The Polish LFS: A Rotating Panel with Attrition

Francesco Pastore, Professor
Seconda Università di Napoli, Santa Maria Capua Vetere (Caserta), Italy and IZA, Bonn
Mieczysław Socha, Professor
Department of Economics, University of Warsaw

Introduction

A large literature on Poland's economic transition has been based on the Polish Labour Force Survey (PLFS), which is the primary labour market data collection effort of the *Główny Urząd Statystyczny* (GUS, the Polish Central Statistical Office). The PLFS has a longitudinal structure, belonging to the family of rotating panels with a 2–2–2-rotation scheme, and covers the period from 1992 until now.

In addition to the usual sample attrition, longitudinal data is also affected by panel attrition. This is the loss of observations in the matched data file relative to two consecutive surveys. *Ex ante* causes of panel attrition include residence changes, refusals to answer the questionnaire due, for instance, to sampling (or panel) fatigue[1] and, to a lesser extent, the death of the interviewee. Measurement and reporting errors can also cause *ex post* failures to match observations in two consecutive surveys.

Panel attrition could generate sample selection bias, if the (observed or unobservable) characteristics of the individuals selected in the matched sample systematically differ from the rest, therefore undermining also the degree of representativeness of the data. Assume, for instance, that attrition is particularly frequent among the unemployed. This implies that the unemployment stock will appear smaller than it actually is in any sample survey. In turn, the flows in and out of unemployment will be over- or underestimated, according also to the relative frequency of attrition among job finders, on the one hand, and job losers and quitters, on the other hand. When attrition is not systematic, it can be ignored. Systematic attrition due to observed characteristics can be cured using well defined weights, whereas that due to unobservable characteristics can be cured using sample selection correction econometric procedures (Lindeboom and Van den Berg, 1998; Dolton, Lindeboom and Van den Berg, 2004).

No analysis of panel attrition in the Polish LFS exists in the literature. This paper aims to partly fill this gap in three ways. First, it reckons that attrition amounted to about 6.5% in the years 1995–'96. This figure is much

---

[1] Filer and Hanousek (2002) argue that non response is one of the main problems of data relative to transition countries, due to the allergy of people to telling the authorities anything.

smaller than that typical of pure panels and similar to that of surveys with a similar design carried out in other countries. Compared to pure panels, the PLFS partly corrects for attrition with its rotating design. Secondly, the paper shows that panel attrition can cause biased estimates of transitions among labour market statuses, leading, in particular, to underestimate the extent of numerical flexibility.

Thirdly, this paper shows that albeit limited in the PLFS case, panel attrition is systematic. The results of a logistic estimate of the determinants of the probability of an observation not to be missed suggest that overall such probability is not randomly distributed. Significant differences arise especially across regions and labour market statuses. Age seems to be the most prominent determinant of attrition and the age profile of attrition is u-shaped. Moreover, men with a high level of educational attainment and residing in low unemployment regions have higher probability of non-response. However, attrition is very low among prime-aged workers. These results are similar to those obtained in Paull (1996) in a study carried out on the British Household Panel Survey and confirm that special caveats should be taken when studying labour market transitions using longitudinal data.

The remainder of the paper is as follows. Section 1 provides an introductory description of the data set. Section 2 deals with the specific structure and design of the PLFS. Section 3 focuses on measurement errors as a specific source of bias. Section 4 gives a definition of attrition and discusses its causes and consequences for the analysis of the PLFS. Section 5 discusses statistical and econometric cures for attrition. Some summary remarks follow.

1. The Polish LFS

1.1. Origin, definitions and period covered

The PLFS is administered to 55,000 individuals circa, representing about 0.17 per cent of those aged 15 or more. The household is the unit. The GUS claims that the sample is representative of the Polish population, namely of its spatial and demographic distribution (Witkowski and Szarkowski, 1994). Interviews have been conducted during the third week of the middle month

Table 1.
The quasi-panel design of the Polish LFS

A93	N93	F94	M94	A94	N94	F95	M95	A95	N95	F96	M96	A96	N96
6	6	–	–	6	6								
	7	7	–	–	7	7							
		8	8	–	–	8	8						
			9	9	–	–	9	9					
				10	10	–	–	10	10				
					11	11	–	–	11	11			

A93	N93	F94	M94	A94	N94	F95	M95	A95	N95	F96	M96	A96	N96
						12	12	–	–	12	12		
							13	13	–	–	13	13	
								14	14	–	–	14	14
									15	15	–	–	15

Note: Each number represents a cohort of individuals. It is possible to follow their participation to the survey along the rows. Each column provides a snapshot of the composition of each survey carried out from May 1992 to November 1997.
Source: own elaboration.

of every quarter of a year, in February, May, August and November, starting from May 1992. The survey has a longitudinal structure. It was organised as a pure panel in the first waves and has become a typical rotating or rounding panel with a 2–2–2 rotation scheme from May 1993 (Table 1).

As noted in Witkowski and Szarkowski (1994) and in Socha and Weisberg (1999), the PLFS data follows the general rules and definitions recommended by the ILO-OECD. This would allow internationally comparable statistics, making it possible to apply most of the available techniques of analysis. For instance, unemployment is not defined simply as joblessness, as it was the case in the pre-transition era. Unemployed are those jobless workers actively seeking a job during the last four weeks of the survey and available to take a job in the reference week of the survey.

Furthermore, starting from May 1994, also sectors and occupations are defined according to the OECD classifications and become more detailed, up to 32 sectors (R.25 NACE)[2] and to over 100 occupations (ISCO–88). The classification of services includes also non-tradable goods and services[3].

1.2. The aims of the PLFS

Most CEECs and former Soviet Union republics decided to collect LFS data in the aftermath of transition (see Filer and Hanousek, 2002, Table 2). They served the need for reliable and internationally comparable information on short term changes affecting the labour market, the place where the most dramatic social effects of the transition recession were expected to happen[4] (Góra *et al.*, 1993). In fact, all the predictions made before and during the first years of transition agreed that the restructuring process would have produced high and persistent unemployment, a remarkable reduction in activity rates and increase in non-participation (Aghion and Blanchard, 1994; Góra, 1994; Boeri, 1994; Svejnar, 1999, p. 2815).

---

[2] Consider that the sectors are over 80 in Italy's and over 200 in the UK surveys.

[3] This is a more detailed level of aggregation than that adopted over the period 1992–'94, but it is still insufficient for some important purposes. Moreover, the difference with the previous definitions causes unrecoverable break in the series.

[4] A further incentive was the fact that satisfactory and reliable time series data would have been available only after at least a decade.

To be dealt with, these issues needed a complex, complete, consistent and flexible data source. Pudney (1993) outlines the importance of survey data in transition countries. The alternative to the LFS could have been administrative data. In the case of Poland, this includes also unemployment registration and direct employment reporting provided by Labour Offices (Góra, 1994). Nevertheless, as noted in Kemp (1991), data generated from administrative sources are often highly inaccurate, inconsistent and based on definitions, which do not correspond to those used in economic analysis.

The slovenliness of administrative data would have been especially likely in the case of unemployment and labour market transitions[5]. Table 2 shows that administrative data tends to overestimate the unemployment rate with respect to LFS data, among other reasons, because at least some workers employed in the grey economy tend to declare their activities to LFS interviews, but not to labour offices[6, 7]. Official statistics tend to overestimate also the actual duration of unemployment due to the tendency of unemployed workers to remain registered after finding ILO employment not to lose eligibility to unemployment benefits. Although, they often leave the unemployment registers lose when unemployment benefits expire, even if they have not found any job (Góra, 1994; OECD, 1997; Adamchik and King, 1999).

Table 2.

Registered and survey unemployment (in per cent, end of the year)

	1992	1993	1994	1995	1996	1997
Registered unemployment rate	13.6	16.4	16.0	15.2	14.3	11.5
LFS unemployment rate	13.7	14.9	14.6	14.4	12.7	11.0
Structure of unemployment by duration						
Long-term unemployed (more than 12 months)						
Registered	45.2	44.8	44.2	37.4		
LFS	39.6	35.8	41.6	39.9	40.5	33.8
Very long term unemployed (more than 24 months)						
Registered			19.9[a]	20.0	17.7	
LFS	12.9	13.4	18.9	19.2	17.4	11.5

Note: a March 1994.
Source: Data on registered unemployment and on LFS relative to the period from 1992 to 1994 is from OECD (1997); data on registered unemployment up to 1997 is from OECD (2000); the rest is own elaboration on PLFS data.

---

[5] As noted in Filer and Hanousek (2002), the Polish is one of the few LFS in transition countries that collects wage data. However, it is questionable whether information provided in the LFS on wages are more reliable than those based on administrative data.

[6] Nonetheless, LFS data still tends to underestimate the real amount of informal employment.

[7] Sestito (1990) analyses the discrepancies between measures of unemployment obtained from administrative and labour force survey data in the Italian case, where notoriously the size

Finally, before the introduction of the LFS, flow data were almost non-existent or were of low quality and relative only to few industries in the state sector (Góra *et al.*, 1993; Boeri and Sziraczki, 1993; Socha and Sztanderska, 1997).

1.3. The shortcomings

The advantages largely offset the limits of the PLFS, which are partly typical of any longitudinal panel study. Here is a list of such shortcomings.

The survey starts in May 1992. This deprives the researcher of important information about the most dramatic phase of economic transition that is the years immediately after the Big Bang[8]. This is of much detriment, as, among other things, 4.6 million workers circa lost their jobs and the participation rate dramatically shrank from 1989 to 1992. A track of the dramatic changes happened in the early stages of transition might be obtained from retrospective life history data provided in the PLFS. For instance, Lehmann and Woodsworth (2000) study the effects of worker reallocation on job tenures in Poland, compared to other countries, also using retrospective information.

The discontinuity in the questionnaire adopted over the years undermines the possibility of carrying out studies comparable over time of some important variables. Two main methodological breaks affect the data: in May 1994, the most eminent change regards the classification of sectors and occupations; further minor modifications were made to the questionnaire in 1997. Overall, since May 1992, five versions of the questionnaire have been adopted.

The degree of sectoral disaggregation of data is low. Also in the post–1994 surveys, the degree of sectoral and occupational disaggregation is not sufficient to many important analytical purposes. In fact, major structural change has taken place within industries. However, a trade-off exists here: on the one hand, the higher the degree of data aggregation, the lower the possibility to register the changes that occurred; on the other hand, the higher the degree of disaggregation, the lower the reliability of data[9].

----

of moonlighting and of the unofficial economy is conspicuous. He reckons that, as national account data measures labour supply in terms of standard units of labour rather than work positions, factors such as moonlighting and migration generate a 10 per cent gap in the unemployment rates computed using those different sources of information.

[8] Until recently, there was a substantial lack of statistical information on firms as well. All the available studies were based either on anecdotal evidence or on interviews to firm managers (Dyker, 1996) or on case studies (see, for instance, Pinto *et al.*, 1993; and Estrin *et al.*, 1995). Administrative data covers the early 1990s. They have provided the material for important studies (see, among other, OECD, 1994). Nonetheless, administrative data are scantly reliable, especially in periods of major structural change. Recently, new data sources on firms at a quite detailed level of aggregation (up to 3digit) have become available, based on elaboration of administrative data, allowing new insights into the crucial years of the early 1990s (see Barbone *et al.*, 1999).

[9] Keeping constant the sample size, an excessive level of disaggregation reduces the statistical significance of the variables obtained and increases the share of classification errors. This is the case, for instance, of occupations.

There are only five classes of firm's size. Especially noticeable is the loss of information relative to firms with more than 100 employees, where most changes were likely to happen because of privatisation. As noted in Blanchard (1994), privatisation in large state-owned and cooperative enterprises almost never involved downsizing to less than 250 employees.

Non response. Some questions either have a relatively low response rate, especially on wages, or are inconsistently answered, as it is the case, for instance, for firm ownership. More generally, classification errors, measurement errors and attrition are much worrisome.

Missing information. Information is missing, for instance, on smaller regions and on trade union membership (Socha and Weisberg, 1999).

2. The quasi-panel nature of the PLFS

The PLFS is a specific purpose micro-economic study, as it collects information on the labour market status and history of individual workers aged 15 or more. This class of individuals represents the unit of analysis[10]. The survey, elicited quarterly, has a longitudinal structure, since it is based on interviews carried upon a large number of agents over a period of fifteen months.

The longitudinal dimension is that typical of the family of *rotating panels*, also called *quasi-panels*, due to the very short period of time, one and a half year in this case, on which information is collected on the same individuals. At any point in time, information is elicited on four cohorts of agents. The cohorts sampled are organised in waves, which start at a given point in time and remain in the survey for a limited period of time before exiting and being substituted by new cohorts[11]. Each cohort remains in the survey for the time needed to carry out six observations, although, as noted later on, each individual is interviewed only four times.

Moreover, variables may be computed using the information provided in the survey, in such a way to make up different panels, some of which with a temporal dimension. In this case, the location, the sector or the occupation of the agents may give the cross-section dimension. Also panels with no temporal dimension, such as cross-sections of cross-sections, may be obtained from the survey at any point in time. Nonetheless, in these cases, the number of observations dramatically reduces.

---

[10] The sampling procedure is based on households though.

[11] Rotating panels have *comparative advantages* relative to repeated cross-sections of individuals changing from one survey to the other. Even if from the latter it is possible to build a panel, either using retrospective questions or computing the values of aggregate variables over time, nonetheless, some important properties of the former, such as the possibility of computing flow variables, are lost. Surveys based on a rotation scheme have also important *absolute advantages* relative to repeated cross sections of individuals. In fact, the former reduces: a) the costs of the survey, since it makes possible to renew 25 per cent of the sample only at any interview; b) and provides more efficient estimates of the variables of interest. Nonetheless, this last hypothesis has not been fully verified yet.

2.1. Structure and design of the PLFS

Table 1 shows the structure and design of the PLFS. Four features are worth mentioning.

At any time, four cohorts are included in the survey in such a way that a cohort of individuals is interviewed for the first time; a second cohort was already interviewed three months earlier; a third cohort was interviewed two times, nine and twelve months earlier; a fourth cohort was interviewed three times, of which the first time three months earlier, the second time twelve months earlier and the third time fifteen months earlier.

Every cohort is interviewed four times within fifteen months. According to the so-called 2–2–2 scheme, the individuals belonging to each cohort are interviewed two consecutive times when they enter the survey; then, they exit the survey for two consecutive quarters; and, finally, they are interviewed again two more times before definitely going out of the survey[12].

If we compare two points in time far from each other exactly one year, two cohorts are common. For instance, in November 1994 and November 1995, the tenth and eleventh cohorts are common.

Comparing two subsequent points in time of the survey, it is possible to find two common waves of the labour force, one entering the survey for the first time and a second one ready to exit the survey. This feature allows the study of quarterly transition rates.

As it should now be clear, four different types of panel are possible with the rotation scheme adopted. Panel one would include an individual cohort (25% of the sample), followed over its entire participation to the survey and hence observed 4 times. Panel two can be obtained combining the two cohorts common to two following quarters (50% of the sample) and would be based on two observations. Panel three can be built combining the two cohorts common to two surveys (50% of the panel) one year apart. Panel four is made up adding to the two observations one year apart a further observation for each cohort obtained in the period in between. The third observation would refer to two different interviews for the two cohorts. An example will clarify this point. Cohorts 10 and 11 are common to the surveys of November 1994 and November 1995. During this period, the two cohorts were surveyed again, in August and in February 1995 respectively. A panel could be based on the three observations and still include 50% of the sample.

Each of the aforementioned features of the enquiry accommodates with a specific purpose. The last three features specifically reflect the panel nature of the data set, whereas the first feature is aimed at maintaining the representativeness of the sample population at any point in time, which is after

---

[12] Many European countries adopt a 2–2–2 structure for their LFS. Other countries adopt a different scheme. The American CPS follows a 4–8–4 scheme, whereas in the case of the Canadian LFS, every cohort stays in the survey for six consecutive months before exiting it definitely.

all the main aim of a LFS. The sampling rule is decided to ensure that such correspondence be as close as possible. Thanks to this feature, the LFS provides more accurate and reliable static measures of many important variables than those obtainable in the case of a pure panel[13]. This is the most important advantage of sampling schemes based on the overlapping of some groups of individuals over schemes with no overlapping.

In the case of data with a panel structure, two important factors inherent the data generating process itself may undermine the accuracy of the stock estimates of many variables of interest for economists, such as employment and unemployment:

— *attrition* will be discussed at length in section four[14]. It typically affects any type of longitudinal data, and especially pure panels.
— *the ageing of the sample population* generates natural, continuous flow of workers from one state to another of the labour market over time.

An example will help me clarify this last point. Suppose the share of individuals aged 15–30 represent 20% of the population. Assume also that as many as 40% of them are unemployed, 30% are employed and 30% are not in the labour force, as they are in education, in training or in search for their best job[15]. The longer the period between two consecutive interviews the higher the share of young workers in the sample who have found or are in search for a job, simply because they are ageing. Some of them have got their degree and start seeking a job. A smaller portion has already found a job, soon after college. Others, instead, have found their job to be below their expectations and have decided to search for another one or have gone back to education or training. All these transitions are typical of the labour market behaviour of young workers (Clark and Summers, 1982), but they may alter the true stocks, if the opposite flows don't cancel out each other. Usually, this is not the case, since with time passing, the number of those finding a job is generally higher than that of those losing a job. Similarly, the number of those flowing into unemployment from non-participation is higher than that of those flowing the other way around. One could express this phenomenon observing that age is a time varying covariate. In other words, the changes in the shares of the labour force are not happening in the underlying population. They are simply caused by the fact that with time passing some individuals in the survey sample have become older than 30 and the number of those aged 15 has decreased. Rotating panels provide a partial solution to this

---

[13] An important example of survey carried out with the structure of a pure panel is the British Household Survey. In this case, the longitudinal dimension is given by a group of individuals interviewed once a year for many years.

[14] Albeit similar in its effects, the loss of information due to attrition is conceptually different from that inherent the matching procedure. The issue will be analysed later on in this section.

[15] The example could go further breaking down, for instance, the sub-sample of young workers by sector of activity, occupation and so on. Such details would not change the point to be made here.

problem, maintaining the composition of the sample more stable than in a pure panel.

The effects of attrition are similar. As an example, consider a cohort entering the survey at a given time. If the sample is random and representative of the underlying population, one may obtain unbiased measures of key variables, say employment, at that time. However, if we continue to interview the same cohort, the longer is the time length, the higher is also the number of agents who exit the survey. Thus, because of attrition, not only the measure of employment will be biased. Also the composition of the sample changes in such a way to undermine its capacity to represent the underlying population. For the same reasons, attrition could affect not only stock, but also flow variables.

2.2. The matching procedure

The databases relative to different points of observation can be merged in order to catch the two cohorts common to the surveys. The *matching (or merging) procedure* is based on a variable, usually obtained as a linear combination of other variables, able to identify each and every individual of the survey, also called *identity variable*. The procedure requires that the variables to be combined should be time invariant. Examples of such variables are demographic or individual characteristics, such as gender, birth date, civil status, education and so on. The identity variable is used to detect the contemporaneous presence of an individual in two different point observations. The main shortcoming of this procedure is that large data sets are often affected by reporting errors. Moreover, there are individuals with similar characteristics in the survey.

In the case of the PLFS, the identity variable is based on the rank number attributed to each individual and a province (*voivodship*) code. The criterion adopted is hence of a deterministic type, since it is aimed at catching all and exactly the same individuals participating to both surveys. It is not error free, though, as reporting errors and attrition still affect the data, as shown in a later section. When the identity variable is not available, a probabilistic procedure is needed. In this case, the share of errors can be partly controlled by the researchers[16]. Two types of errors can be defined:

— A *negative error* happens when two observations relative to the same individual do not match. In two cases, the match can be missed. *Type one* of missing match is due to negative false. When the data relative to an interviewee, for instance the region or the rank number, is misreported in any

---

[16] In fact, the availability of a rank number makes it unnecessary to compute an identity variable. The procedure used by Favro-Paris, Gennari and Oneto (1996) for the Italian case is *probabilistic* and is based on an identity variable obtained as a linear combination of variables, such as the region, the province, the local authority, the rotation group, the family code, sex and the birth date. A restrictive law on privacy prevents the Central Statistical Office from using individual codes.

of the two surveys, then the match does not happen. Type two of negative false happens because of attrition or non-response.

— A *positive error* happens when the information relative to two different individuals is matched, as they end up with the same identity number. This may be due either to the fact that the information relative to one of them is ill reported, as in the case of a deterministic procedure, or to the fact that the two individuals have the same characteristics, at least those used to compute the identity variable, as in the case of a probabilistic procedure.

The consequences of the errors due to the matching procedure are very similar to those due to attrition. First of all, the loss of observations reduces the efficiency of the estimates. Secondly, if the errors are not randomly distributed, there is a possibility that bias affects the estimates, as the estimated parameters could catch the error probability rather than behavioural rules.

In section four, I will attempt to analyse the distribution of missing observations in the Polish data to verify whether it correlates with that of relevant variables, thus undermining the econometric results. Before then, I will discuss the case of measurement errors in longitudinal data.

3. Measurement errors in longitudinal data

Measurement errors may seriously undermine the effectiveness of the matching procedure. Socha and Weisberg (1999) reckon measurement errors in the PLFS are a major concern for the Central Statistical Office. They may essentially arise because of five factors. Firstly, *the participants could misunderstand the questionnaire.* It is important that the questionnaire be clear and well understood by the respondents. The issue is crucial in the case of transition countries. The definition of unemployment is the typical example, as that inherited by citizens used to living in a formerly socialist country where full employment was enforced by law is different from that adopted in Western economies. Another example is reported in Filer and Hanousek (2002): over transition there has been much confusion on firm's ownership. A firm was often considered private, although a share of only 10 per cent of its capital was privately owned. This type of error can generate time inconsistencies in the answers to principal questions and, in some cases, force the investigator to drop cases out, unnaturally altering stock and flow measures.

A further source of error is the *inaccuracy of memory recollection*, which is especially serious in the case of retrospective questions. Memory shortcomings especially affect duration variables.

Thirdly, there is also a phenomenon of *response conditioning*, sometimes called *response variability*. This consists of the fact that the respondents learn about the questionnaire after answering it more than once and tend hence to "adjust" their answers over time. As a consequence, the answers provided by individuals belonging to different cohorts, but interviewed at the same time tend to be partly different, simply because some individuals have already been interviewed and are more aware of the meaning and the aims of the sur-

vey. Response conditioning can produce artificial changes in the stock and flow variables.

Fourthly, the larger is the number of individuals included in the survey, the higher is the probability of *miscoding of answers by the personnel conducting the study*. Coding errors are typical not only of longitudinal, but also of cross-section data and according to Griliches and Hausman (1986, quoted in Kemp, 1991) are, in fact, less conspicuous in the former than in the latter case.

Finally, there are *missing answers*. Attrition can be regarded as a particular type of missing observation. Also in this case, dropping observations out means that the sample loses its ability to represent the underlying population.

The effects of all these sources of error are similar to those of attrition and will be considered in the following section.

4. Attrition

This section provides a definition and studies the consequences of attrition in the PLFS case. First, I discuss the nature and possible consequences of attrition in general. Then, I provide evidence on the size and the distribution of attrition in the PLFS, using the November 1995–'96 rounds. The analysis suggests that albeit limited, attrition is systematic and depends, among others, on some demographic variables, such as age, gender, residence, education. Age seems to be the most important of these factors, suggesting that attrition is less serious among prime-aged workers.

4.1. Definition and typical consequences

Attrition can be defined as the natural and systematic tendency of every cohort to change over time, because of a change of residence, the refusal to answer to further interviews due to sampling (or panel) fatigue, or death of some individuals in the sample. Reporting errors are a further factor of attrition. Attrition is a typical feature of any longitudinal data set, although it arises also in (repeated) cross-section studies, because of sample non-response. In pure panels, attrition is particularly difficult to deal with, because it cumulates from one survey to the other. Moreover, the longer is the time interval between two interviews, the higher is the rate of attrition.

In principle, it is possible that the distribution of attrition among different subgroups of the population be purely random. In this case, as explained in Hsiao (1986), attrition is simply *scaling down* the sample size, so to reduce the efficiency and the power of the tests, but not the consistency of the estimated measures of central tendency and dispersion of key variables. Nonetheless, more frequently, attrition is differential, affecting stock and flow variables, if those selected for the panel systematically differ from those excluded. Generally speaking, the category of respondent whose rate of non-response is

higher can become under-represented within the sample. It is hence funda-
mental also for economists to study the determinants of attrition.

The components of attrition can be randomly and/or non-randomly dis-
tributed. While *measurement errors* are more likely to be random, sampling
fatigue, residence change and the death of respondents can hardly be
thought as independent of other variables. *Sampling fatigue* usually accounts
for a major share of attrition. Evidence exists that young people, especially
men, tend on average to be less accurate in responding to questionnaires
(Dex and McCullock, 1997). *Residence changes* can be due to personal or work
reasons[17]. Respondents who change their address, without leaving any track
of the new address, are not interviewed anymore. There is much evidence to
suggest that residence changes due to labour mobility tend to be more fre-
quent among the youngest and the best-educated male segments of the sam-
pling population. The *death of the interviewee* is more frequent among old
workers. All this considered one would expect attrition to be less common
among prime-aged least educated women.

4.2. The size of attrition in the PLFS

For the first four quarters, from May 1992 to November 1993, the survey
was administered to the same individuals as in a pure panel. The conspicu-
ous share of attrition (10%) convinced the Central Statistical Office to intro-
duce a rotation scheme in May 1993. Socha and Weisberg (1999, p. 17) report
that in the first year of the survey attrition was mainly due to non-response,
either because of refusal to answer the questionnaire or of inability to locate
the respondent. There were large differences between large cities—where
attrition was sizeable, with a maximum of 29.4 per cent in Warsaw—, and ru-
ral areas (3.7 per cent on average).

Góra and Lehmann (1995) estimate that attrition amounted to an almost
constant share of 7.5 per cent circa of the sample when matching the May's
rounds of the PLFS relative to the years from 1992 to 1994. According to the
Authors, attrition bias is almost irrelevant, due to the small scale of the phe-
nomenon.

The file obtained merging the November 1995 and the November 1996
rounds of the PLFS accounted for a total number of 25,459 observations out of
54,469. As about 50% of the cases should be common to the two waves, one
would expect to find 27,234 cases circa in the matched file. This suggests that
about 6.52 per cent of the cases were lost in the matching procedure for vari-
ous reasons (Table 3)[18].

---

[17] Usually the questionnaire includes a question to assess the reason why the interviewee
left the survey. In case of change of residence, no further enquiry is done to understand whether
it was due to personal or work reasons. Including such a question could provide important in-
formation on the determinants of labour mobility within a country.

[18] Similar shares were found relative to the previous and following year. The results are
available from the authors on request.

Table 3.

**The overall attrition rate in the November 1995 and November 1996 rounds of the PLFS**

	Observations	%[a]
Total number of observations in November 1995	54,469	100
Expected number of observations in the matched data file	27,234.5	50
Actual number of observations in the matched data file	27,183	99.81
Number of respondents selected in the in both samples	25,459	93.48
Non response due to misreported rank code	51	0.19
Non response due to attrition	1724	6.33

Note: a The figures in the first two rows refer to the total number of observations. The figures in the following four rows are shares of the expected number of observations in the matched data file.
Source: own elaboration on Labour Force Survey data.

4.3. The determinants of attrition

Paull (1996) noted that age is the most common and important determinant of attrition in the BHPS. Fig. 1 shows that the distribution of attrition by age is u-shaped also in the Polish case. It peaks for the individuals aged 20–29 and then declines gently down to the age of 60–64, when it raises sharply again. This distribution shows little difference across groups of regions with a different unemployment rate. It is likely that changes of residence and sampling fatigue are the reasons of the high attrition rate among young workers, whereas death is the reason among the over–60.
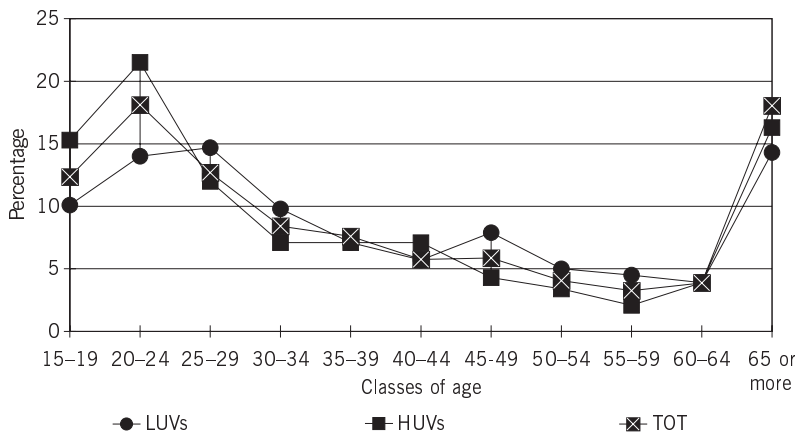
Fig. 1.

Age profile of attrition by group of regions

Note: LUVs and HUVs indicates the groups including the voivodships with the lowest and those with the highest unemployment rate in 1994. Each group represents about one third of the population.
Source: own elaboration on Labour Force Survey data.

Attrition is remarkably more frequent among unemployed than among employed individuals (Table 4). This is probably due to the tendency of the former group to change residence more frequently and to have a higher rate of panel fatigue. Strangely enough, though, the regions with the highest unemployment rate have a lower share of non-response than the regions with the lowest unemployment rate. A possible explanation of this finding is the concentration of large cities in the group of low unemployment regions.

Table 4.

Attrition by group of regions and labour market status

Labour Market States	LUVs		HUVs		Total		
	Total	Attrition	Total	Attrition	Total	Attrition	Average
Employed	52.2	51.6	46.9	36.1	49.8	44	5.6
Unemployed	6.6	10.5	11.6	22.4	9.1	16.5	11.5
Inactive workers	41.2	38	41.5	41.5	41.2	39.5	6.1
	100	100	100	100	100	100	
Average		7.6		5.5		6.3	

Note: LUVs and HUVs indicate groups of voivodships with the lowest and with the highest unemployment rate respectively in 1994. Each group represents about one third of the population. Source: own elaboration on Labour Force Survey data.

Table 5 shows the composition of panel attrition by levels of education attainment and groups of regions. It is clear that individuals with a higher educational level have also a higher probability of quitting the survey. This is probably due to residence move for work reasons. The small group of individuals who have not completed primary school is the only exception: they have a higher than average rate of non-response.

Table 5.

Attrition for classes of individuals with different education attainment

Education attainment	LUVs		HUVs		Tot	
	Tot	Attrition	Tot	Attrition	Tot	Attrition
University	7.9	10.8	5.9	5.2	6.8	7.3
Post-secondary	2.3	9.3	2.3	5.3	2.4	6.7
General secondary	18.8	8.0	17.9	5.1	18.0	6.5
Vocational secondary	7.4	8.2	6.7	6.7	6.8	7.3
Low vocational	27.9	8.0	25.3	6.6	26.3	7.2
Primary	32.3	5.6	37.3	4.7	35.2	4.9
Below primary	3.3	11.9	4.7	7.3	4.5	9.3
Average		7.6		5.5		6.3

Source: own elaboration on PLFS data.

Table 6 and 7 show an important, but under researched issue when studying the consequence of attrition, namely its effect on labour market stocks and flows. Table 6 reports the flow from and to every labour market status as a percentage of the overall sample. Notice also that all the numbers are smaller in the case of transitions computed without taking into account attrition, simply because they are based on a smaller sample. The column "with attrition" cannot take into account the contribution of attrition to each flow, because it is impossible to know whether those who do not answer the questionnaire: a) are staying into their origin status or are moving to another status; b) and to which status they are actually moving. The figures suggest that the effect of sample non-response on the change in the percentage of employed and unemployed workers is almost negligible in percentage terms, but remarkable in terms of absolute numbers. Consider, in fact, that a percentage point represents a difference of thousands of individuals. The stock of the employed is overestimated, while the stock of the unemployed is underestimated. Considering that panel attrition is relatively greater among the latter group, it is likely that the stock of unemployment would be larger considering also the attriters.

Table 7 provides labour market transition matrices with and without attrition. All the transitions are lower in the latter case. In principle, without detailed information on the final status of attriters, it would be impossible to say which flow is under- and which is over-estimated. However, the transition probability in and out of each labour market status is greater, if panel attrition affects with relatively higher frequency those individuals who are changing their labour market status. In other words, panel attrition might give an impression of lower than actual labour market flexibility. This is suggestive of the need to take the due caveats when using transition analysis to measure labour market flexibility.

Table 6.

Labour market transitions with and without attrition in low and in high unemployment regions (1995–'96; November)

Gross flows	LUR		HUR	
	Without Attrition	With Attrition	Without Attrition	With Attrition
Into employment	4.4	4.1	5.1	4.8
Out of employment	3.1	2.9	4.3	4.1
Employment change	+1.3	+1.2	+0.8	+0.7
Into unemployment	2.1	2.0	3.2	3.0
Out of unemployment	3.2	3.1	5.0	4.8
Unemployment change	−1.1	−1.1	−1.8	−1.8
Into non participation	2.8	2.7	3.9	3.8

Gross flows	LUR		HUR	
	Without Attrition	With Attrition	Without Attrition	With Attrition
Out of non participation	3.0	2.8	2.9	2.7
Change in non participation	−0.2	−0.1	+1	+1.1

Source: own elaboration on PLFS data.

Table 7.

Labour market transitions with and without attrition in Poland (1995–'96; November)

1995	1996			
	Employed	Unemployed	Non participating	A
Employed	43.5	1.6	1.9	2.8
Unemployed	2.6	4.2	1.2	1.0
Non participating	1.9	1.0	35.8	2.5
1995	1996			
Employed	46.2	1.7	2.3	
Unemployed	2.8	4.4	1.3	
Non participating	2.4	1.1	37.9	

Source: Own calculation on PLFS.

4.4. A logistic analysis of attrition

So far, only few determinants of attrition have been considered. In this section, a more systematic analysis of the determinants of attrition is carried out estimating the probability of being selected in the sample, rather than quitting it. Table 8 shows the results of a LOGIT model for panel selection in terms of various individual characteristics of the survey sample. The dependent variable is a dummy taking the value of one in case the individual is selected and zero in case the individual quits the survey. The table presents the estimated coefficients. The exponential of these coefficients measures the odds ratio, i.e. the probability of being selected rather than quitting the survey. Two models are included. The main difference is the way of treating regional dummies. In model one, 47 *voivodship* dummies have been included, using Warsaw as the baseline. The coefficients were all positive and significant, suggesting that non-response is especially strong in the capital city. This result is in line with that reported in Socha and Weisberg (1999, p. 17), according to whom panel attrition concentrated in Warsaw. Model two differs from model one in as much as it substitutes the *voivodship* dummies with dummies representing groups of regions homogeneous by unemployment rate. The coefficients confirm that the regions with the lowest unemployment rate, including the most urban areas of the country are those with the highest

rate of panel attrition, due either to sampling fatigue or to residence change (Table 4).

Table 8.

LOGIT model for panel selection

Variable	(1)	(2)	Means
Constant	2.55***	3.25***	
	(0.12)	(0.10)	
Aged 15–19	–0.72***	–0.66***	0.1166
	(0.12)	(0.11)	
Aged 20–24	–1.22***	–0.75***	0.0850
	(0.09)	(0.09)	
Aged 25–34	–0.81***	–0.75***	0.1549
	(0.09)	(0.09)	
Aged 35–44	(baseline)	(baseline)	0.2172
Aged 45–54	–0.08	–0.08	0.1547
	(0.11)	(0.11)	
Aged 55–64	–0.16	–0.14	0.1408
	(0.12)	(0.12)	
Aged 65 or more	–0.92***	–0.87***	0.1308
	(0.11)	(0.11)	
Women	0.22***	0.22***	0.5342
	(0.05)	(0.05)	
University education	–0.43***		0.0671
	(0.11)		
Post-secondary diploma	–0.24	–0.30*	0.0236
	(0.17)	(0.17)	
General and vocational secondary diploma	–0.15**	–0.20***	0.2469
	(0.07)	(0.08)	
Low vocational diploma	–0.21***	–0.19***	0.2605
	(0.08)	(0.08)	
Low secondary school or below	(baseline)	(baseline)	0.2469
Disabled	–0.20**	–0.14*	0.1538
	(0.08)	(0.08)	
Unemployed workers	–0.75***	–0.81***	0.0856
	(0.09)	(0.10)	

Variable	(1)	(2)	Means
Inactive workers	0.03	−0.13*	0.4129
	(0.08)	(0.08)	
Employed workers	(baseline)	(baseline)	0.5015
Long term unemployed	0.25*	0.26*	0.0293
	(0.15)	(0.15)	
Available to change address for work reasons	no	−0.26*	0.0187
	(0.14)		
Employed in the private sector	no	−0.35***	0.1127
	(0.08)		
Voivodship dummies[a]	yes	no	
Low unemployment voivodships	no	(baseline)	0.3552
Medium unemployment voivodships	no	0.32	0.2911
		(0.06)	
High unemployment voivodships	no	0.38***	0.3585
		(0.06)	
Number of observations	27183	27183	
R^2 of Nagelkerke	0.069	0.048	

Note: Dependent variable is a dummy taking value 1 in case of selection in the sample and value 0 in case of non-response. The table reports the coefficients of a logistic estimate. They are significantly different from zero at 1% (***), 5% (**) and 10% (*) level. The figures between brackets are standard errors. The exponential of the coefficient gives the odds ratio, i.e. the probability to be selected in the sample, rather than withdrawing from it at November 1995, for individuals belonging to the sample in November 1995.

[a] 47 voivodship dummies have been included in the estimates, using Warsaw as baseline. Almost all the dummies have a positive and highly significant coefficient, confirming the high attrition rate of the capital city.

Source: own elaboration on Labour Force Survey data.

Confirming the analysis based on unconditional means, the probability of being selected in the sample rather than being attriters follows an inverse u-function, increasing with age up to the age of 35 from when it becomes stable until the age of 65, when it reduces again dramatically. Women, in turn, have much lower probability of non-response than men. Respondents with high educational levels tend to have a lower, not a higher probability of being selected in the sample, perhaps due to their higher tendency to migrate or to change residence. This is in contrast with what was found in the BHPS, where education is a positive determinant of the rate of response to surveys (Paull, 1996; Laurie *et al.*, 1997).

Unemployed workers have a higher rate of non-response than employed workers. However, individuals not in the workforce do not have significantly different behaviour from employed workers. Against the evidence relative to the BHPS (Paull, 1996), individuals with long-term unemployment spells tend to have lower, rather than higher non-response rate.

Interestingly enough, a dummy for disabled people is also significant, suggesting that the probability of quitting the survey is higher for this group.

Two other control variables have been added to the estimate in model (2): a dummy to catch the declared availability to move of the worker and employment in private firms, where much of the turbulence typical of the Polish labour market concentrates. Both variables significantly affect the probability of non-response. Against the general tendency of this labour market status, particularly strong is the impact of employment in the private sector. The individuals available to move and those involved in private activities are more likely to become attriters.

Overall, the results of the logistic analysis of attrition confirm the observations contained in the previous section. The determinants of attrition in the PLFS are generally similar to those reported in similar studies relative to surveys carried out in Western countries, with two exceptions. Individuals with low education attainment and long term unemployment spells tend to have a lower, rather than a higher probability of non response in the Polish case. A possible explanation of this peculiarity is that residence changes, more frequent among highly educated workers, are relatively more important than sampling fatigue in the Polish case, compared to other surveys. The overall significance level and the large number of significant coefficients suggest that attrition is systematic also in the case of the PLFS. Economists should proceed with caution when analysing PLFS data.

5. Cures

The cures for attrition differ according to the research aims in *ex ante* and *ex post* types. The latter need a careful study of the distribution of attrition. Laurie *et al.* (1997) report weighting techniques are used by the BHPS to compensate for attrition bias on observables. Paull (1996) suggests *ad hoc* two step procedures to be adopted in the estimates to correct for unobservable factors. Arulampalam *et al.* (2000) implement such a type of procedure to estimate the probability of job finding and of unemployment persistence, using the English BHPS. Other important examples of studies implementing sample selection procedures to control systematic panel attrition due to unobservables are Lindeboom and Van den Berg (1998) and Dolton, Lindeboom and Van den Berg (2004).

Among the *ex ante* cures one should mention the definition of the design of the survey. Attrition and natural turnover within the work force are the most important reason why the PLFS has become a rotating panel in May 1993 (Socha and Weisberg, 1999, p. 17). The presence in each survey of cohorts en-

tering at a different stage is meant to correct, at least in part, for "attrition" and get less biased measures of the mean value of each variable at any point in time. In fact, correcting attrition is one of the main reasons of the rotating structure of many LFSs.

Lairy *et al.* (1997) discuss various strategies for reducing attrition, especially that due to panel fatigue, in the BHPS. This is a pure panel, with a much smaller sample than the PLFS, but it is clear that some of the field-work related procedures and survey systems to maintaining high response rates would also apply to the case of LFSs. Also in this case, the study of the categories and determinants of panel attrition helps targeting the interventions.

Other statistical cures can be implemented at the time of the survey. The first intervention could include the substitution of the agents exiting the survey with others with similar characteristics. Information on the reasons of withdrawal from the survey could also be of much interest, at least to control for residence changes. Other statistical methods, such as the introduction of economic incentives for individuals to stay in the study, have proved ineffective or, worse, pejorative, as particular groups of agents may respond better to incentives, which may be a further source of bias. Alternatively, one could think of dropping out a proportional number of observations for each group. The risk would be then to produce further bias and to undermine the representative nature of the survey (Johnston and Di Nardo, 1997, p. 402).

From this short survey of the possible cures for attrition, it is apparent that no definitive cure is available. It is to be expected that National Statistical Offices implement corrections at the time of the survey to control at least the observable factors. In the meanwhile, it is necessary that also economists be aware of the problem of attrition when interpreting the results of analysis of micro-data based on LFS.

Concluding remarks

Various advantages and shortcomings of the PLFS have been analysed. A special focus has been on attrition and measurement errors. These are in principle particularly worrisome when analysing labour market dynamics. Flow data based on the PLFS has been the subject of a large strand of literature, but no formal analysis of attrition is available in the case of the PLFS. This paper aims to partly fill this gap and to raise awareness of the problem. The method adopted here is of interest for any individual level survey based on a rotating scheme, the most common in EU countries.

The previous discussion suggests that no definitive statistical or econometric remedy exists against panel attrition. However, a study of attrition should be considered a necessary preliminary step of any research based on individual level panel data, also with a rotating design. First, it is important to assess the size of attrition. Second, the study of the distribution and of the determinants of attrition will make the researcher aware of some possible

sources of bias, providing guidelines to the data analysis. This study suggests that in the PLFS case panel attrition depends mainly on age and is almost unnoticeable among prime-aged female workers with a low level of education. This distribution of attrition is similar to that of the BHPS, with few exceptions: in the Polish case, attriters are more frequent among individuals with a high level of education and experiencing short unemployment spells. Last, but not least, unobservable factors might produce sample selection bias and specific correction procedures should be implemented to control it.

References

Adamchik, Vera, and King, Arthur E., "The Impact of the Private Sector on Labour Market Flows in Poland." *Lehigh University*, mimeo, 1999. Bethlehem, Pennsylvania.

Aghion, Philippe, and Blanchard, Olivier J., "On the Speed of Transition in Central Europe." In *NBER Macroeconomics Annual*, pp. 283–320, 1994. The MIT Press.

Arulampalam, Wiji, Booth, Alison L., and Mark, P. Taylor, "Unemployment Persistence." *Oxford Economic Papers*, **52**, 1:24–50, January 2000.

Barbone, Luca, Marchetti, Domenico J. and Paternostro, Stefano, "The Early Stages of Reform in Polish Manufacturing. Structural adjustment, Ownership and Size." *The Economics of Transition*, 7, 1:157–177, March 1999.

Blanchard, Olivier J., "Transition in Poland." *The Economic Journal*, **104**, 4:1169–1177. September 1994.

Boeri, Tito, "Labour Market Flows and the Persistence of Unemployment in Central and Eastern Europe." In Organisation for Economic Co-operation and Development, *Unemployment in Transition Countries: Transient or Persistent?*, pp. 13–56. Paris: OECD, 1994.

Boeri, Tito, and Sziraczki, Gyorgy, "Labour Market Developments and Policies in Central and Eastern Europe: A Comparative Analysis." Organisation for Economic Co-operation and Development, *Structural Change in Central and Eastern Europe: Labour Market and Social Policy Implications.* Paris: OECD, 1993.

Clark, Kim B., and Summers, Lawrence H., "The Dynamics of Youth Unemployment." In Freeman, Richard B. and Wise David A., Eds., *The Youth Labour Market Problem: Its Nature, Causes and Consequences*, pp. 199–235. Chicago: The University of Chicago Press, 1982.

Dex, Shirley, and McCullock, Andrew, "The reliability of Retrospective Unemployment History data", Working Paper No. 17. Colchester: University of Essex/ISER, 1997.

Dolton, Peter J., Lindeboom Marteen, and Van den Berg Gerard J., "Survey Non-response and Unemployment Duration", Discussion paper No 1303. IZA, September 2004.

Dyker, David A., "The Computer and Software Industries in the East European Economies—A Bridgehead to the Global Economy?", Discussion Paper No 27. Brighton: University of Sussex/STEEP, February 1996.

Estrin, Samuel, Schaffer, Mark E., and Singh, Inderjit J., "The Provision of Social Benefits in State-Owned, Privatised and Private Firms in Poland", Discussion Paper No 223. London: LSE/CEP, February 1995.

Filer, Randall K., and Hanousek, Jan, "Data Watch: Research data from Transition Economies", *Journal of Economic Perspectives*, 2002, forthcoming.

Favro-Paris, Maria, Gennari, Pietro, and Oneto, Gianpaolo, "La durata della disoccupazione in Italia: un'applicazione della struttura longitudinale dell'indagine sulle forze di lavoro." *Quaderni di ricerca* ISTAT, **1**, 4:1–79, 1996.

Góra, Marek (1994), "Labour Market Policies in Poland." In Organisation for Economic Co-operation and Development, *Unemployment in Transition Countries: Transient or Persistent?* Paris: OECD, 1994.

Góra, Marek and Lehmann, Hartmuth, "How Divergent is Regional Labour Market Adjustment in Poland?" Organisation for Economic Co-operation and Development, *The Regional Dimension of Unemployment in Transition Countries. A Challenge for Labour Market and Social Policies.* Paris: OECD, 1995.

Góra, Marek, Kotowska, Irena, Panek, Tomasz, and Podgorski, Jan, "Poland: Labour Market Trends and Policies." Organisation for Economic Co-operation and Development, *Structural Change in Central and Eastern Europe: Labour Market and Social Policy Implications.* Paris: OECD, 1993.

Hsiao, Cheng, "Analysis of Panel Data". Cambridge: Cambridge University Press, 1986.

Johnston, Jack, and Di Nardo, John, *Econometric Methods.* New York: McGraw Hill, 1997.

Kemp, Gordon C. R., "The Use of Panel Data in Econometric Analysis: a Survey", Working Paper No 4. Colchester: University of Essex/ISER, 1991.

Laurie, Heather, Smith Rachel and Scott Lynne, "Strategies for Reducing Non-response in a Longitudinal Panel Survey." Working Paper No 12. Colchester: University of Essex/ISER, 1997.

Lehmann Hartmut, and Wadsworth Jonathan, "Tenures that Shook the World: Worker Turnover in Russia, Poland and Britain." *Journal of Comparative Economic*, **28**, 4:639–664. December 2000.

Lindeboom Marteen and Van den Berg Gerard J., "Attrition in panel survey data and the estimation of multi-state labor market models." *Journal of Human Resources* 33, 458–478, 1998

Organisation for Economic Co-operation and Development, *Unemployment in Transition Countries: Transient or Persistent?* Paris:OECD, 1994.

Organisation for Economic Co-operation and Development, *Economic Surveys: Poland.* Paris:OECD, 1997.

Paull, Gillien, "Dynamic Labour Market Behaviour in the British Household Panel Survey: The Effect of Recall Bias and Panel Attrition", Discussion paper No 10. London: LSE/CEP, 1996.

Pinto, Brian, Belka, Marek and Krajewski, Stefan, "Transforming State Enterprises in Poland: Evidence on Adjustment by Manufacturing Firms." *Brookings Papers on Economic Activity*, 1: 213–70, 1993.

Sestito, Paolo, "Misurazione dell'offerta di lavoro e tasso di disoccupazione", Temi di discussione del Servizio Studi No 132. Rome: Banca d'Italia, March 1990.

Socha, Mieczysław, and Sztanderska, Urszula (1997), "Employment and Labour Market Policies in Poland", University of Warsaw, mimeo.

Socha, Mieczysław, and Weisberg, Jacob, "Poland in Transition: Labour Market Data Collection." *Monthly Labour Review*, **122**, 9:9–21, September 1999.

Svejnar, Jan, "Labour Markets in the Transitional Central and East European Economies." In Ashenfelter, Orley C., and Card, David (eds.), *Handbook of Labour Economics*, vol. 3B, Ch. 42, pp. 2807–2857. Amsterdam: North-Holland, 1999.

Witkowski, Jan, and Szarkowski, Arthur (1994), "The Polish Labour Force Survey". In Chernyshev, Igor, Ed., *Labour Statistics for a Market Economy. Challenges and Solutions in the Transition Countries of Central and Eastern Europe and the Former Soviet Union.* Budapest: Central European University Press, 1994.