UNIVERSITY OF SOUTHAMPTON

Essays in Applied Factor Analysis with Structural Breaks

by

Ronaldo Nazaré

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
FACULTY OF SOCIAL AND HUMAN SCIENCES
DEPARTMENT OF ECONOMICS

October 2013

Declaration of Authorship

I, Ronaldo Nazaré, declare that this thesis titled, 'Essays in Applied Factor Analysis with Structural Breaks' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____________________________________

Date: ____________________________________

"For instance, the correlation between the length of the skull and the weight of the brain must, obviously, be very far from linear. But linearity is at once restored (supposing all the skulls to belong to one type) if we change the second term from the brain's weight to the cube root of the weight."

Charles Edward Spearman, 1907.

"A brief, easily applicable test for determining the number of factors to extract in factor analytic experiments has long been in demand."

Raymond Bernard Cattell, 1966.

"If you can't solve a problem, then there is an easier problem you can solve: find it."

George Pólya, 1945.

UNIVERSITY OF SOUTHAMPTON

Abstract

FACULTY OF SOCIAL AND HUMAN SCIENCES
DEPARTMENT OF ECONOMICS

Doctor of Philosophy

by Ronaldo Nazaré

This thesis offers an analysis on factor models and related topics such as forecasting, the choice of the number of factors, model selection and structural breaks.

Acknowledgements

In the moments of difficulties and unstable hope, my family was always there to motivate me with love. Without them, simply nothing would be possible and my parents are the pillars of this. My wife followed me in this dream, whilst we missed everyone in Brazil. For all them, my eternal gratitude.

I am enormously indebted and grateful to Jean-Yves Pitarakis, my PhD advisor. He guided me with mastery throughout this journey. I believe that my project was really challenging, and for this reason I needed his expertise in a great proportion. I tried my best to enjoy the opportunity to have such a high level professional as my advisor. He patiently offered me innumerable discussions and suggestions which allowed me to develop myself as an economist. In this regard I also thanks the academic staff of the Department of Economics of the University of Southampton for the excellent working environment. Particularly, I would like to thanks Hétor Calvo Pardo with whom I had a chance to have very good conversations. Furthermore, I would like to thank the administrative staff of the University.

With respect to the evaluation of the present work, I would like to thank the examiners of this thesis, Jesús Gonzalo (external examiner) and José Olmo (internal examiner), for their insightful questions during the *viva voce*, in addition to pointing out for necessary corrections.

My colleagues at University of Southampton were very important to smooth the long days of work. In the economics department I wish to thank my friends Reza Boostani, Jurica Brajkovic, Dimitrios Gkountanis, Derya Tas and Juan Correa. I also wish to thank the great friendship of Brian Moore, Juan José Fraire Zamora, Javier Ortiz Santiago, Kate Lyle, Jaimie Ellis, Katie Bruce. Moreover, I am grateful to have met and learned from the art of Mestre Poncianinho, Baris Yazar and Cordão de Ouro - London.

I am very lucky to have received a scholarship from the Banco de Desenvolvimento de Minas Gerais, for which I am extremely thankful. I am also thankful for the financial support of the Graduate Teaching Studentship from the School of Social Sciences and Economics of the University of Southampton. The teaching opportunities were of great value in my learning process. Parts of my research received a financial award from the Santander Abbey Internationalisation Fund for a research collaboration between the University of Southampton and Fundação João Pinheiro, Brazil. This award allowed me to take extra courses and buy relevant books.

I have presented parts of the first and second chapters of this thesis at the research seminar of Department of Statistics, UFMG, Brazil, and at the Annual Doctoral Conference of the Association for the Development of Research in Economics and Statistics (ADRES) Groupement de Recherche en Economie Quantitative d'Aix-Marseille, France. Both events were in 2012, and I acknowledge that I received useful comments.

Last but not least, in no particular order, I also would like to thank very much the immeasurable support of Sérgio Luís Guerra Xavier, Afonso Henriques Borges Ferreira, Fábio Augusto Reis Gomes, Viviane Luporini, Marcelle Chauvet and Prasanta Pattanaik.

Contents

List of Figures

List of Tables

To my parents and sister,

to my wife and to our baby.

This is for you!

Chapter 1

Introduction

Factor models have a long history in the field of Statistics, with the seminal paper of Spearman (1904) on measuring intelligence, commonly referred to as one of the precursors of this literature. Since this early work and the extensive research that has followed on, factor models have become an important research area for both theorists and practitioners. The main usefulness of factor models is their ability to reduce the dimensionality of a large data set into a smaller set of common drivers or factors. This is achieved through the decomposition of a large panel data set into a small linear combination of uncorrelated common factors (latent variables) and an idiosyncratic component.

Factor models have also attracted considerable interest from Economists. Early influential papers go back to Sharpe (1964) who proposed a common risk factor for modelling asset returns in his Capital Asset Pricing Model and Ross (1976)'s Arbitrage Pricing Theory with multiple factors used to explain the comovement of asset returns. Naturally, within this early work, the techniques used for modelling and extracting factors operated under strong assumptions that ruled out features commonly encountered in economic time series, such as heteroskedasticity, cross-sectional correlation, structural breaks or the handling of settings in which both the time series and individual dimensions could be very large. Equally importantly, the issue of the determination of the number of common factors in a formal and statistically rigorous way remained an open question until the recent work of Bai and Ng (2002) and others.

The goal of this thesis is to make a contribution to the literature on factor models by offering practitioners some new techniques designed to improve the modelling of a large data set as a factor model with the potential presence of structural breaks. The thesis also contains an important applied component that aims to better understand the usefulness of factor models for forecasting purposes in the context of the Brazilian unemployment rate. The various toolkits we propose in this thesis rely on the approximate factor model setting recently explored in the work of Bai (2003) and Bai and Ng (2002).

In matrix notation a factor model can be viewed such as $X_t = F_t\Lambda' + e_t$, where X_t is a $T \times N$ matrix of observations, Λ is a $k \times N$ matrix of factor loadings, F is a $T \times k$ matrix of common factors, with k denoting the number of factors, and, e_t is a $T \times N$ matrix idiosyncratic components.

The estimation procedure used in this thesis lies on asymptotic principal components. The problem requires the minimization of $V(k) = \min_{\Lambda F^k} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \Lambda_i^{k'}F_t^k\right)^2$ which does not lead to a unique solution but indeed to a unique sum of squared residuals. We cannot estimate Λ and F simultaneously because they are not uniquely identifiable. To facilitate it computationally, if $T < N$, we calculate the covariance of the dataset as $\Sigma = XX'$, to have it as a $(T \times T)$ matrix and let the normalization $\widehat{F}\widehat{F}'/T = I$ an identity matrix. We achieve the estimates as follows: the common factors $\widehat{F} = V$ (we choose the k eigenvectors of Σ to be the common factors, corresponding to the k largest eigenvalues), and the factor loadings are given by $\widehat{\Lambda} = (V'X)/N$; the common components $\widehat{\chi} = \widehat{F}\Lambda' = (VV'X)/T$ and the idiosyncratic errors, $\widehat{e} = X - \widehat{\chi}$.

If instead we have $T > N$, we could normalize the optimization problem differently to facilitate computationally. We calculate $\Sigma = X'X$ to have it as a $(N \times N)$ matrix. The normalization now makes $\overline{\Lambda\Lambda}'/N = I$, an identity matrix. Then the estimates of the common factors are $\overline{F} = (XV)/N$ and the factor loadings $\overline{\Lambda} = V$ (we choose k eigenvectors of Σ to be the factor loadings, corresponding to the k largest eigenvalues); common components $\overline{\chi} = \overline{F\Lambda}' = (XVV')/N$; and the idiosyncratic errors $\overline{e} = X - \overline{\chi}$.

The choice of the method does not matter because factors estimated by either eigendecomposition schemes have equivalent column spaces. They are scaled versions of each other. The sign of the factors are not identifiable which allows one to sign-adjust the factors accordingly (see *Theorem 1* of Stock and Watson (2002a)). The reason for this is that the estimated factors and factor loadings using either schemes will not be numerically the same. However, the common components are identical: $\widehat{\chi} \equiv \overline{\chi}$.

Chapter 2 proposes a new toolkit for uncovering the presence of breaks in the structure of a factor model. Our proposed method relies on the observation that an omitted break in factor loadings translates into an overestimated number of factors when the latter is estimated via Bai and Ngs model selection based approach (see Breitung and Eickmeier (2011)). We therefore introduce a recursively implemented model selection based approach that is designed to take advantage of this feature. Our recursively estimated number of factors displays a clear jump at the time when a structural breaks occurs and this allows us to propose a graphical approach for uncovering breaks in loadings.

Chapter 3 introduces a novel model selection based estimator of the number of factors that is robust to the presence or absence of structural breaks in the underlying factor model. More specifically, our estimator remains consistent even if the underlying model contains breaks in its loadings. This subsequently allows us to introduce a decision based rule for detecting whether a break is truly present.

Our results in Chapters 2-3 are closely related to the recent literature initiated by Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012). A common feature of those papers is the introduction of techniques designed to detect the presence of a structural breaks in the loading matrix of a factor model. Breitung and Eickmeier (2011) developed a SupWald type test, assuming cross-sectional independence and operating on a series by series basis while Chen et al. (2012) have shown that a structural break in the factor loadings of a factor model can be viewed as a structural break in the coefficients of a regression model linking one of the estimated factors to the remaining ones. The former can then be detected via a conventional SupWald type test statistic. Finally, Han and Inoue (2012)'s approach relies on a Wald type test of stability on the covariance matrix of the estimated factors.

The motivation of the analysis of structural breaks in factor models can be thought in occasions when over time dimension a factor loses its importance or a new factor kicks in. In multifactor models describing asset returns for instance the importance of some factors such as size may vanish over time while other risk factors may emerge (e.g. liquidity factor, macroeconomic risk factor). Furthermore, this motivation can be illustrated with Chen et al. (2012) forecasting simulations. They show the impact of large structural breaks in forecast exercises when estimated factors are used as predictors. In these simulations the mean square errors (MSEs) increase as structural breaks also increase, in such a way that the forecasts neglecting structural breaks in the factor models generate inferior forecasts (larger MSEs).

Chapter 4 of this thesis is an applied exercise exploring the use of factor analysis for forecasting purposes. We explore a rich dataset that includes labor market related historical time series across all major Brazilian metropolitan areas as well as aggregate monthly macroeconomic and monetary related variables in order to forecast the Brazilian unemployment rate. The number of factors is estimated via Bai and Ngs model selection criteria and forecast accuracy comparisons are made across a large number of estimated models including methods that involve forecast combinations and forecast encompassing. Our findings suggest that a simple forecast combination approach using a diffusion index model is able to provide very accurate predictions of Brazilian unemployment rates. Chapter 5 concludes.

Chapter 2

Detecting Structural Breaks in Factor Models: A Recursive Model Selection Based Approach

2.1 Introduction

With the growing availability of large datasets, factor models have attracted considerable attention from both theorists and practitioners due to their ability to reduce dimensionality in a convenient and optimal way. A factor model aims at reduce the dimensionality of a large dataset by formulating the variables involved as linear combinations of a small number of common factors. In economics, early work in applied factor analysis goes back to Sharpe (1964) who formulated the Capital Asset Pricing Model using the market return as the common risk factor linking a large number of individual returns and Ross (1976) who modeled the Arbitrage Pricing Theory along the idea of multiple common factors explaining asset returns. More recently, factor models have regained popularity within the forecasting literature due to the advantages of using of composite predictors (or common factors) summarising information from across hundreds of series (see Stock and Watson (2002a,b)). Similarly, factor augmented VAR models, commonly referred to as FAVARs, have allowed practitioners to use very large datasets in a convenient and feasible way thanks to the dimensionality reduction features of factor models.

In most recent applications, factor models have typically been constructed using a large number of economic and financial time series spanning very long periods. When using common factors extracted from hundreds of macroeconomics time series as predictors of quantities such as inflation it is common practice to use quarterly or monthly series from as early as the post war period. Similarly, it is not uncommon to see the estimation of factor models in finance (e.g. the Fama and French (1993) three factor model) being performed using data from the 60s or 70s onwards. The use of such long data spans naturally raises the issue of model stability and may raise concerns about the suitability of imposing time invariance of factor loadings in such settings (see Banerjee and Marcellino (2008)). The structure of a factor model may also change over time due to a factor losing its importance or a new factor kicking in. In multifactor models describing asset returns for instance the importance of some factors such as size may vanish over time while other risk factors may emerge (e.g. liquidity factor, macroeconomic risk factor). The goal of this chapter is to develop a series of methods designed to uncover the presence of such instabilities within otherwise standard linear factor models. Our goal is solely that of uncovering the presence of a structural break rather than being also able to distinguish between different sources of breaks which we leave for further research (e.g. breaks in loadings versus breaks in the covariance matrix of factors through new factors kicking in for instance).

Our work is inspired by the model selection based methods developed in Bai and Ng (2002) and which have become one of the most commonly used approaches for estimating the number of common factors linking a large number of time series. The main novelty of this chapter however is based on a recursive implementation of the Bai and Ng type model selection criteria as opposed to their more commonly used full sample implementation. We estimate the common factors and their number using asymptotic principal component methods and model selection principles implemented recursively. The recursive profile of the model selection criteria evaluated à la Bai and Ng allows us to obtain a sequence of optimal number of factors in each recursion/time period. This profile of recursively estimated number of factors is the key ingredient in our decision based rule for deciding between stability and break scenarios and relies on the observation that a jump in the estimated number of factors is associated with a break in either the loadings or the number of factors driving the system. There is in fact a one to one relationship between the presence of a break in the structure of a factor model and the jumps that characterize our recursive estimator of the number of common factors. We use this information to propose a simple to use and easily implementable graphical approach for uncovering structural instabilities in factor models. This then allows us to construct a modified estimator of the number of factors that corrects for the distortions induced by the presence of structural breaks. At this stage it is also interesting to recall that early approaches used to determine the number of common factors were also based on graphical methods and Cattell's Scree Test in particular (Cattell (1966)) which plotted the eigenvalues of a particular sample covariance matrix against the numbers of factors. More recently, Forni and Reichlin (1998) have also relied on a graphical approach for estimating the number of common factors within a sectoral dataset.

The issue of structural breaks in factor analysis is a very recent concern that was initiated by a small number of authors working on factor based modeling methodologies. The issue was first raised in the work of Stock and Watson (2008) who argued that breaks in loadings are of little consequence on forecast accuracy provided that they are small. From a methodological perspective however, the first paper that explored formal methods of uncovering structural breaks in factor models was Breitung and Eickmeier (2011) who introduced a Sup Wald type test for detecting a structural break in factor loadings. The test is implemented on a series by series basis assuming cross-sectional independence. Breitung and Eickmeier's work was closely followed by Chen et al. (2012) who developed a more comprehensive and more widely applicable methodology by showing that the presence of a break in the loadings of a factor model translates into a break in the regression coefficients of an auxiliary regression linking one factor with the remaining ones. This break can subsequently be detected through a standard SupWald type test procedure. Similarly, Han and Inoue (2012) have also proposed a new test for detecting breaks in the loadings of a factor model that relies on constructing a Wald type test of stability for a particular sample covariance matrix obtained using estimated factors.

At this stage it is very important to emphasize that all of the above mentioned methods for uncovering breaks rely on the availability of an estimator of the number of common factors, typically taken to be Bai and Ng's model selection based estimator. The latter is typically estimated using full sample information and ignores the impact of underlying instabilities. In this chapter we argue that our proposed methodology and the resulting estimator of the number of common factors can also be used as a valuable input in the implementation of the methods developed in Chen et al. (2012) or Han and Inoue (2012).

The plan of our chapter is as follows. Section 2.2 introduces the factor model of interest and outlines the type of model instabilities we will be considering. Section 2.3 introduces our recursive model selection based approach for estimating the number of common factors and highlights its properties. Section 2.4 is an extensive simulation study exploring the properties of our proposed method and its implementation across a variety of scenarios while Section 2.5 compares the performance of our approach with Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012). Section 2.6 proposes an empirical application using the existing dataset from Section Stock and Watson (2005) and Section 2.7 concludes.

2.2 Modeling Changes in the Factor Structure

We consider the following linear factor model

$$\underset{(N\times T)}{X} = \underset{(N\times r)}{\Lambda}\underset{(r\times T)}{F'} + \underset{(N\times T)}{e} \tag{2.1}$$

where X is the $N \times T$ matrix of the observed dataset, F is a $T \times r$ matrix of common factors, Λ is the $N \times r$ matrix of factor loadings and e is the $N \times T$ matrix of idiosyncratic errors. The above specification views the series in X as being linearly linked to r common factors and an idiosyncratic error term with $\Lambda F'$ denoting the common component. It is in this sense that factor models are able to reduce the dimensionality of a large system. For later use it is also convenient to view each component of X as stacking the elements $X_{it} = \Lambda_i' F_t + e_{it}$ so that $F = (F_1, \ldots, F_T)'$ and $\Lambda = (\Lambda_1, \ldots, \Lambda_N)'$. Note that both the loading matrices Λ_i and the factors F_t are unobserved by the researcher.

Structural change in a factor model specification such as (2.1) may take multiple forms. As in Breitung and Eickmeier (2011) we could think of the factor loadings changing over time as in

$$X_{it} = \begin{cases} \Lambda_{i,1} F_t' + \epsilon_{it} \; t = 1, ..., \ell \\ \Lambda_{i,2} F_t' + \epsilon_{it} \; t = (\ell+1), ..., T. \end{cases} \tag{2.2}$$

or more compactly, using indicator functions, as

$$X_{it} = \Lambda_{i,1} \; F_t' \; I(t \leq \ell) + \Lambda_{i,2} \; F_t' \; I(t > \ell) + \epsilon_{it}. \tag{2.3}$$

Here the factor model in (2.1) is subject to a structural break in its loadings. The break occurs at some unknown time ℓ after which the loading matrix switches from $\Lambda_{i,1}$ to $\Lambda_{i,2}$. Following Chen et al. (2012) it is also useful to reformulate (2.2) or (2.3) as

$$X_{it} = \Lambda_{i,1} \; F_t' + B_i \; G_t' + \epsilon_{it} \tag{2.4}$$

with $B_i = \Lambda_{i,2} - \Lambda_{i,1}$ referring to the size of the break and $G_t \equiv F_t I(t > \ell)$. The above notation also makes clear the fact that a factor model with a break in its loadings and say r factors can be rewritten as a factor model with constant loadings but a greater number of factors (see Breitung and Eickmeier (2011)).

Our goal in this chapter is to propose a method of detecting the presence of a break as in (2.2) or (2.3) by taking advantage of the fact that a model with a break in its loadings can be equivalently written as a model with time invariant loadings associated with a larger number of factors (e.g. $r_0 + 1$ versus r_0) and hence the most commonly used Bai and Ng type model selection criteria for estimating the number of factors will lead to an overestimation of the true number of factors. As mentioned above this point was made in the earlier work of Breitung and Eickmeier (2011) and formally proved in Chen et al. (2012, Proposition 2).

The key idea proposed in this chapter is to introduce a recursive approach for the determination of the number of factors. This allows us to determine whether a break has occurred by taking advantage of the fact that following the occurrence of break the Bai and Ng type model selection criteria will tend to overestimate the true number of factors. When estimated in a recursive manner therefore, we should expect the number of factors to exhibit a jump following the occurrence of the break. Although our methodology is unable to disentangle breaks in loadings from breaks in the deeper factor structure (e.g. in the moments of the factors themselves) it is able to accurately detect the presence of a break under both scenarios.

2.3 Recursive Estimation of the Number of Factors: Recursive Implementation of Bai and Ng's Model Selection Criteria

Bai and Ng (2002) have introduced a wide range of information theoretic criteria designed to estimate the number of factor within linear factor models. Letting k index the number of factors assumed to be between zero and some given upper bound k_{max}, the generic form of Bai and Ng's criteria for each value of k is given by

$$IC(k) = \ln V(k, \widehat{F}^k) + kg(N,T) \tag{2.5}$$

where $V(k, \widehat{F}^k) = \sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \widehat{\Lambda}_i' \widehat{F}_t^k)^2/NT$ and $g(N,T)$ is a deterministic penalty term (e.g. $g(N,T) = \ln C_{NT}^2/C_{NT}^2$ for $C_{NT}^2 = \min(N,T)$). For each possible k, the estimated factors $\widehat{F}_t$ and their respective factor loadings $\widehat{\Lambda}_i$, are estimated via asymptotic principal components. The estimated number of factors is given as the solution to the following optimization problem

$$\widehat{k} = \arg \min_{0 \leq k \leq k_{max}} IC(k). \tag{2.6}$$

Here we propose to assess the presence or absence of a break within (2.1) by implementing the above model selection procedure in a recursive manner. We let

$$V_\ell\left(k_\ell, \widehat{F}_t^{k_\ell}\right) = \frac{1}{N\ell}\sum_{i=1}^{N}\sum_{t=1}^{\ell}(X_{it} - \widehat{\Lambda}_i' \widehat{F}_t^{k_\ell})^2 \tag{2.7}$$

denote the recursively estimated error variance for each possible value of ℓ starting with a given lower bound, say $\underline{\ell}$ so as to ensure estimability and with $\ell = \underline{\ell}, \underline{\ell}+1, \ldots, T$. In what follows we set $\underline{\ell} = [T\underline{\pi}]$ with $\underline{\pi}$ denoting the minimum fraction of observations we wish to allow. Suppose for instance that $T = 100$ and $\underline{\pi} = 0.3$. The above recursive sum of squared errors will generate 71 possible values for V_ℓ under $(t = 1, \ldots, 30)$, $(t = 1, \ldots, 30, 31)$, ..., $(t = 1, \ldots, 30, 31, \ldots, T)$. For each possible value of $\ell = \underline{\ell}, \ldots, T$ our approach requires estimating the factors via asymptotic principal components and computing all possible value of the estimated error variance function in (2.7). Note that we let $k_\ell \in [0, k_{max}]$ for each possible value of ℓ. This then allows us to construct the following recursive model selection criteria à la Bai and Ng

$$IC_{p1,\ell}\left(k_\ell\right) = \ln(V_\ell) + k_\ell \left(\frac{N+\ell}{N\ell}\right) \ln\left(\frac{N\ell}{N+\ell}\right) \tag{2.8}$$

$$IC_{p2,\ell}\left(k_\ell\right) = \ln(V_\ell) + k_\ell \left(\frac{N+\ell}{N\ell}\right) \ln C_{N\ell}^2 \tag{2.9}$$

$$IC_{p3,\ell}\left(k_\ell\right) = \ln(V_\ell) + k_\ell \left(\frac{\ln C_{N\ell}^2}{C_{N\ell}^2}\right) \tag{2.10}$$

with $C_{N\ell}^2 = \min\left(N, \ell\right)$. The above criteria are computed for each possible value of ℓ and we let $\widehat{k}_\ell$ denote the following optimal choice of the number of factors obtained à la Bai and Ng

$$\widehat{k}_\ell = \arg \min_{0 \le k_\ell \le k_{max}} IC_\ell(k_\ell) \tag{2.11}$$

with IC_ℓ referring to one of the above three criteria. In words, for each possible value of ℓ we compute all possible values of the model selection criteria for $k_\ell \in [0, k_{max}]$ and subsequently estimate $\widehat{k}_\ell$ as the value of k_ℓ that leads to the smallest model selection criterion.

Before proceeding further it is important to highlight some basic feasibility requirements for the validity of our procedure. Since our recursions are assumed to start at some known lower bound, say $\underline{\ell} = [T\underline{\pi}]$, it is naturally understood that the true break fraction π_0, if present, must be such that $\pi_0 > \underline{\pi}$.

Before focusing on the properties of $\widehat{k}_\ell$ under the structural break model it is important to also highlight the type of breaks that we expect our procedure to be able to uncover. As discussed in Stock and Watson (2008) for instance, only big jumps in the loading coefficients can be expected to be detectable. This concept has been formalized for the first time in Chen et al. (2012).

The following Proposition provides the key rationale for the use of our proposed recursive approach. It shows that in large samples our recursively estimated number of factors converges to the true number of factors, say r_0, prior to the occurrence of a break but after the occurrence of a break the convergence is towards a quantity that is greater than r_0 if and only if large breaks take place. Borrowing the framework and definitions of structural breaks in Chen et al. (2012, p. 5) we operate under the assumption that the loadings may be subject to large breaks or small breaks. More specifically, if we partition B as (4.2) into two components, say $B = [B_{large}\ B_{small}]$ with B_{large} being an $N \times d_1$ matrix associated with large breaks and B_{small} being an $N \times d_2$ associated with small breaks and such that $d_1 + d_2 = r$ we have

Proposition 2.1: *(i) Assuming that the loading coefficients in* (2.3) *are only subject to small breaks as defined in Assumption 1 of Chen et al. (2012) and operating under the same assumpions as in Chen et al. (2012), we have* $P[\widehat{k}_\ell = r_0] \to 1$ *for* $\pi < \pi_0$ *ans* $P[\widehat{k}_\ell = r_0] \to 1$ *for* $\pi \geq \pi_0$ *as* $(N, T) \to \infty$*. (ii) Assuming that the loading coefficients in* (2.3) *are subject to both small and large breaks as defined in in Assumption 1 of Chen et al. (2012) we have* $P[\widehat{k}_\ell = r_0] \to 1$ *for* $\pi < \pi_0$ *and* $P[\widehat{k}_\ell = r_0 + d_1] \to 1$ *for* $\pi \geq \pi_0$ *as* $(N, T) \to \infty$*.*

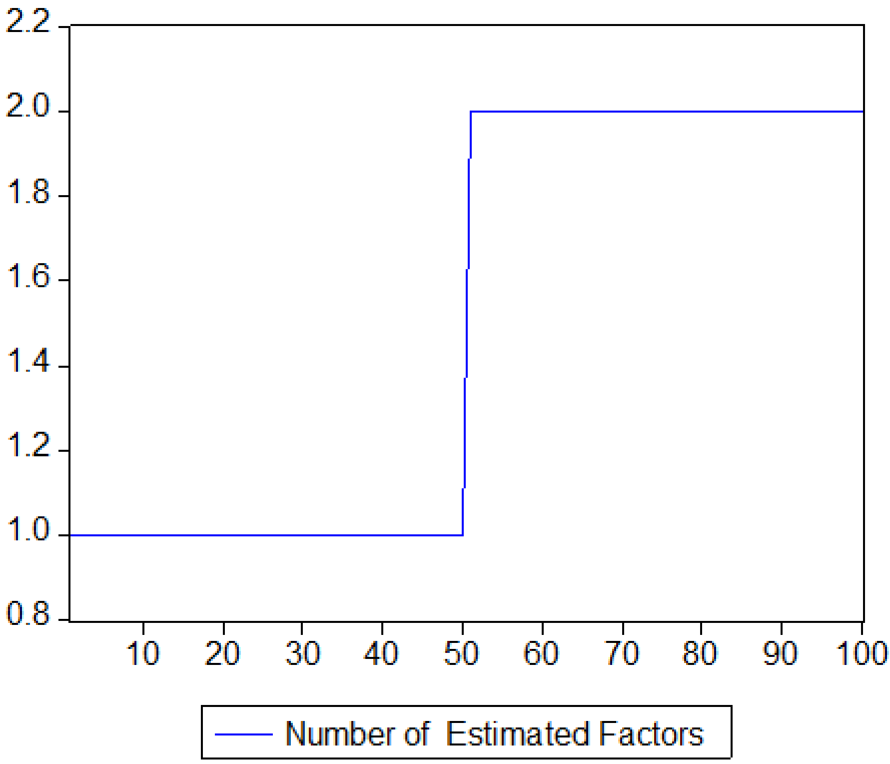
The above proposition whose proof follows directly from Proposition 2 in Chen et al. (2012) highlights the fact that our recursive approach will lead to an estimated number of factors that is inflated by d_1 after the occurrence of the break at time ℓ_0. This is also the main rationale for our decision rule based approach to detecting the occurrence of such breaks. Our approach relies on comparing the sequence of recursively estimated number of factors $\widehat{k}_\ell$ with the full sample counterpart $\widehat{k}(\equiv \widehat{k}_T)$ defined in (2.6). Plotting the sequence of the recursively estimated number of factors $\widehat{k}_\ell$ against ℓ yields a simple graphical approach to uncovering a break.

Decision Rule (Graphical Approach): *If the* $\widehat{k}_\ell$ *sequence plotted against* $\ell = \underline{\ell}, \ldots, T$ *exhibits at least one jump occurring between* $\underline{\ell}$ *and* T*, select model* (2.2)*. Select model* (2.1) *otherwise.*

In Figure 2.1 below we present a simple simulated example to illustrate the usefulness of the above decision rule.

For this single realization the data is of length N=T=100 assuming a simple one-factor model ($r_0 = 1$) across the entire sample length and a change in the number of factors occurring at time $\ell_0 = 50$. This artificially generated example illustrates very clearly the jump in $\hat{\ell}$ taking place at time $t = \ell_0 = 50$. The estimated number of factors is such that $\hat{k}_\ell = 1$ for $t < 50$ and switches to $\hat{k}_\ell = 2$ for $t \geq 50$.

FIGURE 2.1: DGP: $r_0 = 1$ with a large break at $t = \ell_0 = 50$



A modified estimator of the number of factors Perhaps more interestingly we can also use our decision rule to propose a modified estimator of the number of factors that is adjusted for the presence of breaks (jumps). More formally if ϕ denotes the number of jumps that occur in the $\hat{k}_\ell$ series plotted against ℓ, we can define our modified estimator of the true number of factors as

$$\widehat{k}^* = \widehat{k}_T - \phi. \tag{2.12}$$

In the context of Figure 2.1 for instance we have $\widehat{k}_T = 2$ and $\phi = 1$ so that $\widehat{k}^* = 1$ is our modified estimator of the true number of factors. This clearly corresponds to the underlying DGP which had $r_0 = 1$ and a single break in its loadings. When $\phi = 0$ our modified estimator is clearly identical to Bai and Ng's full sample based estimator.

Naturally our proposed graphical approach should be viewed as a preliminary diagnostic tool since it does not provide a formal way of conducting inferences about r_0 due to the unknown distributional properties of the discrete sequence $\widehat{k}_\ell$. This shortcoming is in a way similar to the shortcomings that characterize all estimators of discrete quantities such as $\widehat{k}$. The latter are typically taken as equal to their true counterparts and are not tested for. A more formal approach to assessing the presence of a jump in our $\widehat{k}_\ell$

sequence could be to implement a standard mean shift test using a SupWald type test statistic.

2.4 Monte Carlo Simulations for the Recursive Model Selection based Information Criteria

We initially consider a model characterized by the presence of a single break in its loadings occurring at time ℓ_0 and a fixed number of factors r_0 across time (DGP1)

$$X_{it} = \begin{cases} \sum_{j=1}^{r_0} \Lambda_{i,1} F'_{tj} + \epsilon_{it} \text{ for } t = 1, ..., \ell_0 \\ \sum_{j=1}^{r_0} \Lambda_{i,2} F'_{tj} + \epsilon_{it} \text{ for } t = (\ell_0 + 1), ..., T \end{cases} \tag{2.13}$$

Our Monte-Carlo simulations rely on alternative parameterizations of our specification in (2.13) and are conducted using 1000 replications, taking all random variables as standard normally distributed. Regarding the sample size, cross-sectional dimension and location of the break point we set $\{T, N, \ell_0, \underline{\ell}\} = \{100, 100, 50, 30\}$ throughout. Alternative choices for the break point location have also been considered and led to outcomes very similar to the ones presented below. One may also wonder what would happen in different sample sizes. As we increased N and/or T our graphical approach was able to pick up the jump in a timelier manner (at or around ℓ_0) while with smaller sample sizes the jump occasionally took place at later periods. As expected, there were also instances where $\hat{k}_\ell$ was not characterised by a jump when the true DGP contained a break in its loadings.

Since our goal is to highlight the reliability of our approach based on observing a jump in $\hat{k}_\ell$ we proceed as follows. For each realization of our DGP in (13) we compute our recursive estimator of the number of factors $\widehat{k}_\ell$ across $\ell = 30, \ldots, 100$ i.e. $\widehat{k}_{30}$, $\widehat{k}_{31}$, $\widehat{k}_{33}$, $\ldots$, $\widehat{k}_{100}$. The latter are subsequently averaged across our 1000 replications as $\sum_{h=1}^{1000} \widehat{k}_{30}^{(h)}/1000$ etc. Those averages are subsequently plotted against ℓ and displayed in Figures 2.2-2.10 below.

In the context of model (13), in DGP1 we set $r_0 = 1$ and experiment with breaks of size zero (no break), $B = 0.3$, $B = 0.5$ (Figure 2.2), $B = 0.7$ (Figure 2.3), $B = 0.9$ (Figure 2.4) and $B = 1$ (Figure 2.5); Our results corroborate unequivocally the results presented in Table 1 of Breitung and Eickmeier (2011) who considered a similar experiment. Despite the fact that $r_0 = 1$ throughout the entire sample size, the model selection based estimators tend to point to a greater number of factors as predicted by

Proposition 2 of Chen et al. (2012). Looking at Figure 2.2 for instance we note that the number of factors estimated using the full sample (i.e. $\widehat{k}_{100}$) is on average equal to $2(=2r_0)$ when using the criterion IC_{p3} at $T=100$; about 1.6 on average across 1000 replications when using IC_{p1} at $T=100$; and, about 1.2 on average across 1000 replications when using IC_{p2} again at $T=100$.

It is important to note that since the plots involve averages across 1000 replications there were specific instances in which $\hat{\ell}$ displayed no jump even though the underlying DGP had $B=0.5$. Said differently, we have that occasionally the recursive estimator is not picking up the break yielding this smooth looking transition.

Overall we observe a clearly visible smooth transition in our recursively estimated number of factors $\widehat{k}_\ell$ starting at or around $t=\ell_0$ and which we interpret as evidence of parameter instability (recall that our plots display averages across replications). This transition from one to a greater number of factors starts to occur after about five periods under IC_{p1}, eleven periods under IC_{p3} and only one period for IC_{p3}. Similar patterns also characterize our experiments presented in Figures 2.2-2.10. In all cases, either when the time dimension increases and/or as the size of the break also increases, the smooth looking transition becomes steeper indicating a sharper ability to detect the break. We should stress that, in tables to be presented later, we display precisely the number of times a break is detected (analogously one could think of the number of times the break is not detected) out of 1000 simulations, in different time locations.

When the break is large as when $B=0.7$, $B=0.9$ or $B=1$ (see Figures 2.3-2.5) all three information criteria detect the structural break immediately at the time it takes place. Looking at Figure 2.5 for instance we note that the recursively estimated number of factors display an abrupt jump at or in the very close vicinity of $\ell=50$ and all three criteria tend to display a very similar behavior on average. Table 2.1 below also displays the frequencies of correct decision associated with each criterion. We note that if our concern is that of detecting a break (rather than pinpointing its precise location), our recursive method has accuracy rates close to 100% under large breaks (here for $B\geq 0.7$).

FIGURE 2.2: DGP1 Comparing the performance of the recursive information criteria for a structural break of size $B = 0.5$.

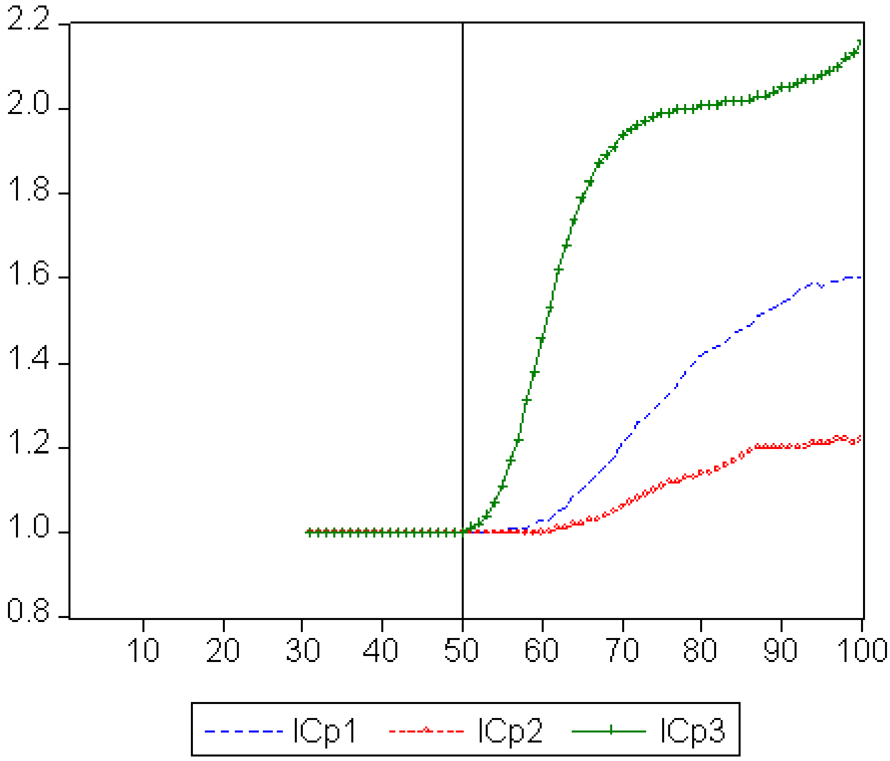


FIGURE 2.3: DGP1 Comparing the performance of the recursive information criteria for a structural break of size $B = 0.7$.

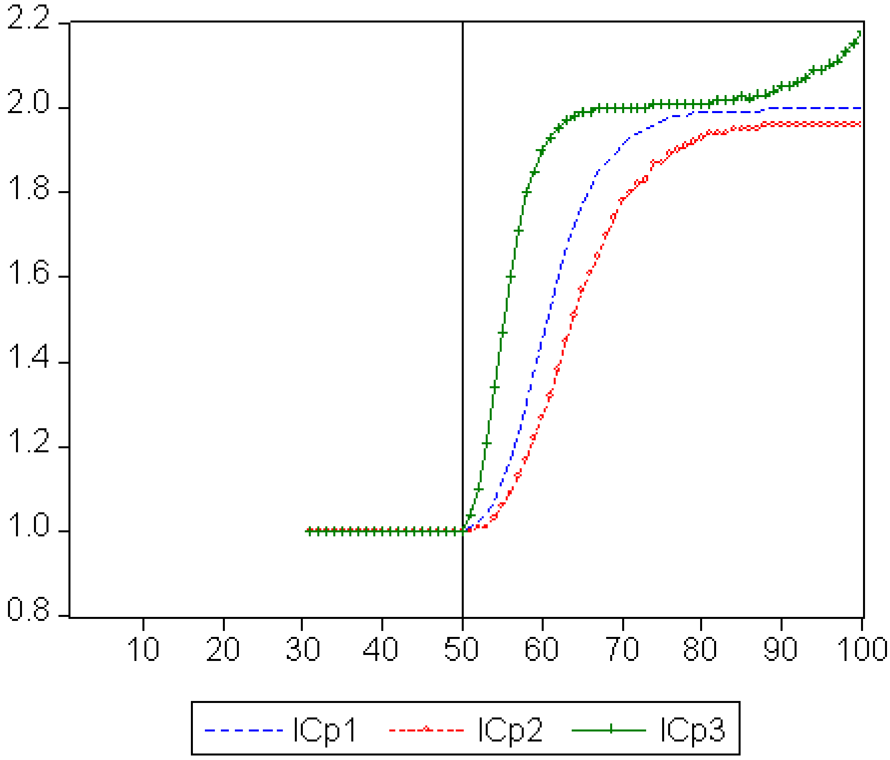


FIGURE 2.4: DGP1 Comparing the performance of the recursive information criteria for a structural break of size $B = 0.9$.

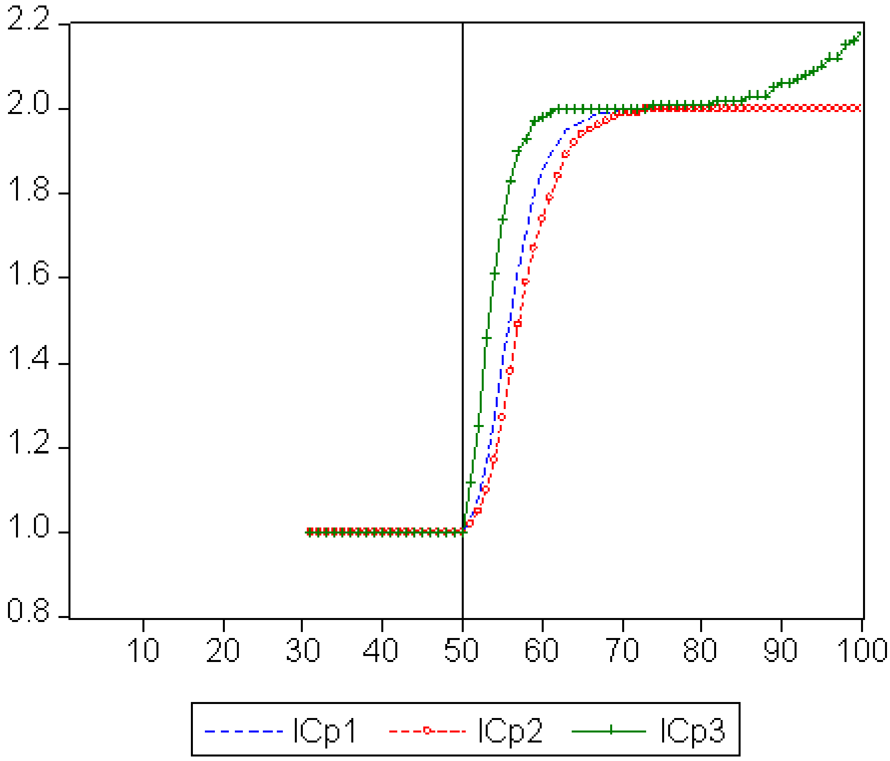
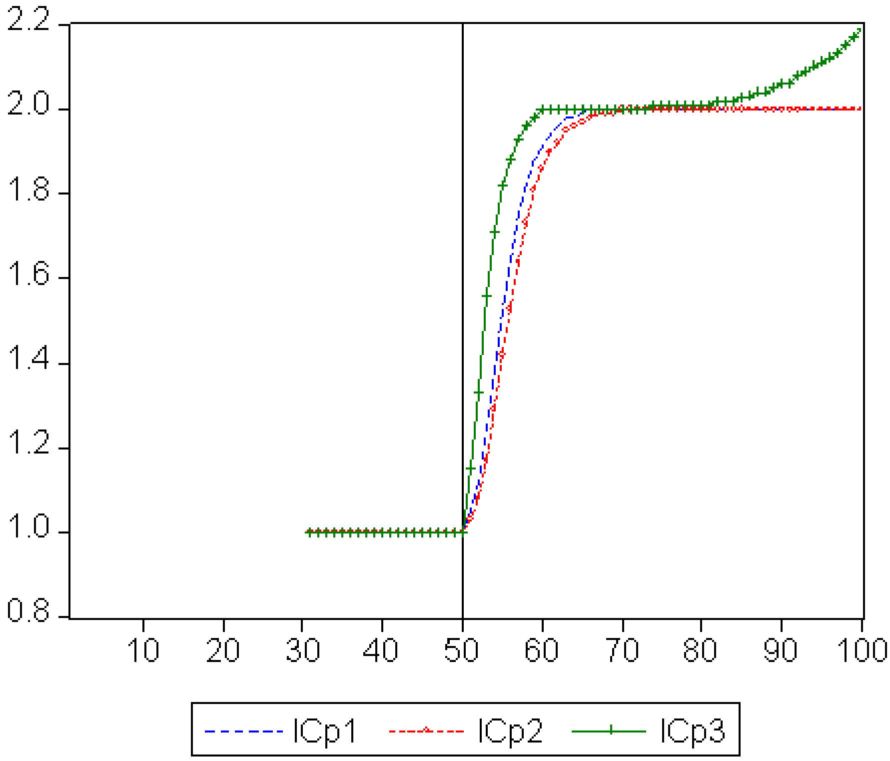


FIGURE 2.5: DGP1 Comparing the performance of the recursive information criteria for a structural break of size $B = 1$.



Our second specification considers a factor model with a change in its factor structure modeled as a new factor kicking in after some time period ℓ_0 (DGP2)

$$X_{it} = \begin{cases} \Lambda_{i,1} F_{tr_0} + \epsilon_{it} \text{ for } t = 1, ..., \ell_0 \\ \Lambda_{i,1} F_{tr_0} + \Lambda_{i,2} G_{tq_0} + \epsilon_{it} \text{ for } t = \ell_0 + 1, ..., T \end{cases} \tag{2.14}$$

Up to time ℓ_0 the model is characterized by r_0 factors while after time ℓ_0 the number of factors jumps to $r_0 + q_0$. Unlike our $DGP1$ we do not assume a break in the factor loadings. We consider two alternative parameterizations of (2.14).

$$\text{DGP2a (Figure 2.6)} : X_{it} = \begin{cases} \Lambda_{i,1} F_{tr_0} + \epsilon_{it} \text{ for } t = 1, ..., 50 \\ \Lambda_{i,1} F_{tr_0} + \Lambda_{i,2} G_{tq_0} + \epsilon_{it} \text{ for } t = 51, ..., T \end{cases}$$

where $r_0 = 1$, $q_0 = 1$. And,

$$\text{DGP2b (Figure 2.7)} : X_{it} = \begin{cases} \sum_{j=1}^{r_0} \Lambda_{i,1} F_{tj} + \epsilon_{it} \text{ for } t = 1, ..., 50 \\ \lambda_{i,2} G_{tq_0} + \epsilon_{it} \text{ for } t = 51, ..., T \end{cases}$$

where $r_0 = 2$, $q_0 = 1$.

Results for the above two parameterizations are presented in Figures 2.6-2.7 which display the profile of the recursively estimated $\widehat{k}'_\ell s$. We again note that under both scenarios our recursively estimated number of factors exhibit a jump that occurs in the vicinity of $t = \ell_0 \equiv 50$.

Our proposed modified estimators of the number of factors associated with DGP2a and DGP2b are given by

$$\widehat{k}^*_{DGP2a} = 2 - 1 = 1$$

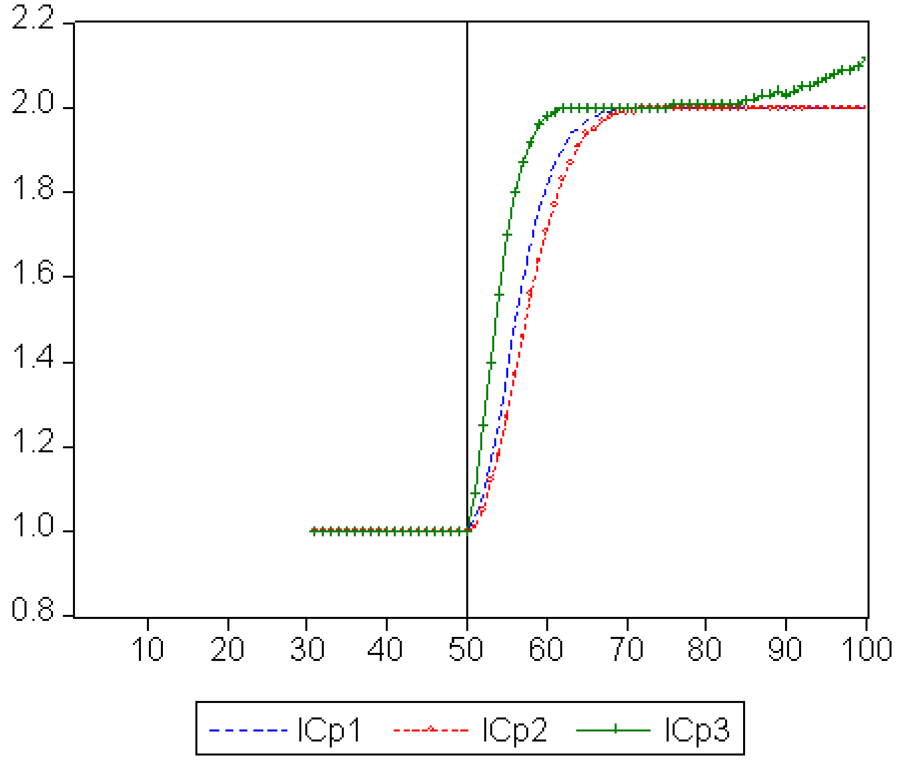
and

$$\widehat{k}^*_{DGP2b} = 3 - 1 = 2$$

As before, the recursive information criteria overestimate the number of factors, and the graphical analysis is similar to larger structural breaks of DGP1. All of the recursive information criteria detect the change in the factor structure right after its occurrence, and the number of factors moves away from one to two.

Now, extending the analysis, not only we allow for the number of factors to change at some point in the time dimension, but we also allow the number of factors to decrease, resulting in the graph of Figure 2.7. In this DGP two factors ($r_0 = 2$) disappear when the changes in the factor structure take place, and one new factor ($q_0 = 1$) kicks in. For

Figure 2.6: DGP2a Comparing the performance of the recursive information criteria when there is only factor along the model but a new factor takes places after the change in the factor structure.



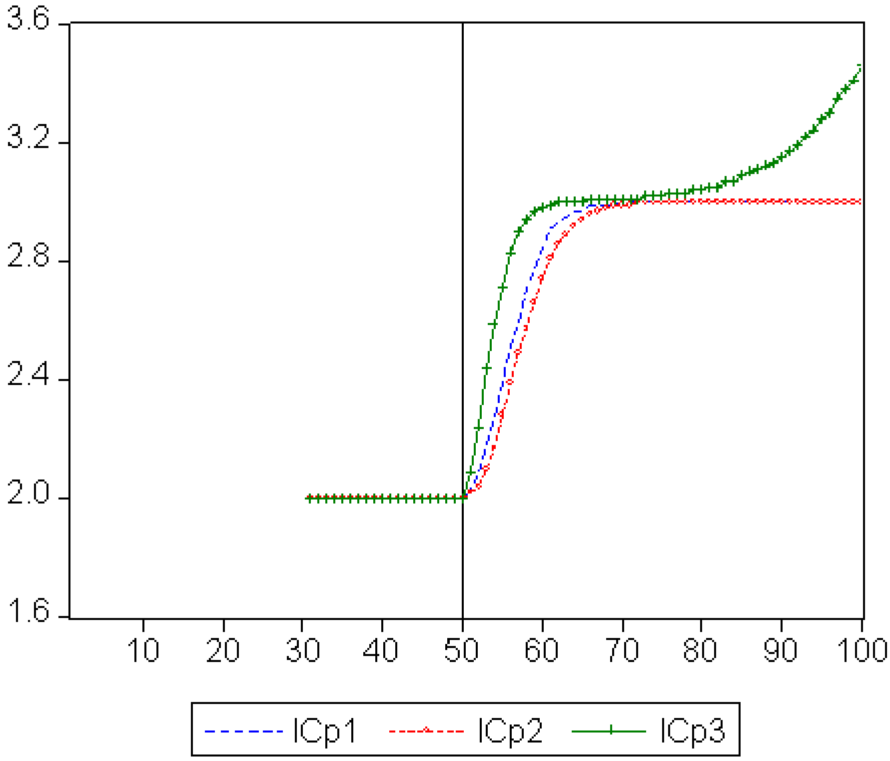
instance, this situation could be thought in cases when a set of independent economies form an economic blocs of any type (monetary, customs or trade blocs). Perhaps, a smaller quantity of factors explains the comovement of a dataset after a bloc is formed.

The information criteria perform correctly before the change in the factor structure pointing out to two factors, but after that the recursive information criteria move to three factors when in reality there is only one. Instead of decreasing the number of factors as it is the case in the true model, it is actually doing the opposite. The result of DGP2b highlights the fact that the information criteria overestimate the number of factors whenever there is any type of relatively large temporal instability.

The purpose of the above experiments was to highlight the timely nature of the recursive estimator of the number of factors $\widehat{k}_\ell$, which tends to jump exactly at or at the close vicinity of $t = \ell_0$. The jump occurs regardless of whether the structural break occurs in the loadings or in the deeper factor structure characterizing the DGP.

The goal of this chapter is to propose a method for detecting the presence of a break and not necessarily that of estimating the number of factors. For this reason it is important to conduct a more comprehensive simulation study designed to assess the performance

FIGURE 2.7: DGP2b Comparing the performance of the recursive information criteria when there are two factors before the change in the factor structure and only one (new) factor after the change in the factor structure: the number of factors decrease after the structural change.



of our decision rule based detection of a structural break. More specifically, using the above DGP we computed the number of times our decision rule led to a linear versus a structural break type decision across our 1000 replications. Results are displayed in Tables 2.1-2.2 below and focus solely on the use of the criteria $IC_{p1,\ell}$ and $IC_{p2,\ell}$.

Note that the detection of a break through our decision rule is equivalent to deciding for a structural break in the factor model so that the figures presented in the top right panels of Tables 2.1-2.2 can be interpreted as the power performance of our decision rule when $B \neq 0$ and as its *size* when $B = 0$.

The true break is set at $t = 50$. From the figures presented in Tables 2.1-2.2, we note that even under a medium sized break of $B = 0.5$, our decision based rule points to a break 60% of the time, and under large breaks 100% of the time when the recursive method is allowed to cover the entire data span (i.e., when $T = 100$).

Examining Table 2.1 more closely, where $IC_{p1,\ell}$ has been used, for DGP1 $B = 0.0$ (no break), the frequency of break detection is zero across the entire time dimension. Considering the case of a large break (e.g. $B = 1$) we note that at $t = \ell_0 + 1 = 51$, the frequency of break detection is only 4.6% but it quickly reaches a frequency of about 91%

at $t = 60$. Full detection is achieved at $t = 75$. For DGP2a and DGP2b the associated correct decision frequencies are very large; full precision is achieved with DGP2b when the number of factors decreases after the break. Similar results are achieved with the use of $IC_{p2,\ell}$ albeit with slightly less accuracy.

TABLE 2.1: Frequencies of break detection (in %) in each of the DGPs using $IC_{p1,\ell}$

$N = 100$, $T = 100$, break at $t = 50$, 1000 simulations
DGP a la Breitung and Eickmeier (2011)

	Size of the Break in the Loadings	True Number of Factors	Frequency of break detection (in %) at			
			$t = 51$	$t = 60$	$t = 75$	$t = 100$
DGP1	$B = 0.0$	1 before and 1 after the break keeping the same factor	0	0	0	0
	$B = 0.5$		0	2.6	31.1	60
	$B = 0.7$		0.5	45.4	97.3	99.9
	$B = 0.9$		2.7	84.8	99.9	100
	$B = 1.0$		4.6	90.9	100	100
DGP2a	-	1 before the break that dies out 1 new after the break	2.6	81.4	100	100
DGP2b	-	2 before the break that die out 1 after the break	100	100	100	100

2.5 Comparing with the Test Based Approaches of Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012)

At this point it is useful to highlight the key differences between our approach and the methods proposed by Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012). Firstly, our approach does not require prior knowledge of the number of factors. The need for prior knowledge of the number of factors, typically estimated according to the Bai and Ng model selection criteria, may adversely impact these three existing methods and may lead to misleading results. In this regard, the tests proposed in Han and Inoue (2012) and Chen et al. (2012) most likely need the factors to be overestimated to achieve better performance.[1] Secondly, to detect nonlinearities in factor models, contrary to the methods available so far, our approach seems to be reliable even under very small sample sizes. Finally, our approach is based entirely on model selection principles and does not require any distributional results.

[1] See Remark B of Breitung and Eickmeier (2011).

Table 2.2: Frequencies of break detection (in %) in each of the DGPs using $IC_{p2,\ell}$

$N = 100$, $T = 100$, break at $t = 50$, 1000 simulations
DGP a la Breitung and Eickmeier (2011)

	Size of the Break in the Loadings	True Number of Factors	Frequency of break detection (in %) at			
			$t = 51$	$t = 60$	$t = 75$	$t = 100$
DGP1	$B = 0.0$	1 before and 1 after the break keeping the same factor	0	0	0	0
	$B = 0.5$		0	0.2	10.6	21.5
	$B = 0.7$		0.4	26.9	87.3	96.3
	$B = 0.9$		1.6	73.8	99.6	100
	$B = 1.0$		2.9	85.8	99.9	100
DGP2a	-	1 before the break that dies out 1 new after the break	1.3	71.2	100	100
DGP2b	-	2 before the break that die out 1 after the break	100	100	100	100

2.5.1 Breitung and Eickmeier (2011)

a Breitung and Eickeimer's method accounts for two possible scenarios: if the break location is known, and if it is unknown. Firstly, if the location is known if the break location is known, their test is equivalent to a Chow (1960) type of test for each of the series on the estimated number factors using Bai and Ng (2002) information criteria to select the number of factors. If the break location is unknown, they use a Sup-type test, as in the fashion of Andrews (1993). Breitung and Eickeimer's null hypothesis tests the stability of the factor loadings individually, by regressing each variable onto the estimated factors.

As discussed in Chen et al. (2012), the method proposed by Breitung and Eickmeier (2011) faces the following three issues. Firstly, their test may fail to detect the presence of a break since the number of factors is very likely to be overestimated. This situation is likely to occur under large breaks and when Bai and Ng (2002) information criteria is used as a prior method to estimate the number of factors. Factor models subjected to a large break can be viewed as a factor model with additional factors (an augmented model), as a linear model. However, factor models that are subjected to a large break are in fact non-linear. In these situations, their test may face an augmented model and will fail to detect the presence of the break. Secondly, Breitung and Eickmeier (2011) test is applied in a series by series basis. The authors impose strong operating assumptions on the idiosyncratic error term of the factor model which are not realistic for economic time series. They impose this to make their pooled test statistics, and thus go on to proceed with an analysis of the stability of the factor loadings. Thirdly, as their test is constructed according to a series by series structural break test framework, and since the factor space is overestimated with the conventional information criteria, their method is also susceptible to incorrectly identifying which of the time series are subjected to a break.

The type of structural break studied here so far is borrowed from Breitung and Eickmeier (2011). Our method has shown promising results because we are not misled by any overestimation of the conventional information criteria.

Comparison with Breitung and Eickmeier (2011) Comparing our approach to that of Breitung and Eickmeiner, when no breaks are assumed, our method does not detect any break in any time in the recursive estimate. This can be seen in the first line of table 2.1 (DGP1, B=0). The estimate is time consistent, as expected, and this result corroborates the simulations of Bai and Ng (2002). On the other hand, Breitung and Eickmeier (2011) tests may find it difficult to detect the linearity of the model with full precision. Nevertheless, in their Table 2.2, amongst their proposed tests, one appears to show better performance for the empirical size in most of sample sizes.

When dealing with breaks of known location (Table 3 of Breitung and Eickmeier (2011)), their tests yields more empirical power in cases of larger sample sizes and larger breaks. In their best performance they achieve $44,60\%$ average frequency of rejecting of the null hypothesis of a break in $N = 50$ variable-specific tests for structural breaks, with $T = 200$. Fixing $N = 50$, smaller time dimesions T never give more power than the latter percentage we mentioned.

Breitung and Eickmeier's tests are not recursive, which means that the results apply to the entire sample size and moreover, their tests are applied individually to each variable. Chen et al. (2012) highlight the possibility of arriving at incorrect conclusions arising from the use of this procedure. For instance, according to simulations performed by Chen et al. (2012), by submitting the Breitung and Eickmeier (2012) tests to relatively larges breaks, in a fixed proportion of 50% of the variables, rejection of the null hypothesis of no break can be raised in up to 90% of the variables. Additionally, with larger breaks, the number of factors will be overestimated and their tests will lose power with, at most, 30.2% rejection of the null hypothesis of no break with a sample size of N = 100 and T = 300 (Table 3 in Chen et al. (2012)).

Furthermore, when the recursive method is allowed to cover the entire data span (i.e., when t = 100), for B = 0.5, our decision based rule suggests a break 60% of the time, and 100% of the time for larger breaks . In the latter cases, our method produced 100% correct detection of the break in simulations performed (rejecting the null hypothesis of no break 100% of the time).

2.5.2 Chen et al. (2012)

Chen et al. (2012) offer two tools. The first one, which is the main one, focuses on the detection of large breaks. This is developed in a two-step framework. In the first step the factors and the respective number of factors are estimated according to the conventional Bai and Ng (2002) information criteria. Chen et al. (2012) are aware of the fact that those information criteria will overestimate the number of factors in break scenarios, as they demonstrate. In the second step, using the factors and the number of factor, the procedure is to investigate the dependence properties of the estimated factors. They do it implementing a well-known structural breaks test such as Andrews (1993), but adapting it to detect breaks in the coefficient of the factors.

The structural break test is implemented by regressing one of the estimated factors onto the set of other factors, and the structural break is verified in the coefficients of these regressors. If a structural break is detected in one of the coefficients, then the factor model is subjected to structural breaks, either in the factor loadings or in the factor structure (e.g., the appearance of a new factor). A methodological advantage of Chen et al. (2012) is that other regression-oriented methods used to detect structural breaks can also be implemented, allowing their method not only to be easily implemented but also to be tested by other regression-oriented structural tests.

The second tool offered by Chen et al. (2012) is used to distinguish whether the break is either a change in the actual number of factors, or only in the factor loadings. This test is analyzed by ranking the covariance matrix of the estimated factors. If the matrix it is a full rank matrix, then the break is in the factor structure; if the matrix rank is reduced, it highlights a break in the factor loadings. In the present study, however, the aim is to detect the break not the source of the break, the Chen et al. (2012) first tool used.

While comparing our approach to that of Chen et al. (2012) we believe that our approach could complementary. This is because we are not required to choose a number of factors prior to implementing the structural break test as they do. Chen et al. (2012) choose the number of factors with the conventional information criteria taken from Bai and Ng (2002) conventional information criteria. As explained by Chen et al. (2012), the size and power of their tests may vary considerably as a different number of factors is chosen. In this regard, an overestimation of the number of factors is preferable for them.

Comparison with Chen et al. (2012) To compare the graphical approach used in this study to Chen et al. (2012), two DPGs were taken from them, performing 1000 RMCS, with $N = T = 100$. The first DGP contains no break, as follows:

$$\begin{aligned}
\text{DGP3}: X_{it} &= \sum_{j=1}^{r} \lambda_i F_{tj} + \epsilon_{it} \\
F_{tj} &= \beta_j F_{(t-1)j} + \upsilon_{tj} \\
&\text{where } r = 3,\ \beta_1 = 0.8,\ \beta_2 = 0.5,\ \beta_3 = 0.2; \\
&\text{with } \lambda_i,\ \epsilon_{it} \text{ and } \upsilon_{tj} \text{ i.i.d. standard normal variables.}
\end{aligned}$$

The DGP3 specification contains three factors (as in Chen et al. (2012)). Our approach detects three factors in all 1000 simulations. Hence, under the null hypothesis of no large breaks, the approach used here outperforms the results obtained by Chen et al. (2012). It is unnecessary to show these results graphically, since the number of estimated factors always equals three. In this case, Chen et al. (2012) verify the empirical size of their tests using different numbers of estimated factors. Contrasted to the approach used in this study, the tests performed by Chen et al. (2012) are highly dependent on both the dimension of the sample sizes, and also on the choice of the number of factors in the regression-oriented structural break tests. Chen et al. (2012) evaluate their test with two, three and four factors. When we generated a linear model with no breaks, the recursive information criteria achieves the expected outcome. That is, selecting the correct number of factors throughout the whole sample. This enables the correct inference that there are no large breaks (no jumps).

$$\widehat{k}^*_{DGP3} = 3 - 0 = 0$$

where $\widehat{k}^*_{DGP3} = 3$ is in fact the true number of factors. In other words, the recursive information criteria detects no large breaks.

One of the alternatives of Chen et al. (2012) is described in DGP4. It deals with structural breaks in the factor loadings of sizes $B = 0.2$ and $B = 0.4$. The structural breaks are located in the middle of the time dimension. This DGP can be expressed as follows:

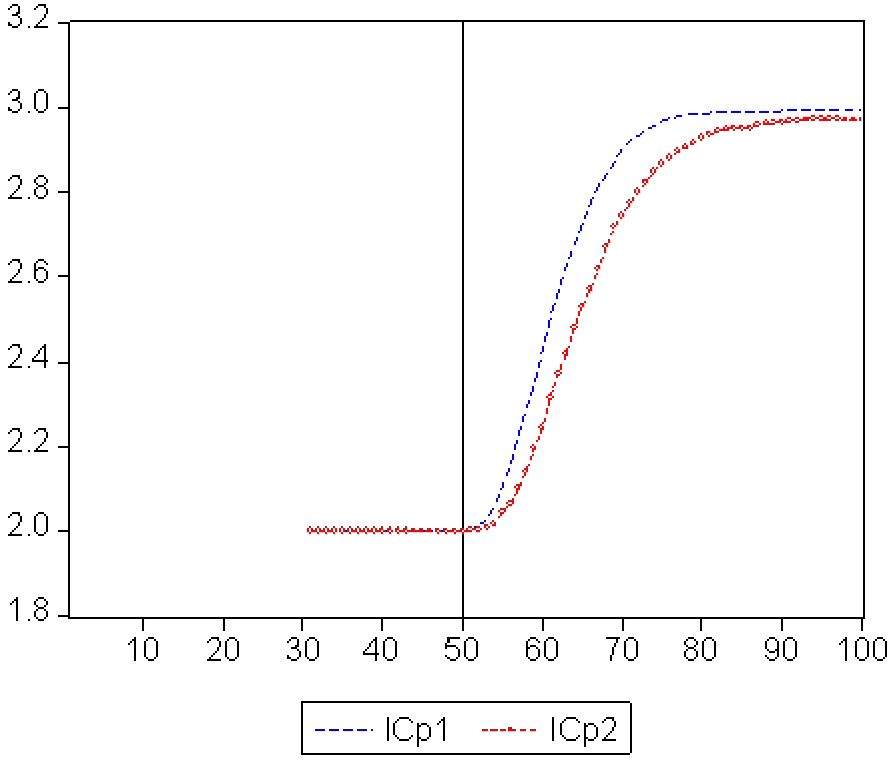
$$
\begin{aligned}
\text{DGP4 (Figure 2.8) :} X_{it} &= \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \epsilon_{it} \text{ for } t = 1, ..., 50 \\
X_{it} &= \sum_{j=1}^{r} (\lambda_{i,1} + b) F_{tj} + \epsilon_{it} \text{ for } t = 51, ..., T \\
F_{tj} &= \beta_j F_{(t-1)j} + \upsilon_{tj} \\
&\text{where } r = 2,\ \beta_1 = 0.8,\ \beta_2 = 0.2, \quad B = 0.2 \text{ and } B = 0.4; \\
&\text{with } \lambda_i,\ \epsilon_{it} \text{ and } \upsilon_{tj} \text{ i.i.d. standard normal variables.}
\end{aligned}
$$

When the break is relatively small, such as when $B = 0.2$, the recursive method is unable to move away from the linear model. That is, the graphical method suggests that there is no change in the number estimated factors. This is to be expected, since the Bai and Ng (2002) information criteria correctly estimate the true number of factors under mild temporal instabilities as evidenced in this case. Using our graphical approach the changes in the number of factors become visible immediately after the structural break, in the occasion of larger breaks such as when $B = 0.4$. This result applies regardless of the information criteria. As it can be seen in our Tables 2.3-2.4, the number of factors starts to become overestimated immediately after the break, and increases monotonically as the time dimension increases. In other words, the break is captured immediately after it occurs. For $IC_{p1,\ell}$, when $t = 51$, one period after the structural break, 0.2% of the simulations detect a large break (overestimation of the number of factors). This percentage increases to 42.8% at $t = 60$ to 96.9% at $t = 75$ only 25 periods after the occurrence of the break, and then to 100% at $t = 100$. The analysis for the $IC_{p2,\ell}$ is analogous (Figure 2.8).

Contrasting with this, under these DGPs, the Chen et al. (2012) tests have different results depending on the selected number of factors. In their framework the performance is improved when the number of factors is overestimated. Additionally, the performance is also improved when the dimension of the sample sizes are larger.

As Chen et al. (2012) is not a recursive test, in the last time observation for $N = T = 100$, the power of their test is very weak when assuming the number of factor equals two. Most of their tests perform well when the number of factors equals three or four, with the Wald test being the best one for rejecting the null hypothesis of no break with 100% precision. As the true number of factors is not known, one has to test across different numbers of factors to infer where breaks occur. A benefit of our recursive approach is that no prior estimation of the number of factors is needed.

FIGURE 2.8: DGP4 Comparing the performance of the recursive information criteria using a Chen et al. (2012) DGP, $B = 0.4$.



$$\widehat{k}^{*}_{DGP4} = 3 - 0 = 3 \text{ for } B = 0.2$$

where $\widehat{k}^{*}_{DGP4} = 3$ is in fact the true number of factors for $B = 0.2$. In other words, the recursive information criteria detect no large breaks.

This result is also in line with Bai and Ng (2002) and Stock and Watson (2002a). These authors explain that the estimate of the factors remains consistent under small temporal instability. However, as we see below, larger instabilities affect the recursive information criteria estimates:

$$\widehat{k}^{*}_{DGP4} = 4 - 1 = 3 \text{ for } B = 0.4$$

where $\widehat{k}^{*}_{DGP7} = 3$ is in fact the true number of factors for $B = 0.4$.

TABLE 2.3: Frequencies of break detection (in %) in each of the DGPs using $IC_{p1,\ell}$

A break detection means a wrong estimate of the true number of factors
$N = 100$, $T = 100$, break at $t = 50$, 1000 simulations

	Size of the Break in the Loadings	True Number of Factors	Frequency of break detection (in %) at			
			$t = 51$	$t = 60$	$t = 75$	$t = 100$
DGP3	-	DGP a la Chen et al. (2011) 2 factors with no breaks	0	0	0	0
DGP4	$B = 0.2$	DGP a la Chen et al. (2011)	0	0	0	0
	$B = 0.4$	2 factors with one break	0.2	42.8	96.9	99.7
DGP5	$B = 1/3$	DGP a la Hi and Inoue (2011)	0.1	30.6	91.4	99.6
	$B = 2/3$	3 factors	5.5	92.0	100	100
	$B = 1$	with	6.8	95.6	100	100
	$B = 2$	one break	2.0	78.6	99.9	100
DGP6	$B = 1$, $\alpha = 0.2$	DGP a la Hi and Inoue (2011)	0	20.2	80.9	98.2
	$B = 1$, $\alpha = 0.4$	3 factors with one break	1.4	71.9	100	100
	$B = 1$, $\alpha = 0.6$	across different fractions	3.6	89.4	100	100
	$B = 1$, $\alpha = 0.8$	of the factor loadings	5.8	93.6	100	100

TABLE 2.4: Frequencies of break detection (in %) in each of the DGPs using $IC_{p2,\ell}$

A break detection means wrong estimation of true number of factors
$N = 100$, $T = 100$, break at $t = 50$, 1000 simulations

	Size of the Break in the Loadings	True Number of Factors	Frequency of break detection (in %) at			
			$t = 51$	$t = 60$	$t = 75$	$t = 100$
DGP3	-	DGP a la Chen et al. (2011) 2 factors with no breaks	0	0	0	0
DGP4	$B = 0.2$	a la Chen et al. (2011)	0	0	0	0
	$B = 0.4$	2 factors with one break	0.1	24.4	86.8	97.2
DGP5	$B = 1/3$	DGP a la Hi and Inoue (2011)	0	30.6	91.4	99.6
	$B = 2/3$	3 factors	2.9	84.7	99.9	100
	$B = 1$	with	5.0	91.0	100	100
	$B = 2$	one break	1.1	64.5	99.3	99.9
DGP6	$B = 1$, $\alpha = 0.2$	DGP a la Hi and Inoue (2011)	0	7.2	57.6	87.8
	$B = 1$, $\alpha = 0.4$	3 factors with one break	0.6	55.1	99.6	100
	$B = 1$, $\alpha = 0.6$	across different fractions	1.5	81	100	100
	$B = 1$, $\alpha = 0.8$	of the factor loadings	3.7	88.5	100	100

2.5.3 Han and Inoue (2012)

The test proposed by Han and Inoue (2012) lies on a null hypothesis of a covariance matrix of the estimated factor loadings according to Bai and Ng (2002) that is constant over time. As previously discussed, their tests perform better when the number of factors is overestimated. In order to compare performances, some of their DGPS were submitted to the recursive information criteria used in this study.

The first one, DGP5, adds the possibility for the size of the structural break to be larger than those previously considered. The results remain with the expected intuition. The second one, DGP6, considers the case in which not all of the factor loadings are subjected to structural breaks, but only a fraction of them. Still, in this case, the results show no surprises, as will be explored below.

Comparing with Han and Inoue (2012) DGP5 is the A1 DGP of Han and Inoue (2012). It is designed to verify the power of their tests when the magnitude of the break in the factor loadings increases, with an additional feature that the variance of the idiosyncratic error term is different than the value of one, and chosen such that $R^2 = \text{trace}E(\epsilon\epsilon')/\text{trace}E(XX')$ is 50%.

$$\text{DGP5 (Figure 2.9)}: X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \kappa\epsilon_{it} \text{ for } t = 1, ..., 50$$

$$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} - b) F_{tj} + \kappa\epsilon_{it} \text{ for } t = 51, ..., T$$

$$F_{tj} \sim \ i.i.d.\ N(0,1),\ \kappa = \sqrt{(1 + b^2/4)r},\ r = 3$$

$$\lambda_{i,1} \sim \ i.i.d.\ N(b/2, 1),\ B = \{1/3,\ 2/3,\ 1,\ 2\}$$

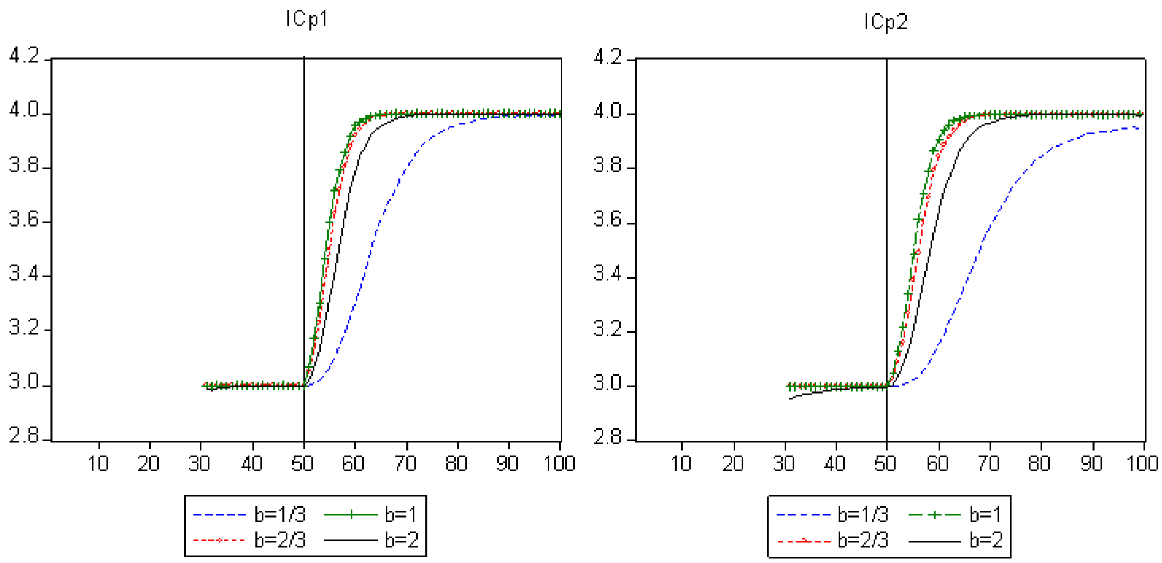
In their simulations, Han and Inoue (2012) to verify the power of their tests, with $N = T = 100$, in the cases of smaller breaks ($B = 1/3$), they reject the null of no break in only less than 10% of the simulations. In this situation they improve performance when the sample sizes increase significantly, reaching full precision of rejecting the null hypothesis of no break with $N = T = 500$. The tests used by Han and Inoue (2012) need larger breaks to improve performance under smaller datasets.

When DGP5 is subjected to our recursive approach, all of the structural breaks are detected without depending on either the dimension of the sample size or on the size of the structural break. At $N = T = 100$ and $B = 1/3$, 99.6% of rejection of the null hypothesis of no break is achieved. Han and Inoue (2012) achieve less than 10% of rejection. More importantly, as our method is recursive, the structural break is detected immediatly its occurrence, with a frequency of 0.1% at $t = 51$, 30.6% at $t = 60$, and 91.4% at $t = 75$. Furthermore, by adjusting for the occurrence of the large structural breaks, the true number of factors can be retrieved, such that $\widehat{k}^*_{DGP5} = 3$.

$$\widehat{k}^*_{DGP5} = 4 - 1 = 3 \text{ for } B = \{1/3,\ 2/3,\ 1,\ 2\}$$

where $\widehat{k}^*_{DGP5} = 3$ is in fact the true number of factors.

FIGURE 2.9: DGP5 Comparing the performance of the recursive information criteria using a Han and Inoue (2012) DGP



Han and Inoue (2012) depend on the prior estimate of Bai and Ng (2002). As the sizes of the breaks increase ($B = 2/3$, $B = 1$ or $B = 2$), the number of factors becomes overestimated in all of the simulations. This helps their test to achieve full precision in rejecting the null hypothesis of no break in most of the sample sizes.

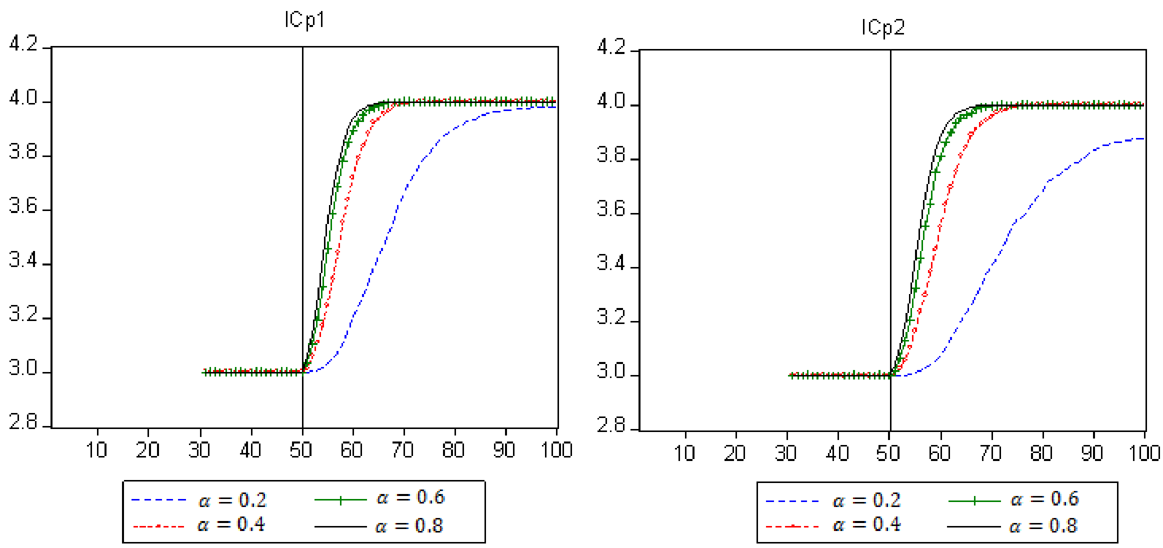
In our recursive case the same intuition was also found with the additional benefit that the recursive frequencies increased substantially, pointing out the occurrence of the break.

The results for DGP5 in Tables 2.3-2.4 ($IC_{p1,\ell}$ and $IC_{p2,\ell}$) clearly show that as either the size of the break increase, or the time dimension increase, or both, the proportion of large break detection also increases. An exception arises here when the break is 'very large'. For example, say for example when $B = 2$. In this case, the detection of the break is smaller, but still follows the previous reasoning.

DGP6 follows the same structure as DGP5 and it is the DGP A2 of Han and Inoue (2012). The difference here is that the structural breaks take place only in a fraction of the factor loadings, and not in all of them. We chose the proportions α of the factor loadings to undergo a structural break, with $\alpha = \{0.2, 0.4, 0.6, 0.8\}$, of size $B = 1$, as in Han and Inoue (2012).

$$\widehat{k}^*_{DGP6} = 4 - 1 = 3 \text{ for } \alpha \in \{0.2, 0.4, 0.6, 0.8\}, \text{ and } B = 1$$

FIGURE 2.10: DGP6 Comparing the performance of the recursive information criteria using a Hi and Inoue (2011) DGP, where α is the fraction of factor loadings subjected to a structural break of size $B = 1$



$$\text{DGP6 (Figure 2.10)}: X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \kappa \epsilon_{it} \text{ for } t = 1, ..., 50$$

$$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} - M) F_{tj} + \kappa \epsilon_{it} \text{ for } t = 51, ..., T$$

$$F_{tj} \sim \ i.i.d. \ N(0,1), \ \kappa = \sqrt{(1 + b^2/4)r}, \ r = 3$$

$$\lambda_{i,1} \sim \ i.i.d. \ N(b/2,1), \ B = 1, \ \alpha \in \{0.2, 0.4, 0.6, 0.8\}$$

$$M \text{ an } N \times \widehat{k}_\ell \text{ with the first proportion } \alpha \text{ of rows equals to } b,$$

$$\text{and the elements of the other rows equals zero}$$

Compared to the previous simulations, we have established that $B = 1$ can be considered a large structural break in the factor loadings due to its impact on the conventional criteria. The results shown in Figure 2.15 demonstrate that for a fraction of at least $\alpha = 0.4$ of the factor loadings undergoing a structural break, the overestimation of the number of factors begins immediately after the structural break. In the recursive estimates, we have the following frequencies of rejecting the null of no break: 0.6% at $t = 51$, 55.1% at $t = 60$, 99.6% at $t = 75$ and 100% at $t = 75$. All the frequencies increase monotonically for larger αs (Figure 2.10).

When comparing this with the results of Han and Inoue (2012) for $\alpha = 0.4$, our test is more powerful as it achieves full precision in the recursive exercise for $N = T = 100$. For larger fractions full precision of rejecting the null hypothesis was achieved even before $T = 100$ (in some cases for $t = 75$), whereas Han and Inoue (2012) do not achieve this outcome in any of their cases for $T = 100$.

What has been found is that not all the factor loadings (i.e. not all the variables) are required to undergo a structural break for the factor model to change its structure over the time. Then adjusting for the occurrence of breaks to reestimate the number of factors $\widehat{k}^*_{DGP6} = 3$.

In all cases, for larger proportions of factor loadings undergoing a structural break, as the time dimension is further away from the structural break, the frequency of structural break detections also increases. Again, Han and Inoue (2012) will need larger datasets to improve their results. However, are not required to obtain a larger data set to achieve this.

2.6 Empirical Application

We use the dataset from Stock and Watson (2005) to apply the recursive model selection based information criteria.[2] A pretreatment is needed as required for factor analysis: logarithms taken when necessary, differences taken until stationarity is achieved, seasonal adjustment, outliers[3] removal and normalization (zero mean and unit variance). The procedure of Stock and Watson (2005) was followed using a balanced monthly dataset from 1960:01 to 2003:12, consisting of 528 time series observations for each of the 132 variables. Following the suggestion in Breitung and Eickmeier (2011), who use the same dataset, we assessed the dataset with outliers adjusted and without adjusted outliers.

We first assessed the conventional information criteria for the entire dataset and for subsamples of candidate structural break of 1984:01 for the US economy, when output growth and inflation showed lower volatility, starting the so-called Great Moderation. Results for both outliers adjusted and not adjusted datasets, are displayed in Table 2.5.

The results in Table 2.5 corroborate the RMCS in the sense that IC_{p2} is the most parsimonious criterion, followed by IC_{p1} and IC_{p3}; the latter fails to do as required.

[2]This dataset is available on their website and we downloaded it in the 10^{th} of June, 2011: http://www.princeton.edu/ mwatson/wp.html

[3]We define an outliers exactly as footnote 11 of Breitung and Eickmeier (2011). An observation of each (stationary) variable with absolute median deviations larger than six times the interquartile range is considered an outlier and is replaced by the median value of the preceding five observations.

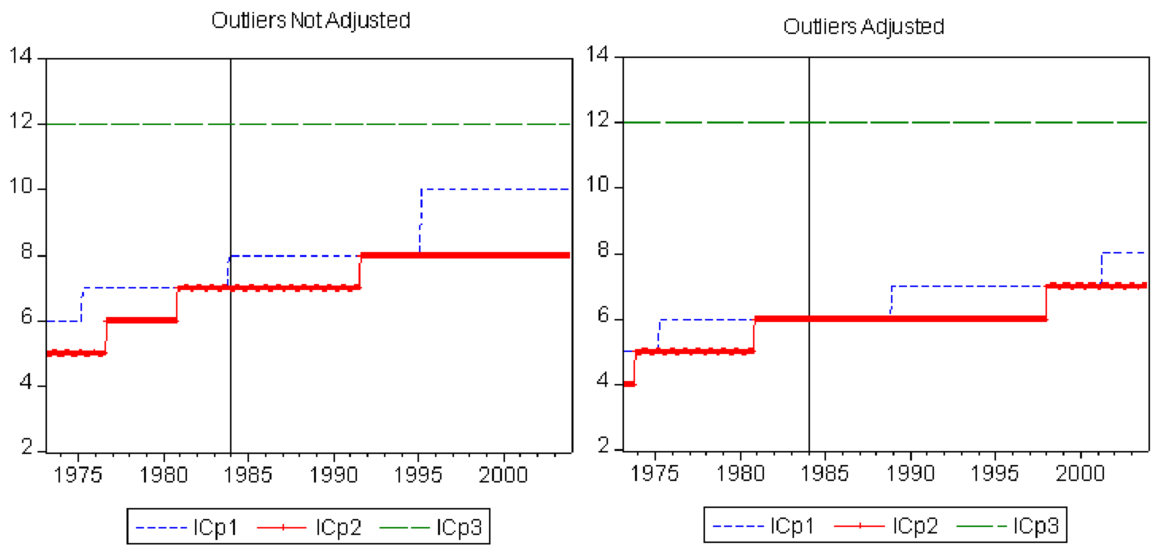
TABLE 2.5: Conventional Bai and Ng (2002) Information criteria

	Outliers not adjusted			Outliers adjusted		
	IC_{p1}	IC_{p2}	IC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
1960:01 to 2003:12	12	12	12	7 (7)	6	12
1960:01 to 1983:12	9	6	12	6 (4)$^{\mp}$	6	12
1984:01 to 2003:12	12	9	12	10 (6)$^{\diamond}$	6	12

Notes:
Results of Breitung and Eickmeier (2011) shown in brackets.
The maximum number of factors allowed is 12. Symbols:
$^{\mp}$ 5-8 factors resulted in very close criteria in their numerical statistical terms; and,
$^{\diamond}$ 7-11 factors resulted in very close criteria in their numerical statistical terms.
By "very close" we mean that each of the information criterion statistic used to estimate the number of factors differs from each other by a value less than or equal to 0.001; hence they are very close.
Breitung and Eickmeier (2011) also found results along similar numbers.

The maximum number of estimated factors was 12 and IC_{p3} reached this upper limit number in all empirical exercises.

FIGURE 2.11: Recursive information criteria with Stock and Watson (2005) data 1960:01 to 2003:12. Results start to be estimated at 30% of the sample, that is, at 1973:03



The entire dataset Our recursive information criteria starts to display results at 30% of the time dimension, and the number of factors estimated are plot in Figure 2.11. The dataset of Stock and Watson (2005) contains 528 monthly observations for each of the 132 variables. Hence, the number of factors starts to be estimated at observation 158, which corresponds to 1973:04. The results of the recursive exercise for outliers not-adjusted (left-hand side) and adjusted outliers (right-hand side) are shown. The results are not strong enough to suggest that the conventional information criteria are sufficiently reliable since it is clear that the recursive information criteria shown graphically displays jumps in the number of estimated factors as the time dimension increases ($IC_{p1,\ell}$ changes the number of factors in 1973:08, 1977:09 and 2000:08, and, $IC_{p2,\ell}$ changes the number of factors in 1973:09, 1976:07, 1977:08 and 1983:06).

Oil Crises In Figure 2.11, the adjusted outliers (right-hand side), the number of estimated factors is highly unstable (many jumps) before 1984:01 (vertical line), for both $IC_{p1,\ell}$ and $IC_{p2,\ell}$. This instability could be explained by the two oil crises (1973 and 1979), which were sources of nonlinearities in this US dataset. In other words, the instability during the 1970s is an explanation of this event, and therefore these big economic events are captured as changes in the factor structure.

Great Moderation Again in Figure 2.11, from 1983:06 onwards, the number of estimated factors becomes rather stable in both information criteria. That is, in both of the recursive estimates the occurrence of the Great Moderation can be inferred. The last jump in $IC_{p1,\ell}$ occurred at 1977:09 (from 5 to 6 factors) before the Great Moderation. Afterwards, a jump took place only at 2000:09, from 6 to 7 factors. The $IC_{p2,\ell}$ criterion may suggest that a structural break took place at 1983:06, from 5 to 6 factors, and no changes in the factor structure is seen afterwards. There is a vertical line drawn at 1984:01 where the Great Moderation is known to have started, this is to help with identification.

It has been found that from 1984:01 until early in the 2000s, there are no changes in the number of factors. This suggests a period of stability in the comovement of this dataset. However, the non-adjusted outliers (left-hand side) result in too many factors and jumps, which makes inference difficult.

The results show that the number of factors changes considerably along the time dimension, describing the usefulness of the recursive approach for uncovering the factor structure. We suggest that in order to determine the number of factors the recursive information criteria should be applied in the entire sample, and then our method from equation (2.12) should be used.

Using our modified estimator of the number of factors Using the method described in equation (12) on the adjusted outliers (right-hand side) of Figure 2.11, the number of estimated factors are the following:

$IC_{p1,\ell}$ Locations of changes in the number of factor estimates: 1973:08, 1977:09 and 2000:08. Then, $\widehat{k}_T = 7$ is the number of factors estimated at the last time dimension (equivalently using the conventional Bai and Ng (2002) criteria; the notation $\widehat{k}_T$ is used henceforth for this purpose), and $\phi = 3$ is the number of jumps using the recursive graphical approach. Then the modified estimator to retrieve a more parsimonious number of factors aiming to achieve an estimator closer to the true number of factors (if not the actual number), can be used as follows:

$$\widehat{k}^{*}_{IC_{p1,\ell}} = 7 - 3 = 4 \tag{2.15}$$

Applying the same reasoning for $IC_{p2,\ell}$ we have:

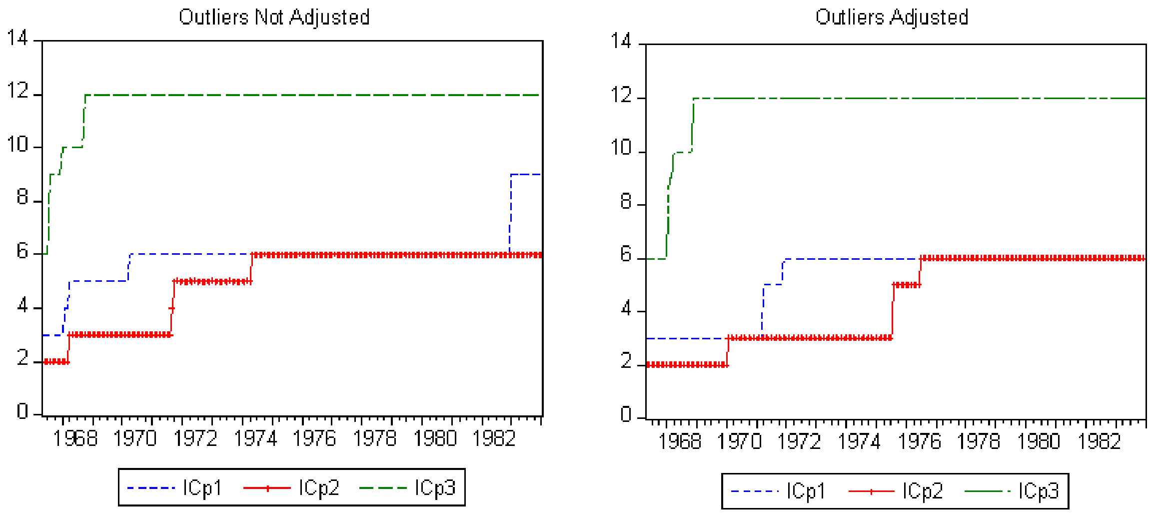
$IC_{p2,\ell}$ Locations of changes in the number of factor estimates: 1973:09, 1976:07, 1977:08 and 1983:06. Then, $\widehat{k}_T = 6$, $\phi = 4$.

$$\widehat{k}^{*}_{IC_{p1,\ell}} = 6 - 4 = 2, \tag{2.16}$$

Splitting the dataset The sample was also split as in Stock and Watson (2008) and Breitung and Eickmeier (2011). This approach can be considered *ad hoc* since splitting the sample can lead to a different comovement of the remaining datasets. Furthermore, one could still think of an existing structural break in a series, individually around a certain date, but, it is still unknown if this structural break in this series could indeed lead to a possible change in the factor structure. Nevertheless, we split the sample to investigate this alternative.

To follow Stock and Watson (2005) and Breitung and Eickmeier (2011) closely, we split the sample in two at 1984:01. Then we perform the recursive information criteria on the subsamples with results shown in Figures 2.12 and 2.13. Splitting the dataset into smaller subsets at predetermined dates is not conclusive, because changes in the number of factors may appear at other time locations. Analyzing the dataset with adjusted outliers (right-hand side), Figure 2.12 shows two changes in the factor structure for $IC_{p1,\ell}$ and three changes for $IC_{p1,\ell}$. They both result with the same number of factors, but $IC_{p2,\ell}$ is subjected to a bigger adjustment in this case, as shown below.

FIGURE 2.12: Recursive information criteria with Stock and Watson (2005) data 1960:01 to 1984:01; results start to be estimated at 30% of the sample, that is, 1973:03



Using the method described in equation (12), with the adjusted outliers (right-hand side) of Figure 2.12, the number of estimated factors are the following:

$IC_{p1,\ell}$ Locations of changes in the number of factor estimates: 1971:05 and 1971:12. Then, $\widehat{k}_T = 6$, $\phi = 2$.

$$\widehat{k}^*_{IC_{p1,\ell}} = 6 - 2 = 4 \tag{2.17}$$

$IC_{p2,\ell}$ Locations of changes in the number of factor estimates: 1970:02, 1975:08 and 1976:07. Then, $\widehat{k}_T = 6$, $\phi = 3$.

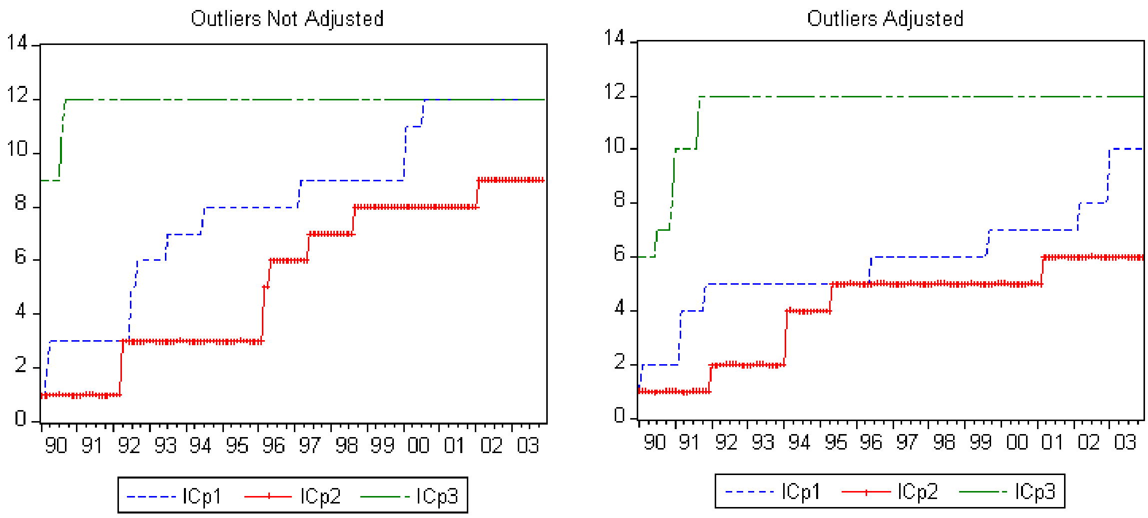
$$\widehat{k}^*_{IC_{p1,\ell}} = 6 - 3 = 3, \tag{2.18}$$

Again, using the method described in equation (12), on the adjusted outliers (right-hand side) of Figure 2.13, the number of estimated factors are the following:

$IC_{p1,\ell}$ Locations of changes in the number of factor estimates: 1990:02, 1991:03, 1991:11, 1996:06, 1999:09, 2002:03, 2003:01. Then, $\widehat{k}_T = 10$, $\phi = 7$.

$$\widehat{k}^*_{IC_{p1,\ell}} = 10 - 7 = 3 \tag{2.19}$$

FIGURE 2.13: Recursive information criteria with Stock and Watson (2005) data 1984:02 to 2003:12; results start to be estimated at 30% of the sample, that is, 1990:01



$IC_{p2,\ell}$ Locations of changes in the number of factor estimates: 1992:01, 1994:02, 1995:05 and 2001:03. Then, $\widehat{k}_T = 6$, $\phi = 4$.

$$\widehat{k}^*_{IC_{p2,\ell}} = 6 - 4 = 2, \tag{2.20}$$

In Figure 2.13 there are seven jumps with $IC_{p1,\ell}$. The conventional criteria yields ten factors, then, it results in three factors with our modified estimator (10-7=3). In its turn, $IC_{p2,\ell}$ has four jumps, leading to two factors (6-4=2).

Our results also suggest that the overestimation of the number of factors may likely to be the case in many empirical applications that use the conventional Bai and Ng (2002) information criteria. This is possibly due to unknown nonlinearities in the factor structure. Our method can help to recalculate the true number of factors, adjusting for changes in the factor structure. Then, in factor model applications one may apply our recursively information criteria in the entire dataset to uncover nonlinearities in factor models, following with our modified estimator to retrieve a more parsimonious number of factors.

2.7 Conclusion

In this chapter our goal was to propose an alternative approach to determine the number of factors in large approximate factor models under nonlinear structures. The methodology expands on the information criteria proposed by Bai and Ng (2002) by uncovering the time dimension of the factor structure and applying it recursively at each point in the time dimension, rather than applying it once in the entire sample. Then, a simple method to adjust for the overestimation of the number of factors is offered. It is a modified estimator to achieve a more parsimonious number of factors. Overall, our approach aims to enable the detection of changes in the factor structures, likely to result in changes in the number of factors along the same factor model. We suggested that our method should be used in the entire sample, without the need for splitting the sample at predetermined dates. In this manner, a practitioner could estimate the factor space and use our approach to uncover nonlinearities of a factor model, without running the risk of inadvertently overestimating the number of factors.

Chapter 3

Structural Changes in Large Dimensional Factor Models: Model Selection Based Inference

3.1 Introduction

The primary usefulness of factor models is to condense the information included in a large dataset into a small number of common factors. Typically, the underlying large data set is expressed as a *linear* combination of a small set of common factors which may or may not have a direct economic interpretation. Examples include empirically derived asset pricing model aiming to decompose stock returns into a linear combination of risk factors (e.g. the Fama and French three factor model) amongst numerous others.

The goal of a factor model based investigation typically involves estimating the number of factors associated with a dataset X_{it} $i = 1, \ldots, N$, $t = 1, \ldots, T$ together with the loadings or sensitivities corresponding to each factor. In this context a popular method with wide applicability and generally weak underlying assumptions is Bai and Ng's model selection based approach (see Bai and Ng (2002)). The latter views the problem of the determination of the number of factors as a model selection problem where one chooses an optimal model via the optimisation of an information theoretic criterion. Estimation is performed using asymptotic principal components and the method is valid under both large N and large T.

More recently numerous authors have also raised the possibility that linear factor models may occasionally be subject to structural breaks. One such example involves a factor model whose factor loadings undergo a structural change after some unknown period.

Three important papers that have formally tackled the issue of the detection of the number of breaks in factor models are Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012). All three papers aim to develop methods for uncovering the occurrence of breaks in the loadings of an otherwise linear model through the development of formal test based methods (e.g. SupWald, Chow).

The goal of this chapter is to instead propose a model selection based approach for the detection of breaks in the loadings of factor models. We view the problem of the estimation of the number of factors and the detection of the potential presence of a break in their loadings jointly.

We initially develop a novel estimator of the number of common factors that is robust to the presence or absence of a break in the factor loadings. This is particularly important since it is well known that conventional criteria are unable to point to the correct number of factors if the underlying model has a break. Typically when the presence of a break is ignored, the Bai and Ng type of model selection criteria lead to an overestimated number of factors both in small and large samples.

We subsequently use our novel estimator to introduce a simple decision rule based approach to detecting whether a break is truly present or not. This is achieved through a comparison of our novel estimator of the number of factors with the conventional one obtained under the assumption of linearity (i.e. the standard Bai and Ng criteria).

Since the existing test based methods for detecting the presence of a break (e.g. Breitung (2011), Chen et al. (2011), Han and Inoue (2011)) all require an estimated number of factors as an input we also view our proposed estimator as a valuable input for the implementation of such tests since our estimator remains consistent regardless of whether the underlying model has a break or not.

The plan of the chapter is as follows. In Section 3.2 we introduce the core factor model and our proposed model selection based estimator of the number of factors. Section 3.3 focuses on a decision rule based approach for deciding between a linear or break based factor model specification. Section 3.4 evaluates the performance of our proposed methods within a comprehensive simulation exercise. Section 3.5 concludes.

3.2 Detecting a Break in Loadings: A Model Selection Based Approach

Following standard notation in the literature we formulate our core linear factor model as

$$X_{ti} = F_t \Lambda_i' + e_{ti} \tag{3.1}$$

where $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ are the cross-section and time series dimensions respectively; X is an $N \times T$ matrix of the observed dataset; $\Lambda_i = [\lambda_{i1}, ..., \lambda_{ir}]$, $F_t' = [f_{1t}, ..., f_{rt}]'$, ϵ_{it} and r represent, respectively, the factor loadings, the factors, the idiosyncratic components and the corresponding number of factors.

The type of nonlinearity we are interested in this chapter is given by a single structural break that shifts the loading coefficients at some unknown time period. More specifically, we write

$$X_{it} = \begin{cases} \Lambda_{i,1} F_t' + \epsilon_{it} \text{ for } t = 1, ..., \ell \\ \Lambda_{i,2} F_t' + \epsilon_{it} \text{ for } t = (\ell + 1), ..., T \end{cases} \tag{3.2}$$

with ℓ denoting the break point location. Here $\Lambda_{i,1} = [\lambda_{i,1}, ..., \lambda_{N,1}]$, $\Lambda_{i,2} = (\Lambda_{i,1} + B)$ with B denoting the size of the break. For further use we also define the break fraction time location π as $\ell = [T\pi]$ with $\pi \in (0, 1)$. Note also that the number of factors remains the same before and after the break. Using indicator functions model (3.2) can also be reformulated as

$$X_{it} = \Lambda_{i,1} F_t' I_1 + \Lambda_{i,2} F_t' I_2 + \epsilon_{it} \tag{3.3}$$

Our initial goal is to propose an estimator of the true number of factors regardless of whether the true model is (3.1) or (3.2)-(3.3). This will then allow us to introduce a decision rule based approach for distinguishing between (3.1) and (3.2).

Letting $\Sigma(k, F^k)$ denote the error variance obtained from (3.1) when k factors have been imposed we recall that the standard linear based model selection criteria used for estimating the number of unknown factors are expressed as

$$IC_L(k_L) = \ln |\widehat{\Sigma}(k_L, \widehat{F}^{k_L})| + k_L g(N, T) \tag{3.4}$$

with $g(N, T)$ denoting a deterministic penalty term that is a function of both the cross sectional and time series dimension. As a matter of notation we index the number of factors obtained from (3.5) as k_L to highlight the fact that the model selection criterion is evaluated from the linear factor model specification. Here $0 \leq k_L \leq k_{Lmax}$ and following Bai and Ng, the optimal number of factors is estimated as the minimiser of

$IC_L(k_L)$. Formally $\hat{k}_L = \arg\min_{0 \le k_L \le k_{Lmax}} IC_L(k_L)$. At this stage it is also important to recall the now well known result that if the factor model is characterised by a break in its loadings the estimator given by $\hat{k}_L$ will typically overestimate the true number of factors r_0 (see Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012)).

We now introduce a generalized model selection criterion designed to be suitable under both linear and structural break specifications. Specifically, our criterion is denoted $IC_{NL}(.)$ and specified as

$$\begin{aligned} IC_{NL}(\ell, k_{NL}) = & \left\{ \frac{\ell}{T} \ln |\widehat{\Sigma}_1(k_{NL}, \widehat{F}^{k_{NL}})| + k_{NL} g(\ell, N) \right\} \\ & + \left\{ \frac{T-\ell}{T} \ln |\widehat{\Sigma}_2(k_{NL}, \widehat{F}^{k_{NL}})| + k_{NL} g(T-\ell, N) \right\}. \end{aligned} \tag{3.5}$$

Here $\widehat{\Sigma}_1(k_{NL}, \widehat{F}^{k_{NL}})$ refers to the residual variance obtained from a fitted linear model with k_{NL} factors using $t = 1, \ldots, \ell$ and $\widehat{\Sigma}_2(k_{NL}, \widehat{F}^{k_{NL}})$ is the residual variance corresponding to the post break regime. Note that the number of underlying factors is not allowed to change before and after the break location.

This information criterion allows for a joint estimation of the location of the break and of the number of factors. For each $0 \le k_{NL} \le k_{NL,max}$ and each $\ell = \ell_1, ..., \ell_2$ (say, $\ell_1 = [T\ 0.10]$ ad $\ell_2 = [T\ 0.90]$) the factor models are estimated and $IC(\ell, k_{NL})$ evaluated. The optimal number of factors and break point location are then obtained as the joint minimisers of $IC(\ell, k_{NL})$.

3.2.1 Theoretical Properties of the Information Criterion

Our estimation procedure follows exactly the ones developed in Bai and Ng (2002) and Bai (2003). Briefly speaking, the main feature of their estimation procedure in a factor model framework is increment of the flexibility that allows for cross-sectional and serial dependence, together with heteroskedasticity in the idiosyncratic component (e_{it}), in addition to weak dependence between the factors and these errors.

The assumptions of Bai and Ng (2002) and Bai (2003) ensure consistent asymptotic principal component estimation of the factors and factor loadings required for the model in (3.1). However, following the same arguments as they did, our nonlinear estimator k_{NL} is also consistent for the true number of factors.

Using a factor model as in equation (3.2), and as mentioned above, with the same estimation procedure of Bai and Ng (2002) and Bai (2003), we are still capable of consistently estimating the correct number of factors when structural breaks take place, without the danger of inadvertently overestimating the number of factors. The following proposition states this consistency.

Proposition 1. Suppose that Assumptions of Bai (2003) hold in addition to Assumption 1 of Chen et al. (2012) (Breaks). As $(T,N)\to\infty$ we have $\lim_{\ell,N\to\infty} P(\widehat{k}_{NL}=r_0)=1$, where r_0 is the true number of factors, provided that the penalty terms satisfy:

(i) $g_1(\ell,N)\to 0$ and $\min(\ell,N)g_1(\ell,N)\to\infty$; and,

(ii) $g_2(T-\ell,N)\to 0$ and $\min(T-\ell,N)g_2(T-\ell,N)\to\infty$.

Proposition 1 establishes the consistency of our nonlinear information criterion to estimate the true number of factors r_0, regardless whether (3.1) or (3.2) hold. The important point to make here is that $\widehat{k}_{NL}$ is a very useful estimator since it is not biased upwards as it is for the case of $\widehat{k}_L$ when the underlying model has a break. It is this feature of $\widehat{k}_{NL}$ relative to $\widehat{k}_L$ that we will use to propose a decision rule based approach for distinguishing between (3.1) and (3.2). Said differently, Bai and Ng (2002) information criteria ($\widehat{k}_L$) are not overparameterized if the model is like in equation (3.1), but it is so in cases like (3.2).[1]

Defining the criterion: the penalty term We borrow a penalty term from Bai and Ng (2002) which satisfies the above conditions and accommodate it nonlinearly as follows: $g_1(\ell,N)$ can be defined as $\frac{\ln\min(\ell,N)}{\min(\ell,N)}$ and $g_2(T-\ell,N)$ as $\frac{\ln\min(T-\ell,N)}{\min(T-\ell,N)}$. Then our criterion becomes

$$
\begin{aligned}
IC(\ell,k_{NL}) = &\left\{\frac{\ell}{T}\ln|\widehat{\Sigma}_1(k_{NL},\widehat{F}^{k_{NL}})| + k_{NL}\left(\frac{\ln\min(\ell,N)}{\min(\ell,N)}\right)\right\}\\
&+\left\{\frac{T-\ell}{T}\ln|\widehat{\Sigma}_2(k_{NL},\widehat{F}^{k_{NL}})| + k_{NL}\left(\frac{\ln\min(T-\ell,N)}{\min(T-\ell,N)}\right)\right\}
\end{aligned} \tag{3.6}
$$

This information criterion displays the same advantages as those of Bai and Ng (2002) (such as being a function of both N and T), but with the additional feature to accommodate nonlinearities in the factor loadings, as we described above.

3.3 Detecting Structural Breaks in Factor Models: A Model Selection Approach

Is the Factor Model Linear or Nonlinear? A decision rule can be offered that enables one to tell whether the model is linear or nonlinear.

[1]See Bai and Ng (2002) theorem 2.

For the linear model, Bai and Ng (2002) explore various data generating processes (DGP) to show that their information criteria consistently estimate the true number of factors under at only mild instabilities in the factor loadings.

From now on, let the linear information criterion from Bai and Ng (2002) be their $IC_{p2}(k_L)$,

$$IC_{p2}(k_L) = \ln|\widehat{\Sigma}(k_L, \widehat{F}^{k_L})| + k_L\left(\frac{N+T}{NT}\right)\ln\left(\min\left(T, N\right)\right) \tag{3.7}$$

where $|\widehat{\Sigma}(k_L, \widehat{F}^{k_L})|$ is the average residual variance when k_L factors are estimated using the linear information criterion, and $\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right)$ is the deterministic penalty term of this criterion.

The Decision Rule As stated above, the linear information criteria will overestimate the number of factors when a structural break takes place, while the nonlinear information criterion consistently estimates the correct number of factors under structural breaks in the factor loadings. With this in mind, the decision rule compares the number of factors estimated by the nonlinear $IC(\ell, k_{NL})$ and the linear information criteria $IC_{p2}(k_L)$.

The practical implementation of the proposed decision rule follows a three-step guideline. In the first step, one should estimate the number of factors according to Bai and Ng (2002), assuming linearity to obtain $\widehat{k}_L$. Hence, for each $k_L = 0, ..., k_{Lmax}$, we obtain $IC_{p2}(k_L)$ as in Bai and Ng (2002), to select the k_L that leads to the smallest $IC_{p2}(k_L)$, for example $\widehat{k}_L$. Using this $\widehat{k}_L$ the magnitude of $IC_{p2}(\widehat{k}_L)$ is stored.

In the second step, one should estimate the number of factors assuming a structural break in the factor loadings, using our information criterion to obtain $\widehat{k}_{NL}$. Analogously, each $k_{NL} = 0, ..., k_{Lmax}$ and each $\ell = \ell_1, ..., \ell_2$ are used to obtain $IC(\ell, k_{NL})$, and to choose the ℓ and k_{NL} that jointly minimize $IC(\ell, k_{NL})$. That is, $(\widehat{\ell}, \widehat{k}_{NL}) = \arg\min_{\ell, \widehat{k}_{NL}} IC(\ell, k_{NL})$ will be found, feeding $(\widehat{\ell}, \widehat{k}_{NL})$ into $IC(\ell, k_{NL})$ we obtain a numerical value for $IC(\widehat{\ell}, \widehat{k}_{NL})$.

Finally, in the third step one should compare the estimates and observe whether $IC_{p2}(k_L) < IC(\ell, k_{NL})$ to determine whether we have a linear model; if not, the model is nonlinear. It can also be put as follows:

$$\text{Decision Rule}\begin{cases} \widehat{k}_{NL} < \widehat{k}_L & \text{the model is nonlinear} \\ \widehat{k}_{NL} = \widehat{k}_L & \text{the model is linear} \end{cases}$$

Expressed in words, if the number of factors estimated with the nonlinear criterion is smaller than the linear one, the model is nonlinear; otherwise, the model is linear. A possibility not considered in the decision rule is the case of $\widehat{k}_L < \widehat{k}_{NL}$, which never occurred in the experiments we performed. We explored various DGPs from Bai and Ng (2002), Breitung and Eickmeier (2011), Chen et al. (2012) or Han and Inoue (2012) and on none of the occasions did this possibility occur.

Then one should note the relationship of the decision rule with the linear and nonlinear information criteria. The decision rule is solely based on the comparison of the estimated number of factors of the linear and the nonlinear information criteria. Since the linear information criterion is designed for linear frameworks only, it will overestimate the number of factors in structural break scenarios. Meanwhile, the nonlinear information criterion will not. Hence, the nonlinear information criterion yields a more parsimonious factor models. This is the key information in our decision rule.

3.4 Monte Carlo Simulations for Consistency of the Results

The following tables show by simulations that not only is it possible to consistently estimate the true number of factors under structural breaks in the factor loadings, but also to detect whether the factor model is linear or nonlinear. The DGPs are inspired by those constructed in Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012).

The primary goal of these papers is to detect structural breaks in factor models. However, they all depend on the choice of the number of factors which are estimated according to the information criteria of Bai and Ng (2002). The Breitung and Eickmeier (2011) method faces difficulties detecting correctly the occurrence of a structural break in occasions when the factor space estimated by the aforementioned information criteria because a factor model with a structural break can be written as a linear model with an additional factor. On the other hand, the tests proposed by Chen et al. (2012) and Han and Inoue (2012) perform better in circumstances when the number of factors is overestimated. In their framework they apply different number of factors with their methods before diagnosing the possibility of a structural break.

We show in simulations below that our information criterion consistently estimates the correct number of factors regardless the true model either linear or nonlinear. Additionally, we are also able to tell whether the model is linear and nonlinear. That is, our decision rule systematically selects the linear (nonlinear) specification when the true model is linear (nonlinear). It is also important to emphasize that our method does not require an *ex-ante* estimation of number of factors since it is itself a method to estimate the number of factors.

For each DGP below we offer two sets of results. The first one presents the number of estimated factors using $IC_{p2}(k_L)$ and $IC(\ell, k_{NL})$. The entries in each respective tables are the averages of the estimated number of factors on 1000 simulations, using these information criteria. With this output one is able to visualize the impact of the estimate of the number of factors as a consequence of structural breaks in the factor loadings, using $IC_{p2}(k_L)$ and $IC(\ell, k_{NL})$.

The second one presents the average frequency (in %) of the 1000 simulations that aims to identify whether the model is linear or nonlinear. Precisely, these entries yields the frequencies of nonlinear models detected in the 1000 simulations using our proposed decision rule. An entry of 0 (100) means that in 1000 simulations this decision rule has detected 0% (100%) nonlinear models.

DGP of Breitung and Eickmeier (2011) This analysis begins with a DGP in the fashion of Breitung and Eickmeier (2011), with results shown in Tables 3.1-3.3. The analysis of Tables 3.1 and 3.2 emphasizes that $\widehat{k}_L$ overestimates r_0, whilst $\widehat{k}_{NL}$ behaves nicely. Table 3.3 shows the performance of the decision rule.

In Table 3.1, when the structural break is $B = 0.5$ or larger, $IC_{p2}(k_L)$ clearly overestimates the number of factors. When the structural breaks are larger than this, almost all simulations overestimate the number of factors to two factors, when in fact there is only one factor in the model.

TABLE 3.1: Estimating the Number of factors

DGP3.1: $X_{it} = \Lambda_{i,1} F'_{t,1} I_1 + \Lambda_{i,2} F'_{t,1} I_2 + \epsilon_{it}$
$r_0 = 1$, B is the size of the breaks, and $\Lambda_{i,2} = \Lambda_{i,1} + B$

Number of factors using $IC_{p2}(k_L)$								
Dimensions		Size of the structural break						
N	T	$B = 0$	$B = 0.1$	$B = 0.3$	$B = 0.5$	$B = 0.7$	$B = 0.9$	$B = 1$
100	40	1	1	1	1.097	1.691	1.936	1.970
100	60	1	1	1	1.124	1.852	1.986	1.997
100	200	1	1	1	1.814	2	2	2
100	500	1	1	1	2	2	2	2
200	60	1	1	1	1.354	1.957	1.997	2
500	60	1	1	1	1.543	1.989	2	2
1000	60	1	1	1	1.599	1.993	2	2
40	100	1	1	1	1.066	1.806	1.988	1.997
60	100	1	1	1	1.118	1.905	2	2
200	100	1	1	1	1.698	1.999	2	2
500	100	1	1	1	1.937	2	2	2
60	200	1	1	1	1.348	1.997	2	2
60	500	1	1	1	1.784	2	2	2
60	1000	1	1	1	1.935	2	2	2
20	20	1.004	1.004	1.003	1.031	1.187	1.438	1.549
50	50	1	1	1	1.028	1.485	1.873	1.937
80	80	1	1	1	1.093	1.856	1.991	1.998
100	100	1	1	1	1.223	1.955	2	2
200	200	1	1	1	1.987	2	2	2

TABLE 3.2: Estimating the Number of factors

DGP3.1: $X_{it} = \Lambda_{i,1}F'_{t,1}I_1 + \Lambda_{i,2}F'_{t,1}I_2 + \epsilon_{it}$
$r_0 = 1$, $\Lambda_{i,2} = \Lambda_{i,1} + B$

Number of factors using $IC(\ell, k_{NL})$								
Dimensions		Size of the structural break						
N	T	$B = 0$	$B = 0.1$	$B = 0.3$	$B = 0.5$	$B = 0.7$	$B = 0.9$	$B = 1$
100	40	1	1	1	1	1	1	1
100	60	1	1	1	1	1	1	1
100	200	1	1	1	1	1	1	1.011
100	500	1	1	1	1	1	1	1.003
200	60	1	1	1	1	1	1	1
500	60	1	1	1	1	1	1	1
1000	60	1	1	1	1	1	1	1
40	100	1	1	1	1	1	1	1.002
60	100	1	1	1	1	1	1	1.003
200	100	1	1	1	1	1	1	1
500	100	1	1	1	1	1	1	1
60	200	1	1	1	1	1	1	1
60	500	1	1	1	1	1	1	1
60	1000		1	1	1	1	1	1
20	20	1	1	1	1	1	1.004	1.006
50	50	1	1	1	1	1	1	1
80	80	1	1	1	1	1	1	1
100	100	1	1	1	1	1	1	1
200	200	1	1	1	1	1	1	1.006

In Table 3.2 the entries have the same meaning but now under $IC(\ell, k_{NL})$ to select the number of factors. The results enable us to correctly determine the number of factors for any size of structural breaks in the factor loadings.

Although the results in Table 3.2 estimate the true number of factors consistently, whether the model is linear or nonlinear in the sense of being subjected to structural breaks in the factor loadings. Using the decision rule, 1000 simulations were run to detect the decision frequencies of a model with a structural break.

Table 3.3: Detecting whether the model is linear or nonlinear

$$\text{DGP3.1: } X_{it} = \Lambda_{i,1}F'_{t,1}I_1 + \Lambda_{i,2}F'_{t,1}I_2 + \epsilon_{it}$$
$$r_0 = 1,\ \Lambda_{i,2} = \Lambda_{i,1} + B$$

Decision Frequencies of a model with a structural break								
Dimensions		Size of the structural break						
N	T	$B=0$	$B=0.1$	$B=0.3$	$B=0.5$	$B=0.7$	$B=0.9$	$B=1$
100	40	0	0	0	9.7	69.1	93.6	97.0
100	60	0	0	0	12.4	85.2	98.6	99.5
100	200	0	0	0	81.4	100	100	98.9
100	500	0	0	0	100	100	100	99.7
200	60	0	0	0	35.4	95.7	99.7	100
500	60	0	0	0	54.3	98.9	100	100
1000	60	0	0	0	59.9	99.3	100	100
40	100	0	0	0	6.6	80.6	98.8	99.7
60	100	0	0	0	11.8	90.5	100	100
200	100	0	0	0	69.8	99.9	100	100
500	100	0	0	0	93.7	100	100	100
60	200	0	0	0	34.8	99.7	100	100
60	500	0	0	0	78.4	100	100	100
60	1000	0	0	0	93.5	100	100	100
20	20	0.4	0.4	0.3	3.1	18.7	43.8	54.9
50	50	0	0	0	2.8	48.5	87.3	93.7
80	80	0	0	0	9.3	85.6	99.1	99.8
100	100	0	0	0	22.3	95.5	100	100
200	200	0	0	0	98.7	100	100	99.4

The results in Table 3.3 show that, as the sizes of the structural break increase, the decision frequencies which indicate the model is subjected to a structural break also increase. Comparing Table 3.3 with the results in Breitung and Eickmeier (2011)[2] it can be seen that the frequency rates at which the decision rule is detecting nonlinearities is satisfactorily high.

[2]See their Table 3 and also note that in their test they impose strong operating assumptions on the idiosyncratic error term of the factor model, assuming cross-sectional independence

DGPs of Chen et al. (2012) We now deal with two DGPs related to Chen et al. (2012). The first one is a linear factor model, containing no breaks. The specification is a three-factor model described below.

$$\begin{aligned} \text{DGP3.2}: X_{it} &= \sum_{j=1}^{r} \lambda_i F_{tj} + \epsilon_{it} \\ F_{tj} &= \beta_j F_{(t-1)j} + \upsilon_{tj} \\ &\text{where } r_0 = 3,\ \beta_1 = 0.8,\ \beta_2 = 0.5,\ \beta_3 = 0.2; \\ &\text{with } \lambda_i,\ \epsilon_{it} \text{ and } \upsilon_{tj} \text{ i.i.d. standard normal variables.} \end{aligned}$$

In all simulations of the DGP3.2, both the nonlinear information criterion, $IC(\ell, k_{NL})$, and the linear criterion $IC_{p2}(k_L)$ detect three factors. With respect to the decision rule, it tells with 100% precision that the factor model is linear, which means the rule is identifying the specification correctly. It is important to hihglight that when the true model is linear, nonlinear information criterion is also able to detect the correct number of factors, performing equally well with respect to the linear information criterion.

The other specification is nonlinear, and is inspired by Chen et al. (2012). It is designed in DGP3.3, with results shown in Tables 3.4 and 3.5. It considers two sizes of structural breaks, $B = 0.2$ and $B = 0.4$, respectively. For the smaller break, $B = 0.2$, both $IC_{p2}(k_L)$ and $IC(\ell, k_{NL})$ achieve a relatively good accuracy in terms of estimating the correct number of factors. In this case it was learned that $B = 0.2$ is mildly unstable since it does not affect the estimation of $IC_{p2}(k_L)$.

For $B = 0.4$ with results shown in Table 3.4, the performances of $IC_{p2}(k_L)$ and $IC(\ell, k_{NL})$ clearly differ. For a break of this size, the number of factors is overestimated in all cases with $IC_{p2}(k_L)$. As the dataset dimensions increase, the estimate moves to three factors (the true number of factors is two). On the other hand $IC(\ell, k_{NL})$ gives a precise estimate of the true number of factors.

In Table 3.5, the performance of the decision rule is verified considering both sizes of structural breaks. For the smaller one, $B = 0.2$, the frequencies of break detections are very small. This corroborates the results shown in Table 3.4. However, no frequency pattern of break detections, for this small structural break could be indentified. Intuitively it can be said that since the structural break is small, the decision rule becomes puzzled as the dataset dimension increases, without suggesting that the model is nonlinear.

Analyzing the case of $B = 0.4$, the frequencies of break detection increase as the dataset dimensions increase (either N or T). For larger dataset dimensions (for example, a

TABLE 3.4: Estimating the Number of factors

DGP3.3: $X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \epsilon_{it}$ for $t = 1, ..., T/2$
$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} + B) F_{tj} + \epsilon_{it}$ for $t = (T/2 + 1), ..., T$
$F_{tj} = \beta_j F_{(t-1)j} + \upsilon_{tj}$
where $r_0 = 2$, $\beta_1 = 0.8$, $\beta_2 = 0.2$, $B = 0.2$ and $B = 0.4$;
with λ_i, ϵ_{it} and υ_{tj} i.i.d. standard normal variables.

		Number of factors using			
		$IC_{p2}(k_L)$		$IC(\ell, k_{NL})$	
Dimensions		Size of the structural break			
N	T	$B = 0.2$	$B = 0.4$	$B = 0.2$	$B = 0.4$
100	40	2.001	2.673	1.994	1.996
100	60	2.001	2.852	2	2
100	200	2.002	2.999	2	2
100	500	2.009	3	2	2
200	60	2	2.941	2	2
500	60	2.001	2.989	2	2
1000	60	2.002	2.992	2	2
40	100	2.004	2.848	2	2
60	100	2	2.922	2	2
200	100	2.003	2.997	2	2
500	100	2.023	3	2	2
60	200	2	2.998	2	2
60	500	2	3	2	2
60	1000	2	3	2	2
20	20	2.013	2.256	2.021	2.075
50	50	2	2.506	1.996	1.997
80	80	2	2.858	2	2
100	100	2	2.970	2	2
200	200	2.01	3	2	2

minimum of $N = 80$ and $T = 80$), the frequencies of break detection are relatively large, indicating that the model is nonlinear in at least 80% of the simulations.

In Chen et al. (2012), the frequencies of rejecting the null hypothesis of no break are also high (see their Table 2); but, the tests have to be implemented using different numbers of factors, and this is an extra decision to the practitioner.

DGPs of Han and Inoue (2012) Tables 3.6-3.9 show the results of $IC_{p2}(k_L)$ and $IC(\ell, k_{NL})$, and of the decision rule, but with two DGPs taken from Han and Inoue (2012).

DGP3.4 in Tables 3.6-3.7 analyzes four different sizes of structural breaks. In the factor models of this DGP the variance of the idiosyncratic error term is different than one, and is chosen such that $R^2 = \operatorname{tr} E(\epsilon\epsilon') / \operatorname{tr} E(XX')$ is 50%, including a relatively larger structural break.

TABLE 3.5: Detecting whether the model is linear or nonlinear

DGP3.3: $X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \epsilon_{it}$ for $t = 1, ..., T/2$
$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} + B) F_{tj} + \epsilon_{it}$ for $t = (T/2+1), ..., T$
$F_{tj} = \beta_j F_{(t-1)j} + \upsilon_{tj}$
where $r_0 = 2$, $\beta_1 = 0.8$, $\beta_2 = 0.2$, $B = 0.2$ and $B = 0.4$;
with λ_i, ϵ_{it} and υ_{tj} i.i.d. standard normal variables.

Decision Frequencies of a model with a structural break			
Dimensions		Size of the structural break	
N	T	$B = 0.2$	$B = 0.4$
100	40	0.7	67.4
100	60	0.1	85.2
100	200	0.2	99.9
100	500	0.9	100
200	60	0	94.1
500	60	0.1	98.9
1000	60	0.2	99.2
40	100	0.4	84.8
60	100	0	92.2
200	100	0.3	99.7
500	100	2.3	100
60	200	0	99.8
60	500	0	100
60	1000	0	100
20	20	15.6	29.9
50	50	0.4	50.7
80	80	0	85.8
100	100	0	97
200	200	1	100

In all four sizes of structural breaks the linear criterion $IC_{p2}(k_L)$ overestimates the number of factors. It gives four factors in most of the cases. However, our $IC(\ell, k_{NL})$ shows consistency. Its consistency can be seen as it gets closer to three factors (the true number of factors) as the increases with dataset dimension.

In Table 3.6, for the case of a relatively larger structural break ($B = 2$) the $IC_{p2}(k_L)$ still selects the number of factors incorrectly. The $IC(\ell, k_{NL})$ achieves the more accurate results in the sense of getting close to the true number of factors as the time dimension increases. For cases of $T \leq 80$, the $IC(\ell, k_{NL})$ slightly underestimates the number of factors. For the other cases, the results become again become quite consistent, but never achieve full precision as in structural breaks of other sizes.

The decision rule with the DGP3.4, with results shown in Table 3.7, is able to determine whether the model is linear or nonlinear with high frequencies in all cases. As the size of the structural break increases, the precision of our decision rule also increase, getting very close to 100% in most cases. Nevertheless, we slightly lose some precision for the

Table 3.6: Estimating the Number of factors

DGP3.4: $X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \kappa\epsilon_{it}$ for $t = 1, ..., T/2$
$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} - B) F_{tj} + \kappa\epsilon_{it}$ for $t = (T/2+1), ..., T$
$F_{tj} \sim$ *i.i.d.* $N(0,1)$, $\kappa = \sqrt{(1+b^2/4)r_0}$, $r_0 = 3$
with b an element of matrix B
$\lambda_{i,1} \sim$ *i.i.d.* $N(b/2,1)$, $B = \{1/3,\ 2/3,\ 1,\ 2\}$

		Number of factors using							
		$IC_{p2}(k_L)$				$IC(\ell, k_{NL})$			
Dimensions		Size of the structural break							
N	T	$B=1/3$	$B=2/3$	$B=1$	$B=2$	$B=1/3$	$B=2/3$	$B=1$	$B=2$
100	40	3.685	3.998	3.998	3.974	2.966	2.903	2.717	1.227
100	60	3.999	3.999	4	3.974	2.999	2.999	2.987	1.890
100	200	4	4	4	4	3	3	3	3.004
100	500	4	4	4	4	3	3.001	3.001	3.006
200	60	3.946	4	4	3.997	3	2.997	2.984	1.750
500	60	3.979	4	4	4	3	3	2.985	1.640
1000	60	3.983	4	4	4	3	2.998	2.990	1.620
40	100	3.809	4	3.999	3.934	3	3	3.007	2.716
60	100	3.872	4	4	3.991	3	3	3.005	2.931
200	100	3.997	4	4	4	3	3	3	2.904
500	100	3.999	4	4	4	3	3	3	2.887
60	200	3.996	4	4	4	3	3	2.984	2.983
60	500	4	4	4	4	3	3	3	2.994
60	1000	4	4	4	4	3	3	3	2.994
20	20	3.262	3.549	3.375	2.263	3.999	3.685	3.313	2.050
50	50	3.450	3.995	3.992	3.537	2.994	2.986	2.920	1.701
80	80	3.799	4	4	3.985	3	3	2.999	2.622
100	100	3.944	4	4	3.999	3	3	3	2.913
200	200	4	4	4	4	3	3	3	3.007

relatively larger break ($B = 2$), but the decision rule is still able to disentangle a linear from a nonlinear model.

In DGP3.5 the size of the structural break is fixed to $B = 1$. We pay attention to the behavior of the information criteria and of the decision rule, when the proportion of variables with a structural break increases.

We compare the information criteria in Table 3.8. In the case of $IC_{p2}(k_L)$, the overestimation of the number of factors becomes clear as the proportion of variables with a structural break increases. That is, when α increases. But as before, the overestimation is also clear as the dataset dimensions increases. Contrarily, our $IC(\ell, k_{NL})$ does is not negatively affected as α increases. Furthermore, as the dataset dimensions increase the precision of selecting the correct number of factors becomes fully accurate, regardless of the proportion of variables with a structural break.

Table 3.7: Detecting whether the model is linear or nonlinear

DGP3.4: $X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \kappa\epsilon_{it}$ for $t = 1, ..., T/2$
$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} - B) F_{tj} + \kappa\epsilon_{it}$ for $t = (T/2+1), ..., T$
$F_{tj} \sim$ *i.i.d.* $N(0,1)$, $\kappa = \sqrt{(1+b^2/4)r}$, $r = 3$
with b an element of matrix B
$\lambda_{i,1} \sim$ *i.i.d.* $N(b/2,1)$, $B = \{1/3,\ 2/3,\ 1,\ 2\}$

Decision Frequencies of a model with a structural break					
Dimensions		Size of the structural break			
N	T	$B = 1/3$	$B = 2/3$	$B = 1$	$B = 2$
100	40	70.4	100	100	100
100	60	99.9	99.9	100	99.9
100	200	100	100	100	99.6
100	500	100	100	100	99.4
200	60	94.6	100	100	100
500	60	97.9	100	100	100
1000	60	98.3	100	100	100
40	100	80.8	99.9	99.2	97.9
60	100	87.2	100	99.5	99.3
200	100	99.7	100	100	100
500	100	99.9	100	100	100
60	200	99.6	100	100	100
60	500	100	100	100	100
60	1000	100	100	100	100
20	20	16	31.9	42	46.6
50	50	45.4	99.5	100	97.1
80	80	79.9	100	100	99.6
100	100	94.4	100	100	100
200	200	100	100	100	99.3

Table 3.8: Estimating the Number of factors

DGP3.5: $X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \kappa \epsilon_{it}$ for $t = 1, ..., T/2$
$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} - M) F_{tj} + \kappa \epsilon_{it}$ for $t = (T/2 + 1), ..., T$
$F_{tj} \sim$ *i.i.d.* $N(0,1)$, $\kappa = \sqrt{(1 + b^2/4)r}$, $r = 3$
$\lambda_{i,1} \sim$ *i.i.d.* $N(b/2, 1)$, $B = 1$, $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$
M an $N \times \widehat{k}_{\ell}$ with its first proportion α of rows equals b,
with b an element of matrix B
and the elements of the other rows equals zero

		Number of factors using							
		$IC_{p2}(k_L)$				$IC(\ell, k_{NL})$			
Dimensions		α is the fraction of N variables with a structural break of size $B = 1$							
N	T	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
100	40	3.582	3.977	3.996	3.9980	2.742	2.747	2.742	2.746
100	60	3.693	3.998	3.999	4	2.988	2.987	2.988	2.987
100	200	3.999	4	4	4	3	3	3	3
100	500	4	4	4	4	3	3	3	3
200	60	3.918	3.999	4	4	2.985	2.987	2.988	2.986
500	60	3.970	4	4	4	2.985	2.985	2.985	2.986
1000	60	3.981	4	4	4	2.990	2.989	3	2.990
40	100	3.551	3.995	3.999	3.999	2.999	2.999	2.999	3.001
60	100	3.706	4	4	4	3	3	3	3
200	100	3.993	4	4	4	3	3	3	3
500	100	3.999	4	4	4	3	3	3	3
60	200	3.953	4	4	4	3	3	3	3
60	500	3.992	4	4	4	3	3	3	3
60	1000	4	4	4	4	3	3	3	3
20	20	2.716	3.069	3.317	3.417	3.159	3.2	3.239	3.281
50	50	3.279	3.920	3.989	3.991	2.917	2.920	2.922	2.926
80	80	3.649	3.999	4	4	3	3	3	3
100	100	3.878	4	4	4	3	3	3	3
200	200	4	4	4	4	3	3	3	3

In terms of the decision rule, it follows the expected intuition: as more variables contain a structural break and/or the dataset dimensions increase, the decision rule exhibits consistency and it is fully correct in most of the cases.

Table 3.9: Detecting whether the model is linear or nonlinear

DGP3.5: $X_{it} = \sum_{j=1}^{r} \lambda_{i,1} F_{tj} + \kappa\epsilon_{it}$ for $t = 1, ..., T/2$
$X_{it} = \sum_{j=1}^{r} (\lambda_{i,1} - M) F_{tj} + \kappa\epsilon_{it}$ for $t = (T/2+1), ..., T$
$F_{tj} \sim$ *i.i.d.* $N(0,1)$, $\kappa = \sqrt{(1+b^2/4)r}$, $r = 3$
$\lambda_{i,1} \sim$ *i.i.d.* $N(b/2,1)$, $B = 1$, $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$
with b an element of matrix B
M an $N \times \widehat{k}_\ell$ with its first proportion α of rows equals b,
and the elements of the other rows equals zero

Decision Frequencies of a model with a structural break					
Dimensions		α is the fraction of N variables with a structural break of size $B = 1$			
N	T	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
100	40	71	99.1	100	100
100	60	69.8	99.9	99.9	100
100	200	99.9	100	100	100
100	500	100	100	100	100
200	60	92	99.9	100	100
500	60	97.2	100	100	100
1000	60	98.1	100	100	100
40	100	55.2	99.6	100	99.8
60	100	70.6	100	100	100
200	100	99.3	100	100	100
500	100	99.9	100	100	100
60	200	95.3	100	100	100
60	500	99.2	100	100	100
60	1000	100	100	100	100
20	20	23.1	33.1	38.6	42.6
50	50	34.3	93.4	99.7	99.8
80	80	64.9	99.9	100	100
100	100	87.8	100	100	100
200	200	100	100	100	100

3.5 Conclusion

This chapter offers two tools to investigate nonlinearities in factor models. The first one is a nonlinear information criterion that enables one to determine the correct number of factors in either a linear or a nonlinear factor model. The nonlinearities take place in the factor loadings only, without changes in the number of factors. The second tool is understood as a decision rule to decide whether a factor model is in fact linear or nonlinear. Perhaps our contributions can complement to the Bai and Ng (2002) information criteria, which is designed only for linear datasets, and also to Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012), who model tools to detect structural breaks in factor models. A topic for parallel research is the case that occurs when the number of factors change at some point in the time dimension. For simplicity it has not been considered in this chapter.

Chapter 4

Forecasting Brazilian Unemployment Rates with Diffusion Indexes

4.1 Introduction

An accurate prediction of the unemployment rate is of crucial importance for economists, policy specialists and the wider business community. The problem is of particular interest for the Brazilian case, an emerging economy that ranks amongst the ten largest GDPs in the world, whilst at the same time displays particularly high unemployment rates. Over the past two decades the Brazilian economy has experienced double-digits unemployment rates combined with high inflation rate and often hyperinflation following Cagan's concept.[1] For instance, between 1979 and 1994 the Brazilian economy underwent thirteen stabilization plans devoted mainly to overcome the problem of high rates of inflation. All of those plans were unsuccessful with the exception of the *Real Plan*[2] implemented in July 1994. The latter was successful in bringing inflation rates to reasonable levels and since then the Brazilian monetary policy has relied on a tough Taylor rule[3] in the sense of being heavily weighted on the inflation rate coefficient.

The Taylor rule of the *Real Plan* was first implemented with a currency board policy until January 1999; for the time being it has followed an inflation targeting policy. The behavior of the Taylor rule in the Brazilian economy can be well highlighted during

[1] Cagan (1956) considers a scenario of more than 50% inflation rate per month as hyperinflation and the Brazilian economy has faced it particularly in the late 80's and early 90's.

[2] See Salgado et al. (2005).

[3] See Taylor (1993) seminal paper. The use of a Taylor rule for the Brazilian economy has been recognized by Salgado et al. (2005) and Moura and de Carvalho (2010) amongst others.

the events of the financial crises fn Mexico 1994, Russia 1997, Southeast Asia 1999, Argentina 2001 and the Brazilian presidential electoral process in 2002. During all these events the Brazilian Central Bank increased the short term interest rates significantly, aiming primarily to avoid undesired high inflation rates. If from one hand the inflation rate has been under control, the unemployment rate has been high ever since. This said, the direct effect of an increasing of short terms interest rates on unemployment rates is a common sense, and not surprisingly, high unemployment rates in Brazil have been of interest to policy makers and researchers, and we are motivated from this context.

The objective of this chapter is to forecast unemployment rates using diffusion indexes (DI), as originally proposed by Stock and Watson (1998). By diffusion indexes it is meant to use the estimated unobserved factors in the forecast exercise. With results from both forecasts and factor analysis we additionally offer some intuition to the policy implication of the context. On the forecast side, the results tell us not only how useful our dataset (formed with labor market and/or macroeconomic and monetary related variables) is in the construction of DI forecasts, but also when they are useful in the sense of different forecast horizons, and in the use of different blocks (i.e., subsets of the dataset) to estimate DI. We explore various blocks to investigate their different usefulness in the forecast exercise. The point is that more data does not necessarily improve DI forecasts, as explained by Ng and Boivin (2006). Factors estimated with different blocks (possibly with different number of variables) may span a different factor space and consequently may underly a different comovement and have different usefulness to the forecaster. As seen in Stock and Watson (1998, 2002a,b, 2006), variables such as GDP (and other measures of output) or inflation rate have been largely used as the target variable. Presumably, these applications use a dataset that attempt to capture the behavior of the whole economy as it would be expected for instance to forecast GDP, meaning that a proper dataset should include all, if possible, of the sectors of the economy.

On the factor analysis side - despite the that fact that our central goal is forecast-oriented - these results are used to contrast the economic history of our context to uncover possible patterns of covariability of the dataset. The latter results of course do not imply any sort of causality since this is not regression-oriented either.

The novelty of our work are twofold. Firstly, it lies in the use of a rich dataset that includes labor market related historical time series across all major Brazilian metropolitan areas as well as aggregate monthly macroeconomic and monetary related variables to forecast Brazilian unemployment rate using factor models, a methodology not used so far for this particular context of that country. Secondly and also related, at the best of our knowledge there is no work devoted to forecast the Brazilian unemployment rate.

The main objective of the factor models is to allow a dataset to be explained by small number of unobserved common forces. In a time series framework, a factor model allows a matrix formed of N time series to be represented by two unobserved orthogonal components: a common and an idiosyncratic. In this way, the N time series are driven by a small number of common forces (say $k < N$ common factors) and idiosyncratic components. Applications of factor models have motivated economists in various ways (e.g., arbitrage pricing theory, building economic indicators, monetary policy analysis or forecasting economic variables).[4]

Furthermore, factor models estimation and inference approaches can be distinguished into two groups: classical factor models (CFM) and modern factor models (MFM).[5] CFM is well documented in the statistics literature, such as in Anderson (2003), whereas MFM can be exemplified for instance with the approximate factor model initiated by Chamberlain and Rothschild (1983). In line with Chamberlain and Rothschild (1983), we make use of the approximate dynamic factor model via asymptotic principal components (APC) written in the static version. We will refer henceforth to the approximate static factor model (ASFM), which is related to the Stock and Watson (1998, 2002a,b, 2006) and Bai and Ng (2002, 2010) and Bai (2003) estimation and inference procedures. The limitations of CFM to accommodate economic data characteristics motivated the MFM (hereafter ASFM). This is briefly explained in section 2.

We forecast unemployment rates in a three-step procedure. First, we estimate the factor model in order to extract the DI. Secondly, we select the number of factors via information criteria of Bai and Ng (2002). Finally, in the third step, we forecast unemployment rates, constructing a DI forecast. We have Brazilian unemployment rate as our target variable (the variable to be forecast), say y^r_{t+h} (where h is the forecast horizon and the superscripts refer to the time dependence of the projection), as a function of the estimated factors and possibly of the lagged values of the target variable. We promote competition of forecast accuracy amongst the estimated DI forecasts against a benchmark. Our estimated choice of benchmark is an AR(4) with a constant, chosen via Box and Jenkins type of model selection for a univariate process. The accuracy is measured through a symmetric loss function, a mean squared error (MSE) of each forecast, to be compared with the MSE of the benchmark. Furthermore, statistical significance evaluation and comparison of the MSEs are made via the celebrated Diebold and Mariano (2002) method. Yet, we verify the possibility of forecast encompassing and forecast combination as suggested by Diebold and Lopez (1996) and Timmermann (2006), and references therein. In the case of forecast combination we also test statistical significance

[4]A review of applications and theoretical results can be seen in Bai and Ng (2008c), Stock and Watson (2010), Reichlin (2003) and Breitung and Eickmeier (2006).

[5]Following Bai (2008) terminology.

evaluation and comparison of the MSEs against the benchmark. More precisely, forecast encompassing is made via Chong and Hendry (1986) and forecast combination is made in three methods: simple averaging, variance-covariance as in Bates and Granger (1969), and regression-oriented as in Granger and Ramanathan (1984).

In the forecasting exercise we have also divided our dataset into six blocks. These blocks are divided accordingly to the characteristic of the variables. In this way we have also promoted competition within factor-based forecasts that contain factors estimated from blocks that do not have information in common (not a single time series in common). This procedure allows some factor-based forecasts to have two factors estimated independently. In this regard we have estimated factors that underly different comovements to investigate the usefulness of block estimates to better explore the dataset.

Policy implications are twofold. Firstly, from factor analysis results (i.e., the ranking of commonalities which accounts for the covariability of the variables in the constructions of the factors), comparing the commonalities of labor market variables with macroeconomic and monetary related variables suggests that in general those factors estimated do not follow the same patter since they are well defined in the way they are weighted by their commonalities. We are dealing with the factors resulted from the model selection based information criteria. Checking the ranking of the commonalities of the estimated factors encompassing the entire dataset, labor market variables dominate the commonality ranking. Even when we include six lags of interest rates the results do not change much. Therefore, in terms of covariability captured by the factor analysis we could say that labor market, and macroeconomic and monetary related variables do not follow a similar patterns (or have different commovement). Further, we calculate the correlation coefficient matrix and a discrepancy statistics[6] of the space spanned amongst the estimated factors across the different blocks. The results from these exercises corroborate with the intuition from the ranking of commonalities in the sense that the factors with similar leading commonalities display higher correlation and less discrepancy (and vice-versa, specially, as one would expect, in those cases that factors are estimated with dataset that do not have information in common).

Secondly, from the forecasting exercises, policy implications are such that factor-based forecasts improve when forecast combination with the benchmark is used, in the following cases: two-month-ahead forecast using the factors estimated with a block of macroeconomic and monetary related variables; four-month-ahead forecast using both the factors estimated with a block of macroeconomic and related variables and the factors estimated with a block unemployment rate related variables; and six-month-ahead forecast for all

[6]We calculated it similar to Ng and Boivin (2006). These authors, however, use this type of statistics to verify discrepancy between simulated true model variables with their corresponding estimates.

estimated factors. Following Diebold and Mariano (2002) evaluation method, overall our findings suggest that the factor-based forecasts with forecast combination produce a substantial improvement with respect to the benchmark model, particularly for two-, four- and six-month-ahead forecasts (we have forecasts for one-, two-, four-, six-, eight- and twelve-month-ahead).

The remaining of the chapter is organized as follows. In section 4.2 we discuss the methodological approach used: factor analysis, information criteria, DI forecasting and forecasting accuracy. In section 4.3 the these results are presented. Finally, section 4.4 concludes. 4.2

4.2 Forecasting Methodology

Our forecasting methodology embraces some modeling techniques that we explain below in the following order: DI forecast, the factor model, estimation of the factor model, information criteria to choose the number of factors, choice of a benchmark model, out-of-sample loss function and forecast accuracy, and finally, forecast encompassing and combination.

DI Forecast We forecast using unobserved estimated factors and lagged values of the variable to be forecast as explanatory variables. The model can be put as follows:

$$y^r_{t+h} = \alpha + \beta'_F F_t + \beta'_w w_t + \varepsilon^r_{t+h} \tag{4.1}$$

where $t = 1, ..., r, ..., T$ is the time series dimension; F_t are the common factors to be estimated (the number of factors to be chosen via information criteria); h is the forecast horizon, whilst the superscript r refer to the dependence of the projection (forecast at time r); w_t is a vector of useful observed variables that helps in the forecasts of y_{t+h} (e.g., lags of y); and, ε_{t+h} is the forecast error.

We forecast estimating equation (4.1) recursively. We use data from $1 + h, ..., r$ for the target variable, and $1, ..., r - h$ of the explanatory variables. The model is estimated recursively up to r, where the forecast of horizon h is made; and then the target variable is estimated again up to $1, ..., r + 1, r + 2, ..., T$, with the explanatory variables being estimated up to $r - h + 1, r - h + 2, ..., T - h$, where the forecasts are made, with respect to the dependence of the projection. For instance, let $(T = r + p)$ be our entire sample. Then we keep reestimating the model using dataset from r to $T - h$ to forecast h steps ahead; we end up with $(p-h+1)$ h-step-ahead forecasts of $\hat{y}$, and $(p-h+1)$ forecast errors (say, $e^r_{t+h} = y^r_{t+h} - \hat{y}^r_{t+h}$).

The h-step-ahead forecast is a "point forecast" (point $r+h$), assuming a linear relationship in the forecast equation. Following Stock and Watson (2002a) the feasible forecasts can be estimated using the estimated factors and parameters, converging to the optimal infeasible forecast. Moreover, the forecast is well-behaved under OLS estimation, leading to asymptotically efficient factors (*Theorem 2* of Stock and Watson (2002a)).

The Factor Model A factor model can be viewed as follows:

$$x_{it} = \lambda_i^{'} F_t + e_{it} \tag{4.2}$$

where x_{it} are the observed variables described in a matrix with i cross sectional variables across t time series observations ($i = 1, ...N$; and $t = 1, ...T$); $\lambda_i^{'}$ is a vector $(r \times 1)$ of factor loadings; F_t is a vector $(r \times 1)$ of common factors; $\lambda_i^{'} F_t = C_{it}$ where C_{it} is the common component; and e_{it} is the idiosyncratic component of x_{it}. Both $\lambda_i^{'} F_t$ and e_{it} are not observable.

In matrix notation the factor model can be described as follows:

$$X_t = F_t \Lambda' + e_t \tag{4.3}$$

where $X_t = [X_1', X_2', ..., X_N']$ is a $T \times N$ matrix of observations; $\Lambda = [\lambda_1, \lambda_2, ..., \lambda_N]'$ is a $k \times N$ matrix of factor loadings; $F = [f_1, f_2, ..., f_T]$ is a $T \times k$ matrix of common factors, with k the number of factors; and $e_t = [e_1', e_2', ..., e_N']$ is a $T \times N$ matrix idiosyncratic components.

To accommodate economic times series we need to distinguish the estimation and inference approach of CFM and an ASFM. A CFM requires that N, the number of variables (or T, the number of time series observations) to be fixed and small. Moreover, CFM requires that: e_t and F_t to be i.i.d. random variables with zero means; e_t and F_t to be orthogonal, and also independent cross-sectionally; $E(f_t f_t') = I$, where I is an identity matrix (i.e., factors are orthogonal); and, the factor loadings are fixed constants Λ. These assumptions lead us to a covariance matrices of e_t and X_t to be represented respectively as $E(e_t e_t) = \Phi$ (where Φ is a diagonal matrix) and $E(X_t X_t') = \Lambda\Lambda' + \Phi = \Sigma$.

As explained in Bai and Ng (2002), CFM is not appropriate for economic applications. This is because of the following: (i) dataset dimensionality in economics should not be restricted to size dimension; (ii) we frequently have cross-sectional and time series dependence in the error structure of economic dataset; (iii) estimation of CFM is via maximum likelihood estimator which is not tractable when the number of parameters become very large; and finally, (iv) in economics we are mostly interested in the estimation of common factors that underlie the observed variables and this is not consistently estimated in the CFM.

We work with ASFM framework with time domain.[7] An advantage of the ASFM is that it requires only the choice of the number of factor as an auxiliary parameter, whereas a disadvantage is the lack of the ability to capture possible dynamics in the the factors (leads and lags of the factors). If we are to include such dynamics (i.e., a lagged factor) in the static models, it is viewed as the inclusion of another factor. As highlighted in Boivin and Ng (2005), the use of dynamics (in frequency domain) or static models (in time domain) is purely a decision of the practitioner. Nonetheless, these author point out arguments in favor of DI forecasting with the static model, due to the fact that it is more flexible to the dataset, as there are less parameters to estimate, and it is easier to implement.[8]

Results in Bai and Ng (2002) have shown the asymptotic results of the estimates: factor estimates converges in probability to its true value up to a rotation, the covariance matrix of the true factors also converges in probability; the non-random factors loadings also have be bounded and each factor one has a unique contribution to the variance of X_t (for random factor we have to add that $\lambda_i's$ have to be independent of the factors and of the idiosyncratic errors); the factors are not the same (non-degenerate) and each has its own distinguish contribution to the variance of X_t; the idiosyncratic component are allowed to be dependent serially and cross-sectionally (non-diagonal covariance matrix of the idiosyncratic terms) and heteroskedasticity can as well be found in these dimensions.[9]

In fact we deal with dynamic factors, such that F_t is dynamic in the sense that $A(L) F_t = u_t$, where $A(L)$ is a polynomial matrix of the lag operator. However, the model is not dynamic in the X_t side directly and hence the model is viewed as static with respect to the relationship between X_t and F_t. In this way the factors are carried out contemporaneously such that the factor loadings are real numbers. That is the reason that the

[7] For frequency domain see the generalized approximate dynamic factor model developed in Forni et al. (2000). By generalization it is meant that both N and T tend to infinity. The key characteristic of this model is that it allows the factors to be written with their leads and lags which can be an advantage for economic applications. A drawback is that it requires the need of the determination of more auxiliary parameters, in addition to the complication to determine the number of factors. Extensions of Forni et al. (2000) can be found in Forni et al. (2004) and Forni et al. (2005).

[8] Stock and Watson (1998, 2002a,b, 2006), Bai and Ng (2002, 2006, 2008a,b,c, 2010) and Bai (2003) describe the estimation, inference and possible applications (including forecast) of the ASFM.

[9] The approximate dynamic factor model of Chamberlain (1983) and Chamberlain and Rothschild (1983) allow for cross-section correlation of the idiosyncratic component; whereas the factor model of Geweke (1977) and Sims et al. (1977) have orthogonal idiosyncratic component.

model can be given as in (4.2) and a dynamics can be viewed as the inclusion of an additional factor. Finally, we should assume a balance panel for the dataset.

Estimation of the Factor Model The estimation consists of the optimizing the following function:

$$V(k) = \min_{\Lambda F^k} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(X_{it-} \lambda_i^{k'} F_t^k \right)^2 \tag{4.4}$$

where $V(k)$ is the sum of squared residuals, considering a k-factor model.

Information Criteria (IC) to Choose the Number of Factors We choose the number of factors via IC Bai and Ng (2002).

The main result of Bai and Ng (2002) is that the proposed IC consistently estimates the number of factors, for both the cross-section N and time series T dimensions. The advantage of their information criteria is that the penalty for overfitting is now a function of N and T simultaneously, differently from the standard IC such as the Akaike and the Schwartz which are either a function of N or T.[10] The information criteria are the following:

$$IC_{p1}(k) = \ln\left(V\left(k, \widehat{F}^k\right)\right) + k\left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right) \tag{4.5}$$

$$IC_{p2}(k) = \ln\left(V\left(k, \widehat{F}^k\right)\right) + k\left(\frac{N+T}{NT}\right) \ln C_{NT}^2 \tag{4.6}$$

$$IC_{p3}(k) = \ln\left(V\left(k, \widehat{F}^k\right)\right) + k\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right) \tag{4.7}$$

where, $C_{NT}^2 = \min(N, T)$.

Choice of benchmark model As explained in Stock (2001) one faces a trade-off between the choice of more sophisticated versus simpler models in a forecast exercise. By using more sophisticated models (say, with more parameters to estimate) one can reduce the estimated errors. On the other hand, one increases the estimation error of the parameters themselves, as having more parameters bring more uncertainty of those estimates. As highlighted by Stock (2001), competing models under these circumstances can be chosen via the use of a simple benchmark model, reliable enough to help model comparison.

[10]The information criteria of Bai and Ng (2002) which is now very popular have been applied for instance, by Stock and Watson (2006, 2008 and 2010) and Breitung and Eickmeier (2006), amongst many others.

We assume a linear time series AR model type as the benchmark, $y_t = c+\sum_{i=1}^{6} \phi_i y_{t-i}+\epsilon_t$, where $\phi_i, ..., \phi_6$ are the parameters of the model, c is a constant and ϵ_t the error term. We select the AR order via Box and Jenkins type of model selection for a univariate process - standard methods such as significance level of coefficients, Akaike and Bayesian information criteria (AIC and BIC), autocorrelation function (ACF) and partial autocorrelation function (PACF), and Breusch-Godfrey serial correlation LM test of residuals. Our attempt goes up to an order up to six in the AR model.

Out-of-Sample Loss Function and Forecast Accuracy The forecast of model (4.1) is made via OLS which makes it convenient to assume a squared loss function as a baseline to evaluate the performance of our forecasts. We first calculate the MSE given by $l = E(y^r_{t+h} - \widehat{y}^r_{t+h})^2/(p-h+1)$ of each of the competing models and compare with the benchmark. We then calculate a relative-MSE (Rel.MSE) as follows: Rel.MSE(factor-based) = MSE(factor-based)/MSE(AR(s)) where s is the order of the AR. A Rel.MSE greater than one implies that the AR(s) forecast is better than the Rel.MSE for that factor-based and vice-versa.

As expected whilst comparing MSEs (or equivalently Rel.MSEs), this exercise will always give us a 'better' model since it is very unlikely two forecast exercise have precisely the same MSE, and an ambiguous conclusion can be achieved. To overcome ambiguity we test the statistical significance of the MSEs to promote competition amongst forecasting models. We apply the evaluation method of Diebold and Mariano (2002).

Let m_1 be our model one to compete against model two, m_2; that is, y^{r,m_1}_{t+h} and y^{r,m_2}_{t+h} respectively. The estimation of these models lead us to two loss squared functions, $l^{m_1} = (y^{r,m_1}_{t+h} - \widehat{y}^{r,m_1}_{t+h})^2$ and $l^{m_2} = (y^{r,m_2}_{t+h} - \widehat{y}^{r,m_2}_{t+h})^2$, where $d = l^{m_1} - l^{m_2}$. We are interested in a linear regression model of $d_{t+1} = c + u_{t+1}$ where c is a constant. The hypothesis testing is $H_0 : c = 0$ against $H_1^{m_1} : c < 0$ or $H_1^{m_2} : c > 0$. In order to overcome the error heteroskedasticity and autocorrelation of the error structure we apply the Newey and West (1987) consistent (HAC) standard errors. If a calculated t-ratio (henceforth DM-statistics) is statistically significant and the constant is negative m_1 defeats m_2 or vice-versa. If the test is not statistically significant, nothing can be said about m_1 and m_2 defeating each other.

Forecast Encompassing and Combination If we cannot achieve a conclusion after using Diebold and Mariano (2002) evaluation method, we proceed with a pairwise encompassing following Diebold and Lopez (1996) and Timmermann (2006) and references therein. We run a regression of the observed dataset on the forecast of the competing models, m_1 and m_2, as in Chong and Hendry (1986), as follows:

$$y^r_{t+h} = \beta_{m_1}\widehat{y}^{r,m_1}_{t+h} + \beta_{m_2}\widehat{y}^{r,m_2}_{t+h} + \epsilon^r_{t+h} \tag{4.8}$$

The test is the interpretation of the significance of the coefficients. Whenever $(\beta_{m_1}, \beta_{m_2}) = (1, 0)$ model one encompasses model two. Contrarily, if $(\beta_{m_1}, \beta_{m_2}) = (0, 1)$ model two encompasses model one. Any other values of the coefficients we do not have forecast encompassing.

When forecast encompassing still uncertain, we do a forecast combination of the forecasts since none of the forecasting models are dominating each other. Forecast combination can be viewed as putting a weight ω and $(1-\omega)$ on models m_1 and m_2 respectively, in such a way that the resulting is a composite forecast, y^c_{t+h}, is the following:

$$y^c_{t+h} = \omega\widehat{y}^{r,m_1}_{t+h} + (1-\omega)\widehat{y}^{r,m_2}_{t+h} \tag{4.9}$$

An issue in forecast combination is to find a reliable method for computing the weights. We explore it in three ways. Firstly we carry out a simple averaging method giving equal weights. In our pair-wise case the weights are simply 0.5 and we refer this forecast combination to y^{c1}_{t+h}. Secondly, we follow the variance-covariance by Bates and Granger (1969), referring this forecast combination to y^{c2}_{t+h}, in which the weights are $(1/\sigma_{m_1})/[(1/\sigma_{m_1}) + (1/\sigma_{m_2})]$ for model one and $(1/\sigma_{m_2})/[(1/\sigma_{m_1}) + (1/\sigma_{m_2})]$ for model two, where $\sigma_{m_{1,2}}$ are the MSEs of models one and two, respectively. The third case, the forecast combination is referred to y^{c3}_{t+h}, following Granger and Ramanathan (1984). For this case we have a regression of observed dataset onto the forecasts m_1 and m_2. After doing these three forecast combination we use again Diebold and Mariano (2002) test for individual statistical comparison between y^{c1}_{t+h}, y^{c2}_{t+h} and y^{c3}_{t+h} against the benchmark to verify if any of the forecast combination has brought performance improvement.

4.3 Empirical Work

4.3.1 Dataset Specifications

Our dataset consists of 69 monthly time series (including interest rates lagged from one to to six months) ranging from 1998:01 to 2009:01, described in Tables 4.1 and 4.2 in the Appendix A.[11] The dataset is divided into six blocks, as follows: all 63 contemporaneous

[11] The usual Taylor rule uses one lag interest rates as a dependent variable of contemporaneous interest rate, amongst other dependent variables. Since many studies diverge in terms of what is the precise monetary mechanism transmission of the short term interest rates, and this is out of the scope of this chapter, we decided to include up to six lags of interest rates in the factor analysis exercise.

variables (Tables 4.1 and 4.2, excluding lags of interest rate) coded (FC); all dataset including the 63 contemporaneous dataset and 6 lags of nominal interest rate (Tables 4.1 and 4.2) having a total of 69 variables coded (FL); labour market dataset with 43 variables (Table 4.1) coded (LB); 35 variables (Table 4.1 - restricted to unemployment rates) coded (UN); only contemporaneous monetary variables, 20 variables (Table 4.2, excluding lags of interest rate) coded (MC); and the 20 contemporaneous monetary variables adding 6 lags of nominal interest rates, having a total of 26 variables (Table 4.2) coded (ML). Full description of the dataset treatment is in the Appendix A.

4.3.2 Forecasting Results

We report short-horizon results for h equals one-, two- and four-month-ahead forecasts and long-horizon for six-, eight- and twelve-month-ahead forecast. The dependent variables is the Brazilian total unemployment rate and it undergoes a logistic transformation (see Koop and Potter (1999)). The unemployment rate U_t is transformed to $Y_t = \ln\left(\frac{U_t}{1-U_t}\right)$. This transformation makes the bounded $[0, 1]$ time series of unemployment rates to be now unbounded. The point is that the error structure of a linear forecast regression with a bounded dependent variable may not result in bounded errors. However, with the transformed series we overcome this issue in the forecasting exercises. An issue Koop and Potter (1999) left unanswered is about imposing stationarity to the transformed series. In our case Y_t is integrated of order one, I(1), and we induce stationarity by taking first difference. Alternatively we could use unemployment growth Estimated factors are all considered I(0). The benchmark is an AR(4) with a constant. [12]

Our dataset ranges from 1998:01 to 2009:01, which provides 133 observations. With the dataset treatment we end up with 132 observations as we lose the first one to induce stationarity. In the forecasting model we first estimate using data until 2006:11 (estimation window). We are using approximately 20% of the full sample size to be generate our forecast (a well-accepted rule of thumb is about 10% to 15%). We then forecast unemployment rate using data raging from 2006:12 to 2009:01 (forecasting window). The number of forecasts is of the size $(p - h + 1)$, where $p = 26$. That is, the number of forecast values decrease as the horizon increases.

Results of forecasts are in Appendix D and Figures 4.1-4.3 in Appendix E. Table 4.6 displays the MSEs of the factor-based forecasts and of the benchmark AR(4); and the Rel.MSE is given by the former divided by the latter. Most of the results in these Tables show that the Rel.MSE are less than one, with six-month-ahead forecast with the

[12] Results are available upon request.

lowest Rel.MSE. If we were to take into consideration only Rel.MSE for our forecasting conclusion, we could already choose a 'winning' model already. However, as explained in section 2.1 we shall verify the statistic significance of the MSEs using the DM-statistics to compare competing forecasts.

In Table 4.7 we analyze the DM-statistics to compare the competing models. For six-month-ahead forecast, the factor-based forecasts containing factor estimated from blocks FL, FC, LB and UN have shown better forecast than the AR(4) benchmark. The factor-based forecasts with factors estimated from a dataset with macroeconomic and monetary related variables (MC and ML) do not show any improvement in any horizon. In addition to this, none of all of the other factor-based forecast are worst off than the AR(4) benchmark.

A question that arises is how the factor-based forecasts behave whilst competing amongst themselves. Looking at DM-statistics, with results in Table 4.8, all possible combinations of pair-wise competition of factor-based forecasts display least one winning factor-based forecast in each of the forecast horizons. As a result, none of the factor-based forecasts from factors estimated with a dataset with macroeconomic and monetary related variables (MC and ML) defeat any of the other factor-based forecasts (FL, FC, LB and UN). But as we will see later on all of the factor estimates have their different usefulness across different horizons, particularly whilst making use of forecast combination.

For now we proceed attempting a factor-based forecast using two estimated factors that contain blocks that do not have information in common to compete with the AR(4) benchmark with results in Table 4.9. We have the following factor-based forecast with two estimated factors: LB and ML, LB and MC, UN and ML, and finally, UN and MC. By doing this the results are not visually improved in terms of Rel.MSE with respect to Table 4.6 cases (that is, with only one factor included). But again, six-month-ahead forecast shows a better result. However, analyzing the DM-statistics for these forecasts in Table 4.10 better results are achieved. Despite the fact that only six-month-ahead forecasts are better with the inclusion of two factor in the forecast, the factors including blocks of macroeconomic and monetary related variables are now all useful in the forecast exercise. In contrast with Table 4.6 when the factor-based forecasts using the estimated factors MC and ML individually did not perform better than the AR(4) benchmark, now these factors become useful as we use each of them together with UN and these forecasts perform better than the AR(4) benchmark.

To further understand the results we verify forecast encompassing with all factor-based forecasts against the AR(4) benchmark. These results are described in Tables 4.11 and 4.12 where no forecast encompassing is seen. As explained in section 2.1 this outcome motivates forecast combination. We evaluate the performance of forecast combination in terms of the Rel.MSE and the DM-statistics test; the forecast combination is the factor-based forecast combined with the AR(4) benchmark, against the AR(4) benchmark.

Using a simple averaging (equal weights) forecast combination, results displayed in Table 4.13, there is substantial performance improvement in all the Rel.MSE (except for the forecast combinations using macroeconomic and related variables for eight-month-ahead in which the Rel. MSE are greater than one); in magnitude, the rest of the Rel. MSE are smaller than previous factor-based forecast. Applying DM-statistics test for this case, described in Table 4.14, results under forecast combination for forecast horizons of two, four an six-month-ahead are now defeating the AR(4) benchmark. Notably, for two-month-ahead forecast, MC and AR(4), and ML and AR(4), are now both defeating the AR(4) benchmark. For four-month-ahead forecast, ML and AR(4) is the winning model; whilst for six-month-ahead forecast, all factors-based forecast used in the forecast combination have shown to be statistically better than the benchmark AR(4).

Making used of variance-covariance weight method, with results displayed in Tables 4.15 and 4.16, we achieve quiet similar results as in the simple averaging case. However, now the forecast combination of UN and AR(4) appears to perform better than the AR(4) benchmark for four-month-ahead forecast. Additionally, we also have 'winners' in terms of DM-statistics test for the combinations of MC and AR(4), and of ML and AR(4) in the two-month-step ahead forecast of UN and AR(4), and of ML and AR(4) for the four-month-step ahead forecast and for all factor-based forecast used in the forecast combination for the six-month-step ahead forecast. Finally, the last forecast combination we attempt is the regression method. The results are in Tables 4.17 and 4.18. In terms of Rel.MSE. we cannot deny we have improved against the benchmark AR(4). Nevertheless, less improvement is achieved when compared to the the variance-covariance method for the DM-statistics test.

Overall we found that factor-based forecasts have improved substantially against the choice of an AR(4) benchmark forecast particularly for the two-, four- and six-month-ahead forecasts generated from the forecast combination using the variance-covariance method (Table 16) of Bates and Granger (1969). The 'winners' of Table 4.16 are displayed in levels against the historical dataset in Table 4.19 (Appendix D) and in Figures 4.1-4.3 (Appendix E).

4.3.3 Factor Analysis Results

Choice of Factors via IC Information Criteria of Bai and Ng (2002) indicate one factor for each of the blocks of dataset. The results are in the Appendix B, Table 4.3.[13] Comparing these results with the tools developed in the preceding chapters allow us to conclude this is a linear factor model, that is, the dataset does not contain relevant structural breaks. We ended up with one factor, hence, no overestimation of the number of factors in a recursive exercise (hence no jumps).

The Nature of the Factors (commonalities) In the Appendix C, Table 4.4 we show the corresponding covariability explained by each factor with respect to each block of dataset. For expositional purposes we show the covariability of the first ten factors. Obviously, blocks with fewer variables will display higher covariability of the dataset set explained by the first factors, because there are less factors. For instance, let us consider the block FL (69 variables), in which a factor model would originally calculate 69 factors. If each factor had the same importance they would individually explain 1.45% of the covariability of the dataset; however, the first factor for this block explains 10.10%.

To understand the nature of the factors we check the commonalities (the corresponding squared of each element of the factor loadings with respect to a variable - which tells the covariability of that variable explained by that factor).[14] The complete output for commonalities for the factors chosen via IC is in Table 4.5; since IC results accounts for one factor in each block of dataset, this is the the factor taken into consideration in Table 4.5.

Overall an analysis of FL and FC suggest that these factors are strongly leaded by labor market variables, whereas macroeconomics and monetary related variables appear to be less important, concentrated in the second half of the commonalities. Using only labor market variables, the factors LB and UN display results with similar intuition in the sense that both of their commonalities are well characterized with better ranked variables of metropolitan areas that are more important economically. For ML and MC, both are factors are leaded by contemporaneous interest rates (NIR). These results suggest the factors we estimated with different blocks are not describing the same comovement.

For instance, for the factor FL, the top five commonalities have either variables of the Brazilian aggregate unemployment and employment (TBR, PEBR and OBR), and about Sao Paulo unemployment and employment variables (TSP and POSP). These

[13]Codes taken from Professor Serena Ng's website and the results generated from Matlab 2009b.

[14]These values can also be achieved by the statistical R^2 of a linear regression of each individual variable against the respective factor.

results bring an intuitive interpretation as the unemployment and employment variables in aggregate terms for Brazil and Sao Paulo metropolitan area (Sao Paulo is the largest economy amongst the six metropolitan areas) are leading the ranking of commonalities of that factor that contains the entire dataset, including lags of interest rate.[15] Following the ranking of highest commonalities of FL, variables TBH, TPOA, HBR, OSP and TDF complete the top ten. What this is saying is that this factor is leaded by Brazilian aggregate and the variables of largest economies of the metropolitan areas. Amongst the other metropolitan areas, such as Salvador and Recife, they do not appear in the top twenty commonalities. Furthermore, the first macroeconomic and monetary related variable to appear is Exports, ranked twenty-third followed by SI, NIR, PS and M1 lying between thirtieth and thirty-third positions in the ranking. In the factor FC the commonalities related to macroeconomic and monetary related variables concentrate again in the second half of the ranking, together with less economically important metropolitan areas (Recife and Salvador).

The factors estimated with macroeconomic and monetary related variables, ML and MC, both have NIR leading the ranking of commonalities followed by SM2 and SI. Imports is ranked in fourth place, followed by exports and velocity of money, international reserves and M4. Except NIR1 that is the ninth highest commonality in ML, another lagged nominal interest rates appear only in the second-half of FL ranking. For instance, NIR6, NIR5 and NIR2 are in the last five positions.

In the Appendix E, Figures 4.4-4.9 we plot each estimated factor against the variable ranked in the first commonality (used here as a measure of the variability of that variable explained by the respective factor). That is: FL, FC, LB and UN, each of them are plotted against total Brazilian unemployment Rate (TBR); and ML and MC, each of them are plotted against NIR. These factors could be interpreted as indexes of the those variables. In Figure 4.7 use a sign-adjust to the factors whenever necessary, such that the factor should follow the variable reasonably well (or better than all of the other variables).

Correlation and Discrepancy amongst the Estimated Factors In Table 4.20 we show a simple correlation coefficient matrix of the estimated factors. In terms of these correlations the estimated factors can be divided into three groups: (i) FC and FL; (ii) LB and UN, and (iii) MC and ML.

The factors estimated with the entire dataset including lagged interest rates and without lagged interest rates, factors FL and FC respectively, are highly correlated with each

[15]The ranking of GDP amongst the metropolitan areas our dataset account for are the following: 1. Sao Paulo (SP); 2. Distrito Federal (DF); 3. Belo Horizonte (BH); 4. Porto Alegre (POA); 5. Salvador (SA); and 6. Recife (RE).

other (0.9995); and these factors are also relatively well correlated with the factor LB, estimated with labour market variables (0.9622 and 0.9670, respectively). When it comes to the factor UN, estimated with only unemployment rate variables, this factor is still relatively well correlated with FL (-0.8668), FC (-0.8762) and of course even higher correlated with LB (-0.9550).

The factors estimated with macroeconomic and monetary variables, ML and MC, are very low correlated with the other factors, such as FL (-0.3420 and -0.3267, respectively), FC (-0.3305 and -0.3166, respectively), LB (-0.1365 and -0.1273, respectively) and UN (0.0619 and 0.0584, respectively). Finally, as expected, the correlation between MC and ML is very high (0.9965).

The results of the discrepancy statistics of the space spanned amongst the estimated factors is displayed in Table 4.21. It is close to one (small discrepancy) when factors span nearly the same space, and close to zero (high discrepancy) otherwise. This statistics has been adapted from Ng and Boivin (2006), and can be given as follows:

$$S_{\hat{F}_i,\hat{F}_j} = \frac{\mathrm{tr}(\hat{F}_j'\hat{F}_i(\hat{F}_i'\hat{F}_i)^{-1}\hat{F}_i'\hat{F}_j)}{\mathrm{tr}(\hat{F}_j'\hat{F}_j)} \tag{4.10}$$

where $\mathrm{tr}(A)$ denote the trace of matrix A; $i, j =$ FC, FL, LB, UN, MC and ML; $S_{\hat{F}_i,\hat{F}_j} \in [0, 1]$; and, $S_{\hat{F}_i,\hat{F}_j} = 1$ if $i = j$.

The discrepancy statistics amongst estimated factors corroborate with the intuition brought by the correlations coefficients: the above three groups of factors are still valid here. FC and FL nearly span the same space (0.9990); FC and FL with respect to LB, have a relatively slightly higher discrepancy (0.9351 and 0.9258); for UN, both FC and FL display even higher discrepancy (0.7677 and 0.7513). And as expected, LB and UN display a small discrepancy (0.9119) since they are estimated with very similar blocks; the same intuition applies for MC and ML (0.9931) when result are even stronger.

The factors MC and ML are the ones that display the highest discrepancy statistics with respect to the others. For instance, with respect to FC (0.1002 and 0.1092, respectively), and to FL (0.1067, 0.1170, respectively) the discrepancies are very high, even considering the fact that a large number of variables used to estimate these factors have information in common, but the results suggest that these factors do not span the same factor space. Finally, MC and ML have even higher discrepancy with respect to LB and UN. This is not a surprise since their blocks do not have information in common. Then, the statistics are: MC and LB (0.0162), MC and UN (0.0038), ML and LB (0.0186), and, ML and UN (0.0034).

Robustness of Factor Model Specification As explained in Ng and Boivin (2006), there is no correct way to organize a dataset for factor analysis. The inclusion of more variables may lead to an increase of correlated errors whilst the common component estimation can be affected by the decrease of its average size. In this way the inclusion of variables (highly correlated) may not bring any benefit. In this regard, estimations of different blocks may affect the structure of the factors, as well as the size of dataset of different groups bringing different interpretation to the factors.

As described earlier, for our factor estimation we have separated the dataset judgmentally into blocks containing less variables. The reason for this is to explore our data dataset alternatively, to estimate different factors in the sense that they possibly underly different comovements within the same dataset, and finally use these estimated factors to improve our forecast exercise.

4.4 Conclusion

The goal of this chapter was to present the forecast of the Brazilian unemployment rate using diffusion index. Good results are found for two-,four- and six-month-ahead forecasts. We explored various methods of evaluation of forecast accuracy, forecast encompassing and combination. In our forecast exercise we have considered a linear model with constant coefficient throughout the sample (no time-varying coefficients). This perhaps brings a limitation to be explore in a future work since it is not unlikely that nonlinearities and/or structural breaks may occur. Moreover, our work can also be extended exploring the literature on factor analysis (i.e., estimation in frequency domain) or possibly newer versions of information criteria to determine the number of factors. Nevertheless, with the techniques explored in this chapter we have achieved satisfactory results when comparing our forecasts with the historical dataset.

Appendix A: Dataset Description

We use 69 monthly time series ranging from January 1998 to January 2009. We first took care of seasonality of each time series whenever it is necessary. Then we take log of all series not in rates or negative. Next, we induce stationary by taking differences.[16] In this regard we have taken first difference for all of the time series which has shown to be enough to induce stationary. Finally we standardized our series in the sense that they display zero mean and unit variance. These procedures are common use in a factor models framework.

Our dataset with respect to unemployment and employment (total, open, hidden, underemployment and discouraged) were obtained from the Fundação Sistema Estadual de Análise de Dados, in which is under the responsibility of the Pesquisa de Emprego e Desemprego (Seade/PED - "*Employment and Unemployment Research*"). This dataset has information for metropolitan areas cross country, which are major cities in Brazil: Sao Paulo (SP), Salvador (RJ), Porto Alegre (POA), Recife (RE), Belo Horizonte (BH) and Distrito Federal (DF) in addition to the aggregate dataset for the country. The rest of our dataset was obtained from the Instituto de Pesquisa Economica Aplicada website (www.ipeadata.gov.br - IPEA - "*Institute of Research Applied Economics*") which is an institution led by the Brazilian government to provide economic research. All data were freely obtained.

We have separated our dataset amongst two groups (labor market and monetary variables). They can be read as follows: a code, a brief description, (sa) to denote that the respective series has been seasonally adjusted, (nsa) if not seasonally adjusted and (log) for the dataset not in rates or negative values in which we applied logarithm. It is noteworthy to remind the reader that a first difference has been taken from each time series in order to induce stationary.

[16] We used the *Census X12* multiplicative and additive X11 methods to verify seasonality of the series. All series have been seasonally adjusted via additive method. We used Phillips-Perron and Augmented Dickey-Fuller tests for stationary (verifying models for intercept, trend and intercept and no trend and no intercept), and we found that we cannot accept stationary in any of the series. Reports of these tests (via E-Views 6) are available upon request.

Table 4.1: Labour Market Dataset

Code	Description	Code	Description
HWRJ	worked hours for the RJ industry (2006=100) (sa, log)	TSA	totalunemployment of SA (sa)
HWBR	worked hours for the Brazilian industry (2006=100) (sa, log)	OSA	open unemployment of SA (sa)
PEBR	employed workers for the Brazilian industry (2006=100) (sa, log)	HSA	hidden unemployment of SA (sa)
HPSP	paid hours for the SP industry (2006=100) (sa, log)	USA	underemployment of SA (sa)
POSP	occupied workers for the SP industry (2006=100) (sa, log)	DSA	discouraged unemployment of SA (sa)
HWSP	worked hours for the SP industry (2006=100) (sa, log)	TDF	total unemployment of DF (sa)
RMW	real minimun wage (sa, log)	ODF	open unemployment of DF (sa)
TBH	total unemployment of BH (sa)	HDF	hidden unemployment of of DF (sa)
OBH	open unemployment of BH (sa)	UDF	underemployment of DF (nsa)
HBH	hidden unemployment of BH (sa)	DDF	discouraged unemployment of DF (sa)
UBH	underemployment of BH (sa)	TPOA	total unemployment of POA (sa)
DBH	discouraged unemployment of BH (sa)	OPOA	open unemployment of POA (sa)
TSP	total unemployment of SP (sa)	HPOA	hidden unemployment of POA (sa)
OSP	open unemployment of SP (sa)	UPOA	underemployment of POA (sa)
HSP	hidden unemployment of SP (sa)	DPOA	discouraged unemployment of POA (sa)
USP	underemployment of SP (nsa)	TBR	total unemployment of Brazil (sa)
DSP	discouraged unemployment of SP (sa)	OBR	open unemployment of Brazil (sa)
TRE	total unemployment of RE (sa)	HBR	hidden unemployment of Brazil (sa)
ORE	open unemployment of RE (sa)	UBR	underemployment of Brazil (sa)
HRE	hidden unemployment of RE (sa)	DBR	discouraged unemployment of Brazil (sa)
URE	underemployment of RE (sa)	DRE	discouraged unemployment of RE (sa)
HWSP	worked hours for the SP industry (2006=100) (sa, log)	PS	payslip for Brazilian industry (sa, log)

Table 4.2: Monetary Dataset

Code	Description	Code	Description
ER	Exchange Rate (Brazilian Real×US Dollar)	M0	Monetary Base (sa, log)
IM	Imports (sa, log)	M1	M0 and money supply (sa, log)
EXP	Exports (sa, log)	DS	Deposits in Saving Accounts (sa, log)
M0A	Broad defininition of M0 (compulsory deposit) (sa, log)	NIR	Nominal Interest Rate
SI	Saving yields (sa)	RI	International Reserves (in US Dollars) (sa, log)
NIR1	1 lag Nominal Interest Rate (sa)	NIR2	2 lags Nominal Interest Rate (sa)
NIR3	3 lags Nominal Interest Rate (sa)	NIR4	4 lags Nominal Interest Rate (sa)
NIR5	5 lags Nominal Interest Rate (sa)	NIR6	6 lags Nominal Interest Rate (sa)
M1	Demand Deposit (sa, log) + M0	CPU	Credit to the Public Sector (sa, log)
CPR	Credit to the Private Sector (sa,log)	M2	Deposit in Saving Accounts + M1 (sa, log)
M3	Certificate of Deposit + M2 (sa,log)	M4	Public bonds + M3 (sa,log)
SM1	Equity Fund of the main financial investments (sa)	INF	Inflation Rate (sa)
SM2	Mutual Fund of the main financial investimets (sa)	TJLP	Long-term interest rates (sa)
D	Demand Deposits: velocity of money (withdraw/balance) (sa, log)		

Appendix B: Information Criteria

Table 4.3: Choice of Number of Factors via Bai and Ng (2002)

Block and IC	Number of Factors	Block and IC	Number of Factors
FC $IC_{p1}(k)$	1	MC $IC_{p1}(k)$	1
FC $IC_{p2}(k)$	1	MC $IC_{p2}(k)$	1
FC $IC_{p3}(k)$	1	MC $IC_{p3}(k)$	1
LB $IC_{p1}(k)$	1	UN $IC_{p1}(k)$	1
LB $IC_{p2}(k)$	1	UN $IC_{p2}(k)$	1
LB $IC_{p3}(k)$	1	UN $IC_{p3}(k)$	1
FL $IC_{p1}(k)$	1	ML $IC_{p1}(k)$	1
FL $IC_{p2}(k)$	1	ML $IC_{p2}(k)$	1
FL $IC_{p3}(k)$	1	ML $IC_{p3}(k)$	1

We have also tested for zero factors.

Appendix C: Factor Analysis

Table 4.4: Covariability Explained by First Ten Factors (in %)

Blocks	FL	FC	LB	UN	MC	ML
1^{st} Factor	10.10	10.99	15.02	16.81	17.63	13.78
2^{nd} Factor	7.58	8.16	10.72	12.56	11.72	9.56
3^{rd} Factor	7.19	7.75	8.62	9.50	10.83	8.76
4^{th} Factor	5.63	6.13	7.07	8.06	8.01	6.82
5^{th} Factor	4.76	5.12	6.48	7.66	7.28	6.47
6^{th} Factor	4.16	4.43	6.05	6.46	6.78	5.96
7^{th} Factor	4.00	4.21	5.37	5.64	5.25	4.93
8^{th} Factor	3.84	4.09	4.66	4.99	4.70	4.69
9^{th} Factor	3.63	3.81	4.11	3.94	4.46	4.51
10^{th} Factor	3.07	3.31	3.29	3.65	3.90	3.98

Table 4.5: Rank of Commonalities of the First Factor by block of dataset (in % and variables in brackets)

Rank	FL	FC	LB	UN	MC	ML
1	8.00(TBR)	8.10(TBR)	9.60 (TBR)	9.32(TBR)	21.92(NIR)	21.458 (NIR)
2	5.34(PEBR)	5.30(PEBR)	5.22 (TBH)	6.44(TDF)	19.31(SM2)	18.95 (SM2)
3	5.07(OBR)	5.03(OBR)	5.21 (TSP)	5.77(HBR)	16.70(SI)	15.965 (SI)
4	4.72(TSP)	4.77(TSP)	5.07 (OSP)	5.40(UBR)	11.64(M)	12.018 (M)
5	4.50(POSP)	4.50(POSP)	5.02 (PEBR)	4.84(TBH)	9.34(X)	9.699 (X)
6	4.43(TBH)	4.46(TBH)	4.91 (HBR)	4.78(HSP)	5.92(D)	6.255 (D)
7	3.99(TPOA)	4.02(TPOA)	4.87 (TDF)	4.78(HDF)	4.67(RI)	4.555 (RI)
8	3.72(HBR)	3.83(HBR)	4.64 (TPOA)	4.69(TSP)	3.61(M4)	3.138 (M4)
9	3.40(OSP)	3.47(TDF)	4.44 (POSP)	4.51(HPOA)	1.67(M3)	1.523 (NIR1)
10	3.36(TDF)	3.38(OSP)	3.91 (UBR)	4.24(TPOA)	1.35(SM1)	1.424 (M3)
11	2.95(HWSP)	2.89(UBR)	3.60 (HPOA)	4.10(TSA)	1.23(TJLP)	1.207(SM1)
12	2.83(HPSP)	2.84(HWSP)	3.29 (OSP)	4.04(UPOA)	0.82(M1)	1.162(TJLP)
13	2.79(UBR)	2.77(HPOA)	3.11 (HBH)	3.99(DSP)	0.47(CPU)	0.498(M1)
14	2.67(HPOA)	2.75(HPSP)	3.06 (UPOA)	3.99(DDF)	0.32(ER)	0.419(CPU)
15	2.63(HBH)	2.65(HBH)	3.01 (UBH)	3.76(OBR)	0.27(M2)	0.375(NIR4)
16	2.60(M)	2.53(UBH)	2.83 (TSA)	3.27(UBH)	0.26(DS)	0.318(DS)
17	2.50(UBH)	2.52(M)	2.58 (HSP)	3.26(HBH)	0.21(CPR)	0.254(M2)
18	2.37(HWBR)	2.43(UPOA)	2.58 (HDF)	2.47(OSP)	0.18(M0)	0.220((ER)
19	2.34(UPOA)	2.28 (HWBR)	2.52 (OBH)	2.38(HSA)	0.07(M0A)	0.145(NIR3)
20	2.16(OBH)	2.17(OBH)	2.15 (DSP)	2.07(OBH)	0.04(IR)	0.140(M0)
21	1.82(TSA)	1.90(TSA)	2.15 (DDF)	1.77(USA)		0.129(CPR)
22	1.72(HWRJ)	1.67(HWRJ)	1.61 (HPSP)	1.60(USP)		0.103(NIR6)
23	1.59(X)	1.66(HSP)	1.48 (ODF)	1.60(UDF)		0.015(IR)
24	1.53 (HSP)	1.66(HDF)	1.30 (OPOA)	1.42(DPOA)		0.013(M0A)
25	1.53 (HDF)	1.54(X)	1.28 (HWSP)	1.25(ODF)		0.006 (NIR5)
26	1.39 (DSP)	1.49(DSP)	1.26 (HSA)	1.13(OSA)		0.000 (NIR2)
27	1.39 (DDF)	1.49(DDF)	1.23 (DPOA)	0.72(DBH)		
28	1.38 (OPOA)	1.35(OPOA)	0.99 (OSA)	0.67(OPOA)		
29	1.27 (ODF)	1.24(ODF)	0.88 (PS)	0.44(DBR)		
30	1.21 (SI)	1.16(SI)	0.83 (USP)	0.42(URE)		
31	1.05 (NIR)	1.01(PS)	0.83 (UDF)	0.42(DSA)		
32	1.01 (PS)	0.98(NIR)	0.76 (HWBR)	0.37(HRE)		
33	0.91 (M1)	0.91(DPOA)	0.75 (USA)	0.05(TRE)		
34	0.88 (DPOA)	0.87(M1)	0.74 (DBH)	0.03(DRE)		
35	0.78 (OSA)	0.81(OSA)	0.72 (DBR)	0.01(ORE)		
36	0.68 (SM2)	0.64(SM2)	0.63 (HWRJ)			
37	0.60 (DBR)	0.61(DBR)	0.41 (DSA)			
38	0.59 (DBH)	0.59(DBH)	0.27 (URE)			

Table 4.5: continuation

Rank	FL	FC	LB	UN	MC	ML
39	0.54 (HSA)	0.58(HSA)	0.14(HRE)			
40	0.52 (D)	0.48(D)	0.08(ORE)			
41	0.43 (USP)	0.48(USP)	0.04(RMW)			
42	0.43 (UDF)	0.48(UDF)	0.01(DRE)			
43	0.37 (CPR)	0.36(M3)	0.01(TRE)			
44	0.35 (M3)	0.36(CPR)				
45	0.33 (USA)	0.35(USA)				
46	0.32 (DS)	0.31(DS)				
47	0.28 (M0A)	0.29(URE)				
48	0.27 (M2)	0.28(M2)				
49	0.27 (URE)	0.27(CPU)				
50	0.26 (CPU)	0.26(M0A)				
51	0.23 (TJLP)	0.22(TJLP)				
52	0.21 (NIR4)	0.21(IR)				
53	0.20 (IR)	0.21(DSA)				
54	0.20 (DSA)	0.19(HRE)				
55	0.19 (NIR1)	0.09(SM1)				
56	0.17 (HRE)	0.08(ORE)				
57	0.17 (NIR3)	0.08(ER)				
58	0.16 (NIR2)	0.04(RMW)				
59	0.09 (ORE)	0.03(M0)				
60	0.09 (SM1)	0.02(M4)				
61	0.07 (ER)	0.01(DRE)				
62	0.04 (RMW)	0.01(TRE)				
63	0.04 (NIR5)	0.00(RI)				
64	0.03 (M0)					
65	0.03 (M4)					
66	0.02 (DRE)					
67	0.02 (TRE)					
68	0.01 (NIR6)					
69	0.00 (RI)					

Appendix D: Forecasts

TABLE 4.6: $h-$month-ahead Out-of Sample Forecasting against the benchmark

$h=1$	AR(4)	FL	FC	LB	UN	MC	ML
MSE	0.00075	0.00071	0.00071	0.00072	0.00077	0.00076	0.00076
Rel.MSE	1	0.9423	0.9417	0.9581	1.0201	1.0145	1.0142
$h=2$	AR(4)	FL	FC	LB	UN	MC	ML
MSE	0.00082	0.00074	0.00074	0.00075	0.00082	0.00082	0.00082
Rel.MSE	1	0.9130	0.9114	0.9257	1.011	0.9999	1.0003
$h=4$	AR(4)	FL	FC	LB	UN	MC	ML
MSE	0.00084	0.00082	0.00082	0.00081	0.00080	0.00084	0.00084
Rel.MSE	1	0.9766	0.9740	0.9604	0.9518	1.0032	1.0017
$h=6$	AR(4)	FL	FC	LB	UN	MC	ML
MSE	0.00065	0.00059	0.00059	0.00059	0.00058	0.00065	0.00066
Rel.MSE	1	0.8964	0.8968	0.8968	0.88873	1.0022	1.0034
$h=8$	AR(4)	FL	FC	LB	UN	MC	ML
MSE	0.00025	0.00025	0.00025	0.00025	0.00025	0.00026	0.00025
Rel.MSE	1	0.99311	0.9921	0.9895	0.9993	1.0216	1.0194
$h=12$	AR(4)	FL	FC	LB	UN	MC	ML
MSE	0.00073	0.00074	0.00074	0.00073	0.00072	0.00074	0.00074
Rel.MSE	1	1.0123	1.0123	0.9971	0.9818	1.0081	1.0056

Relative-MSE (Rel.MSE): Rel.MSE(factor-based) = MSE(factor-based)/MSE(AR(4)).

Table 4.7: Diebold and Mariano (2002)[a] factor-based forecast against the benchmark

$h-$month-ahead out-of sample forecasting
m_1 (model one) vs m_2 (model two)
y_{t+h}^{r,m_1} vs y_{t+h}^{r,m_2}; $l^{m_1} = (y_{t+h}^{r,m_1} - \widehat{y}_{t+h}^{r,m_1})^2$ and $l^{m_2} = (y_{t+h}^{r,m_2} - \widehat{y}_{t+h}^{r,m_2})^2$
$d = l^{m_1} - l^{m_2}$; $d_{t+1} = c + u_{t+1}$, where c is a constant
$H_0 : c = 0$ against $H_1^{m_1} : c < 0$ or $H_1^{m_2} : c > 0$

$h = 1$	DM-Stats	p-value	Winner	$h = 2$	DM-Stats	p-value	Winner
FL vs AR(4)	-1.3095	0.2022	-	FL vs AR(4)	-1.6198	0.118	-
FC vs AR(4)	-1.3287	0.1959	-	FC vs AR(4)	-1.6224	0.1177	-
LB vs AR(4)	-0.9868	0.3331	-	LB vs AR(4)	-1.2078	0.2388	-
UN vs AR(4)	0.3765	0.7096	-	UN vs AR(4)	0.1683	0.8677	-
MC vs AR(4)	0.4711	0.6416	-	MC vs AR(4)	-0.0070	0.9944	-
ML vs AR(4)	0.4678	0.6439	-	ML vs AR(4)	0.0774	0.9389	-
$h = 4$	DM-Stats	p-value	Winner	$h = 6$	DM-Stats	p-value	Winner
FL vs AR(4)	-0.5928	0.5593	-	FL vs AR(4)	-2.1064	0.0479	FL
FC vs AR(4)	-0.6673	0.5114	-	FC vs AR(4)	-2.0912	0.0494	FC
LB vs AR(4)	-1.0681	0.2970	-	LB vs AR(4)	-2.2415	0.0364	LB
UN vs AR(4)	-1.3619	0.1870	-	UN vs AR(4)	-2.2673	0.03460	UN
MC vs AR(4)	0.1089	0.9142	-	MC vs AR(4)	0.1139	0.9104	-
ML vs AR(4)	0.0676	0.9466	-	ML vs AR(4)	0.8459	0.1968	-
$h = 8$	DM-Stats	p-value	Winner	$h = 12$	DM-Stats	p-value	Winner
FL vs AR(4)	-1.3469	0.1946	-	FL vs AR(4)	0.5710	0.5770	-
FC vs AR(4)	-1.3678	0.1881	-	FC vs AR(4)	0.5171	0.6131	-
LB vs AR(4)	-1.6393	0.1184	-	LB vs AR(4)	-0.1729	0.8651	-
UN vs AR(4)	-0.1509	0.8817	-	UN vs AR(4)	-1.6211	0.1272	-
MC vs AR(4)	0.6382	0.5313	-	MC vs AR(4)	0.6139	0.5490	-
ML vs AR(4)	0.6000	0.5559	-	ML vs AR(4)	0.4587	0.6534	-

[a] Newey and West (1987) HAC results of $d = l^{m_1} - l^{m_2}$ regressed on a constant, as explained in section 2.3
If a calculated t-ratio is statistically significant and with a negative constant results in m_1 defeating m_2 ,
and vice-versa; if the test is not significant, nothing can be said about m_1 and m_2 defeating each other.

Table 4.8: Diebold and Mariano (2002)[a] within factors

$h-$month-ahead Out-of Sample Forecasting

$h=1$	Rel.MSE	DM-Stats	p-value	Winner	$h=2$	Rel.MSE	DM-Stats	p-value	Winner
FL vs FC	1.0006	0.2348	0.8162	-	FL vs FC	1.0017	0.5277	0.6025	-
FL vs LB	0.9834	-0.6862	0.4988	-	FL vs LB	0.9862	-0.4931	0.6264	-
FL vs UN	0.9237	-1.6178	0.1182	-	FL vs UN	0.9026	-1.6690	0.1081	FL
FL vs MC	0.9368	-1.5738	0.1280	-	FL vs MC	0.9131	-1.6947	0.1030	FL
FL vs ML	0.9362	-1.6026	0.1215	-	FL vs ML	0.9127	-1.7042	0.1012	FL
FC vs LB	0.9828	-0.7840	0.4403	-	FC vs LB	0.9845	-0.6259	0.5372	-
FC vs UN	0.9231	-1.6911	0.1032	FC	FC vs UN	0.9010	-1.7542	0.0921	FC
FC vs MC	0.9362	-1.5870	0.1250	-	FC vs MC	0.9114	-1.6934	0.1033	FC
FC vs ML	0.9356	-1.6148	0.1188	-	FC vs ML	0.9110	-1.7020	0.1016	FC
LB vs UN	0.9392	-2.1972	0.0374	UN	LB vs UN	0.9151	-2.1421	0.0425	LB
LB vs ML	0.9525	-1.1307	0.2688	-	LB vs ML	0.9254	-1.2411	0.2265	
LB vs MC	0.9525	-1.1307	0.2688	-	LB vs MC	0.9258	-1.2362	0.2283	
UN vs ML	0.9519	-1.1512	0.2605	-	UN vs ML	1.0111	0.1610	0.8734	
UN vs MC	1.0141	0.2553	0.8005	-	UN vs MC	1.0116	0.1673	0.8684	
ML vs MC	1.0006	0.6013	0.5530		ML vs MC	1.0004	1.0004	0.4280	
$h=4$	Rel.MSE	DM-Stats	p-value	Winner	$h=6$	Rel.MSE	DM-Stats	p-value	Winner
FL vs FC	1.0026	2.2271	0.0364	FC	FL vs FC	0.9994	-0.4140	0.6832	-
FL vs LB	1.0168	0.8846	0.3859	-	FL vs LB	1.0086	0.7379	0.4691	-
FL vs UN	1.0260	0.6197	0.5418	-	FL vs UN	0.9637	-1.3256	0.1998	-
FL vs MC	0.9734	-0.4892	0.6295	-	FL vs MC	0.8943	-1.8570	0.0780	FL
FL vs ML	0.9748	-0.4770	0.6380	-	FL vs ML	0.8933	-1.9062	0.0710	FL
FC vs LB	1.0141	0.7808	0.4432	-	FC vs LB	1.0091	0.8365	0.4127	-
FC vs UN	1.0233	0.5672	0.5762	-	FC vs UN	0.9642	-1.3228	0.2008	-
FC vs MC	0.9708	-0.5426	0.5928	-	FC vs MC	0.8948	-1.8417	0.0803	FC
FC vs ML	0.9722	-0.5320	0.6000	-	FC vs ML	0.8937	-1.8914	0.0731	FC
LB vs UN	1.0090	0.3522	0.7280	-	LB vs UN	0.9555	-1.9186	0.0694	LB
LB vs ML	0.9587	-0.9241	0.3654	-	LB vs ML	0.8856	-2.052	0.0534	LB
LB vs MC	0.9573	-0.9327	0.3610	-	LB vs MC	0.8867	-1.9990	0.0593	LB
UN vs ML	0.9501	-1.2634	0.2196	-	UN vs ML	0.9269	-1.8631	0.0771	UN
UN vs MC	0.9487	-1.2739	0.2159	-	UN vs MC	0.9280	-1.7680	0.0923	UN
ML vs MC	0.9985	-0.4171	0.6805		ML vs MC	1.0011	0.4840	0.6336	
$h=8$	Rel.MSE	DM-Stats	p-value	Winner	$h=12$	Rel.MSE	DM-Stats	p-value	Winner
FL vs FC	1.0009	1.3325	0.1992	-	FL vs FC	1.0012	2.1879	0.0461	FC
FL vs LB	1.0036	1.2946	0.2117	-	FL vs LB	1.0152	2.2649	0.0399	LB
FL vs UN	0.9938	-1.3248	0.2017	-	FL vs UN	1.0310	1.7793	0.0968	UN
FL vs MC	0.9720	-0.7480	0.4640	-	FL vs MC	1.0041	0.1340	0.8952	-
FL vs ML	0.9742	-0.7188	0.4814	-	FL vs ML	1.0066	0.2173	0.8310	-
FC vs LB	1.0026	1.0797	0.2945	-	FC vs LB	1.0140	2.1883	0.04609	LB
FC vs UN	0.9928	-1.4619	0.1609	-	FC vs UN	1.0298	1.7370	0.1043	UN
FC vs MC	0.9711	-0.7642	0.4546	-	FC vs MC	1.0029	0.0950	0.9255	-
FC vs ML	0.9732	-0.7360	0.4712	-	FC vs ML	1.0054	0.1784	0.8609	-
LB vs UN	0.9902	-2.3881	0.0281	LB	LB vs UN	1.0155	1.3840	0.1880	-
LB vs ML	0.9706	-0.8227	0.4214	-	LB vs ML	0.9914	-0.3416	0.7377	-
LB vs MC	0.9685	-0.8492	0.4069	-	LB vs MC	0.9890	-0.4330	0.6716	-
UN vs ML	0.9802	-0.5988	0.5567	-	UN vs ML	0.9763	-1.3540	0.1971	-
UN vs MC	0.9780	-0.6376	0.5317	-	UN vs MC	0.9738	-1.4644	0.1651	-
ML vs MC	0.9977	-1.1058	0.2833	-	ML vs MC	0.9975	-1.4845	0.1598	-

Table 4.9: $h-$month-ahead Out-of Sample Forecasting

using more than one factor against the benchmark

$h = 1$	AR(4)	LB and ML	LB and MC	UN and ML	UN and MC
MSE	0.00075	0.00072	0.00072	0.00077	0.00077
Rel.MSE	1	0.9644	0.9640	1.0247	1.0240
$h = 2$	AR(4)	LB and ML	LB and MC	UN and ML	UN and MC
MSE	0.00082	0.00075	0.00076	0.00083	0.00083
Rel.MSE	1	0.9260	0.9263	1.0121	1.0119
$h = 4$	AR(4)	LB and ML	LB and MC	UN and ML	UN and MC
MSE	0.00084	0.00081	0.00081	0.00080	0.00080
Rel.MSE	1	0.9604	0.9623	0.9530	0.9547
$h = 6$	AR(4)	LB and ML	LB and MC	UN and ML	UN and MC
MSE	0.00065	0.00058	0.00057	0.00061	0.00061
Rel.MSE	1	0.8813	0.8800	0.9287	0.9274
$h = 8$	AR(4)	LB and ML	LB and MC	UN and ML	UN and MC
MSE	0.00025	0.00025	0.00025	0.00025	0.00026
Rel.MSE	1	1.0060	1.0087	1.01972	1.0220
$h = 12$	AR(4)	LB and ML	LB and MC	UN and ML	UN and MC
MSE	0.00073	0.00074	0.00074	0.00072	0.00073
Rel.MSE	1	1.0046	1.0074	0.9871	0.9898

Relative-MSE (Rel.MSE): Rel.MSE(factor-based) = MSE(factor-based)/MSE(AR(4)).

TABLE 4.10: Diebold and Mariano (2002)[a] factor-based (with two factors) forecast against the benchmark

$h-$month-ahead out-of sample forecasting

m_1 (model one) vs m_2 (model two)

y^{r,m_1}_{t+h} vs y^{r,m_2}_{t+h}; $l^{m_1} = (y^{r,m_1}_{t+h} - \widehat{y}^{r,m_1}_{t+h})^2$ and $l^{m_2} = (y^{r,m_2}_{t+h} - \widehat{y}^{r,m_2}_{t+h})^2$

$d = l^{m_1} - l^{m_2}$; $d_{t+1} = c + u_{t+1}$, where c is a constant

$H_0 : c = 0$ against $H^{m_1}_1 : c < 0$ or $H^{m_2}_1 : c > 0$

$h = 1$	DM-Stats	p-value	Winner	$h = 2$	DM-Stats	p-value	Winner
LB and ML vs AR(4)	-0.8420	0.4077	-	LB and ML vs AR(4)	-1.2041	0.2402	-
LB and MC vs AR(4)	-0.8512	0.4027	-	LB and MC vs AR(4)	-1.2016	0.2412	-
UN and ML vs AR(4)	0.4753	0.6386	-	UN and ML vs AR(4)	0.1768	0.8610	-
UN and MC vs AR(4)	0.4662	0.6450	-	UN and MC vs AR(4)	0.1743	0.8630	-
$h = 4$	DM-Stats	p-value	Winner	$h = 6$	DM-Stats	p-value	Winner
LB and ML vs AR(4)	-0.7163	0.4813	-	LB and ML vs AR(4)	-2.2344	0.0370	LB and ML
LB and MC vs AR(4)	-0.6528	0.5206	-	LB and MC vs AR(4)	-2.2360	0.0369	LB and MC
UN and ML vs AR(4)	-0.8808	0.3879	-	UN and ML vs AR(4)	-2.1938	0.0402	UN and ML
UN and MC vs AR(4)	-0.8048	0.4295	-	UN and MC vs AR(4)	-2.1849	0.0409	UN and MC
$h = 8$	DM-Stats	p-value	Winner	$h = 12$	DM-Stats	p-value	Winner
LB and ML vs AR(4)	0.1918	0.8500	-	LB and ML vs AR(4)	0.2546	0.8027	-
LB and MC vs AR(4)	0.2653	0.7937	-	LB and MC vs AR(4)	0.3950	0.6987	-
UN and ML vs AR(4)	0.5938	0.5599	-	UN and ML vs AR(4)	-0.7375	0.4729	-
UN and MC vs AR(4)	0.6331	0.5345	-	UN and MC vs AR(4)	-0.5484	0.5920	-

[a] Newey and West (1987) HAC results of $d = l^{m_1} - l^{m_2}$ regressed on a constant, as explained in section 2.3

If a calculated t-ratio is statistically significant and with a negative constant results in m_1 defeating m_2, and vice-versa; if the test is not significant, nothing can be said about m_1 and m_2 defeating each other.

TABLE 4.11: Forecast Encompassing factor-based forecast against the benchmark

$$y^r_{t+h} = \beta_{m_1}\widehat{y}^{r,m_1}_{t+h} + \beta_{m_2}\widehat{y}^{r,m_2}_{t+h} + \epsilon^r_{t+h}$$

If $(\beta_{m_1}, \beta_{m_2}) = (1, 0)$ model one encompasses model two;
If $(\beta_{m_1}, \beta_{m_2}) = (0, 1)$ model two encompasses model one;

$h = 1$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass	$h = 2$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass
FL vs AR(4)	−0.70 (1.82)	1.69 (1.63)	-	FL vs AR(4)	−1.68 (2.03)	1.89 (1.70)	-
FC vs AR(4)	−0.69 (1.81)	1.68 (1.63)	-	FC vs AR(4)	−1.66 (2.03)	1.88 (1.70)	-
LB vs AR(4)	−0.66 (1.80)	1.65 (1.59)	-	LB vs AR(4)	−1.74 (1.91)	1.90 (1.60)	-
UN vs AR(4)	−0.92 (1.71)	1.85 (1.52)	-	UN vs AR(4)	−1.94 (1.62)	1.96 (1.37)	-
MC vs AR(4)	−1.12 (1.60)	2.21 (1.38)	-	MC vs AR(4)	−1.36 (1.78)	1.76 (1.33)	-
ML vs AR(4)	−1.13 (1.60)	2.22 (1.38)	-	ML vs AR(4)	−1.36 (1.78)	1.76 (1.33)	-
$h = 4$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass	$h = 6$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass
FL vs AR(4)	−0.27 (2.41)	0.86 (1.77)	-	FL vs AR(4)	1.25 (2.58)	−0.26 (2.26)	-
FC vs AR(4)	−0.27 (2.40)	0.86 (1.76)	-	FC vs AR(4)	2.25 (2.57)	−0.26 (1.25)	-
LB vs AR(4)	−0.44 (2.37)	0.93 (1.76)	-	LB vs AR(4)	1.14 (2.52)	−0.20 (2.20)	-
UN vs AR(4)	−0.63 (2.23)	0.98 (1.64)	-	UN vs AR(4)	0.77 (2.36)	0.11 (2.07)	-
MC vs AR(4)	−0.79 (1.94)	1.17 (1.34)	-	MC vs AR(4)	0.68 (2.36)	0.32 (2.06)	-
ML vs AR(4)	−0.78 (1.93)	1.16 (1.34)	-	ML vs AR(4)	2.06 (2.35)	0.32 (0.67)	
$h = 8$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass	$h = 12$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass
FL vs AR(4)	−2.95 (1.83)	2.62 (2.00)	-	FL vs AR(4)	−0.63 (3.23)	0.96 (3.05)	-
FC vs AR(4)	−2.94 (1.83)	2.61 (2.00)	-	FC vs AR(4)	−0.63 (3.24)	0.96 (3.06)	-
LB vs AR(4)	−2.83 (1.84)	2.51 (2.01)	-	LB vs AR(4)	−0.44 (3.32)	0.80 (3.13)	-
UN vs AR(4)	−2.66 (1.92)	2.33 (2.09)	-	UN vs AR(4)	−0.03 (3.41)	0.43 (3.21)	-
MC vs AR(4)	−3.08 (1.98)	2.74 (2.15)	-	MC vs AR(4)	−0.23 (3.33)	0.60 (3.13)	-
ML vs AR(4)	−3.05 (1.98)	2.72 (2.15)	-	ML vs AR(4)	−0.23 (3.34)	0.60 (3.14)	-

Standard errors in parentheses. * significant at 5% level; ** significant at 1% level.

TABLE 4.12: Forecast Encompassing factor-based forecast (with two factors) against the benchmark

$$y^r_{t+h} = \beta_{m_1}\widehat{y}^{r,m_1}_{t+h} + \beta_{m_2}\widehat{y}^{r,m_2}_{t+h} + \epsilon^r_{t+h}$$

If $(\beta_{m_1}, \beta_{m_2}) = (1, 0)$ model one encompasses model two;

If $(\beta_{m_1}, \beta_{m_2}) = (0, 1)$ model two encompasses model one;

$h = 1$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass	$h = 2$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass
LB and ML vs AR(4)	2.79 (2.76)	−3.10 (2.85)	-	LB and ML vs AR(4)	0.20 (1.09)	0.15 (0.77)	-
LB and MC vs AR(4)	−0.62 (1.78)	1.61 (1.59)	-	LB and MC vs AR(4)	0.20 (1.09)	0.16 (0.78)	-
UN and ML vs AR(4)	−1.15 (1.67)	2.06 (1.48)	-	UN and ML vs AR(4)	−0.08 (1.13)	0.25 (0.81)	-
UN and MC vs AR(4)	−1.14 (1.67)	2.06 (1.48)	-	UN and MC vs AR(4)	−0.08 (1.13)	0.25 (0.81)	-
$h = 4$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass	$h = 6$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass
LB and ML vs AR(4)	0.46 (0.65)	0.09 (0.57)	-	LB and ML vs AR(4)	−0.02 (0.72)	0.94 (0.63)	-
LB and MC vs AR(4)	0.46 (0.65)	0.08 (0.57)	-	LB and MC vs AR(4)	−0.02 (0.72)	0.95 (0.64)	-
UN and ML vs AR(4)	0.30 (0.69)	0.04 (0.61)	-	UN and ML vs AR(4)	0.01 (0.68)	0.92 (0.60)	-
UN and MC vs AR(4)	0.30 (0.69)	0.04 (0.61)	-	UN and MC vs AR(4)	0.00 (0.68)	0.93 (0.61)	-
$h = 8$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass	$h = 12$	$\widehat{\beta}^1$	$\widehat{\beta}^2$	Encompass
LB and ML vs AR(4)	−0.80 (0.64)	1.43 (1.09)	-	LB and ML vs AR(4)	−1.01 (1.20)	1.23 (1.01)	-
LB and MC vs AR(4)	−0.81 (0.64)	1.44 (1.08)	-	LB and MC vs AR(4)	−1.01 (1.20)	1.23 (1.01)	-
UN and ML vs AR(4)	1.31 (0.64)	1.31 (1.14)	-	UN and ML vs AR(4)	−0.65 (1.05)	0.95 (0.81)	-
UN and MC vs AR(4)	−0.75 (0.64)	1.33 (1.13)	-	UN and MC vs AR(4)	−0.65 (1.05)	0.95 (0.81)	-

Standard errors in parentheses. * significant at 5% level; ** significant at 1% level.

TABLE 4.13: Forecast Combination - Method Simple Averaging (equal weights)

$h = 1$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00075	0.00068	0.00068	0.00070	0.00074	0.00072	0.00076
Rel.MSE	1	0.9122	0.9117	0.9275	0.9837	0.9613	0.9619
$h = 2$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00082	0.00070	0.00070	0.00071	0.00077	0.00075	0.00075
Rel.MSE	1	0.8642	0.8627	0.8761	0.9501	0.9241	0.9245
$h = 4$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00084	0.00078	0.00078	0.00077	0.00075	0.00079	0.00079
Rel.MSE	1	0.9320	0.9293	0.9148	0.8971	0.9379	0.9365
$h = 6$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00065	0.00055	0.00055	0.00054	0.00056	0.00061	0.00061
Rel.MSE	1	0.8386	0.8390	0.8297	0.8639	0.9314	0.9321
$h = 8$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00025	0.00025	0.00025	0.00025	0.00025	0.00025	0.00025
Rel.MSE	1	0.9911	0.9906	0.9899	0.9988	1.0175	1.0152
$h = 12$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00073	0.00072	0.00072	0.00071	0.00070	0.00072	0.00072
Rel.MSE	1	0.9802	0.9790	0.9665	0.9535	0.9820	0.9798

Relative-MSE (Rel.MSE): Rel.MSE(factor-based) = MSE(factor-based)/MSE(AR(4)).

TABLE 4.14: Diebold and Mariano (2002)[a] factor-based against the benchmark

$h-$month-ahead Out-of Sample Forecasting - Forecast Combination - Simple Averaging (equal weights)
m_1 (model one) vs m_2 (model two)
y^{r,m_1}_{t+h} vs y^{r,m_2}_{t+h}; $l^{m_1} = (y^{r,m_1}_{t+h} - \widehat{y}^{r,m_1}_{t+h})^2$ and $l^{m_2} = (y^{r,m_2}_{t+h} - \widehat{y}^{r,m_2}_{t+h})^2$
$d = l^{m_1} - l^{m_2}$; $d_{t+1} = c + u_{t+1}$, where c is a constant
$H_0 : c = 0$ against $H_1^{m_1} : c < 0$ or $H_1^{m_2} : c > 0$

$h = 1$	DM-Stats	p-value	Winner	$h = 2$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-1.25	0.22	-	FL and AR(4) vs AR(4)	-1.54	0.13	-
FC and AR(4) vs AR(4)	-1.25	0.2220	-	FC and AR(4) vs AR(4)	-1.54	0.13	-
LB and AR(4) vs AR(4)	-1.01	0.3186	-	LB and AR(4) vs AR(4)	-1.30	0.20	-
UN and AR(4) vs AR(4)	-0.2103	0.8351	-	and AR(4) UN vs AR(4)	-0.51	0.61	-
MC and AR(4) vs AR(4)	-1.048	0.3043	-	MC and AR(4) vs AR(4)	-1.75	0.09	MC and AR(4)
ML and AR(4) vs AR(4)	-1.029	0.3130	-	ML and AR(4) vs AR(4)	-1.74	0.09	ML and AR(4)
$h = 4$	DM-Stats	p-value	Winner	$h = 6$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-1.01	0.31	-	FL and AR(4) vs AR(4)	-2.36	0.02	FL and AR(4)
FC and AR(4) vs AR(4)	-1.06	0.29	-	FC and AR(4) vs AR(4)	-2.35	0.02	FC and AR(4)
LB and AR(4) vs AR(4)	-1.31	0.20	-	LB and AR(4) vs AR(4)	-2.51	0.02	LB and AR(4)
UN and AR(4) vs AR(4)	-1.75	0.09	-	UN and AR(4) vs AR(4)	-2.85	0.00	UN and AR(4)
MC and AR(4) vs AR(4)	-1.60	0.12	-	MC and AR(4) vs AR(4)	-1.76	0.09	MC and AR(4)
ML and AR(4) vs AR(4)	-1.71	0.09	ML and AR(4)	ML and AR(4) vs AR(4)	-1.83	0.08	ML and AR(4)
$h = 8$	DM-Stats	p-value	Winner	$h = 12$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-0.22	0.82	-	FL and AR(4) vs AR(4)	-0.94	0.36	-
FC and AR(4) vs AR(4)	-0.22	0.82	-	FC and AR(4) vs AR(4)	-0.99	0.33	-
LB and AR(4) vs AR(4)	-0.24	0.80	-	LB and AR(4) vs AR(4)	-1.39	0.18	-
UN and AR(4) vs AR(4)	-0.02	0.97	-	UN and AR(4) vs AR(4)	-1.58	0.13	-
MC and AR(4) vs AR(4)	0.80	0.42	-	MC and AR(4) vs AR(4)	-0.44	0.66	-
ML and AR(4) vs AR(4)	0.72	0.48	-	ML and AR(4) vs AR(4)	-0.49	0.62	-

[a] Newey and West (1987) HAC results of $d = l^{m_1} - l^{m_2}$ regressed on a constant, as explained in section 2.3
If a calculated t-ratio is statistically significant and with a negative constant results in m_1 defeating m_2, and vice-versa; if the test is not significant, nothing can be said about m_1 and m_2 defeating each other.

TABLE 4.15: Forecast Combination - Method Bates and Granger

$h = 1$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00075	0.00068	0.00068	0.00070	0.00074	0.00072	0.00072
Rel.MSE	1	0.9129	0.9125	0.9280	0.9834	0.9612	0.9617
$h = 2$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00082	0.00071	0.00070	0.00072	0.00077	0.00075	0.00075
Rel.MSE	1	0.8662	0.8647	0.8777	0.9498	0.9241	0.9245
$h = 4$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00084	0.00078	0.00084	0.00077	0.00076	0.00079	0.00079
Rel.MSE	1	0.9325	0.9298	0.9156	0.8982	0.9378	0.9364
$h = 6$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00065	0.00055	0.00055	0.00054	0.00057	0.00061	0.00061
Rel.MSE	1	0.8415	0.8419	0.8329	0.8662	0.9313	0.9320
$h = 8$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00025	0.00025	0.00025	0.00025	0.00025	0.00025	0.00025
Rel.MSE	1	0.9911	0.9906	0.9899	0.9988	1.0176	1.0152
$h = 12$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00073	0.00072	0.00072	0.00071	0.00070	0.00072	0.00072
Rel.MSE	1	0.9800	0.9789	0.9666	0.9537	0.9819	0.9797

Relative-MSE (Rel.MSE): Rel.MSE(factor-based) = MSE(factor-based)/MSE(AR(4)).

TABLE 4.16: Diebold and Mariano (2002)[a] factor-based against the benchmark

$h-$month-ahead Out-of Sample Forecasting - Forecast Combination - Bates and Granger relative performance averaging
m_1 (model one) vs m_2 (model two)
y^{r,m_1}_{t+h} vs y^{r,m_2}_{t+h}; $l^{m_1} = (y^{r,m_1}_{t+h} - \widehat{y}^{r,m_1}_{t+h})^2$ and $l^{m_2} = (y^{r,m_2}_{t+h} - \widehat{y}^{r,m_2}_{t+h})^2$
$d = l^{m_1} - l^{m_2}$; $d_{t+1} = c + u_{t+1}$, where c is a constant
$H_0 : c = 0$ against $H_1^{m_1} : c < 0$ or $H_1^{m_2} : c > 0$

$h = 1$	DM-Stats	p-value	Winner	$h = 2$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-1.25	0.22	-	FL and AR(4) vs AR(4)	-1.55	0.13	-
FC and AR(4) vs AR(4)	-1.25	0.22	-	FC and AR(4) vs AR(4)	-1.55	0.13	-
LB and AR(4) vs AR(4)	-1.01	0.31	-	LB and AR(4) vs AR(4)	-1.30	0.20	-
UN and AR(4) vs AR(4)	-0.21	0.83	-	UN and AR(4) vs AR(4)	-0.51	0.60	-
MC and AR(4) vs AR(4)	-1.04	0.30	-	MC and AR(4) vs AR(4)	-1.75	0.09	MC and AR(4)
ML and AR(4) vs AR(4)	-1.03	0.31	-	ML and AR(4) vs AR(4)	-1.74	0.09	ML and AR(4)
$h = 4$	DM-Stats	p-value	Winner	$h = 6$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-1.01	0.32	-	FL and AR(4) vs AR(4)	-2.36	0.02	FL and AR(4)
FC and AR(4) vs AR(4)	-1.06	0.29	-	FC and AR(4) vs AR(4)	-2.35	0.02	FC and AR(4)
LB and AR(4) vs AR(4)	-1.31	0.20	-	LB and AR(4) vs AR(4)	-2.50	0.02	LB and AR(4)
UN and AR(4) vs AR(4)	-1.75	0.09	UN and AR(4)	UN and AR(4) vs AR(4)	-2.85	0.00	UN and AR(4)
MC and AR(4) vs AR(4)	-1.60	0.12	-	MC and AR(4) vs AR(4)	-1.76	0.09	MC and AR(4)
ML and AR(4) vs AR(4)	-1.71	0.09	ML and AR(4)	ML and AR(4) vs AR(4)	-1.83	0.08	ML and AR(4)
$h = 8$	DM-Stats	p-value	Winner	$h = 12$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-0.22	0.82	-	FL and AR(4) vs AR(4)	-0.94	0.35	-
FC and AR(4) vs AR(4)	-0.23	0.82	-	FC and AR(4) vs AR(4)	-0.99	0.33	-
LB and AR(4) vs AR(4)	-0.24	0.80	-	LB and AR(4) vs AR(4)	-1.39	0.18	-
UN and AR(4) vs AR(4)	-0.029	0.97	-	UN and AR(4) vs AR(4)	-1.59	0.13	-
MC and AR(4) vs AR(4)	0.80	0.43	-	MC and AR(4) vs AR(4)	-0.44	0.66	-
ML and AR(4) vs AR(4)	0.71	0.48	-	ML and AR(4) vs AR(4)	-0.49	0.62	-

[a] Newey and West (1987) HAC results of $d = l^{m_1} - l^{m_2}$ regressed on a constant, as explained in section 2.3
If a calculated t-ratio is statistically significant and with a negative constant results in m_1 defeating m_2, and vice-versa;
if the test is not significant, nothing can be said about m_1 and m_2 defeating each other.

TABLE 4.17: Forecast Combination - Method Granger and Ramanathan

$h = 1$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00075	0.00066	0.00066	0.00067	0.00071	0.00067	0.00067
Rel.MSE	1	0.8825	0.8825	0.8986	0.9415	0.8937	0.8939
$h = 2$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00082	0.00065	0.00065	0.00066	0.00070	0.00059	0.00059
Rel.MSE	1	0.7998	0.7992	0.8148	0.8622	0.7258	0.7263
$h = 4$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00084	0.00074	0.00073	0.00072	0.00067	0.00067	0.00067
Rel.MSE	1	0.8757	0.8731	0.8566	0.8017	0.7988	0.7970
$h = 6$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00065	0.00044	0.00044	0.00043	0.00043	0.00044	0.00044
Rel.MSE	1	0.6761	0.6773	0.6681	0.6602	0.6694	0.6693
$h = 8$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00025	0.00025	0.00025	0.00025	0.00025	0.00025	0.00025
Rel.MSE	1	0.9895	0.9888	0.9873	0.9966	1.0163	1.0140
$h = 12$	AR(4)	FL and AR(4)	FC and AR(4)	LB and AR(4)	UN and AR(4)	MC and AR(4)	ML and AR(4)
MSE	0.00073	0.00065	0.00065	0.00065	0.00065	0.00067	0.00067
Rel.MSE	1	0.8906	0.8901	0.8876	0.8871	0.9090	0.9081

Relative-MSE (Rel.MSE): Rel.MSE(factor-based) = MSE(factor-based)/MSE(AR(4)).

TABLE 4.18: Diebold and Mariano (2002)[a] factor-based against the benchmark

$h-$month-ahead Out-of Sample Forecasting - Forecast Combination - Granger and Ramanathan (pair-wise)
m_1 (model one) vs m_2 (model two)
y^{r,m_1}_{t+h} vs y^{r,m_2}_{t+h}; $l^{m_1} = (y^{r,m_1}_{t+h} - \widehat{y}^{r,m_1}_{t+h})^2$ and $l^{m_2} = (y^{r,m_2}_{t+h} - \widehat{y}^{r,m_2}_{t+h})^2$
$d = l^{m_1} - l^{m_2}$; $d_{t+1} = c + u_{t+1}$, where c is a constant
$H_0 : c = 0$ against $H_1^{m_1} : c < 0$ or $H_1^{m_2} : c > 0$

$h = 1$	DM-Stats	p-value	Winner	$h = 2$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-0.81	0.42	-	FL and AR(4) vs AR(4)	-1.04	0.30	-
FC and AR(4) vs AR(4)	-0.81	0.42	-	FC and AR(4) vs AR(4)	-1.04	0.30	-
LB and AR(4) vs AR(4)	-0.69	0.49	-	LB and AR(4) vs AR(4)	-0.94	0.35	-
UN and AR(4) vs AR(4)	-0.35	0.72	-	UN and AR(4) vs AR(4)	-0.63	0.52	-
MC and AR(4) vs AR(4)	-0.75	0.45	-	MC and AR(4) vs AR(4)	-1.93	0.06	MC and AR(4)
ML and AR(4) vs AR(4)	-0.74	0.46	-	ML and AR(4) vs AR(4)	-1.93	0.06	ML and AR(4)
$h = 4$	DM-Stats	p-value	Winner	$h = 6$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-0.85	0.39	-	FL and AR(4) vs AR(4)	-1.64	0.11	-
FC and AR(4) vs AR(4)	-0.88	0.38	-	FC and AR(4) vs AR(4)	-1.64	0.11	-
LB and AR(4) vs AR(4)	-1.03	0.13	-	LB and AR(4) vs AR(4)	-1.73	0.09	LB and AR(4)
UN and AR(4) vs AR(4)	-1.53	0.1870	-	UN and AR(4) vs AR(4)	-1.82	0.08	UN and AR(4)
MC and AR(4) vs AR(4)	-1.54	0.13	-	MC and AR(4) vs AR(4)	-1.59	0.12	-
ML and AR(4) vs AR(4)	-1.56	0.13	-	ML and AR(4) vs AR(4)	-1.59	0.12	-
$h = 8$	DM-Stats	p-value	Winner	$h = 12$	DM-Stats	p-value	Winner
FL and AR(4) vs AR(4)	-0.39	0.69	-	FL and AR(4) vs AR(4)	-0.71	0.48	-
FC and AR(4) vs AR(4)	-0.42	0.67	-	FC and AR(4) vs AR(4)	-0.72	0.48	-
LB and AR(4) vs AR(4)	-0.58	0.56	-	LB and AR(4) vs AR(4)	-0.75	0.46	-
UN and AR(4) vs AR(4)	-0.15	0.88	-	UN and AR(4) vs AR(4)	-0.78	0.44	-
MC and AR(4) vs AR(4)	0.74	0.46	-	MC and AR(4) vs AR(4)	-0.56	0.58	-
ML and AR(4) vs AR(4)	0.67	0.50	-	ML and AR(4) vs AR(4)	-0.57	0.57	-

[a]Newey and West (1987) HAC results of $d = l^{m_1} - l^{m_2}$ regressed on a constant, as explained in section 2.3
If a calculated t-ratio is statistically significant and with a negative constant results in m_1 defeating m_2, and vice-versa;
if the test is not significant, nothing can be said about m_1 and m_2 defeating each other.

Table 4.19: Point Forecasts (in %)

generated from the winners of the forecast combination using the variance-covariance method (Table 4.16) of Bates and Granger (1969); results are also displayed in Figures 4.1-4.3

	History	two-month-ahead		four-month-ahead		six-month-ahead					
		MC	ML	ML	UN	FL	FC	LB	UN	ML	MC
Jan-07	15.30	15.32	15.32								
Feb	15.90	15.43	15.43								
Mar	16.60	16.01	16.01	15.94	15.94						
Apr	16.90	16.64	16.64	16.80	16.76						
May	16.40	16.79	16.79	17.05	17.13	17.03	17.03	17.04	17.06	17.07	17.07
Jun	15.90	16.20	16.20	16.44	16.46	16.52	16.52	16.52	16.52	16.49	16.49
Jul	15.70	15.74	15.74	15.85	15.90	15.91	15.91	15.91	15.90	15.95	15.95
Aug	15.60	15.73	15.73	15.59	15.56	15.68	15.68	15.66	15.65	15.65	15.67
Sept	15.50	15.63	15.63	15.47	15.57	15.5	15.50	15.51	15.53	15.56	15.56
Oct	15.00	15.51	15.51	15.49	15.47	15.25	15.25	15.26	15.29	15.26	15.26
Nov	14.60	15.11	15.11	14.74	14.79	14.66	14.65	14.65	14.69	14.76	14.76
Dec	14.20	14.69	14.69	14.46	14.44	14.42	14.42	14.42	14.42	14.4	14.4
Jan-08	14.20	14.29	14.29	14.39	14.35	14.3	14.30	14.30	14.32	14.35	14.35
Feb	14.50	14.32	14.32	14.36	14.39	14.45	14.45	14.45	14.47	14.45	14.45
Mar	15.00	14.59	14.59	14.58	14.61	14.5	14.50	14.50	14.51	14.48	14.48
Apr	15.00	15.03	15.03	15.09	15.11	14.84	14.84	14.86	14.87	14.9	14.9
May	14.80	14.95	14.95	15.03	15.08	14.97	14.97	14.95	14.96	14.98	14.98
Jun	14.60	14.65	14.65	14.84	14.85	14.79	14.79	14.79	14.79	14.80	14.8
Jul	14.60	14.53	14.53	14.53	14.59	14.67	14.68	14.68	14.7	14.74	14.73
Aug	14.50	14.58	14.58	14.55	14.65	14.66	14.66	14.66	14.67	14.68	14.68
Sept	14.10	14.48	14.48	14.55	14.58	14.41	14.41	14.37	14.4	14.43	14.44
Oct	13.40	14.08	14.08	14.00	14.02	13.86	13.86	13.88	13.9	13.69	13.96
Nov	13.00	13.39	13.39	13.19	13.23	13.18	13.18	13.19	13.19	13.21	13.21
Dec	12.70	13.13	13.13	12.88	12.89	12.87	12.87	12.87	12.89	12.91	12.9
Jan-09	13.10	12.88	12.88	12.75	12.78	12.78	12.78	12.78	12.8	12.83	12.83

The estimation window uses data until 2006:11 and the forecasting window up to 2009:01.

TABLE 4.20: Correlation Coefficient Matrix of the Estimated Factors

	FC	FL	LB	UN	MC	ML
FC	1					
FL	0.9995	1				
LB	0.9670	0.9622	1			
UN	-0.8762	-0.8668	-0.9550	1		
MC	-0.3166	-0.3267	-0.1273	0.0584	1	
ML	-0.3305	-0.3420	-0.1365	0.0619	0.9965	1

TABLE 4.21: Discrepancy Statistics ($S_{\hat{F}_i,\hat{F}_j}$) amongst the Space Spanned of the Estimated Factors

	FC	FL	LB	UN	MC	ML
FC	1					
FL	0.9990	1				
LB	0.9351	0.9258	1			
UN	0.7677	0.7513	0.9119	1		
MC	0.1002	0.1067	0.0162	0.0038	1	
ML	0.1092	0.1170	0.0186	0.0034	0.9931	1

Appendix E: Figures

Figures 4.1-4.3 are forecast results in level from forecast combination for two-, four-, and six-month-ahead forecast using the variance covariance method of Bates and Granger (1969). These are the good results we have achieved as displayed in Table 4.16. In Figure 4.1, the forecast starts at 2007:01, in Figure 4.2 at 2007:03 and in Figure 4.3 at 2007:05. In Figures 4.4-4.9 are plots of the first estimated factor of each block, following the information criteria, against the variable that display the highest commonality for that specific factor; variables in Figures 4.4-4.9 are treated as described in Appendix A.

Figure 4.1: Forecast results in level from forecast combination for two-month-ahead

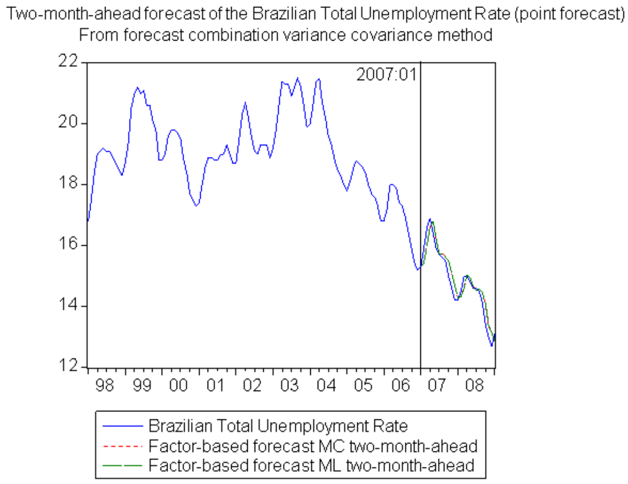


Figure 4.2: Forecast results in level from forecast combination for four-month-ahead

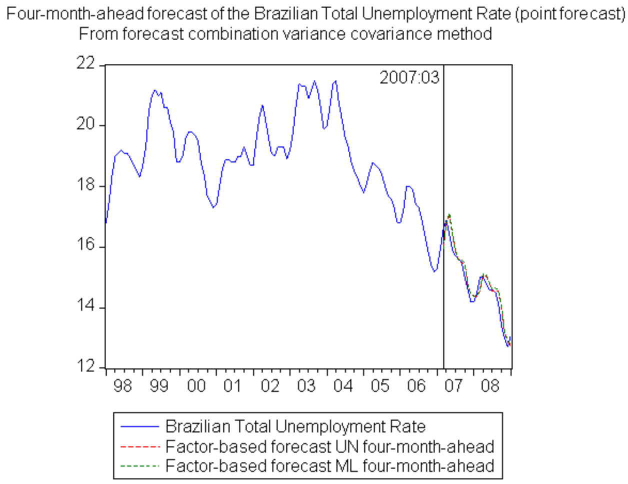


FIGURE 4.3: Forecast results in level from forecast combination for six-month-ahead

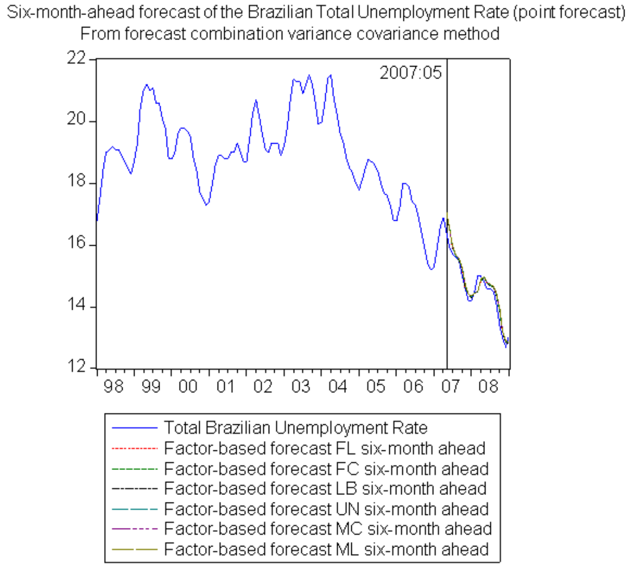


FIGURE 4.4: FL Estimated Factor vs Brazilian Unemployment Rate

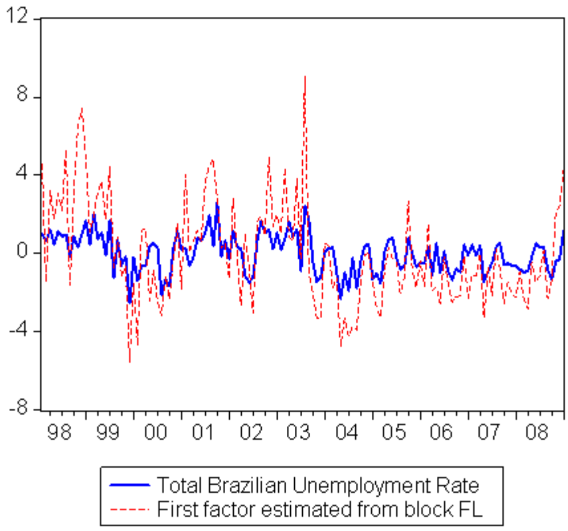


Figure 4.5: FC Estimated Factor vs Brazilian Unemployment Rate

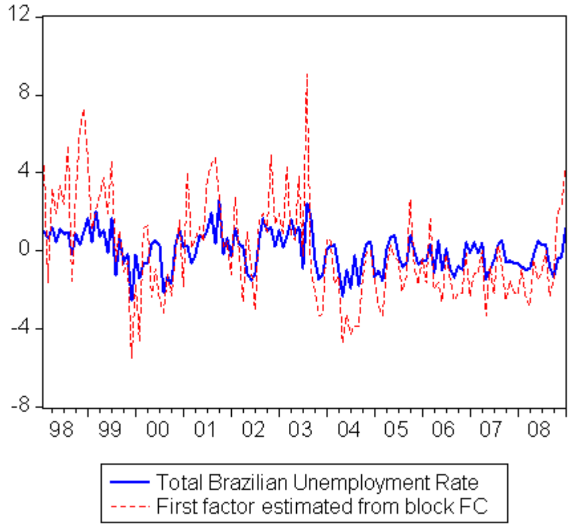


Figure 4.6: LB Estimated Factor vs Brazilian Unemployment Rate

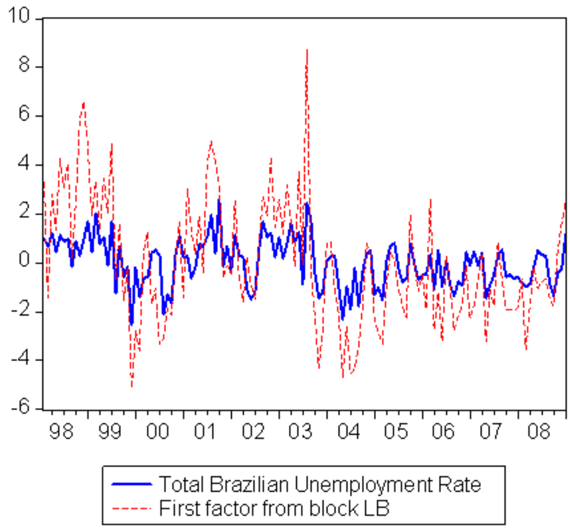


Figure 4.7: UN Estimated Factor vs Brazilian Unemployment Rate

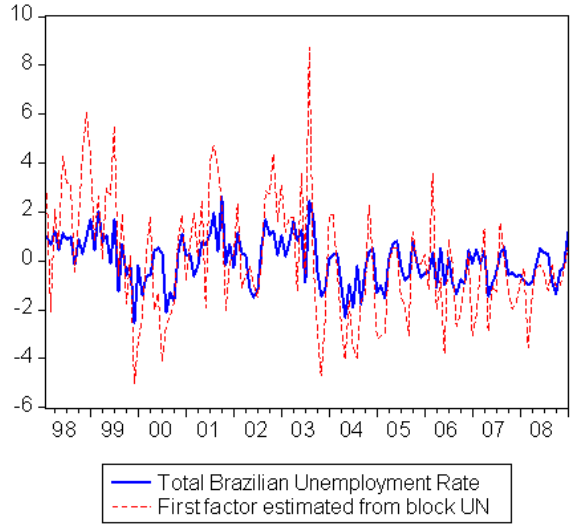


Figure 4.8: ML Estimated Factor vs Brazilian Nominal Interest Rate

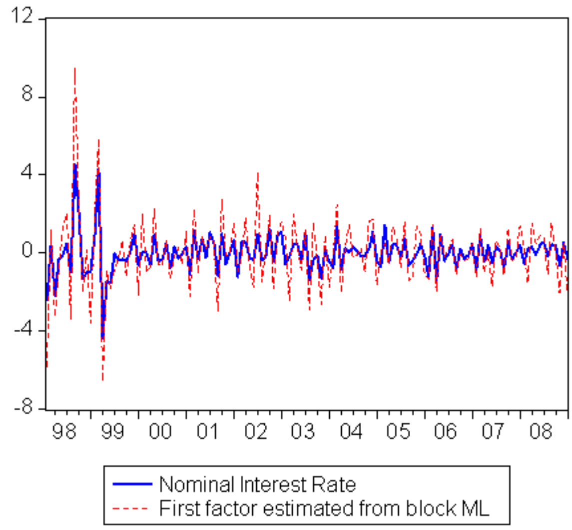
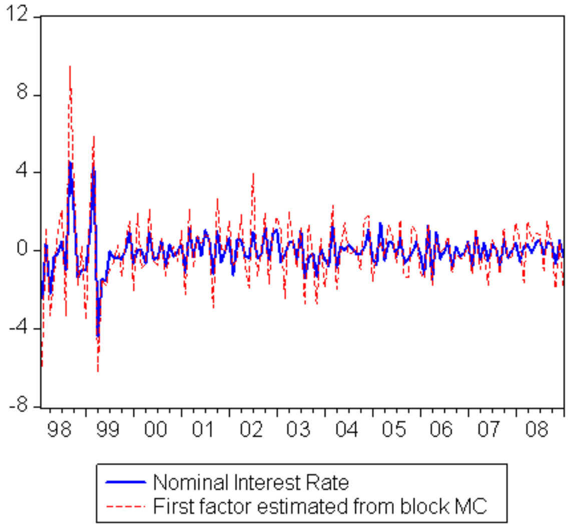


Figure 4.9: MC Estimated Factor vs Brazilian Nominal Interest Rate



Chapter 5

Conclusion

In this thesis we contributed to the literature on factor models by offering practitioners some new techniques designed to improve the modelling of a large data set as a factor model with the potential presence of structural breaks. In addition to this, we also applied factor models for forecasting purposes in the context of the Brazilian unemployment rate. The various toolkits we propose in this thesis rely on the approximate factor model setting recently explored in the work of Bai (2003) and Bai and Ng (2002).

In chapter two our goal was to propose an alternative approach to determine the number of factors in large approximate factor models under nonlinear structures. The methodology expands on the information criteria proposed by Bai and Ng (2002) by uncovering the time dimension of the factor structure and applying it recursively at each point in the time dimension, rather than applying it once in the entire sample. Then, a simple method to adjust for the overestimation of the number of factors is offered. It is a modified estimator to achieve a more parsimonious number of factors. Overall, our approach aims to enable the detection of changes in the factor structures, likely to result in changes in the number of factors along the same factor model. We suggested that our method should be used in the entire sample, without the need for splitting the sample at predetermined dates. In this manner, a practitioner could estimate the factor space and use our approach to uncover nonlinearities of a factor model, without running the risk of inadvertently overestimating the number of factors.

In chapter three we offered two tools to investigate nonlinearities in factor models. The first one is a nonlinear information criterion that enables one to determine the correct number of factors in either a linear or a nonlinear factor model. The nonlinearities take place in the factor loadings only, without changes in the number of factors. The second tool is understood as a decision rule to decide whether a factor model is in fact linear or nonlinear. Perhaps our contributions can complement to the Bai and Ng (2002) information criteria, which is designed only for linear datasets, and also to Breitung and Eickmeier (2011), Chen et al. (2012) and Han and Inoue (2012), who model tools to detect structural breaks in factor models. In this regard, our information criterion and decision rule can be tools that possibly complement to their contributions. A topic for parallel research is the case that occurs when the number of factors change at some point in the time dimension. For simplicity it has not been considered in this chapter.

The goal of the fourth and last chapter of this thesis was to present the forecast of the Brazilian unemployment rate using diffusion index. Good results are found for two- ,four- and six-month-ahead forecasts. We explored various methods of evaluation of forecast accuracy, forecast encompassing and combination. In our forecast exercise we have considered a linear model with constant coefficient throughout the sample (no time-varying coefficients). This perhaps brings a limitation to be explore in a future work since it is not unlikely that nonlinearities and/or structural breaks may occur. Moreover, our work can also be extended exploring the literature on factor analysis (i.e., estimation in frequency domain) or possibly newer versions of information criteria to determine the number of factors. Nevertheless, with the techniques explored in this chapter we have achieved satisfactory results when comparing our forecasts with the historical dataset.

Bibliography

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. New Jersey: Wiley Series in Probability and Statistics.

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica 71*(1), 135–171.

Bai, J. (2008). factor models. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence Intervals for Diffusion Index Forecasts Inference for Factor-Augmented Regressions. *Econometrica 3*(2), 89–163.

Bai, J. and S. Ng (2008a). Boosting Diffusion Indices. *Journal of Applied Econometrics 24*(4), 607–629.

Bai, J. and S. Ng (2008b). Forecasting Economic Time Series Using Targeted Predictors. *Journal of Econometrics 146*(2), 304–317.

Bai, J. and S. Ng (2008c). Large Dimensional Factor Analysis. *Foundations and Trend in Econometrics*, 89–163.

Bai, J. and S. Ng (2010). Principal Components Estimation and Identification of the Factors. Working paper available at http://www.columbia.edu/ sn2294/papers/rotate.pdf.

Bates, J. M. and C. W. J. Granger (1969). The Combination of Forecasts. *Operations Research 20*, 451–468.

Boivin, J. and S. Ng (2005). Understanding and Comparing Factor-Based Forecasts. *International Journal of Central Banking 1*(3).

Breitung, J. and S. Eickmeier (2006). Dynamic Factor Models. *Journal of the German Statistical Society 90*(1), 27–40.

Breitung, J. and S. Eickmeier (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics 163(1)*, 71–84.

Cagan, P. (1956). The Monetary Dynamics of Hyperinflation. In M. Friedman (Ed.), *Studies in the Quantinty Theory of Money*. Chicago: University of Chicago Press.

Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research 1*(2), 245–276.

Chamberlain, G. (1983). Funds, Factors, and Diversification in Arbitrage Pricing Models. *Econometrica 51*(5), 1305–23.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Market. *Econometrica 51*(5), 1281–304.

Chen, L., J. J. Dolado, and J. Gonzalo (2012). Detecting Big Structural Breaks in Large Factor Models. UC3, working papers.

Chong, Y. Y. and D. F. Hendry (1986). General Intelligence, Objectly Determined and Measured. *The Review of Economic Studies 53*(4), 671–690.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*(3), 591605.

Diebold, F. X. and J. A. Lopez (1996). Forecast Evaluation and Combination. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, Volume 14. Amsterdam: North-Holland.

Diebold, F. X. and R. S. Mariano (2002). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics 20*(1), 134–44.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Forni, M., M. Hallin, M. Lippi, and L. Reichilin (2000). The Generalized Dynamic-Factor Modes: Identification and Estimation. *The Review of Economic and Statistics 82*(4), 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The Generalized Factor Models: consistency and rates. *Journal of Econometrics 119*(2), 231–255.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model, one sided estimation and forecasting. *Journal of the Americam Statistical Association 100*, 830–840.

Forni, M. and L. Reichlin (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies 65*(3), 453–73.

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.* Amstendan: North-Holland.

Granger, C. W. J. and R. Ramanathan (1984). Improved Methods of Forecasting. *Journal of Forecasting 3*, 197–204.

Han, X. and A. Inoue (2012). Tests for Parameter instability in Dynamic Factor Models. working paper.

Koop, G. and S. M. Potter (1999). Dynamic asymmetries in U.S. unemployment. *Journal of Business and Economic Statistics 17*(3), 298–312.

Moura, M. L. and A. de Carvalho (2010). What can Taylor rules say about monetary policy in Latin America? *Journal of Macroeconomics 32*(1), 392–404.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–08.

Ng, S. and J. Boivin (2006). Are more data always better for factor analysis? *Journal of Econometrics* (132), 169–194.

Reichlin, L. (2003). Factor Models in Large Cross Section of Time Series. In S. M. Dewatripoint and L. Hansen (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Vol.111, 8th World Congress of the Econometric Society.* Cambridge: Cambridge University Press.

Ross, S. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory* (3), 341–360.

Salgado, M. J. S., M. G. P. Garcia, and M. C. Medeiros (2005). Monetary Policy During Brazils Real Plan: Estimating the Central Banks Reaction Function. *Revista Brasileira de Economia 59*(1), 61–79.

Sharpe, W. F. (1964). Capital Asset Prices - A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance XIX*(3), 425–442.

Sims, C. A., , and T. J. Sargent (1977). Business Cycle Modeling without Pretending to have too uch a priori Economic Theory. In D. J. Aigner and A. S. Goldberger (Eds.), *New Methods in Business Research.* Minneapolis: Federal Reserve Bank of Minneapolis.

Spearman, C. (1904). The American Journal of Psychology. *The Review of Economic Studies 15*(2), 201–292.

Stock, J. H. (2001). Forecasting Economic Time Series. In B. Baltagi (Ed.), *A Companion to Theoretical Econometrics.* Basil: Blackwell.

Stock, J. H. and M. W. Watson (1998). Diffusion Indexes. *NBER Working Papers 6702.*

Stock, J. H. and M. W. Watson (2002a). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics 20*(2), 147–162.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. NBER Working Papers 11467.

Stock, J. H. and M. W. Watson (2006). Forecasting with Many Predictors. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1. Elsevier.

Stock, J. H. and M. W. Watson (2008). Forecasting in dynamic factor models subject to structural instability. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry.* Oxford: Oxford University Press.

Stock, J. H. and M. W. Watson (2010). Dynamic Factor Models. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook of Economic Forecasting.* Oxford: Oxford University Press.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy 73*, 195–214.

Timmermann, A. (2006). Forecast combinations. In A. Timmermann (Ed.), *Handbook of Economic Forecasting*, Volume 1. Elsevier.