11-1996

Statistical Techniques for Labor Market Modelling: Seasonal Adjustment and Local Area Unemployment

Christopher J. O'Leary
W.E. Upjohn Institute for Employment Research, oleary@upjohn.org

Authors
Christopher J. O'Leary, *W.E. Upjohn Institute for Employment Research*

Upjohn Author(s) ORCID Identifier
https://orcid.org/0000-0002-3372-7527

Statistical Techniques for Labor Market Modelling: Seasonal Adjustment and Local Area Unemployment

A report on activity B.1 in the project to provide technical assistance to improve labor market analyses in Hungary, under the agreement between the United States Department of Labor and the Hungarian Ministry of Labor.

November, 1996

Prepared for:

The Ministry of Labor
H-1051. Budapest
Roosevelt tér 7-8.
Hungary

Prepared by:

Christopher J. O'Leary, Senior Economist
W.E. Upjohn Institute for Employment Research
300 South Westnedge Avenue
Kalamazoo, Michigan 49007
U.S.A.

Administered by:

The Bureau of International Labor Affairs
U.S. Department of Labor
200 Constitution Avenue
Washington, DC 20210
U.S.A.

Funded by a loan from the World Bank to the Hungarian Ministry of Labor.

ACKNOWLEDGEMENTS

The work summarized in this report was done by the Hungarian private consulting firm Multi-Racio. The principal contributors to this work were Dr. Miklos Banai who is a physicist by training and Dr. Istvan Varga who is a sociologist. In March of 1993, about the time of the great blizzard which paralyzed North America east of the Mississippi, Drs. Banai and Varga traveled to Washington, DC to learn about local area unemployment estimation from Sharon Brown and her colleagues in the U.S. Bureau of Labor Statistics and to learn about seasonal adjustment methods from Dr. David Findley at the U.S. Census Bureau, and also traveled to Ottawa, Ontario to learn about seasonal adjustment of labor force time series from Marietta Morrey of Statistics Canada. The consultants survived the snowstorm and returned to Hungary armed with a wealth of practical information and ideas about how to apply the North American methods to the situation in Hungary.

In addition to Drs. Banai and Varga other experts, originally trained as theoretical and applied physicists, who contributed to the work summarized here were: Bela Lukacs, Miklos Prisznyak, Kazmer Koleszar, Laszlo Szeute, Tibor Becze Deak, and Maria Surveges.

From the U.S. Department of Labor, assistance was administered by Ambassador John Ferch in the Bureau of International Labor Affairs and his associates Jim Perlmutter, Michael McManus, and Hank Guzda. Elizabeth Taylor of the Bureau of Labor Statistics also helped to organize activities.

At Statistics Canada I thank Marietta Morrey and her colleagues, and at the Bureau of Labor Statistics I thank Sharon Brown and her colleagues. I also thank Dr. David Findley of the U.S. Census Bureau for his guidance on seasonal adjustment methods.

The experts at Multi-Racio have provided practical and well documented methods for use by the National Labor Center in Hungary, I am fortunate to have worked with them and Gyorgy Lazar of the Hungarian National Labor Center on the project.

Christopher J. O'Leary
Kalamazoo, Michigan
November, 1996

Statistical Techniques for Labor Market Modelling:
Seasonal Adjustment and Local Area Unemployment

TABLE OF CONTENTS

SECTION PAGE

APPENDICES

1. INTRODUCTION

Under the most recent phase of services provided by the Bureau of International Labor Affairs in the U.S. Department of Labor to the Hungarian Ministry of Labor, the W. E. Upjohn Institute for Employment Research served as a sub-contractor providing technical assistance to the Hungarian Ministry of Labor.

Beginning in 1994 the W. E. Upjohn Institute for Employment Research worked with the National Labor Center in Budapest on an agenda of four activities under the heading of "Labor Market Modelling." The activities under this project included two efforts to support management and planning of active labor programs: the development of an adjustment methodology for performance indicators and a proposal for a budget allocation model which incorporates program performance as a factor. Results of these two efforts are summarized in a report entitled: *Methods for Performance Based Management of Active Labor Programs in Hungary: An Adjustment Methodology for Performance Indicators and a Proposal for Budget Allocation.*

The other two activities were undertaken by the Hungarian consulting firm Multi-Racio and involved: development of a seasonal adjustment methodology for labor market time series data, and the development of methods for estimation of local area unemployment statistics. This report briefly summarizes the work of Multi-Racio.

Seasonal adjustment and local area unemployment estimation techniques are important for the full elaboration of a modern national labor market information system. The information is crucial to the effective management of monetary and fiscal policy in a modern market economy. Furthermore, Hungary was recently asked to report on the level of development of these techniques when completing questionnaires as part of the negotiation over Hungary's application to join the European Union (EU). These procedures are expected as part of the minimal standard for labor market information systems in EU member states.

2. SEASONAL ADJUSTMENT OF LABOR FORCE TIME SERIES

Fluctuations in labor force time series may result from seasonal, cyclical, or trend reasons. In gauging the health of a labor market it is important to be able to sort out these causes of fluctuations. This information is crucial for economic planning by both government decision makers and private entrepreneurs.

After participating in a study visit in March, 1993, to the U.S. Bureau of Labor Statistics, the U.S. Census Bureau and Statistics Canada, in 1994 the Hungarian consulting firm Multi-Racio began to work on a World Bank financed project of seasonal adjusting labor force data under contract to the National Labor Center in Hungary.

Documentation of Multi-Racio's computer software development and design work for their proposed information system is contained in the following reports available in Hungarian:

1. The methodology of the internationally used seasonal adjustment methods and its application to domestic economic data series - (1994)

2. The Methodology of Seasonal Adjustment and a User Manual for Win-X11-ARIMA - (1995)

3. A proposal for setting up an information system aimed at seasonally adjusting national and county-level registered unemployment data - (1995)

4. An experimental application of X-11 ARIMA/88 to the registered unemployment time series - (1995)

The final project report is their fifth report in the series. An abridged English version of this report is given in Appendix A to this report with the title: *Seasonal Adjustment of Registered Unemployment Time Series in Hungary*

Multi-Racio has also prepared practical and understandable instructions for application of their proposed seasonal adjustment methods. Furthermore, their innovation in personal computer (PC) based software is truly remarkable. Multi-Racio has developed a PC based Windows version of the Canadian X-11 ARIMA system for decomposing seasonal, cyclical, and trend factors in time series. This system is being made available to the National Labor Center in Hungary. Multi-Racio will provide training in the proposed seasonal adjustment methods to analysts in the National Labor Center in the near future. Publication of seasonally adjusted time series based on these methods may begin in the very near future.

3. ESTIMATION OF LOCAL AREA UNEMPLOYMENT

Information on the condition of local and regional labor market conditions is crucial to the formation and effective application of employment policy. It is also necessary to satisfy more general political and economic management aims in modern market economy. The information is important for decisions about public finance and monetary policy.

Currently the *Labor Force Survey* which is conducted quarterly on a representative sample of households by the Central Statistical Office in Hungary provides estimates of national unemployment rates quarterly and county unemployment rates annually with an acceptable level of precision. The CSO definitions of unemployment conform to the International Labor Office (ILO) standard which is the same as that applied by the Current Population Survey (CPS) which is conducted monthly in the United States. The definition determines whether an out of work individual is able, available, and actively seeking work.

An alternative to the survey based estimates of unemployment are register based data. The unemployment register is operated by the national system of public labor centers which provide placement services, unemployment compensation, and perhaps referral to active labor policies. Estimates of unemployment based on register data are more timely, but fail to conform to the ILO definition.

Multi-Racio in their report *Development of A Small Area Unemployment Statistical System* present practical methods, based on techniques used by the U.S. Bureau of Labor Statistics to estimate both national and county unemployment rates monthly with a level of reliability equal to or better than that achieved by the CSO for quarterly estimates of the national unemployment rate based on the Labor Force Survey in Hungary. In addition, methods for estimation of unemployment in 190 small sub-county areas will be possible on an annual basis with an acceptable level of precision.

The methods and procedures, which are more clearly spelled out in the more detailed Hungarian version of the report, call for combining data from the LFS, the unemployment register, and the periodic census of the population. The criterion for selecting an algorithm for each area is to minimize the relative variance of the estimator or the coefficient of variation. The procedure involves combining small area unemployment estimation methods time series methods using Kalman filtering with the starting point for the variance estimates provided by a Jackknife method.

A practical and understandable handbook for these methods has been prepared, and personal computer based software has been developed. Multi-Racio will provide training in the proposed local area unemployment estimation methods to analysts in the National Labor Center in the near future. Publication of local area unemployment statistics based on these methods may begin in the very near future.

APPENDIX A

REPORT ON SEASONAL ADJUSTMENT OF LABOR FORCE TIME SERIES

entitled

Seasonal Adjustment of
Registered Unemployment Time
Series in Hungary

SEASONAL ADJUSTMENT OF REGISTERED UNEMPLOYMENT TIME SERIES IN HUNGARY

FINAL REPORT ON THE RESULTS OF ADAPTING THE SEASONAL ADJUSTMENT METHODOLOGY WITHIN THE "DEVELOPEMENT OF HUMAN RESOURCES" WORLD BANK PROJECT (EMPLOYMENT COMPONENT, SUB-COMPONENT B/1)

by

MultiRáció Ltd.

Budapest, June 1996

FOREWORD

Speaking of adapting seasonal adjustment two questions arise: Is this so important, considering Hungary's present state of affairs, that among the number of problems to be solved this particular one should merit resources? Can only those countries that are wealthier than Hungary afford these statistical systems?

We ourselves were fairly surprised in 1993 on our study visit paid to the Census Bureau of the Bureau of Labor Statistics (USA) and Statistics Canada that departments full of many highly skilled experts work continuously on seasonal adjustment of economic data, while they also keep tuning these procedures and further research in focus.

One evident explanation is indeed what the second question implies: these countries can afford using seasonal adjustment. However considering the tight budget limits set up in these countries, it turns out that these departments exist because they satisfy a vital need in a market economy. So the answer to the above questions may read as follows: The wealth of more well-to-do countries than ours may be attributed in part to developing, and continuously using, methods like the seasonal adjustment of economic data, i.e. the use of seasonal adjustment does not follow from being wealthy but, on the contrary, being wealthy follows from applying this sort of methods.

The main user of seasonally adjusted data is the government itself in market economies. One of the most important fields of use is the planning of the budget yearly, both on the expenditure and revenue sides. A good budget requires good input data. To extract trend in the case of data showing seasonal fluctuations, adjustment is unavoidable. The interim timing of budget related financial transactions assumes knowledge of the seasonal effects in time.

Therefore, an accurate identification of seasonal effects is demanded for efficient cash-flow control and their filtering out is necessary to establish trends. Poorly designed and prodigal budgets cost an order of magnitude more than maintaining an expert group for seasonal adjustment. This revelation and need prompted to create information systems for seasonal adjustment.

Foreword

We think that in the ongoing reform of public finances, aimed at shaping an efficient state, one of the key elements of a statistical system suplying reliable, well-defined data should be a subsystem that performs seasonal adjustments. One of the prerequisites of Hungary's joining the European Union is to set up a new, reliable and accurate statistical system. So our planned integration into the European Union also demands the establishment of the above mentioned subsystem.

In the field of labor statistics it is also very important to know the latest trends by cleaning the time series of the seasonal movements. Without this it is impossible to decide about things like whether growth is moderate or just normal for the off-season. To make the necessary preparations for handling unemployment and to make the appropriate decisions soon enough, it is important to adjust the time series when the conditions are given (i.e. the time series are long enough to get relible results).

In 1994, our company began work on a World Bank financed project to develop methods for seasonal adjustment of labor force data, after the above mentioned study visit to North America in 1993, as a contractor for the National Labor Center (in hungarian OMK).

The documents about our computer software development work, the design work for the information system, and our research work done under this contract in 1994, 1995, and the first half of 1996 and the tests of the information system to be developed are the following:

1. **The methodology of the internationally used seasonal adjustment methods and its application to domestic economic data series - report (1994)**
2. **The Methodology of Seasonal Adjustment and a User Manual for Win-X11-ARIMA - user manual (1995)**
3. **A proposal for setting up an information system aimed at seasonally adjusting national and county-level registered unemployment data - report (1995)**
4. **An experimental application of X-11 ARIMA/88 to the registered unemployment time series - report (1995)**

This present final report is based on these documents. The English version is an extract of the original Hungarian final report [5].

Miklós Banai, István Varga *Budapest, May 1996*

ACKNOWLEDGEMENT

We acknowledge and deeply appreciate the invaluable contributions made by György Lázár, Christopher O'Leary and Marietta Morry. We would also like to express our gratitude for their helpfulness and professional support.

CONTENT

INTRODUCTION

In market economies, month-to-month labor force data occur primarily because of seasonality. Consequently, in order to learn about the underlying trends in economic processes, one needs to remove fluctuations due to seasonal effects. This is what seasonal adjustment, a now standard statistical procedure, is aimed at.

The Federal Reserve Bank, the Bureau of Economic Research, the Treasury, the analyses on Wall Street, and advisors to the President on economic matters are all among the main consumers of seasonally adjusted data in the United States of America. Practically all studies of employment and unemployment at national or regional levels use seasonally adjusted data series. The time series published by executive agencies of the American government (about 5,000 in total) are all seasonally adjusted time series. Consequently, trends can be clearly seen from these data.

This study summarizes the adaptation and research work, aimed at applying seasonal adjustment to Hungarian labor force data according to international practice, which was done for the National Labor Center (OMK), within the framework of Word Bank Program "Development of Human Resources."

The adaptation and research work was based on investigations, made in 1994, into seasonal adjusting Hungarian unemployment data as described in our Report titled "The methodology of internationally applied seasonal adjustment and its application to domestic economic data series". To prepare this report, we carried out the following tasks.

We selected the fundamental procedure to recommend for seasonal adjustment in OMK, developed a user-friendly computer implementation, specified the process and methods of data evaluation, made plans for an information system, and we also started to implement and test the system.

The realization has been or is to be done in the following steps:

1. According to Report [1], we chose to recommend X-11-ARIMA/88, developed by Statistics Canada, as the fundamental procedure for seasonal adjustment for OMK. This procedure is used for the official seasonal adjustment of labor force and other economic data throughout North America. Successful Tests made on more than one hundred time series generated from the OMK database support this recommendation.

2. We obtained the source code for X-11-ARIMA/88 from Statistics Canada, on the

condition that we may use it in this program only and we may not disclose it to third parties. Having studied the code, we made a detailed plan for a user-friendly, computerized solution for seasonal adjustment. Since we could not quickly acquire a program module that identifies ARIMA models and estimates the parameters, we have undertaken to develop such a program module. The detailed plans are ready now, coding and testing are well in the works.

3. We have developed program modules (input data filtering, a flexible user interface, ARIMA model identifier, a model estimator and forecasting module, postprocessing text output files, and graphic visualisation) and tested them. We have also compiled a user manual [2].

4. We have selected more than one hundred 1995 time series from the OMK database and, for testing purposes, applied X-11-ARIMA/88 to them. This exercise demonstrated good applicability of the recommended procedure. We also recommend a standarized process for data evaluation which we propose as a customized checklist, based on the one used in BLS, adapted to the characteristics of Hungarian labor force time series [4].

5. We have planned an information system that would produce seasonally adjusted unemployment data both at the national and county levels [3]. Moreover, based on our investigations, we proposed elementary time series for composing national and county level time series. The seasonally adjusted versions of the main national and county-level time series can be prepared from these seasonally adjusted elementary time series [5].

We commenced to do the following in 1996:

1. Implementation and testing of the information system as specified by the plans. We now do seasonal adjustment monthly for both national and county level elementary time series. We also do direct seasonal adjustment of the aggregated principal time series in order to check adjustment by composition. In addition we adjust more than one hundred national level time series in each month.
2. Reliability tests are being done monthly along with the yearly needed post-revision due at the end of the year.
3. We are developing a program module to aid us in constantly evaluating the reliability of estimates. This will collect the necessary data from the output files of X-11-ARIMA and compute the specified reliability measures.
4. Plans for system enhancements include the following: First we plan to build in a new ARIMA segment which increases productivity by automatically finding the best fitting model, instead of requiring the user to choose among five time series models, and then performs seasonal adjustments. Secondly, we add intervention analysis which can handle changes in the time series that are due to administrative changes.
5. In the near future we will provide a seminar for OMK staff on the significance and the how-to of seasonal adjustment.
6. After carrying out these tasks, seasonal adjustment can be officially introduced by OMK in January, 1997. From 1997 on seasonally adjusted data may be published by the OMK. For the counties prime time series may be published, while for the nation S.A. time series on elementary variables and more than one hundred others will be available for publication.

1. REQUIREMENT FOR COMPUTER SOFTWARE TOOLS

We took into account the following requirements, as outlined in Section B) of the Introduction, while we were shaping system plans and we developed tools accordingly.

Enhancement of X-11-ARIMA tools:

The division between the program module carrying out the analysis and the user interface must be maintained as in the original version.

1. The analyzing module receives instructions through a configuration file. The configuration file is assembled by the user interface.

2. The menu-driven user interface should be

   ___ graphic,
   ___ standard,
   ___ multilingual (at least Hungarian and English),
   ___ equipped with on-line help

   (all in all: user-friendly). Programs running under the operating system MS-Windows usually confirm to these requirements.

3. A modernised display of text-formatted results. MS-Windows wordprocessors and spreadsheets must be capable of using the output files.

4. For the graphical display of results, there were two choices:

   a) Use general purpose, MS-Windows based graphing capabilities. Implementation of widely used file formats are needed to support this.
   b) Use a custom version graphical unit which is integrated to the analyzing module. The unit certainly must provide MS-Windows basic functionality (use of clipboard, printing control, zooming, etc.).

   We chose the second possibility.

5. The preparation of the configuration file that contains instructions should be independent of the user interface.

6. Capability for batch runs, i.e. when automatically several analyses are done

according to pre-made configurations. In this case user interaction is needed for the preparation of configurations and the evaluation of the results.

7. The analyzing program should be modular so that the system can be improved, e.g. adjustment of extreme points (e.g. outliers).

8. Preparation of a user manual which solves syntactical problems in the user manual of X-11-ARIMA.

Processing results:

A condition for applicability in a standard and controllable way is a checklist, which may serve as a starting point for the official publication, and which is adapted to the domestic environment. The prototype was the checklist used in BLS. The checklist was developed, simultaneously with program development, by utilising experiences with more than a hundred time series made from OMK data (see below).

The program system and the checklist must be such that a user with minimal computer skills only, after running the program, must be capable of filling out the checklist properly.

When the source code of X-11-ARIMA/88 was granted to us by Statistics Canada, all preconditions for fulfillment were met. Since Statistics Canada, in an initiative, gave a commission for rewriting X-11-ARIMA mainframe FORTRAN code into the C language, with their kind permission, the present program version is compiled with this C language code which also contains options not available in the original version. For example, the present program offers five, instead of four, ARIMA models for model identification.

The user manual [2] gives detailed information about the analysing system as it was operational at the end of 1995.

The user manual consists of two main parts.

Part I contains the theoretical discussion of analyzing methods and procedures used by the program. Consequently, we present identification of seasonal ARIMA models, the estimation procedure for the parameters and forecasts obtainable with the classic Box-Jenkins methods [6]. This section contains also a detailed algorithmic description of the

ARIMA module of the program which allows for an interpretation and assessment of results. Method X-11-ARIMA/88 is presented in a similar manner.

Part II contains the user manual itself which is also available as on-line help.

The computer program development was made in a close cooperation with HT-Szoft Ltd. Our work was greatly facilitated by the helpfulness of the Time Series Analysis and Research Division of Statistics Canada.

We especially acknowledge the contribution of the leader of the Division, Ms. Marietta Morry.

2. APPLICATION OF THE X-11-ARIMA/88 SEASONAL ADJUSTMENT PROCEDURE, FOR TESTING, ON THE DATABASE OF OMK

In accordance with our proposal for the seasonal adjustment of labor force data series described in the last part of the basic study [1] we applied the X-11-ARIMA/88 method on the time series selected from the register database of National Labor Center (OMK) and the checklist of BLS were used for the evaluation of the results.

We present here the results of the tests (in details see [4, 5]). The presentation will include:

- The decomposition of time series applied for seasonal adjustment, and the correlation matrix of the time series studied.
- The identification of the time series under investigation.
- Tables to overview the runs.
- The evaluation of the checklist of the runs.
- Figures.

2.1 Conditions for disaggregation to the seasonal adjustment of the number of registered unemployed

We have listed the conditions needed to calculate the elementary time series, the aggregate of which gives the basic time series (the number of registered unemployed), in the Appendix III. of the Study titled "A proposal for setting up an information system aimed at seasonally adjusting nation-wide and county-level registered unemployment data" [3].

These conditions are as follows:

a) the time series must be identifiable
b) the decomposed time series cannot show large positive or large negative correlations
c) the elementary time series chosen should possibly not be mere mathematical abstractions, instead the decompositions chosen should have clear economic and sociological content.

The correlation matrix of the seasonal factors of the time series in question allow for studying the correlations between the selectable time series. Based on this and considering the criteria formulated in condition c), we have selected decompositions as they follow:

1. Decomposition by sex and age
   - male to 25 years old (ffkor_a)
   - male from 26 to 45 (ffkor_b)
   - male from 46 to 55 (ffkor_c)
   - male from 55 years old (ffkor_d)
   - female to 25 years old (nokor_a)
   - female from 26 to 45 (nokor_d)
   - female from 46 to 50 (nokor_c)
   - female from 50 years old (nokor_d).

2. School leaver or not, decomposed by sex
   - school leaver male (fpkezdi)
   - non school leaver male (fpkezdn)
   - school leaver female (npkezdi)
   - non school leaver female (npkezdn).

3. Entrant code by age groups
   - non new entrant to 20 years old (k1nemuj)
   - non new entrant from 21 to 45 (k2nemuj)
   - non new entrant from 45 years old (k3nemuj)
   - new entrant to 20 years old (k1ujbe)
   - new entrant from 21 to 45 (k2ujbe)
   - new entrant from 45 years old (k3ujbe)
   - re-entrant to 20 years old (k1ujra)
   - re-entrant from 21 to 45 (k2ujra)
   - re-entrant from 45 years old (k3ujra).

The selected elementary time series form an exhaustive partition of the total number of registered unemployed. Correlation matrices in Tables 1-3 show that the various time series are highly correlated.

We use the evaluations of the revisions to study the different decomposition possibilities.

These basic time series had to be composed, using criteria b). Thus, obtained elementary time series were already suitable for studying the effects of decomposition.

Table 1
The correlation matrix for the time series in Decomposition 1.

	ffkor_a	ffkor_b	ffkor_c	ffkor_d	nokor_a	nokor_b	nokor_c	nokor_d
ffkor_a	1.00000	0.18741	0.27684	-0.23156	0.90973	0.32132	0.21628	-0.26454
ffkor_b	0.18741	1.00000	0.94305	-0.07967	-0.09397	0.86044	0.78304	-0.76317
ffkor_c	0.27684	0.94305	1.00000	-0.09548	0.03167	0.91773	0.91908	-0.73882
ffkor_d	-0.23156	-0.07967	-0.09548	1.00000	-0.04584	0.03242	-0.03688	0.49422
nokor_a	0.90973	-0.09397	0.03167	-0.04584	1.00000	0.06808	0.02171	0.06533
nokor_b	0.32132	0.86044	0.91773	0.03242	0.06808	1.00000	0.87859	-0.71134
nokor_c	0.21628	0.78304	0.91908	-0.03688	0.02171	0.87859	1.00000	-0.59666
nokor_d	-0.26454	-0.76317	-0.73882	0.49422	0.06533	-0.71134	-0.59666	1.00000

Table 2
The correlation matrix for the time series in Decomposition 2.

	fpkezdi	fpkezdn	npkezdi	npkezdn
fpkezdi	1.00000	-0.11102	0.97057	0.17851
fpkezdn	-0.11102	1.00000	-0.19689	0.50443
npkezdi	0.97057	-0.19689	1.00000	0.06626
npkezdn	0.17851	0.50443	0.06626	1.00000

Table 3

The correlation matrix for the time series in Decomposition 3.

	k1nemuj	k1ujbe	k1ujra	k2nemuj	k2ujbe	k2ujra	k3nemuj	k3ujbe	k3ujra
k1nemuj	1.00000	-0.51100	0.14443	0.19310	-0.19149	-0.09095	0.25941	-0.34486	-0.12196
k1ujbe	-0.51100	1.00000	0.28803	-0.14701	0.30962	0.11561	-0.05507	0.12144	0.03087
k1ujra	-0.14443	0.28803	1.00000	-0.38723	0.58667	0.93908	-0.46111	0.34263	0.90383
k2nemuj	0.19310	-0.14701	-0.38723	1.00000	-0.17711	-0.38163	0.48098	0.04021	-0.35737
k2ujbe	-0.19149	0.30962	0.58667	-0.17711	1.00000	0.62734	-0.74709	0.79338	0.62858
k2ujra	-0.09095	0.11561	0.93908	-0.38163	0.62734	1.00000	-0.58448	0.45124	0.96995
k3nemuj	0.25941	-0.05507	-0.46111	0.48098	-0.74709	-0.58448	1.00000	-0.70175	-0.64903
k3ujbe	-0.34486	0.12144	0.34263	0.04021	0.79338	0.45124	-0.70175	1.00000	0.52091
k3ujra	-0.12196	0.03087	0.90383	-0.35737	0.62858	0.96995	-0.64903	0.52091	1.00000

Taking into account reductions and the requirement that the time series should be exhaustive partition, furthermore they should sum up to the total number of the unemployed, in addition the correlation of the basic time series should not be "too high", we made up three sets of basic time series:

Version I. :

1. registered unemployed school leaver (opk),
2. registered unemployed non-school-leaver men, up to 45 years old (fnkab),
3. registered unemployed non-school-leaver men, over 45 (fnkcd),
4. registered unemployed non-school-leaver women, up to 45 years old (nnkab),
5. registered unemployed non-school-leaver women, over 45 (nnkcd).

Version II. :

1. registered unemployed school leaver (opk),
2. registered unemployed non-school-leaver, up to 25 years old (onka),
3. registered unemployed non-school-leaver men, between 26-55 years, plus registered unemployed non-school-leaver women, between 26-50 years (onkbc),
4. registered unemployed non-school-leaver men, over 55 , plus registered unemployed non-school-leaver women, over 50 (onkd).

Version III:

1. registered unemployed school leaver (opk),
2. registered unemployed non-school-leaver men, up to 25 years old (fnka),
3. registered unemployed non-school-leaver women, up to 25 years old (nnka),
4. registered unemployed non-school-leaver men, between 26-55 years, plus registered unemployed non-school-leaver women, between 26-50 years (onkbc),
5. registered unemployed non-school-leaver men, over 55 years, plus registered unemployed non-school-leaver women, over 50 (onkd).

Tables 4-6. show the correlation matrices for the three versions:

Table 4
The correlation matrix for the time series in Version I

	fnkab	fnkcd	nnkab	nnkcd	opk
fnkab	1.00000	0.67474	0.54919	-0.62975	-0.07213
fnkcd	0.67474	1.00000	0.82312	-0.32356	-0.00312
nnkab	0.54919	0.82312	1.00000	-0.30463	0.15354
nnkcd	-0.62975	-0.32356	-0.30463	1.00000	0.14026
opk	-0.07213	-0.00312	0.15354	0.14026	1.00000

Table 5
The correlation matrix for the time series in Version II

	onka	onkbc	onkd	opk
onka	1.00000	0.80254	-0.59659	-0.07563
onkbc	0.80254	1.00000	-0.58645	-0.09878
onkd	-0.59659	-0.58645	1.00000	0.11004
opk	-0.07563	-0.09878	0.11004	1.00000

Table 6
The correlation matrix for the time series in Version III

	fnka	nnka	onkbc	onkd	opk
fnka	1.00000	0.71283	0.87610	-0.66717	-0.15420
nnka	0.71283	1.00000	0.47952	-0.36074	0.09685
onkbc	0.87610	0.47952	1.00000	-0.58645	-0.09878
onkd	-0.66717	-0.36074	-0.58645	1.00000	0.11004
opk	-0.15420	0.09685	-0.09878	0.11004	1.00000

We are going to investigate all these three sets of time series in 1996. The investigation focuses, on one hand, on how much the disaggregated time series deviate from the directly adjusted basic time series, on the other hand, the size of revisions necessary when using various decomposition methods. We will decide, at the end of the experimental period in year 1996, with the help of results of these investigations on which decomposition methods we will use [5].

2.2 Time series of the registered unemployed tested in 1995 and in the first quarter of 1996 and their file identification codes

1. All registered

1. male and female together (osszrm)

2. Disaggregated by sex

2. male (ffrm)
3. female (norm)

3. Disaggregated by sex and age

4. male, up to 25 years old (ffkor_a)
5. male, 26-45 years old (ffkor_b
6. male, 46-55 years old (ffkor_c)
7. male, over 55 years old (ffkor_d)
8. female, up to 25 years old (nokor_a)
9. female, 26-45 years old (nokor_b)
10. female, 46-50 years old (nokor_c)
11. female, over 50 years old (nokor_d)

4. Disaggregated by education

12. primary school or lower (iv01)
13. technical school or comprehensive school (iv23)
14. comprehensive school or grammar school (iv456)
15. high school or university (iv78)
16. comprehensive school or grammar school plus high school or university (iv45678)

5. Disaggregated by sex and education

17. male, primary school or lower (ffiv01)
18. male, technical school or comprehensive school (ffiv23)
19. male, comprehensive school or grammar school (ffiv456)
20. male, high school or university (ffiv78)
21. male, comprehensive school or grammar school plus high school or university (fiv45678)
22. female, primary school or lower (noiv01)
23. female, technical school or comprehensive school (noiv23)
24. female, comprehensive school or grammar school (noiv456)
25. female, high school or university (noiv78)
26. female, comprehensive school or grammar school plus high school or university (niv45678)

6. Disaggregated by sex, age and education

27. male, up to 25 years old, primary school or lower (fkaia)
28. male, up to 25 years old, technical school or comprehensive school (fkaib)
29. male, up to 25 years old, comprehensive school or grammar school (fkaic)
30. male, up to 25 years old, high school or university (fkaid)
31. male, up to 25 years old, comprehensive school or grammar school plus high schoolor university (fkaicd)
32. male, 26-55 years old, primary school or lower (fkbcia)
33. male, 26-55 years old, technical school or comprehensive school (fkbcib)
34. male, 26-55 years old, comprehensive school or grammar school (fkbcic)
35. male, 26-55 years old, high school, university (fkbcid)
36. male, 26-55 years old, comprehensive school or grammar school plus high school, university (fkbcicd)
37. male, over 55 years old, primary school or lower (fkdia)
38. male, over 55 years old, technical school or comprehensive school (fkdib)
39. male, over 55 years old, comprehensive school or grammar school (fkdic)
40. male, over 55 years old, high school or university (fkbid)
41. male, over 55 years old, comprehensive school or grammar school plus , high school or university (fkbicd)
42. female, up to 25 years old, primary school or lower (nkaia)

43. female, up to 25 years old, technical school or comprehensive school (nkaib)
44. female, up to 25 years old, comprehensive school or grammar school (nkaic)
45. female, up to 25 years old, high school or university (nkaid)
46. female, up to 25 years old, comprehensive school or grammar school plus high school or university (nkaicd)
47. female, 26-50 years old, primary school or lower (nkbcia)
48. female, 26-50 years old, technical school or comprehensive school (nkbcib)
49. female, 26-50 years old, comprehensive school or grammar school (nkbcic)
50. female, 26-50 years old, high school or university (nkbcid)
51. female, 26-50 years old, comprehensive school or grammar school plus high schoolor university (nkbcicd)
52. female, over 50 years old, primary school or lower (nkdia)
53. female, over 50 years old, technical school or comprehensive school (nkdib)
54. female, over 50 years old, comprehensive school or grammar school (nkdic)
55. female, over 50 years old, high school or university (nkdid)
56. female, over 50 years old, comprehensive school or grammar school plus high school or university (nkdicd)

7. School leaver or not

57. school leaver (pkezdi)
58. non school leaver (pkezdn)

8. School leaver or not, disaggregated by sex

59. school leaver male (fpkezdi)
60. non school leaver male (fpkezdn)
61. school leaver female (npkezdi)
62. non school leaver female (npkezdn)

9. Type of compensation

63. unemployment benefit (mkj)
64. income support (jpt)
65. benefit for school leavers (pks)
66. no-compensation (nemrv)

10. Type of compensation disaggregated by sex

67. Male with unemployment benefit (fmkj)
68. Male with income support (fjpt)
69. Male with benefit for school leavers (fpks)
70. Male with no-compensation (fnemrv)
71. Female with unemployment benefit (nmkj)
72. Female with income support (njpt)
73. Female with benefit for school leavers (npks)
74. Female with no-compensation (nnemrv)

11. Disaggregated by skills

75. Skilled worker (szmunk)
76. Unskilled worker (bsmunk)
77. Manager (vez)
78. Other white collar worker (eszell)

12. Disaggregated by skills and sex

79. Skilled worker, male (ffszmunk)
80. Unskilled male (ffbsmunk)
81. Manager, male (ffvez)
82. Other white collar male (ffeszell)
83. Skilled worker, female (noszmunk)
84. Unskilled female (nobsmunk)
85. Manager, female (novez)
86. Other white collar, female (noeszell)

13. Entrant code by age

87. New entrant male plus female, up to 20 years old (k1ujbe)
88. New entrant male plus female, 21-45 years old (k2ujbe)
89. New entrant male plus female, over 45 years old (k3ujbe)
90. New entrant male plus female, over 20 years old (k4ujbe)

91. Reentrant male plus female, up to 20 years old (k1ujra)
92. Reentrant male plus female, 21-45 years old (k2ujra)
93. Reentrant male plus female, over 45 years old (k3ujra)
94. Reentrant male plus female, over 20 years old (k4ujra)
95. Not a new entrant male plus female, up to 20 years old (k1nemuj)
96. Not a new entrant male plus female, 21-45 years old (k2nemuj)
97. Not a new entrant male plus female, over 45 years old (k3nemuj)
98. Not a new entrant male plus female, over 20 years old (k4nemuj)

14. Entrant code by sex and age

99. New entrant male, up to 20 years old (fk1ujbe)
100. New entrant male, 21- 45 years old (fk2ujbe)
101. New entrant male, over 45 years old (fk3ujbe)
102. New entrant male, up to 20 years old (fk4ujbe)
103. Reentrant male, up to 20 years old (fk1ujra)
104. Reentrant male, 21-45 years old (fk2ujra)
105. Reentrant male, over 45 years old (fk3ujra)
106. Reentrant male, up to 20 years old (fk4ujra)
107. Not a new entrant male, up to 20 years old (fk1nemuj)
108. Not a new entrant male, 21-45 years old (fk2nemuj)
109. Not a new entrant male, over 45 years old (fk3nemuj)
110. Not a new entrant male, up to 20 years old (fk4nemuj)
111. New entrant female, up to 20 years old (nk1ujbe)
112. new entrant female, 21-45 years old (nk2ujbe)
113. New entrant female, over 45 years old (nk3ujbe)
114. New entrant female, up to 20 years old (nk4ujbe)
115. Reentrant female, up to 20 years old (nk1ujra)
116. Reentrant female, 21-45 years old (nk2ujra)
117. Reentrant female, over 45 years old (nk3ujra)
118. Reentrant female, up to 20 years old (nk4ujra)
119. Not a new entrant female, up to 20 years old (nk1nemuj)
120. Not a new entrant female, 21-45 years old (nk2nemuj)
121. Not a new entrant female, over 45 years old (nk3nemuj)
122. Not a new entrant female, up to 20 years old (nk4nemuj)

2.3 The overview tables of X-11-ARIMA runs

We present below the results of the 5 runs in a summarized form. The tables contain 6 columns. The first column contains the file identification codes of the time series and the other 5 columns contain in order the main results of the runs. These are:

- the model form chosen is multiplicative or additive according to the 1. page of the checklist,
- whether the program fits an ARIMA model to the time series in the run corresponding to the model form chosen, if it is yes what is the form of the ARIMA model, and what is the type of the prior transformation of the time series (None means that there was no prior transformation, Log means that there was a logarithmic prior transformation of the time series, i.e. the ARIMA model was fitted to the logarithm of the time series),

  is the result of the run accepted or rejected by the checklist of the BLS.

File name	I. evaluation	II. evaluation	III. evaluation	IV. evaluation	V. evaluation
osszrm	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	multiplikatív (2,1,0)(0,1,1) Log fails
ffrm	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	multiplikatív (2,1,0)(0,1,1) None fails
norm	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	multiplikatív (2,1,0)(0,1,1) None fails
ffkor_a	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	multiplikatív (2,1,0)(0,1,1) Log acceptable
ffkor_b	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
ffkor_c	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
fkor_d	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails
nokor_a	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails
nokor_b	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
nokor_c	additive none fails	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None acceptable

nokor_d	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	additive (0,1,2)(0,1,1) None fails
iv01	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
iv23	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	multiplikatív (2,1,0)(0,1,1) Log fails
iv456	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) Log acceptable
iv78	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,2,2)(0,1,1) Log acceptable
iv45678	multiplikatív none acceptable	multiplikatív none cceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,2,2)(0,1,1) Log acceptable
iv01	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
ffiv23	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (2,1,0)(0,1,1) Log fails
ffiv456	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (2,1,0)(0,1,1) Log acceptable

ffiv78	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,2,2)(0,1,1) Log acceptable
fiv45678	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,2,2)(0,1,1) None acceptable
noiv01	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None acceptable
noiv23	additive (2,1,0)(0,1,1)None acceptable	additive (2,1,0)(0,1,1)None acceptable	additive (2,1,0)(0,1,1)None acceptable	additive (2,1,0)(0,1,1)None acceptable	additive none fails
noiv456	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) None acceptable
noiv78	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
niv45678	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) None acceptable
fkaia	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	multiplikatív (2,1,0)(0,1,1) Log fails
fkaib	additive none fails	additive none acceptable	additive none acceptable	additive none acceptable	multipliakaitív (0,2,2)(0,1,1) Log acceptable

fkaic	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) Log fails
fkaid	additive none fails	additive none fails	multiplikatív none fails	multiplikatív none fails	additive (0,1,1)(0,1,1) None fails
fkaicd	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) Log fails
fkbcia	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
fkbcib	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
fkbcic	additive (2,10)(0,1,1) None acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,2,2)(0,1,1) Log fails
fkbcid	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (0,2,2)(0,1,1) None fails
fkbcicd	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable
fkdia	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails

fkdib	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
fkdic	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
fkdid	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails
fkdicd	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
nkaia	additive none fails	additive none fails	additive none fails	additive none acceptable	additive none fails
nkaib	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	additive (0,1,1)(0,1,1) None fails

nkaic	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	additive (0,1,1)(0,1,1) None fails
nkaid	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none fails
nkaicd	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	additive (0,1,1)(0,1,1) None fails
nkbcia	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,10)(0,1,1) None acceptable
nkbcib	additive (0,2,2)(0,1,1) None acceptable	additive (0,2,2)(0,1,1) None acceptable	additive (0,2,2)(0,1,1) None fails	additive none fails	additive (2,1,0)(0,1,1) None fails
nkbcic	additive (0,2,2)(0,1,1) None fails	additive (0,2,2)(0,1,1) None fails	additive (0,2,2)(0,1,1) None fails	additive (0,2,2)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails
nkbcid	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
nkbcicd	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails
nkdia	additive none fails	multiplikatív none acceptable	multiplikatív none acceptable	additive none fails	additive none fails

nkdib	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	additive (0,1,1)(0,1,1) None fails
nkdic	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none fails

nkdid	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	additive (2,1,0)(0,1,1) none fails
nkdicd	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none fails
pkezdi	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	additive (0,1,1)(0,1,1) None fails
pkezdn	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) NOne fails
fpkezdi	multiplikatív none acceptable	multiplikatív none fails	multiplikatív none fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None acceptable
fpkezdn	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
npkezdi	multiplikatív none fails	additive none fails	multiplikatív none acceptable	additive none fails	additive (0,1,1)(0,1,1) None fails
npkezdn	additive none fails	additive none fails	additive none fails	additive none fails	additive (2,1,0)(0,1,1) None fails
mkj	additive none acceptable	multiplikatív none acceptable	additive none acceptable	multiplikatív none acceptable	additive none fails

jpt	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails
pks	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive none fails
nemrv	multiplikatív none fails	multiplikatív none fails	additive none fails	multiplikatív none fails	additive none fails
fmkj	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable
fjpt	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails
fpks	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive none fails
fnemrv	additive none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails
nmkj	multiplikatív none fails	additive none fails	additive none fails	additive none fails	additive none acceptable
njpt	additive none fails	additive none fails	additive none fails	additive none fails	additive (2,1,0)(0,1,1) None fails

npks	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive (0,1,1)(0,1,1) None fails	additive none fails
nnemrv	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails
bsmunk	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
szmunk	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	multiplikatív (2,1,0)(0,1,1) Log acceptable
vez	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails
eszell	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,2,2)(0,1,1) Log acceptable
ffbsmunk	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
ffszmunk	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	multiplikatív (2,1,0)(0,1,1) None acceptable
ffvez	additive none fails	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails

ffeszell	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,2,2)(0,1,1) Log acceptable
nobsmunk	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None acceptable
noszmunk	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable	additive (2,1,0)(0,1,1) None acceptable
novez	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None ails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails	additive (2,1,0)(0,1,1) None fails
noeszell	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (2,1,0)(0,1,1) Log fails
k1ujbe	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) None fails
k2ujbe	multiplikatív none fails	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
k3ujbe	multiplikatív none fails	multiplikatív none acceptable	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails
k4ujbe	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	multiplikatív none acceptable

k1ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
k2ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
k3ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
k4ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
k1nemuj	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	additive none fails
k2nemuj	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,1)(0,1,1) None fails
k3nemuj	multiplikatív none fails	multiplikatív none fails	additive none fails	additive none fails	additive none fails
k4nemuj	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,10)(0,1,1) None acceptable
fk1ujbe	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) None fails
fk2ujbe	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
fk3ujbe	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	mulitplikatív none fails	mulitplikatív none fails

fk2ujbe	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
fk3ujbe	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	mulitplikatív none fails	mulitplikatív none fails
fk4ujbe	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable
fk1ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
fk2ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
fk3ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
fk4ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
fk1nemuj	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none fails
fk2nemuj	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails

fk3nemuj	additive none fails	additive none fails	additive none fails	additive none acceptable	additive (0,2,2)(0,1,1) None fails
fk4nemuj	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (0,2,2)(0,1,1) None fails
nk1ujbe	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív (0,1,1)(0,1,1) Log fails
nk2ujbe	multiplikatív none fails	multiplikatív none fails	additive none fails	mulitplikatív none acceptable	mulitplikatív none acceptable
nk3ujbe	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	mulitplikatív none fails	mulitplikatív none fails
nk4ujbe	multiplikatív none fails	multiplikatív none fails	multiplikatív none fails	mulitplikatív none fails	mulitplikatív none fails
nk1ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
nk2ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
nk3ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails

nk4ujra	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
nk1nemuj	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	multiplikatív none acceptable	additive (0,1,1)(0,1,1) None fails
nk2nemuj	additive none acceptable	additive none acceptable	additive none acceptable	additive none acceptable	additive (2,1,0)(0,1,1) None fails
nk3nemuj	additive none fails	additive none fails	additive none fails	additive none fails	additive none fails
nk4nemuj	additive none fails	additive none fails	additive none fails	additive none fails	additive (2,1,0)(0,1,1) None fails

2.4 Evaluation of the measures of the checklists

We executed the test of the X-11-ARIMA/88 procedure with 5 runs based on the time series selected from the database of NLC (National Labor Cencer) (in Hungarian OMK). The first 4 runs were done on monthly time series shorter than 6 years long, while the 5th run was done on exactly 6 years long time series. We present the results of the test according to this composition.

2.4.1 Evaluation of the checklists of runs ended with September, 1995, proposal for the Hungarian version of checklist

We selected more than 170 time series containing national aggregates from the OMK regsiter database to test the seasonal adjustment method and to complete the conditions of its official application. However, it turned out that the time series cntaining data disaggregated by branch are not complete mainly for the starting years. Therefore, it is yet not possible to select time seies of five years long by branches. One can study this kind of series later when at least five year long time series will be available.

Finally, we executed the test using 125 time series in four runnings: (1.) includes the time period January, 1990 - December, 1994, (2.) the time period January, 1990 - March, 1995, (3.) the time period January, 1990 - June, 1995, (4.) January, 1990 - September, 1995. The results of these four runds are summarized as follows.

We ran the program X-11-ARIMA/88 in default setting and evaluated the outputs with the aid of the X-11-ARIMA CHECKLIST developed in the BLS to evaluate the runs of X-11-ARIMA for the seasonal adjustment of labor force series.

Choice of additive or multiplicative decoposition

Additive decomposition was chosen for 66 time series in all of the four runs, i.e. in a stable way, while multiplicative was best in a stable way for 47 time series define stability. The choice of model decomposition varied (i.e. it was unstable) for 13 time series: multiplicative model was chosen more ties for 7 time series while the contrary was for 5 time series (cf. The overview tables in section 2.3).

The choice of ARIMA model

The program fitted an ARIMA model for 20 among the 125 time series and for the remaining 105 time series. The ARIMA model fit was stable (i.e. the same ARIMA model was fitted in all runs) for 15 time series, while the fit was unstable (i.e. the program chose occasionallly, but the same ARIMA model in every case during the four runs) for 5 time series.

The fitted model types were:

(2, 1, 0)(0, 1, 1) None for 12 cases
(0, 1, 1)(0, 1, 1) None for 6 cases
(0, 2, 2)(0, 1, 1) None for 2 cases.

(None means that the program fitted the ARIMA model without the prior transformation of the original time series.) (Cf. The overview tables in section 2.3.)

The results of the seasonal adjustment according to the X-11-ARIMA/88

The acceptance criteria built in the program based on the so-called Q statistics were acceptable for 110 time series in all of the four runs. Thus, the seasonal adjustment was successful for 88% of the all time series under study, i.e. it turned out that the X-11-ARIMA/88 method is unambiguously applicable for the seasonal adjustment of the time series selected from the OMK's register database. (Cf. The arranged checklists of the runs in {4}.)

The results of the seasonal adjustment according to the BLS checklist

The BLS checklist formulates a stronger criteria as the condition of the pubication than the built-in criteria of the X-11-ARIMA/88. The program evaluates all the measures of BLS checklist so one can fill in the list with the aid of output of the program. By overviewing the checklists filled in the four runs we obtain the following results.

The acceptance of seasonal adjustment was stable for 67 time series (54% of all). The non-acceptance of seasonal adjustment was stable for 48 time series (38% of all). The results were not stable (sometimes acceptable, sometimes not) for 10 time series (8% of all). For 3 time series in the latter group, only the checklists of run 1 provided non-acceptable results, the 3 other yielded acceptable results.

In these cases, perhaps the time series used in run 1 were not long enough to allow a stronger criteria of seasonal adjustment acceptance be satisfied. It seems that as the time series got longer and longer the problem is resolved. Therefore, we can insert these three time series into the set of successful time series, thus can be obtained acceptable result for 70 time series (56% of the all) in a stable way.

We summarized the statistical measures of the checklists of 55 (=48+7) time series problematic in the sense of the BLS checklist in 4 tables arrnaged tothe 4 runs. (The statistical measures of problematic time series, Tables I.- IV. in {4}).

We present the numbers of failures of the statistical measuresin the BLS checklist for the problematic time series according to the runs in Table I. We list the conditions denoted by the symbols of the statistical measures in the left column of the table. We set first the conditions, denoted by an asterisk in the BLS's checklist, they must be true for the acceptance in the checklist.

Table I.

How many times are fail the conditions of the checklist for the problematic time series

Condition/Run	Run 1	Run 2	Run 3	Run 4
$Q \leq 1.0$	14	13	14	13
$M7 \leq 1.0$	39	35	37	35
F-Moving < 2.2	35	28	26	28
F-Stable > 7.0	40	38	41	38
$M4 \leq 1.0$	9	9	7	7
$MCD \leq 3$	12	12	12	12
$S \geq I$ or $I \leq 50$	13	7	8	5
O < CI	10	7	8	8

We collected the values of the measures from the conditions V. and VII. of the checklists (which are denoted by asterisk in the BLS checklist) for the unstable time series by runs in Table II.

Table II.
The values of measures belonging to the critical conditions of the checklist for the unstable time series (F-S=F-Stable, F-M=F-Moving)

Time series code/Run	Run 1	Run 2	Run 3	Run 4
nokor _c	F-S=6.994	OK	OK	OK
fkaib	F-M=2.24	OK	OK	OK
nkaia	F-S=6.71, F-M=2.95	F-M=2.96, M7=1.05	F-S=6.37	OK
nkbcib	OK	OK	F-S=6.29, F-M=7.62, M7=1.54	F-S=6.78, F-M=5.49, M7=1.32
nkdia	F-M=3.68, M7=1.09	OK	OK	F-M=3.15
ffpkezdi	F-M=2.22	F-M=3.04	F-S=7.0	F-S=3.97, M7=1.05
k2ujbe	F-M=2.511	OK	OK	OK
k3ujbe	F-S=6.52	OK	F-S=7.0	F-S=6.28
nk2ube	F-S=5.89	F-S=6.59	F-S=5.39, F-M=2.4, M7=1.15	OK

One can read from Table II. the limit values of the three critical measures and the first values next tothe limit values following from our investigation (the limit value is a minimum for F-Stable, it is a maximum for F-Moving and M7). These are:

Min F-S=3.97, the next F-S=5.39
Max F-M=7.62, the next F-M=5.49, the next to this F-M=3.68
Max M7=1.54, the next M7=1.32

Based on these values we haev two alternatives to weaken the BLS criteria of the critical measures both of which fit better to the time series under study than the BLS criteria (of course in our proposal the built-in acceptance criteria ($Q \leq 1.0$) of the X-11-

ARIMA/88 remain valid!). These are:

I. Q≤1.0, M7≤1.4, F-Moving<3.7, F-Stable>5.0,

II. Q≤1.0,M7≤1.6, F-Moving<7.7, F-Stable>3.5

One can collect the number of times the new criteria holds true with the aid of Tables I.- IV. in [4] containing the statistical measures of the problematic time series by runnings and thus the improvement can be determined. Table III. contains the numbers of times the conditions of case I. hold true, while Table IV. contains them for case II.

Table III.

The number of problematic tie series satisfying criteria I. by conditions and runs

Condition/Run	Run 1	Run 2	Run 3	Run 4
M7≤1.4	32	35	35	41
F-M<3.7	39	40	39	38
F-S>5.0	24	27	25	28
All 4 together	16	18	16	17

The set I. of conditions satisfies for 11 problematic time series in all of the 4 runs, thus accepting this proposal 81 time series are not problematic among the 125 time series, i.e. 65% of all.

Table IV.

The number of problematic time series satisfying the criteria II. according to the conditions and runs

Condition/Run	Run 1	Run 2	Run 3	Run 4
M7≤1.6	39	44	43	46
F-M<7.7	52	50	50	52
F-S>3.5	38	36	35	37
All 4 together	32	30	31	34

Set II. of conditions satisfies for 27 problematic time series in all of the 4 runs, thus accepting this proposal 97 time series are not problematic among he 125 time series, i.e. 77.6% of all.

We can choose between the two sets of conditions by studying later adustments (post-revisions). Now we suggest set II. of criteria for preliminary use in such a way that it is reasonable to study both the criteria of BLS and of set I. and II. for basic time series during the experiemental use in 1996.

Statistics M7 is designed to measure the presence of identifiable seasonality in the time series. If its value is greater than 1 then this signals that the seasonablity is not identifiable in the time series. However, the cut-off point of M7 was based on **10-year monthly series** and it corresponds to a combination of F-S and F-M values that indicate 50% distortion in the seasonal factor estimate. In this case, the M7 test statistic takes the following form [7]:

$$M7 = \sqrt{\frac{1}{2}\left(\frac{7}{F\text{-}S} + \frac{3\,(F\text{-}M)}{F\text{-}S}\right)}\,.$$

we insert the limit values of F-M and F-S given in set II. of the conditions, then we obtain the value 2.07 for M7, i.e. the limit value 1.6 for M7 is a stronger condition. On the contrary, we get the limit value, e.g. for F-M, F-M=3.64 from the above relation if we accept the M7=1.6 value.

The time series we studies are all shorter than 6 years. Thus, the exact procedure would be to derive the above relation again for monthy time series shorter than 6 years, based on the distortion test of M7. However, it is not reasonable to execute this study for the following two reasons:

1. The time series under consideration are getting longer and longer. They will be more than 7 years long from the planned date of official application of seasonal adjustment.
2. We will apply the intervention analysis in this year to handle the effect of administrative interventions which affect the shape of the time series. Then this method will improve the possibility to detect the stable seasonality in the time series by subtracting the intervention effects from the seasonal effects.

Our expectation is that getting the time series longer and longer and applying the intervention analysis, we can gradually disregard the weakened acceptance criteria for the M7 statistic and reaching the 10-year length we can recover the criteria of the BLS checklist.

The role of ARIMA model fit to the time series in this seasonal adjustment procedure is that one can forecast and backcast the time series and handle in this way the end point problems appearing by taking moving averages. The number of revisions can be reduced in this way. Since it turned out in these tests that the set of ARIMA models built in the program is not sufficient to describe the time series under consideration with ARIMA models. The solution of this problem could be a development of the program which includes a comprehensive ARIMA model identification, estimation and forecasting module instead of a selected set of ARIMA models. In this version of the X-11-ARIMA software the ARIMA model fit and the forecast and backcast of the time series would be done with this module, while the other module of the program produces the seasonal adjustment of the forecasted and backcasted time series with the X-11 procedure.

2.4.2 Evaluation of the checklists of runs completed for the quarter IV. of 1995

We got the data of quarter IV. of 1995 in January 1996. Thus, we supplemented our runs taking into account the data of the last quarter of 1995. Thus, this fifth run analyzed time series spanning the time interval January, 1990 - December, 1995. This interval includes six complete years, i.e. the time series under consideration are 6 years long.

We ran in this case the X-11-ARIMA/88 program again in default option, we used to be evaluated the outputs of the runs by using the X-11-ARIMA CHECKLIST applied in the BLS for the evaluation of the outputs of the X-11-ARIMA in seasonal adjustment of labor force series and the Hungarian checklist based on the former four runs. We summarize the results as follows.

Additive or multiplicative model form

Additive model form was obtained for 74 time series (61% of all) while *multiplicative* for 48 tme series (39%). The choice of model form changed since the latest run (i.e. it was unstable) in the case of 22 time series (18%) (cf. the "Overview table" of runnings). Consequently, the choice of model form has not changed essentially, only the number of unstable time series rose to 22 from 13 [1].

Choice of ARIMA model

The program did choose ARIMA-model for 74 cases (61%) of the 122 time series, while it did not for 48 cases (39%). It is a significant change comparing to the earlier runs, where the corresponding percentages were 16% and 84% respectively. The explanation of this may be that time time series under investigation had become 6 years long which means that the years after the starting year 1990 also provide 5-year long time series. Thus, the starting year when the shape of the time series are unformed is present with less weight in fitting the ARIMA model. The model type chosen are as follows:

(2, 1, 0)(0, 1, 1) None for 36 cases

(2, 1, 0)(0, 1, 1) Log for 8 cases

(0, 1, 1)(0, 1, 1) None for 15 cases

(0, 1, 1)(0, 1, 1) Log for 4 cases

(0, 2, 2)(0, 1, 1) None for 4 cases

(0, 2, 2)(0, 1, 1) Log for 6 cases

(0, 1, 2)(0, 1, 1) None for 1 case.

(None means that the program fits the ARIMA model to the time series without prior transformation, while Log means that with logarithmic prior transformation.) (CF. the "Overview table" of the runs in section 2.3)

The result of the sasonanal adjustment by the X-11-ARIMA/88

The seasonal adjustment produced was acceptable for 101 time series according to the criteria based on the Q statistical test built in the program, while it failed for 21 cases. Thus, the seasonal adjustment was successful for 83% of the all time series, therefore, the X-11-ARIMA/88 method was unambiguously proved applicable to provide the seasonal adjustment of the labor force series selected from the register database of the National Labor Center.

The result of the seasonal adjustment bythe checklist of BLS

The checklist of BLS requires a stronger acceptance criteria as the condition ofthe publication than the built-in acceptance criteria of the X-11-ARIMA/88. The program evaluates all the measures included in the criteria, thus, one can fill out the checklist by means of the output of the program. In the present run the BLS criteria is true only for 33 time series (27%), while it fails for 89 time series (73%). This result shows essential difference from the result of the former 4 runs, which canbe explained by the completion of the time series to 6 years long.

As a consequence of this completion, on the one hand, the starting year appears in the seasonal adjustment with less weight **and** on the other, an ARIMA model is more likely to fit the time series from the ARIMA model set. In these cases the program applies the X-11 method to the extrapolated time series and **the weight of the observations causing non-identifiable seasonality** (e.g. observations resulted by administrative interventions) **raises** in the extrapolated time series.

The critical measures of the BLS checklist (beside the Q statistic) are satisfied in the run as follows:

- The condition M7≤ 1.0 fails for 42 time series (34%) and obtains 80 time series (66%). For every case where the Q statistic fails there, the M7 statistic also fails.
- The condition F-M < 2.2 (measuring the moving seasonality) fails for 75 time series (61%) and obtains 47 time series (39%). This statistic fails the most frequently.
- The condition F-S > 7.0 (measuring the stable seasonality) fails for 40 time series (33%) and obtains 82 time series (67%).

The conditions of the **Hungarian checklist** proposed are satisfied as follows:

- The seasonal adjustments are accepted according to this checklist for 87 time series **(71% success)** and rejected for 35 time series (29%).

The critical measures of the BLS checklist (beside the Q statistic) with the new limit values prescribed in the Hungarian checklist are satisfied in the run as follows:

- The condition M7≤ 1.6 fails for 21 tme series (17%) and obtains 101 time series (83%).
- The condition F-M < 7.7 (measuring the moving seasonality) fails for 13

time series (11%) and fullfils 109 time series (89%).

- The condition F-S > 3.5 (measuring the stable seasonablity) fails for 19 time series (16%) and fullfils 103 time series (84%).

Further essential notes:

- We found 23 cases in the first 4 runs where the seasonal adjustments were rejected by the checklist criteria.
- In run 5 we found 35 rejected cases. Among the 23 and 35 time series 19 were the same.

Therefore, 4 time series became not problematic from the earlier problematic time series but at the same time 16 earlier not problematic became problematic.

Our expectation is that we can filter out the effects of administrative interventions by means of the **intervention analysis**, while we can obtain a more accurate description ofhte movement of the time series with proper ARIMA model identification and estimation. By implementing these new options inthe program, we will be able to improvethe seasonal adjustments significantly.

Finally, we list the time series of which the seasonal adjustment were rejected in run 5 by the criteria of the Hungarian checklist. These are by their file code:

ffkor_d	nkdia	njpt	fk3ujra	nokor_d	nkdib
vez	fk4ujra	noiv23	nkdid	novez	nklujra
fkbcib	jpt	klujra	nk2ujra	fkdia	pks
k2ujra	nk3ujra	nkaic	nemrv	k3ujra	nk4ujra
nkaicd	fjpt	k4ujra	nk3nemuj	nkbcib	fpks
k3nemuj	nk4nemuj	nkbcic	fnemrv	fklujra	npks
nkbcicd	nmkj	fk2ujra			

2.4.3 The average values of statistics M1-M11 describing the quality of the seasonal adjustment of the common 122 time series of shorter than 6 years long and of exactly 6 years long

For completing the evaluation of the runs we present the average values of the statistics M1-M11 used by the X-11-ARIMA procedure to describe the quality of the seasonal adjustment, and the average value of the Q statistic determined as a weighted average of the M1-M11 statistics for the 122 time series common in the 5 runs. One can think of these average values as the statistical measures of an average time series composed (by equal weights) from the 122 time series [7].

The quality control statistics M1-M11 measure the following quantities [1,6]:

M1: measures the relative contribution of the irregulars to the variance over a three-month span,

M2: measures the relative contribution of the irregular component to the variance of the stationary portion of the series,

M3: measures the amount of month-to-month change in the irregular as compared to the amount of month-to-month change in the trend-cycle,

M4: measures the amount of autocorrelation in the irregular as described by the average duration of run,

M5: measures the number of months it takes the average absolute change in the trend-cycle to dominate that in the irregular,

M6: measures the amount of year-to-year change in the irregular as compared to the amount of year-to-year change in the seasonal,

M7: measures the amount of stable seasonality present relative to the amount of moving seasonality.

The last four quality control statistics describe the year-to-year movement in the seasonal component.

M8: measures the size of the flutuations in the seasonal component throughout the whole series,

M9: measures the average linear movement in the seasonal component throughout the whole series,

M10: measures the size of the seasonal component fluctuations in the recent years (for four years before the last two years),

M11: measures the average linear movement in the seasonal component in recent years (for four years before the last two years).

The X-11-ARIMA/88 calculates onyhe M1-M7 measures for time series shorter than 6 years, it calculates all theM1-M11 measures for time series *at least six* years long (in accordance with thedefinition of M10 and M11).

The Q statistic is obtained as the weighted average of the M statistics for time series shorter than 6 years long as follows:

$$Q = 0.17M1 + 0.17M2 + 0.1M3 + 0.05M4 + 0.11M5 + 0.1M6 + +0.3M7,$$

and for the tme series at least 6 years long:

$$Q = 0.13M1 + 0.13M2 + 0.1M3 + 0.05M4 + 0.11M5 + 0.1M6 + +0.16M7 + 0.07M8 + 0.07M9 + 0.04M10 + 0.04M11.$$

Then the average values of the M1-M7 measures and the average value of Q in run 4 for the 122 time series (shorter than 6 years long) are:

M1 = 0.5896
M2 = 0.3690
M3 = 0.0843
M4 = 0.4329
M5 = 0.1792
M6 = 0.3860
M7 = 0.8025
Q = 0.4919

The average values of the M1-M11 measures and the average value of Q in run 5 for the 122 time series (6 years long) are:

M1 = 0.6214
M2 = 0.3751
M3 = 0.1117
M4 = 0.5439
M5 = 0.1940
M6 = 0.4013
M7 = 0.9571
M8 = 0.8786
M9 = 0.8545
M10 = 0.8801
M11 = 0.8671
Q = 0.5737

If we compare the values of these measures with the values of these measures obtained in Statistics Canada by seasonally adjusting 421 time series varied in length from 5 to 30 yeas as reported in [7], then we can see that the effect of the irregular component is smaller in our time series. This difference can be explained easily because the time series under consideration are the results of an administrative registration, i.e. they do not come from an estimation using survey data, thus, they do not include sampling error. On the contrary, the measures controlling the stability of the seasonal component and the change of its size show higher values. The explanation of these differences may be the more frequent interventions and as a result of an economy in transition. We expect the improvement of the quality ofthe seasonal adjustment, as we mentioned earlier, from the application of the intervention analysis and of proper ARIMA modeling.

2.5. Figures showing graphically the outputs of the seasonal adjustments of some representative time series

We present the result of the sesonal adjustment for 8 time series selected from the 122 time series tested in 5 runs in 1995 and 1996. We present the original time series together with the seasonally adjusted time series and the trend-cycle component produced by the X-11-ARIMA/88.

We note that similar figures were enclosed to the report [4] for all the 122 time series based on the run 4.

Insert Header

Figure 1

MULTI-RÁCIÓ

Figure 2

Figure 3

Figure 4

Figure 5

MULTI-RÁCIÓ

Figure 6

MULTI-RÁCIÓ

Figure 7

3. A PROPOSAL FOR SETTING UP AN INFORMATION SYSTEM AIMED AT SEASONALLY ADJUSTING NATIONAL AND COUNTY-LEVEL REGISTERED UNEMPLOYMENT DATA

3.1 Design issues

Upon organizing an information system which supplies seasonally adjusted data for registered unemployment, reliably and in a controlled manner, one needs to allow for the standpoints as follows:

1, Specify how input data needed for seasonal adjustment should be provided

2, Ensure software that works reliably for mathematical or statistical computations in seasonal adjustment

3, Guarantee manpower conditions to evaluate results

4, Decide whether seasonal factors needed in seasonal adjustment would be determined using current data or forecasted seasonal factors

5, Ensure that checking the obtained results is continuous and an appropriate documentation is made

6, Guarantee controlling and documenting reliability of adjustments for long periods of time

7, Provide for periodic revisions of seasonally adjusted data and the publication of revised data

8, Determine what level of administration hierarchy would perform the tasks prescribed in paragraphs 1-7

9, Schedule tasks prescribed in paragraphs 1-7 and consider time requirements

10, Plan how data publication is to be scheduled

Figure 8

To achieve tasks in paragraph 1-10 one needs to consider practical constraints like:

a) The present order of data acquisition and data processing
b) The method and the place for data archiving
c) The reliability of the mathematical-statistical procedures used
d) Time constraints, technical and personnel requirements of data processing

Constraints a) - c) are externally given which restrict possible solutions, while the technical and manpower requirements of constraint d) are primarily a question of finance and economic feasibility.

Below we first make a proposal how to carry out tasks in paragraphs 1-7. Then we present a possible data processing scheme in terms of regional structure and time schedules. Altogether four possible regional and temporal schemes are possible, but reliability consideration reduces these possibilities to two. We also present possible publication patterns. Regional and temporal structures are shown on flow charts as well.

3.1.1 Source and availability of input data

The main source for data to be processed are the closing data (stock figures) of regional offices. The closing date for them is the 20th of each month. Data are collected in county labor centers at the county level. Collected data are sent into OMK (National Labor Center) on floppy disks. The National database is completed on the 5th of each subsequent month. In addition a computer network (X-25 protocol) is also in use, though at the moment, because of expense and time considerations, floppy disk based data collection is preferred. (Data for a single month amounts to about 80-90 Mb. Transferring 1 Mb lasts about an hour on this network.) Data validation is done at the regional offices only. In OMK data is checked syntactically, for example, records with nonsense codes are excluded. The national and county level data yielding the time series to be adjusted are only extracts of the above mentioned client service database. Therefore there is a separate software program needed in order to produce indirect aggregated input data.

Some comments about archiving are now in order. At present, an archiving of the client service database, in a consistent and systematic way, is being done only in OMK. There is also an archive of the extracted database so that one should be able to do testing.

There exists a database of the time series for the time series to be used, and a filtering mechanism that produces current data, as well, for the same purposes.

3.1.2 Ensuring the reliability of the mathematical and statistical software

As we will see, the software needed for seasonal adjustment are the following:

a) X-11-ARIMA program

b) A validation program to be developed [3].

c) A filtering program that supplies data for time series

So far experience has shown that, as for the basic functions, X-11-ARIMA works satisfactorily. However, having the original source code examined, we discovered certain disfunctions. Therefore if one does not refrain from using only the basic functions, there arises a possibility of getting incorrect results. Note that, according to observations, the five basic built-in ARIMA models do not suffice to identify an ARIMA model for a significant number of time series.

3.1.3 Manpower requirement

The tasks, as detailed above, need minimally

a) one qualified computer expert who runs X-11-ARIMA with basic functions, the validation and the filter programs and maintains the programs

b) one assistant who does manual jobs such as data acquisition, filling out checklists and likewise manual things in connection with the program runs

c) one mathematician analyst who does model identification and confidence tests

3.1.4 Determination of adjusting seasonal factors

Having taken into consideration the methodology presented in Appendix 1 of Ref. [3]

and the alternatives provided by paragraphs 9-11, one may choose between two procedures of adjustment:

a) the determination of the seasonal adjustment factors is done by using all available information (adjusting time series simultaneously)

b) the determination of the seasonal adjustment factors is done by forecasting for a half year time period (forecasting time series adjustment)

3.1.5 Continuously verifying obtained results and documenting the verification

The reliability of adjustment can be checked with methods as given in Appendix 2 of Ref. [3]. Checking reliability depends mainly on the extent of necessary post-revision which is a result of the continuous data update. It is a result of the methodological analysis presented in Appendix 1 of Ref. [3] that the greatest revision is expected to occur after re-adjusting a time series forecast one month later. Consequently, to examine reliability one has to compare the adjustment factors for a time series, those referring to a given month, in subsequent months and look for their change. One may evaluate the reliability of a model with the help of the relative differences of the values coming from the original and later revisions; also with the distribution of the sign of these differences. Certainly one has to perform these tests for the differences between the current and forecasted adjustments when one uses the forecasting method.

As a result of the reliability tests one needs a "reliability checklist" for each times series examined. The current values, and their trends, would yield input for the internal reliability reports.

3.1.6 Checking reliability for longer periods of time

Reliability tests presented in Section 5 also include, in some sense, a long term testing for reliability. However one must extend these tests in two ways. Firstly record the relative differences of the original and the revised data in each month for a certain period of time, say for one year. (See Appendix 2 of Ref. [3].) One must evaluate the results of the annual (or semi-annual) revisions as well in a similar way.

Secondly revising the feasibility of the chosen disaggregation must also be done periodically. Appendix 3 of Ref. [3] contains the criteria which must be met to decide what indirectly adjusted time series should be aggregated to obtain a direct adjustment of aggregated time series. The fulfilment of these criteria must be checked each year. These long term reliability reports have to also be included in the internal reliability reports.

3.1.7 Data revision and publication

According to Appendix 1 of Ref. [3] the seasonal adjustment of principal unemployment time series, extended with new data, must be carried out annually. Though experience has shown that the revised data would deviate from the originally adjusted series only after the very first year, to ensure reliability a revision is made for five consecutive years in both the USA and Canada. Hungarian practice should be similar. The revised data would be published in the end of January each year.

3.1.8 Assignment of adjusting and control tasks to office hierarchy levels

An important issue in designing the information system that produces seasonally adjusted figures is what office hierarchy levels should execute the adjustment procedure and the related revision tasks. It would be natural for the adjustment of the principal national time series must be done by OMK since the national data originates here. Publication, documentation of national reliability measures and their regular evaluation must also fit in here. However it is well known that labor force issues, like the extent and characteristics of unemployment, are significantly different in the various regions in Hungary. So the seasonal adjustment of county level unemployment times series are also well worth the effort. It is especially important to distinguish between trend and seasonal changes of unemployment, i.e. a sensitive indicator of economic tendencies, in those counties which show critical unemployment figures.

Independent analyses of county unemployment figures at the county level may assume a special role in the future. One must consider manpower need, expenses, reliable functioning and the speed of processing and publication in centralised vs. decentralized versions of the system plan. Details can be found in Ref. [3].

periodically. Appendix 3 of Ref. [3] contains the criteria which must be met to decide what indirectly adjusted time series should be aggregated to obtain a direct adjustment of aggregated time series. The fulfilment of these criteria must be checked each year. These long term reliability reports have to also be included in the internal reliability reports.

3.1.7 Data revision and publication

According to Appendix 1 of Ref. [3] the seasonal adjustment of principal unemployment time series, extended with new data, must be carried out annually. Though experience has shown that the revised data would deviate from the originally adjusted series only after the very first year, to ensure reliability a revision is made for five consecutive years in both the USA and Canada. Hungarian practice should be similar. The revised data would be published in the end of January each year.

3.1.8 Assignment of adjusting and control tasks to office hierarchy levels

An important issue in designing the information system that produces seasonally adjusted figures is what office hierarchy levels should execute the adjustment procedure and the related revision tasks. It would be natural for the adjustment of the principal national time series must be done by OMK since the national data originates here. Publication, documentation of national reliability measures and their regular evaluation must also fit in here. However it is well known that labor force issues, like the extent and characteristics of unemployment, are significantly different in the various regions in Hungary. So the seasonal adjustment of county level unemployment times series are also well worth the effort. It is especially important to distinguish between trend and seasonal changes of unemployment, i.e. a sensitive indicator of economic tendencies, in those counties which show critical unemployment figures.

Independent analyses of county unemployment figures at the county level may assume a special role in the future. One must consider manpower need, expenses, reliable functioning and the speed of processing and publication in centralized vs. decentralized versions of the system plan. Details can be found in Ref. [3].

3.1.9 Scheduling tasks

Scheduling tasks may depend on co-operation between office hierarchies, the chosen adjustment methods, the revision procedures and publication obligations. Attention is paid to these considerations in Ref. [3].

3.1.10 Publication

Since publication is also dependent on the characteristics and time requirements of tasks to be carried out, we refer to the presentation of the versions, too.

3.2 Versions of the information system

Seasonal adjustment can be done either using forecasted or simultaneous factors. The realisation of possible processing and publication depends on this. Also, as indicated in Paragraph 8, centralisation vs. decentralisation, assigning tasks to office hierarchical and regional levels are also vital. We describe two possible scenarios below.

3.2.1 Centrally executed adjustment using forecasted seasonal factors

3.2.1.1 Schedule for January and July

Forecasted seasonal factors would be computed in these months. Certainly the forecasting is valid for the next five months.

According to this plan, the national and county unemployment time series would be seasonally adjusted in a central location, namely in OMK. Evidently the reliability control computations and analysis pertaining to this would be carried out here as well. The models assigned to the principal national time series and the forecasted adjusting factors can be published in a publicly available newsletter.

The schedule of processing is given in terms of days: (Time zero is the national and county level database is complete. According to present practice, this is the 5th of the current month.)

1st day: Supplementing disaggregated (elementary) time series, needed to produce aggregated national and county level time series (number of the registered unemployed) , with current data. This amounts to extending about 100-200 time series. (The time series for the registered unemployed is being produced by the composition of 4-6 component time series.)

2nd day: The extended time series is made suitable for input to adjustment.

3-4th days: Running X-11-ARIMA

5-6th days: Model identification to establish current month's seasonal factors.

7th day: The principal seasonally adjusted national and county level unemployment figures become available for publication. Since the above time limits are meant in work days, public announcement of unemployment rates is possible on the 7th working day after obtaining new register data.

8-9th days: Forecasting seasonal adjustment factors for the next five months.

10-15th days: Carrying out computations to evaluate current and long term reliability, as specified in Appendix 2 of Ref. [3].

16-21st days: Seasonally adjusting the chosen 100 principal national time series, computing forecasted adjustment.

Adjusting additional county level times series: Only the principal time series (the number of the unemployed and the related disaggregation series) of counties are adjusted regularly in this version. Certainly, if exceptional need arises, any other county level time series may be analyzed.

Hardware requirements:

Devices needed in OMK and county OMK offices as specified in Ref. [8], along with equipment required for safe storing and archiving of data.

Software requirements:

1. filtering program to extend time series
2. X-11-ARIMA used in seasonal adjustment
3. program module to do computations for reliability tests

Personnel requirements:

1. one qualified computer operator
2. one analyst/statistician
3. one assistant

3.2.1.2 Revision tasks in January

According to international practice, adjusted data must be revised once a year by reprocessing newly extended time series and the thus obtained results must be published as revised data. This revision must be carried out for four years, so the final adjusted figures are available after four revisions.

Consequently the principal nation-wide and county level times series must be adjusted in January in each year. In connection with this, the revised data must be compared to the original data and reliability tests have to carried out.

The above tasks require about 12 work days in the first year, while in the subsequent years, because of the growing number of yearly data to be adjusted, competition of this work needs a whole month (21-22 work days)

Hardware requirement:

see above

Software requirements:

1. X-11-ARIMA
2. reliability testing program

Manpower requirements:

1. one qualified computer operator
2. one analyst/statistician
3. one assistant

3.2.1.3 Schedule for months except January and July

Though using forecasted factors the process of revision and publication becomes simpler and faster, trustworthy work and the need for reliability prescribes the very same tasks to be done.

1st day: Supplementing disaggregated time series, needed to produce aggregated national and county level time series, with current data. Having known current data and forecasted adjusting factors, carry out seasonal adjustment.

2nd day: Publication of adjusted data.

At the same time:

2-7th days: Run X-11-ARIMA with current data, identify model and compute current (simultaneous) adjustment factors.

8-13th days: Carry out computations to evaluate current and long term reliability, as specified in Appendix 2 of Ref. [3]. Evaluate results, make reports about reliability.

14-20th days: Supplementing about 100 times series with current data and seasonally adjust them with forecasted factors.

The **hardware, software** and **personnel requirements** of the above mentioned tasks are identical to that of those in January and July.

3.2.2 Centrally executed adjustment using simultaneous seasonal factors

The schedule is identical in each month and it coincides with the January or July schedule in Version 1. Therefore publication of the first data can be made on the 7th working day after the availability of new data. Certainly the additional revision program must be executed in January.

The decentralized versions are presented in Ref. [3]. However, reliability considerations as it turned out there prefer the two centralized versions presented above.

3.2.3 Flow charts for the presented two versions of the information system

The following two flow charts demonstrate the hierarchical structure of the centralized versions of the information system of seasonal adjustment of registered unemployment data.

Diagram

Diagram

REFERENCES

1. The methodology of the internationally used seasonal adjustment methods and its application to domestic economic data series, report, MultiRáció Ltd. 1994 (in Hungarian)
2. The Methodology of Seasonal Adjustment and a User Manual for Win-X-11-ARIMA, user manual, Multi-Ráció Ltd. 1995 (in Hungarian)
3. A proposal for setting up an information system aimed at seasonally adjusting national and county-level registered unemployment data, report, MultiRáció Ltd. 1995 (in Hungarian)
4. An experimental application of X-11-ARIMA/88 to the registered unemployment time series, report, MultiRáció Ltd. 1995 (in Hungarian)
5. Seasonal adjusment of registered unemployment time series in Hungary, final report, MultiRáció Ltd. 1996 (in Hungarian)
6. G.E.P. Box and G.M. Jenkins: Time series analysis, forcasting and control, Holden-day, London 1970.
7. J. Lothian and M. Morry: A set of quality control statistics for the X-11-ARIMA seasonal adjusment method, Statistics Canada, October 1978, report-78-10-005 E.
8. An adaptation study of a local area unemployment statistical system, report, Multi-Ráció Ltd., November 1993. (in Hungarian)

APPENDIX B

REPORT ON ESTIMATION OF LOCAL AREA UNEMPLOYMENT

entitled

Development of
A Small Area Unemployment
Statistical System

DEVELOPMENT OF
A SMALL AREA UNEMPLOYMENT
STATISTICAL SYSTEM

Closing Report of the World Bank project

Written by
MultiRáció Ltd.
H-1117 Budapest, Szerémi u. 39.
September 1996.

FOREWORD

It is an indispensable precondition for an efficient market economy that the temporal and territorial changes in the principal economic figures be known. Such figures are some indicators of labour force market like the volume of labour force or the rate of unemployment.

Both governmental and entrepreneurantal circles must have information about the aforementioned trends. The government requires continuos information for an effective budget management. Forecasts from monthly or quarterly data of smaller areas supply precious information for planning annual budgets. In turn, an efficient temporal and territorial deployment scheme demands that this information should be available at all times. On the other hand, investors, for example, need data about smaller areas in order to reach well-founded decisions.

To meet these natural demands in market economies, information processing systems have been established working reliably. The budget reforms in Hungary aimed at reshaping the role that the state plays in the economy, also evidently demand information systems which supply temporal and territorial data with satisfactory reliability.

In addition, one of the prerequisites of Hungary's joining the European Union is to establish these systems. The questionnaire, handed over to the Hungarian government, in preparation for the negotiations about Hungary's join, clearly shows as well that the availability of this kind of information is self-evident in the EU member states.

There has also risen need for reliable monthly unemployment rates both for counties and small areas (i.e. the distance in commuters' reach, the areas of local officies).

The more significant clients and potential users of these information systems would be, first of all,

— governmental organisations (for example, to aid endangered or underdeveloped areas financially, to base area development projects upon the output)
— business people (since a creditable description of the labour force situation in the target area forms an integral part of modern investment policies)
— charities or NGOs (so that they can concentrate their efforts on areas in the most need).

The project, supported by the World Bank and aimed at a small area unemployment statistics system, was initiated in 1993 in order to satisfy these needs. We have payed a study visit to the Bureau of Labour Statistics of the USA and learnt the Local Area Unemployment System (LAUS) so that we could have a better understanding of the job to have been done. LAUS produces and publishes about 6500 small area labour force estimates monthly

The adaptation work has been started in Multi-Ráció Co-operative, commissioned by OMK. As a result, the following studies and papers have been prepared:

1. **A study of developing a small area unemployment statistical system, feasibility study, 1993.**

2. **The adaptation study of the cross sectional reggresion analysis and the Handbook method, study report, 1994.**

3. **Determination of county level unemployment rates for three counties and Budapest, study report, 1995.**

4. **Report about the developing and testing the Hungarian small area unemployment statistical system, report, 1996.**

The final report is based on the above.

Budapest, September, 1996 *Miklós Banai, István Varga*

INTRODUCTION

The unemploment rate is the key indicator of the economic climate at the national, regional and local level. Therefore in developed countries, appropriate organisations and analysing systems have been created to produce and publish regularly (usually monthly) unemployment rates for the whole economy and for its various regional parts.

The information produced in this way is used for the following purposes:

1. Planning the goverment budgets in states and municipialities.

2. Guiding the local employment and training plans and programs and in developing the different services.

3. Determining the degree of need, and calculating the share for a given area from the central funds.

4. Informing the decisions of the players in economy regarding investment, development, changing the production profiles, etc..

For example, the estimates of the Local Area Unemployment Statistics (LAUS) Program in the USA are used by private industry and individuals to compare and assess labor market develoments at the State and local level. They are used at all levels of government for planning and budgetary purposes, and by the Federal goverment in allocating Federal funds to States and local areas and in determining eligibility of atreas for benefits from various assistance programs. The LAUS program now encompasses estimation for more than 6500 geographic areas in the United States. On a monthly basis, estimates of employment and unemployment are prepared and published for all States and the District of Columbia, 329 metropolitan areas, 2049 small labor market areas, 3217 counties, and about 1000 cities including all cities and towns in New England and cities of 25000 population or more elswhere.

To achive all of the goals and tasks listed above requires estimating methods and procedures which are *reliable* and *cannot be influenced by subjective opinions*.

It is an important part of the development of the Hungarian labor information system to be able to produce the estimates of unemployment rates for sub-national areas. This development of course needs both theoretical and practical computer solutions of estimation.

The cooperation between OMK and MultiRáció has been aimed at setting up a statistical analysing system which is capable of yielding monthly unemployment rate estimates, their statistical reliability, on-line reports and summaries for about 180 labour force districts (the local officies), all the 19 counties and the capital of Hungary. This document summarizes the results of this work.

The methods involved in the estimating system close to the completion of developement have the following features.

The information system producing small area unemployment rates is an important component of the labour force information system in several countries (e.g. Canada, USA, England, Italy). The estimation methods applied in these countries are based on well-established, standard mathematical statistical estimating procedures. We adapted these methods into the small area estimates system developed by us.

It is a common feature of these methods that they estimate the two main component of labor force:

— the employed and
— the unemployed.

Consequently an up-to-date and reliable unemployment rate may be given for the areas is question. For accuracy, the estimation method uses the most information possible: National Labor Center (NLC) unemployment registration data together with population and labour force survey data collected by the Central Statistical Office (CSO).

The information system allows for studying employed or unemployed social groups from demographical, educational or sociological viewpoints.

A state-of-the-art time series analysing method is also under development which would further reduce the error of the estimates. This Kalman-filtering procedure is applied to improve reliability, e.g. in the Bureao of Labour Statstics of the USA as well.

We may state that the susbsystem yielding small area estimates, after testing in the year 1996, can be put to work from the January of 1997. This date is the January of 1998 for the time series analysing subsystem.

The development of the small area unemployment statistical systems proceeded as follows:

We prepared the feasibility study entitled "A study about developing a small area unemployment statistical system" in 1993.

Based on this work, we started to develop and to test our small area labour force estimating and analysing system by applying it to three counties and Budapest in 1994 and 1995. (See "Determination of county level unemployment rates for three counties and Budapest".)

We extended the system and the testing work for all the counties and areas assigned to NLC's local offices. We have worked out and developed an analysing module, via experimental testing work, which computes office level disaggregated estimates from county level estimates with the population claim and census share methods as described in the feasibility study. The estimating module evaluates the coefficient of variation (CV) for the estimates based on the the standard errors of county estimates.

The key to the system is to compute county level estimates. The solution is to combine the small area estimation method and the time series method using Kalman-filtering. The test version of the module uses all the estimation fuctions possible. We also developed the program block computing variance for the estimates with the Jackknife method. An extended investigation for biases has been done in order to find the optimal set of estimation functions for each county. These result in a covariance weighted average estimation function (a weigthed average of the estimators) whose variance and hence its CV is directly computable from the covariance matrix evaluated by the Jackknife method.

In the framework of the time series analysing method the procedure to be applied on county level has been specified and developed for testing purposes. To reliably determine the parameters of the models of each county we require labour force survey data and employment statistical data continuously.

The Hungarian version of the closing study consists of the following parts and documents:

- the outline of the tasks
- a description of the proposal of the feasibility study
- the documentation for the subsystem that applies the small area estimating functions
- the documentation for the study of biases
- the documentation for the averaged estimating function
- a detailed specification of the time series method
- documentation for the input time series currently available to us
- application of the time series method for a set of counties
- system documentation describing the data base, file-systems and data flow within the integrated estimating system under development
- a description of the "Hungarian Handbook-method", not currently implemented yet, as worked out in 1994's study.

1. A SHORT SUMMARY OF THE FEASIBILITY STUDY

In the first study [1] written in the framework of the World Bank project for the National Labor Center the following items were studied: (1) the possibilities available given the data inputs and statistical methods, of creating a system capable of producing monthly unemployment rates for 190 local areas, and the 19 counties and capital, and the whole country, (2) computing the statistical reliability of the estimates, and (3) providing on-line preconditioned reports about the labor market informations obtained in this way.

The report [1] summarizes the results of this study in the following structure.

In Part I. the problem to be studied is defined, the remaining parts of the report are then divided into three blocks.

The first block (Parts II-VI) presents the local area unemployment statistics of the US and studies the possibility of adaptating this system to Hungary. First we give a brief summary of the system containing the essential components. Then we overview it taking into acount the technical details. Two different methods are represented in sepatate parts: (1) estimation method based on time series analysis using the Kalman filter technique, and (2) Denton's benchmarking method. The adaptation study closes this block.

The second block (Parts VII-X) reviews the small area statistical methods applied in international practice and studies the possible adaptation in Hungary. The estimation methods and the different approaches to the sampling error calculation can be found in separate parts. The evaluation of the small area estimation methods from the aspects of applicability in Hungary can be also found in a separate part. The block is closed by the presentation of some applications of the reviewed methods in three countries (Canada, Sweden and Italy).

Then follows a summary of the methods suggested for application in Hungary which results from the two studies, together with the necessary inputs (Part XI).

In the last block (Parts XII-XIII) the preliminary results of the studies connecting to the computer feasibility of the suggested analyzing system are presented.

1.1. Formulation of the problem

The problem under consideration can be shortly formulated in the following way. One has to estimate the unemployment rate of a given area in a given time and then has to repeat this procedure regularly over equal time periods Then one needs the notion of the unemployment rate applied in the international practice. It is given by the formula:

Unemployment rate = (number of unemployed) / (labor force).

There are two principal ways to determine the unemployment rate for a given area in a given time, they are:

1. One estimates the above rate directly for the given area in the given time.

2. One estimates separately the numerator and the denominator of the above rate for the given area in the given time.

Since for a given time the labor force is equal to the sum of employed and unemployed, the second possibility implies two further subcases. Namely:

2a. One estimates directly the volume of the numerator and of denominator.

2b. One estimates the number of unemployed (numerator) and the number of employed, since labor force = unemployed + employed.

In the international literature methods have been developed for all of these possibilities and they are also applied in the practices around the world. The small area statistical methods presented in the second block in [1] belong to subcase 2a., while the methods in the first block in [1] applied in the USA belong to subcase 2b.

To accomplish the aims of this project we must study three levels as given areas. These are:

a) The country as a whole.
b) The capital and the counties.
c) The local labor areas.

One has to produce the unemployment rate estimates monthly and yearly and it is necessary to average and to benchmark the estimates thus provided in every year to be in accordance to the monthly and yearly data.

In the first step one has to study the estimation methods applied in international practice, solve the main components of the problem, and determine input requirements of these methods.

In the second step one has to study if the inputs of these methods are avaliable in Hungary, and, if not, when they will be available.

In the third step one has to study on what geographic levels the methods can be applied in Hungary.

The results of the study reported in [1] is summarized in the next section.

We note that now there exist four different unemployment rate notions for small areas in Hungary for application. Two of them are calculated by the CSO and two of them are calculated by the NLC. These are in order:

The unemployment rates of NLC:

1. Rate for counties:

Unemployment rate=(registered unemployed in the county)/(labor force in the county obtained by census share method from the national labor balance on 1. January of the year befor the actual year)

2. Rate for the local labor officies:

Unemployment rate=(registered unemployed in thearea of local office) /(labor force in the area of local office obtained from the last census (in 1990))

Both of them are calculated on mounthly bases.

The unemployment rates of CSO:

3. Rate calculated from the labor force survey (LFS) on mounthly bases:

   Unemployment rate = number of unemployed / labor force

4. Rate calculated from the labor force survey (LFS) on quarterly bases:

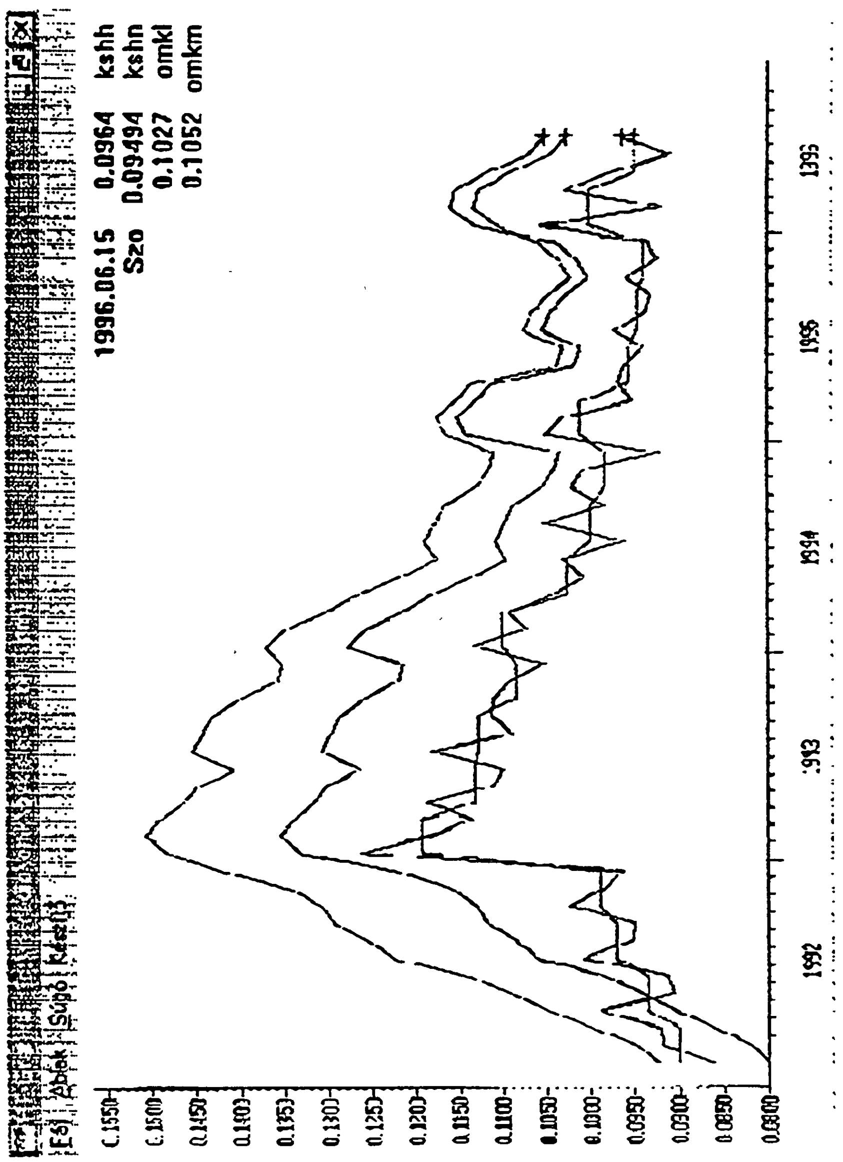
   Unemployment rate = number of unemployed / labor force

These unemployment rates are in accordance with the ILO concept but they do not provide reliable measures becaus of the small sample size for local areas.

We calculated the unemployment rates for the country according to these four different definitions and plotted the correspongding time series in Figure 1. In this figure the different unemployment rates are represented as follow:

— the light blue curve is the unemployment rate 1. of NLC,
— the green curve is the unemployment rate 2. of NLC,
— the dark blue curve is the unemployment rate 3. of CSO, and
— the red curve is the unemployment rate of 4. of CSO.

As we see the mounthly rates of NLC are essencially differ from each other and they lie above countinuously the rates of CSO with varying degree. Furthermore the mounthly rates of CSO show large fluctuations indicating significant variance of this estimations. Thus this figure also show unambigiously the task to be solved: one has to produce *one* consistent unemployment rate for local areas estimated, using available data bases, with acceptable reliability on mounthly basis (or quarterly basis).



The unemployment rate based on the LFS of CSO and calculated on mounthly (blue) and quarterly (red) bases. and the mounthly unemployments rates calculated in NLC (for local labor officies: green, for county level: ligth blue), respectively. January 1992.- June 1996.

1.2. The adaptation study of the small area estimation methods

1.2.1. The small area estimators

We have studied estimation methods developed specially for small area data. These estimation methods apply data for a given time (sectional data).

We pay attention to two essentional points by presenting the small area estimation methods applying sectional data (registration data and survey data referring to a given time). The first one show the need for development of the statistical estimation methods.

It often occurs that one has to study or compare subareas (small areas) using survey data which collected not for this purpose. Often budget constraints prevent observation of a sample which contains all the small areas deserving attention. Thus the evaluations of these surveys have the common aspect that the standard statistical estimation methods cannot apply, or alternatively their results have big errors.

Then one needs such methods which compensate in some way for these shortcomings. The so called *synthetic methods* used for a long time can be considered such methods. The characteristics of these methods are that they use the estimates of a bigger area to produce the estimates of the smaller area by supposing specific conditions hold true. These estimation methods have practical value when one wants to extend the study to such areas where there are no observations, but one can apply these methods to improve the estimates of small areas having only a few observations as well.

The other main component of the small area estimation problem relates to the inclusion of auxiliary information in the estimation procedure. It is an especially important question how to include administrative data (e.g. census, U.I. claims, health insurance data) to improve estimation. The "second generation" of the developed methods tries to solve this task. (It is worthwhile to mention the development of a survey design method, the so called 2-phase survey, which simulates the relation of the survey data and administrative data.)

The third generation of the methods (the methodological innovations of the second half of the last decade) try to find the answer for the question how to decrease the distortion of the deformed estimates with small variance in a way not increasing the variance of the estimates.

For Hungary we propose estimation methods which apply the data of registered unemployed, and the data of Labor Force Survey (LFS), and the data of Census and which have appropriate statistical characteristics (undistorted with relatively small variances).

We propose separate estimators for the numerator (unemployed) and the denominator (labor force) to obtain the unemployment rate.

The suggested corrected generalized regression estimator (see in Part IX of [1] the details) for the number of unemployed applies the registered unemployed as auxiliary information to give the estimates of unemployment for the given small area from the corresponding data of LFS. To give the estimate of labor force for the given area it applies the extrapolated population data of CSO.

The chosen method can produce efficient *quarterly* estimates for *the counties and the capital.*

Our proposal contains the condition of additivity to guarantee consistancy between estimates for the counties and the country. For the yearly benchmarking of the quarterly or monthly and yearly estimates Denton's method is proposed.

The method applicable for *the local labor areas* depends on whether the sample of LFS contains sufficient number of observations for the given area or not. If yes, then the same method can be applied as suggested for county level. If not, then we suggest the synthetic estimator. For the period of the estimation for local labor areas we suggest *the quarterly or monthly* cycles depending on the expected variances of the estimations and the demand of the user.

The detailed description of the suggested methods and their inputs can be found in Part IX and XI of [1].

1.2.2. The simultaneous use of more than one estimators

Small Area Estimators have been developed mainly to diminish variance of the estimation. For this they contain correcting factors in the formulae. However different estimators correct the result into different directions, and without a countrywide and full counting it is impossible to see which estimator is nearest to the true value. Obviously a solution of the problem is to form weighted averages of as many estimators as possible. The weights might be the inverse squared variances, as generally used in evaluation of measurements.

However in the present case this method might lead to serious errors. Namely the estimators are seriously interdependent. Since only the corrections differ in them. So there is a possibility that, by using unbalanced numbers of estimators of various kinds, the weighted average would be unbalanced too. Therefore a generalisation of the weighted averages is needed, which can automatically handle the serious correlations of estimators.

This has been done in such a way that the *centre* of the estimations is formed according to a generalised minimum principle for fluctuations. The method is described in details in Chap. 6 of [3]. Here we give only the final formula and a brief verbal argument.

Let us denote the result of the *i*th estimator by h_i and the centre by h. Then the covariance matrix of the estimators is

$$q_{ik} = \langle h_i h_k \rangle - \langle h_i \rangle \langle h_k \rangle .$$

Now, by requiring that the fluctuations *around the centre* be minimal, one gets

$$\sum_{ik} q^{ik} (h_i - h)(h_k - h) = \min .$$

where q^{ik} is the inverse of the covariance matrix. Hence one gets h and its variance according to Chap. 6 of Ref. 3.

It can be seen that for two very correlated estimators q^{ik} guarantees that the total weight of them in the centre will be only slightly higher than of one of them separately. Therefore the method corrects the errors may have been arisen from simultaneous use of too similar estimators. Of course, no method can guarantee errors from deliberate use of false estimators.

1.2.3. The adaptation study of the LAUS of the USA

The Local Area Unemployment Statistics (LAUS) in the USA uses three basic data input sources to produce the estimates. These are survey data (the monthly data of the Current Population Survey (CPS) and the monthly data of the Current Employment Statistics (CES)), the data of Unemployment Insurance (UI) system, and the data of censuses. The estimation procedures are as follows:

1. The monthly and yearly labor market estimates for the federal level and for one set of states (the "direct" states) are produced from the CPS by applying standard survey evaluation statistical methods. On the basis of the reliability measures of these estimates it was decided which federal states can apply these methods to produce the official estimates.

2. The monthly estimates for another group of the states are produced by applying models developed by means of time series analyzing methods. The latest version of these, the so called TEST models, was introduced to estimate the official data in 1994. These models are signal plus noise type models with structural time series as signal and noise and apply the Kalman filter for signal extraction [5, 6]. The reason for applying these models instead of the CPS estimates is that the statistical error of the model estimates is smaller than that of the CPS estimates.

3. The monthly estimates for the Labor Market Areas (LMA) inside the federal states are produced with the aid of the Handbook method. This method applies other data sources not in the CPS.

4. The monthly or quarterly estimates for smaller areas inside the Labor Market Areas are produced by proportional decomposition of the LMA estimates using the census and UI data.

The report [1] then summarizes the feasible estimating methods for Hungary and their necessary inputs obtained from the study of LAUS according to the hierarchy of area levels.

1.2.4. Proposal of methods for application in Hungary in the hierarchical order of teritories

Monthly estimates for Hungary: One can apply a TEST type model for the monthly time series of the Labor Force Survey (LFS) data of the Central Statistical Office (CSO) to obtain separately both the employment and unemployment estimates *on mounthly basis*. (Now the LFS estimates are base on quarterly basis.) The inputs are the time series of the LFS employment and unemployment monthly estimates, the time series of the employment data obtained from the monthly and quarterly labor statistics, the independent estimates of population and the time series of the monthly numbers of the registered unemployed. The unemployment rate is provided by the application of a TEST type model to the time series of the LFS monthly estimates. Then the statistical error of these new estimates will be smaller than that of the LFS estimates.

Monthly estimates for the capital and the counties in Hungary: In a similar way as for the nation we can apply TEST model versions for the time series of the LFS monthly estimates improved by comined application of the small area methods (so so called centre) for the capital and the counties to obtain the employment and unemployment data. The inputs are the time series of the LFS employment and unemployment monthly estimates improved by the centre methods for the capital and the counties, the time series of the employment data for the capital and the counties obtained from the monthly and quarterly labor statistics, the employment data for the capital and the counties from the census, the time series of the monthly numbers of the registered unemployed for the capital and the counties and the population data for the capital and the counties from census and time series of these data obtained by extrapolation. The unemployment rates are provided by the application of TEST models to the time series of the LFS monthly estimates improved by the centre methods for the capital and the counties. Then the statistical error of these new estimates will be smaller than those of the LFS estimates.

Note:

For demonstration how the signal plus noise estimating method works it is useful to express the model estimate of the signal in time t, $\hat{Signal}(t)$ (as produced by the Kalman filter algorithm) as weighted average of a model-based prediction of the signal, $\tilde{Signal}(t)$, based on historical relationships developed from past data, and the current improved LFS estimate, *Centre(t)*, corrected by a model-based estimate of sampling error, $\tilde{N}(t)$:

$$\hat{Signal}(t) - (1 - w(t))\tilde{Signal}(t) + w(t)(Centre(t) - \tilde{N}(t))$$

The wight, *w(t)*, given to the noise-corrected improved LFS may vary between zero and one. The size of this weight depends upon the accuracy of the model based prediction of the signal relative to the accuracy of the improved LFS in estimating the signal. If the LFS sample is very large, the noise component will be very small, i.e. $\tilde{N}(t) \cong 0$, and full weight, $w(t) \cong 1$, is given to the LFS ($\hat{Signal}(t) = Centre(t)$). At the other extreme, if the LFS sample is very small, the model estimate, which is based on a long historical series, will be much more accurate in relative terms, and full weight, $w(t) \cong 0$ is given to the model prediction ($\hat{Signal}(t) = \tilde{Signal}(t)$). The size of the weight, *w(t)*, depends in part on the reliability of the improved LFS, and in part on the model's ability to predict the signal The relationship is an inverse one: the lower the reliability of the improved LFS (the higher its standard deviation), the less weight is placed on the current improved LFS value in updating the signal estimate and vice versa.

Monthly estimates for the local labor areas: One can apply the "population-claims method" for the county estimates to obtain the employment data. The inputs are the total number of local labor areas in a given county, the monthly estimate of employment for the given county, the population and the number of employed in the local labor area from the census and population of the local area in the mounth of estimation obtained by extrapolation. One can apply a version of the population claims method for county estimates to obtain the unemployment data. The inputs are the monthly estimate of unemployment, stratified by age, for the county containing the local labor area, the number of registered unemployed including the number of unemployed youth who new labor force entrants, stratified by age, in the county containing the local labor area, in the given month and the number of registered unemployed including the number of new entrant youth unemployed in the local labor area in the given month. Then the unemployment rate for the given local labor area in the given month can be calculated by means of the relation

$$unemployed/(employed + unemployed)$$

using the estimates of employed and unemployed for the given labor area in the given month. The error of this estimate can be calculated from the errors of the corresponding estimates for the county.

An alternative way of estimation for local areas is the direct application of the small area estimators to the areas of local labor officies where the conditions of application is satisfied in a stable way in time. Because it turned out that the jackknife method of variance estimation cannot apply on labor office level we have to use another variance estimation method. This is a time series method which determines the statistical fluctuations in the time series of the mounthly values of an applicable estimator by removing the sistematical variations such as trend-cicle and seasonal variations.

Then one can use for official porposes that estimation beween the results of decomposition and small area estimation methods, for a given local area, which has better reliability measures.

APPENDIX A:

Problems with Weighted Least Squares

Take n estimators with values h_i, variances σ_i and correlation coeffitients r_{ik}, fluctuating somehow around an h. Then what happens if we look for some average H with weighted least squares

$$H = \frac{\sum_i \frac{h_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} . \tag{1}$$

First, H inherits all the distortions from h_i. They are somewhat averaged out, but using 2 very similar estimators instead of 1 the average is shifted. Nobody knows a priori. which estimator is nearer to the true h.

Second, H is not optimal statistically. To see this, let us calculate

$$\Delta^2 = \left\langle (h - H)^2 \right\rangle \tag{2}$$

Let us write:

$$h_i = h + \delta_i \tag{3}$$

where h is causal, δ_i is stochastic. Then

$$\Delta^2 = \left\langle \left(\frac{\sum_i \left(\frac{h}{\sigma_i^2} + \frac{\delta_i}{\sigma_i^2} \right)}{\sum_i \frac{1}{\sigma_i^2}} - h \right)^2 \right\rangle = \left\langle \left(\frac{\sum_i \frac{\delta_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} \right)^2 \right\rangle . \tag{4}$$

Now observe that

$$r_{ik} \equiv \frac{\langle h_i h_k \rangle - \langle h_i \rangle \langle h_k \rangle}{\sigma_i \sigma_k} = \left\langle \frac{\delta_i \delta_k}{\sigma_i \sigma_k} \right\rangle . \tag{5}$$

So

$$\langle \delta_i \delta_k \rangle = \sigma_i \sigma_k r_{ik} , \tag{6}$$

and then (4) reads as

$$\Delta^2 = \left\langle \frac{\sum_i \left(\frac{1}{\sigma_i^2} + 2 \sum_{i<k} \frac{r_{ik}}{\sigma_i \sigma_k} \right)}{\sum_{i,k}^{n} \frac{1}{\sigma_i^2 \sigma_k^2}} \right\rangle . \tag{7}$$

Now as an approximation, use a common σ^2 instead of individual ones, and an average $\tilde{r}$. (For estimators all r_{ik}'s are positive. So $\tilde{r}$ does not average out.) Then

$$\Delta^2 \cong \frac{\sum_i \frac{1}{\sigma^2} + 2 \sum_{i<k} \frac{\tilde{r}}{\sigma^2}}{\sum_{i,k}^{n} \frac{1}{\sigma^4}} = \sigma^2 \frac{\sum_i^n 1 + 2 \sum_{i<k} \tilde{r}}{\sum_{i,k}^{n} 1} . \tag{8}$$

So finally

$$\Delta^2 \cong \sigma^2 \frac{n - n(n-1)\tilde{r}}{n^2} = \sigma^2 \left(\frac{1}{n} + \left(1 - \frac{1}{n} \right) \tilde{r} \right) . \tag{9}$$

Therefore

$$\lim_{n \to \infty} \Delta^2 = \tilde{r}\sigma^2, \tag{10}$$

the deviation does not go to 0 if we increase the number of estimators used. The reason of the nonzero final limit is the (positive) correlation $\tilde{r}$. Using σ^{ik} as weights

$$h \cong \frac{1}{2} \frac{\sum_{i,k} \sigma^{ik}(h_i + h_k)}{\sum_{i,k} \sigma^{ik}} . \tag{11}$$

The correlations just appear explicitly and takes care of the problem.