COMENIUS UNIVERSITY, BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS
Mathematics of Economics and Finance

Application of Factor Models in Business Cycle Analysis in the Enlarged EU

MASTER THESIS

Bratislava, 2007

Ivana Bátorová

Application of Factor Models in Business Cycle Analysis in the Enlarged EU

MASTER THESIS

Ivana Bátorová

COMENIUS UNIVERSITY, BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS

Mathematics of Economics and Finance

Doc. Dr. Jarko Fidrmuc

BRATISLAVA, 2007

Modelovanie Hospodárskeho Cyklu v Rozšírenej EÚ pomocou Faktorových Modelov

DIPLOMOVÁ PRÁCA

Ivana Bátorová

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
KATEDRA APLIKOVANEJ MATEMATIKY A ŠTATISIKY

Ekonomická a finančná matematika

Doc. Dr. Jarko Fidrmuc

BRATISLAVA, 2007

I declare this thesis was written on my own, with the only help provided by my supervisor and the referred-to literature.

Bratislava, April 27, 2007 ....

Ivana Bátorová

Acknowledgment

I would like to express my deep sense of gratitude to Doc. Dr. Jarko Fidrmuc for his invaluable help and guidance during the elaboration of this thesis. I am grateful to him for having given me the support and confidence. And also I would like to express special thanks to my family, especially my father, mother and my sister, for their support during my study.

Abstract

Business cycles are characterized by the co-movement of large number of time series at the national and international level. This suggests that the business cycle is driven by relatively small number of common factors, which are not directly observed. The appropriate technique for discovering some of unobserved common factors is called factor analysis and it has become very popular in recent years. Other measures of synchronization in the case of business cycle, which is defined in frequency domain, are dynamic correlation and cohesion. The aim of this thesis is to introduce dynamic correlation analysis and various types of factor analysis and their application to macroeconomic data (GDP of 23 European countries and 5 other countries). We also try to find the common factors for describing the business cycle and finally we compare our results with the leading coincident indicator of the euro area business cycle published by Center for Economic Policy Research (CEPR).

Keywords: business cycle, dynamic correlation, cohesion, principal components, factor models

Abstrakt

Hospodársky cyklus, ako na národnej, tak aj na medzinárodnej úrovni, je charakterizovaný ako spoločný mechanizmus veľkého počtu premenných. Táto skutočnosť poukazuje na to, že hospodársky cyklus je v skutočnosti poháňaný malým počtom spoločných avšak nie priamo pozorovaných faktorov. Vhodnou metódou, ktorá poodhaľuje charakter nepozorovaných spoločných faktorov, je faktorová analýza. Jej prudký rozvoj zaznamenávame najmä v posledných rokoch. Ďalším spôsobom merania synchronizácie hospodárskeho cyklu, ktorý si zasluhuje zvýšenú pozornosť, je aj dynamická korelácia a kohézia, ktoré sú definované na kmitočtovej doméne. Cieľom tejto diplomovej práce je predstaviť analýzu dynamickej korelácie a rôzne typy faktorovej analýzy a aplikovať ich na dostupné dáta (HDP dvadsiatich troch európskych krajín a piatich krajín sveta). Taktiež sa budeme snažiť opísať hospodársky cyklus pomocou spoločných faktorov. Na záver porovnáme naše výsledky s všeobecne uznávaným indikátorom európskeho hospodárskeho cyklu.

Kľúčové slová: hospodársky cyklus, dynamická korelácia, kohézia, analýza hlavných komponentov, faktorové modely

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Eight countries of Central and Eastern Europe (CEECs), namely, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia, and also Malta and Cyprus entered the European Union in May 2004. Furthermore, these countries will all join the European Monetary Union (EMU) as soon as they satisfy the Maastricht convergence criteria. Slovenia, as the first country of the group, introduced the euro in this year. Our analysis, which is based on the sample of data ranges from 1Q1995 to 4Q2005, aims to answer the question whether this step as well as the plans of the remaining countries, was optimal from the point of view of the economic theory.

The successful enlargement of EMU requires that a criteria of a optimum currency area (OCA) have to be satisfied. OCA theory, originated by Mundell (1961), requires that the members of a monetary union to have some common characteristics. The main criterion of OCA is a high degree of synchronization between the monetary union's members. McKinnon (1963) and Kenen (1969) have contributed to OCA theory and they have added an additional OCA criteria which include labor mobility, flexibility of the labor markets, fiscal policy and enhanced trade and integration of financial markets. Earlier research showed that possibly some new member states have already achieved a comparably high degree of business cycle synchronization (see Fidrmuc and Korhonen, 2006), while the remaining criteria are generally not yet fulfilled even by the euro area (see De Grauwe, 2005).

The creation of the monetary union could be highly costly in terms of increased volatility of output and inflation. Accordingly, if the cost outweighs the benefits of the monetary union, EMU enlargement may be premature. It may be caused by asynchronous business cycles between the new member states and the euro area. Therefore, it is highly important to study the business cycle in Central and Eastern European countries and the euro area as an important prerequisite for successful EMU enlargement.

There is a growing literature on business cycle correlation between CEECs and the euro area. In a survey paper, Fidrmuc and Korhonen (2004) report 27 studies dealing with this issue. Darvas and Szapáry (2005) analyse the synchronization for GDP, industry, exports, consumption, services and investment. Artis *et al.* (2005) use concordance measure to acquire an information on whether the business cycles in NMS are in phase with business cycle of the euro area countries. Eickmeier and Breitung (2005) employ dynamic factor model to investigate how important shocks to the euro area business cycle are for NMS in comparison to the current EMU members.

This thesis tries to determine whether the Central and Eastern European countries are ready to join EMU. It assesses the current degree of business cycle synchronization in CEECs in comparison to the euro zone cycle. We put our focus at the countries, aiming to join the EMU. We show that the achieved degree of business cycle synchronization with the euro area is different among the membership candiates.

In the thesis, we outline two approaches for the description of business cycle synchronization that have become popular in the recent years: dynamic correlation analysis and factor analysis. However, there are still only few applications to the process of the euro area enlargement.

First, Croux *et al.* (1999) propose two measures of business cycle synchronization: bilateral dynamic correlations and their multivariate extension, termed cohesion. Both indicators are appropriate for analyses of the nature of dynamics in the co-movement because they account not only contemporaneous covariances, but also covariances at leads and lags. We compute dynamic correlation between output growth of individual countries and the euro area. It provides an information on existence of synchronization between the euro area and CEECs. Then we examine the cohesion as a measure of the degree of business cycle synchronization across seven group of states. Finally, for all European countries we compute ratio that was also suggested by Croux *et al.* (1999). This ratio, termed Borders Measure, describes the co-movement from geographical point of view.

The second approach, factor analysis, is the appropriate technique for discovering the driven factors of business cycle. The factor analysis is a statistical method which reduces a large number of variables that characterize the business cycle to a small number of factors.

The factor analysis has several advantages. Therefore, it is a preferred technique for discovering the latent structure of business cycle. It can cope with many variables without running into scarce degrees of freedom problems often faced in regression-based analysis. Furthermore, the factor models can eliminate idiosyncratic movements which possibly include measurement error and local shocks. Finally, the factor model is that the modelers can remain agnostic about the structure of the economy, which may be the subject of structural changes making standard regression analysis difficult or even impossible. Therefore, we employ the factor analysis to find the common factors for describing the business cycle and to estimate how CEECs are synchronized with EMU countries.

The master thesis is structured as follows. Chapter 2 proposes a measure of co-movement between many economic time series based on dynamic correlation. Chapter 2 also introduces a multivariate measure of co-movement, which is called cohesion. Finally, in order to study the problems of business cycle synchronization and dependence of co-movements from geographical point of view, we estimate dynamic correlation and cohesion of output growth for European countries. Chapter 3 outlines the classical, approximate and dynamic factor model and it is also dedicated to the description of the business cycle by the factor analysis. In Chapter 4 we briefly present the results of estimating the dynamic correlation, cohesion and also the factor model for a output growth for member countries of European Monetary Union (EMU) and Central and Eastern European countries (CEECs). We also assesses the transmission to New Member State. In Appendix we describe our data set and introduce EuroCOIN^{TM}, the leading coincident indicator of the euro area business cycle published by CEPR every month.

Chapter 2

Correlation Analysis

2.1 Introduction

The correlation analysis is the fundamental approach, which has been applied in the literature to study the degrees of synchronization between economic variables.

The most common measure of co-movement between time series is *classical correlation*, which is also used in the literature on the measuring the business cycle correlation between the euro area and the countries from Central and Eastern Europe. Fidrmuc and Korhonen (2006) review these publications and use the meta analysis that confirms the high correlation between the euro area and several CEECs.

The classical correlation, $\mathrm{corr}(x_t, y_t)$, between two random variables x_t and y_t is defined as:

$$\mathrm{corr}(x_t, y_t) = \frac{E(x_t y_t) - E(x_t)E(y_t)}{\sqrt{E(x_t^2) - E(x_t)^2}\sqrt{E(y_t^2) - E(y_t)^2}}$$

Unfortunately, the classical correlation is associated with two main drawbacks: Firstly, it does not allow for a separation of idiosyncratic components and common co-movements. Secondly, it is basically tool of static analysis that fails to capture any dynamics in the co-movement.

An alternative measure of synchronization in the case of business cycles is the *dynamic correlation.*

Croux *et al.* (1999) used the notion of dynamic correlation to construct a multivariate index of co-movement, called *cohesion.* The cohesion provides a measure of the degrees of co-movement within a group of variables or between two group of variables (*cross-cohesion*).

2.2 Dynamic Correlation

Let x and y be a two zero-mean real stochastic processes. Let $S_x(\lambda)$ and $S_y(\lambda)$ be the spectral density functions of x and y and $C_{xy}(\lambda)$ be the co-spectrum, $-\pi \leq \lambda \leq \pi$. The dynamic correlation is defined as

$$\rho_{xy}(\lambda) = \frac{C_{xy}(\lambda)}{\sqrt{S_x(\lambda)S_y(\lambda)}}. \tag{2.1}$$

The dynamic correlation lies between -1 and 1.

If two stochastic processes x and y are obtained by summing the waves of x_t and y_t within a given frequency interval, the dynamic correlation can be defined on the frequency band. Set $\Lambda_+ = [\lambda_1, \lambda_2)$ and $\Lambda_- = [-\lambda_2, -\lambda_1)$, where $0 \leq \lambda_1 \leq \lambda_2 \leq \pi$. Thus, the dynamic correlation within the frequency band Λ_+ is defined as

$$\rho_{xy}(\Lambda_+) = \frac{\int_{\Lambda_+} C_{xy}(\lambda)d\lambda}{\sqrt{\int_{\Lambda_+} S_x(\lambda)d\lambda \int_{\Lambda_+} S_y(\lambda)d\lambda}}. \tag{2.2}$$

In a particular case, if $\lambda_1 = 0$ and $\lambda_2 = \pi$, the $\rho_{xy}(\Lambda_+)$ is reduced to the static correlation between x_t and y_t, $\mathrm{corr}(x_t, y_t)$.

The dynamic correlation within the frequency band, as is defined in (2.2), can be used also for measurement of the co-movement of seasonal components of two economic time series, because we can select the frequency band of our interest and then evaluate the dynamic correlation within this frequency band.

2.3 Cohesion and Cross-cohesion

The cohesion, defined in frequency domain, is a measure of dynamic co-movement between time series. In bivariate case, the measure is reduced to the dynamic correlation (2.1). The cohesion is useful to studying problems of business cycle synchronization and to investigating short-run and long-run dynamic properties of multiple time series. It is an appropriate technique to obtain the facts on co-movements of macroeconomic variables at specified frequency band.

Let $x_t = (x_{1t}, \dots, x_{Nt})'$ be a vector of $N \geq 2$ variables and $w = (w_1, \dots, w_N)'$ be a vector of the non-normalized positive weights to the variables in x_t. The *cohesion* of the variables in x_t is defined as the weighted average of dynamic correlation between all possible pairs of series. Therefore, the cohesion is defined as

$$\mathrm{coh}_x(\lambda) = \frac{\sum_{i \neq j} w_i w_j \rho_{x_i x_j}(\lambda)}{\sum_{i \neq j} w_i w_j}. \tag{2.3}$$

Clearly $\mathrm{coh}_x(\lambda) = 1$ if and only if all the variables in x_t are perfectly co-moved at frequency λ. But the small cohesion index does not need to imply the small pairwise co-movements because it can be originated from large negative and positive covariances canceling out each other.

The measure of cohesion within frequency band $\Lambda_+ = [\lambda_1, \lambda_2]$ is analogously given by

$$\mathrm{coh}_x(\Lambda_+) = \frac{\sum_{i \neq j} w_i w_j \rho_{x_i x_j}(\Lambda_+)}{\sum_{i \neq j} w_i w_j}. \tag{2.4}$$

The cohesion index can be generalized to an index measuring the *cross-cohesion* between the $N-$ vector x_t and $M-$vector y_t. So the cross-cohesion of x_t and y_t at frequency λ is given by

$$\mathrm{coh}_{xy}(\lambda) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} w_{x_i} w_{y_j} \rho_{x_i y_j}(\lambda)}{\sum_{i=1}^{N} \sum_{j=1}^{M} w_{x_i} w_{y_j}}. \tag{2.5}$$

If the x_t and y_t are scalars, then the cross-cohesion is reduced to the dynamic correlation (2.1).

2.4 An Empirical Application of the Correlation Analysis

The correlation analysis, especially dynamic correlation and cohesion, has become very popular in recent years. In comparison with static correlation, dynamic correlation is a modern technique of measuring dynamic co-movement between time series.

In this section we try to analyse our data [1] in terms of business cycle correlation and mainly, we pay attention to find common features between the business cycles of Central and Eastern European countries and the euro area using the correlation analysis.

2.4.1 Classical Correlation

As a starting point, we compute the classical (static) correlation between output growth of the countries from Central and Eastern Europe and output growth of the euro area.

As it is shown in Table 2.1, it is apparent that only four countries from Central and Eastern Europe have positive business cycle correlation with the euro area. In particular, Hungary, Poland and Slovenia have a correlation coefficient with the euro area above 0.3. On the other hand, Lithuania, Slovakia and the Czech Republic stand out as countries with negative correlation coefficients.

	Czech Republic	Estonia	Hungary	Latvia	Lithuania	Poland	Slovakia	Slovenia
EA	-0.2373	0.0652	0.4717	-0.0725	-0.3905	0.4039	-0.3731	0.3283

Table 2.1: Classical correlation between the euro area and CEECs.

This findings are in line with study of Fidrmuc and Korhonen (2006) who argue that business cycle for Hungary, followed by Slovenia and Poland, has the highest correlation with the euro area among the new EU members. This study also points out that Lithuania and Slovakia trail behind other countries.

[1] More information about data are including in Appendix A.

2.4.2 Synchronization of Business Cycles in Europe

Dynamic correlation

In addition, we compute dynamic correlation between output growth of individual countries and the euro area. This analysis may provide an information on whether synchronization between CEECs and EMU countries may exist.

From table 2.2 it is apparent that the output growth in the countries of the euro area is on average more highly correlated with the corresponding output of the euro area than the corresponding variables in CEECs.

The average of the dynamic correlation for euro area countries at all frequencies and also at long run and business cycle frequencies (respectively 0.4267, 0.7172 and 0.5879) is much higher than for CEECs (0.1049, -0.0347 and 0.0656). It is not surprisingly that Germany (0.6018, 0.9057 at all and at the long run frequencies and 0.7474 at BC frequencies), Italy, Belgium and France have the highest dynamic correlation among the current EMU members. And the dynamic correlation in the Netherlands, Finland and Portugal at all frequencies are the lowest.

Among the CEECs, the dynamic correlations of output growth are relatively high for Hungary, Poland and Slovenia, but still lower than for the most countries of the euro area. These findings can be explained by tight trade linkages between Slovenia and euro area and also by big similarity to euro area industry in Hungary. On the other hand, the dynamic correlations between the Czech Republic and Slovakia and the euro area are slightly negative, whereas Lithuania trails behind the others.

Among non-European countries, Canada, followed by the USA, have the highest dynamic correlation with the euro area.

Our findings are in line with existing studies. We found out that Hungary, Slovenia and Poland have achieved relatively high degree of business cycle correlation with the euro area. This is also confirmed by meta analysis realized by Fidrmuc and Korhonen (2006). Also Darvas and Szapáry (2005) found out that GDP and industrial production in Hungary, Poland and Slovenia achieved a

Country	All freq.	Long run freq.	Short run freq.	BC freq.
Austria	0.4012	0.6965	0.3047	0.5980
Belgium	0.5065	0.7594	0.4239	0.6521
Germany	0.6018	0.9057	0.5026	0.7474
Spain	0.4155	0.6647	0.3342	0.5102
Finland	0.2947	0.5491	0.2112	0.4940
France	0.4936	0.8809	0.3671	0.7409
Italy	0.5913	0.8843	0.4956	0.7507
Netherlands	0.0998	0.2157	0.0620	0.1723
Portugal	0.3972	0.7658	0.2769	0.5552
Luxemburg	0.4651	0.8499	0.3395	0.6577
Sweden	0.3825	0.6065	0.3093	0.4997
Switzerland	0.5150	0.7326	0.4440	0.6259
Norway	0.2542	0.3630	0.2186	0.3692
Denmark	0.2855	0.5696	0.1927	0.3613
UK	0.3523	0.7524	0.2216	0.6141
Czech Republic	-0.0539	-0.3231	0.0340	-0.2698
Estonia	0.1155	-0.0390	0.1660	0.1130
Hungary	0.3360	0.5562	0.2640	0.4678
Latvia	0.1107	-0.2348	0.2235	0.0047
Lithuania	-0.1682	-0.5611	-0.0400	-0.3303
Poland	0.3377	0.4231	0.3098	0.4590
Slovakia	-0.0763	-0.5500	0.0783	-0.2804
Slovenia	0.2375	0.4506	0.1679	0.3612
USA	0.2848	0.4826	0.2202	0.3929
Canada	0.4307	0.7608	0.3229	0.6089
Japan	0.1063	-0.0150	0.1459	0.0979
Mean all	0.3044	0.4465	0.2580	0.3950
Mean Europe	0.2998	0.4312	0.2569	0.3858
Mean EA	0.4267	0.7172	0.3318	0.5879
Mean CEECs	0.1049	-0.0347	0.1504	0.0656
Std. all	0.2035	0.4455	0.1374	0.3209
Std. Europe	0.2107	0.4601	0.1426	0.3320
Std. EA	0.1476	0.2095	0.1320	0.1742
Std. CEECs	0.1920	0.4567	0.1190	0.3378

Table 2.2: Average dynamic correlation between output growth in individual countries and the euro area[2].

high degree of correlation with the euro area. We also confirm that the business cycle correlation is higher for the countries of the euro area.

[2]It is referred to (unweighted) average dynamic correlation over all/long run/short run/business cycle frequencies. Business cycle frequencies correspond to 4 to 8 years.

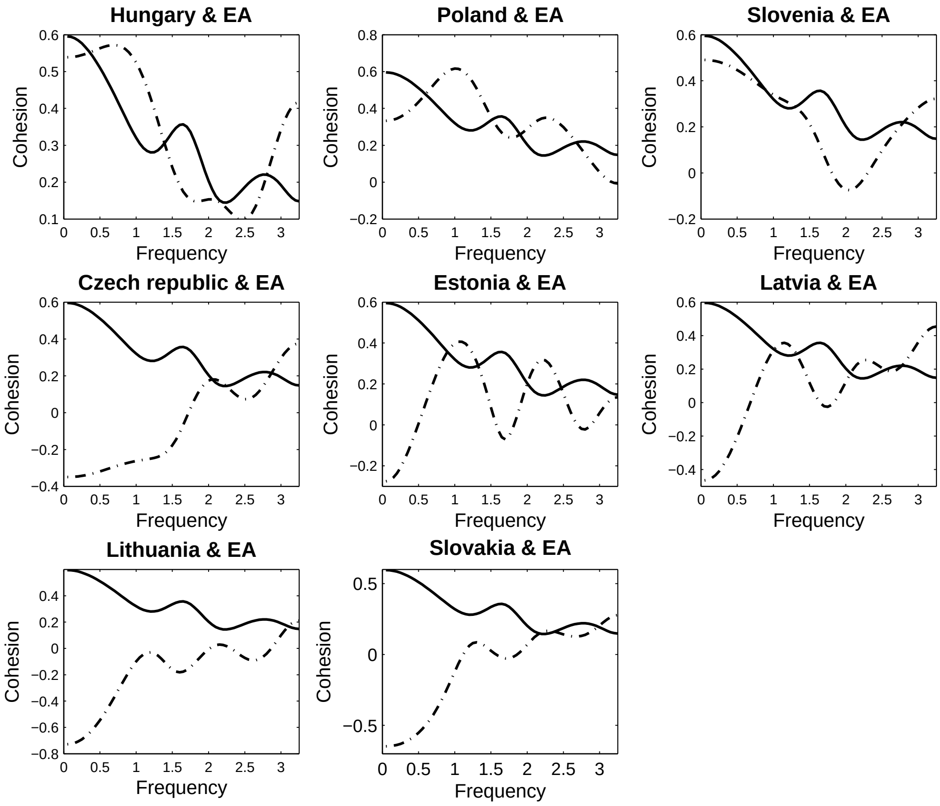

Figure 2.1: Cohesion of the euro area-EA (solid line) and dynamic correlation between output growth of the euro area and individual countries of Central and Eastern Europe (dot-and-dashed line).

The graphical comparisons of the dynamic correlation between euro area and countries from Central and Eastern Europe and cohesion of the euro area are illustrated in Figure 2.1. The findings implied from Table 2.2 are also confirmed by Figure 2.1 where the dynamic correlation and cohesion are represented at all frequencies. Well, from Figure 2.1 it is apparent that Hungary, Poland and Slovenia have a business cycle similar to cycle within the euro area at all frequencies. The other CEECs have a low degree of business cycle synchronization with the euro area. The difference at low frequencies is large, but at high frequencies the difference is much smaller. The dynamic correlation of euro area at business cycle frequencies is similar to dynamic correlation of Hungary, Poland and Slovenia. These findings are also apparent in Table 2.2 and support affirmation that these countries have a business cycle similar to the euro area.

Cohesion and cross-cohesion

In order to illustrate the synchronization across the countries, we compute the cohesion, which is the best technique for measuring of dynamic co-movement between time series.

	EU	EA	CEECs	CEECs 1	CEECs 2	V4	Baltic states
Average cohesion	0.3111	0.3116	0.1182	0.0463	0.2676	-0.0508	0.5899

Notes:
EA: Austria, Belgium, Germany, Spain, Finland, France, Italy, Netherlands, Portugal and Luxemburg.
CEECs: Czech Republic, Hungary, Estonia, Latvia, Lithuania, Poland, Slovakia and Slovenia.
CEECs 1: Hungary, Poland, Slovenia.
CEECs 2: Czech Republic, Estonia, Latvia, Lithuania, Slovakia.
V4: Czech Republic, Hungary, Poland, Slovakia.
Baltic States: Estonia, Latvia, Lithuania.

Table 2.3: Average cohesion for seven groups of countries within Europe.

Table 2.3 represents the unweighted average of the cohesion for seven groups of countries. The average cohesion over all frequencies across the euro area countries is reasonably high and it amounts to 0.3116. But the cohesion across all countries from Central and Eastern Europe is low (0.1182). Following the findings from previous section, we divide the CEECs into to two groups: "CEECs 1", the countries with high correlation with the euro area (Hungary, Slovenia and Poland) and others, "CEECs 2". Well, from cohesion measure, it is apparent that Hungary, Poland and Slovenia are less cohesive than CEECs, even though they are the most correlated with the euro area.

The cohesion across Baltic States (Latvia, Lithuania and Estonia) is the highest (0.5899) which suggests the high degree of business cycle synchronization across these countries. On the other hand, the synchronization across V4 countries (the Czech Republic, Hungary, Poland and Slovakia) is too small that is proved by slightly negative cohesion across them.

Well, our cohesion measures suggest greater synchronization across countries of the euro area than across countries from Central and Eastern Europe. However, the synchronization across Baltic states is the highest.

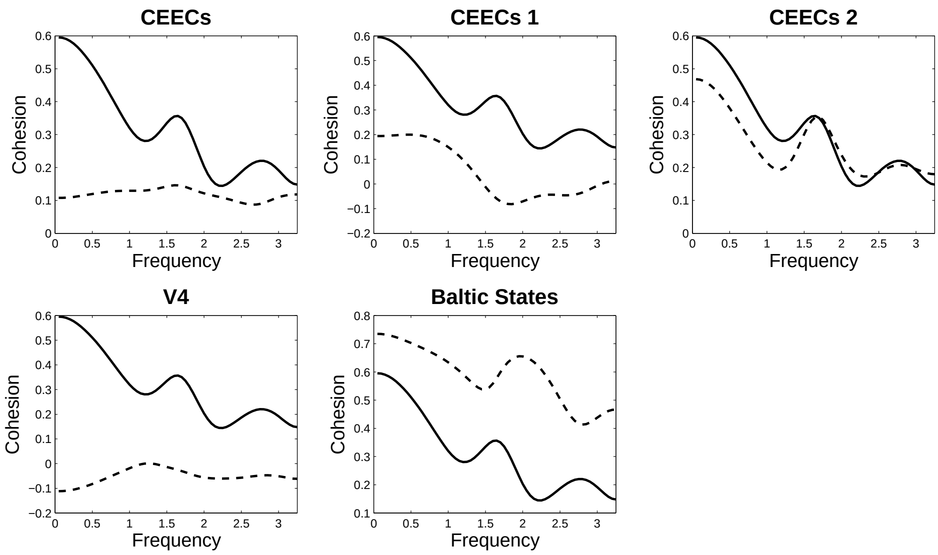

Figure 2.2: Comparison of the cohesion of the euro area (solid line) and the cohesion of the countries of Central and Eastern Europe (dashed line).

Table 2.3 shows only average cohesion over all frequencies for seven groups of countries defined before, but there are important differences at high and at low frequencies. Therefore, the Figure 2.2 illustrates a graphical representation of cohesion at all frequencies. The figure provides a comparison of the cohesion of the euro area countries and cohesion of other groups. The comparison with EU is left out of the figure, because the cohesions are very similar.

As expected, the new member states of EU are less cohesive than EMU countries, which in turn are less cohesive than Baltic states at all frequencies. The group of the countries from Central Europe (V4) is the least cohesive.

The difference between the euro area and CEECs is large especially at business cycle frequencies (around 1.5, corresponding to a period of about 4 years), but at short run frequencies the difference is much smaller. We conclude that as soon as synchronization of short cycles is concerned, the difference between the euro area and CEECs is small and non-significant, while the opposite holds for the business cycle and long run frequencies.

While it is difficult to interpret this behaviour, it seems that the conduct of common monetary policy (which influences especially economic development at high frequencies) will not pose a major problem for the CEECs. In turn, the differences with regard to the long-run development (low frequencies) reflect the convergence process of these countries.

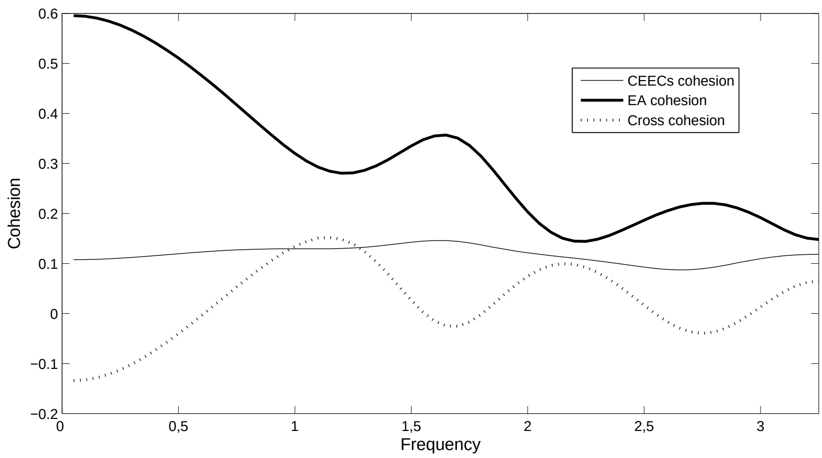

Figure 2.3: Within and Cross-cohesion of the group of CEECs and the group of the euro area states.

The relation between cohesion of the euro area and cohesion of CEECs is illustrated in Figure 2.3, where is also drawn the cross-cohesion between the euro area and CEECs. We can see that the cohesion of EMU countries is larger than cross-cohesion, but again the differences disappear in the short run. We conclude that, the EMU countries are more correlated with other EMU countries than with CEECs and on the contrary, CEECs are less correlated with other CEECs than with EMU countries.

2.4.3 Geographical Aspects of Business Cycle Fluctuations

For all European countries we compute ratio between average cohesion with neighbour states and average cohesion with all states. This ratio was suggested by Croux *et al.*(1999) as a measure of the extent to which “borders matters“. So this measure is defined as

$$\mathrm{BM}_i(\Lambda_+) = \frac{\mathrm{average}_{j\in C_i}\rho_{ij}(\Lambda_+)}{\mathrm{average}_{j\neq i}\rho_{ij}(\Lambda_+)}, \tag{2.6}$$

where C_i is the set of all neighbour states of state i and $\rho_{ij}(\Lambda_+)$ is the dynamic correlation between the state i and j at the selected frequency band Λ_+.

If the ratio is computed for long run frequencies, $\Lambda_+ = [0, \pi/4]$, then BM_i represents the border-correlation measure for the long run of country i. The short run border-correlation measure is obtained by the use of the frequency band $\Lambda_+ = [\pi/4, \pi]$.

Table 2.4 illustrates values of Borders Measure for European states. This ratio (2.6) has been computed for 22 European states with exception of UK, that has no neighbour state (10 countries of EMU, 8 states from Central and Eastern Europe and 4 other European countries). Border Measure has been computed for long run ($\Lambda_+ = [0, \pi/4]$) and short run frequencies ($\Lambda_+ = [\pi/4, \pi]$).

The results from Table 2.4 should be interpreted with caution, because the average co-movements with neighbours are not significantly different from the average co-movements with all states. The Baltic states are the exceptions. These findings are also suggested by Forni and Reichlin (1999) who argue that the core of the most integrated regions in Europe does not have national boundaries.

However, Table 2.4 shows that no difference in patterns emerges between short run and long run ratios and therefore the effect of cohesion with neighbours is the same for long run and short run period. Spain, France and Portugal are the countries which co-move more strongly with their neighbours than with other states, because the ratios are above 1.

Country	Long run BM	Short run BM	Neighbours
Austria	0.9594	0.9770	7
Belgium	1.2817	1.1585	4
Germany	1.2709	1.1366	9
Spain	3.1915	5.1007	2
Finland	1.5720	1.6358	2
France	2.1617	1.9709	6
Italy	1.5927	1.3879	4
Netherlands	-1.541	-0.9683	2
Portugal	2.5105	-	1
Luxemburg	2.0174	1.9927	4
Sweden	0.7777	1.7009	2
Switzerland	1.8078	1.7038	4
Norway	1.2497	1.7427	2
Denmark	2.0156	2.0020	1
Czech Republic	0.3378	-0.4243	4
Estonia	-	2.8956	1
Hungary	-0.2809	0.1513	3
Latvia	-	3.1098	2
Lithuania	-1.3745	-	2
Poland	0.4331	0.2677	4
Slovakia	0.4756	-0.5204	4
Slovenia	0.9282	0.7078	3

Table 2.4: Borders Measure BM_i for the long run and short run computed for 22 European countries. The number of neighbours is in the last column.

2.4.4 Conclusions from the Applications of the Correlation Analysis

The static correlation has some drawbacks, it fails to capture any dynamics in the co-movement, whereas dynamic correlation analysis is the appropriate technique to reveal the degree of synchronization between economic variables. Therefore we use the dynamic correlation and cohesion within Europe in our empirical analysis to illustrate the importance of the dynamic decomposition of co-movements.

Static correlations computed between the euro area and CEECs show that three states (Hungary, Poland and Slovenia) among the new members of EU have the highest correlation with the euro area. Unfortunately, these findings do not define if the countries are correlated with the euro area in the long run or in the short run. In order to specify it, we compute the dynamic correlation between the euro area and other countries. It is obvious that the countries from the euro area are higher correlated with euro area output than the CEECs. The results from dynamic correlation are in line with the conclusion from static correlation. Among the CEECs, Hungary, Poland and Slovenia have a relatively high dynamic correlation and they are stronger correlated with the euro area in the long run.

However, the static correlation as well as the dynamic correlation proves that Lithuania, the Czech Republic and Slovakia have negative business cycle correlation with the euro area.

Our empirical analysis also provides the information about the cohesion within Europe. The Baltic states are the most synchronized since they have the highest cohesion. On the other hand, synchronization across V4 is too small.

In addition, we provide results on the geographical structure of cohesion for Europe. We show, that the long run and short run ratios are similar for the countries and therefore effect of cohesion with neighbours is the same for long run and short run period.

Chapter 3

Factor Analysis

3.1 General Introduction

Factor analysis is a branch of statistics, but because of its development and extensive use in psychology the technique itself is often mistakenly considered as psychological theory. The method came into being specifically to provide mathematical models for the explanation of psychological theories of human ability and behavior.

In 1888, the concept of classical factor analysis was suggested by Galton, but the formulation is generally ascribed to psychologist Carl Spearman (1904) who developed the factor analysis for psychological purposes. He first charged that enormous variety of tests of mental ability could be explained by one underlying *factor* of general intelligence. Although nowadays we know that the Spearman's hypothesis on only one intelligence factor is not true, this research has became the driving mechanism of the development of new statistical technique. After 40 years, the Spearman's model was extended by Thurston (1945) and Lawley (1940) who were interested in the estimating of the factor loadings.

Factor analysis is different from many other statistical methods that are used to study the relationship between independent and dependent variables whereas the factor analysis is used to study the patterns of relationship among many dependent variables. The main goal of factor analysis is to discover something about the nature of the independent variables that affect the dependent ones, nevertheless those independent variables cannot be measured directly. So the information obtained by factor analysis is more tentative and hypothetical than the information received from direct observation of independent variables.

The main applications of the factor analytic techniques are: to reduce a large number of variables to a smaller number of factors for modeling purposes and to uncover the latent structure of a set of variables. Therefore, factor analysis is applied as a data reduction or structure detection method. It could also be used for identifying clusters of cases or outliers.

There are several different types of factor analysis. The simplest factor analytic technique is the principal components analysis (PCA). However, the most popular factor method is the classical factor analysis (CFA) which is more widely used than the principal components analysis.

3.1.1 Applications of Factor Analysis

The application of factor analysis has been chiefly in the field of psychology. Although the factor analysis was developed originally for analyses of mental tests, it is suitable not only for psychological purposes, but also for wider range of cases. For example, the analyses of the set of economic variables or set of physical measurement.

Applications of factor analysis in fields other than psychological purposes have became very popular since 1950. These fields include such varied disciplines as meteorology and medicine, sociology, political and regional science, biology and archeology.

Correspondingly, the first application of the factor model to general economic questions was in the marketing. More recently, factor analysis has been used in finance and macroeconomics.

3.1.2 General Definitions

This section introduces the basic terms used by factor analysis which are common for the principal component analysis and also for the classical factor analysis.

The output of factor analysis is generated as a table in which the rows are observed raw indicator variables and the columns are the *factors*. The cells in the table are called the *factor loadings* and they express the meaning of the factors induced from seeing which variables are most heavily loaded on which factor. The negative coefficient of factor indicates that the variable with negative factor loadings may be regarded as measuring the reversed aspect of the usual type of factor.

Factor loadings, also called *component loadings* in PCA, are the correlation coefficients between the variables and factors. The squared factor loading is the percent of variance in that variable explained by the factor. To get the percent of variance in all the variables accounted for by each factor, add the sum of the squared factor loadings for that factor and divide by the number of variables. This is the same as dividing the factor's eigenvalue by the number of variables.

The sum of the squared factor loadings for all factors for a given variable, which is called *communality*, is the variance in that variable accounted for by all factors. In a complete PCA, with no factors dropped, this will be 1.0, or 100% of the variance. When an indicator variable has a low communality, the factor model is not working well for that indicator and possibly it should be removed from the model. The communality exceeding 1.0 reflects too small sample or the researcher has too many or too few factors.

Uniqueness of variance is the variability of the variable minus its communality. It indicates the extent to which the common factors fail to account for the total unit variance of the variable. Sometimes it is convenient to separate the uniqueness into two portions of variance-the *specificity* and *unreability* of the variable.

Eigenvalue for a given factor measures the variance in all the variables which is accounted for by that factor. If a factor has a low eigenvalue, it is contributing little to explanation of variances in the variables and may be ignored as redundant with more important factors. An eigenvalue of the factor may be computed as the sum of it squared factor loadings for all variables. Thus, eigenvalues measure the amount of variation in the total sample accounted for by each factor.

Factor scores, also called *component scores* in PCA, are the scores of each case on each factor. Factor scores may be used as variables in subsequent modeling. Note also that factor scores are quite different from factor loadings. Factor scores are coefficients of cases on the factors, whereas factor loadings are coefficients of variables on the factors.

3.2 Principal Components Analysis

3.2.1 Introduction

The method of principal components, or principal components analysis (PCA), is a classical statistical method belonging to factor analytic techniques. The PCA is a concept for simplifying a dataset by reduction the dimension of observable random variables which has been widely used in data analysis. The PCA is one of the basic and the simplest factor analytic methods.

The PCA is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data is attributed to the first coordinate (called the first principal component), the second greatest variance to the second coordinate, and so on. The PCA can be used for dimensionality reduction in a dataset while retaining those characteristics of the dataset that contribute most to its variance.

The principal components analytic approach was first proposed by Karl Pearson (1901) for a nonstochastic variables. Then Person's concept, introduced only for the nonstochastic variables, was fully developed for the random variables by Hotelling (1933).

3.2.2 Definition of Principal Components in the Population

Let $x_t = (x_{1t}, \ldots, x_{Nt})'$ be a N-vector with

$$E(x_t) = \mu_t \tag{3.1}$$

$$\mathrm{cov}(x_t) = \Sigma \tag{3.2}$$

So suppose that random vector x_t has known covariance matrix Σ. We shall assume the cases in which the mean vector is $\mathbf{0}$ and the covariance matrix Σ is positive semidefinite matrix or it has multiple roots.

The principal components of x_t are normalized linear combinations of the components of x_t which have special properties in terms of variance. The first principal component of x_t is normalized linear combination

$$p_{1t} = \pi' x_t, \quad t = 1, \ldots, T$$

where $\pi \in E^N$ with $\pi'\pi = 1$ such that

$$\mathrm{var}(\pi' x_t) = \max_{\pi_i} \mathrm{var}(\pi_i' x_t) \tag{3.3}$$

for all $\pi_i = (\pi_{i1}, \ldots, \pi_{iN})' \in E^N$ satisfying $\pi_i'\pi_i = 1$.

The variance of $\pi_i' x_t$ is

$$\mathrm{var}(\pi_i' x_t) = E(\pi_i' x_t x_t' \pi_i) - E^2(\pi_i' x_t) = E(\pi_i' x_t x_t' \pi_i) = \pi_i' \Sigma \pi_i \tag{3.4}$$

To determine the first principal component $\pi' x_t$ it is necessary to find the π that maximizes (3.4) for all $\pi_i \in E^N$ and satisfies $\pi_i'\pi_i = 1$. Let

$$\phi_1(\pi_i) = \pi_i' \Sigma \pi_i - \lambda(\pi_i'\pi_i - 1) \tag{3.5}$$

where λ is a Lagrange multiplier. The goal is to find the π that maximizes the Lagrange function $\phi_1(\pi_i)$ among all choices of π_i that satisfy the condition $\pi_i'\pi_i = 1$. Therefore we deduce that the vector π must satisfy the first derivation of Lagrange function $\phi_1(\pi_i)$ set to equal 0:

$$\frac{\partial \phi_1}{\partial \pi_i} |_{\pi_i = \pi} = 2\Sigma\pi - 2\lambda\pi = 0 \tag{3.6}$$

Therefore since $\pi_i' \Sigma \pi_i$ and $\pi_i'\pi_i$ have derivatives everywhere in region containing $\pi_i'\pi_i = 1$, a vector π must satisfy

$$(\Sigma - \lambda I)\pi = 0. \tag{3.7}$$

By reason that $\pi \neq 0$ (as a consequence of $\pi'\pi = 1$), equation (3.7) has a solution if $\Sigma - \lambda I$ is singular, so if

$$det(\Sigma - \lambda I) = 0. \tag{3.8}$$

That is, λ is a eigenvalue of Σ and π is the corresponding eigenvector. Since Σ is dimension N x N, therefore equation (3.8) has N roots. Let

$$\lambda_1 \geq \lambda_2 \ldots \geq \lambda_N \tag{3.9}$$

are the ordered eigenvalues of Σ and

$$\pi_1 = (\pi_{11}, \dots, \pi_{1N})', \dots, \pi_N = (\pi_{N1}, \dots, \pi_{NN})' \tag{3.10}$$

are the corresponding eigenvectors of Σ. If we multiply the (3.7) by π' on the left, we obtain

$$\pi'\Sigma\pi = \lambda\pi'\pi = \lambda \tag{3.11}$$

This relationship shows that if π satisfies (3.7) and also $\pi'\pi = 1$, then

$$\mathrm{var}(\pi' x_t) = \pi'\Sigma\pi = \lambda. \tag{3.12}$$

Thus for maximization of the variance we have to choose the largest eigenvalue λ_1 of the Σ. So let π_1 be an eigenvector corresponding to the λ_1. Thus the normalized linear function

$$p_{1t} = \pi_1' x_t, \quad t = 1, \dots, T$$

called *the first principal component* of x_t, is a function with maximum variance.

The second principal component is the normalized linear function $\pi' x_t$ with the maximum variance among the all normalized linear functions $\pi_i' x_t$ that are uncorrelated with p_{1t}. So if any function $\pi_i' x_t$ is uncorrelated with p_{1t}, then

$$E(\pi_i' x_t p_{1t}) = 0 \tag{3.13}$$

From (3.13) it is clear that the vectors π_i and π_1 are orthogonal. [1]

Now by maximization of Lagrange function

$$\phi_2(\pi_i) = \pi_i'\Sigma\pi_i - \lambda(\pi_i'\pi_i) - 2\upsilon_1(\pi_i'\Sigma\pi_1) \tag{3.14}$$

where the λ and υ_1 are the Lagrange multipliers, we find the second principal component. So the maximizing π must satisfy

$$\frac{\partial\phi_2}{\partial\pi_i}\Big|_{\pi_i=\pi} = 2\Sigma\pi - 2\lambda\pi - 2\upsilon\Sigma\pi_1 = 0. \tag{3.15}$$

[1] $E(\pi_i' x_t p_{1t}) = E(\pi_i' x_t p_{1t}') = E(\pi_i' x_t x_t' \pi_1) = \pi_i'\Sigma\pi_1 = \pi_i'\lambda_1\pi_1 = \lambda_1\pi_i'\pi_1 = 0.$

Therefore (3.15) implies the relation

$$\pi_1'\Sigma\pi - \lambda\pi_1'\pi - \upsilon_1\pi_1'\Sigma\pi_1 = 0 \tag{3.16}$$

From (3.16) we get

$$\upsilon\lambda_1 = 0$$

since $\pi_1'\Sigma\pi = 0$ and $\pi_1'\Sigma\pi_1 = \lambda_1$. Therefore $\upsilon = 0$ and λ and π must to satisfy (3.6) and (3.7). So *the second principal component* of x_t is

$$p_{2t} = \pi_2' x_t, \quad t = 1, \dots, T$$

where the π_2 is the eigenvector of Σ corresponding to second largest eigenvalue λ_2.

We continue in this way to the Nth step.

Conclusion

If $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_N$ are ordered eigenvalues of Σ and $\pi_1, \dots, \pi_N$ are the corresponding eigenvectors, then

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_N \end{pmatrix}, \Pi = (\pi_1, \pi_2, \dots, \pi_N)$$

are the matrices of ordered eigenvalues and eigenvectors. From relations $\Pi'\Pi = I$ and $\Sigma\Pi = \Pi\Lambda$ we dedicate that $\Pi'\Sigma\Pi = \Lambda$. Thus exist $p_t = (p_{1t}, p_{2t}, \dots, p_{Nt})$ [2], a vector of orthogonal transformation

$$p_t = \Pi' x_t, \quad t = 1, \dots, T \tag{3.17}$$

[2] Properties of p_t:

- $cov(p_t) = \Lambda$ where $\Lambda = diag\{\lambda_1, \dots, \lambda_N\}$.
- ith column π_i of Π satisfies $(\Sigma - \lambda_i I)\pi_i = 0$.
- The components of p_t are uncorrelated and p_{it} has maximum variance among all normalized linear combinations uncorrelated with $p_{1t}, \dots, p_{i-1t}$.

that is called the *vector of principal components* of x_t.

In matrix notation the model is written as

$$P = X\Pi, \tag{3.18}$$

where $X = (x_1, \ldots, x_T)'$ is TxN matrix of data and $P = (p_1, \ldots, p_T)'$ is TxN matrix of principal components.

3.3 Classical Factor Analysis

3.3.1 Introduction

The classical factor analysis (CFA) is a multivariate statistical technique to apply a single set of variables to reduce a large set of variables to a more meaningful, smaller set of variables called *factors* which can account for the correlation of a set. The variables, that are correlated with one another and they are also largely independent of other subsets of variables, are combined into factors. The CFA is one of the most extended forms of factor analysis. The nature of the CFA is using the method principal components analysis that is applied for correlation matrix in which the diagonal elements are not ones, as in the PCA.

3.3.2 Definition of the Model

The each element of the observable vector $x_t = (x_{1t}, \ldots, x_{Nt})'$ can be written in $q-$factor model ($q \leq N$) as [3]

$$x_{it} = a_{i1} f_{1t} + \ldots + a_{iq} f_{qt} + u_{it}, \quad t = 1, \ldots, T \tag{3.19}$$

where x_{it} is the value of the t-th observation on the i-th variable, f_{kt} is the t-th observation on the k-th *common factors*, a_{ik} is the set of linear coefficients called the *factor loadings* associated with f_{kt}, and u_{it} is similar to residual because it is a part of x_{it} not explained by the common factors.

If $a_i' = (a_{i1}, \ldots, a_{iq})$ is a vector of *factor loadings*, then (3.19) can be expressed as

$$x_{it} = a_i' f_t + u_{it}, \quad t = 1, \ldots, T,$$

where $f_t = (f_{1t}, \ldots, f_{qt})'$ is a vector of q common factors.

Let $u_t = (u_{1t}, \ldots, u_{Nt})'$ is a vector of N idiosyncratic (specific) components of x_t and f_t is a vector of q common factors. Then the model can be rewritten

[3] We assume that $E(x_t) = \mu_t = 0$. If the $E(x_t) \neq 0$, then $x_t = \mu_t + Af_t + u_t$.

in matrix notation as

$$x_t = Af_t + u_t \quad t = 1, \ldots, T \tag{3.20}$$
$$X = FA' + U, \tag{3.21}$$

where $X = (x_1, \ldots, x_T)'$ is TxN matrix of data, $A = (a_1, \ldots, a_N)'$ is Nxq matrix of factor loadings, $F = (f_1, \ldots, f_T)'$ is Txq matrix of factors and $U = (u_1, \ldots, u_T)'$ is TxN matrix of specific components.

We shall assume that the vector of uncorrelated errors, u_t, is distributed independently of f_t, $\mathrm{cov}(f_t, u_t) = E(f_t u_t') = 0$, with mean $E(u_t) = 0$ and covariance matrix

$$\mathrm{cov}(u_t) = E(u_t u_t') = \Psi,$$

where $\Psi = \mathrm{diag}(\psi_1^2, \ldots, \psi_N^2)$.

Furthermore, we assume that the vector f_t is taken as a random vector with $E(f_t) = 0$ and

$$\mathrm{cov}(f_t) = E(f_t f_t') = \Omega.$$

When we require $\Omega = I$, the factors are said to be *orthogonal. Oblique* factors we obtain by replacing I by Ω , where Ω is not diagonal, positive definite correlation matrix.

It follows from these assumptions that the covariance matrix of observed X for oblique model is [4]

$$\mathrm{cov}(x_t) = A\Omega A' + \Psi = \Sigma \tag{3.22}$$

and in the case of the orthogonal factor model ($\Sigma = I$) the covariance matrix is

$$\mathrm{cov}(x_t) = AA' + \Psi = \Sigma.$$

[4] $\Sigma = \mathrm{cov}(x_t) = E(x_t x_t') - (E(x_t))^2 = E(x_t x_t') = E((Af_t + u_t)(Af_t + u_t)') = A\Omega A' + \Psi$.

3.3.3 Estimation of the Factor Model

Let the classical factor model in matrix notation

$$X = FA' + U$$

satisfies the assumption defined in previous section. So the matrices A (Nxq) and F (Txq) are both unknown.

There are various criteria for determining the matrix of factor loadings and matrix of factor scores, such as method of principal components, maximum-likelihood, minres method (minimum residual) and unweighted least-squares method. Bellow there are described only two most widely used methods.

Principal Factor Method

The principal factor method is the most commonly used, and is the "default" in most computer programs. In this method, one extracts principal factors from a correlation matrix with communalities in the diagonal. Once the communalities are estimated (for the last time), the analysis proceeds as in principal component analysis. The results are then called principal axes.

The unknown matrix of factor loadings A can be estimated by minimizing the residual sum of squares:

$$\sum_{t=1}^{T}(x_t - Af_t)'(x_t - Af_t) \tag{3.23}$$

subject to the constraint $A'A = I_q$. Differentiating (3.23) with respect to A and F yields the first order condition $(\mu I_N - S)\widehat{\mathcal{A}}_k = 0$ for $k = 1, \dots, q$, where $S = T^{-1}\sum_{t=1}^{T} x_t x_t'$ and $\widehat{\mathcal{A}}_i$ is the i'th column of $\widehat{A}$. The matrix $\widehat{A}$ minimizes the criterion function (3.23). Thus, the columns of $\widehat{A}$ are the eigenvectors of the q largest eigenvalues of the matrix S. So the matrix $\widehat{A}$ is the Principal Components estimator of A[5].

[5]This derivation of the factor loading for principal factor method was originated by Jolliffe (2004).

Maximum-likelihood Method

The method of maximum-likelihood is a well-established and popular statistical procedure for estimating the unknown population parameters. This method yields values of the estimators which maximize the likelihood function of the sample.

Under the assumption of a given number (q) of common factors and normally distributed of x_t, the method of maximum-likelihood is applied to get estimators of the factor loadings from the sample of N variables on the T observations. If $S = T^{-1}\sum_{t=1}^{T} x_t x_t'$, the maximum-likelihood estimator minimizes the function

$$\ell^* = tr(S\Sigma^{-1}) + \log(|\Sigma|)$$

originated by Jöreskog (1969).

Test of significance for the number of factors should be the inseparable part of maximum-likelihood method, because the implicit in the development of the function ℓ^* is an assumption regarding this number. The test procedure is to reject the hypothesis H_0 of q factors if the value of Likelihood Ratio Test $LR = -2[\ell^*(\Sigma = S) - \ell^*(\Sigma = \widehat{A}\widehat{A}' + \widehat{\Psi})]$ exceeds the χ^2 for the desired significance level.

In applying the forgoing test, it is necessary to know the degrees of freedom of χ^2 distribution. For the hypothesis H_0 of q factors, the number of degrees of freedom is given by $v = 0.5[(n-q)^2 - n - q]$.

3.3.4 Approximate Factor Model

The assumptions of the classical, also called static, factor model are restrictive for economic problems because it assumes that N is fixed and much smaller than T, the u_{it} are independent over time and are also independent across i. Thus the variance-covariance matrix of u_t, $\Psi = E(u_t u_t')$, is a diagonal matrix. Further limitation is independence between the factors f_t and idiosyncratic components u_{it}.

Chamberlain and Rothschild (1983) introduced an *approximate factor model* which is more general than the static model. The approximate factor model relaxes the assumption of classical factor analysis and it is also assumed that number of variables (N) increases to infinity. An approximate factor model allows for weakly serial and cross correlation and heteroskedasticity of the idiosyncratic components. Finally, the weak dependence between factors and idiosyncratic components is allowed.

Thus the approximate factor model in sense of Chamberlain and Rothschild must have bounded eigenvalues for covariance matrix Ψ.

3.3.5 Criteria for Determining the Number of Factors

In practice, to determine the unknown number of factors is not so clear. There are some criteria for determining the number of factors which could help us to specify the right number of factors.

The number of common factors, k, is indicated by k largest eigenvalues of correlation matrix R [6] of the sample. A part of total variance explained by the k common factors is $\tau(k) = \left(\sum_{i=1}^{k} \lambda_i\right)/N$, where λ_i is i'th eigenvalues of R. To choose all N factors enables to explain the total variance exactly because $\tau(N) = 1$.[7] Unfortunately, the limit for the explained variance indicates the sufficient fit is unknown.

Kaiser Criterion

Criterion, originated by Kaiser in 1960, is a common rule for determining the number of factors of the factor analysis. The Kaiser rule means to drop all components with eigenvalues under the 1.0. Unfortunately it is a conservative criterion which is not exactly right, it may overestimate or underestimate the true number of factors.

[6] Instead of eigenvalues of correlation matrix it is possible to use the k largest eigenvalues of covariance matrix Σ.

[7] $\tau(N) = tr(R)/N = (\sum_{i=1}^{N} \lambda_i)/N = N/N = 1$.

Scree-test

Scree criterion, designed by Cattell (1966) is, in contrast to the Kaiser criterion, a graphical rule for determining the number of factors. Scree-test illustrtaes the number of factors as the X axis and the corresponding eigenvalues at the Y axis. Cattell suggests to find an elbow on curve of eigenvalues where the smooth decrease of eigenvalues appears to level off to the right of the plot. This point says to drop all components after one starting point of the elbow.

As the Kaiser criterion, the Scree test is not unambiguous criterion because the determining the elbow point can be fairly subjective decision, the researcher may be tempted to set the number of factors desired by his research.

The Kaiser criterion and the Scree test are classical methods for determining the number of factors based on the k largest eigenvalues of correlation matrix R. But it can be shown that all eigenvalues of R increase with N thus the tests based on the sample of eigenvalues are not feasible.

Bai and Ng's Criteria

Further information criteria for specifying the number factors originated by Bai and Ng (2002) are not based on the sample of eigenvalues. It is a method suggested for approximate factor model as N and T converge to infinity.

Let $V(k) = (NT)^{-1}\sum_{t=1}^{T}\widehat{u}_t'\widehat{u}_t$ is the sum of squared residuals (divided by NT) from a k-factor model, where $\widehat{u}_t = x_t - \widehat{\Lambda}\widehat{f}_t$ is the vector of estimated idiosyncratic errors. Bai and Nq suggest the following three information criteria for determining the number of factors:

$$\begin{aligned}
IC_{p1}(k) &= \ln[V(k)] + k\frac{N+T}{NT}\ln\frac{NT}{N+T} \\
IC_{p2}(k) &= \ln[V(k)] + k\frac{N+T}{NT}\ln C_{NT}^2 \\
IC_{p3}(k) &= \ln[V(k)] + k\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right),
\end{aligned}$$

where $C_{NT}^2 = [\min\{N,T\}]$. Minimizing one of these information criteria in the range $k = 0, 1, \ldots, kmax$, where $kmax$ is some pre-specified upper bound for number of factors, can consistently estimate the true number of factors (q).

Although the criteria suggested by Bai and Ng (2002) are not based on the eigenvalues, they have unfortunately some bad features. That is, they require the large dimensional approximate factor model and according to Onatski (2006), Bai-Ng criteria tend to severely overestimate the true number of factors.

3.3.6 Rotation Methods

The idea of rotation is to simplify the factor structure by changing the factor loadings, such that the interpretation of factor analysis is more understandable. Since the alternative rotations may explain the same variance but have different factor loadings, they are used to describe the meaning of factors. Therefore the interpretation of results of factor analysis depends on the applied rotation method. Rotation does not actually change anything but may make the interpretation of the analysis easier.

Two main types of rotation are used: *orthogonal* when the new axes are also orthogonal to each other, and *non-orthogonal (oblique)* when the new axes are not required to be orthogonal to each other. If we choose an oblique rotation the factors are permitted to be correlated with one another ($cov(f_t) = \Omega$). By an orthogonal rotation the factors are not permitted to be correlated (they are orthogonal to one another $\Longrightarrow cov(f_t) = I$). There are various rotational orthogonal or oblique strategies that have been proposed.

Orthogonal Rotation

An orthogonal rotation is specified by a rotation matrix denoted $\mathbf{Q}$, where the rows stand for the original factors and the columns for the new (rotated) factors. A rotation matrix has the important property of being orthonormal because it corresponds to a matrix of direction cosines and therefore $Q'Q = I$.

Varimax rotation developed by Kaiser (1958), is the most popular orthogonal rotation method of factor axes which tries to maximize the variance of each of the factors, so the total amount of variance accounted for is redistributed over the extracted factors.

Oblique Rotation

In oblique rotations the new axes are free to take any position in the factor space, but the degree of correlation allowed among factors is, in general, small because two highly correlated factors are better interpreted as only one factor. Oblique rotations, therefore, relax the orthogonality constraint in order to gain simplicity in the interpretation.

Promax rotation is an alternative non-orthogonal rotation method which has the advantage of being fast and conceptually simple and therefore is appropriate for very large dataset.

3.4 Dynamic Factor Analysis

3.4.1 Introduction

In this section we will review only briefly the dynamic factor analysis because this topic is not the main subject of this master thesis for two reasons. First, the dynamic factor models represent a rich area of empirical analysis which cannot be surveyed in this thesis completely. Second, dynamic factor models require a large number of variables what is the primary justification. Actually, this is the major reason why we do not use this method in this master thesis.

In recent years, large-dimensional factor models have become the most popular type of factor analysis because it has a lot of advantages in various respects in comparison to the other methods.

The dynamic factor model, called also index model, was propossed by Sarget and Sims (1977) and Geweke (1977). Each variable in index model is represented as a sum of common component and an idiosyncratic component, which is orthogonal at any lead and lags both to the common factors and to the idiosyncratic components of all other variables. But the mutual orthogonality of the idiosyncratic components at any lead and lags causes the weakness of this model.

Forni and Lippi (2001) and Forni, Hallin, Lippi and Reichlin (2000) suggested a new model, generalized dynamic factor model, which provides a generalization of index model by allowing for non-orthogonal idiosyncratic terms. Three important features define this model: it is a finite dynamic factor model; it is designed for analysis of large cross section of time series and it allows a correlation between the idiosyncratic terms.

3.4.2 Specification of the Model

The basic idea, in the dynamic factor analysis, is that the every element of the vector x_t, x_{it}, $i \in N$, is represented as the sum of a *common component* χ_{it} and an *idiosyncratic component* ξ_{it}. The common component is driven by

$q-$dimensional vector of *common factors* $\mathbf{f_t} = (f_{1t}, \ldots, f'_{qt})$, which are loaded with possibly different coefficients and lags:

$$\chi_{it} = b_{i1}(L)f_{1t} + b_{i2}(L)f_{2t} + \ldots + b_{q1}(L)f_{qt}.$$

The generalized dynamic factor model can be written in matrix notation as follows:

$$x_t = \chi_t + \xi_t = \mathbf{B}(L)\mathbf{f_t} + \xi_t, \tag{3.24}$$

where $\chi_t = (\chi_{1t}, \ldots, \chi_{Nt})$, $\xi_t = (\xi_{1t}, \ldots, \xi_{Nt})$ and $\mathbf{B}(L)$ is a Nxq matrix of lag operator whose (i, j) entry is $b_{ij}(L)$.

There are three approaches of forecasting in a dynamic factor framework. First, Forni *et al.* (2000) show that the common component can be approximated by projecting the vector x_t on the first q dynamical principal components of x_t. But the disadvantage of this approach is that the dynamic principal components are not available at the beginning and at the end of the sample period. Forni *et al.* (2000) suggest an estimator of the dynamic factors in the frequency domain. The estimator is given by

$$f_x(\lambda) = f_\chi(\lambda) + f_\xi(\lambda),$$

where f_x is the spectral density matrix of x_t, f_χ is spectral density matrix of χ_t and f_ξ is the spectral density matrix of ξ_t.

Second, Stock and Watson (2002b) show that the common component can be approximated by projecting the x_t on the first $r = q(s+1)$ static principal components of x_t, where $(s+1)$ is the number of the current and lagged values.

Third, Forni *et al.* (2005) proposed two-step approach to solve the missing data problem for dynamic principal component at the end of the sample. In the first step the covariance matrix of the common and idiosyncratic component is estimated by the spectral decomposition and then generalized static principal component of χ_t is calculated.

3.4.3 Applications of Dynamic Factor Models

The dynamic factor models can be used to address different economic issues. For instance, it has been successfully applied in a number of papers to construct economic indicator and also for forecasting and in financial and macroeconomic literature to estimate in insurable risk. Recently they have been also applied to macroeconomic analysis to respect in international business cycle.

EuroCOIN™[8] and Chicago Fed National Activity Index (CFNAI)[9] are two most important examples of monthly coincident business cycle indicators constructed by dynamic factor analysis. The former is leading coincident indicator of the euro area business cycle constructed by Altissimo *et al.* (2001). The latter one corresponds to the index of economic activity developed by Stock and Watson (1999).

The factor models are also used as the forecasting tool. Stock and Watson (2002) used an approximate factor model for the estimation of indexes and to construct forecasts for monthly U.S. macroeconomic time series. The various economic variables in European Union have been forecasted by Marcellino (2001) and also by Banerje (2005). Schneider and Spitzer (2004) produced short-term forecasts of real Austrian GDP using the generalized factor model. Artis *et al.* (2001) forecasted various real, nominal and financial variables for UK economy, Schumacher (2005) forecasted the German GDP and Reijer (2005) Dutch GDP using large scale factor models.

The application of factor models in international business cycle is the most important for us. Helbling and Bayoumi (2003) identified the international business cycle among Group of Seven (G-7) countries using the asymptotic dynamic factor model. Eickmeier and Breitung (2005) investigated co-movements between CEECs and the euro area by means of a large-scale dynamic factor model.

[8] For more information see Appendix B.

[9] See http://www.chicagofed.org/economic research and data/cfnai.cfm.

3.5 Empirical Application of the Factor Analysis

Our empirical application addresses the recent discussion on whether the CEECs should join European Monetary Union. One of the criteria that should be satisfied is the synchronization of business cycles. We investigate how important euro area factors are for the CEECs compared to the current EMU members.

We use a classical (static) factor model to study the degree of synchronization between the CEECs and EMU countries. Although the use of a static factor model has some drawbacks, the use of more sophisticated approaches, as approximate factor model or dynamic factor model, is not possible. In the classical factor model, the idiosyncratic components are assumed to be uncorrelated, whereas approximate factor model allows for the idiosyncratic error terms. But it is assumed that the number of variables in approximate factor model tends to infinity. And the primary justification of dynamic factor model is large number of variables as well. Therefore, the assumption of large number of variables for approximate factor model and dynamic factor model is not fulfilled, because our data contain only 28 variables. Therefore, we use the classical factor model.

3.5.1 Determining the Number of Factors

Before the application of the factor analysis to data, it is necessary to determine the number of factors. In section 3.3.5, we specify three criteria for determining the number of factors. But we can use only two of them, because the use of Bai and Ng's criteria is conditional by use of the approximate factor model. Therefore, we use only Kaiser criterion and Scree test that are unfortunately reliable criteria for determining the number of factors.

Kaiser Criterion

From Figure 3.1 that illustrates 28 eigenvalues of the correlation matrix is apparent, that seven eigenvalues of correlation matrix are greater than 1, therefore the Kaiser criterion determines the seven factors used in the factor analysis.

Seven factors specified by Kaiser criterion explain 84.5% of variance if we use the principal factor method and 78.88% of variance in case of the maximum-likelihood method (Table 3.1).

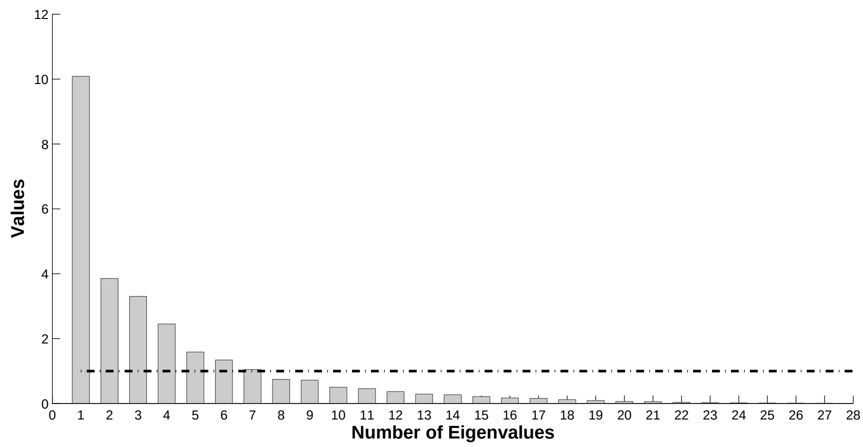

Figure 3.1: Kaiser criterion.

Scree test

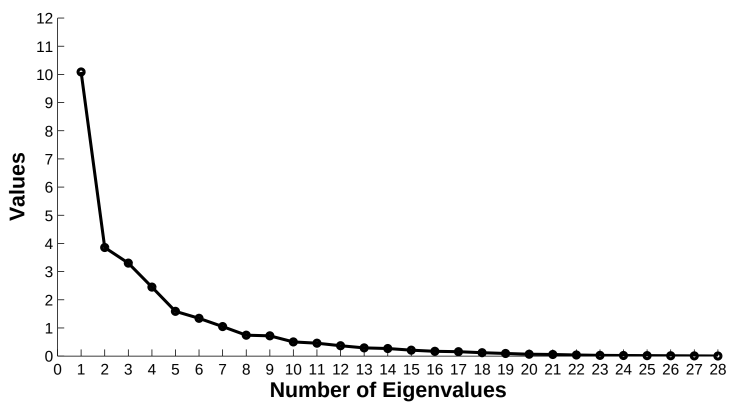

Figure 3.2: Scree test.

Scree test represents the number of eigenvalues in X axis and corresponding values in decreasing order on Y axis. If we choose Scree test for determining the number of factors, we can find a couple of elbows on the curve of eigenvalues (Figure 3.2). Therefore, it is clear, that Scree test is not unambiguous criterion for specifying the number of factors. From Figure 3.2 it is apparent, that we can choose one, three, five or eight number of factors.

Since both criteria (Kaiser and Scree test) determine the different numbers of factors, emerging the question which criterion to use. The Kaiser criterion retains too many factors, while Scree test underestimates the number of factors. An important additional aspect is the extent to which a solution is interpretable. Therefore, both criteria are not reliable.

The number of factors	**Principal factor method**	**Maximum-likelihood method**
1	0.3602	0.3406
2	0.4979	0.4618
3	0.6157	0.5778
4	0.7032	0.6488
5	0.7599	0.7093
6	0.8077	0.7434
7	0.845	0.7888
8	0.8717	0.8092
9	0.8974	0.8346
10	0.9153	0.8674

Table 3.1: Percentage of variance explained by the first ten factors.

The Table 3.1 represents percentage of variance explained by the first ten factors if it is used principal factor method or maximum-likelihood method. The number of factors is specified in first column of the table.[10] Well, the Table 3.1 illustrates how the number of factors affects the percent of explained variance. The percent of the explained variance is the same for rotated and unrotated method, because the idea of the rotation is to simplify the factor structure not to improve the explained variance.

This table can help us to decide how many factors we use in factor analysis. It is clear, that the explained variance increases with increased number of factors.

[10] The variance shares of principal factor method in Table 3.1 is equal to expression from section 3.3.5: $\tau(k) = \left(\sum_{i=1}^{k} \lambda_i\right)/N$, where λ_i is i'th eigenvalues of correlation matrix and N is number of observations.

To use only one factor is not correct, because the business cycle in Europe is not driven only by one factor. We choose three factors, because the differences between the explained variances are fewer and fewer with increased number of factors. Another reason is that two factors explain relatively low share of the total variance (49.79%), whereas three factors account for 61.57%.

3.5.2 Principal Factor Method

First, we applied principal factor method for estimating unknown matrix of factor loadings and factor scores. We estimated three factors as explained before. The factor loadings of these factors are illustrated in Figure 3.3.

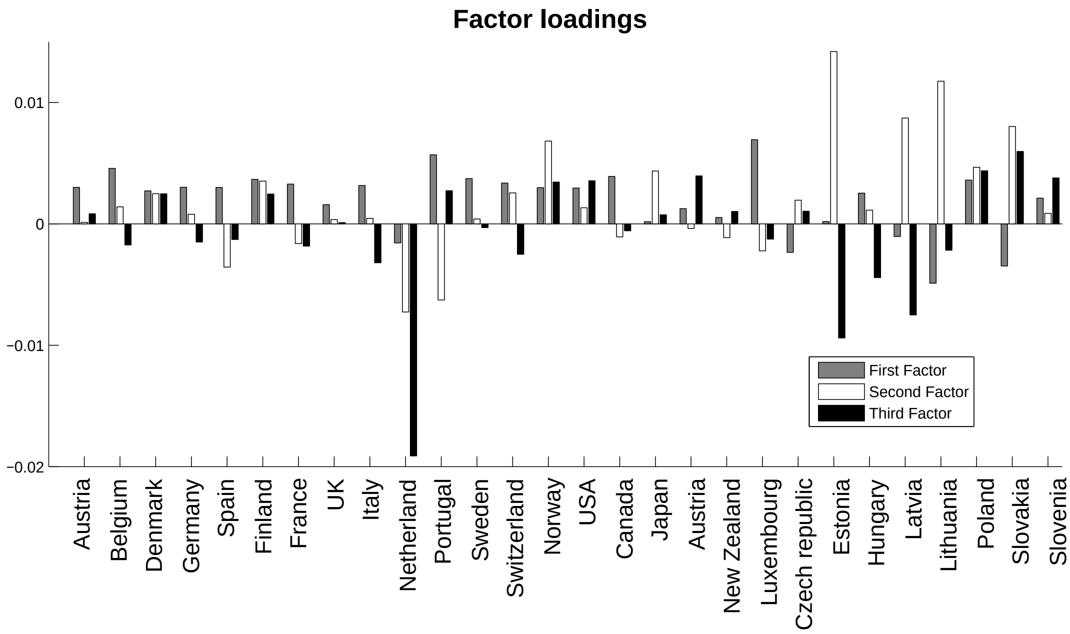

Figure 3.3: Principal factor method: Factor loadings.

Figure 3.3 makes visible that every factor has a particular meaning, because the values of factor loadings are different and depend on the country.

Specifically, the first factor describes euro area countries, because the factor loadings of first factor reach the highest values for countries from the euro area (Luxemburg, Portugal, Belgium, etc.). Among CEECs, Poland, Hungary and Slovenia have the highest value of the first factor. Other countries from Central and Eastern Europe have negative value of the factor loadings of the first factor. The second factor is characteristic for Baltic states and Slovakia. The third factor we can term as idiosyncratic or regional factor, because its factor loadings reach the lowest value among all the factors and the Netherlands, Estonia and Latvia have the highest (negative) value.

The findings implied from Figure 3.3 are also confirmed by Table 3.2. The table shows how much of the variance of output growth in CEECs and EMU countries is explained by 3 common factors.

From Table 3.2 it is clear that on average, the first factor explains a large part of output growth in EMU countries (50%) compared to CEECs (17.43%). Among CEECs the largest variance shares explained by the first factor are exhibited by Hungary, Slovenia and Poland. This implies that these countries are highly synchronized with economies of the euro area. Therefore, from Table 3.2 implies that the first factor can be interpreted as the euro area factor. For example, the first factor accounts for more than 60% of the output growth of France, Luxemburg and Germany. Canada is closely correlated with euro area, because the factor loading and variance share is the highest among the all non-European countries.

On the contrary, the second factor can be interpreted as the factor of CEECs, because it explains a large part of output growth for some countries of Central and Eastern Europe. The highest shares of variance are accounted for by the second factor mainly for the Baltic states and Slovakia. Specifically, 62.43% of the output growth of Estonia, 40.86% of Slovakia, 45.96% of Lithuania are explained by the second factor. On average, the second factor explains a large part of growth (44.89%) in CEECs compared to the first factor (17.44%). The exceptions are Hungary, Poland and Slovenia which are described by the first factor.

The third factor is idiosyncratic, because the variance shares are so low. The

Netherlands has the highest variance share which is not explained by the first two factors. Since the third factor improves the explained part of output growth for some countries (the Netherlands, Latvia and Hungary), it can be interpreted as regional factor.

Country	**Principal factor method**			
	First Factor	Second Factor	Third Factor	All factors
Austria	0.5635	0.0004	0.0142	0.5781
Belgium	0.4528	0.0163	0.0216	0.4907
Germany	0.6615	0.0174	0.0539	0.7327
Spain	0.3450	0.1856	0.0207	0.5513
Finland	0.4296	0.1513	0.0629	0.6438
France	0.7802	0.0736	0.0804	0.9342
Italy	0.5853	0.0045	0.1970	0.7868
Netherlands	0.0070	0.0565	0.3358	0.3993
Portugal	0.5010	0.2333	0.0378	0.7722
Luxemburg	0.6802	0.0267	0.0073	0.7143
Sweden	0.6045	0.0026	0.0014	0.6085
Switzerland	0.5726	0.1256	0.1048	0.8030
Norway	0.1614	0.3237	0.0706	0.5558
Denmark	0.3529	0.1129	0.0963	0.5623
UK	0.6068	0.0114	0.0009	0.6191
USA	0.5123	0.0397	0.2439	0.7960
Canada	0.7327	0.0212	0.0051	0.7591
Japan	0.0010	0.2272	0.0058	0.2340
Australia	0.1256	0.0045	0.4068	0.5370
New Zealand	0.0133	0.0243	0.0165	0.0541
Czech Republic	0.1025	0.0269	0.0065	0.1359
Estonia	0.0003	0.6243	0.2349	0.8594
Hungary	0.3407	0.0257	0.3421	0.7085
Latvia	0.0180	0.4845	0.3073	0.8098
Lithuania	0.2082	0.4596	0.0135	0.6813
Poland	0.2323	0.1480	0.1119	0.4923
Slovakia	0.2003	0.4086	0.1937	0.8027
Slovenia	0.2926	0.0188	0.3074	0.6188
Total variance	0.3602	0.1377	0.1179	0.6157

Table 3.2: The variance shares explained by the first, the second and the third factor and by all factors together in individual countries for the principal factor method.

In general, the first factor is often interpreted as a common business cycle. Following the previous findings about the first factor, we can suppose that the

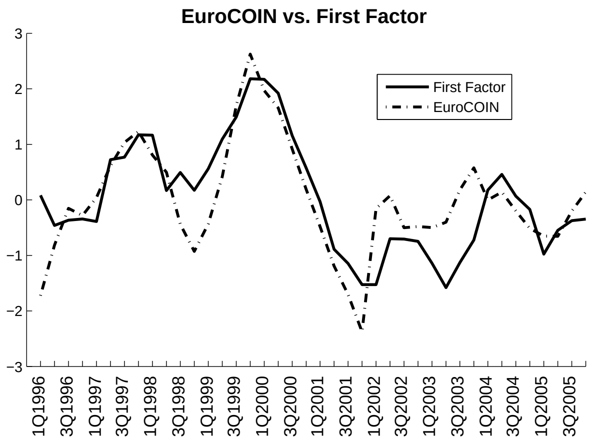

Figure 3.4: Comparison of EuroCOIN^{TM} (dot-and-dashed line) and the first factor (solid line) of the principal factor method.

first factor is closely linked with the euro area business cycle. Therefore, we compare the first factor with leading coincident indicator of the euro area business cycle. Indeed, as it is obvious from Figure 3.4[11], our first factor is highly correlated with EuroCOIN^{TM} [12, 13] and therefore it can be really interpreted as euro area business cycle.

3.5.3 Maximum-likelihood Method

We also applied maximum-likelihood method for estimating the factor loadings and factor scores. We estimated three factors and according to Likelihood Ratio Test defined in section 3.3.2 it is the sufficient number of factors. First, we use the maximum-likelihood method without rotation, but the factor structure is not simply. Therefore we use varimax rotation to change the factor structure and to make the interpretation of factors more understandable.

[11] The monthly EuroCOIN^{TM} series was converted into a quarterly series and it was normalized to have mean of zero and variance of one.

[12] Classical correlation between normalized first factor and normalized EuroCOIN^{TM} equals to 0.7714.

[13] For more information about EuroCOIN^{TM} see Appendix B.

The factor loadings of maximum-likelihood method with rotation are presented in Figure 3.5. The Table 3.3 shows shares of variance of output growth explained by the common factors for maximum-likelihood method with rotation and also without rotation. The maximum-likelihood method accounts for 57.79% of total variance.

Country	**ML Rotate**			**ML Unrotate**			All
	1.Factor	2.Factor	3.Factor	1.Factor	2.Factor	3.Factor	Factors
Austria	0.3434	0.2317	0.0215	0.4725	0.0002	0.1239	0.5966
Belgium	0.4314	0.0342	0.0112	0.4066	0.0678	0.0023	0.4767
Germany	0.5864	0.0863	0.0144	0.5772	0.0924	0.0174	0.6871
Spain	0.4015	0.0026	0.1314	0.4587	0.0447	0.0322	0.5356
Finland	0.1861	0.4198	0.0297	0.2493	0.0872	0.299	0.6356
France	0.9059	0.0063	0.0650	0.9629	0.0008	0.0135	0.9772
Italy	0.7698	0.0004	0.0181	0.654	0.1094	0.0249	0.7883
Netherlands	0.0267	0.2289	0.0003	0.0042	0	0.2517	0.256
Portugal	0.3041	0.0925	0.3139	0.5113	0.1648	0.0343	0.7105
Luxemburg	0.5406	0.078	0.0285	0.6315	0.0002	0.0154	0.6471
Sweden	0.4688	0.1126	0.0007	0.5256	0.0207	0.0359	0.5822
Switzerland	0.5407	0.0917	0.1118	0.4665	0.2559	0.0217	0.7441
Norway	0.0304	0.2794	0.1468	0.0331	0.1908	0.2327	0.4566
Denmark	0.1081	0.3599	0.0122	0.1652	0.044	0.2711	0.4802
UK	0.441	0.1947	0.0020	0.5334	0.0162	0.0882	0.6378
USA	0.1352	0.6803	0.0107	0.2888	0.0004	0.5369	0.8261
Canada	0.5991	0.0893	0.0273	0.6967	0.0008	0.0183	0.7158
Japan	0.0007	0.0242	0.1108	0.0004	0.1137	0.0217	0.1358
Australia	0.0001	0.3833	0.0658	0.0294	0.0499	0.3699	0.4492
New Zealand	0.0135	0.0037	0.0182	0.0169	0.0116	0.007	0.0355
Czech Republic	0.0774	0.0082	0.0088	0.0928	0.0006	0.001	0.0944
Estonia	0.0317	0.0088	0.8708	0.0042	0.8909	0.0162	0.9113
Hungary	0.5239	0.0059	0.0928	0.3646	0.2087	0.0494	0.6227
Latvia	0.0061	0.0522	0.7031	0.0272	0.6772	0.0571	0.7614
Lithuania	0.1600	0.0013	0.4798	0.3006	0.3382	0.0022	0.641
Poland	0.0553	0.3919	0.0199	0.0994	0.0482	0.3194	0.4671
Slovakia	0.4499	0.1538	0.2112	0.4457	0.0982	0.2711	0.8149
Slovenia	0.0385	0.4394	0.0158	0.1204	0.0022	0.371	0.4936
Total variance	0.2920	0.1593	0.1265	0.3264	0.1263	0.1252	0.5779

Table 3.3: The variance shares explained by the first, the second and the third factor and by all factors together in individual countries for the maximum-likelihood method.

From Table 3.3 and from Figure 3.5 it is obvious that the first factor can be interpreted as euro area factor, because the countries of euro area (France, Italy, Luxemburg, Belgium, etc.) reach the highest positive value of its factor loadings. Among CEECs only Hungary, Poland, Slovenia and Estonia have positive factor loadings belonging to the first factor. On the contrary, factor loadings of Slovakia, Latvia and Lithuania are negative, what confirms the negative corelation between these countries and euro area. The first factor explains a large part of output growth in EMU countries (44.96%) compared to CEECs (16.79%). Among CEECs the largest variance shares explained by first factor are exhibited by Hungary, following Slovakia. Let's remember that Slovakia has a negative factor loading.

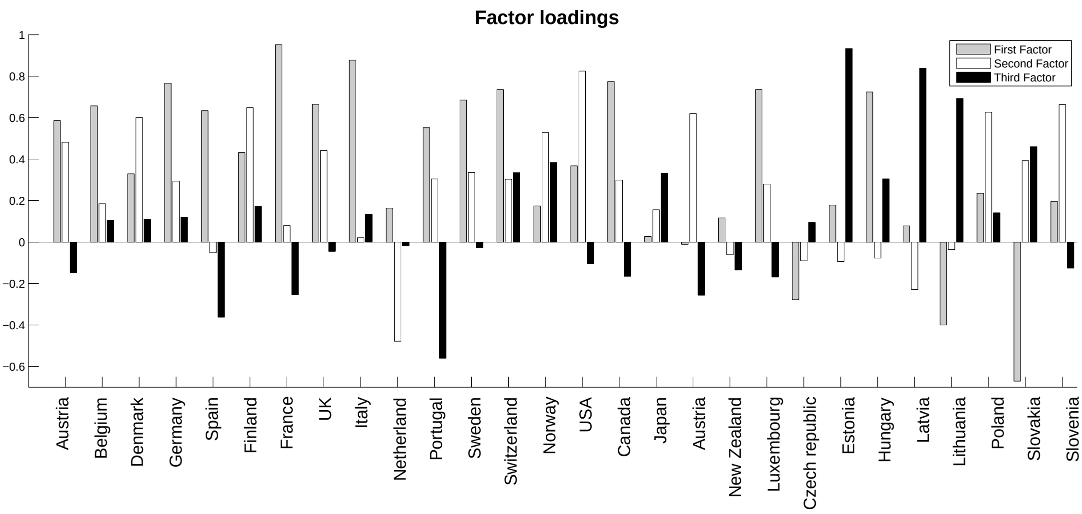

Figure 3.5: Maximum-likelihood method-rotate: Factor loadings.

The second factor represents the relationship between European countries and the USA, because the USA is a leading country of the second factor. The factor loading of the USA is the highest and also variance shares accounted for by the second factor are also the highest.

The large part of output growth in Baltic states (Estonia: 87.01%, Latvia: 70.31% and Lithuania: 47.98%) and Slovakia (21.12%) is explained by the third factor. All of these countries also reach the highest value of the factor loadings. Well, we can interpret the third factor as a factor of Baltic states and Slovakia.

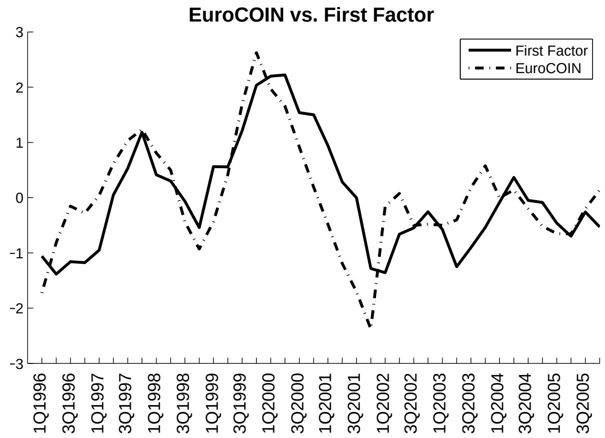

Figure 3.6: Comparison of EuroCOIN^{TM} (dot-and-dashed line) and the first factor (solid line) of the maximum-likelihood method.

As well as in the case of principal factor method, we compare the first factor of maximum-likelihood method with leading coincident indicator of euro area business cycle, EuroCOIN^{TM}. As it is obvious from Figure 3.6, the first factor is highly correlated with indicator of euro area business cycle. The static correlation[14] between EuroCOIN^{TM} and first factor equals 70.61%. Therefore, the first factor can be interpreted as the euro area business cycle.

3.5.4 Conclusions from the Application of the Factor Analysis

The first step to estimate the unknown factor loadings and factor scores is to determine the number of factors. We use two criteria for specifying the number of factors: Kaiser criterion and Scree test. However, these criteria are not reliable,

[14] We use static correlation, because we use classical (static) factor model. Static factor model does not allow for dynamics, therefore we can not use the dynamic correlation.

because they determine the different number of factors. Finally, we decided to use three factors, because three factors sufficiently explained the variance.

The both applied methods, principal factor method and maximum-likelihood method, present look-like results. Three factors generated by principal factor method account for 61.57% of total variance and three factors of the maximum-likelihood factor account for 55.79% of total variance. Among all countries, the explained variance of France, following Estonia, Latvia, Switzerland and Slovakia is higher than 80% in case of principal factor method. Among all countries, France, following Estonia, the USA and Slovakia have the highest variance shares explained by maximum-likelihood method.

The factor that is estimated as the first one by both method can be interpreted as euro area factor, because it mainly accounts for the output growth of EMU countries. The factor estimated by principal factor method is also characteristic only for three countries from Central and Eastern Europe: Hungary, Poland and Slovenia. Other CEECs have a negative factor loadings of first factor. That means, Hungary, Poland and Slovenia have a business cycle similar to euro area business cycle. Hovewer, Slovakia, Latvia and Lithuania have negative correlation with EMU countries.

One of another two factors describes Central and Eastern countries, especially Baltic states and Slovakia. Hungary, Poland and Slovenia is accounted for by the euro area factor and the Czech Republic is a specific case, because it has the lowest variance share among all European countries and the factor loadings reach very low values.

The first factor is interpreted as the euro area factor, therefore we compare the first factor with EuroCOIN^{TM}, which is the indicator of euro are business cycle. The Figures 3.4 and 3.6 and the relatively high static correlation between the first factor and EuroCOIN^{TM} indicate the intensive relation between them.

Chapter 4

Results and Conclusions

This master thesis examines the business cycle synchronization in the new EU members of Central and Eastern Europe and countries of the euro area. From the perspective of common monetary policy, it is relevant to know how the countries are synchronized.

We included in our study the data of GDP at a quarterly frequency for 24 OECD countries and for 3 Baltic countries plus Slovenia. For our analysis we used the softwer MATLAB, because the method of factor analysis is inbuilt in it. The sources of another needs matlab files are web [1] and also some of the authors of related papers who responded to my request for help. Especially, I would like to thank Sandra Eickmeier, Jörg Breitung and Marco Lippi for their valuable advises.

The main goal was to assess the current degree of synchronization of the CEECs and to see what extent they are satisfying one of the OCA criteria, namely, the synchronization of their business cycle with the euro area. We used two approaches for description of business cycle synchronization across Europe: dynamic correlation analysis and static factor analysis.

Firstly, we applied the dynamic correlation analysis which provide an information on existence of synchronization between the euro area and CEECs. On

[1]Bai-Ng criteria: http://www-personal.umich.edu/~ngse/research.html
Correlation analysis: http://www.economia.unimore.it/forni_mario/matlab.htm

average, business cycle correlation between the NMS and the euro area are lower than correlation between EMU countries and the euro area.

We start with EMU countries, which can we spilt into two groups: the "core" countries (Austria, Belgium, Germany, Spian, France, Italy and Luxemburg) which show a higher dynamic correlation with euro area output growth, and "peripheral" countries (Finland, Portugal and the Netherlands) which exhibit a lower synchronization. The reason of these differences could be that the peripheral countries joined to EMU much later.

According to our analysis, we can spilt CEECs into three groups: Hungary, Slovenia and Poland which are more suitable accession candidates than other NMS, because they are the most synchronized; Latvia and Estonia which are less synchronized; and Slovakia, Lithuania and the Czech Republic which are negative correlated with EMU countries.

Our empirical analysis also provides the information about the cohesion within Europe, which illustrates the synchronization across countries. The cohesion across Baltic states is the highest and relatively high cohesion is characteristic also for EMU countries. On the other hand, the synchronization across CEECs and across V4 countries are low.

Secondly, we use the static factor model, which is being increasingly employed. The basic underlying idea is that common movement in a cross-section can be captured by common factors. But the main drawback of using a static factor model is that it does not allow for dynamics in the relationship between the economic variables and factors. Therefore, many studies have used instead of variants of a dynamic factor model.

The first factor estimated by factor analysis can be interpreted as the euro area factor, because it mainly accounts for the output growth of EMU countries. The first factor estimated by principal factor method is also characteristic only for three countries from Central and Eastern Europe: Hungary, Poland and Slovenia. Other CEECs have a negative factor loadings of first factor. That means, Hungary, Poland and Slovenia have a business cycle similar to euro area business cycle. Hovewer, Slovakia, Latvia and Lithuania have negative correlation with EMU countries.

One of another two factors describes Central and Eastern countries, especially Baltic states and Slovakia. Hungary, Poland and Slovenia is accounted for by the euro area factor and the Czech Republic is a specific case, because it has the lowest variance share among all European countries and the factor loadings of all three factors reach very low values.

According to our analysis, Hungary, Poland and Slovenia are more suitable accession candidates than other NMS. Of those countries, Hungary is particularly deeply integrated in terms of trade and FDI and exhibit industry structures are similar to those in the euro area. The Slovenian economy is closely connected through trade with the euro area.

The low synchronization of the Czech Republic and Slovakia is due to the insufficient reforms and macroeconomic imbalance in the first half of the 1990s, leading to currency crisis in the Czech Republic and in Slovakia in 1997 and in 1998. These countries will most probably reach as high level of synchronization as leading CEECs in the coming years.

Appendix A

Data

This appendix describes the main guidelines followed setting up the database which has been used for analysing.

We include in our study the data of GDP at a quarterly frequency for 24 OECD countries and for 3 Baltic countries plus Slovenia. The main source of the data is International Financial Statistics service of the International Monetary Found, Washington (WIFO database)[1].

The sample of OECD countries includes Austria, Belgium, Denmark, Germany, Spain, Finland, France, United Kingdom, Italy, the Netherlands, Portugal, Sweden, Switzerland, Norway, Luxemburg, the Czech Republic, Hungary, Slovak Republic, Poland, the USA, Canada, Japan, Australia, New Zealand. Thus, our data set excludes Greece and Ireland for reason of data unavailability and Mexico, Iceland, Turkey and Korea which are not significant for our analysis. Estonia, Latvia and Lithuania belong to data set of Baltic countries. Overall, we include $N = 28$ quarterly series.

The reliable time series of our data set are unfortunately available from different starting points. They are disposable only from the beginning of the 1990s and for some countries, data availability is even more limited. The sample range for all countries was constrained by availability of data for Hungary and Poland, which start in 1Q1995. So in our estimation the sample ranges from

[1] WIFO - Austrian Institute of Economic Research.

1Q1995 to 4Q2005.

Firstly, the data have been seasonally adjusted using the X12ARIMA method. And then for this analysis we use log differenced data in order to render the data stationarity.

Finally, the series are collected in the Nx1 vectors x_t $(t = 1, 2, \ldots, T)$, where $N = 28$ represents the number of variables and $T = 40$ number of the observations. So prepared data set is appropriate for the analysis.

Appendix B

EuroCOIN^{TM}

"The EuroCOIN^{TM} is the leading coincident indicator of the euro area business cycle available in real time. The indicator provides an estimate of the monthly growth of euro area GDP – after the removal of measurement errors, seasonal and other short-run fluctuations. The indicator is available very quickly, well before the GDP numbers are released." [1]

The existence of the indicator is reasonable, because the only looking at GDP can be misleading. Whereas euro area GDP growth may be influenced by seasonal effects or by factors affecting only a particular sector or a particular country. An additional problem with GDP growth is that it does not provide an information about the monthly economic activity by reason that it is measured at quarterly frequency. Thus, the EuroCOIN^{TM} measured at monthly frequency is the best equipment to describe the euro business cycle.

Therefore, the EuroCOIN^{TM} is often used as a leading indicator for the economic development in the euro area. It is also used for the assessment the current state of the business cycle in the euro area and also for the establishment of historic dating of expansions and recessions.

EuroCOIN^{TM}, monthly coincident indicator of the business cycle of the euro area, was constructed by Altissimo *et al.* in 2001. They used GDP, industrial

[1] Source: http://www.cepr.org/data/eurocoin/.

production and prices for different sectors and countries, financial variables, and other macroeconomic data for six European countries in the estimation of the indicator.

The graph of the indicator represents the euro area business cycle. The interpretation of the graph is so intuitive. If the graph has a positive slope, the rate of growth is increasing. A negative slope indicates decreasing rate of growth. The negative value of the indicator indicates falling economic activity. If the EuroCOIN^{TM} is positive but less (more) than the historical average of GDP, it is rising (decreasing) at a slower rate than average of GDP.

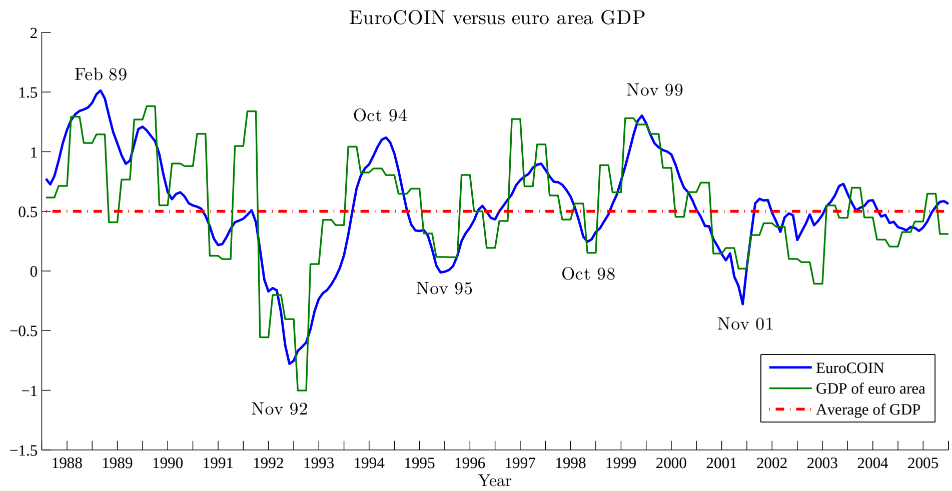

Figure B.1: The comparison of EuroCOIN^{TM} and GDP of the euro area: 1988 – 2005

The Figure B.1 represents the comparison of EuroCOIN^{TM} and GDP of the euro area from 1988 to 2005. This figure proves all propositions about the leading coincident indicator of the euro area business cycle described above.

Thus, we can see from the figure that EuroCOIN^{TM} is clean from seasonal and short-run effects in spite of GDP. It is also clear that there are 4 periods of recession in the euro area from 1988 to 2005. The starting and ending points of recession respectively is defined as the turning points of the cycle and they are also illustrated in figure. By means of the dashed line that represents the historical average of euro area GDP we can detect the periods with low or high growth.

This appendix introduce the main guidelines following the EuroCOIN$^{\mathrm{TM}}$ - leading coincident indicator of euro area business cycle published each month by CEPR[2]. Thus, it should be the best comparing criterion for the results of our analysis.

[2]CEPR: Centrum for Economic Policy Research.

Bibliography

[1] Altissimo F., Bassanetti A., Cristadoro R., Forni M., Lippi M., Reichlin L. and Veronese G. (2001). *EuroCOIN: A real time coincidnet indicator of the Euro area business cycle*, CEPR Working Paper No. **3108**.

[2] Anderson, T.W (1958). *An introduction to multivariate statistical analysis*, John Wiley & Sons: New York.

[3] Artis, M., Banerjee, A. and Marcellino, M. (2001). *Factor forecats for UK*, European University Institute Florence.

[4] Artis M.J., Marcellino M. and Proietti, T. (2005). *Business cycle in the new EU member countries and their confomity with the euro area*, Journal of Business Cycle Measurement and Analysis **2**, 7-42.

[5] Bai B.J. and Ng S. (2002). *Determining the number of factors in approximate factor models*, Econometrica **70**, 191-221.

[6] Banerje, A., Marcellino, M. and Masten, I. (2005). *Forecasting macroeconomic variables for the new member states of European Union*, ECB Working Paper No. **482**.

[7] Cattell, R.B. (1966). *The Scree test for the number of factors*, Multivariate Behavioral Research.

[8] Chamberlain, G. and Rothschild, M. (1983). *Arbitrage, factor structure, mean-variance analysis in large asset markets*, Econometrica **51**, 1305-1324.

[9] Croux, Ch., Forni, M. and Reichlin, L. (1999). *A measure of co-movement for economic variables: theory and empirics*, CEPR Discussion Paper No. **2339**.

[10] Darvas, Z. and Szapáry, G. (2005). *Business cycle synchronization in the enlarged EU: Comovements in the new and old members*, Discussion paper No. **5179**. CEPR, London.

[11] Eickmeier, S. and Breitung, J. (2005). *How synchronized are central and east European economies with the euro area? Evidence from a structural factor model*, Deutsche Bundesbank Discussion Paper No. **20**.

[12] Fidrmuc, J. and Korhonen, I. (2006). *Meta-analysis of the business cycle correlation between the euro area and the CEECs*, Journal of Comparative Economics **34**, 518-537.

[13] Forni, M. and Lippi, M. (2001). *Generalized dynamic factor model: representation theory*, Economic Theory **17**, 1113-1141.

[14] Forni, M., Hallin, M., Lippi, F., Reichlin L. (2000). *The Generalized Dynamic Factor Model: Identification and Estimation*, Review of Economics and Statistics **82**, 540-554.

[15] Forni M., Hallin, M., Lippi, F., Reichlin, L. (2005). *The generalized dynamic factor model: one-sided estimation and forecasting*, Journal of the American Statistical Association **100**, 830-840.

[16] Geweke, J. (1977). *The dynamic factor analysis of economic time series,* in D.J.Aigner and A.S.Goldberger, Eds., Latent Variables in Socio-Economic Models, North Holland: Amsterdam.

[17] de Grauwe, P. (2005). *The Economics of Monetary Union*, Sixth Revised Edition, Oxford University Press.

[18] Harman, H.H. (1960). *Modern factor analysis*, The University of Chicago Press, Chicago.

[19] Helbling, T. and Bayoumi, T. (2003). *Are they all in the same boat? The 2000-2001 growth slowdown and the G-7 business cycle linkages*, International Monetary Fund Eorkin Paper No. **46**.

[20] Hotelling, H. (1933). *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology **24**, 417-441.

[21] Jolliffe, I.T. (2004). *Principal component analysis*, 2nd edition, New York: Springer.

[22] Jöreskog, K.G. (1969). *A general approach to confirmatory maximum likelihood factor analysis*, Psychometrica **34**, 183-202.

[23] Kaiser, H.F. (1958). *The varimax criterion for analytic rotation in factor analysis*, Psychometrika **23**, 187-200.

[24] Kenen, P.B. (1968). *The optimum currency area: an electic view*, In: Mundel, Robert A. and Swoboda, Alexander K.(Eds.), Monetary problems of international economy, University of Chicago Press, Chicago.

[25] Lawley, D.N. (1940). *The estimation of factor loadings by the method of maximum likelihood*, Proc. Roy. Soc. Edinburgh.

[26] Marcellino, M., Stock, J. and Watson M. (2001). *Macroeconomic forecasting in the euro area: Country specific versus area-wide information*, IGIER, Working Paper No. **201**.

[27] McKinnon, R. (1963). *Optimum currency areas*, The American Economic Review **53**, 717-725.

[28] Mundell, R. (1961). *A theory of optimum currency area*, American Economic Review **51**, 657-665.

[29] Onatski, A. (2006). *Determining the number of factors from Empirical Distribution of Eigenvalues*, Columbia University.

[30] Pearson, K. (1901). *On lines and planes of closest fit to systems of points in space*, Phil. Magazine **6**, 559-72.

[31] DEN Reijer, A.H.J. (2005). *Forecasting Dutch GDP using large scale factor models*, DNB Working Paper No. **20**.

[32] Sarget, T. and Sims, C.A. (1977). *Business cycle modelling without pretending to have much a priori economic theory*, in Cristopher A. Sims (ed.) New Methods in Business Research, Minneapolis: Federal Reserve Bank of Minneapolis.

[33] Spearman, C. (1904). *General intelligence objectively determined and measured,* Am. J. Psychol. **15**, 201-93.

[34] Schneider, M. and Spitzer, M. (2004). *Forecasting Austrian GDP using the generalized dynamic factor model*, OeNB Working paper No. **89**.

[35] Schumacher, Ch. (2005). *Forecasting German GDP using alternative factor models based on large datasets*, Deutsche Bundesbank Discussion Paper No. **24**.

[36] Stock, J.H. and Watson, M.W. (1999). *Forecasting inflation*, Journal of Monetary Economics **44**, 293-335.

[37] Stock, J.H. and Watson, M.W. (2002a). *Macroeconomic forecasting using diffusion indexes*, Journal of Business and Economic Statistics **20**, 147-162.

[38] Stock, J. H. and Watson, M.W. (2002b). *Forecasting using principal components from a large number of predictors*, Journal of the American Statistical Association **97**, 1167-1179.

[39] Thurston, L. (1945). *Multiple factor analysis*, Chicago, Illinois: Univ. of Chigaco Press.