-efficient than those described former, since only the fees for interviewers and handling of the database are calling for additional expense.

ADVANTAGES AND DISADVANTAGES OF THE PROPOSED SURVEY METHODS

Although the use of a separate panel (Method 1) of people with higher education degrees would enable in-depth analysis of mobility of the workforce linked with innovation, the discontinuity of regular time series and the cost of the surveys make this method less feasible.

Due to the small sample size, the Labour Force Circulation survey (Method 2) of those with higher education degrees would yield only superficial information and would neglect the smallest organisations, the segment of the economy in which employment changes are most frequent. Admittedly, this is also a problem in the ongoing data collection on Labour Force Circulation, since many employees of small firms work in the underground economy and are thus unlikely to provide information on their movements. Moreover, at the present time, numbers of employees and trends in employment cannot be estimated from the Labour Force Circulation surveys.

In considering the implementation of the supplement to the Labour Force Survey (Method 3), which would appear to be the most reasonable statistical method of the three, it should be borne in mind that supplementary surveys can be extended by ad hoc surveys like the Labour Force Circulation survey. Another advantage of this method is that there would be no need to collect basic information, data gathering and processing costs would be considerably lower than those for Method 1 and the survey questions would only need to be devised once since they would be repeated in succeeding surveys.

APPENDIX

1. Data collection from a separate panel of people with higher education degree

Advantages	Disadvantages	Expenditures per survey
– The new sample would allow an analysis matching most of the necessary needs. – The sampled people would probably not be identical to those having already participated in other surveys, therefore would be more willing to answer. – The new sample could be selected on the basis of the census held in 2001. – Sample size (highly educated, according to age groups): 15–20 thousand people.	– The basic information needs to be gathered, which could be avoided if the survey supplemented an already existing one. – It could not be carried out before the second half of 2002. – Given that the survey is launched with an entirely new sample that cannot be combined with already existing surveys, initial costs would be considerably high especially on the first occasion. – Drop-outs are difficult to predict for the first 5-10 years, the sample may prove to be insufficient.	– Sampling. – Developing statistical methods. – Questionnaire. – Methodological assistance . – Preparation of interviewers. – Fee of interviewers. – Statistical control of the responses. – Material costs. – Software development. – Data processing. – Writing up the survey. **Total expected costs in the first year: euro 170 000 in the following years: euro 34 000.**

2. Supplementing the questionnaire of the Labour Force Circulation surveys

Advantages	Disadvantages	Expenditures per survey
– Adding new questions to the supplementary Labour Force Circulation survey would make the development of a new survey unnecessary.	– The survey would only yield data in the case of people, who change jobs in the given year, therefore insufficiency of sample size would not allow a more detailed analysis. – Sample size (participants of the supplementary Labour Force Circulation survey in the first quarter of 2000): 327 people.	– Preparation of interviewers. – Supplementary fee to interviewers. – Modification of software. – Writing up the survey. **Total expected costs: euro 14 000.**

3. Supplementary questions in the Labour Force Survey

Advantages	Disadvantages	Expenditures per survey
– The supplementary survey would not need to collect basic information. – Data gained from the Labour Force Circulation survey could also be used. – Data gathering and processing costs would be considerably lower than those of Appendix. – Questions in the surveys could be repeated in other surveys following the first occasion.	– The number of people, who refuse answering, is expected to increase with the introduction of a further questionnaire. – Costly sample size (people with higher education degree in the core sample): 6 000 people.	– Questionnaire. – Methodological assistance. – Preparation of interviewers. – Fee of interviewers. – Statistical control of responses. – Material costs. – Software development. – Combining data recording, collection, and control. – Processing. – Writing up the survey. **Total expected costs: euro 24 000.**

Glossary of terms

Gross (or wide) mobility rate: a percentage rate of all employees, it concerns every movement in the employment status, i.e. – beside job-changers – beginners and those, who die, etc. are also included in the gross rate of mobility. To put it simply, everybody, who enter or leave the labour market and change employment, is included.

Highly qualified labour force: 'highly qualified' has two interpretations: one of them is associated with the abilities and experience of the employee, the other is associated with education. In this article and in most statistics 'highly qualified labour force' means employees with a college degree or higher.

ICT (Information and Communication Technology) sectors: high technology sectors. By the OECD definition, they include manufacture of pharmaceuticals, manufacture of office machinery and computers, manufacture of radio, television and telecommunication equipment, telecommunication.

Net (or narrow) mobility rate: a percentage rate of all employees, it concerns only the actual job-changers, i.e. only those employees, who change employment.

REFERENCES

AKERBLOM, M. (1999): *Mobility of highly qualified manpower: A feasibility study on the possibilities to construct internationally comparable indicators*. OECD, December.

The measurement of scientific and technological activities (1995). Manual on the measurement of human resources devoted to S&T. Canberra Manual, OECD, Paris.

CHANGES IN FOREIGN TRADE STRUCTURE AND INTRA-INDUSTRY TRADE IN FIVE CEECS*

ANDREA ÉLTETŐ[1]

The article analyses the structural changes of foreign trade between five Central and Eastern European countries and the EU. It points out that during the nineties, a significant shift has taken place towards technologically more developed products. The former specialisation on low-technology, labour intensive goods strongly weakened and in certain cases new specialisation patterns have been developed among the high-tech products. The level of manufacturing intra-industry trade increased and this in several cases meant product quality improvement as the results of the unit value calculations showed. For these changes in the foreign trade structure those – high and medium-tech – products are responsible which are produced by foreign owned, multinational companies. FDI therefore exerts a determinant effect on the external competitiveness of Central European countries.

KEYWORDS: Foreign trade, Foreign direct investment.

The Central and Eastern European countries are striving for the European Union membership. In this case they will be part of the single internal market experiencing the advantages and drawbacks of the increased competition. The competitiveness of a country is often judged on the basis of its export performance on the foreign markets. The dynamics and the composition of the exports, the complexity and quality of the traded goods reflect the development of their domestic economy. In the era of globalisation and increasing competition this kind of competitiveness is getting more and more important.

Recently, that a decade has passed since the basic political changes in this region, it is instructive to analyse the development of the foreign trade. The paper focused its analysis on those five Central and Eastern European countries (CEEC-5) who were first selected by the EU to begin the accession negotiations: Poland, Estonia, Hungary, Slovenia and the Czech Republic. The subject of the analysis is the foreign trade of these countries with the European Union. In this respect, this paper analyses the changes in the product structure, and specialisation trends emphasising the technological level of the products.

* The article is based on the research supported by the European Union's Phare ACE programme No. P97-8112-R. The content of the publication is the sole responsability of the author and it in no way represents the views of the Commission or its services.

[1] PhD, Senior researcher of the Research Center of Economics of the Hungarian Academy of Sciences.

Among developed countries, a considerable part of trade is intra-industry trade, which means trade with similar, diversified manufactured products. This type of trade can be the exchange of the same goods on the basis of different packing or seasonal effects, can be the exchange of differentiated or substitutive goods or can be induced by intra-industry co-operation. This article examines to what extent this kind of trade increased in the CEEC-5 during the nineties, showing thus similar trends to the developed countries. Based on an in-depth analysis of the intra-industry trade patterns conclusions can also be drawn regarding the quality of the exported products.

The foreign trade performance of a country – especially if its size is small – is often influenced by the strongly export-oriented activity of multinational and foreign firms. Multinationals give a determinant part of world trade at the end of the nineties and they have realised important investments in the CEEC-5 region. According to our hypothesis therefore, the changes taken place in the foreign trade, have been in great part due to the activity of the foreign investment enterprises in the countries.

1. CHANGING FOREIGN TRADE STRUCTURE

At the end of the nineties the foreign trade of the CEECs has already been tied to the European Union, the geographical composition of exports and imports has been modified in favour of the EU. Apart from the geographical factor, there have been important changes also in the product composition of foreign trade. These changes in the manufacturing product structure are in the center of the following analysis.

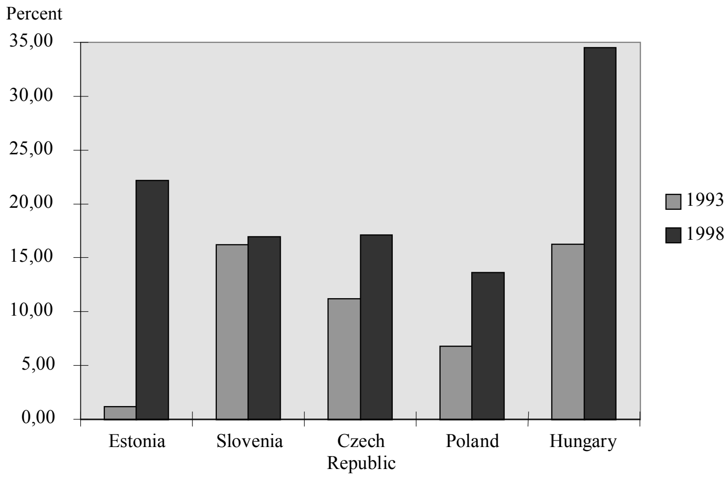
The article uses the industry classification based on the method of OECD (1993) made in ISIC (International Standard Industrial Classification of All Economic Activities) classification. Three groups were created in this classification: high-technology, medium-technology and low-technology intensive products.[2] All the calculations referring to foreign trade dynamics and indices are made at the Standard International Trade Classification (SITC) 5 digit product level (3464 items) given by the EUROSTAT Comext database.[3] This is a level detailed enough to avoid all biases stemming from the aggregation of products. Data were converted afterwards to ISIC Rev3 classification, in order to apply the former mentioned technology grouping of sectors. (This sectoral approach has certainly limitations because of the not sufficiently disaggregated data. Certain products manufactured by high-tech sectors are medium or low tech and vice versa. The OECD therefore developed a product approach, but only for the high-tech products. Though this approach has also its own limitations (*Hatzichronoglou*; 1997) the sectoral approach is applied here.)

In analysing the technology-intensity of traded products, first the development of the high-tech group is followed. According to the OECD definition pharmaceuticals, telecommunication equipment, office machinery, aircraft-spacecraft, precision instruments and electrical machinery branch belong to this group. Figure 1 shows the development of the high-tech exports to the EU in the case of the CEEC-5 between 1993 and 1998.

[2] The indicator of technological intensity (weighted according to sectors and countries) is the share of R&D expenditures in output or value-added. At the end of the nineties based on the experiences the OECD revised the grouping (*Hatzichronoglou*; 1997) and divided the medium-tech group into two parts. Medium-high and medium-low groups were created, precision instruments and electrical machinery were put into the former one. However, the former grouping is applied in the paper.

[3] In this database the EU is the reporter country, so 'CEEC-5 exports to the EU' means the EU imports from the CEECs.

Figure 1. The share of high-tech branches in the exports of the CEEC-5



Source: Own calculations based on EUROSTAT Comext data.

The first striking phenomenon is the extremely rapid increase in the share of high-technology sectors in manufacturing exports in Hungary and in Estonia. This share was the highest in 1998 in Hungary, reaching 34 percent.[4] The definite growth in the case of Poland and the Czech Republic is also remarkable.

If we examine the detailed sectoral composition of the exports and imports it turns out that in all five countries in most cases the same sectors increased their export share during the period examined. The strong increase in high-tech exports is due to the electrical machinery, telecommunication equipment, precision instruments and office machinery. Certain medium-technology sectors also increased their share in all countries, like chemicals and plastics, but above all and the most effectively the motor vehicle branch.

The share of low-technology sectors, however decreased rapidly, mainly due to the food and beverage, textile-clothing industries. These – traditionally strong – sectors in all country decreased significantly their share in the exports to the EU. In some cases it did not mean a decrease also in the absolute value of the exports but in some cases the absolute value decreased too.

In 1998 the share of high-tech products was almost the same in the Slovenian and in the Czech exports to the EU, around 17 percent. However, in the Czech case there was a considerable increase since 1993 while in the case of Slovenian one there is hardly any change. In the Czech exports the electrical machinery sector is responsible for the increase. The share of low-tech product group decreased since 1993, due to the textile-clothing sector.

In the Slovenian exports the share of medium-tech products increased, which has been caused by the motor vehicle and machinery sector. Similarly to the other countries, a decrease in the share of low-tech products can be observed in the exports. This is mainly due to the textile-clothing sector.

[4] This figure can be compared to the extremely high Irish share, which was 37.9 percent in 1998.

Regarding the high-tech products it is interesting that in Hungary the traditionally important pharmaceutical sector lost some of its importance in the EU-relation, although in other, Eastern-European relations it gained positions. The medium-technology sectors also increased their share for which the motor vehicle branch is entirely responsible. The share of low-technology sectors, however decreased rapidly.

Concerning the Polish exports, the share of high-tech products has also increased considerably between 1993 and 1998. The share of medium-technology group remained more or less the same, while the share of low-tech products decreased in the exports to the EU due to the food and textile branches.

As far as Estonia is concerned, the telecommunication equipment (radio-TV sets) branch is the most important among the high-tech groups, but office and electrical machinery sectors also increased considerably their share. At the same time, the role of medium-tech products drastically decreased in the exports to the EU, from 30 percent to 10 percent. As an exception among the examined countries the share of low-tech products remained at the same high level (70%), textile, wood, and refined petroleum products are still important components of the Estonian exports.

Table 1

Share of product groups in the CEEC – EU trade
(percent)

Country, product groups	Exports	Imports	Exports	Imports
	1993		1998	
Hungary				
High technology	16.26	18.43	34.54	25.88
Medium technology	24.62	44.89	37.12	47.38
Low technology	59.12	36.68	28.34	26.74
Manufacturing	100.00	100.00	100.00	100.00
Estonia				
High technology	1.16	8.27	22.17	21.48
Medium technology	29.07	32.28	9.54	24.17
Low technology	69.77	59.45	68.28	54.35
Manufacturing	100.00	100.00	100.00	100.00
Czech Republic				
High technology	11.21	21.42	17.11	23.82
Medium technology	33.45	48.86	44.43	44.76
Low technology	55.34	29.72	38.46	31.42
Manufacturing	100.00	100.00	100.00	100.00
Slovenia				
High technology	16.20	12.19	16.97	14.23
Medium technology	29.47	45.33	41.03	46.47
Low technology	54.33	42.48	42.00	39.29
Manufacturing	100.00	100.00	100.00	100.00
Poland				
High technology	6.78	15.60	13.59	17.58
Medium technology	27.20	46.13	30.55	47.21
Low technology	66.02	38.27	55.86	35.20
Manufacturing	100.00	100.00	100.00	100.00

Source: Own calculations based on EUROSTAT Comext database.

In the structure of the imports from the EU we can find that the share of high-tech products in the exports is considerably higher than their share in the imports in the case of Hungary and slightly in the case of Estonia and Slovenia. However, in absolute value this means only in Hungary more high-tech exports than imports, in the case of Slovenia the value of high-tech exports is slightly lower than the value of imports (although it was much higher in 1993), and Estonia imports considerably more high-tech products than exports.

Generally speaking in terms of foreign trade and in terms of exports on the EU market, the competitiveness of the CEEC-5 countries has improved in the nineties. This means that the export structure has become more similar to that of the developed countries, with more and more high-tech products in the foreground. Apart from the modern structure, another feature of the foreign trade among the developed countries is the relatively high share of intra-industry trade. The following part analyses this kind of development in the case of the CEEC-5.

2. THE DEVELOPMENT OF INTRA-INDUSTRY TRADE

In general, for the analysis of intra-industry trade (IIT),[5] the index of *Grubel–Lloyd* (1975) is used. However, since the work of *Greenaway – Milner* (1994) two types of intra-industry trade can be distinguished. The first is called vertical IIT, when the products traded are of the same type but different in quality, the other is the horizontal IIT, when also the quality of products is comparable.

Separating vertical and horizontal IIT is important for many reasons. Regarding the effect of integration, in the case of countries which are of different development level, integration can enhance vertical IIT. In this case, products of the less developed country, which are of lower quality can be crowded out by better quality imports of more developed countries, thus the costs of adjustment can be high. Regarding this theory, empirical verification of the role of scale economies in creating IIT remained rather poor.[6] Calculations showed that generally vertical IIT is much more significant than horizontal IIT, therefore interest has grown in analysing and explaining vertical IIT.

Falvey (1981) pointed out that difference in quality among similar goods (that is vertical IIT) on the supply side is caused by the differing capital/labour ratio of their production. High-quality products require more capital-intensive production techniques. On the demand side there is an aggregate demand for a variety of differentiated products, low-income consumers will buy lower quality products, high-income consumers high quality products. A relatively labour abundant country will export the lower quality/labour in-

[5] As it is well known, IIT is characteristic for the sophisticated manufactured products. Monopolies, increasing returns to scale, homogeneous consumer preferences in partner countries explain this type of trade. Intra-industry trade is especially intensive among developed countries, which trade with similar, diversified manufactured products. The more similar the factor endowments of the partner countries are, the greater the extent of IIT is. It should be mentioned that intra-industry trade is often mixed with trade within the production vertical. Thus, if a country imports motors and exports cars, it is not intra-industry trade although in a high enough aggregation level both products belong to the 'vehicles and components' category. Therefore proper disaggregation is very important in the measurement of intra-industry trade. It should be clarified that intra-firm trade – between a multinational parent company and affiliates – can be of intra-industry type but not necessarily, intra-firm trade is not a part of intra-industry trade and vice versa.(*Fontagné* et al.; 1995). The propensity for intra-firm trade differs according to industries and also depends on firm and country-specific factors, among them the size of FDI involved (see *Dunning*; 1993). Applying intra-firm trade multinational companies can manipulate the terms of trade, tackle the different tax systems of governments.

[6] For details see *Blanes-Martín* (2000).

tensive version of the product (aiming low-income consumers abroad) and will import the higher quality product (for high-income consumers on the domestic market). Thus IIT is explained by comparative advantages. From another aspect *Davis* (1995) also shows that IIT can take place without increasing returns and imperfect competition. In that case the emphasis is on technical differences among the countries which determine specialisation on one or other types of intra-industry product.

As it was mentioned, the basic indicator used to measure intra-industry trade is the Grubel-Lloyd index. The definition of the index for a given product group '*i*' is the following:

$$B_i = \left[1 - \frac{|X_i - M_i|}{(X_i + M_i)}\right] * 100$$

The index for the whole economy (or a sector group) is the weighted average of the product group indices according to the weight of the product groups in foreign trade (W_i). X and M are exports and imports respectively:

$$B_{iw} = \sum W_i B_i$$

where

$$W_i = \frac{(X_i + M_i)}{\sum (X_i + M_i)}$$

The value of the index can vary between 0 and 100, a higher index means a higher level of IIT. Note that the less detailed aggregation is used, the higher the value of the index is, IIT should therefore be calculated at a very detailed level of classification.[7]

In the method of separation of IIT into horizontal (HIIT) and vertical (VIIT) types quality differences of exports and imports are used.[8] If the export and import unit value differ by less then 15 percent then IIT is horizontal (the traded goods are of the same quality), if the difference is bigger in such a way that export unit values are higher then IIT is high quality vertical, otherwise IIT is low-quality vertical.[9]

In the calculations we apply the SITC 5 digit level classification. The results were converted and grouped according to the technology-intensity level in line with the ISIC classification already used.

Figure 2 gives a 'general view' of the situation in the five examined countries. The first observation one can make is that the share of intra-industry trade is in every country higher in 1998 than it was in 1993. The increase has been especially spectacular in the case of Estonia and Slovenia. Compared to the others, Estonia has still the lowest share, but shows the highest increase. For 1998 Slovenia had the second largest share of IIT in the trade with the EU.

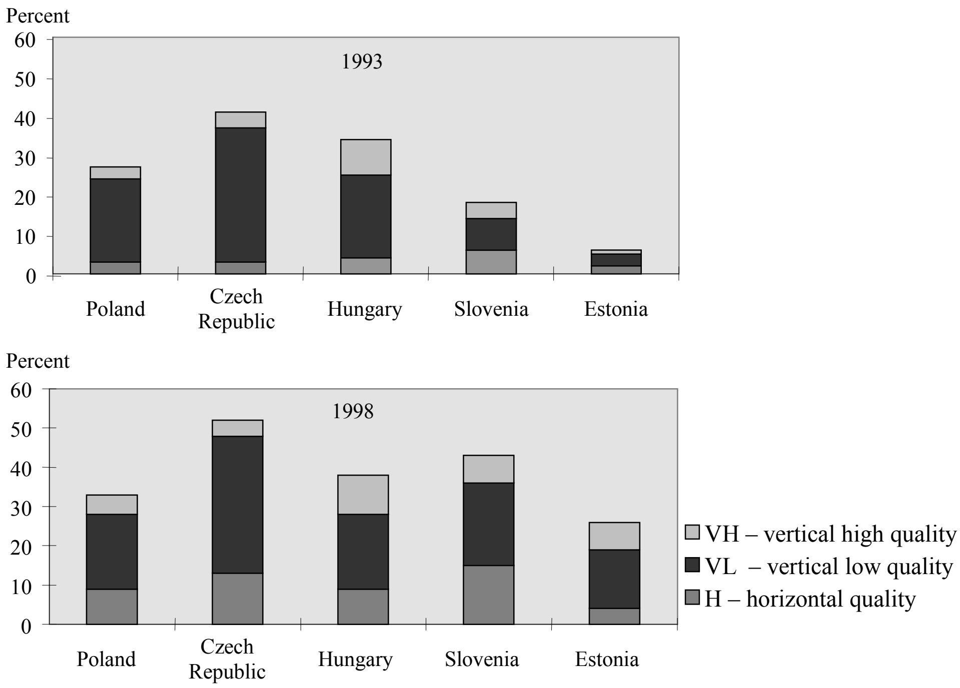
[7] About problems of the GL indicator and other types of measures see *Vona* (1991) for more details.

[8] The principle of 'bigger price, better quality' can be criticised (for example some products can be overpriced), however there is no better method found out for signing quality differences.

[9] If $0.85 \leq UV_X / UV_M \leq 1.15$ then IIT is horizontal.

The level of intra-industry trade was and is the highest between the Czech Republic and the EU. This means that in 1998 more than the half of the manufacturing trade was of intra-industry type. This is almost as high as the Spanish figure (58.8%).[10] Regarding the types of IIT, in every country the vertical low quality IIT dominates, similarly to the international experience.[11] Hungary shows the highest share in the vertical high quality type IIT (where the quality of the exported products is better than that of the imported ones) but Slovenia and Estonia also could increase this type of IIT.

Figure 2. Development of intra-industry trade in the CEEC-5



Source: Own calculations based on EUROSTAT Comext database.

Detailed calculations of IIT at the sectoral level are available in electronic format (www.ksh.hu/statszml). It can be stated that in the CEEC-5 in almost every branch, intra-industry trade increased between 1993–1998.

As regards Hungary, among the high-technology groups a decrease in (mostly vertical low quality) IIT can be observed in the case of pharmaceuticals, but for telecommunications equipment and electronic machinery a significant increase is manifested in

[10] Éltető (2000), p. 159.

[11] Obviously, if we regard the CEEC-5 trade separately with EU-members, we get a heterogeneous picture. Therefore, the level of intra-industry trade depends not only on the aggregation level but also on the country group we define. This phenomenon is called 'geographical bias' by *Fontagné-Freudenberg* (1997) who argue that when different partner countries are put together, the sign of the trade balance for a given product may change from one partner to another and will show up as a 'multilateral' intra-industry flow, which is a pure artefact. Though conscious of this fact, in the calculations EU is assumed to be one geographical unit.

horizontal IIT. Outstandingly high is the share of IIT in the office machinery branch (33.9%). In the case of medical, precision instruments low quality vertical IIT increased.

In the case of low technology sectors, increase in intra-industry trade (although from a low level) is general and in several cases means an increase in horizontal or vertical high quality IIT (textile, paper, metals), which suggests quality upgrading. Regarding medium-technology sectors, the situation is the same. The increase in vertical high quality IIT is especially spectacular in the case of transport equipment (except for railway locomotives).

In the case of Estonia, the considerable increase of IIT was caused by the vertical low and high quality type. The latter increased mainly in the trade of high-tech products, telecommunication, electrical machinery and pharmaceuticals. Among the medium-tech products transport equipment show an increase and plastic products a decrease of vertical high quality IIT. Regarding the low-tech group, intra-industry trade increased significantly in almost every product categories.

In the Czech trade of high-tech products the Grubel-Lloyd index for pharmaceuticals decreased. This is similar to the Hungarian phenomenon though this decrease was due to the vertical high quality IIT. Horizontal IIT increased in the trade of telecommunication equipment and vertical high quality IIT increased in medical precision instruments. In the trade of medium-tech products IIT (of vertical low quality) decreased quite significantly in motor vehicles and manufacture of transport equipment, however in the motor vehicle branch there has been an important increase in horizontal IIT. Horizontal IIT increased in the trade of low-tech products in textile, paper, refined petroleum, fabricated metals, and vertical high-quality IIT increased in the case of wood products; hinting to a quality upgrading.

As it was mentioned, the level of Slovenian IIT with the EU increased to a great extent to 1998, which was mainly due to the horizontal and vertical low type of IIT. Vertical high quality IIT increased in the trade of high-tech products, locomotives and other manufacturing but decreased in the case of non-ferrous metals, motor vehicles and plastics. There are certain low-technology products where horizontal IIT is quite high, like paper, textile, basic and fabricated metals.

Poland experienced a moderate increase in intra-industry trade. There were several product groups where IIT decreased, like pharmaceuticals, precision instruments, rubber, transport equipment, locomotives, etc. Horizontal IIT increased remarkably in telecommunication equipment, electrical machinery, motor vehicles.

Analysing the increase of IIT, it is important to know the 'technical' composition of this increase. Stemming from the properties of the IIT index, an increase of intra-industry trade can be observed in certain cases if the trade balance improves but also if the trade balance worsens. The IIT indicator can increase if imports grow from a lower level than exports but exports remain similar. Therefore it is considered important to analyse what the increase of intra-industry trade caused in these countries. The developments of the trade balance of a sector are manifested in the foreign trade specialisation patterns of the countries.

3. SPECIALISATION PATTERNS

As it was introduced the structure of the foreign trade has changed considerably among the CEEC-5 and the EU during the nineties. Based on this one can presume that there have been changes in trade specialisation patterns too. In this respect the index calculated is a 'specialisation index' or 'net export index' (*Balassa-Noland*; 1987) and its definition is:

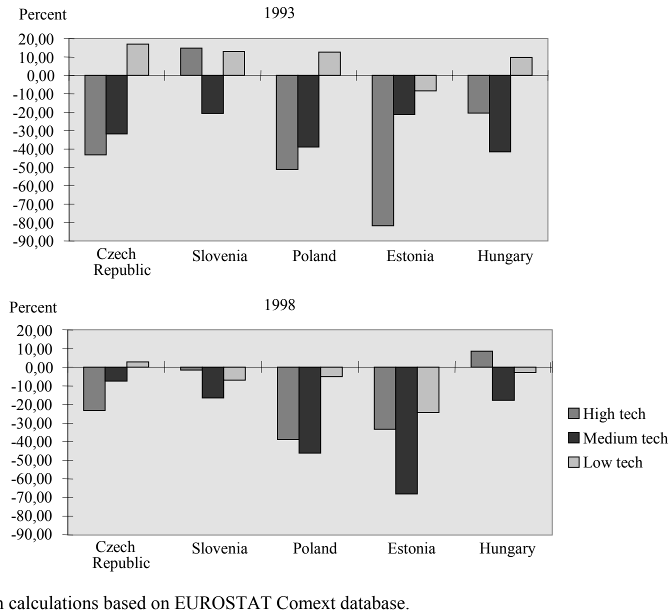
$$SI_i = \frac{X_i - M_i}{X_i + M_i} * 100 \ ,$$

where X_i is the exports and M_i is the imports of sector i. For any aggregates SI is computed as the rario of the aggregated numerator ($\sum(X_i - M_i)$) and the denominator ($\sum(X_i + M_i)$), expressed in a percentage form. One can speak about advantages, or specialisation if the value of the index is positive. Because of the properties of the index, its positive value means a positive trade balance.

Results of the calculation of SI indices for the five CEECs is shown in Figure 3. It can be seen that in 1993 four countries had clear specialisation on low-tech products, Estonia was the only exception. To 1998 this specialisation vanished remaining only to a tiny extent in the Czech Republic. This means that in these low-tech branches the trade balance considerably worsened and turned into a deficit.

Despecialisation on high-tech products at the same time decreased in the countries (except for Slovenia, which was in 1993 specialised on high-tech products but despecialised in 1998). In the case of Hungary, the SI index has even turned to positive for 1998, which already shows a clear specialisation caused by the surplus in trade balance of the high-tech products. Estonia, Poland and the Czech Republic improved their trade balance position considerably in high-tech products but could not turn this into positive. An improvement can be observed in the medium-tech group also, except for Estonia where a strong worsening of despecialisation can be seen.

Figure 3. Specialisation indices in the CEEC-5



Source: Own calculations based on EUROSTAT Comext database.

Regarding Hungary, at those – mainly low-tech – branches where Hungarian specialisation had been traditionally of a high degree (textile, food, wooden products, coke and petroleum, etc.) there are still positive indices but a radical decreasing or stagnating trend can be observed during the nineties. At the same time, new specialisation patterns appeared within the high-tech products, in office machinery and telecommunication equipment. In electrical machinery, specialisation remained existing. On that basis, one can say that Hungary has lost its specialisation in labour-intensive, low technology goods, but at the same time has been able to develop an important export activity in high technology goods. Exports have been increasing considerably faster than imports in these cases.

Table 2

Manufacturing branches where specialisation improved or despecialisation decreased
(1993–1998)

Poland	Slovenia	Czech Republic	Estonia	Hungary
Radio-TV sets	Office machinery	Office machinery	Office machinery	Office machinery
Electrical machinery	Chemicals	Radio-TV sets	Radio-TV sets	Radio-TV sets
Medical, precision instruments	Machinery and equipment	Electrical machinery	Electrical machinery	Medical, precision instruments
Plastic products	Railway locomotives	Medical, precision instruments	Medical, precision instruments	Railway locomotives
Rubber products	Motor vehicles and parts	Plastic	Machinery and equipment	Manufacture of transport equipment
Manufacture of transport equipment		Machinery and equipment	Railway locomotives	
Boats		Non ferrous metals		
		Motor vehicles		
		Boats		
		Fabricated metals		

In the case of the Czech Republic there is a strong improvement in specialisation to medium-tech products. As for of rubber products, locomotives, and motor vehicles the SI index had a positive sign in 1998 (in 1993 it was negative for the motor vehicle products). Among the high-tech groups there is also an improvement, except for pharmaceuticals. In the case of the low technology group although specialisation remained, it has considerably deteriorated since 1993.

Slovenia has also experienced the loss of specialisation on the traditional, low-technology-intensive products. But also in other fields, generally, there is a decrease of the SI value in almost every sectors except for chemicals, locomotives and pharmaceuticals. This hints to a general deterioration of the Slovenian trade balance with the EU.

In the Polish trade there is no specialisation on high-tech products, although an improvement of the SI index can be seen at the electrical machinery branch. Among the medium-tech sectors in two cases we can see a clear specialisation, with positive SI value: by non ferrous metals and manufacture of transport equipment. The specialisation on low-tech products, which was apparent in 1993 has been lost for 1998. The SI index remained positive only for wood products and building of boats.

Estonia also does not show a positive SI index for the high-tech products, although the value for office machinery products improved considerably. Similarly to this, there is no one medium-tech product group either with positive index value. In the case of low-tech products, however, we can find three fields of specialisation: textile-clothing, wood products and refined petroleum. In the latter two cases specialisation is quite strong.

Having an overview on the foreign trade specialisation patterns (and trade balance situation), one can examine which are the sectors where the increase of the intra-industry trade has stemmed from the worsening of the trade balance.

– Czech Republic: chemicals, railway and tramway locomotives, food, beverages, tobacco, textile, clothing, leather, paper and printing, basic metals.

– Slovenia as there are only a few sectors where the trade balance improved, so it is no wonder that in the most cases the increase in IIT was caused by the worsening trade balance: office and electrical machinery, precision instruments, rubber products, motorcycles, food, beverage, tobacco, textile, clothing, leather, wood, paper, basic metals.

– Poland: non ferrous metals, food, beverage, tobacco, textile, clothing, leather, paper, refined petroleum, non metallic minerals, fabricated metals.

– Estonia: pharmaceuticals, chemicals, rubber, plastic, transport equipment, other manufacturing, textile, clothing, leather, wood, paper, non-metallic minerals, boats.

– Hungary: rubber products, food, textile-clothing, wood, paper and printing, petroleum, minerals, non-metallic minerals, basic and fabricated metals, boats.

It is obvious from the list that in every country, at the big majority of low-tech products the increase of intra-industry trade has been due to the worsening of the trade balance. On the contrary, it can hardly be found any high-tech products where the IIT increase stemmed from the growth of the foreign trade deficit. This means that in these cases the reason of the intensification of the intra-industry trade is something else.

4. THE EFFECTS OF FOREIGN DIRECT INVESTMENT

The impact of foreign direct investment (FDI) on foreign trade has been an increasingly important issue in the nineties as for a continuously growing part of the international trade is being realised by multinational enterprises (for an overview of theoretical evidence in this respect see for example (*Cantwell-Bellak*; 2000). In the early theories of foreign direct investment and multinational firms, FDI and foreign trade were considered to be substitutes. First *Mundell* (1957) built a model where both FDI and foreign trade were based on the price differences of products and production factors were determined by the different factor endowments of the countries. The product-cycle theory of *Vernon* (1966) was also based on this substitution principle, where foreign direct investment replaced the export as the product matured.

Still based on the traditional comparative advantage theory, *Kojima* (1975) introduced the concept of trade-oriented (pro-trade) and anti-trade-oriented FDI based on the theory of comparative advantages. According to this, one can speak about trade creating, or pro-trade FDI if the investment is undertaken from the home country's comparatively disadvantaged industries into the host country's comparatively advantaged industries.

Both countries gain from the following trade creation. In the case of anti-trade FDI, however, investment is undertaken by a firm of the home country's comparative advantage industry into the host country's comparative disadvantage industry. In this way the home country has an excess demand for importable goods and an excess supply of exportable goods. The two countries are competing in importing and exporting capacities, thus FDI can even destroy trade.

At the end of the seventies 'new international trade' theories emphasize, however the complementary relationship between FDI and foreign trade (see for example *Krugman*; 1990 and 1991, *Venables*; 1996). This is the result of introducing new aspects in the models like increasing returns to scale, product differentiation, technology-differences among nations. Allowing for these factors and assuming identical relative factor endowments *Markusen* (1983) proved that factor (capital) movements between two economies lead to an increase in the volume of trade.

The different kind of investments have different effects on the foreign trade (*Dunning*; 1993).[12] The resource based investment aims to exploit some kind of natural resources. The efficiency-seeker or export-oriented investment aims to exploit the low cost resources and to provide export markets by concentrating production to a few locations. These are 'internationally integrated' investments (*Cantwell-Bellak*; 2000, p.122.) which involve the local affiliate in the international corporate group. As they state, resource based and internationally integrated investments are generally net trade-creating since they establish international connections by trade.

Those companies belong to the *market-oriented* investors, who invest in a country or into a region in order to supply these markets with their goods or services. The increase of the market or economic policy changes in the aimed country can promote the foreign company to invest. The aim of the investment is to preserve or gain market shares. The products made by the affiliate are sold in the local or regional market. These kind of investments can have both trade creating and trade replacing effects. Market-oriented firms may worsen the trade balance, if their exports are negligible and many of their inputs are imported. In principle, the size of the host country is likely to influence the trade strategy of foreign investors. Thus, big countries tend to be more suitable for market-oriented companies. Whereas small host countries appearing to be more suitable for export-oriented FDI because, apart from having a small domestic market, they use to have a higher degree of openness (ratio of trade to GDP) than large countries. The country-size is just one factor, however. Generally, the export propensity of foreign subsidiaries depends on a lot of industry and firm variables (see more details in *Rojec*; 2000).

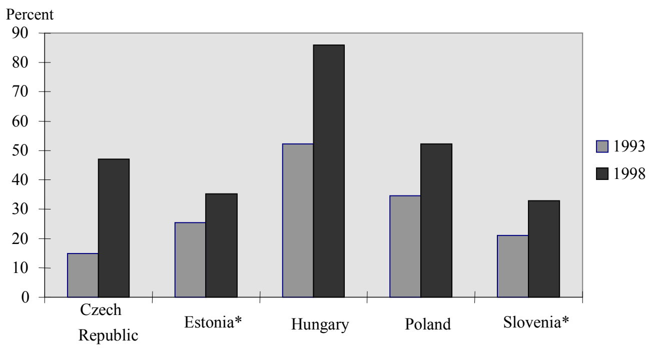
The foreign trade balance situation is and has been an important issue in the economic policy of the CEECs. Presumably the contribution of foreign investment firms to the trade balance in manufacturing has been different in the time being. In the first half of the nineties the import of foreign participation (in many cases greenfield) firms was very high in great part required by the building up of production capacities. Later on, however, the exports of these firms has become so significant that they exerted a positive effect on the foreign trade balance. Best examples for this phenomenon are the so

[12] The author differentiates among four main types: efficiency-seekers, resource-seekers, market seekers and strategic asset seekers.

called industrial free zones,[13] where majority or completely foreign owned firms, multinational affiliates were established. These firms produced USD 2091 million surplus in 1999, giving 43 percent of Hungarian exports and 30 percent of imports. The main direction of the exports is the EU. A considerable part of this trade consists of deliveries to other affiliates or to the mother company in EU countries (mainly Germany).

To the end of the nineties foreign capital has a considerable weight in the manufacturing sectors of the CEEC-5. Foreign penetration in the industry increased rapidly in all fields, let it be gross value-added, net sales revenue, investments, employment, etc. The increase has been really spectacular in the case of exports. In 1993 foreign investment firms were responsible for 15-50 percent of the manufacturing exports in the examined countries and in 1998 this increased to 33-86 percent (see Figure 4). Hungary had and still has by far the highest share of foreign investment enterprises (FIE) in exports. Slovenia and Estonia are similar with around 35 percent and Poland and the Czech Republic with around 50 percent.

Figure 4. Share of foreign investment enterprises in the manufacturing exports



* The base year is 1994 in the case of Slovenia and 1995 in the case of Estonia.
Source: WIIW database on foreign investment enterprises.

There are certain sectors where the share of FIEs is very high in all of the countries: these are the telecommunication equipment, office and electrical machinery (high-tech) and motor vehicle branches (medium-tech). In the case of low-tech products the share of FIEs in exports is generally lower than the average but there are cases of quite high FDI-participation (tobacco, paper).

The available evidence on how FDI affected trade specialisation patterns of Central European host countries is rather scarce. In the case of intra-industry trade, according to certain studies it seems that FDI increases the share of IIT. *Aturupane–Djankov–Hoekman* (1997) for example found on the basis of an econometric analysis of EU-CEEC trade, after controlling for country-specific factors, a positive and significant relationship between FDI and both vertical and horizontal intra-industry trade.

[13] Based on the XXIV/1988 law of foreign investment companies with foreign participation may establish their own industrial-free zones under the control of the customs authorities, within which they are regarded as foreigners for the purposes of exchange control and foreign trade. There are around 100 industrial-free zones spread throughout Hungary, the majority of which belongs to the machinery industry. The regulation of these zones will have to be changed after the accession of the country to the EU (see *Éltető*; 1998).

Table 3

The ten most important export products to the EU in 1998

(thousand euros and percentage of total exports)

Poland			Slovenia			Hungary			Czech Republic			Estonia		
SITC	euro	share	SITC	euro	share	SITC	euro	share	SITC	euro	share	SITC	euro	share
78 120	762 860	4.75	78 120	693 082	13.31	71 322	1 877 844	12.89	78 120	1 472 892	10.10	76 432	144 831	8.23
76 110	444 331	2.77	82 119	208 879	4.01	75 270	637 267	4.37	78 439	346 706	2.38	33 440	82 638	4.69
32 121	426 217	2.65	78 219	103 915	2.00	76 381	507 533	3.48	82 119	306 375	2.10	24 820	81 239	4.61
82 116	410 771	2.56	78 439	92 336	1.77	78 120	431 248	2.96	78 432	227 825	1.56	24 740	74 871	4.25
82 159	280 495	1.75	68 412	76 601	1.47	75 997	369 257	2.53	77 313	192 010	1.32	33 419	72 191	4.10
68 212	276 225	1.72	77 521	64 993	1.25	77 313	362 130	2.49	69 969	164 344	1.13	24 752	64 403	3.66
78 219	256 554	1.60	71 690	60 607	1.16	76 110	327 014	2.24	24 820	154 078	1.06	33 430	57 487	3.27
32 500	219 538	1.37	77 586	57 589	1.11	78 439	261 510	1.80	69 119	152 546	1.05	75 997	57 211	3.25
63 599	200 670	1.25	78 432	52 833	1.01	75 260	210 548	1.45	93 190	129 317	0.89	76 493	33 321	1.89
78 439	199 641	1.24	74 315	51 488	0.99	71 323	207 536	1.42	71 631	109 376	0.75	82 159	30 954	1.76
Total 10	3 477 302	21.66	Total 10	1 462 323	28.08	Total 10	5 191 887	35.64	Total 10	3255469	22.33	Total 10	699146	39.71

Name of SITC numbers: 24 740: coniferous wood in the rough 24 752: non-coniferous wood 24 820: wood of coniferous species 32 121: bituminous coal, not agglomerated 32 500: coke and semicoke of coal, of lignite 33 419: light oils from petroleum or bituminous minerals and products therefrom 33 430: gas oils, 33 440: fuel oils n.e.s. 63 599 manufactured articles of wood 68 212: refined copper 68 412: aluminium alloys, unwrought 69 119: metal structures and parts of iron and steel 69 969: articles of iron or steel n.e.s. 71 322: reciprocating piston engines of a cylinder capacity exceeding 1000 CC, 71 323: compression-ignition engines (diesel or semi diesel) for road vehicles 71 631: electric motors of an output exceeding 37.5W, 71 690: parts for use solely or principally with electric motors, generators 74 315: compressors used in refrigerating equipment 75 260: input–output units in data processing, 75 270: storage units for data processing 75 997: parts of automatic data processing machines, magnetic or optical readers 76 110: television receivers, colour or sound and video recorders, 76 381:video recording or reproducing apparatus 76 432: transmission apparatus for radiotelephony, radio broadcasting, incorporating reception apparatus 76 493: parts of television and radio broadcast receivers, transmission apparatus for radio telephony 76 499: parts of sound recorders and TV image and sound recorders or reproducers, 77 313: ignition and other wiring sets used in vehicles, 77 521: refrigerators, household type 77 586: electric ovens and cookers, cooking plates, grillers 77 821:filament lamps, 78120: motor vehicles for the transport of persons,78 219: motor vehicles for the transport of goods, 78 432: other parts and accessories of motor vehicle bodies 78 439: parts and accessories for motor vehicles 82 116: seats n.e.s. with wooden frames 82 119:parts of seats n.e.s. 82 159: furniture n.e.s. of wood 84 130: men’s jacket, woven, 84 140: men’s trousers, woven, 84 230:women’s jacket of woven textile, 84 270: blouses, shirts of woven textile, 85 148:footwear of leather, 85 190: parts of footwear 93 190: special transactions.

As it was spectaculous, the most important and radical change has been the gain of high-tech products in the exports and in the trade specialisation of the CEEC-5. Let us examine the single products traded. Table 3 shows the first ten export products of the CEEC-5 to the EU.

It can be seen that the most important export products are mainly telecommunication, electronic equipment, office machinery products (high tech) and several motor vehicles and components. In the case of Hungary the ‘first ten’ consists of only machinery and transport equipment products (SITC 7). These are the products which have caused the changes in the export structure, so behind the industries with spectacular development (like office machinery or motor vehicles) in reality there are a small number of products. Exports in the nineties experienced a high and increasing product-level concentration, which is also present at the company level.[14] These products are overwhelmingly produced by multinational affiliates like IBM, Phillips, Renault, Volkswagen, Elcoteq or other major foreign firms.

Keeping this in mind, one can presume that there is a correlation between the participation of FIEs in the exports of the manufacturing sectors and the export performance of the sectors. Table 3 shows a set of possible correlation coefficients. This coefficients relate the 1998 sectoral export-share of the FIEs and the sectoral export value quotient of 1998–1993. Such a way a positive correlation was found, although the extent of the correlation is not too strong and rather different according to countries. What is the reason of the fact that despite the ‘obvious’ determining role of foreign companies in exports (seen on the product level) strong correlation can hardly be detected on the sectoral level?

Correlation between FDI ‘penetration’
in exports and the export performance of the sectors

Country	Correlation coefficient*
Estonia	0.36
Hungary	0.35
Poland	0.56
Czech Republic	0.58
Slovenia	0.46

* Correlation coefficient: between the share of FIEs in the sectoral exports in 1998 and sectoral export value quotient in 1998/1993. Number of observed sectors is 18 (12 in Estonia).

The first reason is the lack of proper data. Available sectoral data on FIEs participation are too aggregated (ISIC 2 digit), they include sub-branches of different behaviour. The concept of Foreign Investment Enterprise is also too aggregated, it consists also of minority FIEs where the control is in domestic hands.

The second reason lies in the nature of foreign investments. As it was mentioned, CEECs experienced a mixture of export-oriented and domestic market oriented FDI inflow. In several sectors, like food, certain chemicals, paper and printing, minerals, etc. FDI penetration is significant (because of domestic sale possibilities, gain of strategic positions, etc.) but the level of exports and the increase during the nineties have been rather small as compared to other sectors.

[14] 74 percent of the total high-tech exports of Hungary originated from three companies (Phillips, IBM, GE) in 1997 and 61 percent was the respective number for 1998. (Calculations based on the economic weekly ‘Figyelő Top 200’ database).

Apart from these, in case of Hungary the relatively low correlation can also be explained by the fact that in certain sectors the share of FIEs in the exports (FDI penetration) and similarly the export 'quantity' was already relatively high in 1993, so the increase has been small up to 1998. (Examples can be the food-tobacco, and other transport equipment sectors.)

5. CONCLUSION

This paper observed the development of the five CEECs' foreign trade with the EU. Apart from the simple export share analysis in the product structure it also examined the specialisation patterns and the development of intra-industry trade. The main results can be summarised as follows.

1. Between 1993 and 1998 the foreign trade structure of the CEEC-5 with the EU went through on considerable changes. The share of the high-tech products increased significantly, first of all in Hungary and Estonia, but also in other countries. Parallel to this process the share of traditional, low technology sectors decreased. Similar development can be observed in the foreign trade specialisation. The former specialisation of the CEEC-5 on low-technology, labour intensive goods strongly lessened and in certain cases new specialisation patterns have been developed among the high-tech products. The export increase of the high-tech products has been accompanied by an import increase in most cases and it is a result of assembly-like, lower value-added activity especially in office machinery, telecommunication equipment and electronic products.

2. The level of intra-industry trade increased in the manufacturing sector in each countries. This increase was the most spectacular in Slovenia and Estonia and the level of IIT is the highest between the Czech Republic and the EU. The increase of IIT in several cases meant product quality improvement at the same time, as the results of the unit value calculations showed. The increase of IIT can be considered a positive phenomenon in the case of several high and medium-tech products. Regarding low-tech products, however, in the majority of the cases the increase of IIT stemmed from the worsening of the trade balance.

3. Foreign direct investments and the activity of foreign investment companies have became increasingly significant in the CEECs in the nineties. FIEs contributed to the restructuring of the economy and of the manufacturing sector. Certain branches have been similarly attractive for foreign investors in the five countries, where the affiliates of major multinational companies are present. After having a look on the foreign trade structure at a very detailed product level it turns out that those – high and medium-tech – products are mainly responsible for the previously described changes in the foreign trade which are produced by these foreign companies. Statistical calculations are uncertain because of the lack of properly detailed data, but they still show a positive correlation between the extent of foreign presence in the exports and the extent of the successful export performance. It can be concluded that FDI exerts a determinant effect on the external competitiveness of the CEECs. The extent of this effect is obviously different among the countries due to their different FDI-attracting policies in the past and due to their different production and location-specific advantage structures. In spite of that one can suggest that the external activity of the foreign investment firms in the CEEC-5 has already integrated these economies to the European Union.

4. As the foreign trade patterns of the CEEC-5 are quite sensitive to the functioning of a few multinational affiliates, the economic policy should focus on the stabilisation and further improvement of recent trade structure and external advantages. Regarding the extensive penetration of foreign capital in these economies, foreign investors should be taken into account on the long run. In this respect the increase of the domestic value added and the maintenance of the attractiveness for foreign investors can be a reasonable policy.

REFERENCES

ATURUPANE, CH. D. – HOEKMAN, S. B. (1997): *Determinants of intra-industry trade between East and West Europe*, World Bank Working Paper, No. 1850.

BALASSA, B. – NOLAND, M. (1987): The changing comparative advantage of Japan and the United States. In: BALASSA, B. (ed): *Nemzetközi kereskedelem és gazdasági növekedés*. Közgazdasági és Jogi Könyvkiadó, Budapest.

BLANES, V. – MARTÍN, C. (2000): The nature and causes of intra-industry trade: back to the comparative advantage explanation? The case of Spain. *Weltwirtschaftliches Archiv*, Vol. 136, No. 3. p. 423–441.

CANTWELL, J. – BELLAK, CH. (2000): The impact of foreign direct investment on international trade. *Journal of International Relations and Development*, Vol. 3. No. 2. p. 120–127.

DAVIS, D. R. (1995): Intra-industry trade: A Heckscher-Ohlin-Ricardo approach, *Journal of International Economics*, No. 39. p. 201–226.

DUNNING, J. H. (1993): *Multinational enterprises and the global Europe*. Addison-Wesley. London.

ÉLTETŐ, A. (1998): Economic policy background to foreign direct investment in Hungary, *WIIW Research Report.* Series No. 244, April.

ÉLTETŐ, A. (2000): Changes in specialisation and intra-industry trade and the effects of FDI: The cases of Hungary and Spain. *Journal of International Relations and Development*, Vol. 3. No. 2. p. 153–172.

FALVEY, R. (1981): Commercial policy and intra-industry trade. *Journal of International Economics*, No. 11. p. 495–511.

FONTAGNÉ, L. – FREUDENBERG, M. – ÜNAL-KESENCI, D. (1995): Régionalisation et échanges de biens intermediaires. *CEPII Document de Travail*, No. 95–11, Décembre.

FONTAGNÉ, L. – FREUDENBERG, M. (1997): *Intra-industry trade methodological issues reconsidered*. CEPII Document de Travail, No. 97–01.

GREENAWAY, D. – HINE, R. (1991): Intra-industry specialisation, trade expansion and adjustment in the European economic space. *Journal of Common Market Studies.* Vol. 29. No. 6. p. 603–622.

GREENAWAY, D. – MILNER, CH. (1994): Country specific factors and the pattern of horizontal and vertical intra-industry trade in the UK. *Weltwirtschaftliches Archiv*, No. 1. p. 76–97.

GRUBEL H. G. – LLOYD P. J. (1975): *Intra-industry trade: the theory and measurement of international trade in differentiated products*. Macmillan. London.

HATZICHRONOGLOU, TH. (1997): Revision of the high-technology sector and product classification, *STI Working Papers*. No. 2. OECD, Paris.

KOJIMA, K. (1975): International trade and foreign investment: substitutes or complements? *Hitotsubashi Journal of Economics*, Vol. 16. No. 1. p. 1–12.

KRUGMAN, P. (1990): *Rethinking the causes of international trade*. MIT Press, Cambridge.

KRUGMAN, P. (1991): *Geography and trade*. MIT Press, Cambridge.

MARKUSEN, J. R. (1983): Factor movements and commodity trade as complements. *Journal of International Economics*, No. 14. p. 341–356.

MUNDELL, R. A. (1957): International trade and factor mobility. *American Economic Review*, Vol. 47. p. 321–335.

OECD (1993): *Industrial policy in OECD countries*. Annual Review, OECD, Paris.

ROJEC, M. (2000): Subsidiary exports: a conceptual framework and empirical evidence. *Journal of International Relations and Development*, Vol. 3. No. 2. p. 128–152.

TÖRÖK Á. – PETZ R. (1999): Kísérlet a K+F-intenzitás és az exportszerkezet közötti összefüggések vizsgálatára a magyar gazdaságban. *Közgazdasági Szemle*, Vol. 46. No. 3. p. 213–230.

VENABLES, A. J. (1996): Equilibrium locations of vertically linked industries. *International Economic Review*, No. 37. p. 341–359.

VERNON, R. (1966): International investment and international trade in the product cycle. *Quarterly Journal of Economics.* No. 80. p. 190–210.

VONA, S. (1991): On the measurement of intra-industry trade: some further thoughts. *Weltwirtschaftliches Archiv*, No. 4. p. 678–700.

ARMS TRANSFERS OF THE POST COLD WAR PERIOD, 1987–1997*

ERVIN R. FRIGYES[1]

In this paper the main methodological problems and the characteristic trends of international arms transfers, are exposed.

It has been documented in a quantitative way, that by the end of the eighties the Soviet Union (Russia) lost both its economic ability and political intention of arming loyal third world countries for sake of influence. This led to the decrease of total arms transfers to the developing countries and therefore to the decrease of world total arms transfers. Arms transfers to the developed countries did not show the same systematic changes. The United States has a leading position in world total arms transfers and especially in transfers to the developed world. And there formed a western dominance in respect of deliveries to the developing countries

However the article is focused on a very decisive period 1987–1997, in the final part it calls attention to the imminent problems caused by the compelling export drive of the still vigorous Russian (and Ukrainian) arms industry faced with the lack of the purchasing power of domestic armed forces. This situation could lead to the destabilization of the present world military situation.

KEYWORDS: Arms transfers; International economy; International comparisons.

Arms transfers are a very special kind of foreign trade. The main motives are not restricted to economic gains. Seeking prestige and influence has a very important role to play. The first part of the article outlines the main methodological problems of the reliability (or completeness) of data, the consideration of dual goods, the discontinuity of time series for Eastern Europe, the composition of main groups of countries and with the estimation of current and constant price USD values. The second part explains the main motives and characteristics of international arms transfers. The mutually interwoven issues of illicit arms trade and the weapons of mass destruction of recent times

* The author is indebted to the Bureau of Verification and Compliance of the United States Department of State. To *Mr. Daniel Gallik*, editor of World Military Expenditures and Arms Transfers (WMEAT) 1998 for giving permission to use the invaluable information of WMEAT98 for his researches, lectures and publication and, nevertheless for mailing him the printed version of that document. He also wants to express his gratitude to the Council of Livable World Educational Fund and personally to *Miss Kate Käufer* for her valuable information. It should be noted that the views and analyses in this article are of the author's responsibility.

[1] Retired associated professor, at Miklós Zrinyi University of National Defense, lecturer on Department of International Military Economics.

are small and light arms. The third part deals with the main characteristics of the changes in arms deliveries for the years of transition from the cold war to present world political and power relations (1987–1997). Supported with several figures and tables a concise analysis is attempted to demonstrate the processes of this decisive period. The fourth part tries to point out the problems and the possible dangers associated with the foreseeable arms exports drive of the major Soviet successor states. Russian (and Ukrainian) arms industry still exist and R&D is still going on. At the same time, domestic armed forces cannot purchase new weapons, thus arms industry is compelled to intensify exports. There is an imminent danger that these weapons could find their ways to rogue countries.

1. SOURCES AND BASIC METHODOLOGICAL PROBLEMS.

In the era of cold war it was quite impossible to collect comprehensive worldwide data on this subject. Now the researcher faces another problem: the abundance of (sometimes incongruent) data. The main sources of statistical analysis now are provided mostly on the internet. Several institutions and non-government organizations (NGOs) publish data and analyses of military expenditures and arms trade on the World–Wide Web. This flow of information includes clippings from the world press.

In respect of basic statistical data this article practically relies upon one single source: World Military Expenditure and Arms Transfers 1987–1997 by the Bureau of Verification and Compliance of the United States Department of State (WMEAT98). WMEAT98 comprises data of 172 countries and information from practically all reliable resources.

This period on which this article is focused (1987–1997) was of very radical changes in the international power structure and this had its effect on the arms trade flows too. Analyzing these changes one is faced with very complicated problems related to statistical methodology.

– The data refer to transfers to armies and law enforcement agencies of legal governments and do not include those items provided by the opposing great powers to 'liberation movements' (as they were called by the providers) or those sent to 'rebel groups' (as they were called by their adversaries). For the beginning of the decade these transfers were – usually – government grants. Soviet arms sent to the PLO, US arms to the Nicaraguan 'Contras' and US and South African arms for the UNITA rebel group in Angola. In the nineties the non-government recipients of arms became supplied mostly by arms fixers. The bills used to be settled either by sympathizer countries (e.g. Libya, Liberia and Burkina Faso) or by special organizations (as in case of UNITA and the RUF – 'Revolutionary United Front' of Sierra Leone).

– One can hardly believe that there is a generally valid uniform treatment of the dual–use articles (jeeps, trucks, light aircrafts, electronics, and explosives.)

– The changes in the international power structure make the comparison of the data of the initial and final years very difficult. Political entities (the Soviet Union, the German Democratic Republic, Czechoslovakia and Yugoslavia, and the Warsaw Pact) ceased to exist and new entities emerged. There are no data for these new entities for the period before the separation and in some cases for the year of separation there are no data at all. For a more or less reliable comparison one of these old entities survived in the statistical tables of WMEAT97 (and some entities in this article too), such as the countries of (former) Warsaw Pact and the successor states of the former Soviet Union.

– The German unification makes further problems. Until 1991 the Eastern provinces of Germany were considered (as they were) a partner country of the Warsaw Pact and a part of the Eastern block. In the latter years they belong to the European part of the NATO and to Western Europe. This (perhaps only feasible) ‘actual state’ solution expresses well the post cold war changes of power relations but, at the same time, makes less comparable the time series of the former Warsaw Pact countries.

– There are other problems with the aggregate groups of the countries. This holds true for the ‘unorthodox’ treatment of European geographical areas as well. According to WMEAT98 Western Europe consists of the European NATO countries (of that time) plus Austria, Finland, Ireland, Malta, Sweden and Switzerland. NATO Europe (and this way Western Europe as well) includes Greece, on the southern tip of the Balkan Peninsula, and Turkey, with 97 percent of her territory in Asia.

There are problems with the two main groups of countries (developed and developing ones) as well. According to WMEAT98 classification the 33 country developed group includes Western Europe – except Malta and Turkey –, Israel, South Africa Japan, Singapore, South Korea, Taiwan, Australia and New Zealand. As regards Eastern Europe the Visegrád Countries (the Czech Republic, Hungary, Poland and Slovakia and from the former Soviet Union only the Russian Federation) belong (in this edition) to the developed group. The rest of Eastern European countries were considered as developing ones. This solution seems to be biased to Slovenia and Estonia with higher per capita GNP and much better living conditions than Russia.

For the necessity of using comparable aggregates in some tables even VMEAT98 uses a no longer existing group: the (former) Warsaw Pact. Since the time series for the Soviet Union ended in 1990 for a better understanding of the provider’s structure and the declination of Eastern European arms deliveries the author composed a group for the latter years: ‘Successor States of the Soviet Union’. The forming of this group has no political aspects. Soviet arms industry was widely scattered in most former Soviet republics. This way Soviet time series until 1990 cannot be compared with Russian time series from 1991 on.

– There are problems with the (current and constant price) dollar estimates too. Current price USD estimates for all countries were estimated in three steps:

a) In the first step the constant (1997) price values were estimated in national currencies.

b) In the second step we estimated the values in (1997) constant USD terms,

c) Finally the values in current USD were estimated.

Thus in WMEAT98 current price USD estimates were derived from constant (1997) dollar estimates. The numerical example for the treatment of price problems in the Statistical Notes of WMEAT98 (p. 207) can be formulated the following way.

a) Estimation of constant (1997) price values in national currencies

Let indicate the original starting value $v_i^t(curr_i)$ the value for the country *i* in national currencies for the year *t*.

$$w_i^t(const_i, 1997_i) = v_i^t(curr_i)\left(Defl_i^{t,1997}\right)^{-1}, \qquad /1/$$

where $w_i^t(const_i, 1997_i)$ is the value in constant 1997 prices for the country *i* in the year *t* in national currency, and $\left(Defl_i^{t,1997}\right)$ is the implicit GNP price deflator for the country *i* for the interval (*t*….1997).

b) Estimation of constant (1997) USD values

$$w_i^t(\$) = w_i^t(const_i, 1997_i)cr(\$, curr_i)^{(1997)}, \qquad /2/$$

where $w_i^t(\$)$ is the value for the country *i* in year *t* in constant 1997 US dollars, and $cr(\$, curr_i)^{(1997)}$ is the currency rate (conversion base) for the national currency of country *i* to USD in 1997[2] , and finally:

c) Estimation of current price USD values

$$v_i^t(\$, 1997_i) = w_i^t(\$)\left(Defl_{US}^{t,1997}\right), \qquad /3/$$

where $v_i^t(\$, 1997_i)$ is the estimated value for the country *i* in the year *t* in current USD, and $\left(Defl_{US}^{t,1997}\right)$ is the implicite US GNP price deflator of the United States for the interval (*t*….1997).

Current price estimates were determined throughout the whole report by US price deflators. Since there were no data available on price indices or PPP[3]-s for the arms deliveries to and from individual countries there was no other way to approach current value data. This method and the 30.7 percent overall price increase between 1987 and 1997 seems to be a low estimation for the articles of arms trade.

Arms industry is probably extremely exposed to increases in factor costs, and at the same time the increasing share of R&D costs could increase prices more than in other sectors of the national economy. Nevertheless, arms industry has quite enough bargaining power to obtain reasonable prices.

[2] For the countries of former Warsaw Pact and Yugoslavia national currencies were converted into dollars on the basis of Purchasing Power Parity (PPP) estimates of the World Bank's World Development Indicators. This equals to the supposition that military expenditures and arms transfers have the same purchasing power parities as the rest of the goods and services.

[3] In WMEAT98 there is a reference to a UN research group studying future availability of military PPPs but , 'practical prospects for the future availability of usable military PPPs are poor due to the lack of underlying national data, especially on military prices'. (*WMEAT98*; 1998. p. 208)

Due to the described methods the use of constant or current prices does not affect the proportions among countries or groups and countries. At the same time, the use of constant 1997 prices gives hints on the considerable changes in the volume of international arms trade. WMEAT 98 uses the following (overall) price deflators for estimating constant 1997 price values and these deflators were used throughout all data of the document.

US Implicit price deflators for estimating current price values

Year	Deflator	Year	Deflator
1987	76.54	1992	90.20
1988	79.12	1993	92.61
1989	82.17	1994	94.56
1990	85.59	1995	96.58
1991	88.28	1996	98.36

This treatment of currencies and price changes, no doubt, provokes most statisticians for a professional criticism. But, at the same time, in lack of necessary detailed information the authors of WMEAT98 had the sole alternative: either do their work this way or do nothing. Fortunately they opted for doing this way.

The figures in this article are based on constant 1997 price data. (Data resulted in the second step of estimations.) Since the relation between constant and current prices is the same throughout the data, structural relations (relative frequencies) are also the same on both price levels. The figures of 1–3 are demonstrating the arms transfers by main (groups of) suppliers to the developed countries and to the developing ones. The initial data in WMEAT98 were in current prices. For transforming them to constant price data the author used the price deflators of columnar-setting to transform (back) the data to constant (1997) price values.

Our data refer to legal trade – to be more exact – to legal deliveries only. (In arms trade there is a distinction between agreements and deliveries. Agreements usually do not mean deliveries in the same year. Deliveries are de facto transactions.)

Legal trade covers the overwhelming part of world arms transfers. Not in quantities but in regard of consequences legal trade represents the tip of the iceberg only. The post cold war Limited Intensity Conflicts or Low Intensity Conflicts (LIC-s) used to be fuelled by illegal (and therefore unregistered) arms trade. It is very characteristic that the keyword: ‘arms fixers’ produces hundreds of items on an internet browser. The illegal trafficking of private arms fixers represents just one type of the clandestine transfers. There are states (the ‘rogue states’) sending arms for their clients involved in interstate collisions or to subversive groups.

2. THE MAIN MOTIVES AND CHARACTERISTICS OF INTERNATIONAL ARMS TRANSFERS

From mere economic points of view arms transfers – like any other type of the international trade flows increases the export revenues of the provider country. This is not true, however, for the arms exports to most of the developing countries in the cold war era. These deliveries were either donations or subjects to hardly recoverable loans.

From the importer's side arms imports represented a necessary burden or (in some cases) the exorbitant ambitions of the rulers of the countries. Table 1 and Table 2 give some information on the relative importance of arms trade in the total foreign trade for some (groups of) countries. From the side of arms producing companies arms exports increase both profits and economics of scale. For governments of the exporting countries there are other considerations too.

From points of view of the recipient countries (especially of the countries of the Third World) arms imports are parts of the 'military burdens'. The great majority of those countries have chronic and acute problems of the current account of the balance of payment and international debts. Arms imports are exhausting their (never sufficient) foreign exchange revenues.

But, it should be emphasized, arms imports provide the unique way for small and less developed countries to maintain their defense capabilities. At the same time, the international division of labor in arms producing has its benefits for the most developed countries too.

It is a very important indicator of the military balance of any country the relative importance of arms trade in their imports and exports. The data for relative export and import participations of arm transfers for the years 1987, 1991, 1994 and 1997 could be seen in Table 1 for the main aggregates and for the United States.

The intensity of arms transfers in imports decreased in the period 1987–1997. This indicator was and (in spite of substantial decrease) remained the biggest for the developing countries. There was a big, however decreasing intensity in case of the OPEC members, and there is a characteristic decrease in the countries of former Warsaw Pact.

Table 1

The intensity of arms transfers in foreign trade

Main aggregates	Arms imports (percent of total imports)				Arms exports (percent of total exports)			
	1987	1990	1994	1997	1987	1990	1994	1997
World total	2.5	1.4	1.0	1.0	2.5	1.6	1.0	1.0
Developed countries	0.9	0.6	0.6	0.6	2.8	1.8	1.1	1.2
Developing countries	11.5	6.5	2.6	2.5	1.2	0.4	0.2	0.2
OECD	0.6	0.5	0.5	0.4	1.7	1.4	1.2	1.3
OPEC	19.8	13.2	8.5	10.3	0.1	0.1	0.2	0.0
NATO	0.5	0.5	0.4	0.3	2.2	1.8	1.5	1.7
USA	0.2	0.3	0.2	0.2	7.0	5.6	4.3	4.6
NATO Europe	0.7	0.5	0.5	0.4	1.1	0.9	0.7	0.8
Warsaw pact	2.5	1.3	0.4	0.3	13.4	9.0	1.6	1.7

Besides the main aggregates it is worthwhile to see the list of the countries with the significant import and export intensities. The highest arms import intensities are observed in five countries of Sub-Saharan Africa and seven Arab-countries, five OPEC countries, one Balkan country, Taiwan, Pakistan, Israel, Myanmar, one NATO-country (Turkey) and one ex-Soviet republic (Kazakhstan).

As regards export intensities the top twenty countries still consists of ten former communist countries (including Serbia and Montenegro and two countries still with communist regimes). The United States, the United Kingdom and France represent the NATO.

The ranking in Table 2 do not indicate the importance of the countries in the international arms transfers. At the same time these data indicate well the role of arms transfers in the national economies of the countries in case.

Table 2

Countries with the highest intensity of arms transfers in their foreign trade in 1997

Rank	Top importer countries	Arms imports (percent of total imports)	Rank	Top exporter countries	Arms exports (percent of total exports)
1	Saudi Arabia	40.4	1	Korea North	8.1
2	Kuwait	24.3	2	Moldova	7.1
3	Burundi	16.5	3	Belarus	6.7
4	Qatar	14.3	4	United States	4.6
5	Myanmar	13.6	5	Ukraine	3.5
6	Egypt	12.1	6	Russia	2.6
7	The Gambia	11.9	7	United Kingdom	2.3
8	Taiwan	8.1	8	Armenia	2.2
9	Rwanda	6.7	9	France	2.0
10	Bosnia and Herzegovina	6.5	10	Uzbekistan	1.7
11	Iran	5.8	11	Serbia and Montenegro	1.7
12	Algeria	5.7	12	Israel	1.6
13	Yemen	5.5	13	Azerbaijan	1.3
14	Pakistan	5.2	14	South Africa	1.2
15	United Arab Emirates	4.7	15	Sweden	1.1
16	Guinea	3.7	16	Swaziland	1.0
17	Israel	3.6	17	China	0.6
18	Angola	3.5	18	Spain	0.5
19	Turkey	3.3	19	Slovakia	0.5
20	Kazakhstan	3.3	20	Czech Republic	0.4

Source: WMEAT98.

Arms exports, or even the use of arms is subject of unilateral or multilateral declared controls including prohibitions and embargoes. These embargoes, however, did not work properly. Perhaps the oldest example is the prohibition by the Lateran Council of 1139 the use against Christians a deadly weapon – the crossbow. (All the same, crossbows were used frequently until firearms replaced them.)

There are effectual restrictive and embargo agreements for the proliferation of ABC (Atomic, Biological and Chemical) weapons and (in very critical stage) for the ABM (Anti-Ballistic Missiles) systems. At the same time, the UN Conference in 9 June 2000, did not lead to breakthrough on the problems of small and light arms. It should be emphasized that small and light arms are the weapons of mass destruction in our time.

The UN secretariat estimates that small arms were the weapons of choice in 46 of the 49 major conflicts in the 1990s which caused over 4 million deaths, 90 percent of which were civilians, and 80 percent of those were women and children. According to the United Nations Children's Fund for the ease of use and worldwide availability, small arms have helped create more than 300 000 child soldiers.

The problem of land mines should be mentioned here. According to estimations there are 87 to 112 million landmines laid world-wide. There are 69 countries affected by landmines and 100 millions of landmines stockpiled. There are approximately 100 companies in 55 countries producing landmines. According to estimations 2–5 million landmines planted, and only 100 000 removed annually.[4] According to Canadian government sources since 1975 one million casualties (mostly civilians) were caused by landmines and there are 250 000 landmine amputees worldwide. It should be emphasized, that (with the exception of the successor republics of the former Yugoslavia) mines in the most affected countries came from abroad.

In the future de-mining could be a very important item of the exports of military services. The problem is that mine affected countries are not solvent ones. Thus the future solution of the mine crisis depends on international solidarity.

3. THE MAIN CHARACTERISTICS OF THE CHANGES IN ARMS DELIVERIES 1987–1997

Table 3 summarizes the main tendencies prevailing in course of the decade 1987–1997. The following interwoven tendencies could be observed:

– The halving of arms transfers to developing countries.
– The increase of arms transfers from the NATO countries and, first of all, from the United States.
– The landslide like decrease of Eastern European arms deliveries.
– The moderate increase of arms transfers to the developed countries.
– The overall substantial (33 percent) decrease of the world's arms deliveries.

Table 3

The main changes in arms deliveries 1987–1997

Main Supplier Groups	Millions of constant (1997) USD		Relative frequencies (percent)		1997 index: 1987=100
	1987	1997	1987	1997	
NATO	38 000	47 500	46.6	87.1	125.0
United States	23 390	31 800	28.7	58.3	136.0
Other NATO-countries	14 610	15 700	17.9	28.8	107.5
Eastern Europe	35 940	3 685	44.1	6.8	10.3
From Eastern Europe					
The Soviet Union*	30 180	3 360	37.0	6.2	11.1
Russia	-	2 300	-	4.2	-
Rest of the former Warsaw Pact**	4 926	325	6.0	0.1	0.7
Rest of the World	7 540	3 365	9.3	6.2	48.0
Main Recipient Groups					
Developed Countries	24 140	26 240	29.6	48.1	108.7
Developing Countries	57 530	28 320	70.4	51.9	49.2
Arms Transfers World Total	*81 480*	*54 550*	*100.0*	*100.0*	*66.9*

* For 1997 the successor states of the former Soviet Union.
** Without The Soviet Union and her successor states and East Germany.

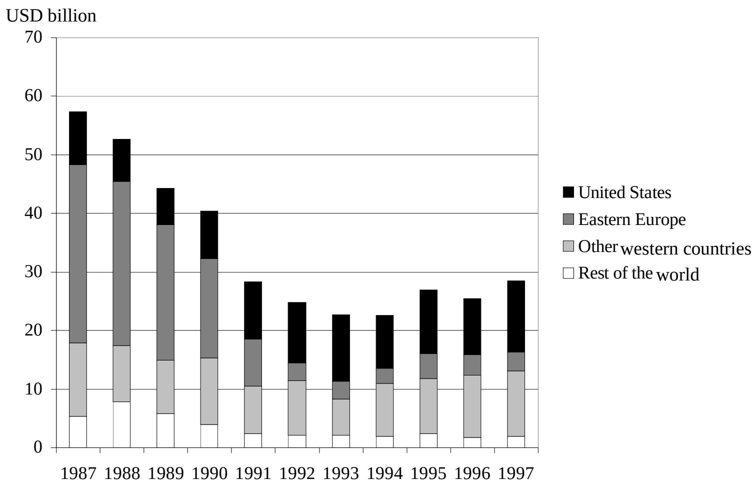
[4] See: www.mines.gc.ca/I_C-e.asp and Land Mine Monitor: www.icbl.org/lm/2000.

In Table 3 we tried to give a bit detailed view about the arms transfers from Eastern Europe. Here the category 'Former Warsaw Pact' as well as in throughout WMEAT97 has no political meaning for the present. The summarizing of the data of the successor states of the former Soviet Union has neither the meaning of a political entity. Only this way can be produced a collection of data describing the process in more depth.

These processes described in Table 3 were interdependent ones. Arms production in the former Soviet Union was a key industry of competitive products. At the same time (besides domestic use and shipments to Warsaw Pact partners) developing countries represented the demand side of the market. And this market did not offered solvent demands. The solvent part of the developing countries (Saudi Arabia, and the Gulf States) bought and buy western, mostly American equipment. The Soviet Union exported weapons by loans or aids for geopolitical purposes. And, as a consequence of this policy in 1997 Libya owed 4 billion USD (2 billion of this sum were remitted) and Syria 10 billion USD, Iraq 7 billion USD to the Soviet Union. Thus the continuation of the arms deliveries could not improve, but would deteriorate the current account of Russia's balance of payments.

Figure 1 describes the dynamic and structural changes in arms transfers to the developing countries.

Figure 1. Arms deliveries to developing countries by main groups of suppliers
(Billions of constant (1997) USD)



Arms deliveries to developed countries practically did not decrease until 1994 (see Figure 2). The decrease of the total sum could be attributed to the drop of Soviet (Russian) deliveries to the other former Warsaw Pact countries which were taken at that time unequivocally as developed ones. The western part of the world and mainly the United States kept its market positions until a transitional decrease in 1994–1995. 1996 was the year of recovery and 1997 was of a rapid increase to the highest values of the total period.

Total arms deliveries formed resultants of the changes on the two (developing and developed) aggregate recipient markets. For the predominance of deliveries to developing countries the dynamic tendencies of total arms deliveries offer the same sub-periods (1987–1990, 1990–1994 and 1994–1997).

Figure 2. Arms deliveries to developed countries by main groups of suppliers
(Billions of constant (1987) USD)

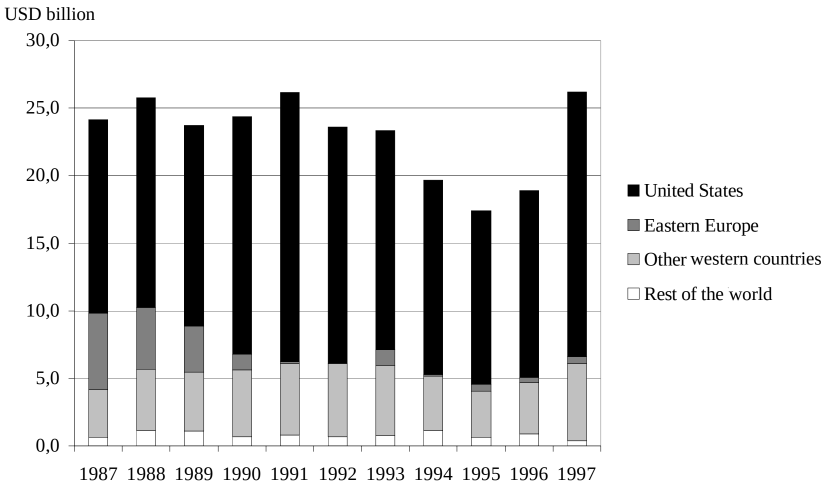
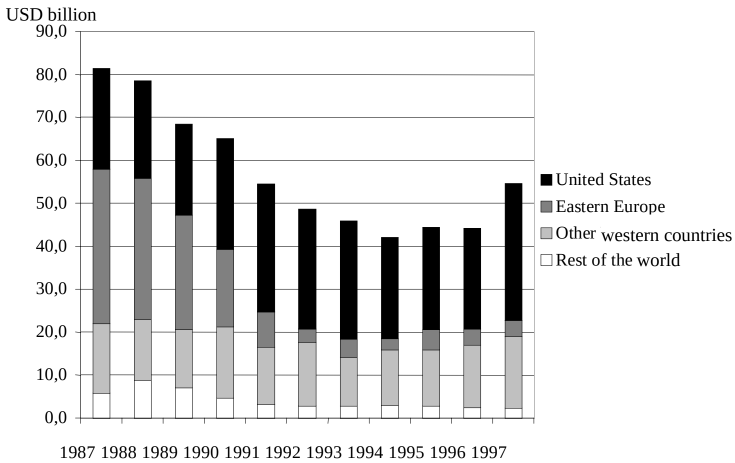


Figure 3. Arms deliveries by main groups of suppliers
(Billions of constant (1997) USD. World total)



For a more detailed description Table 4 provides information about the (annual average) dynamic changes of these periods for both exports and imports detailed by main aggregates, regions and organizations.

It should be mentioned that Eastern European arms deliveries were not restricted to obsolete weapons for 'the native's infantry' as much people could be inclined to believe it. Eastern European, especially the Soviet arms industry was much more developed as other branches of industry of the region. It is enough to have a look at the delivery data for some major weapons.

Table 4

*Percent changes in arms transfers**
(annual average of the periods)

Main aggregates	Arms imports			Arms exports		
	1987–1990	1990–1994	1994–1997	1987–1990	1990–1994	1994–1997
World total	-7.3	-10.2	9.0	-7.3	-10.2	9.0
Developed countries	0.2	-5.1	10.1	-6.7	-9.9	8.9
Developing countries	-10.8	-13.8	7.9	-14.5	-15.7	12.8
Regions:						
Central America and Caribbean	-16.7	-51.0	-16.3	-	-	-
North America NAFTA	18.9	-7.1	1.0	3.6	-2.3	10.4
South America	-17.6	-3.8	24.9	-52.9	15.9	-48.7
Africa, all	-26.9	-16.4	-9.6	19.0	6.3	14.4
North Africa	-21.1	-24.3	22.8	-8.4	-	-
Central Africa	-24.4	-30.3	19.2	-	-	-
Southern Africa	-32.1	-4.9	-46.2	65.1	21.4	14.4
Middle East	-11.4	-8.6	9.7	-10.0	5.0	-27.0
East Asia	-6.2	1.9	23.3	0.7	-26.2	12.1
South Asia	6.6	-44.3	16.0	65.6	-8.1	28.9
Europe, all	-3.7	-11.4	1.5	-13.2	-18.3	9.6
Western Europe	5.9	-7.1	-0.5	-0.1	-5.6	8.9
Eastern Europe	-23.0	-41.3	35.8	-20.4	-38.6	12.6
Oceanic	5.5	-1.7	-10.0	-12.6	42.2	-52.8
Organizations:						
OECD	8.0	-5.1	-0.5	2.0	-3.4	9.6
OPEC	-7.5	-10.5	14.9	-5.1	25.7	-32.4
NATO	8.1	-7.7	-0.7	2.6	-3.7	10.5
NATO Europe	6.4	-7.7	-1.5	0.9	-6.5	10.7
Warsaw pact (former members of)	-19.9	-34.7	14.8	-20.4	-38.3	13.0

* Calculated by geometric means of the indices between the years t_1 and t_2 .
Source: WMEAT98.

Table 5

US and Eastern European deliveries of selected major weapons

Years	World total	US deliveries	Eastern European deliveries	US share (percent)	East European share (percent)
			Tanks		
1986–1988	6 481	1 051	4 490	16.2	44.6
1989–1991	5 199	699	3 600	13.4	58.9
1992–1994	3 333	643	1 120	19.3	21.6
1995–1997	3 045	1 155	780	37.9	16.4

(Continued on the next page.)

(Continuation.)

Years	World total	US deliveries	Eastern European deliveries	US share (percent)	East European share (percent)
			Major surface combatant naval craft*		
1986–1988	47	0	29	0.0	46.8
1989–1991	49	0	19	0.0	38.8
1992–1994	51	2	16	3.9	17.6
1995–1997	85	13	25	15.3	15.3
			Supersonic combat aircraft		
1986–1988	1 556	436	680	28.0	43.7
1989–1991	1 630	340	550	20.9	33.7
1992–1994	680	350	70	51.5	10.3
1995–1997	750	310	140	41.3	17.3
			Helicopters		
1986–1988	1 354	104	860	7.7	43.6
1989–1991	929	159	420	17.1	39.8
1992–1994	681	191	230	28.0	29.4
1995–1997	783	253	290	32.3	34.5

* The deliveries of major combatant naval crafts are not related to newly built ships, but to the sales of old vessels due to the lack of funds of the Russian Navy. Russia is economically unable to maintain her fleet. In the world press there are very often news of selling out of active service Russian battleships to India.

4. SOME SHADOWS OF THE NEAR FUTURE

The time series of WMEAT98 ends at 1997 and so does our short quantitative analysis. The chief aim of this article, however, is to analyse macro-statistical facts, there is quite a great temptation to point out some tendencies which could be decisive ones for the near future. One cannot depend on official statistical data on the most recent period. Therefore we should resort on deductions from the information provided by news agencies and on reliable news of the world press.

It seems to be quite obvious that there is no sound hope for further shrinking of the arms market. And, it is quite possible that Russian (and Ukrainian) arms industry would regain – at least a part of – its position on the world market. The motives would be less prestige, or influence oriented, but the insatiable needs for convertible currencies. (Prestige and influence could be taken as valuable fringe benefits only.)

In the past arms production was the leading (and possibly the only up-to-date) branch of the Soviet economy. Most of the structural problems and the technological lagging back of the Soviet economy came from the ‘concrete wall’ between military and civilian industries. The overzealous protection of the achievements in military technology resulted in the lack of any ‘trickling down’ effect. Thus the technological innovations of the arms industry could not find their ways to boost other branches of the Soviet national economy.

The most of the factories, however broken down, still exist. A considerable part of the scientists and technicians resisted the temptations of better life and still remained home. The Russian school system is still able to provide new generations capable for R&D. One cannot underestimate that fact, that as a contrast to the cold war period, computers become household items in Russia too and the world computer community learnt to fear of Russian hackers and virus makers. (A valuable asset of present and future market of military goods and services.)

The decade 1987–1997 was characterized by the decline and the ceasing of the Soviet Union. At the same time, Russia and Ukraine (the Soviet successor states having the bulk of the armament producing capacities of the former Soviet Union) are in crucial economic and financial situation. And the greatest problem comes from the grave contradiction of the post – soviet military economics. On the one hand there is an existing and expansive – hungry supply potential of the arms industry and, on the other hand, there are the rigid budget constraints of domestic armed forces. As a result in the successor countries of the Soviet Union (especially in Russia and in Ukraine) there is an insoluble gap between the production and R&D capabilities of the arms producers and the miserable purchasing power of the armed forces. Thus Russian (and Ukrainian) arms producers are forced to exports. The government(s) are forced to export the surpluses of the inventories of armed forces too. This situation could bring unforeseeable instabilities not only in the international market of armaments but also in world power relations

According to the report of the *Independent* (17 August 2001.) on the traditional Zhukowsky air show had its debut the new impressing SU-39-‘Berkut’ stealth fighter. But, this precious new fighter has been made in one single copy. The cash strapped Russian defense budget cannot afford to order it. The same holds true for the SU-35 supersonic fighter with variable thrust engines able to hang motionless in air. This aircraft has been offered to South Korea and some Latin American countries. Russia simply cannot order any of them.

There is a quite imminent danger that Russian military export – especially in fields of air defense electronics – would reach the ‘rogue countries’ too. One could feel this imminent danger from the words of *Vladimir Yarashenko*, an air defence expert with Rosoboronexport, who said on the same air show that Iraq and Yugoslavia need not have lost their wars against Western forces in the past decade, if only they had upgraded their defences in time.

Russian arms exports valued 2.8 billion USD in 1999 and will reach 3.2 billion USD this year. An official of the Russian arms exports giant Rosoboronoexport said that they have orders of 13 billion USD and 70 percent of it consists of combat aircraft.

One should listen (with reservations) to an agnostic opinion too. The same article cites *Pavel Felgenhauer*, an independent defense analyst, who said: ‘It’s a garage sale. The whole former arsenal of the Soviet Union is on the block, and the goal is to make money.’ ‘There is stuff from the eighties that’s has been pulled out of old Soviet warehouses. The truth is that Russia’s military industry is incapable any longer of building complete weapons systems. They may sign contracts and take the money, but they cannot deliver’.

For lack of detailed information this article may neither support nor challenge this sceptical opinion. But the modernity of a weapon system cannot be measured simply by calendar years. The best selling fighter plane of most recent times is, no doubt, the F-16 ‘flying falcon’. The F-16A single seat model first flew in December 1976. The first operational F-16A was delivered in January 1979. And, it could be taken for granted, those planes in active service are not the same as those in the first series. The innovations are hidden under the fuselage. Arms exporters of Soviet successor states have the ambition to find markets in the NATO countries too.

*

The writing of this paper was completed at the end of August 2001. At the time of final edition (mid October) it would be very hard, if not impossible, to draw the consequences of the horrifying events of the 11th September 2001. It seems to be that a tighter control could be expected on the illegal arms transfers of small and light arms. At the same time some revision on legal transfers would be on the agenda of several states (e. g. the Slovak government starts a general audit on arms transfers from 1993 on).

On the other hand the imperative need for allies against world terrorism resulted in waiving some restrictions on legal transfers. The most important is that on the 4th October 2001. the Foreign Relations Committee of the US Senate removed the remaining sanctions on U.S. arms exports and military assistance to India and Pakistan. New – and till now unimaginable – alliances are being formed. The bipolar world is definitively over.

Using the cosmological terminology the 11th September 2001 could be taken as a point of singularity in the time-space of international relations, generally and especially of international military relations. We are waiting what kind of time-space continuum would emerge from this black hole.

REFERENCES

COHEN, A. (1997): *The 'Primakov Doctrine': Russia's zero-sum game with the United States*. The Heritage Foundation FYI No. 167. www.heritage.org/library/categories/forpol/fyi167.html

CALHOUN, H. L. (1998): Small arms, light weapons: can they be controlled. In: *World military expenditures and arms transfers*. WMEAT. p. 53–59.)

EASTERN Europe's arsenal on the loose: Managing light weapons flows to conflict zones (1998): In: *Basic papers. Occasional papers on International Security Issues*. No. 26.

GOVERNMENT of Canada Department of Foreign Affairs and International Trade: *Safelane.* www.mines.gc.ca/I_C-e.asp

Hidden Killers (1998). United States, Department of State.

Landmine Monitor Report (2000). International Campaign to Ban Landmines. www.icbl.org

Publications of the Stockholm International Peace Research Institute (1999). www.sipri.se.

The government finance yearbook. International Monetary Fund. www.imf.org

The military balance. International Institute of Strategic Studies. London. www.iss.org/scripts/index.asp

The world factbook. Central Intelligence Agency. www.cia.gov./publications/factbook/index.html

WOOD, B. – PELEN, J. (1999): *The arms fixers. Controlling the brokers and shipping agents*. The Norwegian Initiative on Small Arms Transfers (NISAT).

World military expenditures and arms transfers (1998). Department of State Publication. Washington D. C. www.state.gov/global/arms/bureauvc.html

VARIANCE ESTIMATION FOR STRATIFIED SAMPLES WITH ONE UNIT PER STRATUM

LÁSZLÓ MIHÁLYFFY[1]

A special case of stratified samples is considered where each stratum has the same number of units and from each stratum, one unit is selected in the sample with simple random sampling (SRS). The usual SRS estimator for the variance of a mean is biased under this design, and the size of the bias is estimated in this paper. The problem is related to systematic sampling.

KEYWORDS: Simple random sampling; Stratified samples; Variance estimation.

In this study a special kind of stratified samples and variance estimators is analysed. The paper is organized as follows: section 1 summarises the notations used through the whole paper. Section 2 outlines the problem, while section 3 presents different forms of the variance of the population. In sections 4 and 5 the expectations of the sample variance and the variance estimator are developed in the frame of the investigated sampling design.

1. NOTATIONS

The notations in this paper are borrowed basically from *Cochran* (1977). In particular,

$\mathsf{U} = \{1, 2, 3, ..., N\}$ is a finite universe,
$s = \{1, 2, 3, ..., n\}$ is a sample from U,
$\overline{Y}$ is the population mean of a study variable y,
$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})^2$ is the variance of y in U,
$\overline{y} = \frac{y_1 + y_2 + ... + y_n}{n}$ is the estimate of $\overline{Y}$ under simple random sampling (SRS),
$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$ is the sample estimate of S^2 under SRS (care will be taken to prevent misinterpreting the square root of this for the notation of a sample),

[1] Head of Section of the Hungarian Central Statistical Office.

$V(\bar{y}) = \frac{N-n}{N} \cdot \frac{S^2}{n} = (1-f)\frac{S^2}{n}$ is the variance of $\bar{y}$, where $f = n/N$,

$v(\bar{y}) = (1-f)\frac{s^2}{n}$ is the sample estimate for $V(\bar{y})$,

$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$ is the estimate of $\bar{Y}$ from a stratified sample where the W_hs and the $\bar{y}_h$s are stratum weights and estimated stratum means, respectively,

$V(\bar{y}_{st}) = \sum_{h=1}^{L}(1-f_h)W_h^2 \frac{S_h^2}{n_h}$ is the variance of $\bar{y}_{st}$ where n_h, $f_h = n_h / N_h$ and S_h^2 are sample size, sampling fraction and population variance for stratum *h*, respectively,

$\bar{y}_{sy} = \frac{y_i + y_{i+k} + y_{i+2k} + ... + y_{i+(n-1)k}}{n}$ is the estimate of $\bar{Y}$ from the i^{th} systematic sample provided $N = nk$, $1 \le i \le k$,

$V(\bar{y}_{sy}) = \frac{N-1}{N} S^2 - \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$ is the variance of $\bar{y}_{sy}$ where $y_{i1} = y_i$, $y_{i2} = y_{i+k}$, ..., $y_{in} = y_{i+(n-1)k}$, and $\bar{y}_{i.}$ is the mean of *this* sample.

The relation $N = kn$ is supposed to hold for some integer *k* throughout the paper.

2. THE PROBLEM

The research was motivated by the following modification of systematic sampling: for a fixed order of the units in U, in place of the customary systematic sample

$$s_{sy} = \{i, i+k, i+2k, ..., i+(n-1)k\},$$

use

$$s_{st} = \{ i_1, i_2+k, i_3+2k, ..., i_n+(n-1)k \} \qquad /1/$$

where i_1, i_2, i_3, ..., i_n are different random integers between 1 and *k*. s_{st} is obviously a stratified random sample with one unit per stratum. If *n* and *N* are fixed, the numbers of different systematic samples and stratified samples with one unit per stratum are *k* and k^n, respectively, thus one might expect that the latter are superior to the former. However, the comparison of the two designs in terms of variance turns to be quite hard.

Considering samples of equal size, *Cochran* (1977) specifies situations where systematic sampling is superior to simple random sampling. Among the drawbacks of the method, he mentions that hidden periodicity in the order of units may result in poor precision of $\bar{y}_{sy}$ and no reliable procedure is available to estimate $V(\bar{y}_{sy})$ from the sample. Nevertheless, the sample estimate

$$v(\bar{y}) = \frac{1-f}{n(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad /2/$$

of the variance of an estimated mean in simple random sampling is widely used to estimate $V(\bar{y}_{sy})$. This practice may be supported by a result of *W.G. Madow–L.H. Madow* (1944), quoted also in *Cochran* (1977); this asserts that

$$E(V(\bar{y}_{sy})) = V(\bar{y}) = (1-f)\frac{S^2}{n},$$

where the expectation is taken over all permutations of the units 1, 2, ..., N of U.

As for the stratified sample with one unit per stratum, the theoretical variance in this case is

$$V(\bar{y}_{st}) = (1-\frac{1}{k})\frac{1}{n^2}\sum_{h=1}^{n} S_h^2, \qquad /3/$$

since each $n_h = 1$, the sampling fraction in each stratum is $1/k$, and each stratum weight W_h equals $k/N = 1/n$. Taking the expectation of $V(\bar{y}_{st})$ over all permutations of the units of U leads to the same result as in the case of systematic sampling, i.e. the theoretical variance of the mean under SRS. In other words, comparing systematic sampling and stratified sampling with one unit per stratum on the basis of the expectations of $V(\bar{y}_{sy})$ and $V(\bar{y}_{st})$ results in a draw.

Since S_h^2 in /3/ cannot be estimated on the basis of a single observation, the usual sample-based estimator of $V(\bar{y}_{st})$ breaks down. *Cochran* (1977) enlists a number of different approaches to estimate $V(\bar{y}_{st})$ from the sample in the case of one unit per stratum; some of these work with collapsing adjacent strata, while other methods use auxiliary variables or specific hypotheses on the properties of the units of U. *W.A. Fuller's* unbiased estimator (1970) for $V(\bar{y}_{st})$ does not use collapsed strata or auxiliary variables, but randomizes the strata boundaries. The bulk of this paper is the relation between the expectation of $v(\bar{y})$ in /2/ under the stratified design with one unit per stratum and the variance $V(\bar{y}_{st})$.

3. ALTERNATIVE VARIANCE EXPRESSIONS

Lemma 1. With the notations of section 1, we have

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N} y_i^2 - \frac{1}{N(N-1)}\sum_{i,j}^{N} y_i y_j \qquad /4/$$

and

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i,j}^{n} y_i y_j . \qquad /5/$$

The statement is proved by routine computation. □

Equations /4/ and /5/ can be rewritten in matrix-vector form. Denote e.g. **I** the unit matrix of order N, **E** an $N\times N$ matrix whose entries are all equal to 1, and **y** the N-vector with the components y_1, y_2, ..., y_N. /4/ becomes then

$$S^2 = \frac{1}{N-1}\mathbf{y}'(\mathbf{I}-\frac{1}{N}\mathbf{E})\,\mathbf{y}, \qquad /6/$$

where the prime denotes transpose. Note that $\mathbf{C}=\frac{1}{N-1}(\mathbf{I}-\frac{1}{N}\mathbf{E})$ is symmetric, positive semidefinite, having N-1 eigenvalues equal to $1/(N-1)$ and one eigenvalue equal to 0.

Lemma 2. (*Decomposition of the Variance*). Let n, N and k be integers and $N=kn$. Decompose the universe U in two parts $U_1 = \{1, 2, 3, ..., N-k\}$ and $U_2 = \{N-k+1, N-k+2, ..., N\}$, and denote $S(U_1)^2$ and $S(U_2)^2$ the corresponding variances of the study variable y.

We have

$$S^2 = S(U)^2 = \frac{(N-k-1)(N-k)}{N(N-1)}S(U_1)^2 + \frac{k}{N(N-1)}\sum_{i=1}^{N-k} y_i^2 - \frac{2}{N(N-1)}\sum_{i=1}^{N-k}\sum_{j=N-k+1}^{N} y_i y_j + $$
$$+\frac{k(k-1)}{N(N-1)}S(U_2)^2 + \frac{N-k}{N(N-1)}\sum_{j=N-k+1}^{N} y_j^2 .$$

The proof is done by routine computation. □

Corollary 1. Under the conditions of *Lemma* 2, the variance for the subpopulation U_2 is given as:

$$S(U_2)^2 = \frac{n(N-1)}{k-1}S(U)^2 - \frac{(n-1)(N-k-1)}{k-1}S(U_1)^2 + \frac{2}{k(k-1)}\sum_{i=1}^{N-k}\sum_{j=N-k+1}^{N} y_i y_j -$$
$$-\frac{1}{k-1}\sum_{i=1}^{N-k} y_i^2 - \frac{n-1}{k-1}\sum_{j=N-k+1}^{N} y_j^2 . \qquad /7/$$

This identity can be used to estimate $S(U_2)^2$ if none of the units $N-k+1$, $N-k+2$, ..., N is observed.

4. THE EXPECTATION OF s^2 UNDER STRATIFIED SAMPLING WITH ONE UNIT PER STRATUM

For sample /1/ replace the indices i_1, i_2+k, i_3+2k, ..., $i_n+(n-1)k$ by 1, 2, 3, ..., n, respectively, and compute

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i,j}^{n} y_i y_j . \qquad /8/$$

Lemma 3. Under stratified sampling with one unit per stratum, the expectation of s^2 in /8/ is

$$E(s^2) = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{2}{N(N-k)} \sum_{i<j,\, j-i>k-1}^{N} y_i y_j \,. \qquad /9/$$

Proof: There are n strata, each consisting of k units. Therefore:

– k^n is the number of all different samples, thus the probability of each sample s is $\pi(s) = k^{-n}$,

– k^{n-1} is the number of samples containing a fixed unit i,

– k^{n-2} is the number of samples containing a fixed unit i from stratum h and a fixed unit j from stratum h', $h' \neq h$,

– there is no sample containing two different units i and j from the same stratum h.

Rewrite /8/ as follows:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)} \sum_{i \neq j}^{n} y_i y_j \,, \qquad /8a/$$

multiply both sides by $\pi(s)$, and take the sum over all samples s. By symmetry, the result will be the following:

$$E(s^2) = \lambda\sum_{i=1}^{N} y_i^2 - 2\mu \sum_{i<j,\, j-i>k-1}^{N} y_i y_j \,.$$

The first term on the right-hand side contains a factor $(1/n) \times k^{-n}$, and since each y_i^2 occurs k^{n-1} times, and $nk = N$, it follows that $\lambda = 1/N$. On the other hand, s^2 in /8a/ vanishes for $y_1 = y_2 = ... = y_n$, which implies the similar relation for $E(s^2)$. This results in $\mu = 1/((N(N-k))$. The proof is thereby complete. □

Denote S^{*2} the left-hand side of /9/. The matrix-vector form of that relation is the following:

$$S^{*2} = \frac{1}{N}\mathbf{y}'(\mathbf{I} - \frac{1}{N-k}\mathbf{D})\mathbf{y}, \qquad /10/$$

where **D** is the direct (or *Kronecker*) product of the $n \times n$ matrix

$$\mathbf{a} = \begin{pmatrix} 0 & 1 & 1 & 1 \ldots 1 \\ 1 & 0 & 1 & 1 \quad 1 \\ 1 & 1 & 0 & 1 \quad 1 \\ \vdots & & & \quad \vdots \\ 1 & 1 & 1 & 1 \ldots 0 \end{pmatrix},$$

and the $k\times k$ matrix **b,** whose entries are all equal to 1:

$$\mathbf{D} = \mathbf{a} \otimes \mathbf{b}.$$

D is of the following form:

$$\mathbf{D} = \begin{pmatrix} \mathbf{0} & \mathbf{b} & \mathbf{b} & \mathbf{b} & \dots & \mathbf{b} \\ \mathbf{b} & \mathbf{0} & \mathbf{b} & \mathbf{b} & & \mathbf{b} \\ \mathbf{b} & \mathbf{b} & \mathbf{0} & \mathbf{b} & & \mathbf{b} \\ & & & & \ddots & \\ \mathbf{b} & \mathbf{b} & \mathbf{b} & \mathbf{b} & & \mathbf{0} \end{pmatrix}$$

Lemma 4. **D** has

– n-1 eigenvalues equal to $-k$,
– one eigenvalue equal to $N-k$,
– $N-n$ eigenvalues equal to 0.

Proof: **D** is a symmetric matrix , hence there is an $N\times N$ orthonormal matrix **U** such that $\mathbf{U}'\mathbf{D}\mathbf{U}$ is a diagonal matrix (here and in what follows, the prime denotes transpose). Such a matrix **U** can be defined as follows. Let **u** and **c** be the following $k\times k$ and $n\times n$ matrices, respectively:

$$\mathbf{u} = \begin{pmatrix} 2^{-1/2} & 6^{-1/2} & 12^{-1/2} & \dots & (k(k-1))^{-1/2} & k^{-1/2} \\ -2^{-1/2} & 6^{-1/2} & 12^{-1/2} & & (k(k-1))^{-1/2} & k^{-1/2} \\ 0 & -2\times 6^{-1/2} & 12^{-1/2} & & (k(k-1))^{-1/2} & k^{-1/2} \\ 0 & 0 & -3\times 12^{-1/2} & & (k(k-1))^{-1/2} & k^{-1/2} \\ \vdots & & & & & \\ 0 & 0 & 0 & & -(k-1)(k(k-1))^{-1/2} & k^{-1/2} \end{pmatrix}$$

$$\mathbf{c} = \begin{pmatrix} 2^{-1/2} & 6^{-1/2} & 12^{-1/2} & \dots & (n(n-1))^{-1/2} & n^{-1/2} \\ -2^{-1/2} & 6^{-1/2} & 12^{-1/2} & & (n(n-1))^{-1/2} & n^{-1/2} \\ 0 & -2\times 6^{-1/2} & 12^{-1/2} & & (n(n-1))^{-1/2} & n^{-1/2} \\ 0 & 0 & -3\times 12^{-1/2} & & (n(n-1))^{-1/2} & n^{-1/2} \\ \vdots & & & & & \\ 0 & 0 & 0 & & -(n-1)(n(n-1))^{-1/2} & n^{-1/2} \end{pmatrix},$$

and set $\mathbf{U} = \mathbf{c}\otimes\mathbf{u}$, the *Kronecker* product of **c** and **u**. It is easy to check that $\mathbf{u}'\mathbf{u} = \mathbf{I}_k$ and $\mathbf{U}'\mathbf{U} = \mathbf{I}$ where $\mathbf{I}_k$ and **I** are unit matrices of order k and N, respectively.

It is also easy to see that

$$\mathbf{M} = \mathbf{U'DU}$$

is a diagonal matrix such that

$$\mathbf{M} = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_{n-1}, \mu_n),$$

the μs are $k \times k$ diagonal matrices, $\mu_1 = \mu_2 = ... = \mu_{n-1} = \mathrm{diag}(0, 0, ..., 0, -k)$, and $\mu_n = \mathrm{diag}(0, 0, ..., 0, N-k)$, so the proof is complete. □

Corollary 2. Denote $\mathbf{C}_0$ the matrix of the quadratic form in /10/. $\mathbf{C}_0$ has

– $n-1$ eigenvalues equal to $n/(N(n-1))$,
– one vanishing eigenvalue and
– $N-n$ eigenvalues equal to $1/N$.

Theorem 1. If the conditions of *Lemma* 3 hold, the relative difference between the expectation of the SRS variance estimate /8/ and the population variance S^2 does not exceed $1/(n-1)$, i.e.

$$\frac{|E(s^2)-S^2|}{S^2} = \frac{|S^{*2}-S^2|}{S^2} < \frac{1}{n-1} .$$

Proof: Let $\mathbf{e} = (1, 1, \ldots, 1)'$ be the N-vector with all components equal to 1, and $\mu\mathbf{e}$ the orthogonal projection of $\mathbf{y}$ onto $\mathbf{e}$ where $\mathbf{y}$ represents the values of the study variable on the units of U. Then $\mathbf{y} = \mathbf{z} + \mu\mathbf{e}$ and $\mathbf{z}$ is orthogonal to $\mathbf{e}$, $\mathbf{z'e} = 0$. Consider the matrix-vector representations /6/ and /10/ of S^2 and S^{*2}, respectively. Note that $\mathbf{y'Cy} = \mathbf{z'Cz} = \frac{1}{N-1}\mathbf{z'z}$ and $\mathbf{y'C_0y} = \mathbf{z'C_0z}$, and that the columns of $\mathbf{U}$ in the proof of *Lemma* 4 are eigenvectors of the matrix $\mathbf{E}$, too. Set $\mathbf{\Lambda} = \mathbf{U'EU}$. We have

$$|S^{*2}-S^2| = |\mathbf{z'Cz} - \mathbf{z'C_0z}| = |\mathbf{z'U'}(\frac{1}{N-1}\mathbf{I} - \frac{1}{(N-1)N}\Lambda - \frac{1}{N}\mathbf{I} + \frac{1}{N(N-k)}\mathbf{M})\mathbf{Uz}| =$$

$$= |\mathbf{z'U'}(\frac{1}{(N-1)N}\mathbf{I} - \frac{1}{(N-1)N}\Lambda + \frac{1}{N(N-k)}\mathbf{M})\mathbf{Uz}| \le \rho\, |\mathbf{z'U'Uz}| = \rho\, \mathbf{z'z},$$

where ρ is the maximum of the absolute values of diagonal entries in the diagonal matrix within the brackets. The latter will be denoted by $\mathbf{T} = \mathrm{diag}(\tau_1, \tau_2, ..., \tau_N)$. Let λ_i and μ_i be the i^{th} diagonal entry in Λ and $\mathbf{M}$, respectively. The following cases occur:

λ_i	μ_i	τ_i
0	0	$1/((N-1)N)$
0	$-k$	$1/((N-1)N) - 1/((n-1)N)$
N	$N-k$	0

Ignoring the term $1/((N–1)N)$, we obtain $\rho = 1/((n–1)N)$ and

$$|\, S^{*2} - S^2 \,| \le \frac{1}{(n-1)N}\ \mathbf{z'z} < \frac{1}{n-1}\ \mathbf{z'Cz}$$

as was to be shown. □

5. THE EXPECTATION OF $v(\bar{y})$ UNDER STRATIFIED SAMPLING WITH ONE UNIT PER STRATUM

Recall that the variance of the mean $\bar{y} = \bar{y}_{st}$ under stratified sampling with one unit per stratum is

$$V(\bar{y}) = (1-\frac{1}{k})\frac{1}{n^2}\sum_{h=1}^{n} S_h^2 \qquad /3/$$

and that direct estimation of the stratum variance S_h^2 on the basis of a single observation is not possible. In this section we assume that $n > 2$ and introduce the following estimator for S_h^2:

$$s_{(h)}^2 = \frac{n(N-1)}{k-1}s^2 - \frac{(n-1)(N-k-1)}{k-1}s_{(-h)}^2 - \frac{k}{k-1}\sum_{i\ne h}^{n} y_i^2 - \frac{(n-1)k}{k-1}y_h^2 + \frac{2k}{k-1}\sum_{i\ne h}^{n} y_i y_h\ , \qquad /11/$$

where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i,j}^{n} y_i y_j\ ,$$

$$s_{(-h)}^2 = \frac{1}{n-2}\sum_{i\ne h}^{n} y_i^2 - \frac{1}{(n-1)(n-2)}\sum_{i\ne h, j\ne h}^{n} y_i y_j\ ,$$

i.e. $s_{(-h)}^2$ is computed similarly as s^2 but without the observation from stratum h. The properties of the estimator /11/ will be examined in several steps.

Lemma 5. /11/ can be rewritten as

$$s_{(h)}^2 = y_h^2 + \frac{1}{(n-1)(n-2)}\sum_{i\ne h, j\ne h}^{n} y_i y_j - \frac{2}{n-1}\sum_{i\ne h}^{n} y_i y_h\ ,$$

and adding up $s_{(h)}^2$ for $h = 1, 2, ..., n$ we obtain

$$\sum_{h=1}^{n} s_{(h)}^2 = \sum_{h=1}^{n} y_h^2 - \frac{1}{n-1}\sum_{h\ne h'}^{n} y_h y_{h'}\ .$$

The result follows with routine computation. □

This lemma and /3/ result in the following estimator for $V(\bar{y}_{st})$

$$v(\bar{y}_{st}) = (1-\frac{1}{k})\frac{1}{n}(\frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i,j}^{n} y_i y_j) \qquad /12/$$

which is formally the same as the sample estimate of the variance of an estimated mean $\bar{y}$ under simple random sampling. Note that the indices in /12/ come from those reflecting both stratum and position in the same way as at the beginning of Section 4. By *Lemma* 3 we have

$$E\, v(\bar{y}_{st}) = (1-\frac{1}{k})\frac{1}{n}S^{*2}$$

where S^{*2} denotes the left-hand side of /9/.

Lemma 6. Set $h = n$. The expectation of $s_{(n)}^2$ under the one-unit-per-stratum design is

$$E(s_{(n)}^2) = \frac{n(N-1)}{k-1}S^*(\mathrm{U})^2 - \frac{(n-1)(N-k-1)}{k-1}S^*(\mathrm{U}_1)^2 + \frac{2}{k(k-1)}\sum_{i=1}^{N-k}\sum_{j=N-k+1}^{N} y_i y_j -$$
$$- \frac{1}{k-1}\sum_{i=1}^{N-k} y_i^2 - \frac{n-1}{k-1}\sum_{j=N-k+1}^{N} y_j^2 \, . \qquad /13/$$

where $S^*(\mathrm{U})^2$ and $S^*(\mathrm{U}_1)^2$ denote the expectation of s^2 under the one-unit-per-stratum design for the universe U and the union of the first n–1 strata U_1, respectively.

Proof: The first two terms on the right-hand side of /13/ are obviously the expectations of those on the right-hand side of /11/. The repeat of a part of arguments in the proof of *Lemma* 3 implies the similar statement for the last three terms on the right-hand side of /13/. □

Corollary 3. The bias of $s_{(n)}^2$, or, in general, that of $s_{(h)}^2$ is

$$E(s_{(h)}^2) - S_h^2 = \frac{n(N-1)}{k-1}(S^*(\mathrm{U})^2 - S(\mathrm{U})^2) - \frac{(n-1)(N-k-1)}{k-1}(S^*(\mathrm{U}_1)^2 - S(\mathrm{U}_1)^2) \, , \qquad /14/$$

where U is the universe and U_1 stands for the union of the strata except for stratum h.

Proof: The *Corollary* follows immediately from /13/ and /7/.

Lemma 5 and *Corollary* 3 lead to the following consequence:

Lemma 7.

$$\text{Bias } v(\bar{y}_{st}) = (1-\frac{1}{k})\frac{1}{n^2}\text{Bias}\sum_{h=1}^{n} s_h^2 =$$
$$= -\frac{1}{n^2k^2}\sum_{i=1}^{N} y_i^2 - \frac{1}{n^2k^2}\frac{N-1}{N-k}\sum_{|i-j|>k-1}^{N} y_i y_j + \frac{1}{n^2k^2}\sum_{i,j}^{N} y_i y_j \, , \qquad /15/$$

or, in matrix notations,

$$\text{Bias } v(\bar{y}_{st}) = -\frac{1}{n^2k^2}\ \mathbf{y}'(\mathbf{I} + \frac{N-1}{N-k}\mathbf{D} - \mathbf{E})\mathbf{y} = -\frac{1}{n^2k^2}\ \mathbf{y}'\mathbf{K}\mathbf{y} \qquad /15a/$$

where **y** is the vector of the study variable in the universe.

Proof: /15/ follows from /14/ by adding up the terms $E\left(s_{(h)}^2\right) - S_h^2$ for h = 1, 2, …, n and using the matrix representation for $S^*(\,.\,)^2$ and $S(.\,)^2$. Note that the first term on the right-hand side of /14/ repeats n times. In the second term, different $(N–k)\times(N–k)$ versions of **D** and **E** occur; the coefficients in /15/ depend on the number of occurrences of the $k\times k$ submatrices or blocks of these matrices. Off-diagonal blocks of **D** and **E** occur n–2 times and the diagonal blocks of **E** occur n–1 times. □

Lemma 8. For the orthonormal matrix **U** introduced in *Lemma* 4 we have

$$\mathbf{K} = \mathbf{U}\Delta\mathbf{U}'$$

where $\Delta = diag(\delta_1, \delta_2,\ldots, \delta_{n-1}, \delta_n)$, the δs are $k\times k$ diagonal matrices, and $\delta_1=\delta_2=\ldots=\delta_{n-1}=$ $=diag(1, 1, \ldots, 1, -\frac{n(k-1)}{(n-1)})$, $\delta_n = diag(1, 1, \ldots, 1, 0)$.

Proof: The *Lemma* is an immediate consequence of the previous results. □

Consider an arbitrary study variable represented by its **y** vector and decompose it in two components **x** and **z** such that within each stratum h, z_i is identical to the stratum mean of **y**:

$$z_i = \frac{1}{k}\sum_{j\in h} y_j\ , \quad \text{for } i \in \text{stratum } h,$$

and **x**≡**y** – **z**. It is easy to see that $\mathbf{z} = \mathbf{z_0}+\alpha\mathbf{e}$, where **e**=(1, 1,…,1)', α is the population mean of the study variable, $\alpha = \bar{Y}$, and $\mathbf{Kz_0}=\beta\mathbf{z_0}$ with $\beta\ –n(k–1)/(n–1)$. $\mathbf{z_0}$, **e** and **x** are pairwise orthogonal to each other, and **Ex = 0**, **Dx = 0**. It follows that

$$\text{Bias } v(\bar{y}_{st}) = -\frac{1}{n^2k^2}\mathbf{x}'\mathbf{x} + \frac{1}{n^2k^2}\frac{n(k-1)}{n-1}\mathbf{z_0}'\mathbf{z_0} = -\frac{1}{n^2k^2}\mathbf{x}'\mathbf{x} + (1-\frac{1}{k})\frac{1}{n}\frac{1}{N-k}\mathbf{z_0}'\mathbf{z_0}. \qquad /16/$$

The results of this section are summarised in the following.

Theorem 2.
$$V(\bar{y}_{st}) = (1-\frac{1}{k})\frac{1}{n}(\frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i,j}^{n} y_i y_j) \qquad /12/$$

is an estimator for

$$V(\bar{y}_{st}) = (1-\frac{1}{k})\frac{1}{n^2}\sum_{h=1}^{n} S_h^2$$

with expectation

$$E\, v(\bar{y}_{st}) = (1-\frac{1}{k})\frac{1}{n}(\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{1}{N(N-k)}\sum_{|i-j|>k-1}^{N} y_i y_j \ ,$$

and bias given by /16/.

Remark. From the aspect of stratification, the case where $\mathbf{y} \equiv \mathbf{z}$ is ideal, the variance of $\bar{y}_{st}$ is zero. Unfortunately, this is the worst case for the estimator /12/, the *upward* bias being of the same order as the expectation. If $\mathbf{y} \equiv \mathbf{x}$, the variance of $\bar{x}_{st}$ and the bias of /12/ are

$$\frac{1}{n^2 k}\mathbf{x'x} \quad \text{and} \quad -\frac{1}{n^2 k^2}\mathbf{x'x},$$

respectively, in this case the relative *downward* bias is of the order $1/k$. In principle there are cases where the bias vanishes.

REFERENCES

COCHRAN, W. G. (1977): *Sampling techniques*. (Third edition). John Wiley & Sons, New York-Chichester-Brisbane-Toronto-Singapore.

FULLER, W. A. (1970): Sampling with random stratum boundaries. *Journal of the Royal Statistical Society*, Vol. B32. p. 209–226.

MADOW, W. G. – MADOW, L. H. (1944): On the theory of systematic sampling. *Annals of Mathematical Statistics*, Vol. 15. p. 1–24.

THE ECONOMIC RETURNS TO EDUCATION: FINITE-SAMPLE PROPERTIES OF AN INSTRUMENTAL VARIABLE ESTIMATOR

GÁBOR KÉZDI[1]

This paper evaluates the instrumental variables measurement of the causal effect of education on earnings, with special focus on finite sample issues. A simple well-known theoretical model is presented and the inconsistency of the reduced-form estimator is established. The problem of weak instruments is examined in details and multiple remedies are considered. The relevant issues are illustrated through a particular example from a published paper, including a simulation exercise to inspect weak instruments problems and evaluate the performance of alternative estimators.

KEYWORDS: Returns to education; Instrumental variables; Weak instruments.

Instrumental variables models are very popular in empirical economics for estimating causal effects on observational data.[2] In their clearest form, causal effects can be stated in the framework of thought experiments. Because of nonrandom assignment, however, the non-experimental nature of virtually all economic data makes measurement of the thought experiments difficult. Simple reduced-form models like ordinary least squares (OLS) can be thought of as generalized versions of comparing means in different groups. The problem is that self-selection into those groups is typically not random, and therefore simple between-group comparisons do not measure the intended causal effects. Economics models often help capturing the non-randomness of the assignment and finding the direction of the resulting bias. Instrumental variables (IV) models offer more than that: under the necessary assumptions, IV results can be interpreted as estimates of the causal relationship. Problem is that the required assumptions are quite restrictive, and their validity is often difficult to assess. For accessible reviews of the IV and 'natural experiment' estimators, see *Meyer* (1995), *Angrist, Imbens* and *Rubin* (1996) with the comments and *Heckman* (1997).

The purpose of this paper is to evaluate the IV measurement of an extensively researched question, the causal effect of education on earnings, also known as the (private) economic returns to schooling. It is a good example that there is a simple economic

[1] PhD candidate, Department of Economics, University of Michigan.

[2] The topic of this paper was suggested by *John Bound* and *Jinyong Hahn*. I also thank *László Hunyadi* and *Gábor Körösi* for their helpful comments. All the remaining errors are mine.

model behind the self-selection story but the estimation raises quite a few econometric problems. *Willis* (1986) offers a thorough treatment of the model, and *Willis'* (1986) and *Card's* (1999) works are comprehensive surveys of the empirical literature. In this paper, I illustrate the problems through a specific study by *Harmon* and *Walker* (1995).

As we will see, *Harmon* and *Walker* find OLS to be biased downward. This result is not uncommon: most IV estimates reported by *Card* (1999) that use of similar instruments estimate negative or zero bias. At the same time the conventional theory of educational choice implies that OLS should be biased upwards. Either the theory is wrong or the estimation strategy is flawed. A third possibility might be measurement error in reported education level, which could induce a downward bias. It turns out, however, that the required error is an order of magnitude larger than what is established in the literature (The paper will show that in its example it has to be well over 50 per cent of the total variation.) *Card* (1999) considers a richer model than *Willis* (1986) that has predictions broad enough to incorporate some negative bias of the OLS. Harmon and Walker's estimates, similar to most estimates based on compulsory schooling instruments, are nevertheless a lot larger than all other kind of IV estimates. I am therefore sceptical about their results: I try to find out what could be wrong with the particular strategy they followed.

Among the possible problems, I will focus mainly on the finite sample properties of the estimator. In a controversial study, *Angrist* and *Krueger* (1991) had used an IV strategy similar to Harmon and Walker's to estimate returns to schooling. Their estimates also seem to suggest that OLS is not biased upward. However, *Bound, Jaeger and Baker* (1995) have shown that the instruments they used were weak and thus the results may not tell anything meaningful about the true population relationship of OLS and IV. The fact that Angrist and Krueger's IV was estimated on a very large sample but still suffers from finite-sample problems was suprising to many. The problem of 'weak instruments' was known to econometricians for quite a long time but it was typically ignored by practitioners who thought that large samples are immune to it. *Bound et al.* (1995), *Staiger* and *Stock* (1997) and others have convincingly shown, however, that weak instruments can be a problem in seemingly large samples, too. On the other hand, there exist modified IV estimators that are asymptotically equivalent to the conventional ones but have superior finite-sample properties. I will show the basic intuition behind the problem and consider a few of the alternatives.

The objective of this paper is primarily methodological. I would like to draw attention to the fact that IV can be a powerful strategy if supported by economic theory, but one should not ignore the econometric problems. In particular, the possible weakness of the instruments should be taken seriously. I would also like to show how one can detect the problem and that there are possible remedies. The causal effect of education on earnings serves a paradigmatic example, and my analysis of the *Harmon* and *Walker* (1995) study is intended to be an illustrative exercise. As it turns out, their instruments are not weak, and therefore their results are robust to finite-sample corrections. However, their strategy can be questioned on other grounds. Unfortunately, there is not enough available evidence to address those, but they suggest that it may be a little early to bury the conventional model of educational choice.

The remainder of the paper can be divided into five sections. First, it presents a relatively simple version of the theoretical model. The second part discusses the estimation of the return to schooling. The third section focuses on the finite sample properties of the IV estimator and introduces some alternatives. The fourth discusses *Harmon* and *Walker's* (1995) study and presents results of a Monte-Carlo simulation designed for capturing the finite-sample bias of their estimator. The last part concludes.

1. THE THOUGHT EXPERIMENT

To fix ideas, consider the following thought experiment. Take an individual, assign her a random level of education, and then measure her lifetime earnings. Then assign her a different level of education and measure her lifetime earnings again. The difference of the two earning levels represents the causal effect of education. Repeating the experiment enough times on enough randomly chosen individuals, one can get a good estimate for the average causal effect of education on earnings.

Obviously, this experiment is impossible to carry out. More important is that, as we will see, we can't capture anything close to it in observational data. Therefore, the causal effect of a randomly assigned level of education is impossible to measure. But that may not be a problem after all. As *Willis* and *Rosen* (1979) pointed out, this measure would have 'no significance as guides to the social or private profitability of investment in schooling' (*Willis–Rosen*, 1979, p. 11). The gains of a thorough education in econometrics with all necessary prerequisites would probably exceed its costs for most people, as they would find it a meaningless torture. On the other hand, people with appropriate interest, talent, and endurance probably find it very useful.

Instead of the mean return over all possible schooling levels for all people (the 'average treatment effect'), it makes sense to focus on the effect on those who have actually selected those (the 'average effect of the treatment on the treated'). The economic model of how people chose their education level helps identifying the problem.

A very simple model of education choice is enough to see why self-selection matters. In this model, individuals freely choose their schooling level. The only thing they care about is the present value of their lifetime earnings. They live an infinitely long life and face to a constant interest rate. They do not necessary face the same interest rate but it is constant throughout their lifetime. There are no costs of getting education except that they do not earn while in school. Schooling makes people earn more because it increases their marginal product. On the other hand, the increase in their marginal product is smaller and smaller as their schooling level rises. Again, we allow for heterogeneity in the relationship between marginal product and schooling.

The individuals' schooling choice is, therefore, the solution of an investment problem, where the value of the forgone earnings in the near future are weighted against the value of the increased earnings in the more distant future. The role of the discount rate is crucial. Let s_i denote the years spent in school by person i, r_i the individual-specific interest rate, and $y_i(s)$ the earnings function, also varying from individual to individual. We assume that $y'_i(s)>0$ and $y''_i(s)<0$.

$$s_i^* = \arg\max_s \int_{t=s}^{\infty} e^{-r_i t} y_i(s) dt = \arg\max_s \left(\frac{e^{-r_i s} y_i(s)}{r_i} \right).$$

The solution to this problem is given by the first-order condition (the second-order condition is satisfied by the concavity of y)

$$r_i = \left.\frac{y'_i(s)}{y_i(s)}\right|_{s=s_i^*} = \left.\frac{d \ln y_i(s)}{ds}\right|_{s=s_i^*}.$$

The 'returns to schooling' is defined as the value of the equality as we stated the first-order condition. It is the derivative of the logarithm of the earnings function at the optimum, which is equal to the discount rate of the individual faces. Since both the interest rate and the value of the derivative of the log earnings function can vary across people, the optimal schooling choice is expected to be different. When other things are equal, a higher r_i implies a lower optimal level of education. If two people differ solely in the interest rate they face, then they will have different schooling levels with the same earnings function $y(s)$.

In this case

$$\frac{\ln y_i(s_i^*) - \ln y_j(s_j^*)}{s_i^* - s_j^*} = \frac{\ln y_i(s_i^*) - \ln y_i(s_j^*)}{s_i^* - s_j^*} \approx \left.\frac{d \ln y_i(s)}{ds}\right|_{s=s^*},$$

so a simple reduced-form model (in this case a *Wald* regression) consistently estimates the causal effect, which is the same for both individuals. The estimator is consistent for the average treatment effect, that is the causal effect of randomly assigned education levels on people's earnings. The assignment is not random but that is not a problem, since it is uncorrelated with the effect (because the effect is the same for everybody). There may be a problem if the effect varies at different levels of education, but effects that are local to the different levels can be captured anyway. Obviously, the assumption of homogeneous earnings capacity is not realistic. If we allow for heterogeneous earnings functions the result does not hold anymore.

If two people face the same interest rate but have different earnings capacity, the one with a higher optimal level of schooling must have a higher $(\ln y_i)'$ at the optimum of the other:

$$s_i^* > s_j^* \quad \& \quad r_i = r_j \quad \Rightarrow \quad (\ln y_i)'\big|_{s=s_i^*} = (\ln y_j)'\big|_{s=s_j^*} \quad \Rightarrow \quad (\ln y_i)'\big|_{s=s_j^*} > (\ln y_j)'\big|_{s=s_j^*},$$

because $(\ln y_i)'$ is a decreasing function. For the same reason, we have that

$$\ln y_i(s_j^*) < \ln y_j(s_j^*),$$

and so

$$\frac{\ln y_i(s_i^*) - \ln y_j(s_j^*)}{s_i^* - s_j^*} > \frac{\ln y_i(s_i^*) - \ln y_i(s_j^*)}{s_i^* - s_j^*} \approx \left.\frac{d \ln y_i(s)}{ds}\right|_{s=s^*}.$$

The reduced-form model is biased upward. This is sometimes called the 'ability bias' of the reduced-form estimators, where heterogeneity in ability means heterogeneity of the earnings functions. Uncorrelated heterogeneity in interest rates has no effect on the bias, but there are special cases when a negative correlation might have an opposite effect (see the general setup by *Card*, 1999). In general, however, reduced form estimates will overstate the causal effect of education on earnings.

The model has two important implications. First, given people are free to chose the education they want given r_i and y_i, the derivative of the earnings function, that is, the causal effect of education on earnings can be observed only at the optimal level of education. Second, differences in earnings that correspond to different education levels overstate the causal effect of education.

The first point can be illustrated in the thought experiment to measure the causal effect of education on the earnings of a particular individual. This effect is the derivative of the $\ln y_i$ function, but one can only aim at measuring this derivative at the optimal s^*. The following is one appropriate design: let the individual choose an education level. Observe that: in according to the model, it is s^*. Also observe the corresponding earnings, $\ln y_i(s_i^*)$. Then induce a slightly different schooling level to the individual, and observe that s_i and the corresponding $\ln y_i(s_i)$. The individual has to change her decision, so a new s_i is going to be optimal. One way of doing this would be to change the interest rate she faces, another to constrain her choice to a slightly different previously non-optimal s_i. The difference between the two measured points would approximate $d\ln y_i(s_i^*)/ds$. This thought experiment identifies the local average effect of the treatment on the treated. It is not equal to the average treatment effect (the mean over all different education levels across all people) because for each individual, it is measured at the person-specific optimum.

In real life, of course, one person chooses some education level only once and gets lifetime earnings also once in a lifetime. Therefore, the only way to measure any effect is through inter-personal differences in schooling and earnings. The second implication of the model means that reduced form estimators like OLS overstate the causal effect, even in the local sense.

In addition, real-life measurement might contain measurement errors. Measurement error in the left-hand side variable does not affect consistency of the estimator, but right-hand side errors do. For the reason mentioned previously, measurement error in the schooling level variable has received considerable attention in the literature, and therefore it will be incorporated in the analysis. In the next section, the ways how these two econometric problems affect the OLS estimator and how a valid IV can help will be presented.

2. THE EMPIRICAL MODEL

Let s_i be the observed time (years) spent in school, an imperfect measure of real time spent in school, s_i^*. (The notation is a bit unfortunate: from now on, s_i is the observed value of the optimal choice s_i^*, not just any schooling level as before.) Let x_i be a k–1 dimensional vector of other factors affecting earnings, z_i a vector of factors affecting the schooling outcome but not the earnings, and y_i the logarithm of (lifetime) earnings.

With the measurement error, the model is specified by three equations:

$$y_i = \beta s_i^* + \delta' x_i + u_i \,, \qquad /1/$$

$$s_i^* = \alpha' z_i + \gamma' x_i + v_i \,, \qquad /2/$$

$$s_i = s_i^* + w_i \,, \qquad /3/$$

x_i and z_i are assumed to be uncorrelated with each of the error terms, u_i, v_i, and w_i.

Let us examine the inconsistency of the OLS estimator. OLS estimation of /1/ is inconsistent because s_i^* is endogenous and may be badly measured. s_i^* is endogenous because of a nonzero covariance of u_i and v_i, the unobserved heterogeneity in schooling assignment and earnings. A positive correlation is implied by the former simple model: u_i represents unobserved heterogeneity of the earnings functions, while v_i represents unobserved heterogeneity in the interest rate and the earnings function.

$$E\left[s_i^* u_i\right] = E\left[(\alpha' z_i + \gamma' x_i + v_i) u_i\right] = E\left[v_i u_i\right] = \sigma_{uv} \neq 0 \,. \qquad /4/$$

The measured model is:

$$y_i = \beta s_i - \beta w_i + \delta' x_i + u_i \equiv \beta s_i + \delta' x_i + \varepsilon_i \,, \qquad /5/$$

$$s_i = \alpha' z_i + \gamma' x_i + v_i + w_i = \alpha' z_i + \gamma' x_i + \eta_i \,. \qquad /6/$$

The measurement error, w_i is independent of all exogenous variables and all other error terms, by assumption. Therefore, z_i and x_i are uncorrelated with η_i and ε_i. Estimates of the earnings equation by OLS are inconsistent because of the covariance between the schooling level and the unobserved heterogeneity, and because of the measurement error. The asymptotic bias is a function of σ_{uv}, σ_w^2, and the moments of the covariates in the earnings equation (s and x). Let us derive the probability limit of the OLS estimator.

$$\begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix}_{OLS} \equiv \left[\sum_{i=1}^{n} \begin{pmatrix} s_i^2 & s_i x'_i \\ x_i s_i & x_i x'_i \end{pmatrix} \right]^{-1} \sum_{i=1}^{n} \begin{pmatrix} s_i y_i \\ x_i y_i \end{pmatrix} = \left(\mathbf{X}'_n \mathbf{X}_n\right)^{-1} \mathbf{X}'_n \mathbf{y}_n =$$

$$= \left(\mathbf{X}_n^*{}' \mathbf{X}_n^* + \Sigma_w\right)^{-1} \left(\mathbf{X}_n^*{}' \mathbf{X}_n^* \begin{pmatrix} \beta \\ \delta \end{pmatrix} + \mathbf{X}_n^*{}' \mathbf{u}_n \right) \qquad /7/$$

where

$$\mathbf{X}_n \equiv \left[s_i \quad x_i'\right]_n, \qquad \mathbf{X}_n^* \equiv \left[s_i^* \quad x_i'\right]_n, \qquad \Sigma_n \equiv \begin{bmatrix} \sigma_w^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_n ,$$

$$\mathbf{y}_n \equiv \left[y_i\right]_n, \qquad \mathbf{u}_n \equiv \left[u_i\right]_n .$$

Therefore,

$$plim\begin{pmatrix}\hat{\beta}\\ \hat{\delta}\end{pmatrix}_{OLS} = plim\left(\mathbf{X}_n^*{}'\mathbf{X}_n^* + \sigma_w \mathbf{e}_n \sigma_e \mathbf{e}_n{}'\right)^{-1}\left(\mathbf{X}_n^*{}'\mathbf{X}_n^*\begin{pmatrix}\beta\\ \delta\end{pmatrix} + \mathbf{X}_n^*\mathbf{u}_n\right) =$$

$$= \left(\Phi^{-1} - \frac{1}{1+\sigma_w^2\xi}\Phi^{-1}\Sigma_w\Phi^{-1}\right)\left(\Phi\begin{pmatrix}\beta\\ \delta\end{pmatrix} + E\left[\begin{pmatrix}s\\ x\end{pmatrix}u\right]\right) =$$

$$= \Phi^{-1}\left(\mathbf{I} - \frac{1}{1+\sigma_w^2\xi}\Sigma_w\Phi^{-1}\right)\left(\Phi\begin{pmatrix}\beta\\ \delta\end{pmatrix} + \begin{pmatrix}\sigma_{uv}\\ 0\end{pmatrix}\right),$$

where

$$\Phi \equiv E\left[\left(s^* x'\right)'\left(s^* x'\right)\right], \qquad \mathbf{e}_n = \left[\hat{\varepsilon}_i\right]_n$$

and $\xi>0$ is the upper left element of Φ^{-1}.

One can show that the previous expression implies that

$$\hat{\beta}_{OLS} \to \frac{\beta+\sigma_{uv}}{1+\sigma_w^2\xi} \qquad \text{in probability as } n \to \infty,$$

where

$$\xi \equiv E\left[s_i^{*2}\right] - E\left[s_i^* x_i{}'\right]E\left[x_i x_i\right]^{-1}E\left[x_i s_i^*\right].$$

The OLS estimator is asymptotically biased by the covariance of the structural error terms in an additive way: the sign of the bias is the same as the sign of σ_{uv}. The effect of the measurement error is different: it makes β biased toward zero. In the presence of a positive correlation ($\sigma_{uv}>0$) and measurement error, the two effects are of opposite direction if $\beta>0$.

3. THE INSTRUMENTAL VARIABLE (IV) STRATEGY

The most commonly used empirical strategy to address these econometric problems is instrumental variables estimation. This involves using one or more instruments that affect the schooling decision but are uncorrelated with earnings, conditional on education. In the thought experiment introduced before, good instruments are those that would induce some individuals to choose a different education level than they would choose otherwise. Obviously, the thought experiment itself cannot be carried out, but one can hope to find two groups of otherwise comparable individuals: one that was affected by the instrument, and another one that was not. The IV strategy gets around the endogeneity problem. Moreover, it is consistent regardless of potential measurement error in education.

In general, the IV strategy estimates the parameter of interest for the subpopulation which was affected by the instruments in the sense that they would have changed their behaviour (*Imbens–Angrist*; 1994). The results are therefore local and they correspond to the effect of the treatment on the treated.

The validity of the instruments is an assumption: it cannot be inferred from the sample. Therefore, good instruments are not only hard to find but their quality is not testable. It is a matter of theoretical and speculative discussion.

Before turning to the alternative IV estimators, let us understand why weak instruments can be a problem. To keep things very simple, assume that z is one dimensional and there is no x vector. That is, we have one instrument, z, and one badly measured and endogenous variable, s^*:

$$y_i = \beta s_i^* + u_i \qquad /8/$$

$$s_i^* = \alpha z_i + v_i \qquad /9/$$

$$s_i = s_i^* + w_i \qquad /10/$$

Formally, the IV assumptions state that the instrument is uncorrelated with the structural error terms and have some effect on the schooling choice:

$$E[z_i u_i] = E[z_i v_i] = E[z_i w_i] = 0\,, \qquad E[z_i s_i^*] \neq 0\,. \qquad /11/$$

The IV estimator of β is

$$\hat{\beta}_{IV} \equiv \frac{\hat{\sigma}_{zy}}{\hat{\sigma}_{zs}} \qquad /12/$$

It is consistent for β:

$$plim\hat{\beta}_{IV} = \frac{E[zy]}{E[zs]} = \frac{\beta E[zs^*] + E[zu]}{E[zs^*] + E[zw]} = \beta\,.$$

Why a weak instrument is problematic is not difficult to see. We say that an instrument is weak if it does not predict well the endogenous variable, or, in other words, if the regression of s on z (the first stage regression) has an R^2 close to zero. For given σ_s^2 and σ_z^2, this means that σ_{zs} is small in the limit, and therefore so will be typically in finite samples. Since zero population covariances do not mean zero covariances in finite samples, we can have the IV estimator dominated by covariances with the structural errors if $E[z_i s_i^*]$ is small. The problem of weak instruments is therefore a finite-sample problem. By definition, finite-sample problems are not relevant if the sample is large enough. How large is large enough however depends on the particular problem. *Angrist and Krueger* (1991) use quarter and state of birth as an instrument of schooling level, arguing that state-specific compulsory schooling laws might affect final schooling levels. That is, people born in the fall start school almost a year later than those born in the summer, and therefore some of them might complete one less class than children of the summer. Their sample has 329 000 observations but their instruments predict very little of actual schooling level (the partial R^2 is around 0.0001). Simulations by *Bound et al.* (1995) demonstrate that *Angrist* and *Krueger's* (1991) result can easily be an artifact of small-sample bias and tell nothing about the causal relationship. In fact, the causal relationship may well be zero. The simulation results are theoretically supported in the alternative asymptotics of *Staiger* and *Stock* (1997). The *F*-statistic on the excluded instruments in the first-stage regression is an indicator of how strong the instruments are. An *F*-statistic below 10 is usually seen as a sign of warning.

While intuitively appealing, this rule of thumb is not fully justified by econometric theory. A more careful way to detect weak instruments is by simulation exercises. They have an additional advantage in that they may tell something about the possible remedies if needed. I will present such an exercise on Harmon and Walker's estimator.

The IV estimators

If there are more than one excluded instrumental variables (z), there are more ways to combine them. In what follows, two different methods will be introduced. Three additional estimators will be defined as possible remedies for the weak instrument problem. They are all consistent (under the IV assumptions), but their asymptotic variance and their finite-sample bias and variance can be quite different.

Some new notation will help in the definitions. As before, bold letters denote sample matrices (and vectors). Let S denote the sample sum of squares, and θ the vector of parameters of interest, β and δ:

$$S_{ZZ} \equiv \frac{1}{n}\mathbf{Z}_n{}'\mathbf{Z}_n \equiv \frac{1}{n}\sum_{i=1}^{n}(z_i{}' \quad x_i{}')'(z_i{}' \quad x_i{}'),$$

$$S_{ZX} \equiv \frac{1}{n}\mathbf{Z}_n{}'\mathbf{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n}(z_i{}' \quad x_i{}')'(s_i \quad x_i{}'),$$

$$S_{Zy} \equiv \frac{1}{n}\mathbf{Z}_n{}'\mathbf{y}_n \equiv \frac{1}{n}\sum_{i=1}^{n}(z_i{}' \quad x_i{}')'y_i\,,$$

$$\theta \equiv (\beta \quad \delta')'.$$

Also, let **P** denote the projection matrix onto the column space of **Z**, and **M** matrix creating the residual:

$$\mathbf{P}_n \equiv \mathbf{Z}_n(\mathbf{Z}_n{}'\mathbf{Z}_n)^{-1}\mathbf{Z}_n{}',$$

$$\mathbf{M}_n \equiv \mathbf{I}_n - \mathbf{P}_n\,.$$

The Optimal GMM Estimator

Generalized Method of Moments (GMM) estimators are based on moment restrictions.

Here those involve the covariance of the excluded instruments and the error term in the earnings equation:

$$E[z_i'\varepsilon_i] = E[z_i'(y_i - \beta s_i - \delta' x_i)] = 0$$

The GMM in general is defined in the following way:

$$\hat{\theta}_{GMM} \equiv \left(S'_{ZX} A^{-1} S_{ZX}\right)^{-1} S'_{ZX} A^{-1} S_{Zy} \qquad /13/$$

The estimator is consistent with any positive definite matrix *A*. The optimal GMM estimator is a special case, where *A*, the weighting matrix, is the covariance matrix of the product of the error term and the instruments (in the broad sense):

$$\hat{\theta}_{OGMM} \equiv \left(S'_{ZX} \Omega^{-1} S_{ZX}\right)^{-1} S'_{ZX} \Omega^{-1} S_{Zy}\,, \qquad /14/$$

where

$$\Omega \equiv Var\left[\varepsilon \begin{pmatrix} z \\ x \end{pmatrix}\right] = E\left[\varepsilon^2 \begin{pmatrix} z \\ x \end{pmatrix} (z' \quad x')\right].$$

The implementation requires an estimate of Ω. Since the estimator is consistent with any other positive definite matrix, in the first step one can get a consistent estimate for ε_i by using any appropriate matrix. In particular, the identity matrix meets the required condition. The optimal GMM estimation therefore consists of two steps. The first one consists of estimating $\hat{\theta}_{OGMM_1} \equiv (S'_{ZX} S_{ZX})^{-1} S'_{ZX} S_{Zy}$, and taking the residuals $\hat{\varepsilon}_i = y_i - \hat{y}_{OGMM_1}$. Ω can then be estimated using the estimated residuals and **Z**. The second step is the optimal GMM estimation, by using $\hat{\Omega}$. The optimal GMM estimator is the minimum distance combination of the different instruments if the system is overidentified. It is optimal in the sense that it has the smallest asymptotic variance.

Two Stage Least Squares

The Two Stage Least Squares (2SLS) estimator of θ is defined by

$$\hat{\theta}_{2SLS} \equiv \left(S'_{ZX} S_{ZZ}{}^{-1} S_{ZX}\right)^{-1} S'_{ZX} S_{ZZ}{}^{-1} S_{Zy}\,. \qquad /15/$$

This estimator is equivalent to the two-stage procedure of first regressing *s* on all *z* and *x* (‘first stage’), and then regressing *y* on the first-stage prediction of *s* and the *x* variables (‘second stage’).

$\hat{\theta}_{2SLS}$ is a GMM estimator with weight matrix S_{ZZ}^{-1}, and therefore it is consistent. Moreover, it is equivalent to the optimal GMM estimator if *ε* is homoscedastic. With out homoscedasticity, however, the two estimators are going to be different, and therefore 2SLS is not optimal. In cross-sectional applications, homoscedasticity is a rare exception. On the other hand, the 2SLS estimator is a lot simpler to compute, and it is part of all major statistical packages. For this reason, it remains the most popular IV estimator.

The k-class estimators

k-class estimators are a generalization of the two-stage least square introduced by *Theil* (1958), and are defined as:

$$\hat{\theta}_k \equiv \left[\mathbf{X}_n'\mathbf{P}_n\mathbf{X}_n - (1-k)\mathbf{X}_n'\mathbf{M}_n\mathbf{X}_n\right]^{-1}\left[\mathbf{X}_n'\mathbf{P}_n\mathbf{y}_n - (1-k)\mathbf{X}_n'\mathbf{M}_n\mathbf{y}_n\right]. \qquad /16/$$

2SLS is a *k*-class estimator, with *k*=1. *Nagar*'s estimator (introduced by *Nagar*; 1959) is another special case, where

$$k_{Nagar} = 1 + \frac{q-2}{n} \qquad /17/$$

q being the dimension of z, that is the number of instruments excluded from the earnings equation. Nagar's estimator has the minimum expected bias in finite samples among the *k*-class estimators. One of the assumptions behind this result is that the x_i are nonstochastic. *Donald* and *Newey* (1997) generalize the Nagar results and suggest a modified version

$$k_{Donald-Newey} = 1 + \frac{q-2}{n} \Big/ \left(1 - \frac{q-2}{n}\right). \qquad /18/$$

Limited Information Maximum Likelihood

Some more notation. Let **Y** be the sample matrix of the endogenous variables, and **P**(x) be the projection onto $[x_i']_n$:

$$\mathbf{Y}_n \equiv [y_i \quad s_i]_n, \quad \mathbf{P}(x) \equiv [x_i']_n\left([x_i']_n'[x_i']_n\right)^{-1}[x_i']_n', \quad \mathbf{M}(x) \equiv \mathbf{I}_n - \mathbf{P}(x).$$

The Limited Information Maximum Likelihood (LIML) estimator is derived from the likelihood function assuming normal errors. It can be expressed, however, as another *k*-class estimator with $k=\lambda$, where λ is the smallest eigenvalue of the matrix B defined as

$$\mathbf{B} \equiv \left(\mathbf{Y}_n'\mathbf{M}_n\mathbf{Y}_n\right)^{-1}\mathbf{Y}_n'\mathbf{M}(x)\mathbf{Y}_n.$$

Staiger and *Stock* (1997) show that in their framework, the LIML estimator has the best finite-sample properties.

4. HARMON AND WALKER'S STUDY

Harmon and *Walker* (1995) follow an instrumental variables approach to estimate β on British data. They use a pooled sample of the consecutive cross-sectional waves of the British Family Expenditures Survey (1978–1986). The sample is quite large, $n = 34\,336$. They use the cohort of the individual as an instrument. Their motivation is the following. The minimum school-leaving age was increased two times in the relevant period from 14 to 15 in 1947, and from 15 to 16 in 1973. They provide evidence that these changes indeed changed the behaviour of many individuals: quite a few of those who otherwise would have left school stayed on because of the new law. This strategy is an example of what the natural experiments literature calls a difference in differences estimator (*Meyer*; 1995).

The instrument directly changes the schooling level 'assignment' of the individuals who complete one more class during this time, since it forces them to choose $s_i > s_i^*$, similarly to our thought experiment. It measures the treatment on the treated, those that would have left school at a younger age but had had to stay in school and completed one more class. The instrument is strong enough if this sub-population is significant.

Technically, their instrument is a 3-valued vector indicating whether the person is a member of the cohort that was subject to the first, second, or third minimum school-leaving age (SLA) requirement (SLA=14, 15, or 16). z_i therefore is a vector 2 binary variables, taking SLA=14 the reference group. The other covariates, the x_i-s include a constant, age, age squared, region (10 categories + 1 reference), and year of survey (8 categories + 1 reference). An important implication of the 2-dimensional z_i vector is that 2SLS is equivalent to Nagar's estimator ($k_{Nagar}=1+(q\text{-}2)/n$) and also the *Donald–Newey* estimator. Therefore, in this empirical model, the 2SLS estimator is the IV estimator with the best finite-sample properties in the *Nagar* (1959) and in the *Donald* and *Newey* (1997) setup. It is dominated by the LIML estimator in the framework of *Staiger* and *Stock* (1997).

The OLS estimate is a lot smaller than the IV:

$$\hat{\beta}_{OLS} = 0.06\,,\ \hat{\beta}_{2SLS} = 0.15\,.$$

Taken by face value, the estimated causal effect of education is very large: it states that an additional year spent in school increases earnings by 15 percent for people at the very bottom end of the schooling distribution. This is a surprisingly large effect (most other estimates are at most 10 percent), and one has to be sure that the results are not confounded by small-sample effects or other factors before taking it seriously.

Provided that the IV estimate is correct, one can also conclude that OLS is biased downward by 60 percent. The question our simulation will answer is what combination of endogeneity and measurement error is needed for this result.

The Monte Carlo exercise

To examine the finite-sample properties of the Harmon and Walker's 2SLS estimator and the alternative IV estimators, this paper generated data similar to the original sample. The artificial samples were drawn from a population with moments that match the re ported ones. The Data Generating Process (DGP) of the part of z_i (i.e. the two binary SLA variables) is a multinomial process. The DGP of x_i consists of age (a uniform random variable in the simulations), its square, and two sets of multinomial variables for the mutually exclusive categories of the region and the year of the survey. The covariance of x with z was preserved (except for some simplifying assumptions with negligible consequences). The appendix table compares the simulated moments to the published ones.

The DGP for the structural error terms (u_i and v_i) was modeled as bivariate normal with a correlation of ρ_{uv}. s_i^* was generated following the schooling equation with the reported parameters. Observed schooling attainment variable was $s_i = s_i^* + w_i$, $w_i \sim iid\, N(0,\sigma_w^2)$, representing the measurement error. Finally, y_i was generated by the earnings equation, again, with the published parameters. Several combinations of ρ_{uv} and σ_w^2 were examined.

Formally, the generated vectors z_i and x_i and scalars u_i, v_i, and w_i were used to generate y_i, s_i^*, and s_i the following way:

$$y_i = \beta s_i^* + \delta' x_i + u_i \,,$$
$$s_i^* = \alpha' z_i + \gamma' x_i + v_i \,,$$
$$s_i = s_i^* + w_i \,,$$

where Harmon and Walker's point estimates were used for β, δ, α, and γ. In each run of the simulation new data were generated, and $\hat{\beta}$ was estimated in each dataset. The purpose of the Monte Carlo experiment was to examine the properties of four different estimators (OLS, optimal GMM, 2SLS and LIML) by comparing then to the 'theoretical' value of beta used in the data generating process.

Multiple measures of the bias were considered: the mean error (Bias) and the mean squared error (MSE) are reported in the following, the median and mean absolute error are available upon request. The reported measures are defined as

$$Bias \equiv \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\beta}_j - \beta\right), \qquad MSE \equiv \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\beta}_j - \beta\right)^2 ,$$

where M is the number of Monte Carlo replications, and $\hat{\beta}_j$ is the estimate of β in the j-th replication.

The following tables summarize the results of twelve simulations, each with different parametrization of the econometric problems ρ_{uv} and σ_w^2, each from 1000 replications. With no econometric problems (ρ_{uv} =0, σ_w^2=0), OLS is the best. IV estimators beat OLS in terms of the bias even with relatively small problems (ρ_{uv} =0.05, or σ_w^2=0.1). The difference among the alternative IV estimators are small. All of these indicate that Harmon–Walker's estimator delivers the asymptotic results. LIML seems to slightly outperform the 2SLS (which is here equivalent to the Nagar and also the Donald–Newey estimator), but this result is not robust.

Besides these results, I also carried out simulations on smaller samples. These might be interesting because of two reasons. First, it helps to see what sample size is 'large enough' and in what sense it is large in our setup. Second, the relative performance of the 2SLS (Nagar) and the LIML estimator would probably show better results in smaller samples. The results for n = 1 500, n = 3 000, n = 6 000, n = 9 000, and n = 15 000, are available on request. The simulations suggest that the IV estimators dominate the OLS estimator in terms of bias even with weak endogeneity or measurement error, and even in relatively small samples (n = 1 500). Their superiority disappears in terms of MSE. In fact, a rather serious econometric problem is needed to get smaller MSE for any of the IV estimators, if the sample size is small: $\rho_{uv} = 0.2$ if n = 1 500, and $\rho_{uv} = 0.1$ if n = 3 000. Note that with any reasonable scale of measurement error, all IV estimators underperform OLS in an MSE sense, unless the sample size is very large (n >10 000).

The small-sample results are not clear about the relative performance of the 2SLS (Nagar and Donald–Newey) and the LIML estimators in terms of the bias. On the other hand, the MSE and mean absolute error results seem rather robust. The LIML estimator underperforms 2SLS (Nagar) in small samples in terms of these measures. These results are consistent with the general notion that, even if it is better in expectation, the variance of the LIML estimator does not converge to zero (the asymptotic variance of the root-n magnified estimator is infinite).

Comparison of the performance of different estimators

Endogeneity (ρ_{uv})	Measurement Error (σ_w^2)	OLS	OGMM	2SLS	LIML
			Bias		
0.00	0.00	0.000	0.000	0.000	0.000
0.05	0.00	0.049	-0.001	-0.001	-0.001
0.10	0.00	0.098	0.000	0.000	0.000
0.15	0.00	0.147	0.000	0.000	0.000
0.20	0.00	0.195	-0.001	-0.001	-0.001
0.00	0.10	-0.025	0.000	0.000	0.000
0.00	0.20	-0.046	-0.001	-0.001	-0.001
0.00	0.50	-0.099	-0.001	-0.001	-0.001
0.10	0.10	0.057	-0.003	-0.003	-0.003
0.20	0.10	0.014	0.002	0.002	0.002
-0.10	0.10	-0.107	-0.002	-0.001	-0.001
-0.20	0.10	-0.190	-0.002	-0.001	-0.001
			MSE		
0.00	0.00	0.0000	0.0017	0.0017	0.0017
0.05	0.00	0.0024	0.0013	0.0013	0.0013
0.10	0.00	0.0096	0.0016	0.0016	0.0016
0.15	0.00	0.0216	0.0016	0.0016	0.0016
0.20	0.00	0.0382	0.0012	0.0012	0.0012
0.00	0.10	0.0006	0.0017	0.0017	0.0017
0.00	0.20	0.0022	0.0013	0.0013	0.0013
0.00	0.50	0.0097	0.0015	0.0015	0.0015
0.10	0.10	0.0033	0.0014	0.0014	0.0014
0.20	0.10	0.0195	0.0015	0.0015	0.0015
-0.10	0.10	0.0115	0.0014	0.0014	0.0014
-0.20	0.10	0.0360	0.0015	0.0015	0.0015

5. CONCLUSION

Harmon and Walker's IV estimate is more than twice larger than their OLS result not because their instruments are weak. Their model may give misleading results, though, for a different reason. The instrument they use is people's birth cohort. As it is often the case with similar difference in different natural experiment estimators, there are other things that might be varied between those groups, other than the compulsory schooling laws they faced. For one thing, the quadratic age-earnings profile may not be the right specification to use even if differences in age purely reflect variety in labor market experience (see *Murphy–Welch*; 1990). Moreover, in this cross-sectional setup, the joint effect of age and birth cohort on earnings may reflect time effects in earnings (cycles and trends). Another possible problem is the time path of employment in the unskilled group they focus on: their employment rate fell through the time considered in the developed world, therefore probably in Britain, too. That might have introduced selection problems into observed wages because the sample became more 'able'.

If we accept the results these problems notwithstanding, they tell us that those that would have left school at the minimum school age (14 or 15) but stayed one year longer because of the new law earn 15 per cent more because of this extra year. This is a classical local effect of the treatment on the treated. Generalizing these effects to the rest of the population is not justified by the empirical model itself: we need some theory for that. However, in the model presented previously (or, for that matter, in *Card* 1999), this is not possible without making more structure on the heterogeneity in the costs of education and the earnings function. The large literature on educational choice has not provided enough evidence for that yet.

The instrument in the illustrative example stood well in the simulation exercise but may be problematic for other reasons. I believe that one should carry out a similar analysis if there is enough reason to worry about the finite sample properties of the estimators, and one should check the robustness of the estimators to the other problems. Moreover, one has to be explicit in what exactly the results mean in terms of the locality of the causal effect and the characteristics of the treatment group. There is no free lunch in econometrics either: the real benefits of the instrumental variables estimation strategy can be exploited only by careful analysis.

APPENDIX

Sample moments from Harmon and Walker (1995) and the simulated DGP (1000 replications)

Variable	Whole Sample		SLA=14		SLA=15		SLA=16	
	H&W	Simul	H&W	Simul	H&W	Simul	H&W	Simul
Ln(wage)	1.913	1.907	1.902	1.837	1.995	2.003	1.584	1.644
Std(lnwage)	0.445	1.035	0.434	3.296	0.416	1.983	0.426	4.126
Age	38.7	38.7	55.8	55.8	35.6	35.6	21.6	21.6
Std(age)	12.7	12.8	4.5	6.9	7.3	3.0	2.7	5.2
Region1	0.088	0.088	0.088	0.088	0.085	0.088	0.101	0.088
Region2	0.110	0.110	0.105	0.110	0.109	0.110	0.119	0.110

(Continued on the next page.)

(Continuation.)

Variable	Whole Sample		SLA=14		SLA=15		SLA=16	
	H&W	Simul	H&W	Simul	H&W	Simul	H&W	Simul
Region3	0.075	0.075	0.073	0.075	0.074	0.075	0.082	0.075
Region4	0.099	0.099	0.103	0.099	0.099	0.099	0.090	0.099
Region5	0.037	0.037	0.038	0.037	0.037	0.037	0.032	0.037
Region6	0.306	0.306	0.317	0.306	0.300	0.306	0.311	0.306
Region7	0.074	0.074	0.070	0.074	0.074	0.074	0.080	0.074
Region8	0.051	0.051	0.050	0.051	0.050	0.051	0.054	0.051
Region9	0.089	0.089	0.081	0.089	0.101	0.089	0.050	0.089
Region10	0.013	0.013	0.013	0.013	0.013	0.013	0.017	0.013
Year1	0.116	0.116	0.143	0.143	0.118	0.118	0.062	0.062
Year2	0.116	0.116	0.140	0.140	0.116	0.116	0.075	0.075
Year3	0.121	0.121	0.129	0.129	0.122	0.122	0.102	0.102
Year4	0.118	0.118	0.113	0.113	0.119	0.119	0.118	0.118
Year5	0.101	0.101	0.084	0.084	0.105	0.105	0.113	0.113
Year6	0.104	0.104	0.091	0.091	0.103	0.103	0.135	0.135
Year7	0.101	0.101	0.069	0.069	0.102	0.102	0.159	0.159
Year8	0.102	0.102	0.058	0.058	0.100	0.100	0.192	0.192

REFERENCES

ANGRIST, J. D. – KRUEGER, A. B. (1991): Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, No.106. p. 979–1014.

ANGRIST, J. D.– IMBENS, G. W. – RUBIN, D. B (1996): Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, Vol. 91, No. 434. p.444–455.

BOUND, J. – JAEGER, D. A. – BAKER, R. M. (1995): Problems with instrumental variables estimation when the correlation between the instrument and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, Vol. 90, No.430. p. 443–450.

CARD, D. (1999): The causal effects of education on earnings. In: ASHENFELTER, O. – CARD, D. (eds): *Handbook of labor economics*, Vol. 3A. Elsevier, North-Holland, p. 1801–1868.

DONALD, S. G. – NEWEY, K. W. (1997): *Choosing the number of instruments*. MIT Department of Economics. Working Paper.

HECKMAN, J. J. (1997): Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. *Journal of Human Resources,* Vol. 32. No. 3. p. 441–462.

IMBENS, G. W. – ANGRIST, J. D. (1994): Identification and estimation of local average treatment effects. *Econometrica,* Vol. 62. No.4. p. 467–476.

HARMON, C. – WALKER, I. (1995): Estimates of the economic return to schooling for the United Kingdom. *American Economic Review*, Vol. 85. No. 5. p. 1278–1286.

MEYER, B. (1995): Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics*, Vol. 13. No. 2. p. 151–161.

MURPHY, K. M. – WELCH, F. (1990): Empirical age-earnings profiles. *Journal of Labor Economics*, Vol. 8. No. 2. p. 202–229.

STAIGER, D. – STOCK, J. H. (1997): Instrumental variables regressions with weak instruments. *Econometrica*, Vol. 65. No. 3. p. 557–586.

THEIL, H. (1958): *Economic forecasts and policy*. North-Holland, Chapter 6.

WILLIS, R. J. – ROSEN, S. (1979): Education and self-selection. *The Journal of Political Economy*, Vol. 87. No. 2. p. S7–S36.

WILLIS, R. J. (1986): Wage determinants: A survey and interpretation of human capital earnings functions. In: ASHENFELTER, O. – LAYARD (eds): *Handbook of labor economics*, Vol. 1. p. 525–602.

COMPARISON OF ESTIMATORS FOR PROBABILITY OF DEATH USED IN ACTUARIAL SCIENCE*

KOLOS ÁGOSTON[1]

In this paper a brief summary of the estimators for probability of death is given. Three estimators will be investigated, two of them are parametric ones. To derive a parametric estimator a distribution has to be assumed. Unfortunately the distributions occurring in real life differ from these assumed ones. It will be investigated how can these estimators be applied if we assume other distributions. The bias and efficiency of the estimators are analysed by using Monte Carlo simulation.

The maximum likelihood estimator is the most common in the actuarial practice due to its appealing point estimation properties. The Kaplan–Meier estimation is a better choice, if the purpose is to give a better confidence interval.

KEYWORDS: Actuarial sciences; Estimators; Simulation.

The estimation of probabilities of death (or other failures, for instance disability) is crucial in the life insurance industry. The most often used probability[2] is that a person aged exactly x, will die in a year, denoted by q_x. The main goal of this paper is to give a good estimator for this probability. There is a simple estimator for ratios, and this estimator can be used for estimating the probability of death as well: d/n, where d is the number of deaths, and n is the number of insured lives (where everyone is exactly x year old at the beginning). This estimator has good properties: in the frame of a binomial model it is a maximum likelihood estimator (consistent, asymptotically unbiased, and asymptotically most efficient), and it is unbiased in small samples too.

Though this estimator has very attractive properties, its application has special requirements as well. The most important one says that there is only one reason for exit, and this is death (or failure in general). Let us suppose that there are n elements (persons) at the beginning, d of them die within a year, and the others survive. It means that there is no lapse or censoring, as statisticians say. Furthermore, let us assume that a person – say Mr. Smith – buys an insurance policy, but after half a year he withdraws his policy. From

* The author is indebted to Professor *Erzsébet Kovács*, Professor *Angus Macdonald*, *László Papp* and *Krisztina Szőke* for their help.

[1] Assistant of the Department of Operations Research, Budapest University of Economic Sciences and Public Administration.

[2] These probabilities can be found in the life tables. The premium calculation is based on these probabilities as well.

that time the company has no information whether Mr. Smith has died or survived the year. Given that we do not know this, the ratio estimator cannot be applied any more.

There are many estimators that can handle this deficiency, but there is not an universal one. The paper will compare the properties of the three most widely used estimators. These are the Kaplan–Meier estimator, which is a non-parametric one, the actuarial estimator, and the maximum likelihood estimator.

1. PROPERTIES OF THE THREE COMMONLY USED ESTIMATORS

In this section the derivation and the properties of the three mentioned estimators will be given.

The *Kaplan–Meier estimator* can be derived as follows. Let us divide the one-year interval into sub-intervals, so that only one event occurs in any of these sub-intervals: a failure or a censor event. If within the sub-interval a censor event occurs, one can say that the estimated probability (of failure) is 0 for this interval. If there is only a failure within the sub-interval, the ratio estimator can be applied. The estimated probability for the whole year is one minus the product of the estimated survival probabilities (equals one minus the estimated probability of failure).

It should be noted that the observed lifetime is the difference between the starting time and the time of exit. The cause of exit can either be death or censoring. Let us arrange the observed lifetimes in standing order, and let t_1 be the smallest and t_n is the largest observed lifetime; d_i is a dummy variable which takes the value 1 if the i^{th} person died, otherwise it is 0. Using these variables, the Kaplan–Meier estimator is:

$$\hat{q} = 1 - \prod_{i=1}^{n} \left(\frac{n-i}{n-i+1} \right)^{d_i}$$

If d_n is 1 (i.e. that the last person died) then the estimator gives 1 for the probability of death. In this case it is said that estimation of the probability of death is not defined.

In a small numerical example three persons buy a specified insurance policy. The first one withdraws his contract at the one third of the year. The second dies at the two thirds of the year, while the third survives the year. The estimated probability for the first third year is 0, for the second third year is 1/2 and for the third year is also 0. So the estimated probability for the whole year is: $1 - \left(\frac{2}{3}\right)^0 \left(\frac{1}{2}\right)^1 \left(\frac{0}{1}\right)^0 = \frac{1}{2}$.

The *actuarial estimator* is a parametric one. Here it has to be assumed a rule of how the probability of death in a fraction of year is related to the probability of death in the whole year. According to the common actuarial notation ${}_sq_{x+t}$ means the probability of failure during the interval ($x+t$, $x+t+s$) for a person who is alive at the time $x+t$ ($0 \le t,s \le 1$, $t+s \le 1$). It is also common that if the length of the interval is a whole year and this year begins on the investigated person’s birthday, we do not need to write the left index: $q_x={}_1q_x$. Using these notations the assumed relation between the probabilities is:

$${}_{1-t}q_{x+t}=(1-t)q_x,$$

which is called the *Balducci* assumption. If the probability of death is 0.003 in a year, the probability of death in the interval (x+2/3, x+1) for a person who is alive at the time x+2/3 is: 1/3·0.003=0.001.

The idea behind the actuarial estimator is to adjust the estimator d/n. The number of deaths is given, so n has to be modified. It is better if one says that n means n years instead of n persons. So if somebody withdraws his contract, he does not stand in risk for a whole year, only for a fraction of year. If the person in the previous example withdraws his contract at one third of the year, he stands in risk for one third year. If somebody dies, he stands in risk for the whole year. The explanation for this is that in the case of estimator d/n, if somebody dies, he stands in risk for the whole year as well. Then the actuarial estimator is as follows: the number of deaths divided by 'standing in risk', which is called 'initial exposed to risk' in the actuarial literature. If no censoring event occurs, the actuarial estimator is equal to the common ratio estimator d/n.

Using this estimator for the previous example, the initial exposed to risk is 1/3+1+1=7/3, the number of deaths is 1, so the estimated probability is 1/(7/3)=0.43.

The next question is that how is this connected to Balducci hypothesis. Let us consider a small example. At the beginning there are $l+w$ members, w of them withdraw their contract after one third year. The expected number of deaths is $(l+w)\cdot q_x$, but the insurance company does not know about the deaths of the those who withdrew after the withdrawal. We can calculate the expected number of these deaths by using the Balducci hypothesis: $w\cdot 2/3 q_x$.

The expected number of deaths known by the insurance company equals the expected number of deaths in a year minus the expected number of deaths the company does not know about:

$$d=(l+w)\cdot q_x - w\cdot(2/3)q_x = l\cdot q_x + (1/3)\cdot w\cdot q_x \qquad /1/$$

If equation /1/ is arranged for q_x, we get the actuarial estimator:

$$\hat{q}_x = \frac{d}{l+\frac{1}{3}w}.$$

To derive the actuarial estimator we used the Balducci assumption. It can be proved that if the Balducci hypothesis is true then the actuarial estimator is asymptotically unbiased. (It was believed for quite a long time that the actuarial estimator is a moment estimator, but it was proved 20 years ago that this is not true.)

The *maximum likelihood estimator* is a very popular one too. Its main idea comes from realizing that however the probability of death is a good measure, it has a disadvantage as well, namely it is related to an interval. Sometimes it is better to use a measure for a point instead. We can get this measure as a limit: $\lim_{t\to 0}({}_t q_x / t)$, which is denoted by μ_x and called as the *force of mortality*. The probability of death can be calculated by using the force of mortality:

$${}_t q_x = 1 - \exp\left(-\int_x^{x+t} \mu_\tau d\tau\right).$$

The assumption of constant force of mortality means that $\mu_x = \mu$ in the investigated interval. In other words:

$${}_sq_{x+t}=1-(1-q_x)^s.$$

If we assume the constant force of mortality, we get a maximum likelihood estimator:

$$1-\exp(-d/E_x^c),$$

where E_x^c is called ‘central exposed to risk’ in the actuarial literature. The ‘central exposed to risk’ is the sum of observed times. It can easily be calculated: if somebody withdraws his contract, the calculation is the same as in the case of the initial exposed to risk, but if somebody dies, he stands in risk only until the time of death.

Calculating the estimated probability of death for the previous example:

– Central exposed to risk: (1/3+2/3+1)=2.
– Estimated probability: 1–exp(–1/2)=1–0.61=0.39.

As it was mentioned, in the case of maximum likelihood estimator, we assume a constant force of mortality. Therefore, if the constant force of mortality is appropriate then this estimator is asymptotically unbiased and asymptotically efficient.

Up to this point we wanted to derive an estimator which we can handle easily, this is the reason for using these assumptions instead of believing that the data follow these assumptions. In real life the probability of death is increasing with age[3] so it is a good hypothesis that the force of mortality also increases during the one-year time interval. Unfortunately, the force of mortality does not increase either in the case of Balducci assumption or in the case of constant force of mortality. It is necessary to investigate the behaviour of these estimators under more realistic assumptions such as uniform distribution of deaths[4] or a Gompertz mortality law.[5]

2. MONTE CARLO SIMULATIONS

In order to compare the properties of the former described estimators Monte Carlo simulation are used. Three scenarios for the probability of failure (1%, 5%, 30%), and three scenarios for the sample size (5, 30 and 1 000) are chosen. The uniform distribution of death does not have any parameter, so the situation in the case of these assumption is simple.

[3] This statement holds for all ages greater than 6 in Hungary, and holds for ages greater than 25 in western countries as well.

[4] Uniform distribution of death means that the (expected) number of deaths is the same for all intervals whose lengths are equal. Let us suppose that the probability of death is 0.1 and the sample size is 100 (for instance). It means that 10 deaths will occur (expectedly), 5 of them in the first half-year, 5 of them in the second half-year. The probability of death in the first half-year is: 5/100, and in the second half-year 5/95 (because 5 persons died in the first half-year, so there are only 95 persons alive at the beginning of the second half-year). We can see that the probability of death increases during the year.

[5] Gompertz described an expression for the force of mortality: $\mu_x = Bc^x$. $B > 0$, and B is close to 0, $c > 1$ and c is close to 1. In this case the force of mortality increases so the probability of death increases as well.

The Gompertz mortality law has two parameters but one of them (B) is irrelevant for our analysis.[6] The other parameter was fixed as $c = 1.1$.[7]

Two numbers are simulated for each person. One for failure time[8] and one for censoring time, and results are stored. We will investigate the described assumptions for failure time, but we will use the constant force of mortality for censoring time (to reduce the number of scenarios). The probability of censoring (the probability that a censoring event occurs within a year) is 10 percent (for the same reason). A censoring time and a survival time for each member of the sample will be simulated. So we can calculate simulated estimates. We will repeat it 100 000 times, so 100 000 simulated values for each estimates will be got.

For each estimator the mean and standard deviation will be presented and the estimators will be compared by means of the mean square error (MSE).

It has been mentioned that parametric estimators perform better because their variance is smaller. For this reason we start the simulation by comparing the Kaplan–Meier estimator and the actuarial estimator when the Balducci hypothesis holds for the failure times, then we will compare Kaplan–Meier estimator with the maximum-likelihood estimator when constant force of mortality holds for the failure times. The results can be seen in Tables 1,2,3 and 4.[9]

Table 1

Probability of death
(estimated by actuarial versus Kaplan–Meier estimators)

Sample size	True parameter	Actuarial estimator			Kaplan–Meier estimator		
		Mean	Standard deviation	MSE	Mean	Standard deviation	MSE
5	0.01	.00984106	.04687197	.0021970068	.00989700	.04761310	.0022670183
	0.05	.04951430	.10266833	.0105410215	.04998000	.10514971	.0110564611
	0.30	**.29686623**	.21397758	.0457962241	.29901983	.22096043	.0488244723
30	0.01	.01003072	.01929079	.0003721355	.01004699	.01938276	.0003756935
	0.05	.05005749	.04202622	.0017662061	.05012668	.04229819	.0017891531
	0.30	.29949622	.08721694	.0076070486	.29985433	.08878433	.0078826779
1000	0.01	.00999492	.00332315	.0000110434	.00999542	.00333146	.0000110986
	0.05	.05001927	.00727703	.0000529556	.05001958	.00730579	.0000533750
	0.30	.30005796	.01506360	.0002269155	.30006824	.01529334	.0002338910

Note: Numbers in bold mean that the mean differs from the theoretical parameter value at a 5 percent significance level.

[6] We have to know how ${}_t q_x$ is related to q_x. In case of Gompertz mortality law ${}_t p_x = \exp\left(-\int_x^{x+t} Bc^\tau d\tau\right) = \exp\left(-(B/\ln(c))c^x(c^t-1)\right)$. So $\ln({}_t p_x)/\ln(p_x) = (c^t-1)/(c-1) \Rightarrow {}_t p_x = \exp\left(\ln(p_x)(c^t-1)/(c-1)\right)$.

[7] This value is appropriate for the Hungarian life tables.

[8] In the simulation first a random probability is simulated (*rnd*). If this number is greater than the probability of failure then the person survives the interval, so the failure time is 1. Else a t value is sought so ${}_t q_x$ equals the simulated number (probability). This t will be the simulated failure time.

[9] When the last person died we defined the value of Kaplan-Meier estimator as 1. This event does not occur when the sample size is 30 or 1000. It occurs a few times when the sample size is 5 but the probability of death is small. It occurs in 0.5 percent of the cases when the sample size is 5 and the probability of death is 30 percent.

Table 2

t and p-values for testing means in Table 1

Sample size	Probability of death	Maximum likelihood estimator		Kaplan–Meier estimator	
		t	*p*	*t*	*p*
5	0.01	-1.072	0.284	-0.684	0.494
	0.05	-1.496	0.135	0.060	0.952
	0.30	-4.631	0.000	-1.403	0.161
30	0.01	0.504	0.614	0.767	0.443
	0.05	0.433	0.665	0.947	0.344
	0.30	-1.827	0.068	-0.519	0.604
1000	0.01	-0.484	0.628	-0.434	0.664
	0.05	0.837	0.403	0.847	0.397
	0.30	1.217	0.224	1.411	0.158

Table 3

Probability of death

(estimated by maximum likelihood versus Kaplan–Meier estimators)

Sample size	True parameter	Maximum likelihood estimator			Kaplan–Meier estimator		
		Mean	Standard deviation	MSE	Mean	Standard deviation	MSE
5	00.01	.00998304	.04590550	.0021073153	.01003533	.04634266	.0021476433
	0.05	.04978460	.09999569	.0099991834	.05005150	.10099568	.0102001306
	0.30	**.29751627**	.20653649	.0426634866	.30079983	.21285520	.0453079778
30	0.01	.01008313	.01874080	.0003512246	.01008447	.01875697	.0003518311
	0.05	**.05025774**	.04094795	.0016768008	**.05026694**	.04101747	.0016825041
	0.30	**.29936770**	.08510391	.0072430761	.29965704	.08605093	.0074048795
1000	0.01	.00999778	.00322633	.0000104092	.00999807	.00322852	.0000104234
	0.05	.05002049	.00704600	.0000496465	.05002204	.00705419	.0000497620
	0.30	.30003392	.01479451	.0002188786	.30004775	.01492556	.0002227747

Note: Numbers in bold mean that the mean differs from the theoretical parameter value at a 5 percent significance level.

Table 4

t and p-values for testing means in Table 3

Sample size	Probability of death	Maximum likelihood estimator		Kaplan–Meier estimator	
		t	*p*	*t*	*p*
5	0.01	–0.117	0.907	0.241	0.810
	0.05	-0.681	0.496	0.161	0.872
	0.30	-3.803	0.000	1.188	0.235
30	0.01	1.403	0.161	1.424	0.154
	0.05	1.990	0.047	2.058	0.040
	0.30	-2.349	0.019	-1.260	0.208
1000	0.01	-0.217	0.828	-0.189	0.850
	0.05	0.920	0.358	0.988	0.323
	0.30	0.725	0.468	1.012	0.312

Some conclusions can be made from Table 1 and Table 3. According to the theory there is a bias for each estimator (if the sample size is finite). In case of actuarial and maximum likelihood estimator we can verify this result for small samples and large probabilities. One cannot reveal this bias (with *t*-test) in case of Kaplan-Meier estimator. According to the theory all three estimators are asymptotically unbiased. It can be seen that the simulated estimates are getting closer to their true value as the sample size increases. Not only the bias decreases with the sample size, but also the *t*-value (in absolute terms) decreases (see Table 2 and Table 4), although this statement holds for large probabilities only.

The MSE for the parametric estimators are smaller than those for the Kaplan-Meier estimator. For small samples the Kaplan-Meier estimator is more precise, but as the sample size increases the bias becomes smaller in case of parametric estimators. However the standard deviation is smaller for the parametric estimators and it will result in that the mean square error is smaller for these estimators in any (of these) cases. In both cases the differences in efficiency are tiny.

After investigating the properties of the estimators, when the appropriate distribution was assumed, we can analyse how they behave when other distributions fitting better to real life practice are assumed. The results of this sensitivity analysis can be seen in Table 5 and Table 7, while Table 6 and Table 8 contain *t* and *p* values.

According to the results shown in Table 5 and Table 7, the Kaplan–Meier estimator can be considered as an unbiased estimator, while actuarial estimator and maximum likelihood estimator are biased. This bias becomes significant when both the probability of failure and the sample size are large. The bias is decreasing with the sample size (in case of actuarial and maximum likelihood estimators), but the *t*-statistics is increasing (in absolute terms), so we can conclude that there are significant biases in case of these estimators.

The standard deviations are smaller in case of actuarial and maximum likelihood estimators. The MSE for the actuarial and maximum likelihood are smaller than that for the Kaplan–Meier. However the parametric estimators are not unbiased any more, they are preferred to the non-parametric estimator (with respect to MSE), but the difference is tiny again. When the probability of death is small, the actuarial estimator is more efficient than the maximum likelihood estimator.

Table 5

Probability of death

(estimated by Monte Carlo simulations – uniform distribution of deaths)

Sample size	True parameter	Actuarial estimator			Maximum likelihood estimator			Kaplan–Meier estimator		
		Mean	Standard deviation	MSE	Mean	Standard deviation	MSE	Mean	Standard deviation	MSE
5	0.01	.01004973	.04603766	.0021194683	.01006415	.04615370	.0021301681	.01007950	.04635700	.0021489781
	0.05	.04941703	.09926364	.0098536101	**.04937001**	.09923540	.0098480609	.04969042	.10034615	.0100694465
	0.30	**.29758266**	.20853141	.0434911915	**.29278514**	.20245981	.0410420288	.30011650	.21253080	.0451693556
30	0.01	.01007813	.01872772	.0003507335	.01008056	.01873355	.0003509522	.01009136	.01876892	.0003522808
	0.05	.05016675	.04085613	.0016692513	.05017393	.04086155	.0016696964	.05023810	.04098634	.0016799370
	0.30	**.29839330**	.08518692	.0072593935	**.29619100**	.08362243	.0070072192	.30012069	.08619878	.0074302436
1000	0.01	.00999218	.00323165	.0000104436	.00999311	.00323222	.0000104473	.00999366	.00323391	.0000104582
	0.05	**.04993466**	.00706568	.0000199281	**.04994595**	.00706792	.0000499584	.04997909	.00708180	.0000501523
	0.30	**.29846589**	.01476271	.0002202910	**.29663580**	.01451429	.0002219826	.30004328	.01492269	.0002226885

Note: Numbers in bold mean that the mean differs from the theoretical parameter value at a 5 percent significance level.

Table 6

t and p-values for testing means in Table 5

Sample size	Probability of death	Actuarial estimator		Maximum likelihood estimator		Kaplan–Meier estimator	
		t	*p*	*t*	*p*	*t*	*p*
5	0.01	0.342	0.732	0.446	0.656	0.542	0.588
	0.05	-1.857	0.063	-2.008	0.045	-0.976	0.329
	0.30	-3.666	0.000	-11.269	0.000	0.173	0.863
30	0.01	1.319	0.187	1.360	0.174	1.539	0.124
	0.05	1.291	0.197	1.346	0.178	1.837	0.066
	0.30	-5.964	0.000	-14.404	0.000	0.443	0.658
1000	0.01	-0.765	0.444	-0.674	0.500	-0.620	0.535
	0.05	-2.924	0.003	-2.418	0.016	-0.934	0.350
	0.30	-32.861	0.000	-73.291	0.000	0.917	0.359

Table 7

Probability of death

(estimated by Monte Carlo simulations – Gompertz mortality law)

Sample size	True parameter	Actuarial estimator			Maximum likelihood estimator			Kaplan–Meier estimator		
		Mean	Standard deviation	MSE	Mean	Standard de-viation	MSE	Mean	Standard deviation	MSE
5	0.01	.00993604	.04561098	.0020803653	.00994494	.04571968	.0020902923	.00998842	.04607912	.0021232851
	0.05	.04977052	.09976533	.0099531747	.04968104	.09965294	.0099308098	.05002325	.10075283	.0101511326
	0.30	**.29766343**	.20906922	.0437154001	**.29520474**	.20511519	.0420952368	.29964883	.21270307	.0452427209
30	0.01	.00996515	.01861666	.0003465813	.00997956	.01865877	.0003466912	.00997956	.01865877	.0003481502
	0.05	.05000431	.04072089	.0016581909	.04999852	.04070856	.0016571870	.05009891	.04087626	.0016708781
	0.30	**.29891286**	.08491974	.0072125444	**.29871793**	.08441351	.0071272851	.30007567	.08581999	.0073650771
1000	0.01	.00998985	.00322569	.0000104052	.00998997	.00322570	.0000104052	.00999882	.00323103	.0000104396
	0.05	**.04993295**	.00704448	.0000496292	**.04993493**	.00704485	.0000496342	.04999675	.00706198	.0000498715
	0.30	**.29898562**	.01478491	.0002196227	**.29905454**	.01472266	.0002176507	.29998348	.01492052	.0002226223

Note. Numbers in bold mean that the mean differs from the theoretical parameter value at a 5 percent significance level.

Table 8

t and p-values for testing means in Table 5

Sample size	Probability of death	Actuarial	Estimator	Maximum likelihood	Estimator	Kaplan–Meier	Estimator
		t	*p*	*t*	*p*	*t*	*p*
5	0.01	-0.443	0.658	-0.381	0.703	-0.079	0.937
	0.05	-0.727	0.467	-1.012	0.312	0.073	0.942
	0.30	-3.534	0.000	-7.393	0.000	-0.522	0.602
30	0.01	-0.592	0.554	-0.582	0.561	-0.346	0.729
	0.05	0.033	0.974	-0.012	0.990	0.765	0.444
	0.30	-4.048	0.000	-4.803	0.000	0.279	0.780
1000	0.01	-0.995	0.320	-0.984	0.325	-0.116	0.908
	0.05	-3.010	0.003	-2.921	0.003	-0.145	0.885
	0.30	-21.696	0.000	-20.307	0.000	-0.350	0.726

In real life practice, the exact variances of the estimators are unknown, so they have to be estimated as well using the following variance estimators:

Kaplan–Meier: $(1-\hat{q})^2\left(\sum_{i=1}^{n}((n-i)(n-i+1))^{-1}d_i\right)$

Actuarial : $\dfrac{\hat{q}(1-\hat{q})}{E_x}$

Maximum likelihood: $(1-\hat{q})^2\left(d/\left(E_x^c\right)^2\right)$

In case of actuarial estimator the estimated variance is appropriate when there is no censoring event (binomial model). In practice it is common to say that this estimation also holds when there are censoring events as well. The estimated variance of maximum likelihood estimator is derived by using the Cramer–Rao lower bound. Estimated standard errors can be seen in Table 9.

Table 9

Estimated standard errors

Sample size	Probability of death	Actuarial estimator	Maximum likelihood estimator	Kaplan–Meier estimator*
		Uniform distribution of deaths		
5	0.01	.04063340	.04062139	.04076844
	0.05	.08851087	.08822361	.08898437
	0.30	.18657172	.18334180	.18753764
30	0.01	.01839034	.01839380	.01842987
	0.05	.04017906	.04017366	.04028098
	0.30	.08406086	.08312117	.08474512
1000	0.01	.00322649	.00322677	.00322877
	0.05	.00706318	.00706393	.00707581
	0.30	.01480634	.01465886	.01492114
		Gompertz mortality law		
5	0.01	.04041854	.04037235	.04064237
	0.05	.08870958	.08835855	.08916791
	0.30	.18640674	.18383433	.18713953
30	0.01	.01828577	.01828596	.01832763
	0.05	.04012449	.04010929	.04024556
	0.30	.08410439	.08355575	.08466530
1000	0.01	.00322613	.00322615	.00323082
	0.05	.00706322	.00706269	.00707853
	0.30	.01481115	.01473583	.01490333

* If the last person died, the expression of estimated variance is not correct (we have to divide by 0). In this case we set the variance 0. As it was mentioned earlier, it did not occur when the sample size is 30 or greater.

In Table 9 we can see that the variance is under-estimated for small samples. This bias decreases as the sample size increases, but it does not disappear.

Our further purpose is to give an interval estimation for the probability of death. So for each case a 95 percent probability interval will be calculated for the estimated prob ability, i.e. we have 100 000 confidence intervals for each scenario. Then we count the number of cases when the confidence interval does not contain the estimated parameter. If the estimation process is appropriate, the number of these cases has to be very close to 5000, that is their share must be close to 5 percent. Table 10 presents the results.

Table 10

Share of cases when the confidence interval does not contain the actual parameter

Sample size	Probability of death	Actuarial estimator	Maximum likelihood estimator	Kaplan–Meier estimator
		percent		
		Uniform distribution of deaths		
5	0.01	95.32	95.33	95.33
	0.05	78.59	78.60	78.60
	0.30	21.56	21.60	21.90
30	0.01	74.99	74.99	74.99
	0.05	23.30	23.31	23.30
	0.30	6.13	6.58	6.21
1000	0.01	9.63	9.62	8.72
	0.05	5.48	5.46	5.47
	0.30	5.17	5.61	5.01
		Gompertz mortality law		
5	0.01	95.35	95.36	95.36
	0.05	78.49	78.50	78.50
	0.30	21.63	21.92	21.95
30	0.01	75.21	75.21	75.21
	0.05	23.21	23.21	23.21
	0.30	6.08	6.62	6.12
1000	0.01	9.55	9.55	8.56
	0.05	5.36	5.37	5.38
	0.30	5.11	5.18	5.09

We can see that the results are very poor, especially for small sample sizes. This is an important reason why these estimators cannot be used for small samples. According to Table 10, maximum likelihood estimator achieves the worst results. If the probability of death is small (which is the most relevant case in insurance problems) and the sample size is large enough, the performance of Kaplan–Meier estimator is the best. In case of Kaplan–Meier estimator the confidence interval is wider (since the estimated variance is larger), but this wider interval adheres more to the facts.

3. CONCLUSION

All these three estimators have almost the same properties. The parametric estimators are more robust, i.e. they perform well when the inappropriate distributions are assumed. They are preferred to Kaplan–Meier estimator with respect to MSE, but the difference in efficiency is rather small. As we have seen, the confidence interval for Kaplan–Meier estimator is wider, but this wider interval covers the true value more frequently. The differences in frequency are small again. If the sample size tends to infinity, the MSE for Kaplan–Meier estimator tends to 0 (since it is a consistent estimator), but the MSE for parametric estimators keeps to a positive value, since there is a (significant) bias in this case. Based upon these results the use the Kaplan–Meier estimator can be suggested in life insurance statistics.

REFERENCES

ÁGOSTON, K. – KOVÁCS, E. (2000): *Halandósági modellek*. Budapesti Közgazdaságtudományi és Államigazgatási Egyetem, Operációkutatási Tanszék. Budapest. Aktuárius jegyzetek 3.

BATTEN, R.W. (1978): *Mortality table construction*. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

BENJAMIN, B. – POLLARD, J. H. (1992): *The analysis of mortality and other actuarial statistics.* Butterworth-Heinemann Ltd. Linacre House, Jordan Hill, Oxford.

DORRINGTON, R. E. – SLAWSKI, J. K. (1993): A defence of the conventional actuarial approach to the estimation of the exposed-to-risk. *Scandinavian Actuarial Journal*, p. 107–113.

HOEM, J. M. (1984): A flaw in actuarial exposed-to-risk theory. *Scandinavian Actuarial Journal*, p. 187–194.

KAPLAN, E. L. – MEIER, P. (1958): Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, Vol. 53. p. 457–481.

Lee, E. T. (1992) : *Statistical methods for survival data analysis.* John Wiley & Sons, Inc. New York, Chichester, Brisbane, Toronto, Singapore.

MACDONALD, A. S. (1996): An actuarial survey of statistical models for decrement and transition data. I: Multiple state, Poisson and binomial models. *British Actuarial Journal*, No. 2. p. 129–155.; II: Competing risks, non-parametric and regression models. *British Actuarial Journal*, No. 2. p. 429–448.; III: Counting process models. *British Actuarial Journal*, No. 2. p. 703–726.

ISSN 0039 0690

Published by the Central Statistical Office
Editor in charge: dr. Tamás Mellár
Executive editor: dr. László Hunyadi
Printed by the Akadémiai Nyomda
3340 – Martonvásár, 2001
Director in charge: Lajos Reisenleitner

Managing editor: Orsolya Dobokayné Szabó
Editors: dr. Attila Domokos, Andrea Polyák, Mariann Szűcsné Bruckner, Mária Visi Lakatos
Technical editors: Éva Bálinthné Bartha, Ágnes Simonné Káli

Editorial office: Budapest II., Keleti Károly utca 5–7. Postal address: P.O.B. 51. Budapest, 1525.
Phone: (361)-487-4343, Telefax: (361)-487-4344,
Internet: www.ksh.hu/statszml E-mail: statszemle@ksh.gov.hu
Publishing office: Central Statistical Office, Budapest II., Keleti Károly utca 5–7.
Postal address: P.O.B. 51. Budapest, 1525. Phone: (361)-345-6000
The publication can be purchased at the Statistical Special Bookshop:
Budapest II. Keleti Károly utca 10. Phone: (361)-212-4348