Global, Regional and Country Factors for the World Economy: a dynamic factor approach

Stefano Grassi[(a)*] **Borus Jungbacker**[(b)] **Siem Jan Koopman**[(b,c)]

[(a)] S.E.F. e ME.Q., University of Rome "Tor Vergata"
[(b)] Department of Econometrics, VU University Amsterdam
[(c)] Tinbergen Istitute, Amsterdam

Abstract

We implement the maximum likelihood estimation method for high-dimensional dynamic multi-factor models in presence of missing data. The exact treatment of missing values with reduction technique proposed by Jungbacker and Koopman(2008) allows the estimation of the factors and a large number of parameters in very fast and efficient way. We apply this new methodology for estimating the world cycle and area specific cycles, plus country specific effects. The data set concerns the GDPs of a large number of countries, consequently the speed and reliability of the state space algorithms is crucial in this framework. The paper provides a systematic assessment of the estimation strategy and discusses the empirical evidence in the light of the previous literature.

Keywords: State space form; Missing Values; Maximum Likelihood based analysis; World Business Cycle; Likelihood stability; Variance Decomposition.

1 Introduction

Large scale factor models aim at extracting the main economic signals from a large number of time series. The main idea is to reconstruct the comovements in a panel of economic time series at the basis of a limited number of common factors. Factor analysis has been widely used in economics and in finance. For example factor models can be used to study business cycle synchronization between regional areas. Different areas of applications are surveyed in Stock and Watson (2006). Large scale factor models can be estimated with principal component methods, see Bai and Ng(2008) for a good review.

*Stefano Grassi would like to thank Tommaso Proietti for the continuous help and suggestions, Giuseppe Garofalo for the comments on the early draft of this paper and the Free University of Amsterdam for the hospitality during the period February - June 2009. Address for Correspondence: Via Columbia 2, 000133 Rome, Italy. Tel +393343534438. E-Mail: stefanograssi@hotmail.com

Recently there has been an increasing interest in likelihood-based approach for the estimation of large scale data set. Since factors are explicitly modelled and the estimation method takes into account the model specification, the factor can represent aspects of economic theory. Doz et al.(2007) show, under mild conditions, that the estimates of the unobserved factors obtained from a likelihood-based analysis are consistent estimators for the true factors when $T \to \infty$ and $N \to \infty$ even if the dynamic factor model is misspecified. Furthermore, they present evidence that in some cases a likelihood-based analysis produces more precise estimates of the factors than a principal component method. In a recent paper Jungbacker and Koopman(2008) proposed a new method for likelihood based analysis for dynamic factor models; they demonstrated that when $N > q$, where N is the number of series and q is the number of factors, the computational efficiency of Kalman filter and Smoother can significantly be improved by a simple computational device, this new device is based on the projection of the data on the reduced dimensional factor space.
The problem of unbalanced panel is usually handled by EM algorithm see Stock and Watson(2002). In case of missing observations Jungbacker et al.(2009) proposed a new state space formulation that is faster and more efficient than other formulation proposed in the literature, see among other Reis and Watson(2007) and Banbùra and Modugno(2008).
The presence of a world business cycle still remains an open question, recent studies have provided evidence that there are many cross-country links in macroeconomic fluctuations, see among other Backus et al.(1995). Baxter (1995) find that business cycles in major industrialized economies are quite similar. Those studies are somehow limited due to *data availability* (missing values), and *econometric intractability* (inefficient estimation due to the presence of a lot of parameters), for this reasons they are focused on the analysis of a small group of countries, or to world aggregates. To overcome the second problem Kose et al.(2004) using Bayesian dynamic factor model studied the dynamic comovement of macroeconomic aggregates in a broad cross section of countries. They provide an analysis of world, regional, even country factors. With this approach they can handle many more series, but, they did not provide and efficient treatment of missing data, that it is a standard characteristic in economic time series.

The main aim of this paper is the efficient maximum likelihood estimation of the factors and the parameters in a large dimensional unbalanced panel. We implement the maximum likelihood estimation for the formulation proposed by Jungbacker et al.(2009) and we demonstrate that an efficient estimate of a lot of parameters, in case of missing data, is possible and computationally efficient. Using an unbalanced panel composed by 146 GDPs we estimate the world, regional and country specific factors. Moreover the flexibility of this new technique together with the reduced computational time opens the way to a more efficient treatment of large-scale mixed-frequency dynamic factor model, for an example of this kind of model see Proietti(2008). The remainder of the paper is organized as follow.

The Analysis by Maximum Likelihood is review in Section 2. In Section 3 we apply this new technique to a different dimension data set. We use R^2 and Box-Ljung statistic to asses the accuracy of our model. Furthermore we provide the variance decomposition to measure the relative contributions of the world, region, and country factors. Section 4 concludes the paper. The data set is presented in the Appendix A. The computational time is presented in Appendix B. The state space form and the derivatives useful for analytical maximum likelihood evaluation are in Appendix C.

2 Analysis by Maximum Likelihood

The estimation of the global and regional factors in a big unbalanced panel is a very challenging task. We address this and related issues by implementing the maximum likelihood estimation for the state space form recently proposed by Jungbacker et al.(2009). Subsection 2.1 describes briefly the dynamic factor model and the strategy to disentangle the global and the regional factors, using a block structure in the loading matrix. Subsection 2.2 presents the likelihood for unbalanced panel. The state space form and the analytical derivative are reported in Appendix C.

2.1 Dynamic Factor Model and Block Structure in the Loading Matrix

Consider a panel of N time series where we denote $y_{i,t}$ as the observation at time t in the i series, then the dynamic factor model is given by;

$$y_{i,t} = \Lambda f_t + u_{i,t} \qquad t = 1,...,T \qquad i = 1,...,N \tag{1}$$

where Λ is the $N \times q$ vector of factor loadings, where q is the number of factors and N is the number of series, $u_{i,t}$ is the *country-specific noise* and f_t is a set of unobserved factors of dimension q. We assume that the f_t are a linear combination of an unobserved $p \times 1$ dimensional vector autoregressive process α_t. In particular we have a $q \times p$ selection matrix G that defines the dynamic factor as

$$f_t = G\alpha_t \tag{2}$$

the factors have the following state space representation

$$\alpha_{t+1} = T\alpha_t + \eta_t \qquad \eta_t \sim N(0, \Sigma_\eta) \tag{3}$$

where the α_t is a time-variant state vector. Typically, one employs the identification assumptions that the factors are independent and have variance restricted to 1, we share the same assumption. Furthermore we assume that the errors component u_t follow a VAR(1) process given by

$$u_{t+1} = \phi u_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon) \tag{4}$$

where ϕ is a $N \times N$ diagonal matrix and the variance matrix Σ_ε is $N \times N$ diagonal matrix of unknown parameters that has to be estimated.

One of the main advantages of our estimation methodology is relatively easy way to impose constraints in the loading matrix that is useful to disentangle the global and regional factor. Those kind of restrictions are not possible using a principal component approach. We consider a block structure in the Λ matrix, of this form:

$$\Lambda = \begin{pmatrix} x & x & 0 & 0 & 0 & 0 & 0 \\ x & 0 & x & 0 & 0 & 0 & 0 \\ x & 0 & 0 & x & 0 & 0 & 0 \\ x & 0 & 0 & 0 & x & 0 & 0 \\ x & 0 & 0 & 0 & 0 & x & 0 \\ x & 0 & 0 & 0 & 0 & 0 & x \end{pmatrix}$$

where the x is a vector corresponds to factor loadings that has to be estimated and the 0 corresponds to the factor loading restricted to be 0. The first column of the Λ matrix represents the global factor that loads to all the series, the non zeros elements in the remaining columns represent the regional factor loadings.
In case of unrestricted Λ, see subsection 3.4, to ensure that all parameters are identified, we set $\Lambda = (\lambda_1^{'}, \lambda_2^{'})'$ where λ_1 is an $q \times q$ lower triangular matrix and λ_2 is an $(N-q) \times (N-q)$ full matrix. Where N is the number of series and q is the number of factors.

2.2 Maximum Likelihood for Unbalanced Panel

We can define the log-likelihood function as

$$l(y) = \log p(y_{1,t}^o, ..., y_{i,t}^o; \theta) \tag{5}$$

where $p()$ is the Gaussian density function, $y_{i,t}^o$ is the observed data for the country i at time t and θ is the vector of parameters. Following Jungbacker et al.(2009) and appendix C, the likelihood of our model can be expressed in this form:

$$l(y) = constant + l(y^L, y^{o,m}) + l(y^H) \tag{6}$$

where

$$y^L = \{A_t^L y_t(o_t, o_{t-1})\}_{t=1}^T \qquad y^{o,m} = \{y_t(o_t, m_{t-1})\}_{t=1}^T \qquad y^H = \{A_t^H y_t(o_t, o_{t-1})\}_{t=1}^T \tag{7}$$

$y^L = \{A_t^L y_t(o_t, o_{t-1})\}_{t=1}^T$ and $y^{o,m} = \{y_t(o_t, m_{t-1})\}_{t=1}^T$ corresponds, respectively, to the reduced part of observed values at time t and $t-1$ and the observed values at time t but missing at time $t-1$. The likelihood corresponding to this two components is evaluated with the Kalman Filter. Kalman Filter is not applied to $y^H = \{A_t^H y_t(o_t, o_{t-1})\}_{t=1}^T$, and this partial likelihood is calculated accordingly to:

$$l(y^H) = -\frac{[(N-m) \times T] - Missing)}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log(|\Sigma_{\varepsilon,t}|) - \frac{1}{2} \sum_{t=1}^{T} e_t^{'} \Sigma_{\varepsilon,t}^{-1} e_t \tag{8}$$

where

$$e_t = (I - \Sigma_t A_t^{L'} A_t^L [y_t(o_t, o_{t-1}) - \phi_t^{(o)} y_{t-1}(o_t, o_{t-1})] \tag{9}$$

and m are the number of projection series used in the "reduced form" KFS, for $t = 1, ..., n.$, see for a detailed discussion Jungbacker et al.(2009).

The maximization of this likelihood with respect to the parameter vector θ, involves a high dimensional maximization problem. Large scale optimizations problems are solved, in general, by quasi-Newton type algorithm as described in Nocedal and Wright(1999). For those algorithms we need the evaluation of the $l(y)$ and the score function at each iteration of the algorithm. Due to the high dimension of the parameter space numerical derivatives are not feasible, fortunately, analytical expression for the score is available. Following Koopman and Shephard(1992) and Appendix

C, we can write the likelihood for our model in the following way:

$$\begin{aligned}\log f(y_n,\theta) = & -\frac{1}{2}\sum_{t=1}^{T}\log \mid H_t(\theta)\mid + \log \mid Q_t(\theta)\mid \\ & -\frac{1}{2}\sum_{t=1}^{T}\Big[H_t(\theta)^{-1}\{(y_t-c_t-Z_t\alpha_{t|n})(y_t-c_t-Z_t\alpha_{t|n})^{'}+Z_tP_{t|n}Z_t^{'}\}\Big] \\ & -\frac{1}{2}\sum_{t=1}^{T}\Big[Q_t(\theta)^{-1}\{\eta_{t|n}\eta_{t|n}^{'}+P_{t|n}-T_tP_{t-1,t|n}-P_{t,t-1|n}T_t^{'}+T_tP_{t-1|n}T_t^{'}\}\Big] \\ & -\frac{1}{2}\log \mid P_0\mid -\frac{1}{2}(\alpha_0-a_0)^{'}P_0^{-1}(\alpha_0-a_0)-\log f(\alpha|Y_n;\theta)\end{aligned} \tag{10}$$

where $\eta_t = (\alpha_{t+1} - d_t - T_t\alpha_t)$, and $P_{t,t-1|n}$ is the covariance between the states, that has to be calculated in order to evaluate the likelihood and the derivative. Those states can contain missing values and we apply the following formula:

$$P_{t,t-1|n} = Cov(\alpha_t, \alpha_{t-1}|n) = T_{t-1}P_{t-1|n} - \Sigma_{\eta,t}N_{t-1}L_{t-1}P_{t-1|t-2} \tag{11}$$

see Proietti (2008). All the quantities are given as output by the Kalman filter and Smoother. Although other solutions could be applied those are, in presence of many missing data, highly inefficient or difficult to implement. For example we can augment the state vector in order to calculate the covariance between the missing values but this huge state vector slows down the Kalman filter and Smoother enormously and may even lead to numerical inaccuracies. Thanks to this parallel covariance calculation and to the device of Jungbacker and Koopman (2008) we can estimate the factors with many series and many missing values in a very efficient way.

3 Empirical application

In this section we apply this new estimation technique to the data set taken from Penn World Table available, free of charge, from the web side *http://pwt.econ.upenn.edu/*. The observation are Annual GDP Constant Price: Laspeyres (base year 2000) and span the period between 1950 and 2004. The time series extracted are 146 with 54 observations available at the most for each series, this leads to an arbitrary pattern of missing data. The logarithm of the series are assumed to be I(1), so we differentiate and standardize all of them. One concern about procedures that extract measures of the world business cycle is that large countries drive the world component simply because of their size. In the procedure used here we are working in growth rates, so the size of the country can have no direct impact on the results. That is, the econometric procedure that extracts common components does not distinguish between a 2-percent growth rate in the United States and a 2-percent growth rate in China. We divide the countries in 6 different areas accordingly to Table 1. This division is somehow different from that proposed by Kose et al.(2004) because we include the Oceania countries, divided accordingly to developed and developing, in the Asia and Oceania Developed and Asia and Oceania Developing and Poor regions. We assume one global factor (α^{global}) and six regional factors ($\alpha^{regional}$ eg. one each for North America, Latin America, Europe, Asia and Oceania Developed, Asia and Oceania Developing and Poor and Africa). Thus for the series i we have:

$$y_{i,t} = \lambda_i^{world}\alpha_t^{world} + \lambda_i^{region}\alpha_{r,t}^{region} + u_{i,t} \tag{12}$$

Table 1: All the 146 countries divided accordingly to six different areas.

North America	1-3				
	United States				
	Mexico				
	Canada				
Latin America	1-7	8-14	15-21	22-28	29-32
	Brazil	Jamaica	Barbados	Guatemala	Netherlands Antilles
	Argentina	Peru	Costa Rica	Honduras	Puerto Rico
	Bolivia	Paraguay	Dominica	Nicaragua	Trinidad Tobago
	Cambodia	Uruguay	Dominica Republic	Panama	Suriname
	Chile	Venezuela	Ecuador	Bahamas	
	Colombia	Antigua	El Salvador	Bermuda	
	Cuba	Belize	Grenada	Haiti	
Europe	1-7	8-14	15-21		
	Germany	Portugal	Finland		
	France	Norway	Denmark		
	United Kingdom	Netherlands	Belgium		
	Italy	Luxembourg	Austria		
	Sweden	Ireland	Cyprus		
	Switzerland	Iceland	Malta		
	Spain	Greece			
Asia/Oceania Developed	1-4	5-8	9-12	13-15	
	Japan	Australia	Singapore	Saudi Arabia	
	China	New Zealand	Thailand	Kuwait	
	Taiwan	Malaysia	Republic of Korea	Quatar	
	Hong Kong	Turkey	United Arab Emirates	Israel	
Asia/Oceania Developing and Poor	1-6	7-12	13-18	19-24	25-27
	Philippines	Papua New Guinea	Micronesia, Fed. Sts	Jordan	Bhutan
	Indonesia	Kiribati	India	Bangladesh	Maldives
	Korea, Dem. Rep	Samoa	Pakistan	Iraq	Mongolia
	Brunei	Solomon Islands	Sri Lanka	Nepal	Syria
	Laos	Tonga	Mauritius	Oman	
	Macao	Vanuatu	Iran	Bahrain	
Africa	1-10	11-20	21-30	31-40	41-47
	South Africa	Senegal	Comoros	Liberia	Sierra Leone
	Egypt	Somalia	Dem. Rep. Congo	Lesotho	Swaziland
	Morocco	Tunisia	Republic of Congo	Malawi	Tanzani
	Nigeria	Uganda	Equatorial Guinea	Mali	Togo
	Algeria	Cameroon	Gabon	Mauritania	Zambia
	Central African Republic	Botswana	Gambia	Namibia	Zimbabwe
	Cote d'Ivoire	Benin	Ghana	Niger	Sudan
	Ethiopia	Burundi	Guinea	Cape Verde	
	Madagascar	Burkina Faso	Guinea - Bissau	Mozambique	
	Rwanda	Chad	Kenya	Sao Tome and Principe	

where $u_{i,t}$ follows AR(1) processes given by

$$u_{i,t} = \phi_i u_{i,t-1} + \varepsilon_{i,t}$$
$$E(\varepsilon_{i,t}\varepsilon_{j,t-s}) = 0 \; for \;\; i \neq j \;\; ,s > 0$$

i denotes the countries and r denotes the regions reported in Table 1. The factor loadings, λ_i, reflect the degree to which variation of $y_{i,t}$ can be explained by each factor.

The estimation process further relies on the starting values for the parameters that in our case are:

$$\lambda_t = random(0,1) \quad \Sigma_\varepsilon = I_N \quad \Sigma_\eta = I_r \quad T = 0.1 \times random(0,1) + 0.1$$
$$\phi^{(o)} = 0.1 \times random(0,1) \quad \phi^{(1)} = 0.1 \times random(0,1) \quad \phi^{(*)} = 0.1 \times random(0,1);$$

see formula (20) and (21) of Appendix C.

The huge number of parameters and the random starting points give multiple solutions for our likelihood, and most of them could be suboptimal. We address this important issues in subsection 3.1. We present the results for the Kose et al. (2004) data set (60 countries with time span 1960-1990) and the full data set (146 countries with time span 1950-2004) in subsection 3.2 and 3.3 respectively. The diagnostic checking and model fit are presented in Subsection 3.4, the variance decomposition in Subsection 3.5 and finally a summary of our findings is presented in Subsection 3.6. The estimation time for all the considered models is reported in table B.1 Appendix B.

3.1 Likelihood Stability

The random starting values give different maxima, this kind of behaviour is quite common with our parameters dimension. To find the highest maximum we repeat the maximization 100 times, every time with random starting values, moreover, we have to investigate the parameters stability. As a measure of parameters variation we use an Euclidean distance normalized by the number of replications:

$$D(\hat{\Theta}^i, \hat{\Theta}^*) = \frac{1}{N} \parallel \hat{\Theta}^i - \hat{\Theta}^* \parallel \tag{13}$$

where N is the number of replications, $\hat{\Theta}^*$ is the parameters value at the chosen highest maximum and $\hat{\Theta}^i$ is the parameters value at the i maximum. We report the likelihood values against iterations and the distance measure in figure 1. The picture shows that the likelihood is moving around a range and the most significant maximum is reached different times for example after 3 and 5 iterations. When we are closed to the highest maximum the distance measure goes to 0, indeed the parameter's values are very near. The results presented in the following sections are taken from the maximum value reached by the likelihood.

3.2 Global and Regional factors using reduced data set

We start this subsection by estimating, with the new technique, the global and regional factors in the same spirit of Kose et al.(2004). The 60 countries used in the study are reported in Appendix A and they are divided accordingly to Table 1. Figure 1 reports the estimated global factor with the 33 and 67 -percent quantile bands. The fluctuations of the factor is very similar to Kose et al.(2004) and reflect the major economic events of those 30 years: the expansion in the period of the 1960's, the recession of the mid-1970's (same period of the first oil price shock), the strong recession in the early 1980's, caused by the debt crisis and the tight monetary policies started in Usa around 1979.

As in previous studies the estimated global factor confirms that the recession in the early 1980's was stronger than the recession of mid 1970's. The inclusion of the Latin America countries that suffered a lot from the debt crisis of the early 1980's, strongly influenced the global factors. Finally it is clear the downturn of the early 1990's.

Thanks to an efficient treatment of the missing values we can estimate the global and regional factor using all the 146 countries, in this experiment the time period is still 1960-1990. In the bottom graph of figure 3 we report the estimated global factor with the 33 and 67 -percent quantile bands. Looking at the graph we can notice that the movement is quite similar to that of figure 2. The inclusion of more countries does not change a lot the estimate of our global factor. The downturn of early 1970's seems to be less strong, this effect could be due to the inclusion of Asia and Oceania Developed countries, that in the 70's experimented a long growth period, see figure 4.

Figure 1: *Upper graph: Likelihood variation against iteration. The straight line indicates the maximum value reached by the likelihood. Bottom graph: Distance measure for parameter's variation*

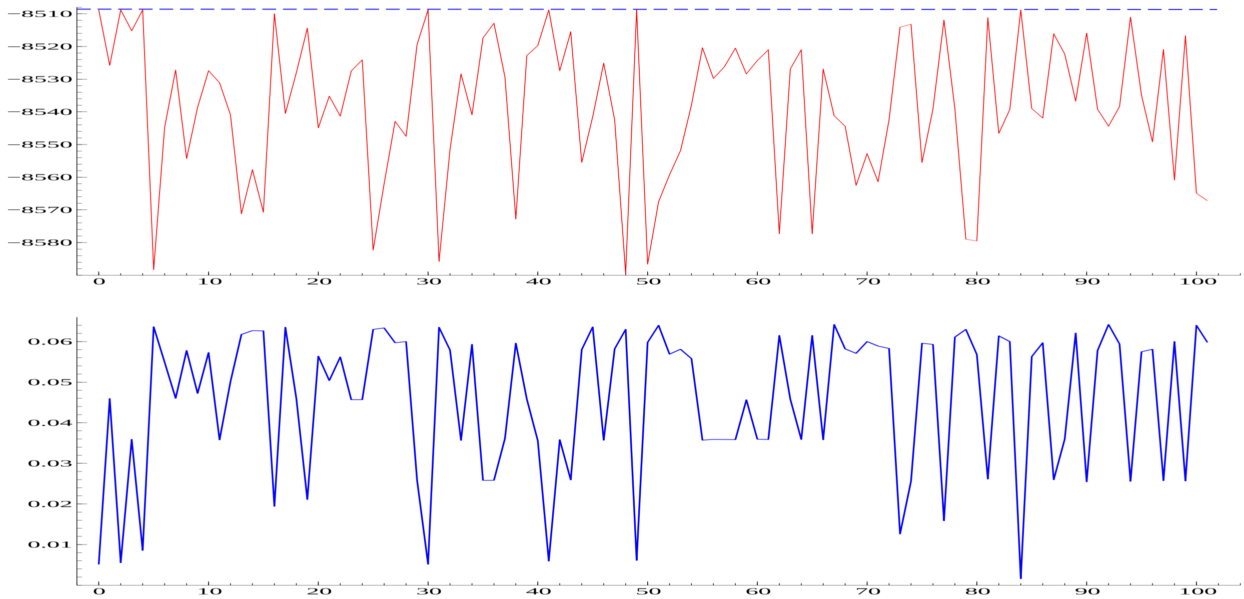
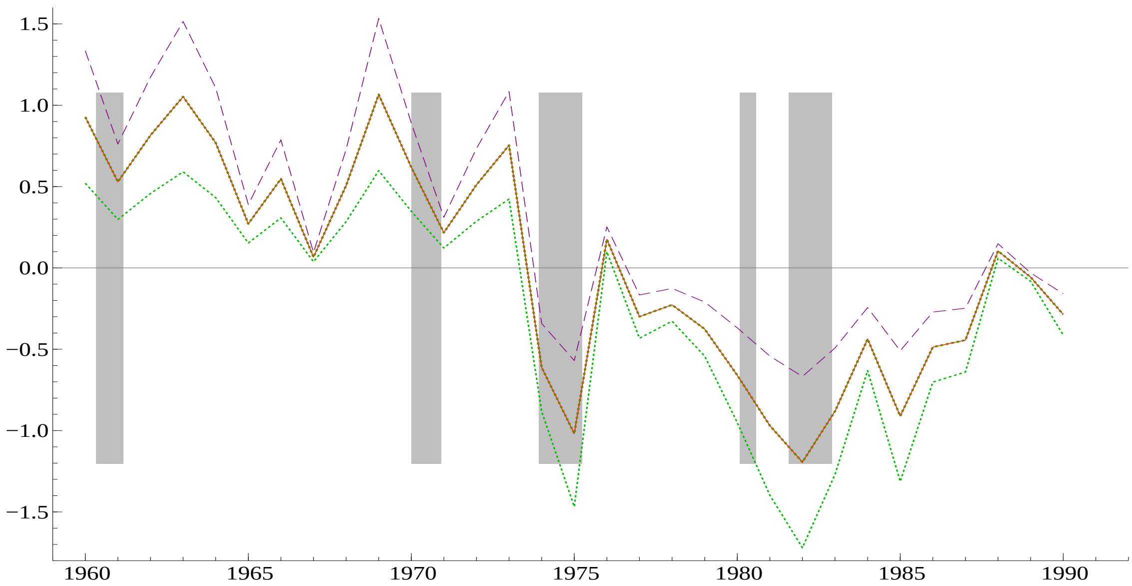


Figure 2: *World Factor plus 33 percent and 67 percent confidence interval, estimated using the Kose et al(2004) data set. We report the NBER recessions with the vertical lines.*



3.3 Global and Regional factors using complete data set and all the time span

In this section we estimate the global and regional factors using all the data set for whole period. Figure 3 reports the estimation of the global factor with the recessions as vertical lines. We use Usa recessions as a proxy of global recessions until 1985, before that year the global recession date are not available, see *http://www.imf.org/*. Figure 4 shows the estimation of all the regional factors.

The global factor reflects the major economic events from the 1950 until 2004. The 1958 recession, the expansion in the period of the 1960's, the downturn of early 1970's, the recession of the mid-1970's and the strong recession in the early 1980's. Moreover the figure shows the Great Moderation period that took place around 1984 and the global recessions that according to the IMF chronology, see *http://www.imf.org/*, corresponds to: 1990-1993, 1998 and 2001-2002. Figure 3, middle graph, we reports the estimated factor using all the 146 countries for the time period 1950-2004, but limited around 1960-1990. This graph is very similar to the global factor estimated using 146 countries for time period 1960-1990, bottom part of figure 3. This shows that using more observations gives no different picture of our global factor, this is useful to asses the robustness of our estimation technique. To study the economic evolution of every single area we report in figure 4 the regional factors.

The upper left part of figure 4 reports the North America regional factor with NBER recessions as grey vertical line, as it is clear, this factor closely follows those recessions. In the last part of the graph we can notice a long period of growth until 2000 that coincides with the Clinton Era that strongly interested this area. After this period we have the 2001 recession caused by the collapsed of Dot-com bubble and September 2001 attacks. This recession seems to be different from the recent one, in fact, it has been neither so strong and nor so persistent. This results is in line with some findings in the literature, see Nordhaus(2002). Interestingly the downturn of 1994-1995 is not a Usa recessions but a strong drop of the Mexico's GDP.

Factor 2, Latin America regional area, shows the debt crisis of the 1980's with decrement in the factor started around 1979 until 1983, see Weeks(2000). The trough clearly showed by the figure corresponds to a severe recession that took place in this area around 1989. Moreover it is clear the slowdown in the factor with the two troughs of 1999, the same timing of Argentina's GDP decrease, and the though of 2002, the same timing of the Argentina's default and the global recession.

About European region, factor 3 shows that those countries suffer a lot from the recession of the mid-70's but less from the recession of the early 1980's. Moreover, it displays the decline starting from 1990 and ending with a deep value of the factor in the 1992-1993, the same timing of the EMS crisis, see Eichengreen(2001). As other areas in the world the European region was interested by the 2001 recession.

Factor 4, Asia and Oceania Developed regional Area, shows the high growth period of the 1970's. Moreover the debt crisis of the early 1980's seems to affect marginally this area. This factor shows clearly the financial crisis that took place in those countries around 1997. This region has been strongly affected by the 2001 global recession.

About Asia and Oceania Developing and Poor and Africa those regional factors seem to follow a different path, these findings are in line with Kose et al.(2004). We will justify this more rigorously in subsections 3.4 and 3.5.

Table 2 reports the estimated VAR coefficients together with the eigenvalues organized in descending order. Look-

Figure 3: *Upper graph: World Factor estimated using all 146 countries for time period 1950-2004 plus 33 percent and 67 percent confidence interval. Gray vertical line, USA recessions, Blue vertical lines global recessions. Middle graph: World Factor estimated using all 146 countries for time period 1950-2004 but zoomed around 1960-1990 plus 33 percent and 67 percent confidence interval. Gray vertical line USA recessions. Bottom graph: World graph estimated using all 146 countries but using sub-sample 1960-1990 plus 33 percent and 67 percent confidence interval. Gray vertical line USA recessions.*

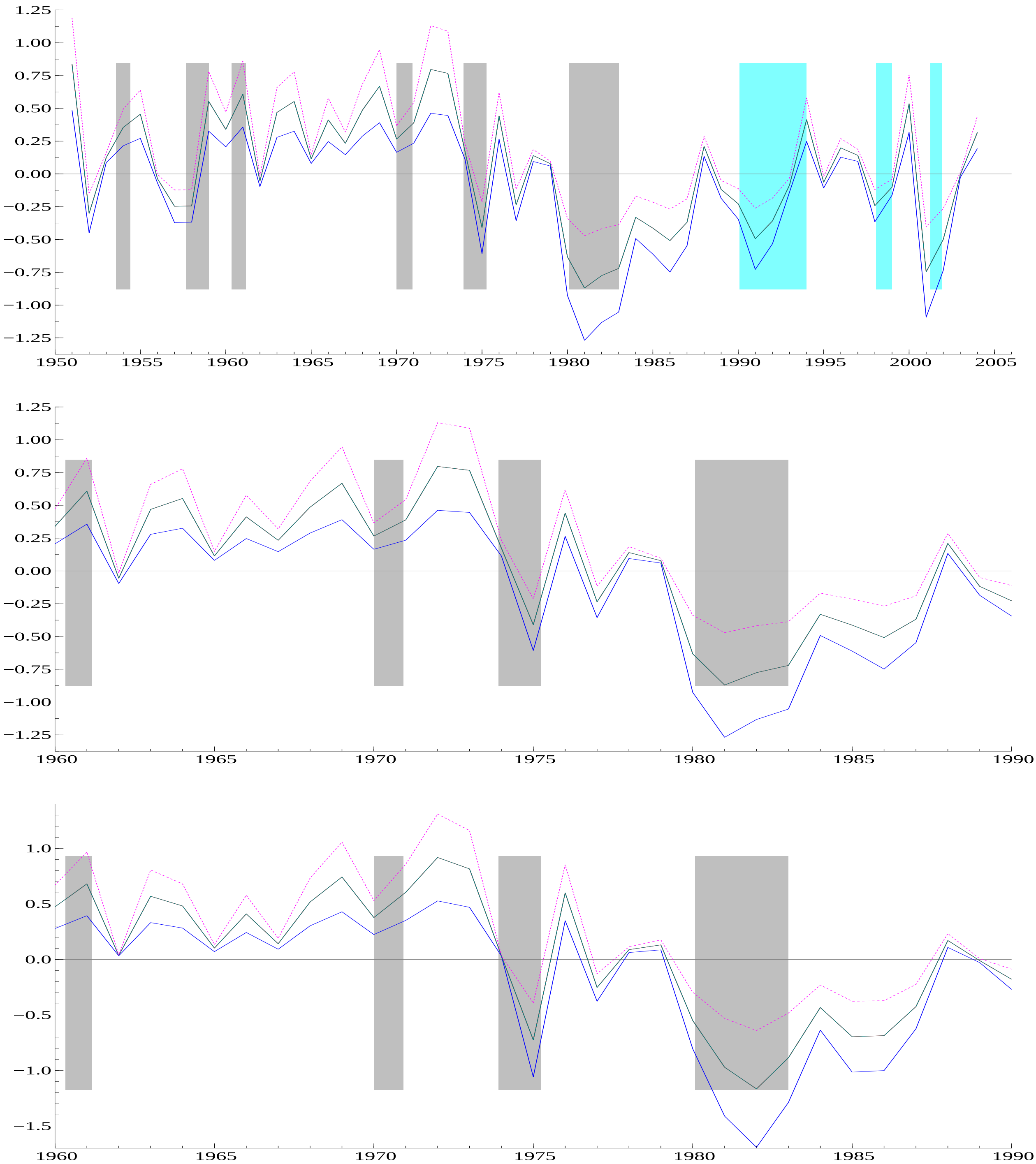


Figure 4: *Regional Factors with 33-percent and 67-percent confidence intervals. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region. The vertical lines are the NBER USA recessions that we report just for the North America Region.*

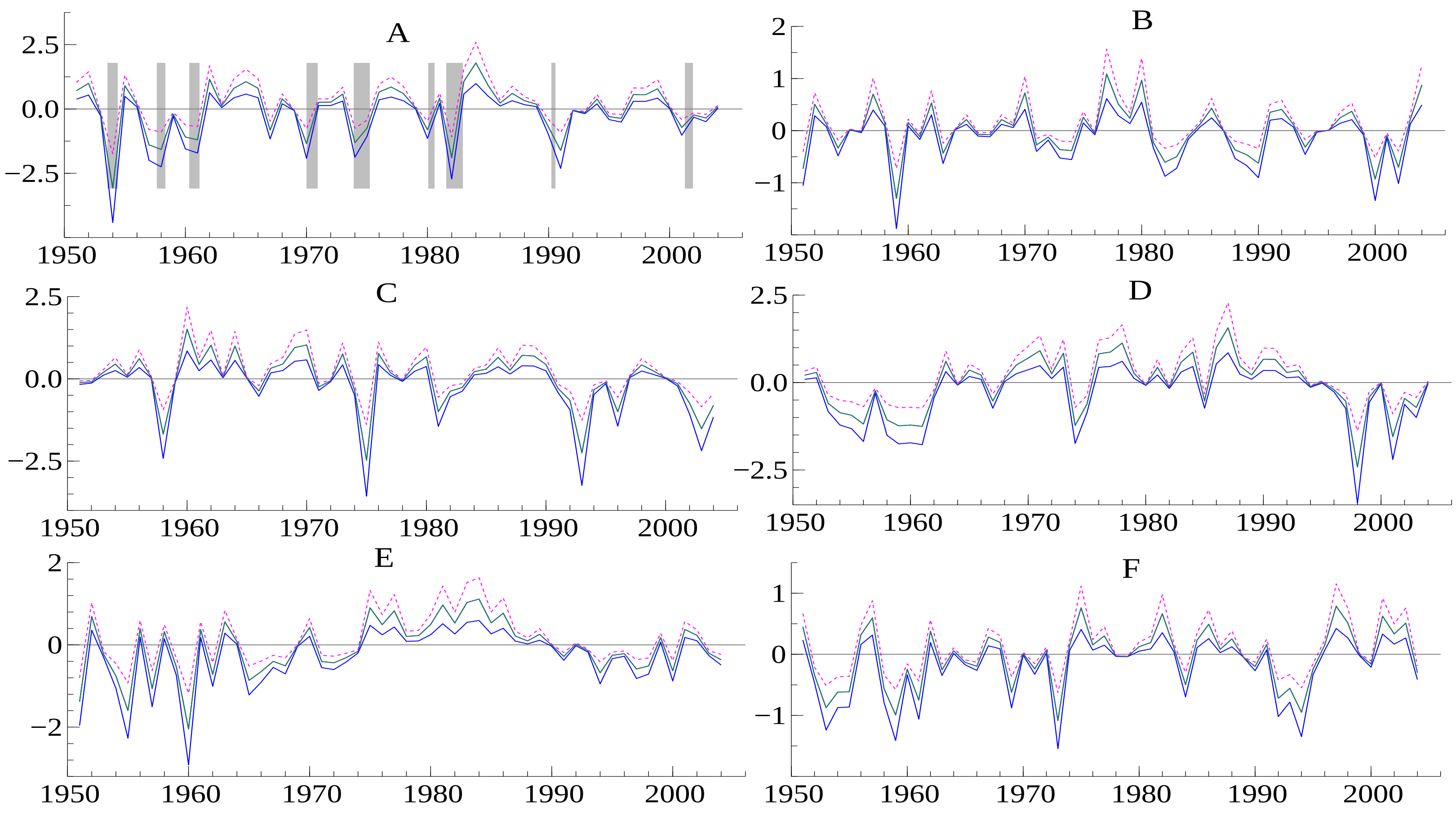


Table 2: Estimated VAR coefficients and Eigenvalues for 146 countries, period 1950-2004. Real is the real part of the eigenvalues, that ranges between a maximum of 0.60 and a minimum of 0.03. Img is the complex conjugate for the eigenvalues.

	Var Coefficients							Eigenvalues		
Factor	1	2	3	4	5	6	7		*Real*	*Img*
	0.228	0.003	0.066	0.002	0.002	0.045	0.091	Eig_1	0.60	0.00
	0.056	0.122	0.036	0.049	0.100	0.095	0.196	Eig_2	0.26	0.00
	0.111	0.008	0.153	0.065	0.025	0.032	0.057	Eig_3	0.12	0.02
	0.224	0.098	0.083	0.231	0.207	0.017	0.053	Eig_4	0.12	-0.02
	0.005	0.177	0.078	0.036	0.234	0.143	0.100	Eig_5	0.06	0.08
	0.093	0.052	0.055	0.109	0.064	0.125	0.197	Eig_6	0.06	-0.08
	0.084	0.046	0.064	0.008	0.051	0.061	0.184	Eig_7	0.03	0.00

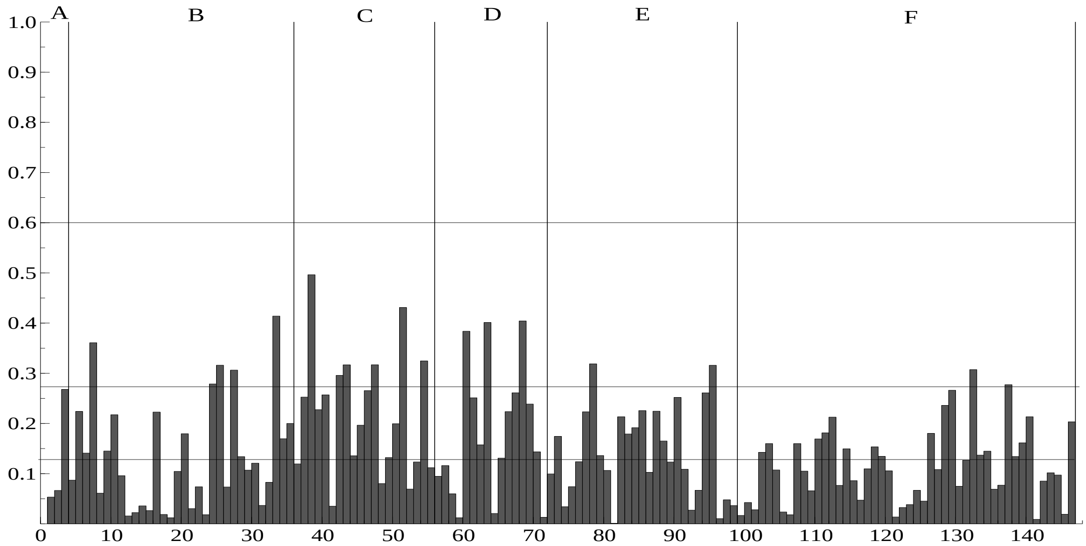
ing at the table we can conclude that the factors are estimated as stationary and they seem to be quite persistent, the largest eigenvalue is around 0.60. We find the presence of two not persistent cyclical behaviour in the factors since conjugate pair of complex eigenvalues are obtained whereas the real part is equal to 0.12 and 0.06. Figure 5 reports in a bar plot the estimate of the autoregressive parameter for our state space form that is used to calculate the percentage of country specific component. To facilitate the reading of this graph we divide it in the six regional areas using vertical lines. The autoregressive parameters range from a minimum around 0.05 to a maximum around 0.5. This figure reports in the horizontal line the 3 biggest eigenvalues of table 2.

3.4 Diagnostic checking and model fit

In this subsection we discuss the model fit and the model diagnostics using the R^2 and Ljung-Box statistic. The actual estimate of Λ is not easy to interpret and therefore Stock and Watson(2002) proposed to focus on the R^2 goodness-of-fit statistics which is obtained by regressing the univariate time series $y_{i,t}$, for each $i = 1, ..., N$, on a constant and a particular principal component estimate. These R^2 statistics are then regarded as proxies for the correlations (in absolute values) between the series and each principal component. In our modelling framework, we can evaluate the correlations between the series and each factor directly. The N regressions can be repeated for each principal component and the resulting N dimensional series of R^2 statistics can be displayed as an index plot for each principal component. We present the N series of R^2 statistics for the seven factors, in case of unrestricted and restricted Λ in figure 6 and in figure 7. To make it more readable the R^2 is split based on the areas.

Figure 7 shows the R^2, in the case of restricted Λ, for the global (left hand side) and the regional (right hand side) factors, divided accordingly to the different areas. The global factor is quite correlated with USA, the value is around 0.20. Canada is less correlated with the global factor, the value is around 0.15. Mexico shows a good link with the global factor with a value for the global R^2 around 0.22. More interestingly the regional R^2, is high correlated with USA and Canada with values around 0.75 for USA and 0.90 for Canada. This is quite natural if we consider that

Figure 5: *Red bars: ϕ_1 parameters for the autoregressive component, see formula (12). The vertical lines divide the ϕ_1 parameters between the areas. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region. The horizontal lines reports the largest eigenvalues estimate using the maximum likelihood, see Table 2.*



Canadian economy is strongly linked to USA. Mexico seems not to share a lot with the regional factor.

The global factor is not very correlated with the Latin America countries, this seems reasonable if we consider the fact that almost all of those countries are very poor. Moreover the regional factor seems not to be very important for this area.

The global factor has an important effect in Europe with highest value around 0.60. Germany presents a good link with the global factor similar to France and less than the UK., the global factor is very important for UK and Italy. The importance of global factor in the Italian economy is due to the export oriented type of this economy. The regional factor is very important for the European countries in fact Germany, France, Italy, and Belgium are very correlated with the regional factor. United Kingdom is not influenced a lot by the European regional factor. We will analyze deeply those findings in subsection 3.5.

The correlation between the Asia and Oceania Developed country with respect to the global factor is quite interesting. The country with more correlation is the Saudi Arabia, this seems reasonable if we consider the fact that Saudi Arabia's economy is petroleum-based and almost the 90 percent of export earnings come from the oil industry. Then, among other, Taiwan, Hong Kong and Japan are correlated with the global factor in this area. Two special cases are: Japan that seems to be quite correlated with the global factor but is not influenced by the regional one, and China that is not correlated with both factors. It seems that all its variability is explained by the country specific component, see subsection 3.5.

The correlation of Asia and Oceania Developing and Poor region with the global factor is small, with null correla-

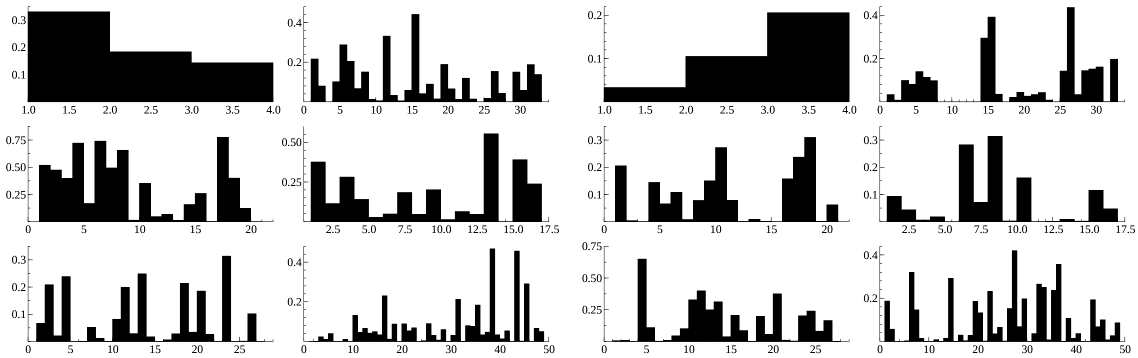


Figure 6: *Global and Regional Factors in the unrestricted case. Left graph: R^2 for the Global Factor divided by areas accordingly to Table 1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region. Right graph: R^2 for the Regional Factor divided by areas accordingly to Table 1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region.*

tion for some countries like North Korea, Pakistan and Iraq. Finally the Africa macro area has not a strong correlation with the global factor, and the regional factor seems not to be important. For most of them the countries specific effect is the prominent element that explains the economic fluctuations, see subsection 3.5. Finally looking at figure 6 and 7 we can notice the difference in the R^2 for the unrestricted and restricted case. The regional R^2 are very different between the two figures, showing the importance of the restriction in the loading matrix in order to disentangle global and regional factors.

Another advantage of this framework is to easily account for model misspecification tests and diagnostics concerning normality, heteroskedasticity and serial correlation and it can be seen as an effective tool for model selection. The Kalman filter allows us to calculate in few seconds the prediction errors for our data set even in presence of missing values. Thanks to this tool we can carry out easily the Ljung-Box test (1978). The Ljung-Box Q(q) statistic is based on the first q sample autocorrelations r_k^*, $k = 1, ..., q$ of the residual series and is computed by $Q(q) = \sum_{k=1}^{q} r_k^2$.

The Ljung-Box statistics for the 146 series is presented as index plot in figure 8 for q = 12. Almost all the series are in the confidence interval, and we can conclude that the our specification is successful in capturing the collective dynamics in our data set. On the other hand some series have very high value for this statistic. For example the highest value is about North Korea, in this case the model does not fit a country with this small economic dimension and untrustable GDP. Another important exception is Japan, the high value of the statistic can be easily explained if we consider that Japanese economy seems to be more detached from other industrialized countries with the domestic shocks that explain big portion of its volatility, see subsection 3.5.

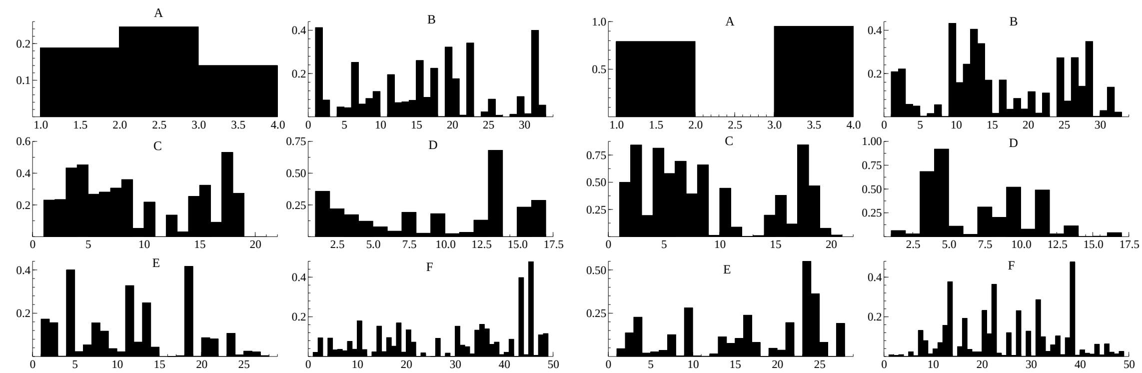


Figure 7: *Global and Regional Factors in the restricted case. Left graph: R^2 for the Global Factor divided by areas accordingly to Table 1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region. Right graph: R^2 for the Regional Factor divided by areas accordingly to Table 1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region.*

3.5 Variance Decomposition

To measure the relative contributions of the world, regional and country factors to variations in aggregate variables for each country, we estimate the share of the variance of each macroeconomic aggregate due to each factor. We decompose the variance of each observable into the fraction that is due to each global, regional and the country specific factor. With orthogonal factors the variance of observable i can be written in the following way:

$$var(y_{i,t}) = (\lambda_i^{world})^2 var(\alpha_t^{world}) + (\lambda_i^{region})^2 var(\alpha_{r,t})^2 + var(country_{r,t}) \tag{14}$$

where r is the region and the variance of the country component is given by the unconditional variance of AR(1) process. The fraction of the volatility explained by the global factor is given by:

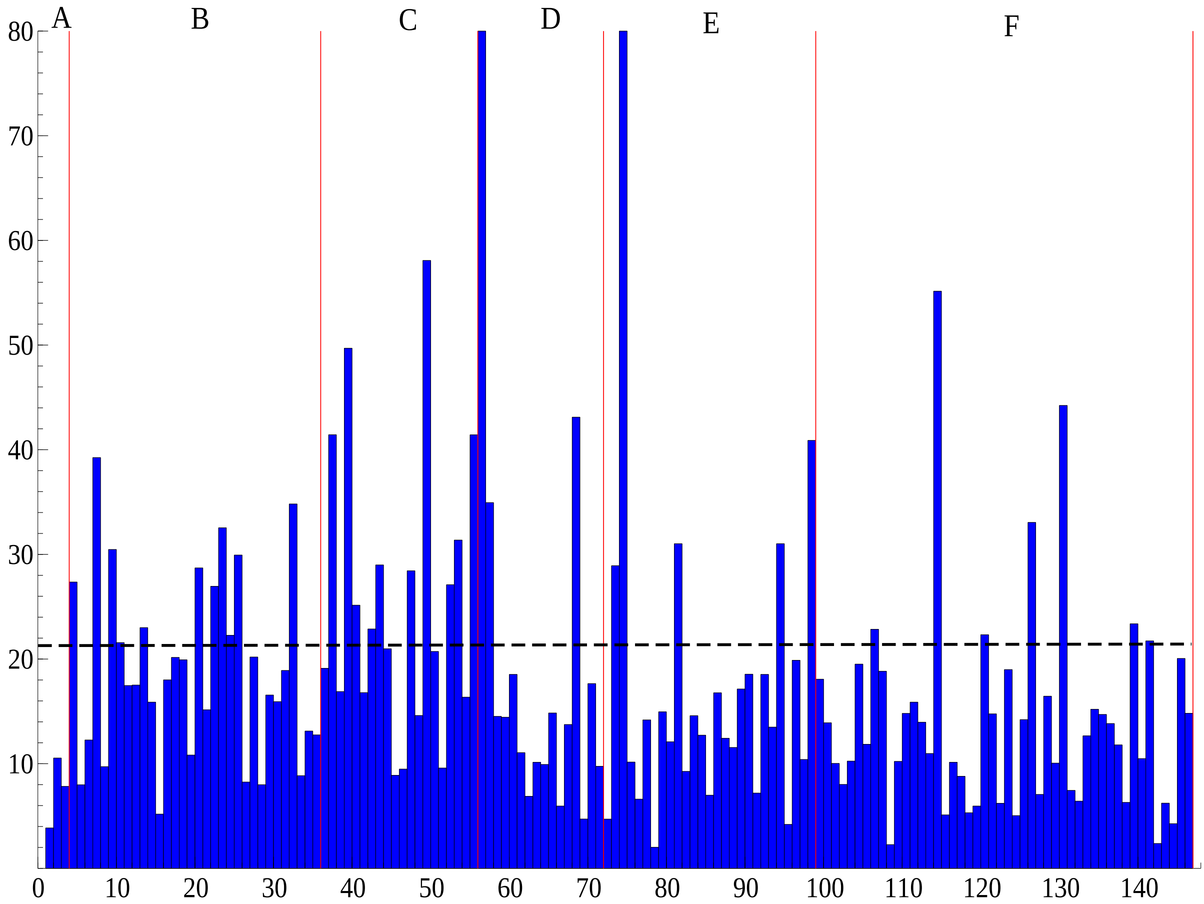
$$\frac{(\lambda_i)^2 var(\alpha_t^{World})}{var(y_{i,t})} \tag{15}$$

this measure is calculated using the parameters estimated using maximum likelihood technique.

We show the variance attributable to each factor for the main countries in table 3. This table displays the variance decomposition for the period 1950-2004 for all the 146 countries and the variance decomposition period 1960-1990 for the Kose et al.(2004) data set. As measure of the importance of the factors we report the 33-percent and the 67-percent quantiles that are calculated based on a Gaussian approximation, therefore it is not guaranteed that they are in the bounds. The full tables are available from the authors upon request.

Table 3, shows that the portion of variability explained by the global factor for the USA economy has increased between the two periods. The global factor seems to be more important for the USA in the period 1950-2004. The portion of variability explained by the regional factor has increased as well. Canada seems to be less correlated with the global factor in the period 1950-2004, but more correlated with the regional one. As far as Mexico is concerned

Figure 8: *Ljung-Box Q(12) for the generalized least squares residuals of dynamic factor model. The displayed Ljung-Box values are truncated at 80. The vertical lines divide the regional areas accordingly to: (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region. The horizontal line is the 95 percent confidence interval*



the share of variability explained by the global factor is almost equal in the two periods, but the percentage of regional factor is decreasing, this fact confirms the findings in subsection 3.4.

As far as EU is concerned we can notice that the volatility explained by the global factor has decreased between the two periods for almost all the countries. We can notice that the variability explained by the regional factor is increased for Germany, Italy and France. One important exception is the UK where the variability explained by the global factor is substantially increased between the two periods and the variability explained by the regional factor has decreased. Those values confirm the findings by Stock and Watson(2005), in particular the UK seems not to be related to European factor any more.

About Latin America countries it is interesting to see the evolution of variance decomposition for Brazil and Venezuela. In Brazil the percentage of variability explained by the global factor has increased between the two periods and the regional factor has slightly decreased. Venezuela experimented a substantial decrease in the global factor, a substantially stable value for the regional factor and an increment in the country specific component. This effect could be due to the Venezuela crisis in the 80's and the corresponding slowdown of the economic activities. Moreover other countries like Chile seems to be more detached from the global factor with an increasing share in the regional one.

As far as Asia and Oceania Developed is concerned, Japan is less influenced by the global factor during the period 1950-2004, and the regional factor has not strongly been influenced as well. Great proportion of its variability is explained by country specific component. If we compare those values with those corresponding to the period 1960-1990 we can notice that during the 1980's and 1990's, cyclical fluctuations in Japanese GDP became almost detached from the global factor, with domestic shocks explaining big portion of the cyclical movement of Japanese GDP. The increment in the portion of variability explained by the regional factor is consistent with Asian trade being increasingly important for Japanese economy. Hong Kong suffered from a substantial decrease in the percentage of the global factor between the two periods and a substantial increase in the regional factor. This suggests a stronger link of this economy with the other countries in this area. One important case is China that seems not to be influenced by the global and regional factor, all its variability is explained by the country specific component. South Korea experimented a substantial increase of the global and the regional factor, between the two periods, moreover the country specific component has decreased substantially. Those results are in line with the Korean economy, see Pecotich and Shultz(2006), in fact Korea is the seventh largest trading partner of the United States and the eighth largest trading partner of the European Union, moreover is the Asia's biggest exporter of refined oil products.
About Asia and Oceania Developing and Poor and Africa regions almost all the variability is explained by the country specific factor, therefore those countries seem to be detached from the world economy. One important exception is the Philippines. In this country the global and regional factor seems to increase between the two periods, and the idiosyncratic component decreased substantially. This effect could be explained by the fact that during the 1960s, the economy was regarded as the second largest in Asia, second to Japan. However, the leadership of Ferdinand Marcos proved disastrous, by transforming the market economy into a centrally planned economy. The country suffered severe economic recession, and only recovered in the 1990s with a program of economic liberalization, see Gargan(1997).

Table 3: *Variance Decomposition for different data sets. The arrows indicates the variation between periods of the Variance decomposition. The country reported with the (*) are the same countries used in Kose et al.(2004). We report the complete countries for the North America and Europe Region and a selection of countries for the other regions.*

Variance Decompositions for North America Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
United States(*)	1960-1990	16.110	28.763	41.417		25.784	46.036	66.288		14.115	25.201	36.287	
	1950-2004	20.511	36.620	52.730	⇑	21.020	37.530	54.039	⇓	14.478	25.850	37.222	⇑
Canada(*)	1960-1990	21.367	38.149	54.931		19.170	34.226	49.283		15.472	27.625	39.777	
	1950-2004	15.502	27.677	39.853	⇓	19.943	35.608	51.272	⇑	20.564	36.715	52.866	⇑
Mexico(*)	1960-1990	13.687	24.438	35.188		0.511	0.912	1.314		41.810	74.650	107.49	
	1950-2004	11.406	20.365	29.324	⇓	0.140	0.251	0.361	⇓	44.462	79.384	114.31	⇑

Variance Decompositions for European Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
Germany(*)	1960-1990	21.312	38.052	54.791		7.296	13.028	18.759		27.400	48.920	70.441	
	1950-2004	21.143	37.749	54.355	⇓	16.973	30.304	43.636	⇑	17.893	31.947	46.001	⇓
France(*)	1960-1990	42.194	75.334	108.470		8.055	14.382	20.709		5.759	10.284	14.808	
	1950-2004	13.379	23.888	34.396	⇓	21.308	38.045	54.781	⇑	21.321	38.067	54.814	⇑
Italy(*)	1960-1990	31.596	56.412	81.228		4.943	8.826	12.709		19.470	34.762	50.054	
	1950-2004	27.927	49.862	71.797	⇓	13.689	24.442	35.194	⇑	14.392	25.696	37.000	⇓
UK(*)	1960-1990	3.323	5.931	8.541		12.217	21.813	31.409		40.469	72.255	104.04	
	1950-2004	27.479	49.063	70.646	⇑	5.192	9.270	13.348	⇓	23.337	41.667	59.997	⇓
Sweden(*)	1960-1990	20.344	36.323	52.302		1.425	2.545	3.664		34.239	61.132	88.024	
	1950-2004	21.695	38.735	55.776	⇑	7.8219	13.966	20.109	⇑	26.492	47.299	68.107	⇑
Switzerland(*)	1960-1990	8.9906	16.052	23.114		17.497	31.240	44.982		29.521	52.708	75.895	
	1950-2004	29.273	52.266	75.258	⇑	9.6299	17.194	24.757	⇓	17.106	30.541	43.976	⇓
Spain(*)	1960-1990	30.670	54.760	78.849		1.833	3.273	4.714		3.505	41.966	60.428	
	1950-2004	5.8527	10.450	15.047	⇓	10.949	19.548	28.148	⇑	39.207	70.002	100.80	⇑
Portugal(*)	1960-1990	8.2862	14.794	21.303		21.374	38.161	54.949		26.349	47.044	67.740	
	1950-2004	20.593	36.767	52.942	⇑	12.368	22.083	31.798	⇓	23.047	41.150	59.252	⇓
Norway(*)	1960-1990	2.7341	4.881	7.029		0.007	0.013	0.019		53.267	95.105	136.94	
	1950-2004	15.646	27.934	40.223	⇑	2.0317	3.6274	5.2232	⇑	38.331	68.438	98.545	⇓
Netherlands(*)	1960-1990	25.933	46.301	66.670		4.3046	7.6856	11.067		25.771	46.013	66.255	
	1950-2004	9.0096	16.086	23.162	⇓	21.853	39.016	56.180	⇑	25.147	44.898	64.649	⇓
Luxembourg(*)	1960-1990	1.241	2.217	3.192		24.082	42.997	61.912		30.685	54.786	78.887	
	1950-2004	0.007	0.013	0.019	⇓	16.731	29.873	43.014	⇓	39.270	70.113	100.96	⇑
Ireland(*)	1960-1990	23.314	41.626	59.938		2.8008	5.0007	7.2005		29.894	53.373	76.853	
	1950-2004	14.072	25.125	36.178	⇓	5.6188	10.032	14.445	⇑	36.318	64.843	93.368	⇑
Iceland(*)	1960-1990	11.216	20.025	28.834		1.061	1.894	2.728		43.732	78.080	112.43	
	1950-2004	12.254	21.879	31.504	⇑	0.282	0.503	0.725	⇓	43.472	77.617	111.76	⇓
Greece(*)	1960-1990	18.717	33.417	48.118		3.4268	6.118	8.809		33.865	60.464	87.063	
	1950-2004	15.223	27.179	39.136	⇓	1.155	2.062	2.969	⇓	39.631	70.759	101.89	⇓
Finland(*)	1960-1990	13.185	23.541	33.897		7.0555	12.597	18.139		35.768	63.862	91.956	
	1950-2004	17.051	30.444	43.837	⇑	6.1953	11.061	15.927	⇓	32.762	58.495	84.227	⇓
Denmark(*)	1960-1990	8.0748	14.417	20.759		17.512	31.267	45.022		30.421	54.315	78.210	
	1950-2004	18.591	33.193	47.795	⇑	10.168	18.155	26.141	⇓	27.249	48.652	70.055	⇓
Belgium(*)	1960-1990	37.097	66.234	95.372		10.170	18.158	26.146		8.7415	15.607	22.473	
	1950-2004	33.461	59.742	86.024	⇓	15.975	28.522	41.069	⇑	6.5730	11.736	16.898	⇓
Austria(*)	1960-1990	20.819	37.171	53.524		7.5165	13.420	19.324		27.673	49.408	71.144	
	1950-2004	25.942	46.317	66.693	⇑	3.916	6.993	10.070	⇓	26.150	46.690	67.229	⇓
Cyprus	1960-1990	N.A	N.A	N.A		N.A	N.A	N.A		N.A	N.A	N.A	
	1950-2004	4.4055	7.8657	11.326		4.7961	8.5632	12.330		46.807	83.571	120.34	
Malta	1960-1990	N.A	N.A	N.A		N.A	N.A	N.A		N.A	N.A	N.A	
	1950-2004	18.985	33.897	48.808		3.9073	6.9762	10.045		33.116	59.127	85.138	

Variance Decompositions for Latin America Region													
		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
Brazil(*)	1960-1990	14.860	26.532	38.204		11.401	20.355	29.310		29.748	53.112	76.477	
	1950-2004	18.887	33.721	48.555	⇑	9.4491	16.871	24.292	⇓	27.673	49.409	71.144	⇓
Argentina(*)	1960-1990	16.601	29.640	42.679		9.6144	17.166	24.717		29.793	53.194	76.595	
	1950-2004	19.456	34.738	50.019	⇑	15.818	28.242	40.667	⇑	20.734	37.020	53.306	⇓
Bolivia(*)	1960-1990	10.028	17.904	25.780		2.330	4.161	5.991		43.650	77.935	112.22	
	1950-2004	8.482	15.145	21.807	⇓	0.071	0.127	0.182	⇓	47.455	84.728	122.00	⇑
Chile(*)	1960-1990	19.588	34.973	50.358		3.729	6.659	9.588		32.691	58.367	84.044	
	1950-2004	2.3150	4.1332	5.9515	⇓	7.2485	12.942	18.635	⇑	34.198	61.059	87.920	⇑
Venezuela(*)	1960-1990	16.939	30.243	43.547		6.5594	11.711	16.863		32.511	58.046	83.581	
	1950-2004	3.2513	5.8049	8.3586	⇓	7.9750	14.239	20.503	⇓	44.782	79.956	115.13	⇑

Variance Decompositions for Asia and Oceania Developed Region													
		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
Japan(*)	1960-1990	37.734	67.372	97.010		0.028	0.051	0.073		18.246	32.576	46.907	
	1950-2004	26.914	48.054	69.194	⇓	1.601	2.860	4.118	⇑	27.492	49.086	70.679	⇑
China	1960-1990	N.A.	N.A.	N.A.		N.A.	N.A.	N.A.		N.A.	N.A.	N.A.	
	1950-2004	0.648	1.157	1.667		3.502	6.254	9.005		51.857	92.588	133.32	
Taiwan	1960-1990	N.A.	N.A.	N.A.		N.A.	N.A.	N.A.		N.A.	N.A.	N.A.	
	1950-2004	16.176	28.882	41.587		18.362	32.783	47.205		21.471	38.335	55.199	
Hong Kong(*)	1960-1990	30.580	54.599	78.618		0.229	0.409	0.590		25.199	44.991	64.783	
	1950-2004	4.031	7.197	10.363	⇑	25.695	45.877	66.060	⇑	26.282	46.925	67.568	⇑
South Korea (*)	1960-1990	0.160	0.286	0.412		10.688	19.083	27.477		45.160	80.631	116.10	
	1950-2004	15.840	28.282	40.724	⇑	17.027	30.401	43.775	⇑	23.141	41.317	59.492	⇓

Variance Decomposition for Asia and Oceania Developing and Poor Region													
		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
India(*)	1960-1990	10.845	19.364	27.882		18.356	32.774	47.191		26.807	47.863	68.918	
	1950-2004	10.035	17.916	25.798	⇓	10.396	18.561	26.727	⇓	35.578	63.523	91.467	⇑
Indonesia(*)	1960-1990	9.2150	16.453	23.691		1.229	2.195	3.161		45.564	81.352	117.14	
	1950-2004	12.194	21.771	31.349	⇑	7.861	14.036	20.211	⇑	35.953	64.192	92.431	⇓
Philip.(*)	1960-1990	19.527	34.865	50.202		0.735	1.313	1.891		35.746	63.822	91.898	
	1950-2004	25.095	44.805	64.516	⇑	4.822	8.611	12.399	⇑	26.091	46.584	67.076	⇓
Pakistan(*)	1960-1990	10.621	18.964	27.306		4.187	7.475	10.765		41.200	73.561	105.92	
	1950-2004	8.4249	15.042	21.659	⇓	6.353	11.343	16.333	⇑	41.231	73.615	106.00	⇑

Variance Decompositions for Africa Region													
		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
S.Africa(*)	1960-1990	11.513	20.556	29.598		5.047	9.011	12.976		39.449	70.433	101.42	
	1950-2004	14.415	25.737	37.058	⇑	16.671	29.766	42.860	⇑	24.923	44.498	64.073	⇓
Egypt	1960-1990	N.A	N.A.	N.A.		N.A.	N.A.	N.A.		N.A.	N.A.	N.A.	
	1950-2004	2.881	5.144	7.406		4.733	8.451	12.169		48.394	86.405	124.42	
Morocco(*)	1960-1990	22.362	39.926	57.491		8.458	15.102	21.746		25.188	44.971	64.755	
	1950-2004	3.188	5.693	8.198	⇓	10.374	18.522	26.670	⇑	42.446	75.785	109.12	⇑
Nigeria	1960-1990	N.A.	N.A.	N.A.		N.A.	N.A.	N.A.		N.A.	N.A.	N.A.	
	1950-2004	8.0802	14.427	20.773		2.173	3.881	5.588		45.755	81.692	117.63	
Algeria	1960-1990	N.A.	N.A.	N.A.		N.A.	N.A.	N.A.		N.A.	N.A.	N.A.	
	1950-2004	5.791	10.340	14.889		0.023	0.042	0.061		50.193	89.617	129.04	

3.6 Summary of the findings

Finally we report in the following table our main findings, which agree with Stock and Watson(2005) and Kose et al.(2004) using respectively sample 1950-2004 with 146 countries and sub sample 1960-1990 with 60 countries.

Stock and Watson (2005)	Kose et al.(2004)
1) There is evidence of an emerging "Euro-zone only" factor.	1) Their results indicate that there is a distinct world business cycle. The world factor seems to account for a significant fraction of output growth fluctuations in many countries. The factor is quite persistent, and reflects many major worldwide economic events.
2) United Kingdom is less correlated with the European Union Countries and more correlated with Usa and Canada, hence there is evidence of an emerging English speaking region.	2) The European regional factor plays a relatively minor role in accounting for the economic activities, there is no strong evidence for a European regional factor.
3) As far as Japan is concerned, international shocks have become unimportant, and domestic shocks explains nearly all of its volatility.	3) Japan is the only outlier in Asia; its world factor is much more important, and the country and idiosyncratic factors less important, than in the rest of the region. 4) The world factor explains little of output variation in most African economies; evidently, African economic fluctuations are not like those in most of the rest of the world.

4 Conclusion

In this paper we employ a new Maximum Likelihood-based inference to estimate the latent factors and to study the dynamic comovement of macroeconomic aggregates in a broad cross section of countries. We provide an analysis of comovement across the world and across regions, for different periods and different cross section dimension. Our paper also makes a methodological contribution as it provides a useful framework to study factors in a large data set with different pattern of missing values. Using variance decomposition we found evidence of an emerging European cycle, it turns out that a big portion of the volatility of the European aggregates can be attributed to a common European factor. Japan seems to be detached from the global factor and great portion of its variability, during the period 1950-2004, is explained by the country specific component. United Kingdom seems to be more related to the global factor and less with the European one. Finally using state space methods we calculate model misspecification test and diagnostics, from one step head prediction error even in presence of missing data. We applied the Ljung-Box statistic to our data set and we find that this model specification is a good enough to represent our series.

5 Appendix A: Dataset

The data set used in the study as been taken from the Penn World Table (http://pwt.econ.upenn.edu/). The observations are Annual GDP Constant Price: Laspeyres (base year 2000) and span the period between 1950 and 2004. The time series extracted are 146 with 54 observations. All the series are assumed to be I(1) in the logarithm so we differentiate the series, moreover we standardize all of them.

Europe	
Country	Sample
Germany[*,1]	1970–2004
France[*,1]	1950–2004
Italy[*,1]	1950–2004
United Kingdom[*,1]	1950–2004
Sweden[1]	1950–2004
Switzerland[1]	1950–2004
Spain[*,1]	1950–2004
Portugal[1]	1950–2004
Norway[1]	1950–2004
Netherlands[1]	1950–2004
Luxembourg[1]	1950–2004
Ireland[1]	1950–2004
Iceland[1]	1950–2004
Greece[1]	1951–2004
Finland[1]	1950–2004
Denmark[1]	1950–2004
Belgium [1]	1950–2004
Austria [1]	1950–2004
Cyprus	1970–2004
Malta	1970–2004

West and Central Asia	
Country	Sample
Israel	1950–2004
Turkey*	1950–2004
India[*,1]	1950–2003
Pakistan[*,1]	1950–2004
Sri Lanka[1]	1950–2003
Mauritius	1950–2004
Iran*	1955–2003
Jordan	1954–2003
Bangladesh[*,1]	1972–2003
Iraq	1970–2003
Nepal	1960–2003
Oman	1970–2003
U.A.E.	1970–2003
Bahrain	1970–2003
Bhutan	1970–2003
Maldives	1970–2004
Mongolia	1970–2003
Saudi Arabia	1970–2003
Kuwait	1970–2003
Qatar	1970–2003
Syria	1960–2003

America	
Country	Sample
United States[*,1]	1950–2004
Mexico[*,1]	1950–2004
Brazil[*,1]	1950–2003
Argentina[*,1]	1950–2004
Bolivia[1]	1950–2003
Cambodia*	1970–2003
Canada[1]	1950–2004
Chile[1]	1951–2004
Colombia[1]	1950–2003
Cuba	1970–2003
Jamaica[1]	1953–2003
Peru[1]	1950–2003
Paraguay[1]	1951- 2003
Uruguay[1]	1950–2004
Venezuela[1]	1950–2004
Antigua	1950–2004
Belize	1970–2004
Barbados	1960–2004
Costa Rica[1]	1950–2004
Dominica	1970–2003
Dominica Republic[1]	1951–2003
Ecuador[1]	1951–2004
El Salvador[1]	1950–2003
Grenada	1970–2003
Guatemala[1]	1970–2003
Honduras[1]	1950–2004
Nicaragua	1950–2004
Panama[1]	1950–2003
Bahamas	1970–2004
Bermuda	1970–2003
Haiti	1970–2000
Netherlands Antilles	1970–2003
Puerto Rico	1970–2003
Trinidad Tobago[1]	1950–2003
Suriname	1970–2003

Table 4: *[(1)]Corresponds to countries that are used in Kose etal(2004), (*)Corresponds to countries with more than 40 million inhabitants, source http://www.citypopulation.de/*

Africa	
Country	Sample
South Africa[*,1]	1950 - 2004
Egypt*	1950 - 2003
Morocco[*,1]	1950 - 2003
Nigeria*	1950 - 2004
Algeria	1960 - 2003
Central African Republic	1970 - 2003
Cote d‘Ivoire[1]	1960 - 2003
Ethiopia	1950 - 2003
Madagascar	1960 - 2004
Rwanda	1960 - 2003
Senegal[1]	1960 - 2003
Somalia	1970 - 2004
Tunisia	1960 - 2004
Uganda	1950 - 2003
Cameroon[1]	1960 - 2003
Botswana	1970 - 2004
Benin	1959 - 2003
Burundi	1960 - 2003
Burkina Faso	1959 - 2004
Chad	1960 - 2003
Comoros	1960 - 2003
Dem. Rep. Congo*	1970 - 2004
Republic of Congo*	1960 - 2003
Equatorial Guinea	1960 - 2003
Gabon	1960 - 2004
Gambia	1960 - 2004
Ghana	1955 - 2003
Guinea	1959 - 2004
Guinea - Bissau	1960 - 2003
Kenya[1]	1950 - 2003
Liberia	1970 - 2003
Lesotho	1960 - 2003
Malawi	1954 - 2004
Mali	1960 - 2004
Mauritania	1970 - 2003
Namibia	1970 - 2003
Niger	1960 - 2004
Cape Verde	1960 - 2003
Mozambique	1960 - 2003
Sao Tome and Principe	1970 - 2004
Sierra Leone	1970 - 2003
Swaziland	1970 - 2004
Tanzani	1960 - 2003
Togo	1960 - 2004
Zambia	1955 - 2003
Zimbabwe[1]	1954 - 2003
Sudan*	1970 - 2003

East Asia and Oceania	
Country	Sample
Japan[*,1]	1950 - 2004
Australia[1]	1950 - 2004
New Zealand[1]	1950 - 2004
Philippines[*,1]	1950 - 2004
Thailand[*,1]	1950 - 2003
Taiwan	1951 - 2004
China*	1952 - 2004
Indonesia[*,1]	1960 - 2004
Hong Kong[1]	1960 - 2004
Malaysia[1]	1955 - 2003
Singapore[1]	1960 - 2004
Dem. Rep. Korea,	1970 -2003
Republic of Korea[*,1]	1953 - 2004
Brunei	1970 - 2003
Laos	1970 - 2003
Macao	1970 - 2004
Papua New Guinea	1970 - 2003
Kiribati	1970 -2003
Samoa	1970 - 2003
Solomon Islands	1970 - 2003
Tonga	1970 - 2003
Vanuatu	1970 - 2003
Micronesia, Fed. Sts.	1970 - 2003

Table 5: *[(1)]Corresponds to countries that are used in Kose et al(2004), (*) Corresponds to countries with more than 40 million inhabitants, source http://www.citypopulation.de/*

6 Appendix B: Evaluation time

This appendix reports the evaluation time for all the estimated models. The program is written in Ox v. 5.10 console (Doornik (2007)) using our source code. The program is running on a standard notebook with Debian Linux, 2 GB memory and a 2.0 GHz two-core processor.

Table B.1: Evaluation Time

Model	BFGS Iter	Time for 100 iter.	Evaluation Time	Data Points	Missing values	Parameters
Seven factors 60 countries period 1960-1990	129	2 minutes and 10 seconds	2 minutes and 50 seconds	1891	30	361
Seven factors 146 countries period 1960-1990	320	6 minutes	20 minutes	4557	559	879
Seven factors 146 countries period 1950-2004	351	12 minutes and 10 seconds	42 minutes and 35 seconds	7938	1399	879

7 Appendix C: Analytical Score

We start this section by review the state space formulation for dynamic factor model in presence of missing values, this section draws heavily for Jungbacker et al. (2009).
The dynamic factor model given in (1) links the observations vector y_t to a set of unobserved factors f_t for $t = 1,...,n$. We assume that the f_t are a linear combination of an unobserved $p \times 1$ dimensional vector autoregressive process α_t. In particular we have a $q \times p$ selection matrix G that define the dynamic factor as

$$f_t = G\alpha_t \tag{16}$$

where we have for the factors the following state space representation

$$\alpha_{t+1} = T\alpha_t + \eta_t \qquad \eta_t \sim N(0, \Sigma_\eta) \tag{17}$$

In our formulation the α_t is a time-variant state vector. The dynamic factor model (1) can be expressed in state space form like

$$y_t = Z\alpha_t + u_t \tag{18}$$

where $Z = \Lambda G$. The factor loading Λ is treated as fixed and depend on some unknown coefficients vector that has to be estimated.
We assume that the error component u_t follow a VAR(1) process given by

$$u_{t+1} = \phi u_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma_\varepsilon) \tag{19}$$

where ϕ is an $N \times N$ diagonal matrix and the disturbance variance matrix Σ_ε is $N \times N$ diagonal matrix of unknown parameters that has to be estimated.

Now take a vector of time series y_t for $t = 1,...,n$ the following expression $y_t(o_t, o_{t-1})$ indicates the observations present at time t and time $t-1$. In the same way we have $y_t(m_t, m_{t-1})$ for missing values at time t and $t-1$. Moreover we have all the other possible combinations: $y_t(m_t, o_{t-1})$, $y_t(o_t, m_{t-1})$, $y_t(m_{t+1}, o_t)$.
We accomplish this notation with the following state space formulation:

$$\begin{pmatrix} y_t(o_t,o_{t-1}) \\ y_t(o_t,m_{t-1}) \end{pmatrix} = \begin{pmatrix} \phi_t^{(o)} y_t(o_t,o_{t-1}) \\ 0 \end{pmatrix} + \begin{pmatrix} Z(o_t,o_{t-1}) & -\phi_t^{(o)} Z(o_t,o_{t-1}) & 0 & 0 \\ Z(o_t,m_{t-1}) & 0 & I & 0 \end{pmatrix} \dot{\alpha}_t + \begin{pmatrix} \varepsilon(o_t,o_{t-1}) \\ 0 \end{pmatrix} \tag{20}$$

with state equation $\dot{\alpha}_t$ given by:

$$\dot{\alpha_{t+1}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \phi_t^{(*)} y_t(m_{t+1},o_t) \end{pmatrix} + \begin{pmatrix} T & 0 & 0 & 0 \\ I & 0 & 0 & 0 \\ 0 & 0 & 0 & J_t \phi_t^{(1)}(m_t,m_t) \\ \phi_t^{(*)} Z_t^* & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \\ u_t(o_t,m_{t-1}) \\ u(m_t) \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \\ J_t \varepsilon_t(m_t) \\ \varepsilon_t(m_{t+1},o_t) \end{pmatrix} \tag{21}$$

where

$$u_t(m_t) = \begin{pmatrix} u_t(m_t,m_{t-1}) \\ u_t(m_t,o_{t-1}) \end{pmatrix} \tag{22}$$

J_t is a selection matrix of 0's and 1's and $\phi_t^{(o)}$ is a diagonal matrix. We include those entries of u_t in the state vector that correspond to missing entries in y_t and/or y_{t-1}. When both element of y_t and y_{t-1} are present we can compute the corresponding element in the c_t matrix, indeed the constant term in the equation (20). The updates for α_t and α_{t-1} are given by following equation:

$$\begin{pmatrix} \alpha_{t+1} \\ \alpha_t \end{pmatrix} = \begin{pmatrix} T & 0 \\ I & 0 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \end{pmatrix} \tag{23}$$

The transition equation of $u_{t+1}(o_{t+1}, m_t)$ and $u_{t+1}(m_{t+1}, m_t)$ are the selection $u_{t+1}(m_t)$ (re-ordered). The transition from $u_t(m_t)$ to $u_{t+1}(m_t)$ is the autoregressive update (19), see Jungbacker et al. (2009) for a more formal treatment of those equations.

We can still apply the computational device of Jungbacker and Koopman(2008) to the missing value state space formulation to get a significant computational gain. Considering the factor model (20) and (21) we carry out a collapsed computation only on the vector $y_t(o_t, o_{t-1})$.
Now define

$$A_t^L = C_t^{-1} Z_t^{+,'} \Sigma_{\varepsilon,t}^{-1} \qquad Z_t^+ = \left[Z(o_t,o_{t-1};.), -\phi_t^{(o)} Z(o_t,o_{t-1};.) \right] \qquad \Sigma_{\varepsilon,t} = \Sigma_{\varepsilon,t}(o_t,o_{t-1};o_t,o_{t-1}) \tag{24}$$

and C_t is chosen such that

$$C_t C_t^{'} = Z_t^{+,'} \Sigma_{\varepsilon,t}^{-1} Z_t^+ \tag{25}$$

for $t = 1,...,n$. The transformation A^L is applied only to the $y_t(o_t, o_{t-1})$ and does not require to consider the element of the state vector associated with the u_t since they do not affect $y_t(o_t, o_{t-1})$.
Now define the matrix

$$A_t = \begin{pmatrix} A_t^L \\ A_t^H \end{pmatrix} \tag{26}$$

where A_t^H is chosen as $A_t^H \Sigma_{\varepsilon,t} A_t^{H,'} = 0$. The state space model for the transformed observation is given by

$$\begin{pmatrix} A_t^L y_t(o_t,o_{t-1}) \\ A_t^H y_t(o_t,o_{t-1}) \end{pmatrix} = \begin{pmatrix} A_t^L c_t^o \\ A_t^H c_t^o \end{pmatrix} + \begin{pmatrix} C_t^{'} G & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \dot{\alpha}_t + A_t \varepsilon_t(o_t,o_{t-1}) \tag{27}$$

where $Var[A_t\varepsilon_t(o_t,o_{t-1})]$ is a block diagonal variance matrix with upper block given by $Var[A_t\varepsilon_t(o_t,o_{t-1})] = I$, and $\dot{\alpha}_t$ is given in formula (21). It follows that the second part $A^H y_t$ does not depend on the state vector and is not considered in the KFS step. From equation (27) we can split the likelihood as in formula (6).

We now derive the exact score for the state space model of formulation of (20) and (21). The model can be rewrited more compactly as:

$$\begin{aligned} y_t &= c_t + Z_t\alpha_t + \varepsilon_t \qquad \varepsilon_t \sim N(0,H_t) \\ \alpha_{t+1} &= d_t + T_t\alpha_t + R_t\eta_t \qquad \eta_t \sim N(0,Q_t) \end{aligned} \tag{28}$$

where $\alpha_1 \sim N(a_{1|0},P_{1|0})$ and $t = 1,...,n$.

Taking the first derivatives of the likelihood respect to the system vectors and matrices c_t, d_t, Z_t, T_t, H_t and Q_t we obtain the following derivatives:

$$\begin{aligned} &\frac{\partial l(y)}{\partial d_t} = \widetilde{R}_t Q_t^{-1} \widetilde{R}_t (a_{t+1|n} - T_t a_{t|n} - d_t) \qquad \frac{\partial l(y)}{\partial T_t} = \widetilde{R}_t Q_t^{-1} \widetilde{R}_t (M_{T_t} + d_t a_{t|n} - T_t M_{Z_t}) \\ &\frac{\partial l(y)}{\partial c_t} = \widetilde{H}_t^{-1}(y_t - Z_t a_{t|n} - c_t) \qquad \frac{\partial l(y)}{\partial Z_t} = \widetilde{H}_t^{-1}((y_t - c_t)a_{t|n}^{'} - Z_t M_{Z_t}) \\ &\frac{\partial l(y)}{\partial Q_t} = Q_t^{-1} M_{Q_t} Q_t^{-1} - \frac{1}{2} diag\{Q_t^{-1} M_{Q_t} Q_t^{-1}\} \\ &\frac{\partial l(y)}{\partial H_t} = \widetilde{H}_t^{-1} M_{H_t} \widetilde{H}_t^{-1} - \frac{1}{2} diag\{\widetilde{H}_t^{-1} M_{H_t} \widetilde{H}_t^{-1}\} \end{aligned} \tag{29}$$

with:

$$\begin{aligned} &M_{Q_t} = E(\eta_t\eta_t^{'}|y_1,...,y_n) - Q_t \qquad M_{T_t} = a_{t+1|n}a_{t|n}^{'} + P_{t+1,t|n} \\ &M_{H_t} = (y_t - c_t - Z_t a_{t|n})(y_t - c_t - Z_t a_{t|n})^{'} + Z_t P_{t|n} Z_t^{'} - \widetilde{H}_t \qquad M_{Z_t} = a_{t|n}a_{t|n}^{'} + P_{t|n} \end{aligned} \tag{30}$$

$\widetilde{R}$ is a selection matrix and the matrix $\widetilde{H}_t$ is constructed using the $y_t(o_t,m_{t-1})$ as index due to singularity of H_t. Finally the matrix M_{Q_t} can be evaluated by $\eta_t = \hat{R}(\alpha_{t+1} - T_t\alpha_t - d_t)$ and by $P_{t+1,t|n}$ using formula (11) and $P_{t|n}$.

The system vectors and matrices further depends on our parameters of interest so applying the chain rule, see Magnus and Neudecker(2007), we have:

$$\begin{aligned} &\frac{\partial vec(l(y))}{\partial vec(T)} = \left(\frac{\partial vec(l(y))}{\partial vec(T_t)}\right)^{'} \left(\frac{\partial vec(T_t)}{\partial vec(T)}\right)^{'} \qquad \frac{\partial vec(l(y))}{\partial vec\left(\phi_t^{(o)}\right)} = \left(\frac{\partial vec(l(y))}{\partial vec(c_t)}\right)^{'} \left(\frac{\partial vec(c_t)}{\partial vec\left(\phi_t^{(o)}\right)}\right)^{'} + \left(\frac{\partial vec(l(y))}{\partial vec(Z_t)}\right)^{'} \left(\frac{\partial vec(Z_t)}{\partial vec\left(\phi_t^{(o)}\right)}\right)^{'} \\ &\frac{\partial vec(l(y))}{\partial vec(Z_t)} = \left(\frac{\partial vec(l(y))}{\partial vec(Z_t)}\right)^{'} \left(\frac{\partial vec(Z_t)}{\partial vec(Z)}\right)^{'} \qquad \frac{\partial vec(l(y))}{\partial vec\left(\phi_t^{(*)}\right)} = \left(\frac{\partial vec(l(y))}{\partial vec(d_t)}\right)^{'} \left(\frac{\partial vec(d_t)}{\partial vec\left(\phi_t^{(*)}\right)}\right)^{'} + \left(\frac{\partial vec(l(y))}{\partial vec(T_t)}\right)^{'} \left(\frac{\partial vec(T_t)}{\partial vec\left(\phi_t^{(*)}\right)}\right)^{'} \\ &\frac{\partial vec(l(y))}{\partial vec\left(\phi_t^{(1)}\right)} = \left(\frac{\partial vec(l(y))}{\partial vec(T_t)}\right)^{'} \left(\frac{\partial vec(T_t)}{\partial vec\left(\phi_t^{(1)}\right)}\right)^{'} \end{aligned} \tag{31}$$

References

[1] Bai, J. and NG, S. (2008) *Recent developments in large dimensional factor analysis*, Working Paper, Mimeo.

[2] Banbura, M. , Modugno M. (2008) *Maximum likelihood estimation of large factor model on data sets with arbitrary pattern of missing data*, Working Paper.

[3] Backus, D.K.; Kehoe, P. J. and Kydland, F.E. (1995) *International Business Cycles: Theory and Evidence* in Thomas F. Cooley, ed., *Frontiers of business cycle research* Princeton, NJ: Princeton University Press, 1995, pp. 33157.

[4] Backus, D.K. and Crucini, M.J.(2000)*Oil Prices and the Terms of Trade*, Journal of International Economics, February, 50 pp 203-31.

[5] Baxter, M. (1995) *International Trade and Business Cycles* in Gene M. Grossman and Kenneth Rogoff, eds., Handbook of international economics, volume 3. Amsterdam: North-Holland, pp. 1801-64.

[6] Doornik, J.A. (2007) *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

[7] Doz, C., Giannone, D., and Reichlin, L. (2007). *A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models*, CEPR Discussion Paper No. 5724

[8] Eichengreen, B. (2001). *The EMS Crisis in Retrospect*, CEPR Discussion Papers 2704.

[9] Engle, R.F., Watson M.W. (1981) *A one-factor multivariate time series model of metropolitan wage rates*, J. American Statistical Association 76, pp. 774-781.

[10] Jarque, C.M., Bera, A.K. *Test for the Normality of the observations and regression residuals*, International Statistical Review 55:pp. 163-172.

[11] Jungbacker B. and Koopman S.J. (2008). *Likelihood-based analysis for dynamic factor models*, University of VU, Working Paper

[12] Jungbacker, B., Koopman S.J an van der Wel, M.(2009). *Dynamic Factor Analysis in the Presence of Missing Data*, Tinbergen Institute Discussion Paper 010-4

[13] Koopman S.J. and Shephard N. (1992). *Exact Score for Time Series Models in State Space Form*, Biometrika, Vol 79 No 4, Dec 1992

[14] Kose, M. Ayhan (2002) *Explaining Business Cycles in Small Open Economies: How Much Do World Prices Matter?* Journal of International Economics, 56(2), pp. 299-327.

[15] Kose, M.A, Otrok, C. and Whiteman C.H (2004) *International Business Cycle: World, Region, and Country-Specific Factors*, The American Economic Review.

[16] Ljung, G.M. and Box, G.E.P (1978). *On a measure of lack of fit in time series models*, Biometrika, 66, pp. 67-72

[17] Gargan, E. A. (December 11, 1997). *Last Laugh for the Philippines; Onetime Joke Economy Avoids Much of Asia's Turmoil*. New York Times.

[18] Magnus, J.R., Neudecker, H. (2007) *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Third Edition, John Wiley and Sons.

[19] Nocedal, J., Wright, S.J.(1999) *Numerical Optimization*, New York:Springer Verlag.

[20] Nordhaus, D. W. (2002) *The Mildest Recession: Output, Profits, and Stock Prices as the U.S. Emerges from the 2001 Recession*, NBER Working Paper n 8938.

[21] Gentry, J.W., Jun, S., Chun, S., Kang H. and Ko, G. (2006) *Korea: Two Countries, Sharp Contrasts, but a Common Heritage* in Pecotich, A., Shultz, C.J. Handbook of Markets and Economies, M.E. Sharpe.

[22] Proietti T. (2008). *Estimation of Common Factors under Cross-Sectional and Temporal Aggregation Constraints: Nowcasting Monthly GDP and its Main Components*, MPRA Paper 6860, University Library of Munich, Germany

[23] Reis, R., Watson M.W. (2007) *Relative goods prices and pure inflation*, NBER Working Paper, No. 13615

[24] Stock, J.H and Watson, M. H (2002) *Macroeconomic Forecasting Using Diffusion Index*, Journal of Business and Economic Statistics, Vol 20, N.2

[25] Stock, J.H and Watson, M. H (2005) *Understanding Changes in International Business Cycle Dynamics*, Journal of European Economic Association, September 2005.

[26] Stock, J.H., and Watson, m.H. (2006) *Forecasting with many predictors*, in Elliott, G., Granger, C.W.J. and Timmermann, A. (eds.) Handbook of Economic Forecasting, Volume 1, Elsevier.

[27] Weeks, J.(2000) *Latin America and the 'high performing Asian economies': growth and debt*, Journal of International Development, 625-654.