HUNGARIAN STATISTICAL REVIEW

JOURNAL OF THE
HUNGARIAN CENTRAL
STATISTICAL OFFICE

VOLUME 82 2004. SPECIAL NUMBER 9

CONTENTS

ISSN 0039 0690

Published by the Central Statistical Office
Editor in charge: dr. Péter Pukli
Executive editor: dr. László Hunyadi
Printed by the Akadémiai Nyomda
4083 – Martonvásár, 2004
Director in charge: Lajos Reisenleitner

Managing editor: Orsolya Dobokayné Szabó
Editors: Mária Visi Lakatos, Soma Várady
Technical editors: Éva Bartha, Ágnes Simonné Káli

Editorial office: Budapest II., Keleti Károly utca 5–7. Postal address: P.O.B. 51. Budapest, 1525.
Phone: (361)-345-6546, Telefax: (361)-345-6594,
Internet: www.ksh.hu/statszemle E-mail: statszemle@ksh.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 Bookshop:
Budapest II. Keleti Károly utca 10. Phone: (361)-212-4348

CONSIDERATIONS ON AN ETHICAL CODE OF STATISTICS

DR. GYÖRGY SZILÁGYI[1]

The Declaration of Professional Ethics will be twenty years old in the following year. There is a general agreement among statisticians that this document is basically appropriate for the future; nevertheless a kind of updating seems to be necessary in order to follow the development of the society and of information technology, and to facilitate the dissemination. The article reviews the existing Declaration and puts forward a number of modifications and new obligations.

KEYWORDS: Ethics.

In 1985 the International Statistical Institute (ISI) adopted its Declaration on Professional Ethics (published in the International Statistical Review[2] and available also on the ISI website[3]). The document is – within and outside ISI – highly appreciated. Nevertheless, two questions arose in the last few years:

1. Dissemination. It has been discovered that the document is not well known enough. Statisticians, including ISI members ignore its content, or even its existence (e.g. there are statistical agencies that drafted own statistical ethical codes, not because they are not satisfied with the ISI Code, rather because they are not aware of it). The situation is even worse among non-statisticians.

2. Necessity of updating. New phenomena – e.g. those connected with the revolutionary development of information technology, with education etc. – need to be incorporated in the Professional Ethics.

The present paper is focusing on the second issue.[4] We have to observe, though, that the two questions are not independent. One of the obstacles of a more efficient dissemi-

[1] The author is Chairman of the Scientific Board of Official Statistics in the Hungarian Statistical Office and Professor of the Budapest University of Economic Sciences.

[2] VoL.54. No. 2. p. 227–242.

[3] http://www.cbs.nl/isi/ethics.htm

[4] The first draft of the revision was introduced by the present author and discussed in the framework of an Open Meeting at the 54th General Conference of ISI in Berlin, August 2003. Positive and negative views expressed at the meeting are taken into account in this article.

nation is the relatively large size of the existing Declaration. With this in mind, three ways are open for a revision:

– Reducing the size of the Declaration (the ways of possible reductions are discussed later under the heading 'A possible size and structure of an updated version').
– Keeping the size as it is, and produce a reduced version in order to promote the dissemination for a large audience.
– Adding more details (considerations, explanations, references etc.) to the reduced version as an annex.

The advantage of the first way is the easy dissemination, of the second version is that it permits to keep a great number of valuable considerations, references etc. The third one is a compromise between the two. This article puts forward a reduced version without opting between the variants outlined above.

GENERAL ISSUES

Whose ethics is it? – I.

The answer to the above question seems to be obvious: it is the ethics of *statistics,* or, in a more personal sense, of statisticians. Moreover: of *all* statisticians, i.e. official, academic, private etc. statisticians (employers or employees) as well as teachers of statistics. Towards this end the preamble of the present (1985) Declaration deserves special attention: 'Statisticians work within a variety of economic, cultural, legal and political settings, each of which influences the emphasis and focus of statistical inquiry. They also work within one of several different branches of their discipline, each involving its own techniques and procedures and its own ethical approach.'…'Thus, no declaration could successfully impose a rigid set of rules to which statisticians everywhere should be expected to adhere…'
In other words the declaration should not be authoritarian or prescriptive. The informative and descriptive nature of the document needs to be maintained. The efficiency of the declaration however can be substantially raised if the reader is aware of his/her own duty.

THE OVERALL STRUCTURE OF THE EXISTING DECLARATION

The structure of a system like a code of professional ethics is not simply a layout; it is part of the objectives and contributes to the efficiency of the document concerned. The present structure of the Declaration arranges the various obligations of statisticians in a comprehensive manner. The code in fact has a multi-level building.

– At the first level it identifies the *broad categories of obligations* (4 altogether) as 'Obligations to society', 'Obligations to founders and employers' etc.
– At the next level these obligations are subdivided into *specific duties* (16), like 'Considering conflicts of interests' or 'Clarifying obligations and roles'.
– The various items of these two levels are then followed by *more detailed definitions* and
– even more detailed *comments, explanations and references.*

As a reminder, the first two levels of the existing Declaration is reproduced below:

1. Obligations to society
1.1. Considering conflicting interest
1.2. Widening the scope of statistics
1.3. Pursuing objectivity
2. Obligations to funders and employees
2.1. Clarifying obligations and roles
2.2. Assessing alternatives impartially
2.3. Not pre-empting outcomes
2.4. Guarding privileged information
3. Obligations to colleagues
3.1. Maintaining confidence in statistics
3.2. Exposing and reviewing methods and findings
3.3. Communicating ethical principles
4. Obligations to subjects
4.1. Avoiding untrue invasion
4.2. Obtaining informed consent
4.3. Modification to informed consent
4.4. Protecting the interests of subjects
4.5. Maintaining confidentiality of records
4.6. Inhibiting disclosure of identities.

A POSSIBLE SIZE AND STRUCTURE OF AN UPDATED VERSION

Taking into account the structure of the 1985 version as outlined above, the reduced version might

– keep the first level i.e. the 4 broad categories of obligations;
– basically keep the second level (specific duties), not necessary exactly with 16 items, but around this size; some of the present duties might be omitted, substituted by a new one (updating), combining two into one etc.;
– shorten the detailed definitions;
– omit the detailed comments etc. from the reduced version.

The four broad categories of obligations

As mentioned before, the fundamental structure of the 1985 Declaration is constituted by four broad categories of obligations:

1. Obligations to society
2. Obligations to funders and employees
3. Obligations to colleagues
4. Obligations to subjects.

These groups of obligations, indeed, cover and classify in a logical manner all relevant aspects of an ethical code. Nevertheless, compared with the ethics of many other professions, one may feel that *obligations to the (own) profession (statistics)* is missing. The prestige of our profession requires such a duty; it is therefore proposed to add such a group of obligations. However, instead of increasing the number of the broad categories, it is convenient to combine it with the obligations to colleagues, so a new item would be 'Obligations to the profession (statistics) and colleagues'. This slight modification may have some consequences, e.g. obligation 'Widening the scope of statistics' (1.2.) belonging now to the 'Obligations to society' (1.) becomes closer to the now enlarged group regarding profession and colleagues, so it is proposed to transfer it to group 3.

One by one analysis of the groups

Let us analyse now some details of the four groups of obligations.

1. Obligations to society. The professional performance of statistics is relevant in many aspects to the society. Four factual duties can be defined within this group:

1.1. Contribution to the extension of knowledge of the members of society *(new)*
1.2. Consideration of conflicting interests
1.3. Objectivity, impartiality.
1.4. When teaching statistics: transmitting ethical principles and values to the students. *(new)*[5]

As mentioned before, a *short* definition is proposed to each obligation. E. g. such a definition to 1.2. might be: 'Statisticians should consider the likely consequences of collecting and disseminating various types of data and should guard against predictable misinterpretations or misuse.'

2. Obligations to funders and employees. Statistical work is suited to the needs and resources of those who are paying for it. It is necessary that funders and employers understand the capabilities and limitations of statistics and that the funder's and employer's information is protected.
In this group of obligations three duties are proposed:

2.1. Clarification of obligations and roles
2.2. Respect of ethics of other professions (new)
2.3. Not pre-empting outcomes

The proposed new item (2.2.) seems to be necessary in interest of mutual respect between statisticians and non-statisticians. Consequently, statisticians who carry out their

[5] I am very greatful to the International Association for Statistical Education, personally to *Mrs. Carmen Batanero* and *Mr. Chris Wild* for their valuable comments made to the first draft of the revision. Most of those comments are reflected in the detailed definitions of the individual duties.

activity in the environment of a profession which has its own ethical code (e.g. medicine, sociology, journalism) are expected to respect those principles to the same extent as the ethical code of statistics.

3. Obligations to the profession (statistics) and colleagues. According to the considerations above, this is a (partially) new item. In addition, this is a group of obligations that needs two types of extensions: *a*) new items in connection of obligations to the profession (e. g.: sharing experiences with colleagues); *b*) new items in connection of development of information technology and development in the society. In such a way, the proposed composition of this group of obligations can be the following:

3.1. Widening the scope of statistics (transferred from group 1.)
3.2. Maintaining confidence in statistics
3.3. Maximum but correct use of benefits of technical development *(new)*
3.4. Participation in lifelong learning *(new)*
3.5. Dissemination of this Code of Ethics
3.6. Sharing experiences with colleagues
3.7. Respect intellectual ownership *(new)*

As before, here is a sample of the short definitions; this time to item 3.3. (Maximum but correct use of benefits of technical development): 'Statisticians should follow the development of information and communication technology, apply it to the maximum, but mainly in favour of data suppliers and users; without any misuse of the advantages offered by these devices.'

4. Obligations to subjects and respondents (persons, households, institutions, enterprises etc.). This is obviously the most difficult group of obligations. It takes about the third of the size of the 1985 Declaration. It contains a considerable number of problematic cases, both from the statisticians and the respondents side. The basic principle of this chapter is that statistical investigations involving the participation of human or institutional subjects should be based on friendly cooperation, consent and protection of subjects and respondents.

A special kind of extension seems to be necessary vis-á-vis the present Declaration where almost the entire text refers to subjects of social statistics (individuals, families) and almost nothing is told on subjects of 'economic statistics'. Businesses, enterprises and other economic entities need similar protection as individuals, in addition to special measures.

On the other hand it seems to be possible to combine some items of the 1985 Declaration; e.g. 'Obtaining informal consent' and 'Modification to informal consent' can be combined into 'Obtaining informed consent (even if direct consent is inhibited)'. Then the detailed definition can be e.g. 'Freely given informed consent by the subjects and respondents is useful even if participation is required by the law. Irrespective of the compulsory or voluntary character of the inquiry, the subjects should be convinced of the value of their contribution. In case if informed consent cannot be obtained directly, the subjects' interests should be safeguarded in one of the numerous indirect ways.'

So the concrete duties of this group of obligations can be:

4.1. Protection of subjects and respondents against excessive risk and excessive imposition on their time and privacy.

4.2. Obtaining informed consent (even if direct consent is inhibited).

4.3. Maintaining confidentiality of records and inhibiting disclosure of identities.

Whose ethics is it? – II.

Why put the same question again? Because it has a facet different from that considered at the beginning of this writing. Then it was enough to state that the ethical code belongs to the statisticians, and all kinds of statisticians. Now the question can be put into in a broader context. The position of statistics and statisticians, their respect of the ethical code does not depend solely on those persons. There are several reasons to think so:

a) Statistics is a science; it constitutes input for other sciences and other activities;

b) Statistical data are published and interpreted also by non-statisticians;

c) Persons outside statistics may limit statisticians to fulfil their duty in the spirit of the ethical code.

Non-statisticians in this context are those who have regular contact with statistics. e.g. secondary publisher of the product of statisticians. Some typical groups of non-statisticians whose behaviour is not independent from statistics are:

– Researchers in different sciences (other than statistics)
– Employers or supervisors of statisticians
– Teachers (other than teachers of statistics)
– Politicians
– Media representatives

It is therefore desirable to formulate a set of ethical principles vis-á-vis this group of agents. Toward this end two questions need closer consideration:

I. The status and treatment of these ethical principles

II. The nature and wording of such principles.

As regards to the *status of the ethical principles for non-statisticians*, one may ask about the competence of ISI concerning non-statisticians. In this respect it should be taken into account that the ethical code is not a low or a decree of an authority. It constitutes a set of guidelines of behaviour in connection of a given profession. With this in mind, the following attitudes can be considered:

a) Integrating the ethical principles of these professions into the Declaration of Professional Ethics as a second target group ‘*non statisticians*’ (the first target group being ‘statisticians’), *b*) Without integration into the system, addition of a kind of annex to the Declaration, e.g. 'Advices to (or expectations from) non-statisticians regarding ethical aspects of statistics'.

c) Forget about the group of non-statisticians.

This article refrains from choosing among these options. It is however obvious that when turning to the above second question (*The nature and wording of such principles*), only items *a*) and *b*) remain relevant.

This set of ethical principles is considerably smaller than the Code discussed so far. In comparison with those of statisticians; three types can be distinguished:

A. Obligations which are the same as those set against statisticians:

1. Consideration of conflicting interest (1.2.)

2. Objectivity, impartiality (1.3.)

3. Maintaining confidentiality of records and inhibiting disclosure of identities (4.3.)

B. Obligations having different interpretation for statisticians and non-statisticians:

4. Clarification of obligations and roles.

In the case of statisticians the definition reads: 'Statisticians should clarify in advance the respective obligation of both sides; for example the relevant parts of this Code to which they adhere' (2.1.).

For non-statisticians it goes the other way round:: 'A non-statistician funder, employer or supervisor of a statistician is expected to respect the stipulations of the ethical code binding the statistician.'

5. Not pre-empting outcomes.

In the case of statisticians the definition reads: 'Do not accept conditions and obligations that are contingent upon a particular outcome from a proposed statistical inquiry' (2.3.).

For non-statisticians: 'When using and/or publishing official or other statistics, use them in the professional context of their actual meaning, rather than for underpinning preconception.'

C. Obligation, special for non-statisticians:

6. Give true interpretation to statistical data, tables or diagrams; in case of doubt consult the producer.

I hope that considerations of this paper bring closer to an agreement upon an updated version of the Declaration on Professional Ethics.

INCOME OR EXPENDITURE?*
THEIR COMPETING ROLE TO CHARACTERIZE THE LIVING CONDITIONS OF HOUSEHOLDS

ÖDÖN ÉLTETŐ[1] – ÉVA HAVASI[2]

The individual data bases of the Hungarian Household Budget Surveys are suited to examine the relation between the incomes and expenditures of the households, and to study which of the two variables characterizes better the living conditions of the households, can separate better the poor from the not poor and, respectively, the well-off households from the not well-off ones. In the study the authors try to answer these questions on the data bases of the HBS in 2001 and 2002. It is also examined whether there was any appreciable change in these topics between the two years considered. Authors conclude that, if possible, both variables are to be taken into account in a complex manner, because the really poor are those poor in both respects and the really wealthy are well-off both in income and expenditure.

KEYWORDS: Income; Expenditure; Inequality; Poverty.

Data on living conditions and especially on consumption patterns of households are, generally, provided by Household Budget Surveys (HBSs). In many countries HBS data on both incomes and expenditures are inquired, while in other countries households are asked to report their expenditures only. Hungary belongs to the former group of countries and its HBS is a continuous survey covering annually about 10 thousand households selected at random by a two and three stage stratified sampling design.

Using the individual data base of households co-operating in the 2001 and 2002 HBS in Hungary, authors investigated whether the *income* or *current expenditures* (disregarding investment type expenditures for production and business operational costs) are in closer relation with the real living conditions of the households, explain better the phenomena characterizing poverty as well as wealth in the Hungarian society today. The paper summarizes the main findings of the research. At this point it must be noted, however, that income data of the Hungarian HBS are, generally, less reliable than expenditure data and therefore income inequality is very probably somewhat underestimated.

* The study is a modified and extended version of the paper presented by the authors in 2003 at the 54th Session of the International Statistical Institute in Berlin.

[1] Retired deputy head of department of the HCSO.

[2] Sociologist of the HCSO.

For the measurement of both income and expenditures we used the well-known OECD1 equivalence scale (first adult is given 1, further adults 0.7, children 0.5). Poverty threshold is defined as usual, i.e. as 60 percent of the *median* income and expenditure, respectively. At the other end of the distribution are those households considered as income or expenditure well-off belonging to the top income or expenditure decile.

The paper proceeds as follows: The first section presents some findings regarding the nature of relation between income and expenditure of households, while inequalities of incomes and expenditures, resp. are investigated and compared in the next section. Then issues of income and expenditure poverty and wealth, resp. are dealt with. In the following section we try to characterize from various aspects the income, expenditure and double poor, as well as the well-off households. The last section draws some conclusions.

RELATION BETWEEN INCOME AND EXPENDITURE

According to the data of the HBS in 2002 the average equalized yearly income of households was about 731 thousand HUF (corresponding to about 2 850 euro at that time) and their average equalized expenditures was about 659 thousand HUF (~2 570 euro) in Hungary, i.e. the former exceeded the latter by about 11 percent indicating that, on the average, households had some saving in that year. The difference between the *median* income and expenditures of households was somewhat larger, it amounted to almost 15 percent. One of our most important findings shows that the relationship between income and expenditure is not too strong, the correlation coefficient amounts to 0.68 only. We can come to the same conclusion if we range household members into quintiles and deciles simultaneously on the basis of their equalized income and expenditures. Only 43.5 percent of the persons can be found in the same quintile, while the quintile positions of more than 17 percent of the persons differ by more than one quintile in the two types of ranging (see Table 1). As to the deciles, only 26 percent of the persons belong to the same deciles and the decile position of almost 42 percent of the persons differ by more than one decile (see Table 2). In both cases the extreme deciles and quintiles in the diagonal show higher agreement than do the rest of deciles and quintiles. There was no appreciable change in these figures from 2001 to 2002.

Table 1

Share of income and expenditure quintiles of persons, 2002

Income (quintiles)	Expenditure					
	1.	2.	3.	4.	5.	*Total*
	quintiles					
1.	**11.4**	4.7	1.9	1.2	0.8	*20.0*
2.	5.0	**6.6**	5.2	2.4	0.8	*20.0*
3.	2.4	5.2	**6.2**	4.8	1.4	*20.0*
4.	1.1	2.7	4.8	**6.9**	4.5	*20.0*
5.	0.1	0.8	1.9	4.7	**12.5**	*20.0*
Total	*20.0*	*20.0*	*20.0*	*20.0*	*20.0*	***100.0***

Table 2

Share of income and expenditure deciles of persons, 2002

Income (quin-tiles)	Expenditure										
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	*Total*
	quintiles										
1.	**5.1**	2.0	1.1	0.4	0.2	0.3	0.3	0.1	0.3	0.2	*10.0*
2.	2.2	**2.2**	1.9	1.3	0.8	0.6	0.4	0.3	0.3	0.1	*10.0*
3.	1.2	2.0	**1.7**	1.7	1.4	0.6	0.7	0.3	0.1	0.2	*10.0*
4.	0.6	1.2	1.6	**1.6**	1.7	1.4	0.8	0.6	0.4	0.1	*10.0*
5.	0.5	1.0	1.4	1.6	**1.4**	1.5	1.2	0.9	0.4	0.1	*10.0*
6.	0.3	0.6	0.9	1.3	1.7	**1.5**	1.4	1.3	0.7	0.2	*10.0*
7.	0.1	0.6	0.8	1.0	1.0	1.6	**1.6**	1.4	1.2	0.5	*10.0*
8.	0.1	0.3	0.4	0.4	1.0	1.1	1.7	**2.2**	2.0	0.8	*10.0*
9.	0.0	0.1	0.2	0.5	0.4	1.0	1.3	1.7	**2.7**	2.0	*10.0*
10.	0.0	0.0	0.0	0.1	0.2	0.4	0.5	1.1	1.9	**5.7**	*10.0*
Total	*10.0*	*10.0*	*10.0*	*10.0*	*10.0*	*10.0*	*10.0*	*10.0*	*10.0*	*10.0*	***100.0***

INCOME AND EXPENDITURE INEQUALITY

Theoretical considerations indicate that expenditures should be more equally distributed than incomes, because many low income households spend above their income by drawing down past savings and high income households generally save part of their income and therefore spend less than it. As can be seen from Figure 1 empirical data corroborate this experience: expenditures exceeded income by almost 50 percent in the first income decile in 2001 and by about 25 percent in 2002, but reach only 80 percent of income in the top income decile in both years. In the rest of the income deciles the ratio of expenditures to income is nearly one. Expenditures exceeded incomes somewhat more in 2001 than in 2002.

Figure 1. Ratio of expenditure to income by income deciles

Still, data clearly indicate that the inequality of expenditures significantly exceeds that of the incomes. This is demonstrated both by the Lorenz curve and the shares of income and expenditure deciles in Figures 2 and 3, as well as by the various inequality measures shown in Table 3.

Figure 2. Share of income and expenditure deciles of persons, 2002

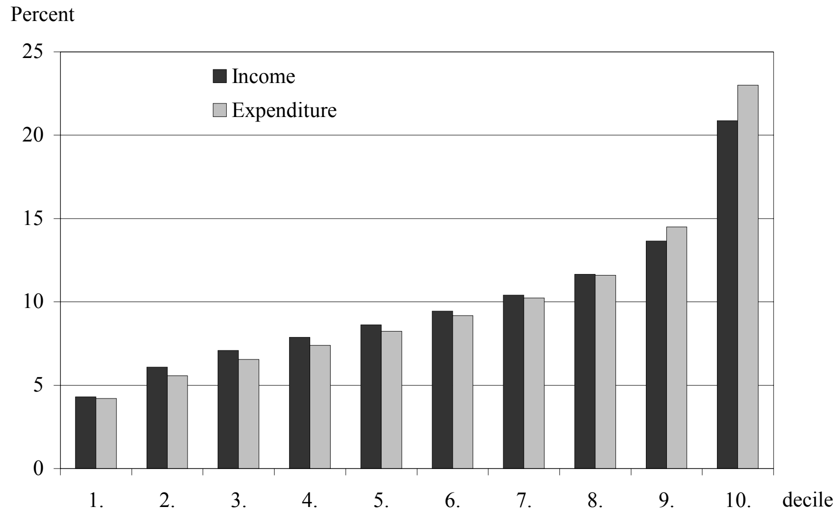
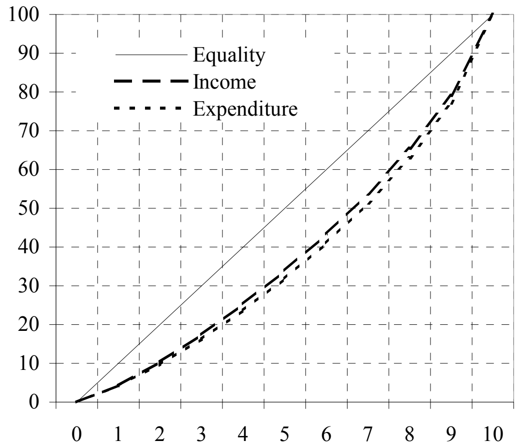


Figure 3. Lorenz curve of income and expenditure



There is especially significant difference in the shares of the top decile in 2002: share amounts to 23.8 percent in case of expenditures, as against the value of 20.9 percent for incomes. The inequality measures of expenditures (see Table 3) exceeded those of the incomes by 10-13 percent. In 2001 the differences were generally a bit smaller, only the coefficient of variation of the expenditures exceeded by more than 17 percent that of incomes. It is remarkable that while S10/S1 and the Gini coefficient indicate a slight decline in the inequality of incomes from 2001 to 2002, the coefficient of variation shows a definite increase in that inequality. In the case of expenditures all but one measures indicate a mild increase in the inequality between the two respective years.

Table 3

Inequality of income and expenditure

Indicator	Income		Expenditure	
	2001	2002	2001	2002
	year			
Share of the 1st decile (S1)	4.17	4.30	4.11	4.20
Share of the 10th decile (S10)	21.11	20.87	22.72	23.81
S10/S1	5.07	4.84	5.53	5.49
Gini coefficient	0.2400	0.2335	0.2616	0.2635
Éltető-Frigyes measure	2.00	1.96	2.15	2.16
CV (percent)	50.94	54.08	59.81	61.19

It is interesting to note that the same phenomenon was found by *Ann Harding* and *Harry Greenwell* [2002] in connection with income and expenditures inequality of households in Australia.

We made some research to find out why expenditures distribute more unequally than incomes. One minor factor may be that the very rich people are, generally, not covered by the HBS, because they tend not to co-operate in the survey. But the main underlying cause seems to be connected with the nature of the expenditures. Not only the relative variance of the expenditures is markedly greater than that of the incomes but also the between deciles part of the variance of expenditures exceeds considerably that of the incomes: this part was 71 percent for expenditures, while it was only 67 percent in the case of incomes in 2002. We tried, in addition, to explain the logarithm of the summed squared deviations from the mean of both incomes and expenditures by means of a linear regression containing the following four explanatory variables:

1. educational attainment of the household head (measured by the number of classes completed)
2. age of the household head
3. whether the household belongs to the top decile or not
4. whether the household lives in Budapest or not.

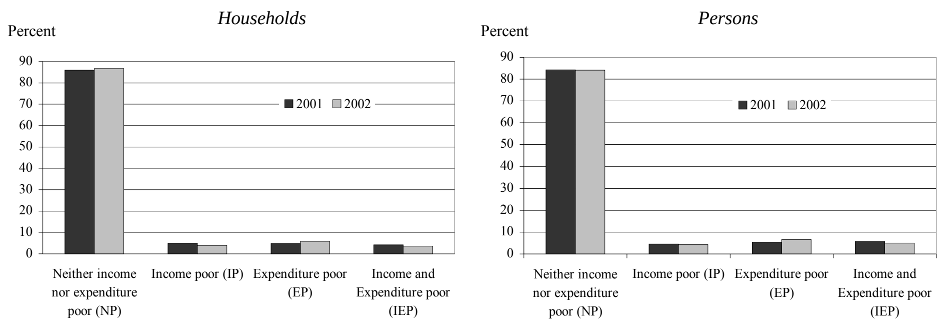
It turned out that the above variables explain the logarithmic variability of incomes less than that of expenditures (adjusted R^2s were 0.09 and 0.22, resp. in 2002) and the determining factor is, in both cases, variable 3; but while the value of the corresponding standardized β coefficient was only 0.286 in the case of incomes in 2002, it was much higher, 0.487 for the logarithmic variance of the expenditures.

It can be concluded that by simple and easily definable variables the expenditures of households can be less explained than that of their incomes. As it was pointed out, earlier variability of expenditures are only partially determined by current income. Moreover they seem to be affected to a greater extent by less easily measurable variables (e.g. traditional attitudes towards saving, accustomed spending patterns, environmental effects, etc.) than incomes. In addition, in recent years consumption does not any more restrained by lack of supply, people can buy anything they want (frequently what they do not really want), if possessing the required financial sources. All these contribute to the greater variability and inequality of expenditures than of incomes, at least in today Hungary.

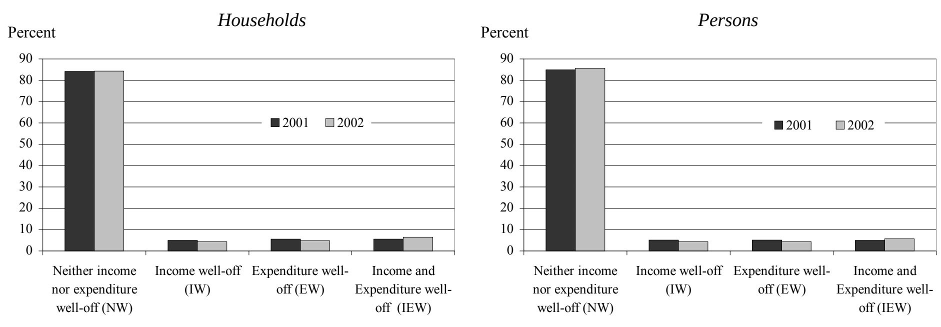
INCOME AND EXPENDITURE POVERTY AND WEALTH

According to the definition of poverty threshold given in the introductory part, 9.2 percent of the Hungarian population could be considered as income poor and 11.6 percent as expenditure poor in 2002. However, one of our most important findings shows that the income poor and the expenditure poor are not the same sets of households. The common part is not a major share: less than 48 percent of those belonging to the income poor was at the same time expenditure poor and about 38 percent of the expenditure poor was also income poor in 2002. As a consequence, only 5.0 percent of the society – 3.6 percent of the households – can be considered as poor from both aspects. Similar statements can be made about the well-off households: only 5.7 percent of the population – 6.4 percent of the households – can be considered as well-off from both aspects.

Figure 4. Percent of poor and non-poor households and persons, 2001 and 2002



Figures 5. Percent of well-off and not well-off households and persons, 2001 and 2002



CHARACTERISTICS OF POOR AND WELL-OFF HOUSEHOLDS

A deeper analysis shows that there are significant differences not only in the proportions but also in a number of important characteristics among the three sets of the poor (income poor, expenditure poor and the poor from both aspects). Moreover, in respect of certain characteristics considerable changes have occurred between 2001 and 2002, especially within the income and double poor. Most of the data indicate that expenditure poverty is more stable, while the structure and characteristics of income poor households can considerably change from one year to the next one. However, this latter phenomenon may follow – at least in part – from the already mentioned fact that income data of the HBS are less reliable then expenditure data. Among the differences in characteristics it is worth mentioning that e.g. while the proportion of large families with at least three children were 8-9 and 10 percent among the income and expenditure poor households in both years, it was more than 22 percent in 2001 and 27 percent in 2002 among households poor from both aspects (large families amount to 4 percent of all households in Hungary). Looking at it from the other side it is remarkable that childless families amount to only 30-31 percent of households poor from both aspects, while to about 55 percent of expenditure poor households. There was a remarkable change in this respect among the income poor households from 2001 to 2002: the proportion of childless households decreased from 60 percent to 44 percent.

This change is only one symptom of the changes in the structure of the income poor households. In a similar manner the proportion of one member households also decreased from 34 percent to 23 percent, that of households consisting only of old persons from 33 percent to 17 percent. At the same time the percentage of young households increased from 5 percent to 8 percent and that of households with unemployed member(s) from 14 percent to 23 percent. In this latter aspect the double poor households are especially at a disadvantage: among them in nearly every second household there was at least one unemployed person in 2002. More detailed data can be found in Table 4 below.

It is surprising to see the large difference between the income poor and expenditure poor in respect of the proportions of households living in Budapest: only 9 percent of the expenditure poor households live in the capital as against the 24 percent of the income poor. It seems that the many temptations and possibilities in the capital to spend induce the income poor living here to spend over their real financial resources.

From Table 4 data it can be concluded that double poor households live mostly in villages (generally in small ones), their heads are low educated, there are many large families among them with more children and, in addition, unemployment is considerably more frequent among them than among other types of households, even among households poor only from one aspect. It is worth mentioning, furthermore, the remarkable difference in the proportion of single person households: 34 percent of income poor households consists of one person as against their 21 percent among the expenditure poor. The difference in the proportion of households living in the capital was already discussed.

Table 4

Characteristics of poor and non-poor households, 2001 and 2002

Household characteristics	Neither income nor expenditure poor (NP)		Income poor (IP)		Expenditure poor (EP)		Income and expenditure poor (I&EP)	
	2001.	2002.	2001.	2002.	2001.	2002.	2001.	2002.
	year							
Average household size	2.6	2.6	2.5	2.9	3.0	3.1	3.6	3.8
	Percentages							
One member households	24.0	24.8	33.9	22.9	21.0	20.6	14.8	11.3
5 or more member households	7.8	7.7	11.2	11.0	16.8	18.5	29.6	32.4
Households with								
no child	60.7	61.0	60.1	44.3	55.4	55.7	29.9	30.8
3 or more children	3.9	3.8	8.9	7.5	10.0	10.0	22.4	27.2
unemployed member(s)	5.3	5.2	14.0	23.2	17.2	14.7	41.5	44.5
adult(s) without job	3.9	3.8	7.2	9.2	8.5	9.1	23.0	19.4
Households								
living in Budapest	21.1	21.9	23.6	9.7	9.3	11.8	9.0	3.4
living in villages	30.7	30.1	42.2	43.5	51.1	41.5	52.9	56.0
consisting only young persons within the household (under 30 years old)	5.7	5.6	5.4	8.2	3.2	5.4	14.5	5.8
consisting only old persons within the household (over 60 years old)	26.9	26.9	32.7	16.7	32.7	29.6	11.2	7.1
with head of low level of education	29.1	27.4	43.4	42.7	59.7	58.1	59.9	63.4
with head of high level of education	14.7	15.2	5.5	6.8	3.4	0.8	0.5	0.3
Subjective poor[a)]	6.6	8.4	6.2	7.4	42.5	49.4	30.1	28.4
Consuming poor[b)]	9.9	10.6	8.0	18.0	18.8	31.7	21.5	44.3
Housing poor[c)]	9.8	8.3	21.4	23.5	30.1	28.4	55.6	51.6
Housing-equipment poor[d)]	7.5	7.3	16.9	18.2	21.6	21.8	38.6	45.5
Multiple deprived[e)]	0.7	1.4	11.7	18.7	32.2	36.2	77.7	78.8

[a)] We asked the households' opinion how much money would be needed for them to a low or very low living standard. If the households had more than 20 percent less income as needed according to their opinion for this minimum living standard, they were defined subjective poor.

[b)] The household is consuming poor if the share of the food expenditure in their total current household expenditure exceeds 45 percent.

[c)] The classification is based on the social environment of the dwelling and/or on the substandard quality of the dwelling.

[d)] It refers to the provision of the household with consumer durables. Near 20 types of high-value domestic appliances were included. The index, based on standardized values for each appliance weighted using their distribution (based on z scores), was used to obtain housing-equipment deciles. Households in the bottom decile are housing-equipment poor.

[e)] It is defined by means of 6 different types of poverty and social exclusion dimensions (e.g. income, expenditure, housing equipment, subjective poverty). If the household is poor from at least 3 aspects, it is considered multiple deprived.

Table 5

Characteristics of well-off and not well-off households, 2001 and 2002

Household characteristics	Neither income nor expenditure well-off (NW)		Income well-off (IW)		Expenditure well-off (EW)		Income and expenditure well-off (I&EW)	
	2001.	2002.	2001.	2002.	2001.	2002.	2001.	2002.
	year							
Average household size	2.7	2.7	2.8	2.6	2.5	2.4	2.4	2.4
	Percentages							
One member family	24.2	23.7	16.1	18.4	25.9	31.6	24.7	26.5
5 or more member family	10.0	10.0	6.7	5.4	5.5	4.2	2.8	5.0
Households with								
no child	58.7	57.6	57.5	67.0	60.1	62.5	65.9	68.3
3 or more children	5.7	5.6	3.0	2.4	3.1	2.8	2.5	3.8
Households								
living in Budapest	17.3	17.2	42.7	39.9	27.2	26.3	37.0	40.8
living in villages	35.4	34.3	19.2	20.0	24.0	23.1	21.2	20.4
consisting only young persons within the household (under 30 years old)	5.2	4.6	8.9	10.2	7.5	12.3	13.4	12.3
consisting only old persons within the household (over 60 years old)	30.0	28.6	4.3	12.1	18.5	16.6	6.4	7.0
with head of low level of education	36.8	35.3	13.1	8.3	12.5	13.4	4.5	4.9
with head of high level of education	8.3	8.8	38.9	37.5	26.7	28.9	50.3	48.8
Subjective well off*	1.8	2.6	10.7	13.3	10.0	13.5	21.8	30.1
Housing-equipment well of**	7.5	7.2	18.5	19.8	19.8	23.4	29.3	30.9
Holiday abroad	13.7	15.6	12.9	23.3	31.1	31.6	30.7	35.5

* It is defined by self-categorization.
** Households in the top decile according to the housing-equipment index, see at Table 4.

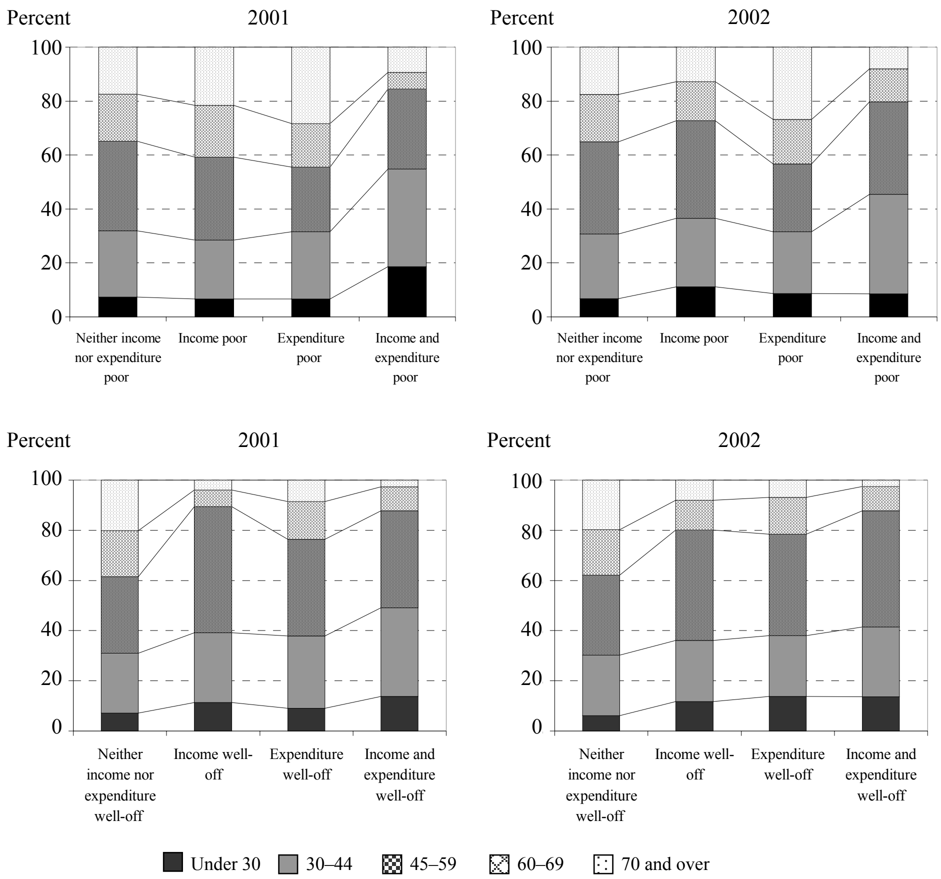
Further striking figures can be found in the last five rows of Table 4 above. While almost half of the expenditure poor households considered themselves poor in 2002, this was much less frequent, less than 30 percent among the double poor households. In respects of other dimensions of poverty, on the other hand, the proportions within the double poor households exceed markedly those within either only income or expenditure poor households. While e.g. the proportion of housing-equipment poor households among the income poor was only 18 percent and 22 percent among the expenditure poor households, it amounted to nearly 46 percent among the double poor households. However, perhaps the most important and striking figures are shown in the last row: these indicate that in both years multiple deprived and double poor households coincide to a great extent, almost four from every five double poor households are at the same time multiple deprived.

Looking now at the opposite end of the income and expenditure distributions there are a number of similarities in the characteristics of the income and expenditure well-off households. However, significant differences can also be experienced in respect of a few characteristics. Thus it is remarkable e.g. that the proportions of single member households and those consisting of old persons only are considerably higher among expenditure well-off households than among income well-off households.

The opposite is true in respect of households with highly educated head and those living in Budapest. Among double well-off households the proportion of households with highly educated head is strikingly high, while that with low level of education is insignificant, much lower than among either groups of households well-off from one aspect only. As Table 5 data below indicate only 10-13 percent of income or expenditure well-off households consider themselves being well-off, among double well-off households, however, this rate is more than double: it was 22 percent in 2001 and 30 percent in 2002. From Table 5 it can be concluded, furthermore, that the proportion of households who spend their holiday abroad is, in today Hungarian circumstances, a good indicator of being really well-off, almost 36 percent of double well-off households gave account of such occurrence in 2002, while this proportion is some what lower among expenditure well-off households and lower, 13-23 percent among income well-off households.

It is instructive, finally, to investigate and compare the structure of the different types of poor and well-off households by the age groups of the household head. First it must be noted that households with elderly head generally do not belong to neither poor nor well-off households. On the other hand, if they are poor or well-off, this relates primarily their expenditures. We can differentiate between two types of old households: one part of them did not yet get accustomed to the consumer type society, they do not spend all their incomes, give preference to save instead. The other type of old households, on the other hand, is of spending type, i.e. their expenditures exceed their current income making use of their past savings. It is noteworthy, furthermore, in connection with Figure 6 that young and middle aged households are over-represented among all types of well-off households, as well as among double poor households, but the bulk of well-off households consist of households where the head is in the second half of his/her economically active life.

Figure 6. Percent of poor and well-off households by age group of the head, 2001 and 2002



CONCLUSION

We deem that the efforts made in connection with our research were not fruitless, in fact they were remunerative. To the question: whether income or expenditure is better to characterize the welfare, the living conditions of households in Hungary today, a definite answer cannot be given, the answer depends on the aim of the investigation. However, our results indicate that we can describe the living conditions of the population, the poor and the well-off households more precisely if using both measures. Thus it can be concluded that the answer to the question in the title is neither income nor expenditure, but if an HBS contains data on both the incomes and the expenditures of the households both variables are to be taken into account in a complex manner when investigating the living conditions, the poverty and the wealth of the households. Data unequivocally indicate that the really, deeply poor are those poor in respect of both income and expenditure and the really wealthy are well-off both in income and expenditure.

REFERENCES

ÉLTETŐ, Ö. – FRIGYES, E. [1968]: New income inequality measures as efficient tools for casual analysis and planning. *Econometrica*. Vol. 36. No. 2.

Household Budget Survey, 2001. Annual Report [2002]. Hungarian Central Statistical Office. Budapest.

Household Budget Survey, 2002. Annual Report [2004]. Hungarian Central Statistical Office. Budapest.

HARDING, A. – GREENWELL, H. [2002]: Trends in Income and Consumption Inequality in Australia. Paper prepared for the 27th Conference of the IARIW. Stockholm.

HARDING, A. – GREENWELL, H. [2001]: Trends in income and expenditure inequality in the 1980s and 1990s. University of Canberra. Natsem.

HAVASI, É. [2003]: Poverty and Social Exclusion in Hungary Today. In: *Human Development Report 2001.* UNDP. Budapest.

THE RETURN AND RISK PROFILE OF EQUITIES AND EQUITY PORTFOLIOS AT THE BUDAPEST STOCK EXCHANGE*

GYÖNGYI BUGÁR[1] – GIANNI BARATTO[2] – ISTVÁN PREHOFFER[3]

This paper examines the risk and return characteristics of equities listed and traded in category 'A' at the Budapest Stock Exchange in the time period of 2001–2002. The performance of two portfolio strategies is also evaluated. It is shown empirically that a systematic portfolio allocation has several advantages to stock picking. Indeed, the portfolio strategies examined performed well not only on an ex post but also on an ex ante basis.

KEYWORDS: Risk and return of equities. Equity portfolios. Budapest Stock Exchange.

In response to the world-wide market downturn since 2000 and because of some unfavourable governmental steps and taxation reasons, the turnover and the capitalisation on the equity market of the Budapest Stock Exchange (BSE) has fallen significantly. Particularly, in 2001 the market turnover fell by 60 percent and the capitalisation of equities decreased by 16 percent as compared to the end of the year 2000 (*Statistical Report* [2001], p. 6–8). The negative trends of the earlier year seemed to take an upward turn in 2002. It is indicated by the fact that both the turnover and capitalisation of the equity market have increased by more than 9 and 3.5 percent, respectively (*Statistical Report* [2002], p. 3–4).

The aim of this paper is to study the risk and return characteristics of equities listed in category 'A' of the BSE over the period 2001–2002. The performance of two portfolio strategies is also analysed. We intend to show that a systematic portfolio allocation has several advantages over the approach of picking some individual equities to invest in, especially in times of undesirable market processes.

The structure of the paper is as follows. The next section provides a description of the database used in the analysis and the methodology applied. Empirical results are presented next, followed by some concluding remarks.

* The first author was generously supported in writing this paper by the Hungarian Academy of Sciences in the framework of the Bolyai Scholarship. The authors wish to thank *Eszter Kocsis* (Information Centre of the Budapest Stock Exchange) for all her help in data collection and providing up-to date information.

[1] Associate Professor, University of Pécs, Faculty of Business and Economy.

[2] Finalist Student, University of Padua, Faculty of Economics at the time of writing the paper: ERASMUS Student at the University of Pécs.

[3] Supply Chain Leader, Elcoteq Hungary. At the time of writing the paper: Finalist Student, University of Pécs, Faculty of Business and Economics, Business Degree Programmes in English.

DATA AND METHODOLOGY

The database for the analysis consists of the daily closing prices of the equities listed and traded in category 'A' at the Budapest Stock Exchange and those of the BUX index within the period of 3 January 2001 to 23 December 2002. In 2001, the category system for equities at the BSE was revised and equities were grouped into two categories, 'A' and 'B', using a modified set of criteria (*Annual Report* [2001], p. 21.). The capitalisation of the equities listed in category 'A' represented more than 90 percent of the total equity capitalisation in both years studied in our current paper (*Statistical Report* [2001] and [2002], p.14.).

Firstly, we excluded from our analysis the time series that were not complete, since it is not possible to compare securities on the base of different (or in any case, excessively different) time series. In particular, we excluded Graboplast and Csopak. The former was removed from the trading list in December 2001 and the latter was delisted in January 2002. Although, Pick was also removed from the trading list on 7 November 2002, in that particular case, we had enough data to perform the analysis.

Therefore, there were altogether 24 securities included in the study: Antenna Hungária, BorsodChem, Danubius, DÉMÁSZ, Egis, Fotex, Globus, Graphisoft, Humet, Inter-Európa Bank, MATÁV, Mezőgép, MOL, NABI, OTP Bank, Pannonplast, Pick, Prímagáz, RÁBA, Richter Gedeon, Synergon, TVK, Zalakerámia and Zwack Unicum. We also considered the BUX index.

Moreover, we had to find and adjust the prices resulting from split and reverse split[4]:

– In case of OTP, there was a 10-to-1 split in March 2002. It means that each shareholder received 10 new shares with the face value of 1/10th of the original for each 'old' (pre-split) shares held, and the old shares were withdrawn. This type of transaction has the direct result that the face value of all shares held remains unchanged. In order to handle the above-mentioned split properly, we multiplied all the after-split values (closing prices) in the time series by 10.

– In case of Humet, a 1-to-10 reverse split was made in September 2002, in which the existing stocks were replaced by new ones, for each 10 'old' share a new share was given with a face value 10 times higher than the original one. Therefore, we divided the after-split prices by 10.[5]

We decided to use weekly returns as the basis for the analysis. In our opinion, to a large extent this time unit is not influenced by the events that have but a limited and only daily impact on the trading of securities and, at the same time, it has the right sensitivity to the changes in the trend of the time series. For this reason, we took the closing prices of Wednesdays, as it is a day in the middle of the week and therefore their prices do not carry the effect of variables related to the beginning nor the end of the week. In those few cases, when Wednesdays were not applicable for ordinary business reasons we took the nearest day at our disposal.

[4] The announcement on split/reverse split can always be found among the BSE News (Press Releases/Orders) on the homepage of the BSE: www.bse.hu.

[5] Throughout the analysis we took the viewpoint of the investor who keeps the security once acquired during the whole period studied, i.e. we considered "buy-and-hold" decisions. Doing it otherwise, especially in case of split/reverse split, one can easily make a mistake when calculating returns based on unmodified security prices.

Based on the Wednesdays' prices, the weekly rates of return were calculated as follows:

$$R_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}} \qquad /1/$$

where $P_{i,t}$ and $R_{i,t}$ are the price and the rate of return on equity *i* on the week *t*, respectively. The weekly return defined above is given in percentage and can be regarded as a relative measure of profitability. After this step we had altogether 51 data in each time series of weekly returns for both years (the only exception was Pick, for which 44 return data were available in 2002).

From the time series of the weekly returns we calculated[6] the following values for each equities:

Average (weekly) return:

$$\overline{R}_i = \frac{1}{T}\sum_{t=1}^{T} R_{i,t} \qquad /2/$$

where *T* is the number of weeks considered.

Standard deviation of return[7]:

$$\sigma_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (R_{i,t} - \overline{R}_i)^2} \qquad /3/$$

Covariance of return with that of the BUX:

$$\sigma_{i,BUX} = \frac{1}{T}\sum_{t=1}^{T} \left(R_{i,t} - \overline{R}_i\right)\left(R_{BUX,t} - \overline{R}_{BUX}\right) \qquad /4/$$

where $R_{BUX,t}$ is the return on the BUX index on the week *t* and $\overline{R}_{BUX}$ is the average return on the BUX.

Risk index beta with respect to the BUX:

$$\beta_i = \frac{\sigma_{i,BUX}}{\sigma_{BUX}^2} \qquad /5/$$

where σ_{BUX} denotes the standard deviation of return of the BUX.

Correlation of returns for each pairs of equities:

$$\rho_{i,j} = \frac{\frac{1}{T}\sum_{t=1}^{T} (R_{i,t} - \overline{R}_i)(R_{j,t} - \overline{R}_j)}{\sigma_i \sigma_j} \qquad /6/$$

[6] Strictly speaking, we estimated the parameters in question relying on the sample, namely on the time series of weekly returns.

[7] To be more exact, instead of the above formula we used the empirical unbiased estimator for the standard deviation.

The risk index beta above is regarded as a measure of market related risk, often referred to as systematic risk (*Levy–Sarnat* [1984], p. 428–451). It can be interpreted and estimated as the slope of the linear time-series regression of the security return ($R_{i,t}$) on the return of the market portfolio, the BUX ($R_{BUX,t}$). In our case:

$$R_{i,t} = \alpha_i + \beta_i R_{BUX,t} + e_{i,t} \qquad /7/$$

where $e_{i,t}$ is the error term (the deviation from the regression line) and α_i is the regression constant (vertical intercept).

Assuming that the error term is uncorrelated with the return on the market portfolio and taking the variance of both sides of Equation /7/ we obtain:

$$\sigma_i^2 = \beta_i^2 \sigma_{BUX}^2 + \sigma_e^2 \qquad /8/$$

At this point we had all the input data to make a portfolio optimisation based on *Markowitz's* [1999] theory, the Mean-Variance criterion. A portfolio is a combination of the different securities selected by the investor. Technically, it is a vector of weights, i.e. the percentages of the total capital invested into the different securities. The return on a security as well as on the portfolio of securities must be handled as a random variable, because it is unknown at the beginning of the investment period when the investment decision making takes place.

In the Markowitz model the rule upon which the selection between different investment options is made is the following: an option F is preferred to an option G if and only if

$$E_F(R) \geq E_G(R) \quad \text{and} \quad \sigma_F^2(R) \leq \sigma_G^2(R) \qquad /9/$$

for all values of *R* (with strict inequality for at least one value of *R*).[8] In /9/ $E(R)$ and $\sigma^2(R)$ denote the expected return and variance (the square of the standard deviation) of return, respectively.

The expected return is taken as an indicator of the investment's average profitability; the variance of return serves as an indicator of its risk. Instead of the variance, the standard deviation of return can also be regarded and interpreted as a risk measure. As an estimator of the expected return, the average return (i.e. the mean) $\overline{R}_i$ can be used. In this context formula /8/ can be referred to as the decomposition of risk, where $\beta_i^2 \sigma_{BUX}^2$ is the systematic risk component and σ_e^2 is the non-systematic risk component (the former can also be called non-diversifiable risk and the latter is cited as diversifiable risk).

The return on a portfolio can be formulated as:

$$R_p = \sum_{i=1}^{N} x_i R_i \qquad /10/$$

[8] Regarding a more detailed discussion of the above mentioned Mean-Variance Efficiency Criterion and its applications see e.g. *Markowitz* [1999], p. 129–201 or *Levy–Sarnat* [1984], p. 235–355.

where R_i is the return on security *i*, x_i is the weight in security *i* (i.e. the proportion of money invested in it) and *N* is the number of securities held in the portfolio.

In order to determine the expected return on a portfolio, the weighted average of the expected returns of the securities it includes need to be calculated. Therefore, the portfolio's expected return can be expressed as:

$$E_p = \sum_{i=1}^{N} x_i \overline{R}_i \qquad /11/$$

where $\overline{R}_i$ is the average return on security *i*.

The variance of a portfolio is influenced both by the variance of the individual securities within and by the correlation between the various pairs of securities:

$$\sigma_p^2 = \sum_{i=1}^{N} x_i^2 \sigma_i^2 + 2\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j>i}}^{N} x_i x_j \sigma_i \sigma_j \rho_{i,j} \; . \qquad /12/$$

The portfolios we seek to identify are the efficient portfolios. A portfolio is efficient if there is no other portfolio preferred with respect to the conditions in /9/. It means that an efficient portfolio is the investment with the highest expected return on a certain level of risk or it is the one with the lowest risk on a certain level of expected return.

In order to determine the combinations of securities that comprise the efficient portfolios one has to solve the optimisation problem as follows:

$$\min \sigma_p^2 = \sum_{i=1}^{N} x_i^2 \sigma_i^2 + 2\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j>i}}^{N} x_i x_j \sigma_i \sigma_j \rho_{i,j}$$

subject to

$$E_p = \sum_{i=1}^{N} x_i \overline{R}_i \qquad /13/$$

$$\sum_{i=1}^{N} x_i = 1 \qquad 0 \le x_i \le 1 \qquad i = 1, 2, ..., N$$

The objective function expresses the aim to identify the portfolio weights (x_i) for each feasible expected portfolio return (E_p) so that the risk of the portfolio (σ_p^2) is minimised. The expected return, the standard deviation of return of each security and the correlation matrix of returns are used as the input parameters estimated from the sample. The last constraint in /13/ means that only long positions are allowed, i.e. short sales are excluded. It implies that it is not possible to sell securities the investor does not own and use the proceeds to invest in other securities. This restriction is in coincidence with the regulation predominant at the BSE where short sales are forbidden.[9]

[9] A detailed discussion of short sales can be found in *Sharpe–Alexander* [1990].

To make the optimisation in practice we used the software Invest[10], which has been developed to accompany 'Modern Investment Theory' by *Haugen* [1997].

EMPIRICAL RESULTS

The main figures on risk and return characteristics of the equities are presented in Table 1 and Table 2, for the year 2001 and 2002, respectively. In addition to the equities, the BUX index is also shown in the tables. The BUX index is used as the proxy of the market portfolio, namely the benchmark with respect to the beta values are estimated. Beyond the average (weekly) return and the standard deviation, we report the performance ratio, defined as the average return per unit of standard deviation. Because the standard deviation is considered as a measure of risk, the performance ratio can be regarded as the risk-adjusted average return.[11] Furthermore, we list the beta, the t-statistic (that shows whether beta is significantly different from 0), the R^2 (that indicates the explanatory power of the model /7/) and also the term labelled by risk ratio. According to our definition, the last one is the ratio of the non-systematic (or the so-called diversifiable) and the total risk component (see the decomposition of risk given by formula /8/).

Table 1

The main risk and return figures of category 'A' equities of the BSE in 2001

Equity	Average return	Standard deviation	Performance ratio	Beta	Risk ratio	*t*-value	R^2
	(percent)						
ANTENNA	–1.23	6.21	–	0.89	0.82	3.30*	0.18
BCHEM	–0.39	7.16	–	1.56	0.58	6.00*	0.42
DANUBIUS	–0.43	4.36	–	0.15	0.99	0.75	0.01
DÉMÁSZ	–0.60	5.20	–	0.22	0.98	0.90	0.02
EGIS	0.18	3.89	0.05	0.57	0.81	3.41*	0.19
FOTEX	–0.54	7.53	–	1.06	0.82	3.26*	0.18
GLOBUS	0.70	6.08	0.11	0.96	0.78	3.73*	0.22
GRAPHISOFT	–1.07	8.53	–	1.32	0.79	3.64*	0.21
HUMET	–0.64	9.15	–	0.38	0.98	0.89	0.02
IEB	0.28	4.93	0.06	0.60	0.87	2.71*	0.13
MATÁV	–0.11	6.60	–	1.87	0.28	11.16*	0.72
MEZŐGÉP	–0.27	4.72	–	0.56	0.87	2.67*	0.13
MOL	0.21	4.36	0.05	1.14	0.39	8.75*	0.61
NABI	–0.30	4.75	–	0.79	0.75	4.01*	0.25
OTP	0.20	3.18	0.06	0.69	0.58	6.00*	0.42
PICK	–0.72	7.09	–	0.78	0.89	2.46*	0.11
PPLAST	–0.92	5.89	–	1.11	0.68	4.76*	0.32
PRÍMAGÁZ	0.59	8.67	0.07	1.12	0.85	2.94*	0.15
RÁBA	–0.81	5.57	–	0.85	0.79	3.57*	0.21
RICHTER	–0.03	3.14	–	0.44	0.82	3.25*	0.18
SYNERGON	0.35	12.96	0.03	1.70	0.85	2.98*	0.15
TVK	–0.34	8.48	–	1.65	0.66	5.03*	0.34
ZALAKERÁMIA	–0.59	6.11	–	0.90	0.80	3.46*	0.20
ZWACK	0.00	3.41	–	0.31	0.93	1.95**	0.07
BUX	–0.05	2.99	–	1.00	–	–	–

* β is significantly different from zero at 5 percent level.
** β is significantly different from zero at 10 percent level.

[10] The software was programmed by David Y. Tan, Joe Dada III, Kim Peters and Craig Lewis.

[11] However, one should be cautious in using the performance measure when the average return is negative because its value is completely misleading. If we compare two securities with the same negative return, the negative performance ratio is higher for the security which is less risky. That is why we simply omitted the values in the case of equities with negative average return.

The first result we got from our research is that the number of stocks with positive average return is only 7 in 2001 and 13 in 2002 out of the total pool of 24 equities analysed (the rows of the tables belonging to the securities with positive returns are highlighted). Indeed, the market passed through a crisis for which the main causes are well-known. The negative trend in security returns was a global phenomenon and therefore we consider it to be a 'systematic' reaction of the Budapest Stock Exchange to a world-wide recession. It is confirmed by a working paper (*A csatlakozás előtt álló*... [2003], BSE Publication, p. 6) that describes the main characteristics of the stock exchanges of the accession countries of Central and Eastern Europe in the period of 2001-2002. As an indicator of the global downturn in Europe, the study refers to a –35 percent change in the value of the FTSE Eurotop 100 index (which consists of the shares of 100 blue-chip companies in the European Union) from 2001 to 2002.

In 2001 the average weekly return ranged between –1.23 and 0.7 percent and the standard deviation of returns altered between 3.14 and 12.96 percent. In 2002, the average return was within the range of –2.08 and 1.47 percent, while the standard deviation of returns was between 2.53 and 12.91 percent.

The performance ratio seems to be quite low on average: few stocks have a good payout for the risk implicit in the share. In 2001, the highest performance ratio (0.11) was registered for the equity with the highest average return (Globus). In 2002 IEB has shown the highest performance (0.2) with a relatively big average return (1.12%) and modest risk (5.7%).

Table 2

The main risk and return figures of category 'A' equities of the BSE in 2002

Equity	Average return	Standard deviation	Performance ratio	Beta	Risk ratio	*t*-value	R^2
	(percent)						
ANTENNA	-0.16	5.90	–	0.89	0.72	4.41*	0.28
BCHEM	0.19	3.67	0.05	0.45	0.81	3.36*	0.19
DANUBIUS	0.41	4.89	0.08	-0.11	0.99	-0.54	0.01
DÉMÁSZ	0.38	3.45	0.11	0.30	0.91	2.25*	0.09
EGIS	0.60	5.42	0.11	0.89	0.66	5.02*	0.34
FOTEX	-0.62	5.80	–	0.94	0.67	4.89*	0.33
GLOBUS	0.16	4.35	0.04	0.37	0.91	2.18*	0.09
GRAPHISOFT	-1.49	5.67	–	0.78	0.76	3.90*	0.24
HUMET	-2.08	12.91	–	0.42	0.99	0.81	0.01
IEB	1.12	5.70	0.20	0.73	0.80	3.55*	0.20
MATÁV	-0.06	4.61	–	1.06	0.33	9.95*	0.67
MEZŐGÉP	-0.50	5.29	–	0.56	0.86	2.80*	0.14
MOL	0.24	3.93	0.06	0.85	0.42	8.26*	0.58
NABI	-0.18	3.93	–	0.37	0.89	2.46*	0.11
OTP	0.75	5.17	0.15	1.33	0.17	15.45*	0.83
PICK	0.39	4.91	0.08	0.02	1.00	0.11	0.00
PPLAST	-1.43	4.69	–	0.65	0.76	3.97*	0.24
PRÍMAGÁZ	1.47	12.42	0.12	0.90	0.93	1.85**	0.07
RÁBA	-1.02	3.61	–	0.34	0.89	2.50*	0.11
RICHTER	0.07	4.81	0.01	1.03	0.43	8.04*	0.57
SYNERGON	-0.45	5.81	–	1.02	0.61	5.57*	0.39
TVK	0.57	4.42	0.13	0.12	0.99	0.68	0.01
ZALAKERÁMIA	-0.27	4.40	–	0.47	0.86	2.83*	0.14
ZWACK	0.15	2.53	0.06	-0.01	1.00	-0.05	0.00
BUX	0.26	3.54	–	1.00	–	–	–

* β is significantly different from zero at 5 percent level.
** β is significantly different from zero at 10 percent level.

Based on the figures reported in Table 1 and Table 2, we can observe some improvement throughout the two-year period. The crisis had its worst effects in 2001 and this line of reasoning can be supported by several facts:

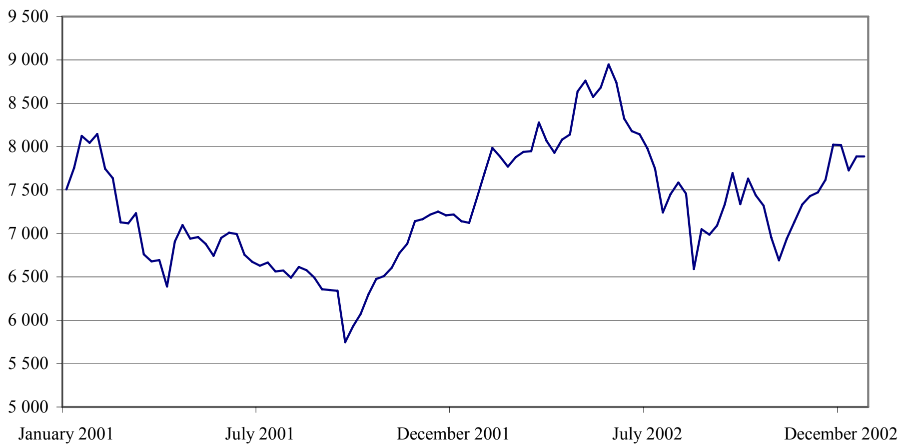
– there were only 7 securities with positive average returns in 2001 against 13 in 2002.

– when comparing the values of average return and standard deviation of return (risk) for the same security we realise that in 16 cases (out of the 24) the average return was higher, and in 16 cases the risk was lower in 2002 than in 2001. 10 equities had a preferable risk and return profile (i.e. the higher return and lower risk at the same time) in 2002 than in the previous year. Consequently, in 2002 there were only two securities with lower average return and higher risk as compared to 2001.

– the grand mean of the average returns was higher in 2002 (–0.07%) than in 2001 (–0.27%) but it was still negative, and it was accompanied by a lower mean risk (standard deviation). For the latter the respective values were 5.35 percent (in 2002) and 6.17 percent (in 2001).

– by looking at Figure 1, where the values of the BUX index are displayed over the period studied, it is easy to recognise that the negative trend of price movements has changed to a positive one. This can be confirmed by calculating and comparing the annual returns. In fact, the annual return of the BUX index was –4.91 percent in 2001 and 10.7 percent in 2002. According to the Annual Report (*BSE* [2002], p. 4), considering the return on the BUX in 2002, the BSE became the fourth best-performing exchange in the world.

Figure 1. The values of the BUX index from January 2001 to December 2002



It is worth mentioning that the annualised standard deviations[12] fall into the range of 22.6–93.5 percent in 2001 and 18.2–93.1 percent in 2002, respectively. The annualised

[12] The annualised standard deviation can be calculated by multiplying the weekly standard deviation by $\sqrt{52}$.

mean values of the standard deviation are 44.5 percent (2001) and 38.6 percent (2002). These values are high even in the context of emerging markets (for comparison see *Bernstein* [2000], p. 6 and 9-28).

When looking at the 'evolution' of betas it can be observed that in 18 cases out of 24 the beta of the same equity has decreased over the period studied. The average value of beta was 0.9 in 2001 and 0.6 in 2002. In 2001 altogether 9 equities fell into the aggressive category, with a beta higher than 1. In 2002 there were only 4 aggressive equities. All in all, the equities seemed to become more defensive in 2002 than they were in 2001.

With only a few exceptions, the *t*-statistics support the notion that beta is significantly different from zero. However, the explanatory power of model /8/ which explains the changes in equity prices through that of the market (represented by the BUX) is very low in general. The equity with the highest R^2 value is MATÁV in 2001 and OTP in 2002. This can be explained by the fact that these are the equities with the highest capitalisation, and also with the highest weight in the BUX basket (*Statistical Report* [2001], p. 10, 18 and also *Statistical Report* [2002], p. 5, 14). Consequently, the overall performance of the market is highly influenced by price fluctuations of these securities (maybe rather than the other way around).

So far we have not discussed the risk ratio, namely the ratio of the non-systematic and the total risk. As shown in Table 1 and Table 2, the risk ratio is very high in general. On average, it is almost the same in the two years studied (0.77 in 2001 and 0.76 in 2002).

It is remarkable that the risk ratio and the value of R^2 sum up to 1. It is simply a technical result, implied by the definition of beta and the derivation of formula /8/ for decomposition of risk.[13]

The results gained from our study on individual securities confirm that the BUX has quite a low influence on the equity prices. It follows that the resulting beta values need to be interpreted and used with caution. Therefore, we do not suggest to apply beta as a risk measure instead of the standard deviation of returns (or equivalently the variance), since a fair amount of volatility in security returns is not accounted for[14].

In the construction of portfolios for each year we decided to involve securities with positive return only. This way, the number of equities involved in the portfolio optimisation was 7 in 2001 and 13 in 2002. As mentioned before, the input parameters for portfolio optimisation are the average returns, the standard deviations of returns (see highlighted values in Table 1 and 2) and the correlations between the different pairs of security returns. It is clear (see formula /12/) that the lower the correlation terms of the different pairs of security returns are, the higher the risk reduction benefit of a portfolio can be.

The correlation terms for the equities with positive average return are reported in Table 3 and Table 4.

In 2001, all the correlation coefficients were positive, with values below 0.5. The highest term was experienced between the returns of MOL and Inter-Európa-Bank (0.47). The average of the correlation terms is 0.23 (by excluding the ones located in the diagonal).

[13] A proof of this statement can be found in *Levy–Sarnat* [1984], p. 436-437.
[14] An overview of the problems related to beta as a risk measure is given in *Bugár* [1998a].

Table 3

Correlation matrix of equity returns in 2001

Equity	EGIS	GLOBUS	IEB	MOL	OTP	PRÍMAGÁZ	SYNERGON
EGIS	1.00	0.32	0.08	0.35	0.20	0.04	0.02
GLOBUS		1.00	0.30	0.36	0.23	0.22	0.20
IEB			1.00	0.47	0.03	0.36	0.07
MOL				1.00	0.31	0.23	0.10
OTP					1.00	0.28	0.33
PRÍMAGÁZ						1.00	0.43
SYNERGON							1.00

Table 4

Correlation matrix of equity returns in 2002

Equity	BCHEM	DANUBIUS	DÉMÁSZ	EGIS	GLOBUS	IEB	MOL	OTP	PICK	PRÍMAGÁZ	RICHTER	TVK	ZWACK
BCHEM	1.00	–0.04	0.04	0.15	0.27	0.02	0.37	0.45	0.06	0.07	0.26	0.42	–0.08
DANUBIUS		1.00	0.05	–0.02	–0.15	0.22	–0.03	–0.13	0.09	–0.02	0.07	0.30	–0.05
DÉMÁSZ			1.00	0.27	0.04	0.33	0.26	0.20	–0.04	–0.03	0.19	0.05	0.24
EGIS				1.00	0.10	0.27	0.36	0.41	–0.08	0.03	0.57	0.09	0.11
GLOBUS					1.00	0.12	0.36	0.32	0.11	0.06	0.23	–0.04	0.19
IEB						1.00	0.29	0.31	–0.12	0.14	0.40	–0.02	0.00
MOL							1.00	0.64	0.01	0.25	0.37	0.17	0.26
OTP								1.00	0.05	0.17	0.65	0.08	–0.08
PICK									1.00	0.07	–0.01	0.01	0.33
PRÍMAGÁZ										1.00	0.23	0.15	0.09
RICHTER											1.00	0.07	–0.14
TVK												1.00	0.02
ZWACK													1.00

In 2002, the average is significantly lower, with the value of 0.15. We can also observe the presence of 17 negative coefficients (out of 78), which amount to about one-fifth of the terms. The highest correlation registered is 0.65 (for the pair of OTP and Richter) and the lowest one is –0.15 (in case of Danubius and Globus).

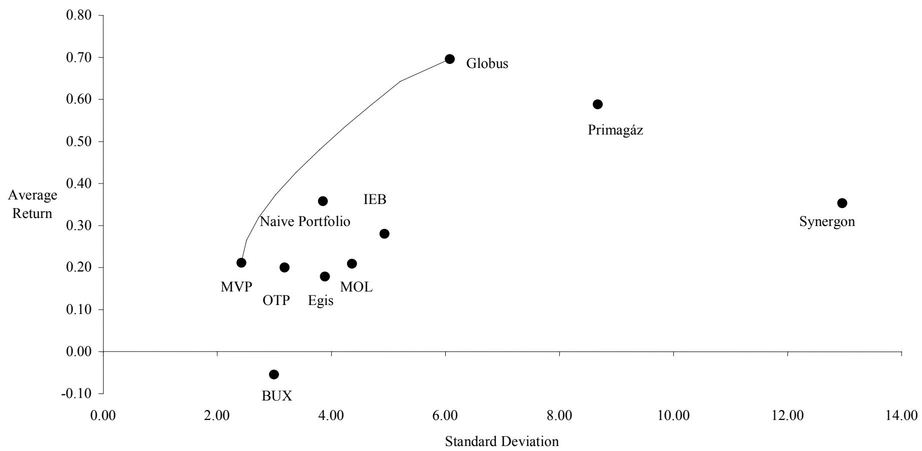
With a view to the portfolio optimisation, we made the calculations for each year with the help of the software Invest, excluding short sales as they are not applied in practice (otherwise, it would be a mere theoretical exercise).

The results in terms of efficient frontiers are shown in Figure 2 and 3. The continuous curve in each figure represents the efficient frontier. The risk and return combination of the individual equities involved in portfolio optimisation is given by the discrete points (each point is labelled by the name of the equity). In addition, the risk and return combination for the so-called Naïve Portfolio and that of the BUX is also plotted in the figures. The Naïve Portfolio is the equally weighted portfolio, which contains all the securities included in the portfolio selection process with equal proportion for each. Clearly, there is a special portfolio denoted by MVP (Minimum-Variance-Portfolio), the efficient portfolio with the lowest possible risk (variance of return).

Prior to performing the analysis of the efficient portfolios there is an important issue that needs clarification. Technically, we have an infinite number of efficient portfolios represented by the different risk-expected return combinations to choose from. As it can be seen in Figures 1 and 2, the average return is a strictly (monoton) increasing function of the risk (measured by the standard deviation). It means that undertaking a higher risk is compensated by a higher level of expected return. The software we applied to conduct the research gives an opportunity to choose any portfolio on the efficient frontier by typing in the required return on the portfolio. However, in reality the investor has to select a particular portfolio among the efficient ones, i.e. to follow a specific investment strategy. We regard it as a systematic portfolio allocation. In this paper we evaluate the performance of two portfolios, namely we simulate two investment strategies. The first one creates the Naïve Portfolio (NP) the second one constitutes the Minimum-Variance-Portfolio (MVP). Constructing the NP is probably the simplest way to benefit from diversification without requiring any sophisticated method for portfolio optimisation. The advantage of the MVP, especially in a risky period with highly volatile equity returns, is that it has the highest potential for reducing the risk. In brief, these were the reasons for choosing the above mentioned two portfolio allocation strategies.

Looking at and comparing Figure 2 and 3 it can be realised that the equity with the highest average return, as the extreme right point of the curve, is contained in the efficient frontier (in 2001 the equity with the highest return was Globus and Prímagáz in 2002). It is always the case when short sales are excluded. In general, the portfolio is becoming more diversified, i.e. contains more securities as we are "going down" on the efficient frontier towards the MVP.[15]

Figure 2. Risk and Return of Equities and Efficient Portfolios in 2001
(percent)

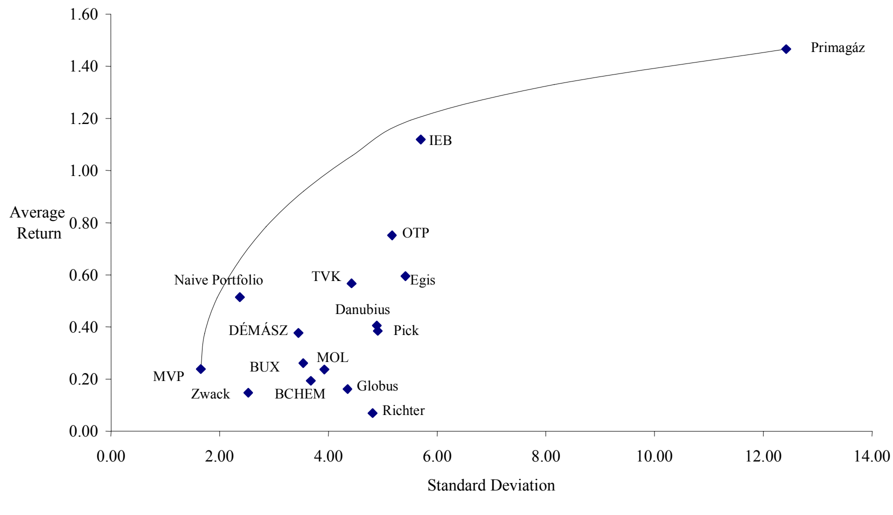


[15] Here we will not report the composition of the efficient portfolios other than the MVP. For readers interested, the results are available from the authors upon request.

In 2001, the average return and the risk of the efficient portfolios was in the range of 0.21–0.70 percent and 2.42–6.08 percent, respectively. Looking at Figure 2 it can be observed that all individual equities (except Globus) and the BUX are far from being efficient. In choosing any of them, there is always an efficient portfolio with dominant risk-return characteristics, namely one with a higher average return for the same level of risk or with a lower risk for the same level of expected return. The above statement is more or less true for 2002 as well, with the exception of Prímagáz (which is efficient). But in that year there is one more exception: Inter-Európa-Bank, which is 'nearly efficient'. Needless to say, it was the equity with the highest performance. In both years the performance of the NP seems to be quite good (below this issue is analysed more in detail). In 2002 the ranges, in which the average return and the risk of the efficient portfolios fell, were wider than those of 2001. The range of the average return was 0.24–1.47 percent and the risk fell into the interval of 1.66–12.42 percent.

It is remarkable that the position of those equities involved in the 2001 and 2002 portfolio as well has changed greatly on the 'risk-return map', referring to a change both in their risk-return profile and in their performance from 2001 to 2002. The only exception is MOL with a relatively stable risk and average return. However, in five cases out of the total six, we can report an improvement in performance.

Figure 3. Risk and Return of Equities and Efficient portfolios in 2002
(percent)



In Table 5 the composition of the MVP is presented for both years. In 2001, the MVP has contained only 3 equities out of the 7 included in the portfolio optimisation. In 2002, it was more balanced in this sense, excluding only 4 out of the total pool of 13 securities considered. Obviously, each year, the equity with the lowest risk (standard deviation of return) had the highest weight in the MVP (see Tables 1 and 2 for comparison). In 2001 OTP and in 2002 Zwack had the highest proportion, 50.1 percent and 40.4 percent, respectively. It might seem surprising that MOL had no part in the portfolios of the years studied, despite the fact that it had lower risk than some of the equities that are included.

Table 5

Composition of the Minimum Variance Portfolio (MVP)
(percent)

Equity	2001	2002
BCHEM	–	16.7
DANUBIUS	–	11.4
DÉMÁSZ	–	11.6
EGIS	28.5	0.0
GLOBUS	0.0	5.6
IEB	21.4	1.2
MOL	0.0	0.0
OTP	50.1	0.0
PICK	–	2.7
PRÍMAGÁZ	0.0	0.0
RICHTER	–	7.0
SYNERGON	0.0	–
TVK	–	3.4
ZWACK	–	40.4

A likely explanation is that MOL's return on average has a higher correlation with the other security returns involved in portfolio optimisation than the other equities. According to formula /12/, when the risk of the portfolio is being minimised, it is not only the risk of the individual equities taken into account but the correlation between the returns as well.

Next, we examine the performance of the MVP and that of the NP. The BUX index is used as a benchmark for evaluation. The results are summarised in Tables 6 and 7.

Table 6

The main characteristics of two portfolio strategies and the BUX index in 2001

Denomination	Average Return	Standard Deviation	Performance Ratio	Beta	Risk Ratio
	percent				
MVP	0.21	2.42	0.087	0.64	0.38
Naïve Portfolio	0.36	3.85	0.094	0.97	0.43
BUX	–0.05	2.99	–	1.00	–

In drawing a comparison of the risk of the MVP in Table 6 to the risk of the individual equities involved in the portfolio optimisation (see highlighted rows in Table 1), it can be calculated that the standard deviation of the MVP is about 24 percent and 81 percent lower than that of the individual equity with the lowest and the highest standard deviation, respectively. Therefore, in 2001 the risk reduction benefit from creating the MVP can be regarded as quite good. Further, the risk and return parameters of the MVP as compared to those of the NP are quite distinct. The return on the MVP is lower and it is a less risky portfolio. In itself, the MVP is designed to reduce the risk, hence its lower exposure to risk through investment does not come as a surprise. Still, it is remarkable that the performance of the NP is superior to that of the MVP. Both portfolio allocation strategies show a low performance but still a better one than the individual equities (except Globus[16]). It is also worth mentioning that the NP has a higher beta[17] and the higher risk ratio as well. Using the BUX index as a benchmark to evaluate the performance of the two portfolios, all in all, we can conclude that their performance is not bad compared to the general state of the stock market in 2001.

In 2002, both the performance of the BUX and the two portfolio allocation strategies have improved greatly as compared to 2001. It is due to the higher average returns and, in case of the two portfolio allocation strategies, to a lower level of risk as well. The performance ratio of the MVP is more than one and a half times higher than it was in 2001.

Table 7

The main characteristics of two portfolio strategies and the BUX index in 2002

Denomination	Average Return	Standard Deviation	Performance Ratio	Beta	Risk Ratio
	percent				
MVP	0.24	1.66	0.145	0.20	0.82
Naïve Portfolio	0.50	2.37	0.211	0.53	0.38
BUX	0.26	3.54	0.073	1.00	–

The NP has produced an even greater impovement in terms of performance: its performance ratio is more than two times higher in 2002 than it was in 2001. The betas of both portfolios are lower than they were in 2001. The risk ratio of the MVP is high, which is due to its low beta, namely its low sensitivity to the market volatility.

Comparing the results shown in Table 7 to those presented in Table 2 we can conclude that both portfolios outperform almost all individual equities. Indeed, for the NP we got a higher performance ratio than for any of the equities. The NVP was outperformed by IEB and OTP only. It is notable that the NP repeatedly had a better performance than the MVP in 2002. In a similar research made on blue-chips traded at BSE, *Bugár* [1998b] has also reported on the very good performance of the NP.

As shown in Figure 3, the NP is located rather close to the efficient frontier. Decidedly, the efficient portfolio with the same average return (0.5%) as the NP has a standard deviation of 1.93 percent. So, it has a performance ratio of 0.259, which is 23 percent higher than that of the NP. The creation of an efficient portfolio is a rather sophisticated process, requiring time and effort to estimate the parameters and to implement the portfolio optimisation. Taking this fact into account, we can safely say, it is not necessarily worth to carry out the process.

One can argue that our analysis was performed on an ex post basis. In fact, the problem of this approach is that it only reveals past the event what should have been done ear-

[16] An investment into Globus is much more risky than the above-mentioned portfolio selection strategies. Here it should be emphasised that portfolio allocation pays in terms of stabilising the return, i.e. reducing the risk but not necessarily in terms of increasing the performance.

[17] The beta of a portfolio is a weighted average of the securities' betas included in it.

lier on. Consequently, the benefits detected are potential only so they cannot be realised. In order to overcome this difficulty we also examined the performance of the MVP and the NP on an ex ante basis, i.e. we determined the returns which would have been realised on average in 2002 on the portfolios set up in 2001 and kept unaltered.

Based on the information in Table 2 and Table 5, the average (weekly) return in 2002 on the MVP(2001) is:

$$\overline{R}_{MVP\,(2001)} = 0.6 \cdot 0.285 + 1.12 \cdot 0.214 + 0.75 \cdot 0.501 \approx 0.79\ (\%)\ .$$

Similarly, the average return can be realised in 2002 on the NP(2001):

$$\overline{R}_{NP\,(2001)} = \frac{1}{7} \cdot (0.6 + 0.16 + 1.12 + 0.24 + 0.75 + 1.47 - 0.45) \approx 0.56\ (\%)\ .$$

As it can be seen the MVP outperformed the NP on an ex ante basis. Comparing these results to the average return on individual equities presented in Table 2, it is clear that in 2002 there are only two equities which performed better than the MVP(2001) and five equities which outperformed the NP(2001). In addition, based on the prior information on the average returns as well as the optimal portfolio weights in 2001, the average return was even higher on the portfolios set up at the end of 2001 than it was in case of the portfolios produced with the help of the data from 2002 and including the equities with positive return into the portfolio (see Table 7). It served as a lesson to prove that a systematic portfolio allocation by using a 'buy and hold' strategy can be more successful than continuously changing the equities selected and included in the portfolio. To conclude, in the long run it might be more profitable to apply the same portfolio strategy on the stable set of securities.

*

In this paper we studied the risk and return characteristics of equities listed and traded in category 'A' at the Budapest Stock Exchange over the time period 2001-2002. We also made a portfolio optimisation based on *Markowitz's* [1999] theory, the Mean-Variance criterion. The expected return was taken as an indicator of the investment's average profitability, and the standard deviation of return served as an indicator of its risk. Furthermore, we estimated the beta values, and tested the explanatory power of the linear regression model of security return on the return of the BUX index. The performance of two portfolio strategies was also evaluated. The major findings of the analysis can be summarised as follows.

Both the analysis of individual equities and the efficient portfolios supported that the stock exchange passed through a crisis which had its worst effects in 2001. Indeed, we experienced an increase of the average return and the performance and a decrease of the risk from 2001 to 2002.

It was found that the influence of the BUX on equity prices is quite low. As a consequence, the beta values we got should only be interpreted and used with care. On the basis of our empirical findings it cannot be recommended to apply beta as a risk measure instead of the standard deviation of returns, because in this case a large part of volatility in security returns would not be explained.

On an ex post basis the Naïve Portfolio had a better performance than the Minimum-Variance-Portfolio in both years. Considering that to create an efficient portfolio is a sophisticated process requiring time and effort to estimate the parameters and to implement the portfolio optimisation, it seems that we can be satisfied with the benefits promised by the naïve way of diversification.

On an ex ante basis the Minimum-Variance-Portfolio has shown a better performance than the Naïve Portfolio. However, both of them resulted in an even higher average return than their ex post counterparts. The ex post portfolios have been set up under the condition of using the data from 2002 and involving the equities with positive return into the portfolio, while their ex ante counterparts have been constructed on the basis of utilising prior information on the average returns as well as the optimal portfolio weights at the end of 2001. Therefore, it has been confirmed that a systematic portfolio allocation by using a 'buy and hold' strategy can be more successful than continuously changing the equities selected and included in the portfolio.

REFERENCES

Annual Report [2001]: *Budapest Stock Exchange*, p. 1–48.

Annual Report [2002]: *Budapest Stock Exchange*, p. 1–36.

BERNSTEIN, W. J. [2000]: *The intelligent asset allocator: How to build a portfolio to maximize returns and minimize risk*. McGraw-Hill. New York.

BUGÁR GY. [1998a]: Az értékpapírok piaci kockázatának méréséhez. *Bankszemle*. Vol. 42. No. 1–2. p. 54–62.

BUGÁR GY. [1998b]: Hatékony részvénykombinációk és kockázatcsökkentési stratégiák a BÉT-en. *Bankszemle.* Vol. 42. No. 8. p. 59–67.

A csatlakozás előtt álló kelet-európai országok részvénypiacai [2003]. *Budapest Stock Exchange Publication*. p. 1–7. (also published in the online business daily paper, Portfolio.hu: www.portfolio.hu/reszveny.tdp (8.15 a.m. 01.04. 2003))

HAUGEN, R. A. [1997]: *Modern investment theory*. Prentice-Hall. Englewood Cliffs. New Jersey.

LEVY, H. – SARNAT, M. [1984]: *Portfolio and investment selection – Theory and practice*. Prentice Hall. New Yersey.

MARKOWITZ, H. M. [1999]: *Portfolio selection: efficient diversification of investments*. Basil Blackwell. Oxford.

SHARPE, W. F. – ALEXANDER, G. J. [1990]: *Investments*. Prentice-Hall. Englewood Cliffs. New Jersey.

Statistical Report [2001]. *Budapest Stock Exchange*, p. 1–28.

Statistical Report [2002]. *Budapest Stock Exchange*, p. 1–23.

DISTANCES AND DIRECTIONS OF INTERNAL MIGRATION IN HUNGARY*

SÁNDOR ILLÉS[1]

The gravity centre is a classic regional analytical method which requires masses. This mass can be a multitude of people but also any other absolute quantity. The gravity centres of the migration can be considered as one of the variants of the gravity centres of the population. Adopting ourselves to the nature of the migration flows we can not grasp the migration itself with one but only with two gravity centres: with the gravity centre of out-migration and the gravity centre of in-migration. The gravity centre of migration must characterise the regional distribution of the migrate subpopulation directly before and after the event of migration. Our purpose was to state the average distance and characteristic direction of the internal migration flows in Hungary. Used the centrographic approach we got a detailed picture on the development of the direction and distance of migration.

Data used for the research were the data of internal migration by settlements for the years 1984-2002 and the geographical co-ordinates of the same settlements, both supplied by the Hungarian Central Statistical Office. The methods were simple. The gravity centre is the weighted arithmetic means – weighted with the migrants – of the co-ordinates of latitude and longitude of the settlement centres. After all the gravity centres of migration are nothing else but the mean values of the regional distribution of the out-migrants and in-migrants. Various methods are known for the characterisation of the situation around the gravity centre. One of them, Bachi's 'd' standard distance was calculated to prove two kinds of spatial selectivity of migration in Hungary. In this contribution we chose the whole country as a spatial unit to be studied, but regional and county gravity centres were computed during the research.

The gravity centres of out-migration and in-migration were separated from one another and were very near to the gravity centres of the total population. But they were not exactly in the same place. Conclusion can be made that there are territorial selectivities: on the one hand between the sending and receiving settlements (type 1), on the other hand the spatial distribution of the migrate subpopulation is not simply a representative sample of the spatial distribution of total population (type 2). In the country as a whole the gravity centres of in-migration located to the west from the gravity centres of out-migration, thus in the period studied the dominant direction of the migrations proceed to west. The average length of the way made by the permanent migrants was greater compared to the temporary migrants all of the investigated period. The distances between the national gravity centres of migration gradually shortened until 1997, after which a slight increase could be observed.

KEYWORDS: Internal migration; Gravity cenre method; Regional selestivity.

* Preliminary version of the paper was discussed at the European Population Conference, Warsaw, Poland, 26-30 August, 2003.

[1] Senior research fellow, HCSO Demographic Research Institute.

The significant social structure changes which had taken place from the 1950's onwards entailed increasing territorial mobility (it is enough to think of the effects of industrialisation and the forced emergence of agricultural co-operatives) (*Sárfalvi* [1965]; *Compton* [1971]; *Kovács* [1985]). The end of the 1980's, however, brought along changes in a very different character (*Rédei* [1986]; *Daróczi* [1998]; *Kovács* [2002]). Since after international migration came into the limelight at the end of the 1980's, examinations of internal migration movements necessarily became sidelined. Besides, internal migration took an unexpected turn in the period of transformation in Hungary.

The effect of political, economic and social changes the transformation of the 1980's appeared not in the increasing of levels but precisely in a sharp decline in the volume of migration within Hungarian society. The transformation affected individuals and families in such a way that they came to discard or postpone their possible plans of move. The dominant experience of this period of 'metamorphosis' was a feeling of insecurity which, however, in most Central and Eastern European countries did not reach the level of utter hopelessness required for inducing a mass scale internal and international migration. The unpredictability of the near future kept people place-bound and they did not willingly make risky plans for migration in the early 1990's. They drew a greater sense of local security from having an established home and a network of personal contacts than might have been afforded even by a solid job offer at a distant location. The relatively small number of people who accepted risky migration in return for considerable advantages tried their luck not mainly within their own country but in one of the Western countries (*Illés* [2000]; *Trócsányi–Tóth* [2002]).

When in the mid-1990's the most difficult period of transformation was over, the extent of internal migration showed a radical turning point. After a period of low intensity, growing migration levels indicated the beginning of a new era. This increasing tendency, varying with odd waves, showed that Hungarian society was recovering from its previous paralysed state. It is hard to predict the end-point of this growing tendency but likely it is going to last until the end of the delay period that stops free movement of persons (the period of derogation lasting for 0 plus 2 plus 3 plus 2 years in EU context) (*Lukács–Király* [2001]).

Level of migratory movements

The use of total migration and residential mobility rates affords a better judgement and understanding of the actual levels of spatial mobility than the use of the absolute figures. From a methodological point of view, total rates are able to remove the distortions arising from the changes of population number and age composition. The meaning of these rates is easily understood – they show the amount of migration and residential mobility that would occur in the life of one average person if migratory conditions of a particular period of time were to become fixed as permanent. This indicator shows the same changes as the changes of absolute numbers and intensity rates, in other words the changes of age composition had no significant influence on the variations in the trends (*Illés* [2002]).

We may state that from the second half of the 1980's onwards more people moved homes within their own town or village than to other settlements. Previously, only migration was taken into account in judging the level of territorial mobility. This made the Hungarian population appear highly immobile. It has been successfully shown that besides the low migration level there is a supplementary movement (almost identical in number) within the given township. Adding the two forms of mobility together, their sum no longer allows us to call the Hungarian population immobile in general within the country's boundaries.

Figure 1. Total migration rate (TMR) and the total residential mobility rate (TRMR) in Hungary, 1990–2002

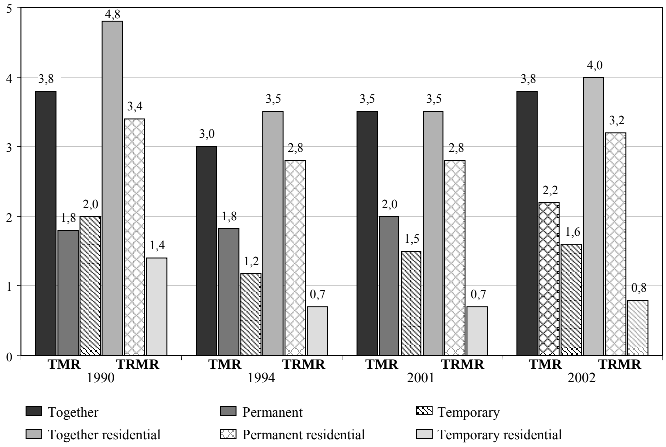
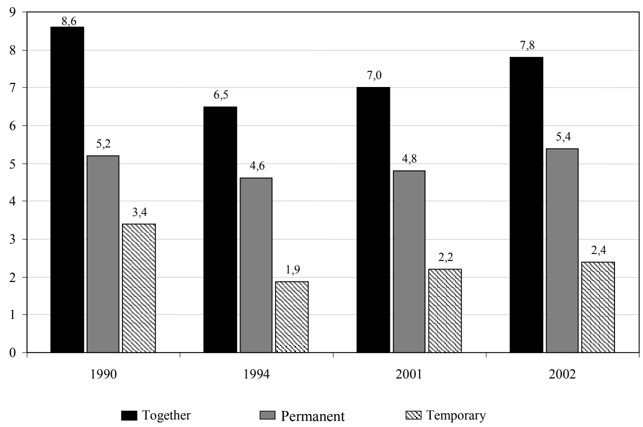


Figure 2. Total spatial mobility rate (TSMR=TMR+TRMR) in Hungary, 1990–2002



Research context

The questions of the gravity centre have an extensive international and Hungarian literature. Here we only mention the calculations regarding Hungary and their subject matters. The method was used first on the epoch of World War II for practical purposes, namely at the preparatory works of the placement of the big railway stations of Budapest, and at the planning of the distribution places of essential foodstuff (*Bene* [1961] p. 99.). In these cases the purpose was obviously to indicate the nearest points reachable to larger masses of people, thus they tried to minimize the length of the way to these points.

Later the calculations concerning the gravity centre were connected with the total population and were made already considering the whole country, or a part of the country as a regional unit (*Bene* [1961]; *Bene–Tekse* [1966]; *Nultsch* [1968]; *Mészáros* [1995]), mainly for scientific purposes. It can be stated that with the increase in the size of the examined area the calculations relating to the gravity centre have a decreasing practical and a growing theoretical significance. For calculations of the gravity centre not only the multitude of people was used as a mass. Using this method, *József Nemes Nagy* [1987] proved the translocation of the industrial gravity centres of the former Soviet satellite states to the east, so finally he could determine the dominant geographical direction of the long-term moving of the industrial production. One of the last users was *Zoltán Sümeghy* [1998a]; [1998b] who presented the shift of the gravity centres of persons of Hungarian and Slovakian nationalities on the area of recent Western Slovakia. It is already of the near past that *Tamás Ábrahám* [2000], *József Nemes Nagy* [2002] and *Zsolt Bottlik* [2002] also used this method at their research.

From the outline of the history of this topic in Hungary it can be assumed that the calculation of the gravity centre is a regional analytical method, which requires masses (absolute numbers). This mass can be a multitude of people but also any other absolute quantity. The gravity centre of the migration can be considered as one of the variants of the gravity centre of the population. Adopting ourselves to the nature of the migration flows (bipolar phenomenon) we cannot grasp the migration itself with one but only with two gravity centres: with that of out-migration and in-migration, respectively (*Compton* [1971]; *Wunsch-Thermote* [1978]; *Sárfalvi* [1991]; *Mészáros* [1994]; *Valkovics* [1998]; *Dusek* [2003]). The gravity centre of migration must characterize the regional distribution of the migrate subpopulation directly before and after the event of migration. After all the gravity centre of migration is nothing else but the mean value of the regional distribution of the out-migration and in-migration. Compared to the gravity centre of the population, in case of that of migration calculations result the dense indicators of the spatial distribution of migrate subpopulation.

Within a regional unit from the location compared to one another of the gravity centre of out-migration and in-migration, respectively, and from the distance we can make a direct conclusion regarding the similarities and differences, respectively, of the sending and receiving areas compared to one another. If we choose the whole country as a regional unit to be studied, then computing the gravity centre of out-migration and also of in-migration we can quantify a certain ‘average’ length of the migration flows. Forming time series from these values we can confirm or deny the hypothesis whether the distances of migrations shortened in the last decades and if proved so we can also quantify to what extent. Thus our purpose is to state the average distance and characteristic direction of the internal migration flows in Hungary. Using the method we get a detailed picture on the development of the direction and distance of migration flows for the years of transition, too. It is probable that while there is a significant decrease in the quantity of regional mobility and a moderate increase in the flow from urban to rural areas during the transformation period, there are also significant modifications in the development of the directions and distances of internal migration.

The method and the data

The method is simple. The gravity centres are the weighted arithmetic means – weighted with the migrants – of the co-ordinates (latitudes and longitudes) of the settlements. The gravity centres of migration are determined with the known methodological apparatus of the calculation of the gravity centres of the population (*Bene–Tekse* [1966] p. 65.). For instance, for the year (period) *k* the co-ordinates of the national (regional, county) gravity centres of the permanent out-migrants can be calculated with the following equations:

$$\overline{x} = \frac{\sum_{i=1}^{n} m_i \cdot x_i}{\sum_{i=1}^{n} m_i} \qquad \overline{y} = \frac{\sum_{i=1}^{n} m_i \cdot y_i}{\sum_{i=1}^{n} m_i}$$

Where $\overline{x}$ is the geographical line of longitude of the gravity centre of the permanent out-migrants, $\overline{y}$ is the geographical line of latitude of the gravity centre of the permanent out-migrants, m_i is the number of persons out-migrated from settlement *i*, x_i is geographical longitude of the settlement *i,* and y_i is the geographical latitude of the settlement *i*. The figures ($i = 1, 2, \ldots n$) mean the settlements of the country. Also the national (regional, county) gravity centres of the permanent in-migrants and of the temporary migration can be calculated with the same method.

Data used for the research were the data of internal migration by settlements for the years 1984–2002 and the geographical co-ordinates of the settlements, both supplied by the Hungarian Central Statistical Office.[2]

The method of calculation being well known and elaborated, its creators considered it as suitable for the analysis of all the vital events – migration included. Migration was yet not analysed in Hungary with this method this far. *Lajos Bene* (1961) calculated the gravity centres of net migration for the 1921–1941 period, but this cannot yet be considered as a pure analysis of the migrations because he did not distinguish the gravity centres of out-migrations and in-migrations. Therefore the calculation of the gravity centres of the net migration does not show the whole migration movement but 'only' the centres

[2] During the period examined there were significant changes in the settlement stock (in the beginning of 1984 3066 settlements, at the end of 2002 3145 settlements). There were unifications and disaggregations (*Szigeti* [1997], [1998]) and their effect cannot be let out of consideration. The geometric gravity centres of the settlements of Hungary were in the same place in the 1984–1988 and in the 1989–1993 periods. In the 1994–1997 period it moved by 234 metres to the east and by 175 metres to the south. Thus we can see that the impact resulting from the modifications in the borders of the settlements was practically insignificant. The method itself as well as the theoretical consideration did not request homogenization of the settlement series for the beginning or the end of the period investigated. Because of the above arguments for each year we worked with the actual topical settlement series of the absolute number of migrants and later we sum them up for the appropriate periods.

of the changes in the population number resulting from the migration of the area examined.[3]

Following these long introductory chapters we can put the question: for what purpose can we use the gravity centres of migration? Responses are as follows:

1. They are very compact mean values, which characterize the regional distribution of the migrate subpopulation.

2. The direction determined by the two gravity centres can be considered as the most characteristic among the many directions of migration.

3. The distance between the two gravity centres as the crow flies (air kilometres, meters) also means a certain average distance of the migration streams in the time interval studied.

National gravity centres of migration

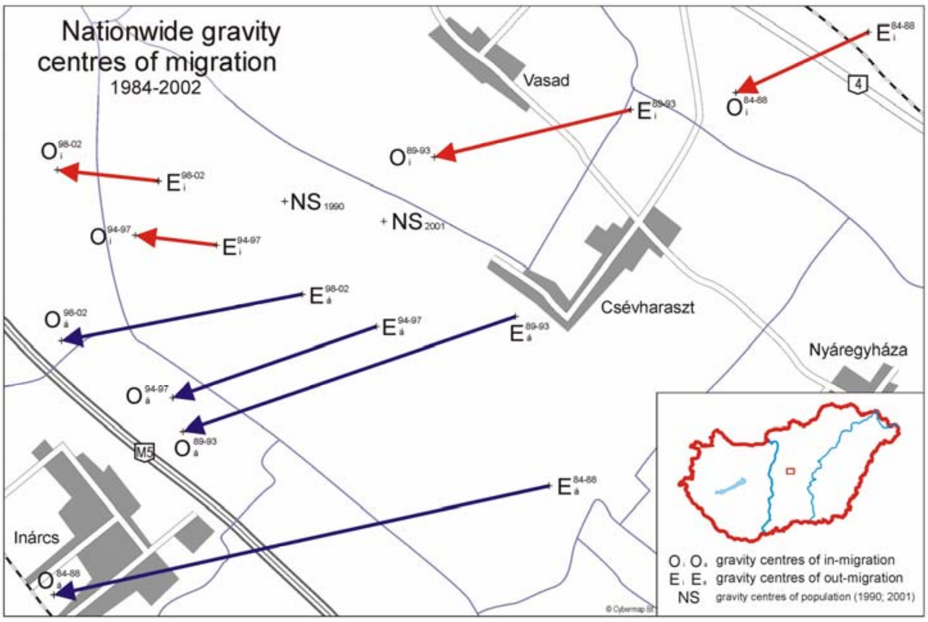
On Figure 3 the gravity centres of permanent and temporary migration, respectively, are indicated for Hungary as a whole, for four periods of time between 1984 and 2002. Moreover, the gravity centres of the total population at the dates of the 1990 and 2001 population censuses are shown. The gravity centres of the permanent and of the temporary migration are separated from one another. It means that flows and counterflows are not equal, so there are main streams of internal migration in Hungary. The gravity centres are very near to the 1990 and 2001 population gravity centre which can be found on the area of the settlement Csévharaszt in Pest county. Yet they are not exactly in the same place, and from these differences the conclusion can be made that there are territorial selectivities: on the one hand in relation to the sending and receiving settlements (type 1), on the other hand the spatial distribution of the migrating subpopulation is not simply a representative sample of the regional distribution of the total population (type 2). The above-mentioned two kinds of selectivity, however, are very small. Their extent decreased uniformly until 1997 because the gravity centres of migration were getting closer and closer to one another and at the same time to the gravity centres of total Hungarian population. In all of the investigated periods the selectivity of first type was bigger for the permanent migration than the temporary one. It means that the separation of the sending and receiving settlements were greater for permanent migrants than the temporary migrants. In the first ten years of the period examined it seemed that a greater regional selectivity of second type could be stated for the temporary migrations compared to the permanent ones. After 1994 this characteristic feature changed and permanent migrations showed a greater regional selectivity of second type because these gravity centres were farther from the gravity centre of the population.

The gravity centres of the permanent out-migrants are on the administrative area of the settlement Csévharaszt, to south-southeast from the gravity centre of the total population. The gravity centres of in-migrations can be found near to the village Inárcs. It can be stated that the permanent out-migrants are more to the east from the gravity centre of the population, the in-migrants however are more to the west, so the dominant flow directed from east to west in the internal migration. The direction observed in the earlier

[3] The study of the change in the population number with the method of the gravity centres would be complete if we calculated also the gravity centres of the natural and total increase and decrease, respectively, and if we also compared them.

period still exists, in fact this is the most characteristic. From Table 1 it can be affirmed uniformly that the gravity centres of permanent migrations get nearer and nearer to the gravity centre of the population until 1997 which shows the decrease in the extent of the territorial selectivity of the second type of the permanent migrations.

Figure 3. Nationwide gravity centres of migration, 1984–2002



Note: O_i, E_i – temporary migration; $O_á$, $E_á$ – permanent migration.

Table 1

Distance and direction between the national gravity centre of population 1990 and the national gravity centres of migration
(meter)

Year	Out-migration		In-migration	
	gravity centres			
	Permanent			
1984–1988	5700	SE	6700	S–SW
1989–1993	4500	E–SE	3600	S–SW
1994–1997	2200	SE	3300	S–SW
1998–2002	2300	S	3900	S–W
	Temporary			
1984–1988	9000	E–NE	6900	E–NE
1989–1993	5400	E–NE	2300	E–NE
1994–1997	1200	SW	2300	W–SW
1998–2002	1900	W–NW	3400	W–NW

Between 1984 and 1993 the gravity centres of the temporary out-migrants and in-migrants can be found mostly farther to the east, on the administrative area of Csévharaszt, Vasad and Monor. After 1994 their location changed and they were farther to the west from the gravity centre of the population. It is not difficult to realise the tendency that also the characteristic sending and receiving areas of the temporary migration are shifted farther and farther to the west. Similarly to the permanent migration the distances of the gravity centres of the temporary migration measured to the gravity centre of the total population show a monotonous decrease until 1997 which also refers to the decline in the extent of the spatial selectivity of second type.

The gravity centres of the permanent and temporary migration moved characteristically in a different direction compared to one another (see Figure 3). The gravity centres of the permanent out- and in-migrations 'advanced' to the north and then to the west, while the temporary gravity centres of the same type moved characteristically to south-west. Despite the moving to the north and south, moving to the west can be considered as a common feature, especially in the second half of the period studied.

The gravity centres of the out-migration and in-migration regarding the same period moved in similar directions and to similar extents, so it can be foreseen that the distances between them could not change radically either.

Distances

With indirect estimation it was proved in the literature already that parallel with the decrease in volume also the average distance of the spatial movements shortens. The process of the regional levelling was one of the most important cause of the decline in distance before the transition period *(Erdősi* [1985]; *Kovács* [1985]; *Rédei* [1991]; *Daróczi* [1998]; *Nemes Nagy* [1998]; *Horváth-Rechnitzer* [2000]; *Dobosi* [2003]). Calculation of the gravity centres of migration of a various degree of spatial aggregation gives direct evidence to prove the decrease in the distance of migration and the extent of the decrease. It seems that – given that regularities are valid with the transition period still holding on – the regional differences becoming more and more marked because of the different pace of development will increase the average distances of movement (*Cséfalvay* [1993]; *Enyedi* [1994], [1996]; *Rees-Kupiszewski* [1999]; *Nagy* [2002]; *Szászi-Hajnal-Reszler* [2003]). Let us examine whether this hypothesis can be verified.

The gravity centre of out-migration can be considered as a mean value of the regional distribution of the out-migrating subpopulation and the gravity centre of in-migration as a mean value of the regional distribution of the in-migrants. If these two points do not coincide or are not very near to one another (i.e. more than 100 metres apart) then the distance between the two points also means a certain average physical distance of the migrations in the regional unit studied (country, region, county) not in absolute but in a relative sense. If these two points are near to one another then we have to conclude that the sending and receiving areas are neighbours of one another and presumably there are mutual flows and couterflows to a similar extent. With the increase in the average distance the migratory relations of the neighbours weaken, the relations of the farther areas strengthen and flows by pairs begin to occur in one direction, therefore separation of the sending and receiving areas increases.

The distances of the nationwide gravity centres of the permanent migrants were 7500 m, 5200 m, 3200 m and 3600 m in 1984–1988, 1989–1993, 1994–1997 and 1988–2002, resp., which are not really great distances. Even so, however, the average length of the way made by the permanent migrants is the multiple, two and three times of that of the temporary migrants (2100 m, 3000 m, 1200 m and 1500 m). Consequently, we can say that the permanent migrants make a longer way in the physical space. The shorter distance characteristic of the temporary migrations can be explained by the fact that the number of the temporary migrants, whose purpose is often to take a job and who are registered temporarily, decreased radically as well. The distance of the temporary migrations for study and housing purposes was presumably shorter earlier, too (Table 2), with the expansion of third level education strengthening this process.

Table 2

Distance between the nationwide gravity centres of out-migration and in-migration
(meter)

Type	Permanent migration					Temporary migration				
		1984–88	1989–93	1994–97	1998–02		1984–88	1989–93	1994–97	1998–02
All migration	$E_{á}^{O} - O_{á}^{O}$	7 500	5 200	3 200	3 600	$E_{i}^{O} - O_{i}^{O}$	2 100	3 000	1 200	1 500
Interregional migration	$E_{á}^{ORK} - O_{á}^{ORK}$	23 200	18 500	11 300	13 900	$E_{i}^{ORK} - O_{i}^{ORK}$	3 500	5 800	3 200	3 700
Intercounty migration within regions	$\overline{E}_{á}^{RBMK} - \overline{O}_{á}^{RBMK}$	4 800	3 000	3 000	1 600	$\overline{E}_{i}^{RBMK} - \overline{O}_{i}^{RBMK}$	1 500	2 600	1 400	–

For the country as a whole we distinguished three types of moves and we also stated the geographical distances between their gravity centres (Table 2). In the first group all the migrations were taken into consideration. On basis of Table 2 we can see that the distances of the migrations crossing the borders of the regions (second type of move) are the longest. Here even multiple lengths can be measured compared to the average distance of the movements within the regions. The migrants leaving their county but remaining in their own regions covered the shortest way.

Our opening hypothesis regarding the permanent migrations until 1997 cannot be confirmed for the country as a whole. The distances of the migrations do not grow in parallel with the increasing regional disparities, we measure characteristic diminution instead. The permanent migrants covered shorter and shorter distances until 1997. In case of the temporary migration between 1988 and 1993 increase in the distance can be confirmed for the country as a whole, but already as far as between 1994 and 1997 the distances are shorter than between 1984 and 1988. The opening hypothesis can be confirmed for both the permanent and temporary migration after 1998, i.e. the distances of migration grow up again. After 1998 slight increases were measured for all the migratory types.

Directions

Following the distance, the other spatial parameter of order, the study of the direction is very important when migration to be examined. The gravity centres of the in-migrants in Hungary are to the west from the gravity centres of out-migrations, thus in the period studied the prevailing direction of the permanent and temporary migrations proceed to the west. We may add that beside the characteristic western direction also a slight southern component can be verified. In case of the permanent migrations we could observe a very stable direction to the west and south-west which is 'worthy of its character' and did not change in the years of the transition (Figure 3 and Table 3). In the period 1994-1997 the most characteristic direction of the temporary migrations proceeded clearly to the west.

Table 3

Direction of migration

Type	1984–1988	1989–1993	1994–1997	1998–2002
	Permanent migration			
All migration	W-SW ↙	W-SW ↙	W-SW ↙	W-SW ↙
Interregional migration	W-SW ↙	W-SW ↙	W-SW ↙	W ←
	Temporary migration			
All migration	W-SW ↙	W-SW ↙	W ←	W ←
Interregional migration	W-SW ↙	W ←	W ←	W ←

Apart from the migrations within the region, taking into consideration only those who are migrants between the regions the flow to the west is still prevailing. Between 1994 and 1997 for the temporary migrations crossing the borders of the regions the western vector became stronger. It may be that this modification shows the beginning of the change in tendency but we find it more real supposition that in this case we can observe a provisional fluctuation and the direction to the west and south west will resume.

It can also be assumed that for the country as a whole there is no significant difference in the basic directions of the permanent and temporary migration (though we cannot be sure whether the situation was the same in time before the periods covered by our study).

Dispersion

In general sense there may be various objections (shortcomings) to the use of gravity centres. The first one may be that the gravity centres themselves have so compact mean values that they do not show anything regarding the regional distribution of the phenomenon examined. Those who have such an opinion are right considering that really it is not possible to do much with a gravity centre alone. We can describe its location, maybe indicate it on a map but we can hardly get farther. For solution of this problem obviously we have to seek points of comparison. Naturally it is possible to compare to the significant cities. Beside this in our case we can compare some gravity centres of migration to favoured points (e.g. Pusztavacs, the geographical centre of Hungary). Other possibilities for comparison are the gravity centre of the population or the geometric centre of the settlements of Hungary. Beside the description we find two ways for a sophisticated analysis. The first one is the dynamism in time. We may calculate the pair of gravity centre referring to one date, to one period and then also for another date and period, i.e. we widen our study in time. Doing so it is possible to make conclusions concerning the modifications in the characteristic directions, the changes in the average distances of migration. The second way of analysis – which in our case also means a widening of the regional details – is to calculate the regional gravity centres, the gravity centres of counties and even of the smaller regional units.[4] Presentation of the two ways of analysis on a map gives a synthesis of the partial result.

The second shortcoming to the gravity centres is that gravity centres are after all weighted arithmetical means and, similarly to the statistical means, they themselves do not illustrate the situation of the phenomenon studied around the gravity centre, in our case the dispersion of the migrating sub-population around the gravity centres. Various methods are known for the characterization of the situation around the gravity centre (*Tekse* [1966]). Of them Bachi's *d* standard distance was calculated (*Bene–Tekse* [1966]) at the study of the gravity centres of the population, and for possibilities of comparison we used this method to state the dispersion coefficient (*Bachi* [1962]).

$$d = \left\{ \frac{\sum_{i=1}^{n} m_i \left[(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2 \right]}{\sum_{i=1}^{n} m_i} \right\}^{\frac{1}{2}}$$

Where d is the standard distance, $\bar{x}$ is the geographical line of longitude of the gravity centre of migration, $\bar{y}$ is the geographical line of latitude of the gravity centre of migration, m_i is the number of persons migrated from the settlement i, x_i is the geographical longitude of settlement i, y_i is the geographical latitude of settlement i. The figures $(i = 1, 2, \ldots n)$ mean the settlements of the country. The d distance is simply the square root of the average – weighted with the population numbers – of the quadratic distances of the gravity centre of migrations and of the individual settlements.

In Hungary the dispersion around the gravity centre of the population indicated with the d distance was 132.8 km in 1900, 129.1 km in 1930, 128.3 km in 1949 and 126.6 km in 1960. *Lajos Bene* and *Kálmán Tekse* [1966] explained the decreasing tendency of the more rapid growth in the population of the urban areas. Unfortunately, at the calculation of the gravity centre of population made on basis of the data of the 1990 population census, dispersion was not calculated. As a consequence of the urbanization process, the population of the urban areas grew after 1960 on, so presumably also the tendency of de-

[4] It should be mentioned that choosing the appropriate regional unit is absolutely necessary to determine our research purposes.

crease in the *d* distance continued. If we accept this supposition then the dispersion must have most probably declined. According to our estimation the dispersion of the gravity centre of the population of 1990 might have been somewhere between 120 and 125 km. On the basis of 2001 population census the dispersion of gravity centre of total population was 123.5 km.

Due to the second type of regional selectivity of the migrants compared to the regional distribution of the total population, the dispersion of the gravity centres of in-migration and out-migration of the country must be higher than the dispersion around the gravity centre of the population (except for the temporary in-migration of 1998–2002 period).

Earlier at the comparison of the distance between the gravity centres of migrations and the gravity centres of population we supposed that on a national scale only since 1994 is the regional selectivity of the permanent migration higher than that of the temporary migration. In case of higher selectivity it can be expected that also the dispersion around the gravity centres of permanent migrations is greater compared to that of the temporary migrations.[5] Though with time the difference decreases more and more. The higher dispersion around the gravity centres of the out-migration refers to the fact that the sending settlements are located more unevenly in Hungary. Opposite to this the location of the receiving settlements and the areas represented by them is more concentrated.

Table 4

Dispersion of nationwide gravity centres of migration, 1984–2002
(meter)

Type	Permanent				Temporary			
	1984–1988	1989–1993	1994–1997	1998–2002	1984–1988	1989–1993	1994–1997	1998–2002
	In-migration gravity centre							
All migration	126 941	126 746	125 575	123 059	123 467	123 250	123 973	124 067
Interregional migration	109 317	112 016	112 727	109 992	117 973	113 631	112 700	113 095
Intercounty migration within individual regions*	39 862	40 136	40 323	50 433	42 632	40 513	41 021	49 761
	Out-migration gravity centre							
All migration	132 900	129 958	127 140	124 974	126 315	126 362	124 810	124 922
Interregional migration	127 921	122 674	118 209	117 110	122 792	120 111	114 921	115 136
Intercounty migration within individual regions*	41 903	41 451	40 089	49 796	43 856	43 163	40 965	49 716

* Rounded arithmetic means of the distances of intercounty migration of the individual regions.

[5] With the calculation of dispersion of the gravity centres of migration we could prove that the dispersion of the permanent migration was higher than the temporary one until 1997. That means the first type of territorial selectivity of permanent migration was higher in degree than that of the temporary one. During the whole investigated period the dispersion around the gravity centres of out-migration were higher that of the in-migration one. For that reason the result of migration flows strenghtened the concentracion of Hungarian population. In Hungary as a whole the process of suburbanization (*Kovács* [2002]; *Izsák* [2003]) was not strong enough to counterbalance the process of concentration led by internal migration.

Studying the gravity centres of the migration between the regions we can still state that the dispersion around the gravity centres of out-migration is higher than the dispersion around the gravity centres of in-migrations and the difference decreases in time. It must be mentioned that the differences by pairs of the regional gravity centres of out-migration and in-migrations are much higher than to the total migration.

*

The gravity centres of out-migration and in-migration were separated from one another and were very near to the 1990 and the 2001 gravity centre of the total population. But they were not exactly in the same place. Conclusion can be made that there are territorial selectivities: on the one hand between the sending and receiving settlements (type 1), on the other hand the spatial distribution of the migrate subpopulation is not simply a representative sample of the spatial distribution of total population (type 2). These two kinds of selectivity, however, are very small. Their extent decreased uniformly in the course of time because the gravity centres of migration got closer and closer to one another and to the gravity centres of total Hungarian population until 1997. After 1998 a small growing up was investigated. Due to the second type of regional selectivity the dispersion of the gravity centres of in-migrations and out-migrations of the country are mostly higher than the dispersion around the gravity centre of the population.

The gravity centres of the migration are separated from one another. It means that flows and counterflows are not equal, so there are main streams of internal migration in Hungary. In the country as a whole the gravity centres of in-migration located to the west from the gravity centres of out-migration, thus in the period studied the dominant direction of the migrations proceed to west.

The distances between the national gravity centres of migration gradually shortened until 1997 (except for the temporary one of the 1989–1993 period and the temporary and permanent of 1998–2002). After that slight increase was measured. The length of the way made by the permanent migrants was greater compared to the temporary migrants in all of the investigated period. On this basis we can say that the permanent migrants make a longer way in the physical space.

Our opening hypothesis (i.e. that increasingly sharp territorial differences would increase the distances involved in migration) cannot be proved on a national level until 1997. From 1998 onwards, however, people involved in migration started to cover ever larger distances in physical space which shows that the hypothesis came to apply. (To explain why the appearance of larger distances delayed almost an entire decade presents a question for further research.)

The gravity centre of the population for 2001 is further east than was the gravity centre of 1990. This fact seemingly contradicts one of the main results covered in the paper, whereby internal migrants move dominantly in a western direction. This paradox may be explained sufficiently by involving two factors. One is the natural increase characteristic of the eastern parts of Hungary because of a higher fertility rate prevailing there. The second factor is connected to the territorial distribution of international migrants. The majority of immigrants and new citizens settled east of the Danube and in Budapest. Thus we may state as conclusion that the sum of the effects arising from fertility and the selective nature of international migration surpassed that of the effect of internal migration on the spatial distribution of the total population and pushed the population gravity centre toward the east.

REFERENCES

ÁBRAHÁM T. [2000]: Az USA fekete lakosságának migrációja a XX. században. *Demográfia*. Vol. 43. No. 1. 161–175. p.

BACHI, R. [1962]: Standard distance measures and related methods for spatial analysis. Regional Science Association. *Zurich Congress Papers*. Vol. 10. 83–132. p.

BOTTLIK, ZS. [2002]: A szlovákok etnikai súlypontjának változásai a Dunántúli-középhegység területén a XVIII. századtól napjainkig. *Területi Statisztika*. Vol. 42. No. 6. 551–561. p.

BENE, L. [1961]: Magyarország népességi súlypontja. *Demográfia*. Vol. 4. No. 1. 91–102. p.

BENE, L.–TEKSE, K. [1966]: *Vizsgálatok a népesség területi eloszlásának alakulásáról Magyarországon 1900–1960*. KSH Népességtudományi Kutató Csoport Közleményei. 9. köt. Budapest. 65. p.

COMPTON, P. A. [1971]: *Some aspects of the internal migration of population in Hungary since 1957*. (Publications of the Demographic Research Institute.) Budapest. No. 33.

CSÉFALVAY, Z. [1993]: Felharmadolt ország. *Valóság*. No. 7. 1–17. p.

DARÓCZI, E. [1998]: Residential moves within Hungary, 1985–1996. *Espace, Populations, Societes*. Vol. 27. No. 3. 381–388. p.

DOBOSI, E. [2003]: A komplex regionális fejlettség matematikai-statisztikai elemzése. *Területi Statisztika*. Vol. 43. No. 1. 15–33. p.

DUSEK, T. [2003]: A gravitációs modell és a gravitációs törvény összehasonlítása. *Tér és Társadalom*. Vol. 17. No. 1. 41–58. p.

ENYEDI, GY. [1994]: Területfejlesztés, regionális átalakulás. *Társadalmi Szemle*. Vol.49. No. 8–9. 133–139. p.

ENYEDI, GY. [1996]: *Regionális folyamatok Magyarországon az átmenet időszakában.* Hilscher Rezső Szociálpolitikai Egyesület. Budapest.

ERDŐSI, F. [1985]: Az ingázás területi-vonzáskörzeti szerkezete Magyarországon. *Demográfia*. Vol. 28. No. 4. 489–498. p.

HORVÁTH, GY.–RECHNITZER J. (ed.) [2000]: *Magyarország területi szerkezete és folyamatai az ezredfordulón*. MTA Regionális Kutatások központja. Pécs.

ILLÉS, S. [2000]: Changing levels of spatial mobility in Hungary. In: *Kovács, Z.* (ed.): *Hungary towards the 21st century.Geographical Research Institute Hungarian Academy of Sciences*. Budapest. 137–149. p.

ILLÉS, S. [2002]: *Költözünk vagy vándorlunk? In Ezredforduló – magyar valóság – cenzusok.* Magyar Statisztikai Társaság. Budapest. 37–51. p.

IZSÁK, É. [2003]: *A városfejlődés természeti és társadalmi tényezői*. Budapest és környéke. Napvilág Kiadó. Budapest.

KOVÁCS, T. [1985]: A népesség területi mozgásának főbb jellemzői az elmúlt 30 évben és valószínű jövőbeni vonásai. In: *Káposztás, F.* (ed.) *A népesség területi elhelyezkedése és mozgása*. KSH Népességtudományi Kutató Intézet Kutatási jelentései 25. kötet. Budapest. 19–30. p.

KOVÁCS, T. [2002]: A területi fejlettségbeli különbségek alakulása Magyarországon. *Területi Statisztika*. Vol. 42. No. 6. 506–517. p.

KOVÁCS, Z. [2002]: Az urbanizáció jellemzői Kelet-Közép-Európában a posztszocialista átmenet idején. *Földrajzi Közlemények*. Vol. 126. No. 1–4. 57–78. p.

LANGERNÉ RÉDEI, M. [1986]: A kistérségi népességmozgások. *Statisztikai Szemle*. Vol. 64. No. 11. 1093–1107. p.

LUKÁCS, É.–KIRÁLY M. (ed.) [2001]: *Migráció és Európai Unió*. Szociális és Családügyi Minisztérium. Budapest.

MÉSZÁROS, Á. [1995]: A népesség területi átrendeződése és települési koncentrációja. *Statisztikai Szemle*. Vol. 73. No. 7. 533–542. p.

MÉSZÁROS, R. [1994]: *A település térbelisége*. JATEPress. Szeged.

NAGY, G. [2002]: Oldódtak-e az öröklött területi különbségek a rendszerváltás éveiben? In: *Abonyiné et al*. (eds.) *A magyar társadalomföldrajzi kutatás gondolatvilága.* Szegedi Tudományegyetem Gazdaság és Társadalomföldrajzi Tanszéke. Szeged. 211–225. p.

NEMES NAGY, J. [1987]: *A regionális gazdasági fejlődés összehasonlító vizsgálata*. Akadémiai Kiadó. Budapest.

NEMES NAGY, J. [1998]: *A tér a társadalomkutatásban. Bevezetés a regionális tudományba*. Hilscher Rezső Szociálpolitikai Egyesület. Budapest.

NEMES NAGY, J. [2002]: Spatial gravity centres of the dynamics and the crisis in Hungary. *Hungarian Statistical Review*. Special number 7. 75–85. p.

NULTSCH, H.-G. [1968]: A népességi súlypontok. *Demográfia*. Vol. 11. No. 2. 260–264. p.

RÉDEI, M. [1991]: Close migration directions under changing propensity. In: *Józwiak, J.–Kotowska, I.* (eds.) *Usefulness of demographic modelling*. Szkola Glówna Handlova-Instytut Statystyki i Demografii. Warszawa. 159–177. p.

REES, P.–KUPISZEWSKI, M. [1999]: *Internal migration and regional population dynamics in Europe: synthesis*. Council of Europe Publishing. (Population studies, No. 32.) Strasbourg.

SÁRFALVI, B. [1965]: *A mezőgazdasági népesség csökkenése Magyarországon*. Akadémiai Kiadó. Budapest.

SÁRFALVI, B. [1991]: *Magyarország népességföldrajza*. Tankönyvkiadó. Budapest.

SZÁSZI, F.–HAJNAL B.–RESZLER G. [2003]: *Tanulmányok Szabolcs-Szatmár-Bereg megye népességének belső (belföldi) vándorlásáról (1869–1995).* MTA Szabolcs-Szatmár-Bereg Megyei Tudományos Testületének Közleményei. Stúdium Kiadó. Nyíregyháza.

SZIGETI, E. [1997]: Urbanizáció, városhálózat, várossá nyilvánítás. *Területi Statisztika.* Vol. 37. No. 1. 66–79. p.

SZIGETI, E. [1998]: Új községek – új önkormányzatok. *Területi Statisztika*. Vol. 38. No. 1. 20–33. p.

TEKSE, K. [1966]: A népesség koncentrációjának jellemzéséről. *Demográfia*. Vol. 9. No. 4. 564–576. p.

TRÓCSÁNYI, A.–TÓTH J. [2002]: *A magyarság kulturális földrajza II*. Pannónia Tankönyvek. Pro Pannónia Kiadó Alapítvány. Pécs.

VALKOVICS, E. [1998]: Kísérlet a belföldi vándorlások x éves korban várható átlagos számának becslésére. In: *Illés, S.–Tóth, P. P*. (eds.) *Migráció*. (Tanulmánygyűjtemény) I. köt KSH Népességtudományi Kutató Intézet. Budapest. 189–216. p.

WUNSCH, G.–TERMOTE, M. [1978]: *Introduction to demographic analysis*. *Principles and methods*. Plenum Press. New York and London.

VARIANCE ESTIMATION WITH THE JACKKNIFE METHOD IN THE CASE OF CALIBRATED TOTALS

LÁSZLÓ MIHÁLYFFY[1]

Estimating the variance or standard error of survey data with the jackknife method runs in considerable difficulties if the sample weights are calibrated. The current methodology used in the household surveys of the Hungarian Central Statistical Office is reviewed, some possible approaches are compared, and a new strategy of using the jackknife method is recommended in the paper.

KEYWORDS: Jackknife method; Raking; Generalised regression estimator.

Sample surveys – especially household surveys – conducted by the statistical agencies of different countries have in common among other things the following two features:

– the final sample weights are the result of some calibration procedure,
– the variance or standard error of survey data are estimated by some method based on the secondary processing of sample data such as the jackknife and the bootstrap method, the method of balanced half-sample replicates, etc. (*Wolter* [1985]).

The Hungarian Central Statistical Office (HCSO) has a considerable tradition in conducting household surveys. The beginning of the household budget survey (HBS) dates back e.g. to the late forties of the 20th century. The use of an up-to-date two-way calibration procedure as well as that of the VPLX Software (*Fay* [1998]) for variance estimation were introduced in the HBS and in the Labour Force Survey (LFS) – which started in 1992 – in the mid-nineties. This has led to the practice that the final sample weights were first determined by calibration, and the standard errors for some data were then estimated by the VPLX Software. Apart from some modification discussed in what follows, this practice has not changed substantially during the last eight years. It should be noted that the jackknife variance procedure currently used at the HCSO does not comply with the jackknife principle. That is, whenever a new jackknife replicate or pseudo value is created, it should be of the same functional form as the original estimate. One could assume,

[1] Statistical advisor of the HCSO. The views in this paper are those of the author and are not necessarily shared by the HCSO. When working on the paper, the author also benefited from the comments of *Mike Hidiroglou* of the ONS, Newport, on an earlier version.

under favourable conditions, that the bias resulting from this procedure is not significant, but, unfortunately, this will not be the case in most practical situations. In a 1996 paper, *W. Yung* and *J. N. K. Rao* have pointed out the following:

– if the variance of calibrated estimates is to be estimated with the jackknife method, the calibration procedure should be repeated whenever a new jackknife replicate is created (the case of *correct weighting*);
– if the previous rule, i.e. the jackknife principle is ignored, the variance estimates can be seriously biased in the case of calibrated estimates (the case of *incorrect weighting*);
– the linearised version of the jackknife formula is practically equivalent to the jackknife with correct weighting, yet the computing time is reasonably small as compared to the latter.

In this paper an improved version of the current HCSO technique to estimate the variance of calibrated LFS estimates will be introduced and compared to the jackknife method with correct weighting. We shall see that, though it is a jackknife application with incorrect weighting, its results are good approximations of those obtained with a version of jackknife with correct weighting. The improved HCSO technique can be regarded as a possible alternative to the linearised jackknife method, though the two techniques are not compared in the paper. Our approach will be more empirical than theoretical, and focuses on the main table of the Hungarian LFS which is – in a somewhat reduced form – as follows:

Table 1

Hungarian labour force survey, September 2003

Age-groups	Employed	Unemployed	In labour force	Not in labour force	Working age population	Participation rate, percent	Unemployment rate, percent
Total	*3 977 107*	*222 662*	*4 199 769*	*3 541 462*	*7 741 231*	*54.25*	*5.30*
15–19	25 045	10 667	35 712	588 930	624 642	5.72	29.87
20–24	328 174	43 196	371 370	326 455	697 825	53.22	11.63
25–29	594 679	39 964	634 643	207 619	842 262	75.35	6.30
30–39	1 062 059	51 759	1 113 818	242 050	1 355 868	82.15	4.65
40–54	1 594 690	67 067	1 661 757	493 766	2 155 523	77.09	4.04
55–59	282 837	8 213	291 050	327 773	618 823	47.03	2.82
60–69	81 614	1 796	83 410	937 141	1 020 551	8.17	2.15
70–74	8 009	0	8 009	417 728	425 737	1.88	0.00
Male	2 159 669	125 964	2 285 633	1 403 286	3 688 919	61.96	5.51
Female	1 817 438	96 698	1 914 136	2 138 176	4 052 312	47.24	5.05

Jackknife estimates of standard errors for the data in this table have been computed, using different versions of incorrect and correct weighting. This paper describes and compares these various jackknife procedures. The basis of the comparison is the deviation from the result obtained by correct weighting on the one hand, and the run time used for the computation on the other. The conclusions drawn from our numerical results depend, among other things, on the design of the Hungarian LFS as well as on the method of calibration in use; hence the reader should be careful when applying them in a different environment.

The structure of the paper is as follows. Section 2 contains a concise description of the sample design of the Hungarian LFS as well as that of the technique of calibration in use. The different applications of the jackknife method – called henceforth strategies – are described and the corresponding numerical results are presented in Section 3. Thereafter a brief paragraph summarises the conclusions of the paper. The Appendix gives an insight into the technique of calibration used in the Hungarian LFS.

SAMPLE DESIGN AND CALIBRATION IN THE HUNGARIAN LABOUR FORCE SURVEY

The *sample design* of the Hungarian LFS (a quarterly survey interviewing individuals in approximately 38 000 non-institutional households) is stratified by locality size, administrative categories, and type of residence. A systematic sample of dwellings is then selected within these strata. In the *old* sample (i.e. up to the fourth quarter 2002) each locality with at least 10 000 inhabitants was *self-representative*; in the *new* sample (i.e. from the first quarter 2003) the corresponding number is 5 000. For both the new and old samples, a stratified *non-self-representative* sample of localities was selected from the rest of the country. For the self-representing localities, primary sampling units (PSUs) were census enumeration districts (EDs) in the case of the old sample and dwellings for the new one. In contrast, localities were the PSUs in the strata of non-self-representing localities for both old and new samples. Given that a locality had been *selected*, EDs were the secondary sampling units for the old sample and dwellings in the new sample. The ultimate sampling units were dwellings in all cases. It follows that the new sample has one less stage of selection than the old sample. Whenever localities or EDs were the sampling units, the method of selection was probability proportional to size (PPS). As noted above, dwellings were selected using systematic random sampling. Prior to this, the dwellings were sorted within localities by type of residence, giving thereby rise to implicit stratification. The old sample had 130 design strata and 753 localities, and the new one has 275 strata and 662 localities. The quarterly sample of the LFS is split in three statistically equivalent *monthly* subsamples, each having 1/3 of the size of the quarterly sample

Estimation

Using the VPLX Software to estimate the standard error for the data in Table 1, it is straightforward to choose the 'stratified jackknife' option. To this end the user has to supply the codes 'stratum' and 'cluster' on each observation of the input data file of the VPLX program. The 'stratum' is obviously the code of the design stratum, while both in the old and the new sample; the 'cluster' was identified as the five-digit standard code of the locality in the case of *non-self-representing* localities. In the *self-representing* localities of the *old* sample, the 'cluster' was identified as the code of the enumeration district (ED). In the case of self-representing localities of the *new* sample, there are no predetermined clusters, thus it is the user's task to define them for the correct application of the stratified jackknife option of the VPLX program. In our experience, this can be done by distributing the sampled dwellings in those localities with some random method in groups containing three or four dwellings. Those triples and quadruples of dwellings are then regarded as clusters and a unique identifier code is assigned to each of them. If the input data file is prepared in this way, the VPLX program creates as many jackknife replicates as the number of different cluster codes observed on the file. On each occasion, all observations belonging to one of the clusters are removed from the sample, and the sample weights of the remaining observations in the corresponding stratum are properly adjusted.

We next introduce the notation to define generalised regression (GREG) which is a special case of calibration (*Deville–Särndal* [1992]). Let

– s be a probability sample consisting of the PSU units 1, 2, …, n,
– w_j the design weight associated with sampled unit j, j = 1, 2, …, n,
– w_j^c the corresponding calibrated weight,
– y_j the value of the study variable observed on unit j,
– $\mathbf{x}_j$ an m-vector of auxiliary variables measured on unit j of the sample,
– $\mathbf{X}$ the m-vector consisting of the known population totals of the auxiliary variables,
– $\hat{\mathbf{X}} = \Sigma_j w_j \mathbf{x}_j$ the sample estimate of $\mathbf{X}$,
– $\mathbf{A} = \Sigma_j w_j \mathbf{x}_j \mathbf{x}_j$’ a nonsingular matrix of order m (the prime denotes transpose).

Using this notation, the resulting system of calibrated weights

$$w_j^c = w_j(1 + \mathrm{x}_j’\mathbf{A}^{-1}(\mathbf{X} - \hat{\mathbf{X}})) \qquad /1/$$

is the unique solution of a constrained minimisation problem. It can be stated as:

$$\text{minimise the distance function } \sum_{j=1}^{n}(w_j^c - w_j)^2 / w_j \qquad /2/$$

$$\text{subject to the system of equations } \sum_{j=1}^{n} w_j^c \mathbf{x}_j = \mathbf{X}\,. \qquad /3/$$

The estimator of the unknown population total Y that uses these calibrated weights given by

$$\hat{Y}^{\mathrm{g}} = \sum_{j=1}^{n} w_j^c y_j$$

is known as the generalised regression estimator.

In spite of the numerous advantages of the GREG such as the explicit expression /1/ for the calibrated weights, calibration in the Hungarian household surveys is mainly carried out using raking. That is it uses the generalised iterative scaling (*Darroch–Ratcliff*

[1972]). The reason for this is twofold. On the one hand, it is very easy to write a code based on the raking algorithm; this is illustrated in the Appendix by a segment of the program used in the jackknife estimations reported in the paper. On the other hand, the experience gained with this method since 1994 has proved satisfactory in every respect, including among other things the speed of the computation, too. The calibration procedure can be interpreted as solving a constrained minimisation problem also in the case of raking (*Darroch–Ratcliff* [1972]; *Deville–Särndal* [1992]): the distance function /2/ is replaced in this case by

$$\sum_{j=1}^{n}(w_j^c \log(w_j^c / w_j) - w_j^c + w_j) \qquad /4/$$

called the information divergence between w_j^c and w_j, i.e. the calibrated and the original design weights. Similarly to the quadratic distance function /2/, the information divergence is also nonnegative, strictly convex and vanishes if and only if $w_j^c \equiv w_j$.

The monthly as well as the quarterly sample of the Hungarian LFS can be regarded as the union of the subsample of the capital city Budapest and the subsamples pertaining to the nineteen counties. For each of these geographical units, the calibration of the weights of the corresponding subsample is performed independently of the other geographical units, and the following controls or benchmarks (i.e. population totals of the auxiliary variables) are used:

– totals of 20 age-sex groups (10 for males and 10 for females),
– the total population living in major cities (i.e. in cities with a county's rights),
– the total number of households.

The totals of age-sex groups relate to the non-institutional population and are updated every month using the demographic components method. The last two controls are derived from the updated population total of the county (or the capital) on the basis of proportions observed in the recent census.

The number of controls used for the full LFS sample is 20×22 = 440. Calibration is usually performed on the basis of monthly data, and the quarterly weights are derived from the monthly ones by division by three. It is worth noting that the entries in the column 'Working Age Population' in Table 1 are all aggregates (i.e. totals) of controls.

JACKKNIFE STRATEGIES TRIED FOR THE LABOUR FORCE SURVEY

The different jackknife strategies we have examined as possible tools of estimating the standard error for calibrated estimates (such as the entries in Table 1) include

– jackknife with incorrect weighting, i.e. using the VPLX with calibrated weights,
– jackknife combined with GREG or raking for correct weighting, furthermore,
– two other strategies called in what follows HCSO_1 and HCSO_2.

The latter identifiers should refer to the current practice of estimating standard errors for LFS data at the Hungarian Central Statistical Office on the one hand and on some improvement of that practice on the other.

The current practice is based on the assumption that standard errors computed with the VPLX for calibrated data are acceptable in the case of ratios, and that the estimation of the standard error of totals should be reduced to the case of ratios. Given that the sample weights are calibrated, consider an arbitrary estimated total $\hat{Y}$. For any auxiliary variable x the estimate $\hat{X}$ is equal to the population total X and

$$\hat{Y}_{\mathrm{RAT}} = \hat{X}\ (\hat{Y} / \hat{X}) = X\hat{R} \qquad /5/$$

where $\hat{R} = \hat{Y} / \hat{X}$ and RAT in the subscript refers to ratio estimate. A standard argument from sampling theory shows that the equality

$$\mathrm{Var}(\hat{Y}_{\mathrm{RAT}}) = X^2 \mathrm{Var}(\hat{R}) \qquad /6/$$

holds for the variances of $\hat{Y}_{\mathrm{RAT}}$ and $\hat{R}$ provided that under the given sampling design, calibration would be carried out in the same way for all possible samples. Nevertheless, if the variances in /6/ are replaced by their jackknife estimates using incorrect weighting, the inequality

$$\mathrm{var}_{jack}(\hat{Y}_{\mathrm{RAT}}) > X^2 \mathrm{var}_{jack}(\hat{Y}/\hat{X}) \qquad /7/$$

is obtained in the majority of cases. In the HCSO_1 strategy, the expression on the right-hand side of /6/ is the basis of estimating standard errors of totals. As we shall see, in this way a part of the bias coming from incorrect weighting is removed, since the effect of the auxiliary variable x on the study variable y is reflected in the estimates of the variance and the standard error.

The idea of the new strategy HCSO_2 is to modify /5/ in such a way that *more* auxiliary variables may have their impact on the *estimated* variance (and thus also on the *estimated* standard error) of $\hat{Y}$. Considering the auxiliary variables in the LFS, it is straightforward to decompose $\hat{Y}$ as follows:

$$\hat{Y} = \hat{Y}_{\mathrm{RAT}} = \hat{Y}_{1,\mathrm{RAT}} + \hat{Y}_{2,\mathrm{RAT}} + \ldots + \hat{Y}_{20,\mathrm{RAT}} ,$$

where $\hat{Y}_{i,\mathrm{RAT}}$ is the contribution of age-sex group i to the total $\hat{Y}_{\mathrm{RAT}}$, $i = 1, 2, ..., 20$. Following the pattern of /5/, each $\hat{Y}_{i,\mathrm{RAT}}$ can be written as

$$\hat{Y}_{i,\mathrm{RAT}} = \hat{X}_i(\hat{Y}_i / \hat{X}_i) = X_i\hat{R}_i ,$$

where X_i is the control for the age-sex group i, i = 1, 2, ..., 20. A straightforward choice for a variance estimate for $\hat{Y}$ is then

$$\text{var}_{jack}(\hat{Y}_{\text{RAT}}) = \sum_{i=1}^{20}\sum_{j=1}^{20}\sigma_{ij}X_iX_j ,$$

where σ_{ij} is the general entry of the 20×20 variance-covariance matrix of the estimated ratios $\hat{R}_i$, which can be estimated with the VPLX Software. Unlike its predecessor, in certain cases the strategy HCSO_2 is suitable to estimate the standard errors of rates, too. For instance, if the total of the working age population is an aggregate of controls, then participation rate and the total of individuals in the labour force are scalar multiples of each other, and the estimated standard error assigned to the latter divided by total working age population can be used as standard error of the participation rate.

The standard errors of the LFS data in Table 1 were estimated with eight different strategies and for three different periods, namely, December 2002, August 2003 and September 2003. Owing to space considerations, only a part of the numerical results will be presented in what follows, namely, the estimates obtained with five strategies using the data of September 2003. No relevant information will be lost in this way since in some cases two similar strategies have yielded practically the same estimates, though with markedly different run times, and the standard error estimates obtained for the different months in consideration show rather similar patterns. The original eight strategies are listed in Table 2.

Table 2

Jackknife strategies used to estimate the standard error of LFS estimates

Number	Description	Run time (min : sec)
1	Incorrect weighting: the use of the VPLX with calibrated weights	≈ 00:04.0
2	Incorrect weighting, the current strategy of the HCSO (HCSO_1)	≈ 00:04.0
3	Correct weighting. Calibration method: raking, convergence criterion: ±0.0001 At each jackknife replicate, the iteration starts with the original design weights	50:55.92
4	Correct weighting. Calibration method: raking, convergence criterion: ±0.0001 At each replicate, the iteration starts with the *recent* calibrated weights	18:19.23
5	Correct weighting. Calibration method: raking, convergence criterion: ±0.001 At each replicate, the iteration starts with the *recent* calibrated weights	6:54.93
6	Correct weighting. Calibration method: GREG. At each jackknife replicate, the calibrated weights are expressed in terms of the original design weights	16:56.69
7	Correct weighting. Calibration method: GREG. At each jackknife replicate, the calibrated weights are expressed in terms of the *recent* calibrated weights	16:20.27
8	Incorrect weighting, an improvement of the HCSO strategy (HCSO_2)	≈ 00:04.0

The CPU times given in the Table were recorded for the data set of September 2003. The programs were run on a machine with a Pentium III processor having the speed of 733 Mhz and a memory of 256 Mb. Strategies 1, 2 and 8 are VPLX applications, and the corresponding run times are approximate values, since the program does not report them. The programs for strategies 3-7 were written by the present author in IML – i.e. Interactive Matrix Language – of the SAS System, Version 8e. The underlying formulae were borrowed from the Yung–Rao paper [1996], first of all the jackknife variance expression which reads as follows:

$$\text{var}_{jack}(\hat{\theta}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (\hat{\theta}_{(hj)} - \hat{\theta})^2$$

where L is the number of design strata, n_h is the number of sampled clusters in stratum h, $\hat{\theta}$ and $\hat{\theta}_{(hj)}$ are estimates of some total or ratio. $\hat{\theta}$ and $\hat{\theta}_{(hj)}$ are based on the whole sample and on the part of the sample that remains after deleting the observations belonging to cluster j in stratum h; respectively. $\hat{\theta}_{(hj)}$ has the same functional form as $\hat{\theta}$.

Strategy 3 is based on the jackknife principle recalled in the preceding paragraph. The convergence criterion refers to the ratio of the left-hand side of the calibration equation /3/ to the right-hand side, i.e. to the given population control. Strategies 4 and 5 are relaxed versions of strategy 3. The former departs from the rigorous principle allowing the use of weights obtained for the recent replicate as starting point for the next replicate. In terms of the distance function /4/, this means that we start closer to our goal, i.e. to the system of final weights of the current replicate than in the case of strategy 3. As a consequence, less iteration is required, and this is reflected in the run times 50 min 55.92 sec and 18 min 19.23 sec, respectively. Strategy 5 contains some additional relaxation, namely, the convergence criterion is set to ± 0.001 instead of ±0.0001, implying a further reduction in the run time down to 6 min 54.92 sec. If $s_i(\hat{Y})$ is the estimated standard error obtained with strategy i (i = 1, 2, ...,8) for some estimated level $\hat{Y}$, define the deviations $d_{ij} = d_{ij}(\hat{Y}) = 100 * (s_i(\hat{Y}) - s_j(\hat{Y}))/s_j(\hat{Y})$ for $1 \leq i, j \leq 8$, $i \neq j$. Over a set of 43 estimated totals from the LFS in September 2003, we have found that min(d_{43}) = –0.22, mean(d_{43}) =0.8, max(d_{43}) = 3.47, min(d_{53}) = –0.33, mean(d_{53}) = 1.27 and max(d_{53}) = 5.45 (all data in percentages). In view of these small deviations, the estimated standard errors $s_4(\cdot)$ and $s_5(\cdot)$ are not displayed, i.e. the results of strategies 4 and 5 will be represented by those of strategy 3.

Strategy 6 is the counterpart of strategy 3 with raking replaced with generalised regression as the method of calibration. Although the GREG procedure is rarely used in Hungarian household surveys, it seemed important to compare it with raking in the context of the jackknife method and calibrated estimates. It should not be surprising that the GREG produces LFS data slightly different from those obtained with raking. The estimator based on raking is not identical to the one based on the GREG. Furthermore, one of the calibration equations was dropped, since quasi-multicollinearity was detected in the system of equations /3/. The GREG estimates corresponding to the entries of Table 1 are given in Table 3. The differences between the data of Table 1 and Table 3 are within the limits of sampling error; in particular, the entries in the column 'Working age population' are, apart from some round-off errors, practically the same in both tables.

Table 3

Hungarian labour force survey, September 2003, calibration method: generalised regression

Age-groups	Employed	Unemployed	In labour force	Not in labour force	Working age population	Participation rate, percent	Unemployment rate, percent
Total	*3 968 524*	*223 331*	*4 191 855*	*3 549 395*	*7 741 250*	*54.15*	*5.33*
15–19	25 227	11 118	36 345	588 311	624 656	5.82	30.59
20–24	327 692	43 197	370 889	326 953	697 842	53.15	11.65
25–29	595 475	39 236	634 711	207 537	842 248	75.36	6.18
30–39	1 059 494	52 428	1 111 922	243 983	1 355 906	82.01	4.72
40–54	1 590 996	67 712	1 658 707	496 798	2 155 505	76.95	4.08
55–59	281 439	8 018	289 457	329 360	618 817	46.78	2.77
60–69	80 100	1 622	81 722	938 823	1 020 544	8.01	1.98
70–74	8 101	0	8 101	417 631	425 732	1.90	0.00
Male	2 156 439	126 744	2 283 183	1 405 747	3 688 930	61.89	5.55
Female	1 812 085	96 587	1 908 672	2 143 648	4 052 320	47.10	5.06

Strategy 7 is a relaxed version of strategy 6. When jackknife replicates are computed, the calibration procedure uses the *recent* calibrated weights belonging to the previous replicate rather than the original design weights. Just as in the case of strategies 3-4, the input weights resulting from calibration in strategy 7 are closer to the calibration result than those resulting from strategy 6: the distance being measured in this case is based on /2/. For strategy 7, this does not yield perceptible gains in run time (16 min 20.27 sec vs. 16 min 56.69 sec.). In terms of the notations introduced above, the comparison of the results of strategies 6 and 7 yields min(d_{76}) = –0.75, mean(d_{76}) = 0.22 and max(d_{76})=1.14 percent, respectively. Therefore, only the estimated standard errors $s_6(\cdot)$ will appear in the tables, but not $s_7(\cdot)$.

In what follows standard error estimates obtained with strategies 1, 2, 3, 6 and 8 will be compared. Note that when generating jackknife replicates for strategy 6 using the GREG, some calibrated weights can be negative, and they were not excluded from the computations in the current research. Table 4 contains the different standard error estimates for the totals of employed, unemployed, individuals in and not in the labour force in different breakdowns by sex and age groups. The corresponding standard error estimates for the rates of participation and unemployment can be found in Table 5. Since the entries in the column 'Working age population' in Tables 1 and 3 are aggregates of controls; they have no sampling variability over the set of possible samples if calibration is performed in the same way for each of them. In other words, the standard errors associated with these estimates vanish. However, this is not the case if the *estimates* of these standard errors are considered. In the case of correct weighting, numerical inaccuracies in inverting matrices or taking limits, result in some small positive estimates of the standard error; that is the corresponding estimated coefficient of variation never exceeded 5×10^{-4}.

In contrast to this, the biggest relative bias associated with the application of incorrect weighting was found just in the case where the estimates agreed with sums of some controls: far from being zero, the estimated variance was practically the same as that obtained when there was no calibration at all.

Table 4

Standard errors for Hungarian LFS data in September 2003, estimated by different jackknife applications

Denomination	Incorrect weighting	Current HCSO strategy	Correct weighting		Improved HCSO strategy
			calibration with raking	calibration with GREG	
Employed					
Total	*49 597*	*32 513*	*21 069*	*26 416*	*25 485*
15–19	3 223	3 186	2 676	3 430	3 167
20–24	12 800	9 142	7 585	9 249	9 188
25–29	19 756	9 518	7 869	9 375	9 485
30–39	27 063	10 169	8 519	10 355	10 247
40–54	30 332	13 795	11 469	14 283	13 765
55–59	12 481	8 602	7 310	8 756	8 524
60–69	6 546	6 225	5 297	6 257	6 232
70–74	2 093	2 086	1 818	2 120	2 080
Male	30 504	20 289	13 287	16 504	15 859
Female	28 051	21 477	15 050	18 442	18 234
Unemployed					
Total	*9 436*	*9 289*	*7 689*	*9 907*	*9 351*
15–19	1 942	1 936	1 501	2 195	1 931
20–24	4 089	3 908	3 177	4 102	3 944
25–29	3 907	3 874	3 281	3 918	3 831
30–39	4 603	4 610	3 763	4 744	4 555
40–54	4 873	4 742	3 891	4 969	4 817
55–59	1 733	1 733	1 429	1 732	1 722
60–69	1 006	1021	884	948	1 007
70–74	n.a.	n.a.	n.a.	n.a.	n.a.
Male	7 070	7 009	5 698	7 343	6 993
Female	5 904	6 078	4 857	6 125	5 887
In labour force					
Total	*50 486*	*32 513*	*20 691*	*25 753*	*24 731*
15–19	3 745	3 623	3 030	3 997	3 650
20–24	13 789	9 421	7 873	9 618	9 483
25–29	20 188	8 928	7 372	8 756	8 813
30–39	27 531	9 355	7 982	9 715	9 469
40–54	30 812	13 149	11 054	13 773	13 211
55–59	12 693	8 725	7 385	8 858	8 606
60–69	6 595	6 327	5 335	6 303	6 273
70–74	2 093	2 086	1 818	2120	2 080
Male	31 000	19 920	12 805	15 788	15 033
Female	28 375	21 477	14 856	18 143	17 977
Not on labour force					
Total	*39 002*	*32 513*	*20 635*	*25 752*	*24 731*
15–19	17 232	3 623	3 031	3 998	3 650
20–24	13 585	9 421	7 878	9 619	9 483
25–29	10 752	8 928	7 372	8 756	8 813
30–39	10 493	9 355	7 974	9 715	9 469
40–54	14 442	13 149	11 037	13 773	13 211
55–59	12 172	8 725	7 395	8 858	8 606
60–69	21 619	6 327	5 351	6 303	6 273
70–74	13 202	2 086	1 822	2 120	2 080
Male	24 051	19 920	12 774	15 788	15 033
Female	26 922	21 477	14 840	18 142	17 977

Turning to Table 4, note that the estimated standard errors obtained with correct weighting and raking are uniformly smaller over the 43 variables in consideration. These results are in line with what has been observed in the literature: for example *Stukel, Hidiroglou* and *Särndal* [1996]. Jackknife variance estimates are generally larger than those obtained using the Taylor expansion. The estimates produced by the current HCSO strategy represent considerable improvement to those obtained using the incorrect jackknife variance procedure. However, these improved estimates are often far from those resulting from correct weighting with raking; examples are total employed (49 597, 32 513, 21 069), employed male (30 504, 20 289, 13 287), etc. For small totals such as total unemployed and unemployed aged 20-24 the differences are smaller: the corresponding triples are 9 436, 9 286, 7 689, and 4 089, 3 908, 3 177, respectively. The correct weighting with GREG, which is also supposed to yield practically unbiased estimates, results in slightly greater standard error estimates than its counterpart that uses raking. We should recall here that the GREG and the raking result in two different estimators; e.g. in the case of individuals aged 15–19 in the labour force the two estimates of the total are 35 712 and 36 343, respectively.

Table 5

Standard errors for Hungarian LFS rates in September 2003,
estimated by different jackknife applications
(percent)

Denomination	Incorrect weighting	Current HCSO strategy	Correct weighting		Improved HCSO strategy
			calibation with raking	calibration with GREG	
Participation rate					
Total	0.42	0.42	0.27	0.33	0.32
15–19	0.58	0.58	0.48	0.64	0.58
20–24	1.35	1.35	1.13	1.38	1.36
25–29	1.06	1.06	0.88	1.04	1.05
30–39	0.69	0.69	0.59	0.72	0.70
40–54	0.61	0.61	0.51	0.64	0.61
55–59	1.41	1.41	1.19	1.43	1.39
60–69	0.62	0.62	0.52	0.62	0.61
70–74	0.49	0.49	0.43	0.50	0.49
Male	0.54	0.54	0.35	0.43	0.41
Female	0.53	0.53	0.37	0.45	0.44
Unemployment rate					
Total	0.22	0.22	0.18	0.23	0.22
15–19	4.74	4.74	3.79	5.25	4.74
20–24	1.02	1.02	0.82	1.06	1.02
25–29	0.61	0.61	0.52	0.62	0.61
30–39	0.41	0.41	0.34	0.42	0.41
40–54	0.29	0.29	0.23	0.30	0.29
55–59	0.59	0.59	0.49	0.59	0.59
60–69	1.20	1.20	1.06	1.16	1.20
70–74	n.a.	n.a.	n.a.	n.a.	n.a.
Male	0.30	0.30	0.25	0.32	0.30
Female	0.31	0.31	0.25	0.32	0.31

Finally, the estimates produced with the improved HCSO strategy (HCSO_2) are surprisingly close to those obtained with the correct weighting using GREG. This is remarkable since HCSO_2 is based on the use of the VPLX Software which uses incorrect weighting. Note that VPLX runs in a very short run time (approximately 4 seconds) as compared to correct weighting with the GREG (16 minutes 57 seconds).

In our experience, the bias resulting from incorrect weighting is not significant for estimated standard error of ratios not exceeding 12 percent. This is shown in Table 5, especially in the rows of unemployment rate where the only outlier is the group of individuals aged 15–19 having an unemployment rate of about 30 percent. In the case of ratios close to 50 percent the biasing effect of incorrect weighting is considerable: this strategy yields the standard errors 0.54 and 0.53 for the participation rates of male and female, respectively, in contrast to the corresponding figures 0.35 and 0.37 obtained with correct weighting with raking. According to Table 5, the performance of the improved HCSO strategy is similar to that of correct weighting using the GREG in the case of standard errors of ratios.

Repeating the computations with the LFS data sets of August 2003 and December 2002 has led to the following experience. Using the data of August 2003 has yielded practically the same tendencies which can be seen in Tables 4 and 5, the actual differences in estimates were clearly due to the changes over time in the variables observed. This is not surprising since the LFS samples in August and September are statistically equivalent, each being one third of the quarterly sample. In addition, the actual changes in the variables from August to September were moderate, though the decrease in the level of unemployment was actually significant.

Table 6

Standard errors for some Hungarian LFS data in December 2002, estimated by different jackknife applications

Denomination	Incorrect weighting	Current HCSO strategy	Correct weighting		Improved HCSO strategy
			calibration with raking	calibration with GREG	
Employed					
Total	56 505	34 134	22 015	29 562	28 908
40–54	31 997	15 266	11 686	15 764	15 234
Male	35 261	21 086	13 795	18 528	17 514
Female	29 703	22 321	14 994	19 623	19 732
In labour force					
Total	59 540	34 134	21 408	28 715	28 029
Male	36 830	20 716	13 050	17 458	16 603
Female	30 927	22 321	14 906	19 604	19 596
Participation rate (percent)					
Total	0.44	0.44	0.28	0.37	0.36
Male	0.56	0.56	0.35	0.47	0.45
Female	0.55	0.55	0.37	0.48	0.48

Turning to the different standard error estimates obtained for the data of December 2002, the deviations from those pertaining to September 2003 are greater. These deviations come from different sources, of which the most important one is probably the difference between the old and the new sampling design. In particular, the number of clusters – the building blocks of creating jackknife replicates – was 2096 in August 2003 and 2123 in the next month, but it amounted to 4730 in December 2002. This made correct jackknife computation even more time-demanding for the old sample than for the new one; with the data of December 2002, correct weighting has used 1 hour 36 min 37.8 sec with raking and 42 min 26.49 sec with the GREG. Nevertheless, the main conclusions remained the same as in the case of the new sample (i.e. August and September 2003). These are as follows: the standard error estimates obtained with correct weighting and raking are uniformly less than those obtained with other strategies, and the improved HCSO technique results in estimates well approximating those obtained with correct raking and the GREG. The HCSO_2 strategy yields in this case relatively less gains in precision than in the case of the new sample; this is best shown by estimates based on the whole sample or on large subsamples, see Table 6.

*

Different strategies of the jackknife method to estimate the variance of some data of the Hungarian Labour Force Survey (LFS) were investigated in the paper. If the jackknife method is used in the case of calibrated estimates, the procedure of calibration should be repeated whenever a new pseudo value or jackknife replicate is created; otherwise the weighting will be incorrect. On the one hand, correct weighting demands unusual long run time even on fast modern personal computers – a case is reported in the paper where solving a medium size problem needed 51 minutes –, on the other hand, the use of incorrect weighting may cause serious bias in the estimated variances and standard errors. It is pointed out in the paper that

– if raking is used to ensure correct weighting, slight modifications can reduce the run time of 51 minutes to 18 or even 7 minutes, at the cost of acceptable loss in precision, and

– with suitable algebraic manipulation, the use of available software for the jackknife can be organised so that the biasing effect of incorrect weighting may be compensated for by the proper use of the controls occurring in the calibration procedure. A new strategy labelled HCSO_2 in the paper has produced similar variance (and standard error) estimates as a strategy of correct weighting based on generalised regression estimation.

The experience described in the paper is based on a series of computations using monthly data sets of the Hungarian LFS from different periods as input file: December 2002, August and September 2003. Our results reveal some promising features of the strategy HCSO_2, but obviously, further research is needed to evaluate this method. On the one hand, its performance should be compared with that of the linearised jackknife, which is widely used and is practically equivalent to jackknife with correct weighting. On the other hand, the relationship between jackknife with correct weighting and the HCSO_2 strategy ought to be examined to see if there is some theoretical reason explaining why the results obtained with the latter approximate so well those produced by the former.

APPENDIX

In what follows a subroutine written in SAS-IML language for performing calibration on the basis of the Darroch-Ratcliff procedure (i.e. generalised iterative scaling) is presented.

```
start scaling;
   work = w;    /* w is the column vector of design weghts (input) */
   eps1=1000.;
   eps2 = 0;
   it = 0;
   do while (it < iter & (eps1 > upper | eps2 < lower));      /* by default, iter (# of iteration steps) =1200,
upper=1.0001, lower=0.9999 */
   it=it+1;
   u = u / u1;     /* u is the row vector of updating factors */
   work = work # u`;     /* updating of the weights; u` is the transpose of u */
   do j=1 to n1;     /* n1 = dimension of w (and also of u) */
   if work[j] < 90. then work[j] = 90.;     /* lower bound of the weights = 90 */
   if work[j] > 1500. then work[j] = 1500.;     /* upper bound of the weights = 1500 */
   end;
   y = q * work;     /* q is the 22*n1 matrix in the calibration equations */
   y = y<>ymin;     /* replacing possible zeroes in y by 1 */
   r = cc / y`;     /* r and cc are the 22-dimensional row vectors of the scaling factors of the equations and
the controls, respectively; y` is the transpose of y */
   do ji=1 to 22;     /* # of controls = 22 */
   if r[ji]=0 then r[ji]=1.0;
   end;
   eps1 = r[<>];
   eps2 = r[><];
   u = r * q;     /* computation of the updating factors */
   u1 = q[+,];
   end;
   work = floor(work+0.5); /* rounding the calibrated weights (output) */
   free q y r;
finish scaling;
```

This subroutine is called in the main program by the ‘call scaling;’ statement. A single call results in calibrated weights for the subsample of some of the 20 adminstrative units of the country. With the notations of the program, the system of calibration equations can be written as q*w =cc or q*work = cc. The variables w, q, n1, ‘upper’,‘lower’ and cc described between the parentheses ‘/*’ and ‘*/’ should be set values before calling the subroutine. The 22-dimensional column vector ymin as well as the n1-dimensional row vectors u and u1 should be initialised before the call in the main program, setting all components equal to 1. The fixed values lower bound = 90, upper bound = 1500 and the numbers of controls (22) should be changed if the input data are other than those of the Hungarian LFS of some month. The maximal number of iteration steps as well as the tolerance limits (‘lower’ = 0.9999, ‘upper’ = 1.0001 and iter = 1200) may be changed optionally.

Remark. In the *Darroch–Ratcliff* algorithm, the updating factor u_j of the weight w_j is the geometric mean of the scaling factors r_1, r_2, ..., etc. weighted by the entries in column j of the matrix q (or, with the notations in /3/, by the components of the vector $\mathbf{x}_j$). In contrast, the above subroutine uses the corresponding weighted arithmetic mean. This causes some slight deviation form the unique solution of the optimisation problem /3/-/4/, which is, however, compensated by some technical advantages.

REFERENCES

DARROCH, J. N. – RATCLIFF, D. [1972]: Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*. No. 43. pp. 1470–1480.

DEVILLE, J.-C. – SÄRNDAL, C.-E. [1992]: Calibration estimators in survey sampling. *Journal of the American Statistical Association*. No. 87. pp. 376–382.

DEVILLE, J. C. – SÄRNDAL, C.-E. – SAUTORY, O. [1993]: Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*. No. 88. pp. 1013–1020.

FAY, R. E. [1998]: VPLX software. Variance estimation for complex surveys. http://www.census.gov.

STUKEL, D. M. – HIDIROGLOU, M. A. – SÄRNDAL, C.-E. [1996]: Variance estimation for calibration estimators: a comparison of jackknifing versus Taylor linearization. *Survey Methodology*. No. 22. pp. 117–125.

WOLTER, K. M. [1985]: Introduction to variance estimation. Springer. New York – Berlin – Heidelberg – Tokyo.

YUNG, W. – RAO, J. N. K. [1996]. Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*. No. 22. pp. 23–31.

DIAGNOSTICS OF THE ERROR FACTOR COVARIANCES

OTTÓ HAJDU[1]

In this paper we explore initial simple factor structure by the means of the so-called '*EPIC*' factor extraction method and the '*orthosim*' orthogonal rotational strategy. Then, the results are tested by confirmative factor analysis based on iteratively reweighted least squares on the one hand and asymptotically distribution free estimation on the other hand. Besides, based on multivariate kurtosis measures, multivariate normality is also investigated to see whether the use of the IWLS method is appropriate or a robust ADF estimator with relatively larger standard error is preferred. Finally, the paper draws attention that confidence intervals for the non-centrality based goodness of fit measures are available.

KEYWORDS: Latent variables; Covariance structural equations; Heteroscedasticity; Goodness of fit.

A model that relates measured variables to latent factors in covariance structure analysis is called a measurement model. These models are mostly factor analysis models and it is standard to distinguish between *confirmatory* and *exploratory* approach. In an exploratory factor analysis, we may not know how many factors are needed to explain the inter-correlations among the indicators. In addition, even if we are sure about the existence of a particular factor, we may not know which variables are the best indicators of the factor. Exploratory factor analysis will give us results: the number of factors, the factor loadings, and possibly the factor correlations.

In contrast, if we anticipate these results, we can do a confirmatory factor analysis. In this type of factor analysis, we presumably have a *hypothesis* about the number of factors, which measured variables are supposedly good indicators of each of the factors, which variables are unrelated to a factor and, how strongly or weakly the factors correlate to each other. In confirmatory models, variables are presumed to be factorially simple. That is, a given indicator is usually expected to be influenced by very few factors, typically only one. In addition, the covariance structure of the error factors can be arbitrary if it is reasonably justified and the model identification permits it. This means that the error (unique) factors are not necessarily uncorrelated but their variances may be equal by the homogeneity hypothesis.[2] Of course, hypothesis may be incorrect hence it must be tested

[1] Associate professor of the Budapest University of Technology and Economic Sciences.
[2] This error covariance structure is analogous to that used in the econometrics literature.

by sample information. Nevertheless, the hypothetical simple factor structure could be explored by orthogonal or oblique rotation of factor loadings carried out on an initial loading matrix produced by some factor extraction method.

The aim of this paper is twofold. In an explorative step we draw attention to a factor extraction method named EPIC (Equal Prior Instant Communalities) and an orthogonal rotation technique called *orthosim*. The EPIC method is used in our analysis as a compromise between two methods: the principal components method, which is computationally simple and the maximum likelihood factor analysis, which frequently leads to convergence problems. The 'orthosim' solution (proposed by *Bentler* [1977]) does not optimize some simplicity criterion within the loading matrix instead it maximizes a generalized variance type factor simplicity index corresponding to the loading matrix as a whole.

The core problem is that the initial unrotated EPIC factor solution (named by *Kaiser* [1990]) is based on a method in which the variances of the *uncorrelated* error factors are initially taken as *equal*, permitting the computations to be done explicitly and untroubled by linear dependencies among the variables (*Anderson* [1984] p. 21.). Nevertheless, the homogeneity assumption of the equal unique variances, as well as the factor pattern itself is merely a hypothesis. Therefore, it must be tested by using a confirmatory factor analysis step. There are two main approaches available to estimate the parameters of a confirmatory factor model. The first is based on some normality assumptions. However, if the normality assumptions are violated, asymptotically distribution free (ADF) approach must be used. This article gives a review of multivariate kurtosis measures to help decision whether the use of ADF method (with relatively larger standard errors) is necessary or not. In addition, the paper draws attention to those goodness of fit measures for which confidence intervals are available.

Finally, the paper illustrates the problems investigated based on microeconomic balance-sheet data. The computations are performed using the statistical programs 'Statistica 6.0' and 'EQS'.

THE ROLE OF UNCORRELATEDNESS IN THE FACTOR MODEL

In factor analysis, one assumes that certain observable variables (indicators) correlate because there are one or several underlying latent factors that generate the observed **x** data. The parametric form of the factor analysis model is given by

$$\mathbf{x}_{(p,1)} = \mathbf{\Lambda}_{(p,m)}\mathbf{f}_{(m,1)} + \mathbf{u}_{(p,1)} \qquad /1/$$

where vector $\mathbf{x}=[x_1, x_2,..., x_p]^T$ consists of p indicators, vector $\mathbf{f}=[f_1,f_2,...,f_m]^T$ consists of m common (latent) factors and $\mathbf{u}=[u_1, u_2,..., u_p]^T$ represents the *error* factors, *unique* to that indicator.[3] The so-called '*pattern matrix*' $\mathbf{\Lambda}$ of order (p, m) consists of λ_{jk} factor loadings. The higher the value of a loading in absolute magnitude the more important the

[3] The 'error factor' and 'unique factor' terminologies are used synonymously in this paper.

factor is. Using /1/ we can express the **C** covariance matrix of order (p,p) among the observed indicators based on covariances as follows:

$$\mathbf{C}=\mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T+\mathbf{C}_{uu}+\mathbf{\Lambda C}_{fu}+\mathbf{C}_{uf}\mathbf{\Lambda}^T . \qquad /2/$$

It is apparent, that **C** has $p(p+1)/2$ distinct elements (including the variances on the main diagonal as well) but the total number of the unknown parameters in /2/ is

$$pm+\frac{m(m+1)}{2}+\frac{p(p+1)}{2}+mp . \qquad /3/$$

The factor model is identified when the number of parameters to be estimated q is less than the number of the distinct observed covariances that is:

$$df=\frac{p(p+1)}{2}-q>0 \qquad /4/$$

where df is the degree of freedom. Hence, it is necessary to reduce substantially the number of parameters to be estimated relative to the number of indicators. This can be achieved by imposing hypothetical restrictions on the parameters and by increasing the number of indicators.

Such straightforward assumption is that the unique factors are uncorrelated with the common factors, i.e. equation $\mathbf{C}_{fu}=\mathbf{C}_{uf}=\mathbf{0}$ holds in /2/. This restriction yields a decomposition of the observed covariance matrix in the following form:

$$\mathbf{C}=\mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T+\mathbf{C}_{uu} . \qquad /5/$$

A further reasonable restriction that can be imposed is that the unique factors are uncorrelated with each other as well. This means that the covariance matrix $\mathbf{C}_{uu}$ is diagonal. Based on this additional restriction the decomposition of the observed covariance matrix takes the form:

$$\mathbf{C}=\mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T+\mathbf{\Psi}^2 \qquad /6/$$

where $\mathbf{\Psi}^2$ is our standard notation for the *diagonal* covariance matrix of the unique factors. In addition, if the unique variances are homogeneous, i.e. all of them are equal to a constant σ^2, the covariance decomposition is as follows:

$$\mathbf{C}=\mathbf{\Lambda C}_{ff}\mathbf{\Lambda}^T+\sigma^2\mathbf{I} . \qquad /7/$$

Using now the conventional notation of $\mathbf{C}_{ff}=\mathbf{\Phi}$ and concerning orthogonal factors, $\mathbf{\Phi}$ is diagonal and, further, assuming standardized factors, $\mathbf{\Phi}$ equals the identity. A non-diagonal $\mathbf{\Phi}$ indicates 'oblique' (i.e. correlated) factors.

Explorative factor analysis is basically aimed at estimating the $\left(\mathbf{\Lambda},\mathbf{\Phi},\mathbf{\Psi}^2\right)$ parameters in /6/ without any presumed knowledge about them except, that the common factors are standardized (i.e. $\mathbf{\Phi}$ is a correlation matrix). In contrast, in a confirmative analysis interpretable parameters are selected to be estimated rather than accepting any computationally convenient assumptions. Our focus in this paper is mainly on testing the hypothesized structure of $\mathbf{\Psi}^2$.

Once a solution is obtained, with any $\mathbf{T}_m$ non-singular matrix of order m equation /1/ is still satisfied in the following form:

$$\mathbf{x} = \left(\mathbf{\Lambda}\mathbf{T}^{-1}\right)\left(\mathbf{T}\mathbf{f}\right) + \mathbf{u} . \qquad /8/$$

Replacing $\mathbf{f}$ by $\mathbf{f}^*=\mathbf{Tf}$ and $\mathbf{\Lambda}$ by $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{T}^{-1}$ we perform an *oblique rotation* which results in the covariance matrix $\mathbf{T}\mathbf{\Phi}\mathbf{T}^T = \mathbf{\Phi}^*$ of the rotated factors and preserves the reproduced covariance matrix being unchanged:

$$\mathbf{\Lambda}^*\mathbf{\Phi}^*\mathbf{\Lambda}^{*T} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T .$$

Specially, an *orthogonal rotation* is performed when the factors are uncorrelated and $\mathbf{T}$ is orthonormal: $\mathbf{T}^T = \mathbf{T}^{-1}$.

Our final goal is to give a pattern of loadings as clear as possible that is factors that are clearly marked by high loadings for some variables and low loadings for others. This general pattern is referred to as „simple structure'. This can be achieved by a two-step approach. First, in the explorative step we estimate the orthogonal loadings and subsequently rotate them. Then, in the confirmative step we fix some parameters at some (typically zero or equal) hypothetical value, reestimate the *free* parameters and test the goodness of fit. Specially, the adequacy (goodness of fit) of a specific orthogonal or oblique factor solution can directly be tested by confirmative factor analysis.

There are various rotational strategies that have been proposed in the field to explore a clear pattern of loadings. The most widely used orthogonal rotational strategy is the so-called varimax method (*Kaiser* [1958], *Ten Berge* [1995]). Despite the popularity of varimax, we shall use another method based on a different approach named *orthosim* (*Bentler* [1977]).

The so-called '*orthosim*' orthogonal rotation is based on a factorial *simplicity* index. Let us start with a known loading matrix $\mathbf{A}$ and transform it with an orthonormal $\mathbf{T}$ into $\mathbf{B} = \mathbf{AT}$, such that $\mathbf{D} = diag\left(\left(\mathbf{B}*\mathbf{B}\right)^T\left(\mathbf{B}*\mathbf{B}\right)\right)$ is a diagonal matrix and * denotes the element-wise (Hadamard) product. Then we seek the rotated pattern matrix which *maximizes* the index of factorial simplicity defined as the *generalized variance* as follows:

$$GV = \det\left(\mathbf{D}^{-1/2}\left(\left(\mathbf{B}*\mathbf{B}\right)^T\left(\mathbf{B}*\mathbf{B}\right)\right)\mathbf{D}^{-1/2}\right) \rightarrow \max .$$

This determinant (based on a symmetric, nonnegative definite matrix with unit diagonal elements) ranges between zero and one. It equals zero when there are linear dependencies among the columns of $(\mathbf{B} * \mathbf{B})$. Such a case when some columns of **B** are proportional or identical except for sign. The maximum value of one occurs if the matrix in *GV* is the identity. This means that the factor pattern is factorially simple. It must be emphasized that provided a diagonal scaling matrix *GV* is invariant with respect to the column rescaling of **B**.

The concept of oblique rotations can be used in order to obtain more interpretable simple structure that best represents the 'clusters' of variables, without the constraint of orthogonal factors. One of the recommended widely used methods is the *direct quartimin method* (*Jennrich–Sampson* [1966]).

THE 'EQUAL PRIOR INSTANT COMMUNALITIES' FACTOR EXTRACTION METHOD

Under the hypothesis of this orthogonal factor model, the unique variances of the covariance (correlation) matrix are presumed to be equal. According to the standard orthogonality and the Kaiser-normalization requirements the matrix

$$\mathbf{\Lambda}^T \mathbf{\Psi}^{-2} \mathbf{\Lambda} = \mathbf{D} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m \end{bmatrix}$$

is diagonal and the maximum likelihood (ML) equations /9/ and /10/ must hold (*Lawley–Maxwell* [1971] p. 27. EQ 4.9; p. 30. EQ 4.19):

$$\mathbf{\Psi}^2 = diag(\mathbf{C} - \mathbf{\Lambda}\mathbf{\Lambda}^{T)}\,, \qquad /9/$$

$$\mathbf{C}\mathbf{\Psi}^{-2}\mathbf{\Lambda} = \mathbf{\Lambda}(\mathbf{I}_m + \mathbf{D})\,, \qquad /10/$$

where **C** is the covariance matrix of the observed indicators. Alternatively, equation /10/ can be written as:

$$(\mathbf{C} - \mathbf{\Psi}^2)\mathbf{\Psi}^{-2}\mathbf{\Lambda} = \mathbf{\Lambda}\mathbf{D}\,.$$

It is apparent that the columns of $\mathbf{\Lambda}$ are the eigenvectors corresponding to the largest m eigenvalues of matrices $\mathbf{C}\mathbf{\Psi}^{-2}$, or $(\mathbf{C} - \mathbf{\Psi}^2)\mathbf{\Psi}^{-2}$.

Let us suppose, that the uncorrelated unique factors are *homogeneous* i.e. $\mathbf{\Psi}^2 = \sigma^2 \mathbf{I}_p$, and consider the standard spectral decomposition of the covariance matrix **C** of the indicators. Taking only the first m eigenvalues on the main diagonal of the diagonal matrix $\mathbf{U}_m$ then:

$$\mathbf{C}\mathbf{W}_m = \mathbf{W}_m \mathbf{U}_m\,, \quad \mathbf{W}_m^T \mathbf{W}_m = \mathbf{I}_m\,, \qquad /11/$$

where the columns of $\mathbf{W}_m$ are the corresponding eigenvectors. After some manipulations we can write /11/ equivalently as:

$$\mathbf{C}\left(\sigma^2\mathbf{I}_p\right)^{-1}\left(\sigma\mathbf{W}_m\left(\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m\right)^{\frac{1}{2}}\right)=\sigma\mathbf{W}_m\left(\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m\right)^{\frac{1}{2}}\frac{1}{\sigma^2}\mathbf{U}_m .$$

Clearly, making the

$$\mathbf{\Lambda}=\sigma\mathbf{W}_m\left(\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m\right)^{\frac{1}{2}}, \qquad /12/$$

$$\mathbf{D}=\frac{1}{\sigma^2}\mathbf{U}_m-\mathbf{I}_m , \qquad /13/$$

$$\mathbf{\Psi}^2=\sigma^2\mathbf{I}_p \qquad /14/$$

substitutions – provided homogeneous unique factors – our all initial ML requirements are met.

The estimation of variance σ^2 happens in the following manner. Given that $tr\left(\mathbf{\Lambda}_m\mathbf{\Lambda}_m^T\right)$ is the sum of the m communalities.

$$p\sigma^2=\sum_{j=1}^{p}\mathrm{var}\left(x_j\right)-tr\left(\mathbf{L}_m\mathbf{L}_m^T\right)$$

and based on equations /12/, /13/ and $\mathbf{W}_m^T\mathbf{W}_m=\mathbf{I}_m$, we obtain

$$p\sigma^2=\sum_{j=1}^{p}u_j-tr\left(\mathbf{W}_m\sigma^2\mathbf{D}\mathbf{W}_m^T\right)=\sum_{j=1}^{p}u_j-tr\left(\left(\mathbf{U}_m-\sigma^2\right)\mathbf{W}_m^T\mathbf{W}_m\right)=$$

$$=\sum_{j=1}^{p}u_j-\left(\sum_{j=1}^{m}u_j-m\sigma^2\right),$$

from which

$$\left(p-m\right)\sigma^2=\sum_{j=m+1}^{p}u_j . \qquad /15/$$

We draw attention that $\mathbf{WU}^{1/2}$ gives the standard principal components loading matrix.

There are two reasons why the $\mathbf{\Psi}^2 = \sigma^2 \mathbf{I}_p$ assumption is not as restrictive as it seems: first, the unique variances for the covariance matrix are not used in the computations and are not presumed to be equal when an explorative factor extraction is carried out on the correlation matrix as it is the usual case. Rather, the ratio of common factor variance to unique variance is hypothesized as equal for all variables under the model. Second, the estimated communalities for the correlation matrix, obtained from the solution, can vary substantially in practice.

AN EXPLORATIVE STUDY BASED ON MICROECONOMIC FINANCIAL INDICATORS

Based on balance-sheet data of 2117 Hungarian economic units from the branch with NACE code '5011' in 1999, the following indicators have been investigated by EPIC factor analysis followed by orthosim and direct quartimin rotations:

Profit after taxation / Liabilities: 'ATPLIAB'
Cash-Flow / Liabilities: 'CFLIAB'
Current ratio = Current assets / Short term liabilities: 'CURRENT'
Adjusted Current ratio = (Current assets-Inventories) / Short term liabilities: 'ACURRENT'
Long term liabilities / (Long term liabilities + Owner's equity): 'DEBT'
Owner's equity / (Inventories + Invested assets): 'EQUITYR'

The cases with observed value smaller than –10 and those with larger than 10 are excluded from the analysis. The covariance and correlation matrices of these 6 variables are given in Table 1.

Table 1

Covariances and correlations of the financial microeconomic indicators (N=2117)

Variable	ATPLIAB	CFLIAB	CURRENT	ACURRENT	DEBT	EQUITYR
			Covariance matrix			
ATPLIAB	0.513	0.501	0.118	0.155	–0.086	0.193
CFLIAB	0.501	0.571	0.155	0.181	–0.110	0.223
CURRENT	0.118	0.155	0.837	0.842	–0.189	0.571
ACURRENT	0.155	0.181	0.842	1.566	–0.289	0.671
DEBT	–0.086	–0.110	–0.189	–0.289	0.596	–0.934
EQUITYR	0.193	0.223	0.571	0.671	–0.934	2.721
			Correlation matrix			
ATPLIAB	1.000	0.927	0.180	0.173	–0.155	0.163
CFLIAB	0.927	1.000	0.225	0.191	–0.188	0.179
CURRENT	0.180	0.225	1.000	0.735	–0.268	0.378
ACURRENT	0.173	0.191	0.735	1.000	–0.299	0.325
DEBT	–0.155	–0.188	–0.268	–0.299	1.000	–0.733
EQUITYR	0.163	0.179	0.378	0.325	–0.733	1.000

The eigenvalues of the *correlation* matrix constitute the main diagonal of the diagonal matrix **U**=< 2.709, 1.589, 1.101, 0.306, 0.222, 0.072 > where the first three largest roots account for a variance explained of 90 percentage. In order to extract 3 unrotated factors named F1, F2, F3, the EPIC factor model is used. The estimated variance of the unique factors (see equation /15/) is the average omitted eigenvalue:

$$\hat{\sigma}^2 = \frac{0.306+0.222+0.072}{3} = 0.2\,.$$

The EPIC loadings based on equation /12/ are shown in Table 2. They are computed from the $\mathbf{WU}^{1/2}$ principal components loadings (see also Table 2) according to the following manner:

$$\Lambda_{11}^{EPIC} = 0.601 = \hat{\sigma}\mathbf{W}_{11}\left(\frac{1}{\hat{\sigma}^2}\mathbf{U}_{11} - \mathbf{I}_{11}\right)^{\frac{1}{2}} = \hat{\sigma}\frac{\Lambda_{11}^{PCA}}{\sqrt{u_1}}\left(u_1\frac{1}{\hat{\sigma}^2} - 1\right)^{\frac{1}{2}} =$$

$$= \sqrt{0.2}\,\frac{0.6245}{\sqrt{2.709}}\sqrt{2.709\frac{1}{0.2} - 1}$$

and

$$\Lambda_{63}^{EPIC} = -0.4271 = \sqrt{\frac{0.2}{1.101}}\left(-0.4722\right)\sqrt{\frac{1.101}{0.2} - 1}\,.$$

Table 2

Initial, unrotated and rotated EPIC factor loadings

Variable	PCA factor loadings $\mathbf{WU}^{1/2}$			EPIC factor loadings			Orthosim solution			Direct quartimin solution		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
ATPLIAB	*0.624*	0.756	–0.036	*0.601*	0.707	–0.033	0.088	**0.921**	0.079	–0.013	***0.933***	0.008
CFLIAB	0.653	0.731	–0.029	0.629	0.684	–0.026	0.118	**0.916**	0.100	0.015	***0.924***	–0.009
CURRENT	0.705	–0.309	0.526	0.679	–0.289	0.475	**0.851**	0.107	0.185	**0.872**	0.012	–0.006
ACURRENT	0.688	–0.324	0.536	0.662	–0.303	0.485	**0.853**	0.085	0.175	**0.878**	–0.009	0.004
DEBT	–0.659	0.358	0.559	–0.634	0.335	0.505	–0.139	–0.088	**–0.862**	0.047	–0.008	***0.893***
EQUITYR	0.698	–0.391	*–0.472*	0.672	–0.366	*–0.427*	0.228	0.079	**0.842**	0.056	–0.008	***–0.854***

Since the first three eigenvalues of the correlation matrix account for a large portion of the total variance, it is clear, that the principal components and the EPIC loadings differ just to a slight extent. On the other hand, when some of the subsequent eigenvalues tend to be more important this tendency is not necessary.

The solutions from the orthogonal *orthosim* and oblique *direct quartimin* rotations are given in Table 2 and are almost identical.

According to the rotated loadings the following factors have been explored:

– F1: liability-based ‘profitability’,
– F2: current ratio-based ‘liquidity’,
– F3: long run ‘debtness’.

After oblique rotation the inter-factor correlations are not negligible because: Corr(F1,F2) = –0.233, Corr(F1,F3) = –0.408 and Corr(F2,F3) = –0.21, respectively. As a consequence, in the confirmative analysis step these correlations need to be estimated, increasing hence the number of free parameters.

The question arises at this stage is whether the hypothetical restriction $\sigma^2\mathbf{I}$ imposed on the covariance matrix of the error factors is acceptable or not. Decision on the model will be based on ‘goodness of fit’ measures, evaluated first retaining and then relaxing the restrictions. Such a method that provides tools for inference via the maximum likelihood theory is the *generalized weighted least squares*. However, when the sample does not come from a multivariate normal distribution, the asymptotically distribution free estimator is still available. A detailed overview of it is as follows.

ASYMPTOTICALLY DISTRIBUTION FREE ESTIMATORS

Based on a sample of size N let $\mathbf{S}$ denote the usual unbiased estimator of the population covariance matrix $\boldsymbol{\Sigma}_{(p,p)}$ whose elements are functions of a parameter vector $\boldsymbol{\theta}$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) .$$

The weighted least squares (WLS) quadratic form discrepancy function measures the discrepancy between the sample covariance matrix $\mathbf{S}$ and the reproduced covariance matrix $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ evaluated at an estimator (*Browne* [1974]):

$$F\left(\mathbf{s},\boldsymbol{\sigma}(\boldsymbol{\theta})\right) = \left(\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta})\right)^T \mathbf{W}^{-1}\left(\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta})\right) \rightarrow \min ,$$

where $\mathbf{s}$ and $\boldsymbol{\sigma}(\boldsymbol{\theta})$ are column vectors, formed from the $p^* = p\left(p+1\right)/2$ non-duplicative elements of $\mathbf{S}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, respectively and $\mathbf{W}$ is a positive definite weight matrix of order $\left(p^*, p^*\right)$. It is optimal to choose the weight matrix based on the covariance matrix of the sample covariances with typical element:

$$w_{jk,lt} = (N-1)\operatorname{cov}\left(s_{jk}, s_{lt}\right) = (N-1)\sigma_{jk,lt} = \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl} + \frac{N-1}{N}\kappa_{jklt} , \qquad /16/$$

where $\sigma_{jl} = \left[\boldsymbol{\Sigma}\right]_{jl}$ and

$$\kappa_{jklt} = \sigma_{jklt} - \left(\sigma_{jk}\sigma_{lt} + \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right)$$

is a fourth-order cumulant with the fourth-order moment

$$\sigma_{jklt} = E\left(x_j - \mu_j\right)\left(x_k - \mu_k\right)\left(x_l - \mu_l\right)\left(x_t - \mu_t\right).$$

Equation /16/ gives the weight matrix for Browne's Asymptotically Distribution Free (ADF) estimator (*Browne* [1984]). Letting N tend to infinity the ADF weight takes the form without specifying any particular distribution:

$$w_{jk,lt} = \sigma_{jklt} - \sigma_{jk}\sigma_{lt} \qquad /17/$$

with consistent (but not unbiased) estimators

$$m_{jklt} = \frac{1}{N}\sum_{i=1}^{N}\left(x_j - \overline{x}_j\right)\left(x_k - \overline{x}_k\right)\left(x_l - \overline{x}_l\right)\left(x_t - \overline{x}_t\right),$$

$$m_{jk} = \frac{1}{N}\sum_{i=1}^{N}\left(x_j - \overline{x}_j\right)\left(x_k - \overline{x}_k\right).$$

Let us consider the heterogeneous kurtosis theory (*Kano–Berkane–Bentler* [1990]) which defines a general class of multivariate distributions that allows marginal distributions to have heterogeneous kurtosis parameters. Let $\kappa_j^2 = \sigma_{jjjj} / 3\sigma_{jj}^2$ represent a measure of excess kurtosis of the *j*th indicator. Then the fourth-order moments have the structure

$$\sigma_{jklt} = \frac{\kappa_j + \kappa_k}{2}\frac{\kappa_l + \kappa_t}{2}\sigma_{jk}\sigma_{lt} + \frac{\kappa_j + \kappa_l}{2}\frac{\kappa_k + \kappa_t}{2}\sigma_{jl}\sigma_{kt} + \frac{\kappa_j + \kappa_t}{2}\frac{\kappa_k + \kappa_l}{2}\sigma_{jt}\sigma_{kl}.$$

Under the assumption that all marginal distribution of a multivariate distribution are symmetric and have the same relative kurtosis, the elliptical (homogeneous kurtosis) theory estimators and test statistics can be obtained. The common kurtosis parameter of a distribution from the elliptical class of distributions with multivariate density[4]

$$c\,|\mathbf{V}|^{-\frac{1}{2}}\, h\left[\left(\mathbf{x} - \boldsymbol{\mu}\right)^T \mathbf{V}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right)\right]$$

is

$$\kappa = \frac{\sigma_{jjjj}}{3\sigma_{jj}^2} - 1.$$

Then, the fourth-order moments are

$$\sigma_{jklt} = \left(\kappa + 1\right)\left(\sigma_{jk}\sigma_{lt} + \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right).$$

[4] Here c is a constant, h is a non-negative function and $\mathbf{V}$ is a positive definite matrix.

Letting again N tend to infinity, substitution into /17/ yields the weight

$$w_{jk,lt} = \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl} + \kappa\left(\sigma_{jk}\sigma_{lt} + \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right) = \left(\kappa+1\right)\left(\sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl}\right) + \kappa\sigma_{jk}\sigma_{lt}$$

Obviously, if $\kappa = 0$, then, multivariate normal distributions are considered and the typical element of the weight matrix takes the form

$$w_{jk,lt} = \sigma_{jl}\sigma_{kt} + \sigma_{jt}\sigma_{kl} \,.$$

Because the size of **W** in practice can be very large it is reasonable to perform computations based on an equivalent form of the discrepancy function. Namely, assuming elliptical distributions, the quadratic form discrepancy function takes the form:

$$F_E = \frac{1}{2(\kappa+1)} tr\left[\left((\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}\right)^2\right] - \frac{\kappa}{4(\kappa+1)^2 + 2p\kappa(\kappa+1)} tr^2\left((\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}\right)$$

which reduces to the normal theory discrepancy function when $\kappa = 0$, i.e. the distributions have no kurtosis:

$$F_N = \frac{1}{2} tr\left[\left((\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}\right)^2\right]. \qquad /18/$$

When **V=I**, one obtains the unweighted least squares estimator F_{ULS}, the substitution **V=S** yields the generalized least squares estimator F_{GLS} and an iteratively reweighted solution F_{IWLS} is obtained when $\mathbf{V} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is the reproduced covariance matrix generated by $\hat{\boldsymbol{\theta}}$ in each iterative step. Finally, asymptotically, F_{IWLS} leads to maximum likelihood estimate F_{ML} for exponential families of distributions.[5]

If $[\mathbf{V}]_{jk}$ is a consistent estimator of $\left[\boldsymbol{\Sigma}\right]_{jk} = \sigma_{jk}$ then $\hat{w}_{jk,lt}$ will be a consistent estimator of cov(**s**,**s**). Further, the *unbiased* estimator of $\hat{w}_{jk,lt}$ is

$$\hat{w}_{jk,lt} = \frac{N}{\left(N-2\right)\left(N-3\right)} \times$$

$$\times\left[\left(N-1\right)\left(m_{jklt} - m_{jk}m_{lt}\right) - \left(m_{jl}m_{kt} + m_{jt}m_{lk} - \frac{2}{N-1}m_{jk}m_{lt}\right)\right].$$

[5] The statistical distribution of the elements of a covariance matrix is not the same as that of a correlation matrix. This is obvious if you consider the diagonal elements of a covariance matrix, which are the variances of the variables. These are random variables – they vary from sample to sample. On the other hand, the diagonal elements of a correlation matrix are not random variables – they are always 1. The sampling distribution theory employed for the case of a covariance matrix is not applicable to a correlation matrix, except in special circumstances. It must be emphasized that it is possible (indeed likely) to get some incorrect results if we analyze a correlation matrix as if it were a covariance matrix. This has been described in the literature (see, for example, *Cudeck* [1989]). In order to analyse of the correlation matrix of the input data correctly, computations are based on the constrained estimation theory developed by *Browne* [1982]. As a result, we give the correct standard errors, estimates, and test statistics when a correlation matrix is analyzed directly.

If **W** consists of these unbiased elements, it may not be positive definite, but it would be unlikely the case when N is substantially larger than p^*.

As it is apparent, the measures of multivariate kurtosis play a key role from the multivariate normality point of view.

THE MEASURES OF KURTOSIS

The statistics described subsequently allow us to examine whether the assumptions of *multivariate* normality have been violated. The consistent estimator of the common relative multivariate kurtosis parameter κ is the *rescaled* Mardia's sample measure:

$$\left(\hat{\kappa}+1\right)=\sum_{i=1}^{N}\frac{\left[\left(\mathbf{x}_i-\overline{\mathbf{x}}\right)^T\mathbf{S}^{-1}\left(\mathbf{x}_i-\overline{\mathbf{x}}\right)\right]^2}{Np\left(p+2\right)}.$$

This measure should be close to 1 if the distribution is multivariate normal.

If the sample comes from a multivariate normal distribution, the Mardia-coefficient of multivariate kurtosis defined as

$$MK=\hat{\kappa}p\left(p+2\right)$$

should be close to zero.[6] Further, the normalized multivariate kurtosis

$$\hat{\kappa}_0=\frac{MK}{\sqrt{8p\left(p+2\right)/N}}$$

has a distribution that is approximately standard normal at large samples.

The elliptical distribution family includes the multivariate normal distribution as a special case. As mentioned in this distribution family all variables have a common kurtosis parameter κ. This parameter can be used to rescale the *Chi*-square statistic if the assumption of an elliptical distribution is valid. The Mardia-based *kappa*

$$\hat{\kappa}_1=\frac{MK}{p\left(p+2\right)}$$

is an estimate of *kappa* obtained by rescaling the Mardia's coefficient of multivariate kurtosis. This number should be close to zero if the population distribution is multivariate normal.

Distribution theory provides a lower bound for *kappa*. It must never be less than $-6/(p+2)$, where p is the number of variables. The adjusted mean scaled univariate kurto-

[6] The expected value and variance of $\left(\hat{\kappa}+1\right)p\left(p+2\right)$ are $\left(N-1\right)p\left(p+2\right)/\left(N+1\right)$ and $8p\left(p+2\right)/N$ respectively.

sis is an alternate estimate of *kappa*, which takes into account this requirement and is obtained simply as the average univariate kurtosis:

$$\hat{\kappa}_2 = \frac{1}{3p}\sum_{j=1}^{p} \max\left\{\left(\frac{m_{jjjj}}{m_{jj}^2} - 3\right), \frac{-6}{p+2}\right\},$$

where

$$\frac{m_{jjjj}}{m_{jj}^2} - 3$$

is the rescaled (i.e. uncorrected), biased estimate of *univariate* kurtosis for variable x_j. The asymptotic variance of this univariate measure is $24/N$, which is used to standardize the uncorrected kurtosis in order to produce the '*normalized*' kurtosis.

The $\hat{\kappa}_2$ estimate averages the scaled univariate kurtosis, but adjusts each one that falls below the bound to be at the lower bound point. This coefficient should be close to zero if the distribution is multivariate normal.

Table 3

Measures of Multivariate Kurtosis

Measure	Value
Mardia Coefficient of Multivariate Kurtosis	403.085
Normalized Multivariate Kurtosis	946.437
Mardia-Based Kappa	8.398
Mean Scaled Univariate Kurtosis	9.580
Adjusted Mean Scaled Univariate Kurtosis	9.580
Relative Multivariate Kurtosis	9.398

Table 4

Univariate measures of skewness and kurtosis

Variable	Measures of skewness			Measures of kurtosis		
	Skewness	Corrected	Normalized	Kurtosis	Corrected	Normalized
ATPLIAB	0.490	0.490	9.201	46.798	46.912	439.524
CFLIAB	2.176	2.178	40.876	50.111	50.232	470.636
CURRENT	4.056	4.059	76.188	22.257	22.313	209.040
ACURRENT	3.311	3.314	62.203	13.932	13.968	130.850
DEBT	4.256	4.259	79.940	27.112	27.179	254.632
EQUITYR	–1.387	–1.388	–26.051	11.809	11.840	110.913

Considering our 6 financial microeconomic indicators, the computed values of the measures discussed are presented in Table 3 and Table 4. Results show that the requirement of zero kurtosis is violated. Nevertheless, the homogeneous kurtosis hypothesis about a common non-zero kurtosis parameter could still be valid. But as we can see from Table 4 the univariate uncorrected kurtosis measures do not justify accepting the case of a common kurtosis parameter.

Finally, if the univariate kurtosis and skewness measures separately reject the assumption of univariate normality, the hypothesis about the multivariate normality must also be rejected as a consequence. Therefore, the *corrected univariate kurtosis* and *skewness* measures are also useful unbiased estimates for investigation of the assumption of normality.[7] They are, respectively:

Corrected univariate kurtosis

$$b^*_{2(j)} = \frac{(N+1)N^2}{(N-1)(N-2)(N-3)}\left[\frac{m_{jjjj}}{(m_{jj})^2}\right] - \frac{3(N-1)^2}{(N-2)(N-3)}\,.$$

Corrected univariate skewness

$$b^*_{1(j)} = \frac{N^2}{(N-1)(N-2)}\left[\frac{m_{jjj}}{(m_{jj})^{3/2}}\right].$$

The asymptotic variance of this latter measure is $6/N$, which is used to standardize the uncorrected skewness to produce the '*normalized*' skewness.

As a consequence of the kurtosis and skewness measures, we prefer the ADF estimator under arbitrary distribution as long as it produces interpretable results.

Estimates of free parameters and their inference statistics (standard error, T-value) are given in Table 5 based on both IWLS and ADF estimators considering both homogeneous and heterogeneous error-variance models. The corresponding converged values of the discrepancy function are also included. (Each of the four model-estimation converged within 10 iteration steps.) The type of the free parameters is indicated by the following scheme in the first column of the table: (.) contains latent variable, [.] includes measured indicator, the numbered -#-> arrow represents directed relationship and the numbered -#- wire represents undirected relationship (i.e. variance, covariance). Finally, the numbered name of an error factor is DELTA#.

As we can see, only the 'ADF, Homogeneous' T-values for parameters #11 and #13 are not significant with P-values 0.336, 0.454, respectively. All other P-values are practically zeros. Obviously, in the case of the ADF estimator (because of the distributional knowledge omitted) the estimated standard errors are higher than those computed by IWLS.

Based on the discrepancy function the results from the ADF method seem to be preferred. Contrary, based on the Root Mean Square (RMS) standardized residual[8], the IWLS results exhibit a better fit. The former results are based on the assumption of multivariate normality, while the latter is not, producing greater standard errors. Nevertheless our main purpose is to compare the homogeneous model with the heterogeneous one.

[7] One can find the uncorrected counterparts closed in the [.] bracket.

[8] Residual is divided by its standard error.

Table 5

Model characteristics from IWLS and ADF estimators

Free parameters	Iteratively reweighted least squares estimator						Asymptotically distribution free estimator					
	Heterogeneous variances			Homogeneous variances			Heterogeneous variances			Homogeneous variances		
	Estimate	St.Error	T	Estimate	St.Error	T	Estimate	St.Error	T	Estimate	St.Error	T
(F1)-1->[ATPLIAB]	0.664	0.012	56.444	0.634	0.014	44.706	0.465	0.060	7.773	0.569	0.064	8.885
(F1)-2->[CFLIAB]	0.756	0.012	65.054	0.673	0.015	46.140	0.530	0.066	8.066	0.585	0.066	8.849
(F2)-3->[CURRENT]	0.837	0.022	37.855	0.765	0.015	49.839	0.556	0.045	12.334	0.859	0.049	17.515
(F2)-4->[ACURRENT]	1.005	0.030	33.997	1.159	0.020	56.991	0.990	0.068	14.671	0.875	0.050	17.528
(F3)-5->[DEBT]	-0.590	0.019	-31.016	-0.600	0.013	-44.7	-0.763	0.044	-17.28	-0.298	0.020	-15.27
(F3)-6->[EQUITYR]	1.581	0.043	36.631	1.589	0.026	60.476	1.203	0.072	16.782	1.297	0.071	18.362
(DELTA1)-7-(DELTA1)	0.072	0.002	32.527	***0.189***	0.003	56.338	0.031	0.010	**3.191**	***0.018***	0.005	3.667
(DELTA2)-8-(DELTA2)	0.000	0.000	-	***0.189***	0.003	56.338	0.000	0.000	-	***0.018***	0.005	3.667
(DELTA3)-9-(DELTA3)	0.137	0.027	5.013	***0.189***	0.003	56.338	0.201	0.033	6.031	***0.018***	0.005	3.667
(DELTA4)-10-(DELTA4)	0.555	0.042	13.081	***0.189***	0.003	56.338	0.273	0.106	**2.579**	***0.018***	0.005	3.667
(DELTA5)-11-(DELTA5)	0.248	0.017	14.668	***0.189***	0.003	56.338	0.030	0.031	**0.962**	***0.018***	0.005	3.667
(DELTA6)-12-(DELTA6)	0.220	0.108	2.028	***0.189***	0.003	56.338	0.878	0.128	6.882	***0.018***	0.005	3.667
(F2)-13-(F1)	0.244	0.022	11.028	0.238	0.024	9.927	0.047	0.063	**0.749**	0.212	0.072	2.944
(F3)-14-(F1)	0.193	0.022	8.778	0.207	0.024	8.651	0.220	0.035	6.300	0.250	0.053	4.701
(F3)-15-(F2)	0.427	0.022	19.541	0.399	0.020	19.842	0.321	0.026	12.540	0.431	0.034	12.807
Discrepancy Function	0.0441			0.973			0.0301			0.0897		
degree of freedom	6			11			6			11		
RMS Stand. Residual	0.0158			0.0625			0.243			0.285		
Chi-Square Statistic	93.3284			2057.94			63.7372			189.81		
Goodness of fit indices	*Confidence intervals at 90 percent level*											
Noncentrality based indices	LB	PE	UB	LB	PE	UB	LB	PE	UB	LB	PE	UB
Population Noncentrality Index	0.026	0.038	0.055	0.500	0.551	0.606	0.017	0.027	0.041	0.065	0.085	0.107
Steiger-Lind RMSEA Index	0.066	0.080	0.095	0.213	0.224	0.235	0.053	0.067	0.083	0.077	0.088	0.099
McDonald Noncentrality Index	0.973	0.981	0.987	0.738	0.759	0.779	0.980	0.986	0.992	0.948	0.959	0.968
Population Gamma Index	0.982	0.987	0.991	0.832	0.845	0.857						
Adjusted Population Gamma Index	0.938	0.956	0.970	0.679	0.704	0.727						
Other fit indices												
Joreskog GFI		0.986			0.844			0.832			0.499	
Joreskog AGFI		0.952			0.701			0.412			0.044	
Akaike Information Criterion		0.058			0.982			0.044			0.099	
Schwarz's Bayesian Criterion		0.098			1.009			0.084			0.126	
Browne-Cudeck Cross Validation		0.058			0.982			0.044			0.099	
Null Model Chi-Square		7989.2			7989.2			291.8			291.8	
Null Model *df*		15			15			15			15	
Bentler-Bonett Normed Fit Index		0.988			0.742							
Bentler-Bonett Non-Normed Fit Index		0.973			0.650							
Bentler Comparative Fit Index		0.989			0.743							
James-Mulaik-Brett Parsimonious Fit Index		0.395			0.544							
Bollen's Rho		0.971			0.649							
Bollen's Delta		0.989			0.743							

Note: Where a baseline model is involved, it is assumed to be the null model, defined as a model without any common factors.

var(F1)=var(F2)=var(F3)=1 and the error factors (Delta1-Delta6) are uncorrelated.

Once a minimized converged value of the discrepancy function has been reached and selected as the best one, the subsequent evaluation of its goodness of fit is necessary. For this purpose a wide selection of fit-indices is available. Some of them are hypothesis theory-based, others are heuristic. On the other hand, we can distinguish noncentrality-based goodness of fit indices and other indices including incremental type indices as well. In the following section we discuss those employed in this paper.

NONCENTRALITY-BASED GOODNESS-OF-FIT INDICES

Let us consider the null hypothesis that the restricted model $\mathbf{\Sigma}(\boldsymbol{\theta})$ holds for the population covariance matrix $\mathbf{\Sigma}$, against the alternative that it does not hold:

$$H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta}),$$

$$H_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}(\boldsymbol{\theta}).$$

In other words, the H_1 hypothesis states that a significant improvement is expected in the discrepancy between the restricted and the unrestricted models due to a simple switch from $\mathbf{\Sigma}(\boldsymbol{\theta})$ to $\mathbf{\Sigma}$. Then, the discrepancy between the *true* and the hypothesized model is

$$F\left(\boldsymbol{\sigma},\boldsymbol{\sigma}(\boldsymbol{\theta})\right) = \left(\boldsymbol{\sigma}-\boldsymbol{\sigma}(\boldsymbol{\theta})\right)^T \mathbf{W}^{-1}\left(\boldsymbol{\sigma}-\boldsymbol{\sigma}(\boldsymbol{\theta})\right) \to \min$$

which could be minimized with respect to the parameter vector $\boldsymbol{\theta}$. Let $F\left(\boldsymbol{\sigma},\boldsymbol{\sigma}\left(\boldsymbol{\theta}^*\right)\right)$ denote the minimized value at some $\boldsymbol{\theta}^*$. Then, asymptotically, $\chi^2 = (N-1)F\left(\mathbf{s},\boldsymbol{\sigma}(\boldsymbol{\theta})\right)$ is distributed as a *noncentral* Chi-square with

$$df = \frac{p(p+1)}{2} - q$$

degrees of freedom and *noncentrality parameter*

$$\tau = (N-1)F\left(\boldsymbol{\sigma},\boldsymbol{\sigma}\left(\boldsymbol{\theta}^*\right)\right)$$

or

$$\frac{\tau}{N-1} = F\left(\boldsymbol{\sigma},\boldsymbol{\sigma}\left(\boldsymbol{\theta}^*\right)\right)$$

rescaled noncentrality parameter, where q is the number of parameters to be estimated for the model. Obviously, when the model holds, $\tau = 0$ and χ^2 is distributed as a *central* Chi-square with *df* degrees of freedom.

Hence, the size of τ can be considered as a *population* measure of model *misspecification*, with larger values of τ indicating greater misspecification. As it follows from the probability theory, the expected value of the noncentral Chi-square statistic is

$$E\left(\chi^2\right) = E\left\{(N-1)F\left(\mathbf{s}, \boldsymbol{\sigma}(\boldsymbol{\theta})\right)\right\} = df + \tau .$$

Hence, based on only one observation for χ^2, the estimated value of the noncentrality parameter is

$$\hat{\tau} = NCP = \chi^2 - df ,$$

where

$$\chi^2 = (N-1)F\left(\mathbf{s}, \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}\right)\right)$$

is the estimated measure of distance between the *currently* investigated model which is the *target* of our hypothesis and the saturated model with $p(p+1)/2$ free parameters, say, **s**. Therefore, the discrepancy function (named also fitting function) is calculated as

$$F = \frac{\chi^2}{N-1} .$$

Note, that *NCP* can be negative when the estimated Chi-square is less than the *df*. Dividing the noncentrality parameter by (N–1) yields the population noncentrality index *PNI* which is a measure of population *badness-of-fit* and depends only on the model, and the method of estimation:

$$PNI = \max\left\{\frac{\chi^2 - df}{N-1}, 0\right\} .$$

The population noncentrality index *PNI* is an unbiased estimate of the rescaled noncentrality parameter and is relatively unaffected by the sample size. However, *PNI* fails to compensate for model complexity. In general, for a given **S**, the more complex the model the better its fit. A method for assessing population fit which fails to compensate for this will inevitably lead to choosing the most complex models, even when simpler models fit the data nearly as well. Because *PNI* fails to compensate for the size or complexity of a model, it has limited utility as a device for comparing models.

The adjusted root mean square error index, first proposed by *Steiger* and *Lind* [1980], takes a relatively simplistic approach to solving these problems. Since model complexity is reflected directly in the number of free parameters, and inversely in the number of degrees of freedom, the *PNI* is divided by degrees of freedom, then the square root is taken to return the index to the same metric as the original standardized parameters.

$$RMSEA = \sqrt{\frac{PNI}{df}} .$$

The RMSEA index can be thought of roughly as a root mean square standardized residual. Values above .10 indicate an inadequate fit, values below .05 a very good fit. Point estimates below .01 indicate an outstanding fit. The rule of thumb is that, for ‘close fit’, RMSEA should be less than $c = .05$ yields a rule that

$$\frac{\chi^2}{df} < 1 + (N-1)c^2 = 1 + \frac{N-1}{400} .$$

With this criterion, if $N = 401$, the ratio of the *Chi-square* to its degrees of freedom should be less than 2. Note that this rule implies a *less stringent* criterion for the ratio χ^2/df as sample size increases.

Rules of thumb that cite a single value for a critical ratio of χ^2/df ignore the point that the *Chi-square* statistic has an expected value that is a function of degrees of freedom, population badness of fit, and *N.* Hence, for a fixed level of population badness of fit, the expected value of the *Chi-square* statistic will increase as sample size increases.

McDonald [1989] proposed an index of noncentrality that represents one approach to transforming the population noncentrality index *PNI* into the range from 0 to 1. The index does not compensate for model parsimony, and the rationale for the exponential transformation it uses is primarily pragmatic. The index may be expressed as

$$MDNI = e^{-\frac{1}{2}PNI} .$$

Good fit is indicated by values above 0.95. Similarly, the *scaled likelihood ratio criterion* is

$$LHR = e^{-\frac{1}{2}F} .$$

Further, the weighted population coefficient of determination can also be defined as

$$\Gamma = 1 - \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}(\boldsymbol{\theta}))^T \mathbf{W}^{-1} (\boldsymbol{\sigma} - \boldsymbol{\sigma}(\boldsymbol{\theta}))}{\boldsymbol{\sigma}^T \mathbf{W}^{-1} \boldsymbol{\sigma}} ,$$

where $\mathbf{W}$ is a positive definite weight matrix. Under arbitrary weighted least squares estimation, the *population gamma index* of *Tanaka* and *Huba* [1985] is given as a general form for the *sample* fit index for covariance structure models. It assumes the covariance structure model has been fit by minimizing the WLS discrepancy function. Then, the index is

$$\gamma = 1 - \frac{\left(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right)^T \mathbf{W}^{-1} \left(\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\right)}{\mathbf{s}^T \mathbf{W}^{-1} \mathbf{s}}.$$

When the distributions have no kurtosis ($\kappa = 0$) based on /18/ we can write γ as the parametric form of the *Jöreskog–Sörbom* [1984] index of fit:

$$JSI_V = 1 - \frac{\frac{1}{2} tr\left(\left(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)\mathbf{V}^{-1}\right)^2}{\frac{1}{2} tr\left(\mathbf{S}\mathbf{V}^{-1}\right)^2}.$$

If **V**=**I**, or **V**=**S**, one obtains the Jöreskog–Sörbom (JS) index for the *ULS* and *GLS* estimators, respectively. Specially, using *IWLS*, i.e. $\mathbf{V} = \hat{\boldsymbol{\Sigma}}$, gives asymptotically the *JSI* index for the maximum likelihood (ML) estimation with

$$F_{IWLS} = \frac{1}{2} tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1} - \mathbf{I}\right)^2 = \frac{1}{2}\left(tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right)^2 - 2tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) + p\right).$$

Hence,

$$\gamma_{IWLS} = 1 - \frac{tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1} - \mathbf{I}\right)^2}{tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right)^2} = \frac{2tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) - p}{2F_{IWLS} + 2tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) - p}.$$

In addition, when the

$$tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right) = p$$

equation holds, the $\gamma_{IWLS,ML}$ index reduces to the classic JS goodness of fit index:

$$GFI = \frac{p}{tr\left(\mathbf{S}\hat{\boldsymbol{\Sigma}}^{-1}\right)^2} = \frac{p}{2F_{IWLS} + p}.$$

As a consequence, *GFI* can be thought of as the sample equivalent of the index defined in the population as

$$\Gamma_1 = \frac{p}{tr\left(\boldsymbol{\Sigma}\left[\boldsymbol{\Sigma}(\boldsymbol{\theta})\right]^{-1}\right)^2} = \frac{p}{2F\left(\boldsymbol{\sigma}, \boldsymbol{\sigma}(\boldsymbol{\theta})\right) + p} = \frac{p}{2\frac{\tau}{N-1} + p}.$$

Any consistent estimate of τ will give a consistent estimate for Γ_1. This index like *PNI*, fails to compensate for the effect of model complexity. Consider a sequence of *nested* models, where the models with more degrees of freedom are special cases of those with less degrees of freedom. For such a nested sequence of models, the more complex models (i.e. those with more free parameters and less degrees of freedom) will always have Γ_1 coefficients as low or lower than those which are less complex.

The adjusted population *gamma* index Γ_2 attempts to compensate for this tendency:

$$\Gamma_2 = 1 - \frac{p(p+1)}{2 \cdot df}(1 - \Gamma_1)$$

and its sample counterpart is

$$AGFI = 1 - \frac{p(p+1)}{2 \cdot df}(1 - GFI).$$

Values of the Joreskog GFI above .95 indicate good fit. This index is a *negatively biased* estimate of the population GFI, so it tends to produce a slightly pessimistic view of the quality of population fit. We give this index primarily because of its historical popularity.

The Population Gamma index is a superior realization of the same rationale. The values of the Joreskog AGFI above .95 also indicate good fit. This index is, like the GFI, a negatively biased estimate of its population equivalent. As with the GFI, the Adjusted Population Gamma Index is a superior realization of the same rationale.

At this stage we have arrived at an important conclusion that the lower and upper bounds of an α level confidence interval of the Chi-square statistic can be inserted into any goodness of fit measure that involves the Chi-square statistic. Consistent estimates and confidence intervals for Γ_1 may thus be converted into corresponding quantities for Γ_2.

OTHER INDICES OF FIT

Rescaled Akaike Information Criterion

In a number of situations the user must decide among a number of competing *nested* models of different dimensions. This criterion is useful primarily for deciding which of several nested models provides the best approximation to the data. The most typical example is the choice of the number of factors in common factor analysis. *Akaike* ([1973], [1974], [1983]) proposed a criterion for selecting the dimension of a model. *Steiger* and *Lind* [1980] presented an extensive Monte Carlo study of the performance of the Akaike criterion. Here the criterion is rescaled (without affecting the decisions it indicates) so that it remained more stable across differing sample sizes. The rescaled Akaike criterion (modified by *Cudeck* and *Brown* [1983]) is as follows.

Let $F_{ML,k}$ be the maximum likelihood discrepancy function and q_k be the number of free parameters for the model M_k. Let N be the sample size.

When trying to decide between several nested models, choose the one with the smallest Akaike criterion:

$$AC_k = F_{ML,k} + \frac{2q_k}{N-1}.$$

Schwarz's Bayesian Criterion

This criterion (*Schwarz* [1978] also modified by *Cudeck* and *Brown* [1983]) is similar in use to Akaike's index, selecting, in a sequence of nested models, the model for which

$$SC_k = F_{ML,k} + \frac{q_k \ln(N)}{N-1}$$

is a minimum.

Browne–Cudeck Single Sample Cross-Validation Index

Browne and *Cudeck* [1990] proposed a single sample cross-validation index as a follow-up to their earlier (*Cudeck–Browne* [1983]) paper on cross-validation. Cudeck and Browne had proposed a cross-validation index which, for model M_t in a set of competing models is of the form $F_{ML}(\mathbf{S}_{cv},\mathbf{\Sigma}_t(\boldsymbol{\theta}))$. In this case, F is the maximum likelihood discrepancy function, $\mathbf{S}_{cv}$ is the covariance matrix calculated on a cross-validation sample, and $\mathbf{\Sigma}_t(\boldsymbol{\theta})$ the reproduced covariance matrix obtained by fitting model M_t to the original calibration sample. In general, better models will have smaller cross-validation indices.

The drawback of the original procedure is that it requires two samples, i.e. the calibration sample for fitting the models, and the cross-validation sample. The new measure estimates the original cross-validation index from a single sample.

The measure is

$$C_t = F_{ML}(\mathbf{S}_{cv}, \mathbf{\Sigma}_t(\boldsymbol{\theta})) + \frac{2q_t}{N-p-2}.$$

Null Model Chi-square and df

This is the *Chi-square* goodness-of-fit statistic, and the associated degrees of freedom, for the hypothesis that the population covariances are all zero. Under the assumption of multivariate normality, this hypothesis can only be true if the variables are all independent. The 'Independence Model' is used as the 'Null Model' in several comparative fit indices.

Bentler–Bonett Type Fit Indices

One of the most historically important and original fit indices, the *Bentler–Bonett* [1980] *normed index* measures the *relative decrease* in the discrepancy function caused by switching from a 'Baseline Model' (typically the null model) to a more complex model. It is defined as:

$$0 \le NFI_{t/b} = \frac{F_b - F_t}{F_b} = \frac{\chi_b^2 - \chi_t^2}{\chi_b^2} \le 1\,,$$

where F_b is the discrepancy function for the 'baseline model', F_t is the discrepancy function for the *target* (typically the current) model. This index approaches 1 in value as fit becomes perfect. However, it does not compensate for model parsimony.

The comparative *Bentler–Bonett* [1980] non-normed fit index takes into account model parsimony. It is defined as

$$NNFI_{t/b} = \frac{\chi_b^2 - \dfrac{df_b}{df_t}\chi_t^2}{\chi_b^2 - df_b} = \frac{\dfrac{\chi_b^2}{df_b} - \dfrac{\chi_t^2}{df_t}}{\dfrac{\chi_b^2}{df_b} - 1} = \frac{\dfrac{F_b}{df_b} - \dfrac{F_t}{df_t}}{\dfrac{F_b}{df_b} - \dfrac{1}{N-1}}$$

or it can be written as

$$NNFI_{t/b} = \frac{\chi_b^2 - df_b - \left(\dfrac{df_b}{df_t}\chi_t^2 - df_b\right)}{\chi_b^2 - df_b} = 1 - \frac{df_b}{df_t}\frac{NCP_t}{NCP_b} = 1 - p_{b/t}\frac{PNI_t}{PNI_b}\,,$$

where $p_{b/t}$ is the so-called parsimony coefficient.

Bentler Comparative Fit Index

The comparative index (*Bentler* [1990]) estimates the relative decrease in the population noncentrality obtained by changing from the 'Baseline Model' to the t model. The index may be computed as:

$$BCFI_{t/b} = 1 - \frac{NCP_t}{NCP_b}\,,$$

where NCP_t is the estimated non-centrality parameter for the target model and NCP_b is that for the base line model.

James–Mulaik–Brett Parsimonious Fit Index

This index was one of the earliest (along with the Steiger–Lind index) to compensate for model parsimony. Basically, it operates by rescaling the Bentler–Bonett Normed fit index to compensate for model parsimony. The formula for the index is:

$$PI = \frac{df_t}{df_b} NFI_t ,$$

where *NFI* denotes the Bentler–Bonett normed fit index.

Bollen's Rho

This comparative fit index computes the relative reduction in the discrepancy function per degree of freedom when moving from the 'Baseline Model' to the *t* model. It is computed as

$$\rho_{t/b} = \frac{\frac{F_b}{df_b} - \frac{F_t}{df_t}}{\frac{F_b}{df_b}} = 1 - \frac{df_b}{df_t} \frac{F_t}{F_b} .$$

Comparing with *NNFI*, we see that, for even moderate *N,* there is virtually no difference between Bollen's *Rho* and the Bentler–Bonett Non-normed fit index.

Bollen's Delta

This index is also similar in form to the Bentler–Bonett index, but rewards simpler models (those with higher degrees of freedom). It is computed as:

$$\Delta_{t/b} = \frac{F_b - F_t}{F_b - \frac{df_t}{N-1}} .$$

EVALUATION OF MODEL FIT

Based on the results given in Table 5 and discussion of the goodness of fit measures presented earlier the following statements can be established:

Both the IWLS and the ADF estimates exhibit an outstanding goodness of fit. The model with more parameters, of course, performs a better fit. Except the *population gamma* and *Jöreskog–Sörbom* indices, the ADF estimator seems to be preferred against the IWLS estimator.

As a brief summary measure the pseudo R-square defined as

$$R^2 = 1 - \frac{\chi^2}{\chi^2_{null\text{-}model}}$$

are 98.83 and 74.24 percent for the IWLS homogeneous and heterogeneous models, respectively. Hence, switching from one model to the other seems to cause a considerable difference. These measures for the ADF estimators are 78.16 and 34.95 percent, respectively.

The null-model goodness of fit Chi-square value (the distance from the saturated model with $p(p+1)/2$ parameters) is substantially smaller in the ADF case. This null model Chi-square means the moving range given for the Chi-square to get closer to the saturated model (to the sample points). The corresponding R-square values thus must be interpreted on this shorter range of improvement for the ADF case. The null model chi-square estimated by the ADF method is 291.781, maybe underestimated to a great extent. Therefore additional goodness of fit measures based on this distance are not published in Table 5.

Only the James-Mulaik-Brett Parsimonious Index prefers the improvement in the degree of freedom versus worsen in the discrepancy function.

Considering any model of our interest, because of the large sample size the model chi-square statistic is relatively large as compared with the small degree of freedom resulted in from the (6,6) order of the sample covariance matrix. As a consequence, in spite of the goodness of fit measures, the chi-square test suggests to reject each of our models at any significance level.

Even a moderately large sample size is given, as it is the present case, it is not possible to choose between the competing homogeneous and heterogeneous models based on chi-square-difference test statistic. Namely, the difference between these chi-square statistics 2057.94–93.33=1964.61 and 189.81–63.74=126.07 are still significant at (11–6=5) degrees of freedom no matter whether the IWLS or the ADF results are considered.

Nevertheless, hypothesis testing can be avoided if we use some so-called incremental goodness of fit index such as the Bentler-type indices. Normed indices that fall into the interval of (0,1) are preferred because of their easy interpretation. In our investigation the Bentler-type incremental indices are as follows

$$NFI_{\text{homogeneous/heterogeneous}} \mid IWLS = \frac{\chi_o^2 - \chi_e^2}{\chi_o^2} = \frac{2057.94 - 93.3284}{2057.94} = 0.9546\,,$$

$$NNFI_{\text{homogeneous/heterogeneous}} \mid IWLS = \frac{\dfrac{\chi_o^2}{df_o} - \dfrac{\chi_e^2}{df_e}}{\dfrac{\chi_o^2}{df_o} - 1} = \frac{\dfrac{2057.94}{11} - \dfrac{93.3284}{6}}{\dfrac{2057.94}{11} - 1} = 0.9218\,,$$

$$BCFI_{\text{homogeneous/heterogeneous}} \mid IWLS = 1 - \frac{NCP_o}{NCP_e} = 1 - \frac{93.3284 - 6}{2057.94 - 11} = 0.9573\,,$$

and in the case of the ADF estimation they take the values as follows

$$NFI_{\text{homogeneous/heterogeneous}} \mid ADF = \frac{\chi_o^2 - \chi_e^2}{\chi_o^2} = \frac{189.81 - 63.7372}{189.81} = 0.6642\,,$$

$$NNFI_{\text{homogeneous/heterogeneous}} \mid ADF = \frac{\dfrac{\chi_o^2}{df_o} - \dfrac{\chi_e^2}{df_e}}{\dfrac{\chi_o^2}{df_o} - 1} = \frac{\dfrac{189.81}{11} - \dfrac{63.7372}{6}}{\dfrac{189.81}{11} - 1} = 0.408\,,$$

$$BCFI_{\text{homogeneous/heterogeneous}} \mid ADF = 1 - \frac{NCP_o}{NCP_e} = 1 - \frac{63.7372 - 6}{189.81 - 11} = 0.6771\,.$$

The smaller the value of an incremental index, the closer the models of interest are to one another. Hence, the results from IWLS may suggest that the assumption of the homogeneous variances is acceptable but, in contrary, based on the ADF method we may conclude that the error factor variances are heterogeneous. Recall here that the use of IWLS is questionable because of rejecting the normality and the zero kurtosis assumption.

Finally, the question of our interest is whether the magnitude of improvement in the discrepancy function due to involving *correlated error* (unique) factors is significant or not. Enabling cov(DELTA3, DELTA5) to be freely estimated provided *heterogeneous* error variances by the ADF method, the discrepancy function reduces to 0.0105 and the Chi-Square Statistic becomes 22.3062 with 5 degrees of freedom and P-value of 0.000458. Parameter estimates are given in Table 6. As compared with the corresponding Chi-Square Statistic 63.7372, the difference is (63.7372–22.3062)=41.431 with 1 degree of freedom which is significant at any level.

Table 6

Parameter estimation by ADF with heterogeneous but correlated error factors

Free parameters	Estimation	Standard error	T-value	Prob.
(F1)-1->[ATPLIAB]	0.635	0.058	10.894	0.000
(F1)-2->[CFLIAB]	0.668	0.061	11.001	0.000
(F2)-3->[CURRENT]	0.818	0.052	15.869	0.000
(F2)-4->[ACURRENT]	0.979	0.050	19.647	0.000
(F3)-5->[DEBT]	–0.633	0.040	–15.822	0.000
(F3)-6->[EQUITYR]	1.463	0.077	18.951	0.000
(DELTA1)-7-(DELTA1)	0.018	0.009	1.975	0.048
(DELTA2)-8-(DELTA2)	–0.000	0.000		
(DELTA3)-9-(DELTA3)	0.123	0.028	4.373	0.000
(DELTA4)-10-(DELTA4)	0.557	0.098	5.687	0.000
(DELTA5)-11-(DELTA5)	0.191	0.032	6.008	0.000
(DELTA6)-12-(DELTA6)	0.567	0.141	4.027	0.000
(F2)-13-(F1)	0.284	0.061	4.648	0.000
(F3)-14-(F1)	0.241	0.032	7.557	0.000
(F3)-15-(F2)	0.465	0.030	15.671	0.000
(DELTA3)-16-(DELTA5)	0.061	0.008	7.741	0.000

Table 7

Goodness of fit confidence intervals for IWLS and ADF measures with heterogeneous but correlated error factors

Noncentrality measures	90 percent IWLS confidence interval			90 percent ADF confidence interval		
	Lower	Point	Upper	Lower	Point	Upper
Population Noncentrality Index	0.007	0.015	0.025	0.003	0.008	0.017
Steiger-Lind RMSEA Index	0.038	0.054	0.071	0.024	0.040	0.058
McDonald Noncentrality Index	0.987	0.993	0.996	0.992	0.996	0.999
Population Gamma Index	0.992	0.995	0.998			
Adjusted Population Gamma Index	0.965	0.980	0.990			

Finally, we conclude that the $\sigma^2\mathbf{I}$ restriction imposed in our factor model is strongly questionable.

REFERENCES

AB MOOIJAART – BENTLER, P. M. [1985]: The weight matrix in asymptotic distribution-free methods. *British Journal of Mathematical and Statistical Psychology*. No. 38. p. 190–196.

AKAIKE, H. [1973]: Information theory and an extension of the maximum likelihood principle. In: *Petrov, B. N. – Csaki, F.* (Eds.) *Second International Symposium on Information Theory*. Budapest. Akadémiai Kiadó.

AKAIKE, H. [1974]: A new look at the statistical model identification. IEEE transactions on automatic control. AC-19, p. 716–723.

AKAIKE, H. [1983]: Information measures and model selection. *Bulletin of the International Statistical Institute: Proceedings of the 44th Session*. Vol. 1. p. 277–290.

ANDERSON, T. W. [1984]: Estimating linear statistical relationships. *The Annals of Statistics*. Vol. 12. No. 1. p. 1–45.

BENTLER, P. M. [1977]: Factor simplicity index and transformations. *Psychometrika*. Vol. 42. No. 2. p. 277–295.

BENTLER, P. M. [1990]: Comparative fit Indexes in structural models. *Psychological Bulletin*. Vol. 107. No. 2. p. 238–246.

BENTLER, P. M. [1983]: Some contributions to efficient statistics in structural models: sand estimation of moment structures. *Psychometrica*. Vol. 48. No. 4. p. 493–517.

BENTLER, P. M. – AB MOOIJAART [1989]: Choice of structural model via parsimony: a rationale based on precision. *Psychological Bulletin*. Vol. 106. No. 2. p. 315–317.

BENTLER, P. M.– BONNET, D. G. [1980]: Significance tests and goodness of fit in the analysis of covariance structures.

BOLLEN, K. A. [1989]: Structural equations with latent variables. Wiley. New York.

BOLLEN, K. A. [1986]: Samle size and Bentler and Bonnet's nonnormed fit index. *Psychometrika*. Vol. 51. No. 3. p. 375–377.

BROWNE, M. W. [1982]: Covariance structures. In: *Hawkins, D. M.* (Ed.) *Topics in Applied Multivariate Analysis.* MA: Cambridge University Press. Cambridge.

BROWNE, M. W. [1974]: Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*. No. 8. p. 1–24.

BROWNE, M. W. [1984]: Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*. No. 37. p. 62–83.

CUDECK, R. – BROWNE, M. W. [1983]: Cross-validation of covariance structures. *Multivariate Behavioral Research*. No. 18. p. 147–167.

BROWNE, M. W. – CUDECK, R. [1990]: Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*. No. 24. p. 445–455.

CUDECK, R. – HENLY, S. J. [1991]: Model selection in covariance structures analysis and the „problem' of sample size: a clarification. *Psychological Bulletin*. Vol. 109. No. 3. 512–519.

GOLUB, G. H. – VAN LOAN, C. F. [1996]: Matrix computations (3rd ed.). MD: Johns Hopkins University Press. Baltimore.

JENNRICH, R. I. – SAMPSON, P. F. [1966]: Rotation for simple loadings. *Psychometrika*. Vol. 31. No. 3. p. 313–323.

KAISER, H. F. [1990]: Outline of EPIC, a new method for factoring a reduced correlation matrix. *Paper presented at Society of Multivariate Experimental Psychology*. Providence. RI.

KAISER, H. F. [1958]: The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. No. 23. p. 187–200.

LAWLEY, D. N. – MAXWELL, A. E. [1971]: Factor analysis as a statistical method. American Elsevier Publishing Company. Inc., New York.

KANO,Y. – BERKANE, M. – BENTLER, P. M. [1990]: Covariance structure analysis with heterogenous kurtosis parameters. *Biometrika*. No. 77. p. 575–585.

LI-TZE HU – BENTLER, P. M. [1992]: Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*. Vol. 112. No. 2. p. 351–362.

MCDONALD, R. P. [1989]: An index of goodness-of-fit based on noncentrality. *Journal of Classification*. No. 6. p. 97–103.

McDonald, R. P. – Marsh, H. W. [1990]: Choosing a multivariate model: noncentrality and goodness of fit. *Psychological Bulletin*. Vol. 107. No. 2. p. 247–255.

Mulaik, S. A. et al. [1989]: An evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin*. Vol. 105. No. 3. p. 430–445.

Raftery, A. [1993]: Bayesian model selection in structural equation models. In: *Bollen, K. – Long, J.* (Eds.) *Testing structural equation models.* Newbury Park. California. p. 163–180.

Raftery, A. [1995]: Bayesian model selection in social research. In: *Marsden, P.* (Ed.) *Sociological Methodology 1995.* San Francisco. p. 111–163.

Schwarz, G. [1978]: Estimating the dimension of a model. *Annals of Statistics*. No. 6. p. 461–464.

Steiger, J. H. – Lind, J. C. [1980]: Statistically-based tests for the number of common factors. Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City. May 30.

Tanaka, J. S. – Huba, G. J. [1985]: A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*. No. 38. p. 197–201.

Tucker, L. R. – Lewis, C. [1973]: The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. No. 38. p. 1–10.

ROBUST STANDARD ERROR ESTIMATION IN FIXED-EFFECTS PANEL MODELS*

GÁBOR KÉZDI[1]

This paper focuses on standard error estimation in Fixed-Effects panel models if there is serial correlation in the error process. Applied researchers have often ignored the problem, probably because major statistical packages do not estimate robust standard errors in FE models. Not surprisingly, this can lead to severe bias in the standard error estimates, both in hypothetical and real-life situations. The paper gives a systematic overview of the different standard error estimators and the assumptions under which they are consistent (in the usual large *N*, small *T* asymptotics). One of the possible reasons why the robust estimators are not used often is a fear of their bad finite sample properties. The most important results of the paper, based on an extensive Monte Carlo study, show that those fears are in general unwarranted. I also present evidence that it is the absolute size of the cross-sectional sample that primarily affects the finite-sample behaviour, not the relative size compared to the time-series dimension. That indicates good small-sample behaviour even when $N \approx T$. I introduce a simple direct test analogous to that of *White* [1980] for the restrictive assumptions behind the estimators. Its finite sample properties are fine except for low power in very small samples.

KEYWORDS: Panel models; Serial correlation.

This paper focuses on Fixed-Effects panel models (FE) with exogenous regressors on pooled cross sectional and time series data with relatively few within-individual observations. Empirical studies that estimate this kind of FE models are abundant, and they routinely estimate standard errors under the assumption of no serial error correlation within individual units. In the past three years, the top three economics journals with a focus on applied empirical research published 42 papers that estimated linear FE models with time series within individual units.[2] Out of the 42, only 6 took serial correlation into

* My first thanks go to *John Bound* and *Gary Solon* for their suggestions and support. *Jinyong Hahn*, *Steven Levitt*, *Shinichi Sakata* and *Douglas Staiger* provided many helpful comments. All remaining errors are mine. Correspondance: kezdi@econ.core.hu.

[1] Budapest University of Economics, Institute of Economics, Hungarian Academy of Sciences (IE/HAS) and Central European University, Budapest (CEU).

[2] The examined journal issues were the following: *American Economic Review*, Vol. 88. No. 4. to Vol. 91. No. 3.; *Journal of Political Economy* Vol. 106. No. 4. to Vol. 109. No. 3.; and *Quarterly Journal of Economics*, Vol. 103. No. 3. to Vol. 106. No. 2. Only papers that estimated linear FE models on panel data with time-series $T > 2$ within the individual units were considered.

account when estimated the standard errors.[3]

Serial correlation in the error process affects standard errors in FE models with more than two observations per individual unit, unless all right-hand side variables are serially uncorrelated. The stronger the correlation and the longer the time horizon is, the larger is the effect. Serial correlation consistent standard error estimators for panel models without Fixed-Effects are covered by most econometrics textbooks. Same is not true, however for FE. Similar estimators were developed explicitly for FE models by *Kiefer* [1980], *Bhargava et al.* [1982], and *Arellano* [1987], but they have been overlooked by practitioners. It seems that worries about finite sample properties are responsible for this fact. Major statistical computer packages do not allow for any robust standard error estimation in FE models. Stata^{TM} for example, calculates standard errors that are robust to serial correlation for all linear models but FE (and random effects). It does so for an analogous model but it explicitly cautions against using robust methods in samples with long time-series within individual units.[4] As we will see, however, even this warning is unwarranted.

In this paper I give a systematic overview of standard error estimation in FE models, together with the assumptions under which the estimators are consistent. I also introduce a very simple test for the assumptions in question (it is analogous to White's 1980 direct test for heteroskedasticity). The asymptotic results consider the case when T is fixed and $N \to \infty$, and they are straightforward applications of *White's* [1984] general results. The novelty in this paper is a thorough examination of the finite-sample properties of the estimators and tests. The Monte Carlo study considers various combinations of the time-series and cross-sectional sample size, and the degree of serial correlation and cross-sectional heteroskedasticity.

The most important result is that the general robust standard error estimator, known in other models as the 'cluster' estimator (introduced to FE by *Arellano* [1987]) is not only consistent in general but it behaves well in finite samples. The Monte Carlo experiments reveal that the cluster estimator is unbiased in samples of usual size although it is slightly biased downward if the cross-sectional sample is very small. The results suggest that it is the cross-sectional dimension itself that matters, not its relative size to the time-series dimension $(N \text{ and not } N/T)$. The variance of the estimator naturally increases as the sample gets small but stays moderate at usual sample sizes. *Kiefer's* [1980] estimator is consistent under the assumption of conditional homoskedasticity across individuals. Quite naturally, when consistent, it is superior to the robust estimator in terms of both variance and small-sample bias. The bias of the estimators that assume no serial correlation is substantial when the assumption is not met, and it is larger than the finite-sample bias of the robust estimators at any sample size. The bias is a function of serial correlation both in the right-hand-side variables and the error term. The test that looks at the restrictive assumptions delivers the desirable size and power properties in relatively large samples. Its power, however, is quite low in small samples unless the serial correlation is very strong.

Bertrand, Duflo and *Mullainathan* [2001] have drawn attention to robust standard error estimation in the context of a special FE model, the 'Difference-in-Differences' (DD) model. Typically, DD models estimate effects of binary treatments on different individ-

[3] Two did that by a parametric specification of the error process, one by using the cluster estimator (see later). The other three did not specify the standard error estimator they used.

[4] 'Why is it dangerous to use the robust cluster ( ) option on areg (areg estimates the same Fixed-Effects model as xtreg, fe)?' http://www.stata.com/support/faqs/stat/aregclust.html. I thank John Bound for this note.

ual units by comparing before and after treatment outcomes. Serial correlation in the error process has especially large effect on standard errors in these models because the main right-hand-side variable is highly correlated through time (the binary treatment variable changes only once in most cases). The problem is irrelevant if only two points in time are compared but it can lead to a severe bias to conventional standard error estimates in longer series. Bertrand et al. report simulation results on frequently used data (yearly earnings for US states) that show 45 to 65 percent rejection rates of a *t*-test on 'placebo' binary treatments instead of the nominal size of 5 percent. This size distortion is probably due to downward biased standard errors. Bertrand et al. suggest an intuitively appealing simulation-based method to overcome the problem. Apart from being a little complicated for applied research, their method is specific to binary treatment effects. The alternative solutions I present here are more conventional, easier to implement, and general to all FE models. They also behave well in finite samples.

The asymptotic results are stated in the main text. To keep things simple, I consider a data generating process that is *i.i.d.* in the individual units. This simplification is justified because our main concern is about the process within the individual units. The usual T fixed, $N \to \infty$ asymptotics is considered for the results. The proofs are straightforward applications of standard *i.i.d.* results (*White* [1984], for example). For this reason they are not presented in the paper. Exceptions are the simplified versions of the asymptotic covariance matrix of the FE estimator under the appropriate assumptions. They are derived in the main text because of their importance.

The remainder of the paper is organized as follows. The first section introduces the assumptions underlying the data generating process, the model, and the Fixed-Effects estimator. The second part presents the sampling covariance matrix of the FE estimator and its simplified versions under restrictive assumptions, and it introduces the estimators. The third part examines the finite sample properties of the four proposed estimators. The fourth part introduces a direct test for the restrictions and examines its finite sample properties, and the last part concludes.

SETUP

Data generating process

Assume that a T dimensional random vector y_i and a $T \times K$ dimensional random matrix x_i are generated by an independent and identically distributed (*i.i.d.*) process. More formally, we assume that the $T \times (K+1)$ dimensional random process $\{y_i, x_i\}_{i \in N}$ on $\{S, F, P\}$ is *i.i.d.,* with finite fourth moments. Note that there is no restriction in the time series dimension. In particular, nonconstant variance, unit roots, an unequal spells are allowed. We can do so because of the T fixed assumption. All asymptotic results will be driven by the cross-sectional properties of the process.

The intuition behind the data generating process (DGP) assumption is that each i is an individual observation that is drawn from a population in a random fashion. The assumption implies that there is one $E[y_i]$ and one $E[x_i]$, which are the population means. The goal of the exercise is to reveal the relationship between y and x in the population.

Model

For estimating this relationship, consider a linear panel model with exogenous regressors and individual-specific constants ('Fixed-Effects').The panel has a cross-sectional dimension *i* and a time-series dimension *t*.

$$y_{it} = x'_{it}\beta + \varepsilon_{it} = \alpha_i + x'_{it}\beta + u_{it} \qquad /1/$$

or, in vector notation,

$$y_i = \alpha_i 1 + x_i\beta + u_i , \qquad /2/$$

where $y_i = [y_{i1},\ldots,y_{iT}]'$ is $T \times 1$, $x_i = [x'_{i1},\ldots,x'_{iT}]'$ is $T \times K$, $\varepsilon_{it} = \alpha_i + u_{it}$, α_i is a scalar, $1 = [1,1,\ldots,1]'$ is $T \times 1$, and $u_i = [u_{i1},\ldots,u_{iT}]'$ is $T \times 1$, $i = 1,\ldots,N$, and $t = 1,\ldots,T$.

For future reference, let x_{ik} be the $T \times 1$ vector of the *k*-th right-hand side variable so that $x_i = [x_{i1},\ldots,x_{iK}]$.

The intuition behind the model is the following. We would like to uncover something about the conditional mean of *y* given *x*, which may be different across individuals. /2/ models the conditional mean of *y* given *x* in a linear fashion. There is an *i*-specific intercept denoted by α_i. It is interpreted as the conditional mean of y_i given $x_i = 0$. The model is restrictive in that apart from the intercept this conditional mean is the same across both the *i* and the *t* dimension. One interpretation of β is that it is a population average of the relationship after accounting for the *i*-specific intercept. The model does not put any restriction on the covariance of x_i and α_i, the latter treated as a random variable itself. Formally, we assume that all relevant moments exist and that $E[x_{ik}u'_i] = 0$ for $k = 1,2\ldots,K$. On the other hand, we allow for $E[\alpha_i x_i] \neq 0$.

We want a consistent estimator for β and its asymptotic covariance matrix. We can take the limit in both the cross-sectional and the time-series dimension, so it is important to be explicit what we mean by consistency and an asymptotic distribution. In this paper, the $N \to \infty$, *T* fixed asymptotics will be considered. In that case, it is the limiting distribution of $\sqrt{N}\left(\hat{\beta} - \beta\right)$ that we are interested in.

The $N \to \infty$, *T* fixed asymptotics is a natural setup for household or individual panels like the PSID (the Panel Study of Income Dynamics of the University of Michigan). It is also a natural approximation for country or regional panels if the time series is relatively short $(N > T)$. The simulation results suggest, however, that the proposed estimators behave well also in the finite $(N < T)$ setup.

The Fixed-Effects estimator

OLS with *N* constants for capturing each of the α_i is a natural candidate for estimation. This estimator is often called the 'least-squares dummy-variables' estimator or

LSDV in order to distinguish it from OLS with only one constant. For computational reasons, however, it is common to use the Fixed-Effects (FE, also known as Within-) estimator instead. FE is OLS on mean-differenced variables, which are defined as

$$\underset{T\times 1}{\tilde{y}_i} \equiv \left[y_{i1}-\bar{y}_i,\ldots,y_{iT}-\bar{y}_i\right]', \quad \underset{T\times K}{\tilde{x}_i} \equiv \left[x_{i1}-\bar{x}_i,\ldots,x_{iT}-\bar{x}_i\right]', \text{ and}$$

$$\underset{T\times 1}{\tilde{u}_i} \equiv \left[u_{i1}-\bar{u}_i,\ldots,u_{iT}-\bar{u}_i\right]'$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ etc.

To simplify notation, let $\underset{T\times T}{M} = I_T - \frac{1}{T}1_T1_T'$. Note that M is idempotent. Then, $\tilde{y}_i = My_i$, $\tilde{x}_i = Mx_i$ and $\tilde{u}_i = Mu_i$. The mean-differenced equation to estimate is $\tilde{y}_i = \tilde{x}_i\beta + \tilde{u}_i$, and the Fixed-Effect estimator for β is defined as

$$\hat{\beta}_{FE} \equiv \left(\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\tilde{x}_i'\tilde{y}_i\right) = \tilde{S}_{xx}^{-1}\tilde{S}_{xy}. \qquad /3/$$

$\tilde{S}_{xx} \equiv \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i$, and $\tilde{S}_{xy} \equiv \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{y}_i$. A standard result is that FE and the LSDV estimator for β on levels are computationally equivalent.

Recall that we assume that the data generating process is *i.i.d.* in the cross-sectional dimension, and therefore the $(\tilde{y}_i, \tilde{x}_i)$ are *i.i.d.*, too. $\hat{\beta}_{FE}$ is consistent for β in the $N \to \infty$, T fixed asymptotics without further assumptions about the time-series dimension. The conditional covariance matrix of $\tilde{u}_i$ affects the asymptotic covariance of $\hat{\beta}_{FE}$. Serial correlation and heteroskedasticity of any kind would also make $\hat{\beta}_{FE}$ inefficient. The rest of the paper focuses on consistent estimation of the sampling covariance of $\hat{\beta}_{FE}$. Efficient estimation of β is not addressed here.[5]

ASYMPTOTIC DISTRIBUTION OF THE FIXED-EFFECTS ESTIMATOR

The covariance matrix of $\hat{\beta}_{FE}$ is easy to derive because of cross-sectional independence and the linearity of the model.

[5] Some of the introduced covariance matrix estimators could be used for efficient estimation (feasible GLS) of the parameters. Although that seems like a natural extension of my analysis, it would introduce other problems that should be dealt with. It could aggravate bias from measurement error or misspecification of the timing of binary variables or lagged effects.

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right)^{-1} \left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{y}_i\right) = \left(\sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i\right)^{-1} \left(\sum_{i=1}^{N} \widetilde{x}_i' \left(\widetilde{x}_i \beta + \widetilde{u}_i\right)\right) =$$

$$= \beta + \widetilde{S}_{xx}^{-1} \left(\tfrac{1}{N} \sum_{i=1}^{N} \widetilde{x}_i' \widetilde{u}_i \right).$$

Proposition 1. Suppose that $\{y_i, x_i\}_{i \in N}$ is *i.i.d.* with finite second moments. Consider the Fixed-Effect (FE) panel model /1/ and /2/ and assume that $E[\widetilde{x}_i' \widetilde{x}_i]$ and $\widetilde{S}_{xx} = \frac{1}{N} \sum_{i=1}^{N} \widetilde{x}_i' \widetilde{x}_i$ are positive definite. The FE estimator defined by /3/ is consistent and asymptotically normal with covariance matrix *D* defined below /5/ and /6/

$$\hat{\beta}_{FE} = \widetilde{S}_{xx}^{-1} \widetilde{S}_{xy} \to \beta \quad prob - P \quad \text{as} \quad N \to \infty, \text{ and}$$

$$D^{-1/2} \sqrt{N} \left(\hat{\beta}_{FE} - \beta \right) \overset{A}{\sim} N(0, I), \text{ where} \qquad /4/$$

$$D \equiv E[\widetilde{x}_i' \widetilde{x}_i]^{-1} V E[\widetilde{x}_i' \widetilde{x}_i]^{-1} \quad \text{and} \qquad /5/$$

$$V \equiv E[\widetilde{x}_i' \widetilde{u}_i \widetilde{u}_i' \widetilde{x}_i]. \qquad /6/$$

The standard errors of the elements in $\hat{\beta}_{FE}$ are therefore the square root of the diagonal elements of *D* divided by *N*, or with some abuse of notation,

$$\hat{\beta}_{FE} \overset{A}{\sim} N\left(\beta, \tfrac{1}{N} D\right)$$

The proof is a straightforward application of Theorems 3.5 and 5.3 in *White* [1984]. Note that the time-series properties of $\{\widetilde{u}_i\}$ or $\{\widetilde{x}_i\}$ are not restricted in any way. Among other things, serial correlation and time-series heteroskedasticity of any kind are allowed, and so are unit roots and unequal spacing. All asymptotic results follow from the fixed length of the time-series and the cross-sectional *i.i.d.* assumption.

The next few subsections will consider simplified versions of $V = [\widetilde{x}_i' \widetilde{u}_i \widetilde{u}_i' \widetilde{x}_i]$ under restrictive assumptions.

Cross-sectional homoskedasticity

Under conditional homoskedasticity in the cross-sectional dimension but no restriction in the time series dimension, we have that $E[u_i u_i' \mid x_i] = E[u_i u_i'] \equiv \Omega$. Since *M* is nonstochastic, $E[\widetilde{u}_i \widetilde{u}_i' \mid \widetilde{x}_i] = M E[u_i u_i' \mid x_i] M$, and so

$$\widetilde{\Omega} \equiv E[\widetilde{u}_i \widetilde{u}_i'] = E[\widetilde{u}_i \widetilde{u}_i' \mid \widetilde{x}_i] = M \Omega M$$

This implies that

$$V \equiv E\left[\tilde{x}_i'\tilde{u}_i\tilde{u}_i'\tilde{x}_i\right] = E\left[\tilde{x}_i' E\left[\tilde{u}_i\tilde{u}_i' \mid \tilde{x}_i\right]\tilde{x}_i\right] = E\left[\tilde{x}_i'\tilde{\Omega}\tilde{x}_i\right].$$

Here, again, no time-series restrictions are used.[6] Notice that

$$E\left[\tilde{x}_i'\tilde{u}_i\,\tilde{u}_i'\tilde{x}_i\right] = E\left[x_i'\,MM\,u_i\,u_i'MM\,x_i\right] = E\left[x_i'Mu_iu_i'Mx_i\right].$$

Using this fact, we can simplify V further to get $V = E\left[x_i'M\Omega Mx_i\right] = E\left[\tilde{x}_i'\Omega\tilde{x}_i\right]$. This result is not used for the present estimator because we naturally want everything to be a function of the mean-differenced variables. The result is important in itself nevertheless. It means that using the levels error covariance matrix or mean-differenced error covariance matrix are equivalent.

No serial correlation

In the absence of serial correlation in the error process $\{u_{it}\}$, we have that $E\left[u_{it}u_{is}\right] = 0 \quad \forall s \neq t$, and therefore $\Omega_i \equiv E\left[u_iu_i' \mid x_i\right] = \langle \omega_{it} \rangle_{T\times T}$ a diagonal matrix, with elements $\omega_{it} = E\left[u_{it}^2 \mid x_i\right]$. Therefore,

$$V = E\left[x_i'Mu_iu_i'Mx_i\right] = E\left[\tilde{x}_i'\Omega\tilde{x}_i\right] = E\left[\sum_{i=1}^{T}\omega_{it}\tilde{x}_{it}\tilde{x}_{it}'\right] = E\left[\sum_{i=1}^{T}u_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right].$$

We would like to express this in terms of the conditional variance of the mean-differenced errors, because we estimate the model on mean-differenced data. One can show that $E\left[\tilde{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right] = \frac{T-1}{T}E\left[u_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right]$ and therefore

$$V = E\left[\sum_{t=1}^{T}u_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right] = \frac{T}{T-1}E\left[\sum_{t=1}^{T}\tilde{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right]$$

The same result is implied by zero serial correlation in the right-hand-side variables, that is if $E\left[x_{it}x_{is}'\right] = 0 \quad \forall t \neq s$. Let $\tilde{\Omega}_i \equiv E\left[\tilde{u}_i\tilde{u}_i' \mid x_i\right]$ and write

$$V = E\left[x_i'Mu_iu_i'Mx_i\right] = E\left[x_i'E\left[\tilde{u}_i'\,\tilde{u}_i \mid x_i\right]x_i\right] = E\left[x_i'\tilde{\Omega}_ix_i\right] =$$

$$= E\left[\sum_{t=1}^{T}\tilde{\omega}_{it}x_{it}x_{it}'\right] = E\left[\sum_{t=1}^{T}\tilde{u}_{it}^2x_{it}x_{it}'\right] = \frac{T}{T-1}E\left[\sum_{t=1}^{T}\tilde{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right],$$

[6] V is basically a seemingly unrelated regressions (SUR) covariance matrix, with T equations and the β constrained to be the same. *Kiefer* [1980] has introduced this estimator in the FE context.

where we used the fact that $E[x'_{it}x_{is}]=0 \quad \forall s \neq t$ and $E\left[\sum_{t=1}^{T} x_{it}x'_{it}\right]=E[x'_i x_i]$, both implied by $E[x_{it}x'_{is}]=0$. The last equality makes use the fact that $E[\tilde{x}'_i\tilde{x}_i]=\frac{T-1}{T}E[x'_i x_i]$.

The assumption we use is zero serial correlation in the error process or in (and across) the right-hand-side variables. The error process may be heteroskedastic in any dimension. This sampling covariance matrix is in fact a $\frac{T}{T-1}$-scaled version of the one that is behind the original White heteroskedasticity-consistent estimator, applied to the mean-differenced data.

Note that it is the error terms or the right-hand-side variables in levels (as opposed to mean-differences) that are assumed to be serially uncorrelated. In the fixed T setup we focus on, mean-differencing induces serial correlation in the first-differenced errors, because all $\tilde{u}_{it}$ are correlated with $\bar{u}_{it}$. Assuming no serial correlation in the mean-differenced error terms would deliver a similar result without the $\frac{T}{T-1}$ factor. We think that assumption has no intuitive appeal. The model is set up in levels, while mean-differencing is only a way to get around the correlation of α_i and x_i. We can already see that the unscaled White estimator is going to be inconsistent in the fixed-T framework. This is an example of the incidental parameter problem (*Lancaster* [2000]). The adjustment is analogous to 'degrees of freedom' corrections for the α_i parameters when the model is estimated in levels.

Homoskedasticity and no serially correlation

If there is no serial correlation and the conditional variance of u_{it} is the same at every t, that is $E[u_{it}^2 \mid x_{it}]=\Omega_i=\Omega=\sigma^2 I_T$ we get back the appropriately scaled *i.i.d.* OLS estimator for V.

$$V=E[\tilde{x}'_i\Omega\tilde{x}_i]=\sigma^2 E[\tilde{x}'_i\tilde{x}_i]\,,\quad \text{where} \quad \sigma^2=E[u_{it}^2].$$

D simplifies in this case to $D=\sigma^2 E[\tilde{x}'_i\tilde{x}_i]^{-1}$. We would like to have an expression in terms of the mean-differenced error term. Analogously to the relationship of the conditional level and mean-differenced variances, we have that $\sigma^2=E[u_{it}^2]=\frac{T}{T-1}E[\tilde{u}_{it}^2]$.

Homoskedastic errors and serially independent right-hand side variable imply the same covariance of $\hat{\beta}_{FE}$. Assume that $E[u_i u'_i \mid x_i]=\Omega$ with $\varpi_{tt}=\sigma^2$, and $E[x_{it}\,x'_{is}]=0 \quad \forall s \neq t$. Recall that no serial correlation across and within right-hand side variables implies that $E[\tilde{x}'_i\tilde{x}_i]=\frac{T-1}{T}E[x'_i x_i]$. Therefore,

$$V=E[x'_i M\,\Omega\, M x_i]=E[x'_i\tilde{\Omega}\,x_i]=E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\varpi}_{st}x'_{it}x_{is}\right]=\sum_{t=1}^{T}\tilde{\varpi}_{tt}E[x'_{it}x_{it}],$$

where the last equality holds because $E\left[\sum_{t=1}^{T}\sum_{s=1}^{T} x'_{it} x_{is}\right] = 0$ if $s \neq t$. Using $\tilde{\omega}_{tt} = E\left[\tilde{u}_{it}^2\right] = \frac{T-1}{T}\sigma^2$ we get the same result as before.

The asymptotic variance of the Fixed-Effect estimator is the $\frac{T-1}{T}$-scaled asymptotic variance of the OLS estimator on the mean-differenced data. Just like before, the zero serial correlation is assumed about u_{it} or x_{it} and not their mean-differenced counterparts. And again, conventional OLS standard errors based on the *FE* residuals are going to be inconsistent because of the incidental parameter problem, with the same bias as in the White estimator.

Estimation

We have considered four cases for *V*. Case /0/ the general, /1/ has cross-sectional conditional homoskedasticity but no restriction in the time dimension, /2/ no serial correlation, and /3/ has cross-sectional and time-series conditional homoskedasticity and no serial correlation. The four asymptotic covariance matrices are, respectively,

$$D_0 \equiv E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} E\left[\tilde{x}_i'\tilde{u}_i\tilde{u}_i'\tilde{x}_i\right] E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} \qquad /7/$$

$$D_1 \equiv E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} E\left[\tilde{x}_i'\Omega\tilde{x}_i\right] E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} \qquad /8/$$

$$D_2 \equiv \frac{T}{T-1} E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} E\left[\sum_{t=1}^{T}\tilde{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right] E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} \qquad /9/$$

$$D_3 \equiv \sigma^2 E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} . \qquad /10/$$

Let $\breve{u}$ denote the FE residuals. By the analogy principle, the proposed estimators for D_0 through D_3 are, respectively,

$$\hat{D}_0 \equiv \left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}, \text{ where} \qquad /11/$$

$$\breve{u}_i \equiv \tilde{y}_i - \tilde{x}_i\hat{\beta}_{FE} , \qquad /12/$$

$$\hat{D}_1 \equiv \left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\breve{\Omega}\tilde{x}_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}, \text{ where} \qquad /13/$$

$$\breve{\Omega} \equiv \frac{1}{N}\sum_{i=1}^{N}\breve{u}_i\breve{u}_i' \quad , \qquad /14/$$

$$\hat{D}_2 \equiv \frac{T}{T-1}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\breve{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}, \qquad /15/$$

$$\hat{D}_3 \equiv \hat{\sigma}^2\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}, \text{ where} \qquad /16/$$

$$\hat{\sigma}^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{u}_{it}^2 . \qquad /17/$$

Under our cross-sectional *i.i.d.* assumption it is straightforward to show that the $\hat{D}_j$ are consistent for the corresponding D_j (j = 0,1,2,3) if T is fixed and $N \to \infty$. The proofs are straightforward application of Theorem 5.3 (v) in *White* [1984]. One should note that the estimators don't correct for degrees of freedom decreased by the dimension of $\tilde{x}_i$. That is only for keeping things as simple as possible. Not surprisingly, the simulation results presented in the next section suggest that such corrections would slightly improve upon the finite-sample bias of the consistent estimators.

$\hat{D}_0$ is known as the 'clustered' covariance estimator, and was introduced by *Arellano* [1987]. It is always consistent in our setup. $\hat{D}_1$, introduced by *Kiefer* [1980], makes use of the covariance matrix of the FE residuals, $\breve{\Omega}$. It is consistent under any time-series behaviour as long as the error term is homoskedastic in the cross-sectional dimension. $\hat{D}_2$ is the original heteroskedasticity-consistent estimator of *White* [1980] scaled by $\frac{T}{T-1}$. It is consistent if the error term or the right-hand-side variables are serially uncorrelated. $\hat{D}_3$ is the scaled version of the homoskedasticity-consistent OLS estimator. It is the conventional sampling covariance estimator of $\hat{\beta}_{FE}$, calculated as the default by all software packages. It is consistent only under cross-sectional and time-series homoskedasticity and if either the error term or the right-hand-side variables are serially uncorrelated and have the same variance.

FINITE-SAMPLE PROPERTIES

In this section Monte Carlo simulation results are presented. To keep things simple, the analysis was restricted to a one-dimensional x variable. The data generating process involved the possibility of serial correlation in both x_{it} and u_{it}. In particular, stationary AR(1) processes were considered with autoregressive parameters 0, 0.1, 0.3, 0.5, 0.7, and 0.9 for each process (all 36 combinations were analyzed). Two separate sets were examined. In one, u was homoskedastic in the cross-sectional dimension, in the other it was heteroskedastic conditional on x. The two data generating processes were the following.

$$x_{it} = \rho_x x_{i(t-1)} + v_{xit}, \qquad x_{it} \sim N(0,1), \qquad \text{DGP /1/}$$

$$u_{it} = \rho u_{i(t-1)} + v_{uit}, \qquad u_{it} \sim N(0,1)$$

Same as DGP /1/, plus

$$v_{uit} = \sqrt{h_{it}}\,\omega_{it} \qquad \omega_{it} \sim iid\ N(0,1) \qquad \text{DGP /2/}$$

$$h_{it} = a_0 + a_1 x_{it}^2, \qquad a_0 = a_1 = 0.5\,.$$

10,000 Monte Carlo simulations were conducted for each of the 2×36 parameter settings. I have estimated the sampling distribution of the $\hat{\beta}_{FE}$ and compared its standard deviation to the mean of the 10,000 estimated standard error estimates ($SE_j = \sqrt{\frac{1}{M}\sum_{m=1}^{M} SE_{mj}}$, $j = 0,1,2,3$). These means were then used to calculate the relative bias $\left(\frac{SE_j - std(\hat{\beta}_{SE})}{std(\hat{\beta}_{SE})}\right)$. In addition to the relative bias, I also present the standard deviation of the SE_j. Several combinations of (N,T) were considered. The $(500,10)$ case establishes large-sample properties while the $(50,10)$ case looks at what happens in relatively small *N* samples. The $(50,50)$ case illustrates what happens when $N = T$ in relatively small samples, and the $(10,50)$ case is an illustration of what happens in a small-sample $N < T$. Finally, a $(10,10)$ example illustrates extreme small sample behaviour. The results are shown in Tables 1 and 2.

Table 1.1

N = 500, T = 10. Homoskedastic errors

ρ_u		ρ_x							
		0.0		0.3		0.5		0.9	
		bias	*CV*	*bias*	*CV*	*bias*	*CV*	*bias*	*CV*
0.0	SE_0	–0.01	0.03	0.00	0.04	–0.01	0.04	–0.01	0.04
	SE_1	–0.01	0.01	0.00	0.02	–0.01	0.02	0.00	0.03
	SE_2	–0.02	0.02	0.00	0.02	0.00	0.02	0.00	0.03
	SE_3	–0.01	0.01	0.00	0.02	–0.01	0.02	0.00	0.02
0.3	SE_0	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
	SE_1	0.00	0.02	0.01	0.02	0.00	0.02	0.00	0.03
	SE_2	0.00	0.02	–0.06	0.02	–0.10	0.02	–0.16	0.03
	SE_3	0.00	0.02	–0.06	0.02	–0.10	0.02	–0.17	0.02
0.5	SE_0	0.01	0.04	–0.01	0.04	0.01	0.04	–0.01	0.05
	SE_1	0.01	0.02	–0.01	0.02	0.01	0.02	–0.01	0.03
	SE_2	0.01	0.02	–0.11	0.02	–0.16	0.02	–0.26	0.03
	SE_3	0.01	0.02	–0.11	0.02	–0.16	0.02	–0.27	0.02
0.9	SE_0	0.00	0.049	–0.01	0.049	0.00	0.049	0.00	0.051
	SE_1	0.00	0.026	–0.01	0.027	0.00	0.027	0.00	0.030
	SE_2	0.00	0.026	–0.17	0.027	–0.25	0.028	–0.39	0.034
	SE_3	0.00	0.021	–0.17	0.021	–0.25	0.022	–0.42	0.026

Tables 1. contain *Relative Bias* ('bias': mean estimated SE over the standard deviation of the simulated distribution of β_{FE}) and *Coefficient of Variation* ('CV': standard error of the estimated SE distribution over its mean) of the four different SE estimators. As of *Homoskedastic errors*: In each cell, the first row corresponds to the general estimator (SE_0), the second row to the Omega-estimator (SE_1 consistent under cross-sectional homoskedasticity), the third row to the scaled version of the original White estimator (SE_2, consistent under no serial correlation), and the fourth row to the scaled version of conventional estimator (SE_3 consistent under homoskedasticity and no serial correlation). Results are from 10,000 Monte Carlo experiments.

Table 1.2

N = 50, T = 10. Homoskedastic errors

ρ_u	Estimator	ρ_x							
		0.0		0.3		0.5		0.9	
		bias	CV	bias	CV	bias	CV	bias	CV
0.0	SE_0	–0.02	0.12	–0.02	0.12	–0.02	0.12	–0.04	0.14
	SE_1	0.00	0.05	0.00	0.05	0.00	0.06	–0.02	0.08
	SE_2	0.00	0.07	0.00	0.07	–0.01	0.07	–0.02	0.08
	SE_3	0.00	0.05	0.00	0.05	0.00	0.05	–0.01	0.07
0.3	SE_0	–0.02	0.12	–0.02	0.12	–0.02	0.12	–0.01	0.14
	SE_1	0.00	0.05	0.00	0.06	0.00	0.06	0.01	0.08
	SE_2	–0.01	0.07	–0.07	0.07	–0.11	0.07	–0.16	0.08
	SE_3	0.00	0.05	–0.06	0.05	–0.10	0.05	–0.16	0.07
0.5	SE_0	–0.01	0.12	–0.03	0.13	–0.02	0.13	–0.04	0.15
	SE_1	0.01	0.06	–0.01	0.06	0.00	0.07	–0.02	0.09
	SE_2	0.00	0.07	–0.12	0.07	–0.17	0.07	–0.27	0.09
	SE_3	0.01	0.05	–0.11	0.05	–0.17	0.06	–0.27	0.07
0.9	SE_0	–0.02	0.141	–0.02	0.145	0.00	0.147	–0.04	0.159
	SE_1	0.00	0.085	0.00	0.083	0.02	0.085	–0.02	0.097
	SE_2	0.00	0.082	–0.17	0.086	–0.24	0.088	–0.40	0.107
	SE_3	0.00	0.066	–0.17	0.067	–0.25	0.070	–0.43	0.082

Table 1.3

N = 50, T = 50. Homoskedastic errors

ρ_u	Estimator	ρ_x							
		0.0		0.3		0.5		0.9	
		bias	CV	bias	CV	bias	CV	bias	CV
0.0	SE_0	–0.02	0.11	0.00	0.10	–0.01	0.11	–0.01	0.12
	SE_1	–0.01	0.02	0.01	0.02	0.00	0.03	0.00	0.06
	SE_2	–0.01	0.03	0.01	0.03	0.00	0.03	0.01	0.05
	SE_3	–0.01	0.02	0.01	0.02	0.00	0.02	0.01	0.04
0.3	SE_0	–0.01	0.11	0.00	0.11	0.00	0.11	–0.02	0.12
	SE_1	0.01	0.03	0.01	0.03	0.01	0.03	–0.01	0.06
	SE_2	0.01	0.03	–0.07	0.03	–0.13	0.03	–0.23	0.05
	SE_3	0.01	0.02	–0.07	0.02	–0.12	0.02	–0.23	0.04
0.5	SE_0	–0.02	0.11	–0.01	0.11	–0.02	0.11	–0.01	0.12
	SE_1	–0.01	0.03	0.00	0.03	–0.01	0.03	0.00	0.06
	SE_2	–0.01	0.03	–0.13	0.03	–0.22	0.04	–0.36	0.05
	SE_3	–0.01	0.03	–0.13	0.03	–0.22	0.02	–0.36	0.04
0.9	SE_0	–0.02	0.120	–0.01	0.123	–0.02	0.123	–0.03	0.135
	SE_1	–0.01	0.059	0.00	0.059	0.00	0.059	–0.01	0.071
	SE_2	0.00	0.048	–0.23	0.047	–0.37	0.047	–0.63	0.065
	SE_3	0.00	0.041	–0.23	0.041	–0.37	0.040	–0.64	0.054

Table 1.4

N = 10, T = 50. Homoskedastic errors

ρ_u	Estimator	ρ_x							
		0.0		0.3		0.5		0.9	
		bias	CV	bias	CV	bias	CV	bias	CV
0.0	SE_0	–0.08	0.24	–0.08	0.25	–0.07	0.25	–0.09	0.27
	SE_1	–0.01	0.08	–0.01	0.09	0.00	0.09	–0.02	0.14
	SE_2	–0.01	0.06	–0.01	0.06	0.00	0.07	0.00	0.10
	SE_3	0.00	0.04	0.00	0.05	0.01	0.05	0.00	0.09
0.3	SE_0	–0.08	0.25	–0.09	0.25	–0.08	0.25	–0.09	0.28
	SE_1	–0.02	0.09	–0.01	0.09	–0.02	0.09	–0.02	0.14
	SE_2	–0.01	0.07	–0.09	0.07	–0.14	0.07	–0.23	0.10
	SE_3	–0.01	0.05	–0.09	0.05	–0.14	0.05	–0.23	0.09
0.5	SE_0	–0.08	0.25	–0.08	0.25	–0.08	0.25	–0.10	0.28
	SE_1	–0.01	0.09	–0.01	0.09	–0.01	0.10	–0.02	0.14
	SE_2	–0.01	0.07	–0.14	0.07	–0.22	0.07	–0.37	0.11
	SE_3	0.00	0.05	–0.13	0.05	–0.22	0.06	–0.37	0.10
0.9	SE_0	–0.10	0.275	–0.09	0.273	–0.09	0.273	–0.11	0.296
	SE_1	–0.02	0.141	–0.02	0.143	–0.02	0.145	–0.03	0.167
	SE_2	–0.02	0.104	–0.24	0.104	–0.37	0.108	–0.64	0.147
	SE_3	–0.01	0.092	–0.23	0.094	–0.37	0.096	–0.63	0.124

Table 1.5

N = 10, T = 10. Homoskedastic errors

ρ_u	Estimator	ρ_x							
		0.0		0.3		0.5		0.9	
		bias	CV	bias	CV	bias	CV	bias	CV
0.0	SE_0	–0.10	0.27	–0.09	0.27	–0.11	0.28	–0.13	0.31
	SE_1	–0.03	0.13	–0.02	0.13	–0.02	0.15	–0.04	0.20
	SE_2	–0.03	0.14	–0.03	0.15	–0.02	0.15	–0.04	0.18
	SE_3	–0.01	0.11	–0.01	0.11	–0.01	0.12	–0.02	0.15
0.3	SE_0	–0.08	0.27	–0.11	0.27	–0.10	0.28	–0.13	0.31
	SE_1	–0.01	0.14	–0.03	0.14	–0.03	0.15	–0.04	0.19
	SE_2	–0.02	0.14	–0.10	0.15	–0.13	0.15	–0.19	0.19
	SE_3	0.00	0.11	–0.08	0.11	–0.11	0.12	–0.17	0.15
0.5	SE_0	–0.10	0.27	–0.09	0.28	–0.11	0.29	–0.13	0.32
	SE_1	–0.02	0.14	–0.01	0.15	–0.03	0.16	–0.04	0.20
	SE_2	–0.03	0.15	–0.13	0.16	–0.18	0.16	–0.27	0.19
	SE_3	–0.01	0.12	–0.10	0.12	–0.17	0.13	–0.26	0.16
0.9	SE_0	–0.09	0.311	–0.11	0.310	–0.11	0.314	–0.14	0.331
	SE_1	–0.02	0.193	–0.04	0.193	–0.03	0.195	–0.05	0.217
	SE_2	–0.02	0.180	–0.19	0.185	–0.28	0.188	–0.42	0.222
	SE_3	0.00	0.150	–0.18	0.154	–0.26	0.155	–0.42	0.185

Tables 2. contain *Relative Bias* ('bias': mean estimated SE over the standard deviation of the simulated distribution of β_{FE}) and *Coefficient of Variation* ('CV': standard error of the estimated SE distribution over its mean) of the four different SE estimators, and

*Conditional heteroskedasticity in the cross-sectional dimension***.** In each cell, the first row corresponds to the general estimator (SE_0), the second row to the Omega-estimator (SE_1 consistent under cross-sectional homoskedasticity), the third row to the scaled version of the original White estimator (SE_2, consistent under no serial correlation), and the fourth row to the scaled version of conventional estimator (SE_3 consistent under homoskedasticity and no serial correlation). Results are from 10,000 Monte Carlo experiments.

Table 2.1

N = 500, T = 10. Cross-sectional conditional heteroskedasticity

ρ_u	Estimator	ρ_x 0.0 bias	0.0 CV	0.3 bias	0.3 CV	0.5 bias	0.5 CV	0.9 bias	0.9 CV
0.0	SE_0	–0.01	0.04	0.01	0.04	0.00	0.05	–0.02	0.05
	SE_1	–0.28	0.01	–0.25	0.01	–0.24	0.02	–0.13	0.03
	SE_2	–0.03	0.02	–0.01	0.02	–0.02	0.02	–0.02	0.02
	SE_3	–0.28	0.01	–0.25	0.01	–0.24	0.02	–0.13	0.02
0.3	SE_0	0.00	0.04	0.00	0.05	0.00	0.05	–0.01	0.06
	SE_1	–0.26	0.01	–0.24	0.02	–0.23	0.02	–0.12	0.03
	SE_2	–0.02	0.02	–0.08	0.02	–0.12	0.02	–0.18	0.02
	SE_3	–0.26	0.01	–0.29	0.02	–0.31	0.02	–0.27	0.02
0.5	SE_0	0.00	0.05	0.00	0.04	0.00	0.05	–0.01	0.06
	SE_1	–0.23	0.01	–0.22	0.02	–0.20	0.02	–0.11	0.03
	SE_2	–0.02	0.02	–0.12	0.02	–0.18	0.02	–0.27	0.03
	SE_3	–0.23	0.02	–0.31	0.02	–0.34	0.02	–0.35	0.02
0.9	SE_0	–0.02	0.05	–0.02	0.05	0.00	0.06	0.00	0.06
	SE_1	–0.14	0.02	–0.12	0.03	–0.10	0.02	–0.05	0.03
	SE_2	–0.02	0.02	–0.17	0.02	–0.25	0.03	–0.38	0.03
	SE_3	–0.14	0.02	–0.26	0.02	–0.34	0.02	–0.45	0.03

Table 2.2

N = 50, T = 10. Cross-sectional conditional heteroskedasticity

ρ_u	Estimator	ρ_x 0.0 bias	0.0 CV	0.3 bias	0.3 CV	0.5 bias	0.5 CV	0.9 bias	0.9 CV
0.0	SE_0	–0.02	0.13	–0.03	0.14	–0.03	0.14	–0.03	0.16
	SE_1	–0.27	0.05	–0.26	0.05	–0.24	0.06	–0.11	0.09
	SE_2	–0.03	0.09	–0.04	0.09	–0.03	0.09	–0.01	0.10
	SE_3	–0.27	0.04	–0.26	0.05	–0.24	0.05	–0.11	0.07
0.3	SE_0	–0.02	0.14	–0.02	0.14	–0.02	0.15	–0.03	0.16
	SE_1	–0.25	0.05	–0.24	0.06	–0.22	0.06	–0.12	0.09
	SE_2	–0.03	0.09	–0.09	0.09	–0.12	0.09	–0.18	0.10
	SE_3	–0.26	0.05	–0.29	0.05	–0.31	0.05	–0.26	0.07
0.5	SE_0	–0.02	0.14	–0.03	0.14	–0.02	0.15	–0.03	0.17
	SE_1	–0.23	0.06	–0.22	0.06	–0.20	0.07	–0.10	0.10
	SE_2	–0.03	0.09	–0.13	0.09	–0.18	0.09	–0.26	0.10
	SE_3	–0.23	0.05	–0.31	0.05	–0.34	0.06	–0.34	0.08
0.9	SE_0	–0.02	0.15	–0.03	0.16	–0.03	0.17	–0.05	0.19
	SE_1	–0.12	0.08	–0.11	0.08	–0.10	0.09	–0.07	0.11
	SE_2	–0.01	0.09	–0.17	0.09	–0.26	0.10	–0.40	0.12
	SE_3	–0.12	0.07	–0.26	0.07	–0.34	0.07	–0.46	0.09

Table 2.3

N = 50, T = 50. Cross-sectional conditional heteroskedasticity

ρ_u	Estimator	ρ_x 0.0 bias	0.0 CV	0.3 bias	0.3 CV	0.5 bias	0.5 CV	0.9 bias	0.9 CV
0.0	SE_0	–0.03	0.11	–0.01	0.11	–0.02	0.11	–0.03	0.13
	SE_1	–0.29	0.02	–0.28	0.02	–0.28	0.03	–0.23	0.05
	SE_2	–0.02	0.04	0.00	0.04	–0.01	0.04	–0.01	0.05
	SE_3	–0.30	0.02	–0.28	0.02	–0.28	0.02	–0.23	0.03
0.3	SE_0	–0.01	0.11	–0.01	0.11	–0.01	0.11	–0.03	0.14
	SE_1	–0.27	0.02	–0.26	0.03	–0.26	0.03	–0.23	0.05
	SE_2	0.00	0.04	–0.08	0.04	–0.13	0.04	–0.24	0.05
	SE_3	–0.27	0.02	–0.32	0.02	–0.36	0.02	–0.41	0.03
0.5	SE_0	–0.01	0.11	–0.02	0.11	–0.03	0.12	–0.03	0.14
	SE_1	–0.23	0.03	–0.24	0.03	–0.24	0.03	–0.22	0.06
	SE_2	0.00	0.04	–0.14	0.04	–0.23	0.05	–0.38	0.05
	SE_3	–0.24	0.02	–0.34	0.03	–0.41	0.02	–0.51	0.04
0.9	SE_0	–0.02	0.13	–0.02	0.13	–0.02	0.13	–0.03	0.16
	SE_1	–0.09	0.06	–0.09	0.06	–0.10	0.06	–0.13	0.07
	SE_2	–0.01	0.05	–0.23	0.05	–0.36	0.05	–0.63	0.07
	SE_3	–0.09	0.04	–0.30	0.04	–0.43	0.04	–0.68	0.05

Table 2.4

N = 10, T = 50. Cross-sectional conditional heteroskedasticity

ρ_u	Estimator	ρ_x 0.0 bias	0.0 CV	0.3 bias	0.3 CV	0.5 bias	0.5 CV	0.9 bias	0.9 CV
0.0	SE_0	–0.09	0.25	–0.09	0.25	–0.08	0.26	–0.12	0.29
	SE_1	–0.27	0.09	–0.27	0.09	–0.26	0.09	–0.23	0.14
	SE_2	–0.02	0.09	–0.02	0.09	–0.01	0.09	–0.02	0.10
	SE_3	–0.29	0.04	–0.29	0.04	–0.28	0.05	–0.23	0.08
0.3	SE_0	–0.08	0.25	–0.09	0.26	–0.09	0.26	–0.10	0.29
	SE_1	–0.25	0.09	–0.25	0.09	–0.25	0.10	–0.21	0.14
	SE_2	–0.02	0.09	–0.10	0.09	–0.15	0.09	–0.23	0.10
	SE_3	–0.27	0.04	–0.33	0.04	–0.37	0.05	–0.39	0.08
0.5	SE_0	–0.07	0.26	–0.09	0.26	–0.09	0.26	–0.12	0.30
	SE_1	–0.22	0.09	–0.22	0.10	–0.23	0.10	–0.21	0.15
	SE_2	0.00	0.09	–0.14	0.09	–0.23	0.09	–0.38	0.11
	SE_3	–0.23	0.05	–0.34	0.05	–0.41	0.05	–0.50	0.08
0.9	SE_0	–0.10	0.28	–0.10	0.28	–0.11	0.29	–0.12	0.33
	SE_1	–0.11	0.14	–0.10	0.14	–0.11	0.15	–0.13	0.18
	SE_2	–0.02	0.11	–0.24	0.11	–0.37	0.11	–0.63	0.15
	SE_3	–0.10	0.09	–0.30	0.09	–0.43	0.09	–0.67	0.12

Table 2.5

N = 10, T = 10. Cross-sectional conditional heteroskedasticity

ρ_u	Estimator	ρ_x							
		0.0		0.3		0.5		0.9	
		bias	CV	bias	CV	bias	CV	bias	CV
0.0	SE_0	–0.11	0.29	–0.12	0.29	–0.13	0.30	–0.14	0.33
	SE_1	–0.26	0.13	–0.25	0.14	–0.24	0.14	–0.14	0.21
	SE_2	–0.07	0.19	–0.07	0.19	–0.07	0.18	–0.05	0.20
	SE_3	–0.27	0.10	–0.26	0.10	–0.25	0.11	–0.12	0.17
0.3	SE_0	–0.11	0.29	–0.13	0.30	–0.13	0.30	–0.14	0.34
	SE_1	–0.24	0.13	–0.25	0.14	–0.23	0.15	–0.13	0.21
	SE_2	–0.06	0.18	–0.14	0.18	–0.16	0.18	–0.20	0.21
	SE_3	–0.26	0.10	–0.30	0.11	–0.31	0.11	–0.25	0.17
0.5	SE_0	–0.12	0.30	–0.12	0.30	–0.12	0.31	–0.14	0.34
	SE_1	–0.24	0.14	–0.22	0.15	–0.20	0.16	–0.13	0.22
	SE_2	–0.07	0.18	–0.16	0.19	–0.21	0.19	–0.29	0.21
	SE_3	–0.24	0.11	–0.30	0.11	–0.33	0.12	–0.33	0.17
0.9	SE_0	–0.12	0.32	–0.13	0.33	–0.13	0.34	–0.16	0.36
	SE_1	–0.14	0.19	–0.13	0.19	–0.13	0.20	–0.11	0.24
	SE_2	–0.05	0.19	–0.20	0.20	–0.28	0.20	–0.43	0.24
	SE_3	–0.13	0.14	–0.26	0.15	–0.34	0.15	–0.45	0.20

In order to assess the results, note that in the first set (Tables 1), SE_0 and SE_1 are always consistent for the true *SE*, and SE_2 and SE_3 are consistent if $\rho_u \rho_x = 0$ (either of the two is zero). In the second set (Tables 2), SE_0 is always consistent for the true *SE*, SE_2 is consistent if $\rho_u \rho_x = 0$, but SE_1 and SE_3 are never consistent because of cross-sectional heteroskedasticity.

Tables 1.1 and 2.1 present the large-sample results. Bias of the consistent estimators is virtually zero. The bias of the inconsistent estimators here increases as ρ_u and ρ_x increase. Unbiasedness of SE_2 and SE_3, when they are consistent, indicates that the unscaled White and OLS estimators are biased also in small samples. In the heteroskedastic setup, the bias of SE_1 and SE_3 is dominated by heteroskedasticity in the small-ρ setups, and serial correlation takes over as ρ_u and ρ_x increase. The variance of the estimators behave the predictable way, with the more restrictive ones having smaller variation. These differences, however, are very small for practical purposes.

Smaller-N samples (Tables 1.2 and 2.2) basically deliver the large-sample results in terms of the bias. SE_0 shows a small-sample bias that is larger than other consistent estimators but it is still negligible. Differences in the variance are magnified, as expected, but they are not extremely large either. Both the small-sample bias and the variance of the consistent estimators increase as ρ_u and ρ_x increase. This reflects the fact that higher serial correlation decreases the variation in the variables if the overall error and RHS variance is fixed, as were throughout the simulation.

Bias due to serial correlation is greater as *T* increases. Small-sample bias of SE_0 stays small in the $(50, 50)$ setup but becomes a significant negative 8-16 percent when in the $N = 10$ setups. The results indicate that it is the overall sample size, and especially *N*, the size of the cross-sectional sample that determines the small-sample bias. Reluctance of using the cluster estimator SE_0 when *T* is large is unjustified. The variance disadvantage of SE_0 is larger in the $(50, 50)$ case than the $(50,10)$ case, as expected, but the differences remain modest. The small-sample properties of SE_1 (the more restrictive serial correlation consistent estimator) are significantly better when it is consistent. Its small sample bias stays close to zero even in the $(10,10)$ sample, and its standard deviation is below 25 percent larger than that of SE_3 when $N = 10, T = 10$ and 50 percent when $N = 10, T = 50$.

The good small-sample behaviour may seem somewhat surprising. But they may simply reflect that the standard error estimators take an average over all NT observations. In this light, even the $(10,10)$ sample is not small: it consists of 100 observations altogether.

The results have the following practical implications. In large samples SE_0 is just as good for applied work as the restricted estimators even when the latter are also consistent. In smaller samples there is some advantage SE_1 (the Omega-estimator) if that is consistent for the true *SE*. The conventional estimator SE_3 has no substantial advantage over SE_1, other than computational simplicity. The simulation results suggest that properties of the estimators don't depend much on the relative size of *T* and *N* but rather on the total sample size *NT* and especially *N* itself. At the same time, an increasing *T* increases the bias due to serial correlation. Cautioning against using the 'clustered' estimator SE_0 when the time-series is long is therefore not simply unnecessary but quite misleading.

A DIRECT TEST FOR HOMOSKEDASTICITY AND NO SERIAL CORRELATION

In this section I propose a direct test for the restrictive assumptions under which the alternative (less robust) estimators are consistent. $\hat{D}_3$ is easier to compute and has the best properties if consistent. $\hat{D}_1$ performs significantly better in terms of variance than $\hat{D}_0$ when both are consistent, especially in smaller samples. Moreover, the properties of $\hat{D}_1$ match closely those of $\hat{D}_3$ if both are consistent. If we can test for the restrictions that make $\hat{D}_1$, $\hat{D}_2$ or $\hat{D}_3$, consistent we can always choose the best consistent estimator. In this section I develop such a test.

Let me introduce the following notation. Recall that the alternative standard error estimators differ only in how they estimate $V = E[\tilde{x}_i' \tilde{u}_i \tilde{u}_i' \tilde{x}_i]$. The assumptions behind the restricted estimators can therefore be tested by comparing the corresponding *V* estimates to one that is always consistent. Define

$$\hat{V}_0 \equiv \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_i' \breve{u}_i \breve{u}_i' \tilde{x}_i \qquad /18/$$

$$\hat{V}_1 \equiv \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\breve{\Omega}\tilde{x}_i , \qquad \breve{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\breve{u}_i\breve{u}_i' , \qquad /19/$$

$$\hat{V}_2 \equiv \frac{1}{N}\frac{T}{T-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}' , \qquad /20/$$

$$\hat{V}_3 \equiv \frac{\hat{\sigma}^2}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i , \qquad \hat{\sigma}^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{u}_{it}^2 . \qquad /21/$$

$\hat{V}_0$ is always consistent for V. $\hat{V}_1$ is consistent under cross-sectional homoskedasticity. $\hat{V}_2$ is consistent under no serial correlation in the (levels) error or the (levels) right-hand-side variables. $\hat{V}_3$ is consistent if both $\hat{V}_1$ and $\hat{V}_2$ are consistent and time-series homoskedasticity also holds. A direct way to test whether the more restrictive assumptions hold is to check whether $V_1 = V$, $V_2 = V$, or $V_3 = V$. In order to formulate the linear hypotheses, let's use the *vech* operator that stacks columnwise the diagonal and sub-diagonal elements of a symmetric matrix.[7]

$$v_j \equiv vech\left(V_j\right) \qquad /22/$$

$$\hat{v}_j \equiv vech\left(\hat{V}_j\right), \qquad j = 0,1,2,3 \qquad /23/$$

The hypotheses are

$$H_0 : \quad v_j - v_0 = 0, \quad j = 1,2,3$$

$$H_1 : \quad v_j - v_0 \neq 0, \quad j = 1,2,3 .$$

The test I propose is analogous to *White's* [1980] test for heteroskedasticity. Since $\hat{v}_0$ is always consistent and the $\hat{v}_j$ are consistent only under the appropriate H_0 , their distance is an intuitive test statistic. If they are close enough, the restrictions probably hold. If they are very far, they probably do not hold.

Proposition 2. Suppose that $\{y_i, x_i\}_{i \in N}$ is *i.i.d.* with finite fourth moments. Consider the Fixed-Effect (FE) panel model (1-2) and assume that $E[\tilde{x}_i'\tilde{x}_i]$ and $\tilde{S}_{xx} \equiv \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i$ are positive definite. The test-statistic h_j defined below using /18-23/ are distributed chi-squared under H_0 .

[7] Suppose that $K = 3$ and $\mathbf{A}$ is symmetric: $\mathbf{A} = \{a_{ij}\}$. Then, $vech(\mathbf{A}) = (a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33})'$. See Magnus and Neudecker (1988), e.g., for more discussion.

Their asymptotic power is 1. That is,

$$h_j \equiv N\left(\hat{v}_j - \hat{v}_0\right)' \hat{C}_j^{-1}\left(\hat{v}_j - \hat{v}_0\right) \sim \chi^2\left(\frac{K(K+1)}{2}+1\right)$$

under H_0 $(j = 1,2,3)$ and

$\lim_{N\to\infty} Pr\left(h_j > c\right) = 1 \quad (j = 1,2,3)$ for any $c \in \Re$ otherwise.

$$\hat{C}_1 \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i - \tilde{x}_i'\breve{\Omega}\tilde{x}_i\right)\left(\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i - \tilde{x}_i'\breve{\Omega}\tilde{x}_i\right)',$$

$$\hat{C}_2 \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i - \sum_{t=1}^{T}\breve{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right)\left(\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i - \sum_{t=1}^{T}\breve{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right)',$$

$$\hat{C}_3 \equiv \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i - \hat{\sigma}^2\tilde{x}_i\tilde{x}_i'\right)\left(\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i - \hat{\sigma}^2\tilde{x}_i\tilde{x}_i'\right)', \text{ where}$$

$$\breve{\Omega} \equiv \frac{1}{N}\sum_{i=1}^{N}\breve{u}_i\breve{u}_i' \quad \text{and}$$

$$\hat{\sigma}^2 \equiv \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{u}_{it}^2 .$$

The proof is straightforward provided the simplifications to V derived earlier and the consistency of the estimators $\hat{V}_j$ for the appropriate V_j. It is therefore skipped here and is available upon request.

Finite-sample properties

Tables 3 and 4 report simulated rejection rates for the three tests in the data generating processes identical to Tables 1 and 2, respectively, based on 10,000 Monte Carlo trials. Results for the $(500,10)$ and $(50,10)$ setups are presented only but all setups from Tables 1-2 were examined. The unpublished results indicate that given *N*, the size does not change but the power increases with *T*, and the test loses almost all of its power on extremely small-*N* samples.

Tables 3. have rejection rates of h_1 (H_0: $V_1 = V_0$), h_2 (H_0: $V_2 = V_0$) and h_3 (H_0: $V_3 = V_0$). Nominal size=0.05. As of *homoskedastic errors*: V_1 and V_0 are asymptotically equivalent always; V_2 and V_0, and V_3 and V_0 are asymptotically equivalent if $\rho_x = 0$, or $\rho_u = 0$. Results are from 10,000 Monte Carlo experiments.

Table 3.1

N = 500, T = 10. Homoskedastic errors

ρ_u		ρ_x 0.0	0.3	0.5	0.9
0.0	h_1	0.05	0.05	0.04	0.04
	h_2	0.04	0.06	0.04	0.06
	h_3	0.05	0.05	0.05	0.05
0.3	h_1	0.04	0.04	0.04	0.04
	h_2	0.05	0.56	0.92	1.00
	h_3	0.05	0.36	0.79	0.99
0.5	h_1	0.04	0.04	0.04	0.04
	h_2	0.06	0.90	1.00	1.00
	h_3	0.05	0.79	0.99	1.00
0.9	h_1	0.04	0.04	0.04	0.03
	h_2	0.06	1.00	1.00	1.00
	h_3	0.05	0.99	1.00	1.00

Table 3.2

N = 50, T = 10. Homoskedastic errors

ρ_u		ρ_x 0.0	0.3	0.5	0.9
0.0	h_1	0.09	0.08	0.08	0.08
	h_2	0.08	0.08	0.08	0.09
	h_3	0.09	0.09	0.09	0.10
0.3	h_1	0.09	0.08	0.08	0.07
	h_2	0.08	0.05	0.07	0.14
	h_3	0.09	0.04	0.04	0.06
0.5	h_1	0.08	0.08	0.08	0.07
	h_2	0.08	0.06	0.18	0.46
	h_3	0.09	0.04	0.09	0.26
0.9	h_1	0.08	0.07	0.07	0.06
	h_2	0.09	0.14	0.44	0.86
	h_3	0.10	0.08	0.29	0.74

Table 4.1.

N = 500, T = 10
Cross-sectional conditional heteroskedasticity

ρ_u		ρ_x 0.0	0.3	0.5	0.9
0.0	h_1	1.00	1.00	1.00	0.77
	h_2	0.10	0.07	0.06	0.06
	h_3	1.00	1.00	1.00	0.71
0.3	h_1	1.00	1.00	1.00	0.53
	h_2	0.08	0.73	0.96	0.99
	h_3	1.00	1.00	1.00	1.00
0.5	h_1	1.00	1.00	1.00	0.37
	h_2	0.05	0.96	1.00	1.00
	h_3	1.00	1.00	1.00	1.00
0.9	h_1	0.60	0.48	0.34	0.04
	h_2	0.04	0.99	1.00	1.00
	h_3	0.58	1.00	1.00	1.00

Table 4.2.

N = 50, T = 10
Cross-sectional conditional heteroskedasticity

ρ_u		ρ_x *0.0*	*0.3*	*0.5*	*0.9*
0.0	h_1	0.36	0.30	0.19	0.03
	h_2	0.05	0.05	0.05	0.08
	h_3	0.36	0.29	0.18	0.03
0.3	h_1	0.28	0.21	0.12	0.03
	h_2	0.05	0.05	0.07	0.11
	h_3	0.28	0.39	0.38	0.15
0.5	h_1	0.19	0.14	0.08	0.03
	h_2	0.06	0.07	0.18	0.32
	h_3	0.20	0.40	0.48	0.36
0.9	h_1	0.03	0.03	0.03	0.04
	h_2	0.08	0.11	0.35	0.74
	h_3	0.04	0.16	0.36	0.65

The results in general reflect the finite-sample properties of the estimators. The tests deliver their asymptotic properties in the $N = 500, T = 10$ setup. The notable exceptions are h_2 and h_3 under conditional homoskedasticity and very weak serial correlation ($\rho_u < 0.3$, $\rho_x < 0.3$, Table 3.1), and h_1 under conditional heteroskedasticity and very strong serial correlation (Table 4.1). The former are quite natural while the latter reflects that strong serial correlation dominates heteroskedasticity in the conditional variance (Table 2.1).

Tables 4. have rejection rates of h_1 (H$_0$: $V_1 = V_0$), h_2 (H$_0$: $V_2 = V_0$) and h_3 (H$_0$: $V_3 = V_0$). Nominal size=0.05. As of *Conditional cross-sectional heteroskedasticity in the errors*: V_1 and V_0, and V_3 and V_0 are never asymptotically equivalent; V_2 and V_0 are asymptotically equivalent if $\rho_x =$ or $\rho_u = 0$. Results are from 10,000 Monte Carlo experiments.

The size is about right in moderate size samples. It is slightly biased upward which makes the test a little too conservative (the actual size varies between 0.06 and 0.09 compared to a nominal size of 0.05). The power varies considerably with the alternatives. In the homoskedastic setup, the power, quite naturally, is a positive function of the serial correlation in *u* and *x*. The heteroskedastic setup yields the same result except against V_1, the heteroskedasticity-inconsistent but serial correlation consistent estimator.

*

The paper examined linear FE models with short time series within individual units. Serial correlation in the error process and the right-hand-side variables was shown to induce severe bias in the conventional standard error estimates. At the same time, the paper has shown that well-known robust ('clustered') estimator applied to the mean-differenced data is not only consistent but also behaves well in finite samples. Applied researchers should, therefore, routinely estimate the robust estimator in moderate-sized and large samples, the same way they already routinely estimate the heteroskedasticity-consistent estimator in cross-sectional models. The robust estimator does not get biased or significantly more disperse as the time-series dimension increases. At the same time, however, the serial correlation bias of the inconsistent estimators increases with the time-series dimension. Therefore, contrary to the intuition of many applied researchers, the advantages of the robust estimator increase as the time-series get longer. It is the cross-sectional size of the sample that primarily affects the finite-sample behaviour of the estimator.

In small samples and under cross-sectional homoskedasticity, there is some advantage of using the alternative serial correlation consistent estimator, the 'Omega'-estimator. The conventional FE standard error estimator (the scaled version of the conventional OLS estimator on the mean-differenced data) has no significant advantage over the Omega-estimator even if both are consistent. In small samples, therefore, the Omega-estimator should be used unless there is evidence for cross-sectional heteroskedasticity. The paper has also introduced a simple direct test for the assumptions under which the restrictive estimators are consistent. The test delivers the appropriate size properties. Its power is quite small in small samples but good enough to detect strong serial correlation.

REFERENCES

ARELLANO, M. [1987]: Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*. Vol. 49. No. 4. p. 431–434.

BERTRAND, M. – DUFLO, E. – MULLAINATHAN, S. [2001]: *How much should we trust differences-in-differences estimates*? Working Paper presented at the UCLA/RAND Labor and Population Workshop. Feb. 20. 2001.

BHARGAVA, A. – FRANZINI, L. – NARENDRANATHAN, W. [1982]: Serial correlation and the Fixed-Effects model. *The Review of Economic Studies*. Vol. 49. No. 4. p. 553–549.

KIEFER, N. M. [1980]: Estimation of Fixed-Effect models for time series of cross-sections with arbitrary intertemporal covariance. *Journal of Econometrics.* Vol. 14. No. 2. p. 195–202.

LANCASTER, T. [2000]: The incidental parameter problem since 1948. *Journal of Econometrics*. Vol. 95. p. 391–413.

MAGNUS, J. R. – NEUDECKER, H. [1988]: *Matrix differential calculus with applications in statistics and econometrics*. New-York. Wiley.

WHITE, H. [1980]: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica.* Vol. 48. No. 4. p. 817–838.

WHITE, H. [1984]: *Asymptotic theory for econometricians*. Academic Press. Orlando.