Dynamic Factor Analysis of High Dimensional Time Series: Application to Indian Economy

Project Report submitted

By

Soumya Ghatak

Reg.No 21MSMS01

Under the guidance of

Prof. Madhuchhanda Bhattacharjee

The project submitted in partial fulfillment of the requirement for the degree of Master of Science in Statistics

From 2021 To 2023

At

University Of Hyderabad
School Of Mathematics And Statistics
Prof. C R Road, Gachibowli, Hyderabad
Telengana, India - 500046

Certificate

This is to certify that the project work entitled "*Dynamic Factor Analysis of High Dimensional Time Series: Application to Indian Economy*" submitted by Mr. *Soumya Ghatak* is a bonafide project work done under my supervision. It is being submitted in partial fulfillment of the requirements of degree in Master of Science in Statistics, University of Hyderabad.

Prof. Madhuchhanda Bhattacharjee	Dean
School of Mathematics and Statistics	School of Mathematics and Statistics
University of Hyderabad, Telangana	University of Hyderabad, Telangana

Date: 23-05-2023
Place: Hyderabad

Declaration

I Soumya Ghatak hereby declared that the project work entitled "*Dynamic Factor Analysis of High Dimensional Time Series: Application to Indian Economy*" is an original record of studied and bonafide work carried out by me under the guidance of Prof. Madhuchhanda Bhattacharjee, School of Mathematics and Statistics, University of Hyderabad, Telangana, India and has not been submitted by me elsewhere for the award of any degree, diploma, title or recognition before.

Name: Soumya Ghatak

Roll No: 21MSMS01

School of Mathematics and Statistics

University of Hyderabad, Telangana

Acknowledgement

I take this opportunity to express my profound gratitude and deep regards to the following for their generous assistance and incorporation of fresh ideas in the completion of this Dissertation Paper.

I would like to thank my institution, University of Hyderabad for giving me chance to do this project.

Most importantly, I would also like to express my sincere gratitude towards my project guide Prof. Madhuchhanda Bhattacharjee for being an idealistic channel without whose paramount guidance and utmost care, this paper would not have been possible.

Lastly, I would like to thank each and every individual who directly or indirectly assisted me in the completion of this dissertation paper, especially my family members, friends and peers who supported me throughout my project.

Soumya Ghatak

21MSMS01

M.Sc. Statistics

School of Mathematics and Statistics

University of Hyderabad

Contents

Dynamic Factor Analysis of High Dimensional Time Series : Application to Indian Economy

Soumya Ghatak

May 23, 2023

Abstract

Classical dimensionality reduction techniques like PCA or classical factor models can not be used to reduce dimensions of time series data as the observations are not independent. In this report we have discussed a popular dimensionality reduction technique Dynamic Factor Model for high dimensional time series data. In this method a dynamic nature of the underlying factors is assumed. Here we also studied different properties of dynamic factor model along with some of the extensions of it. After that we used a small-scale dynamic factor model to model and forecast the Indian GDP growth using other economic variables. Finally we compare this dynamic factor model with other competing methods in various respects.

1 Introduction

In macroeconomics or finance area (also in many cases of biology, ecology etc.) we generally have large number of time series data which are correlated among themselves. In most of the cases we have time series data for 100s or even 1000s variables (indicators) but only for very few number of time points i.e. number of variables in many cases is much more than number of observed data. Analyzing those large number time series variables is computationally expensive and often it is difficult to interpret when there is so many variables. We need to find a lower dimensional representation of that time series data without loosing much of the information of the whole data. There are several statistical method for dimensionality reduction for multivariate data such as Principal Component Analysis, Classical Factor Analysis. But using those methods on time series data is not appropriate. One of the assumption of those methods is that the sample have to be independent of one another. But this assumption is not valid for time series data. Time series observations for different time points are not independent. To solve this problem Dynamic Factor Models are introduced. Similar to classical factor model here also we assume that the time series vector at a time point is described by a few numbers of latent factor variables (number of factor variables assumed to be much lesser than number of time series variables), but here we assume that the underlying factor variable follows some dynamic process.

2 Literature Review

Here we mainly discuss about three broad topics **Theoretical Development**, **Applications** and other **Competing Methods**.

2.1 Theoretical Development

Dynamic factor model first proposed by **Geweke (1977)[21]** in the paper "Dynamic Factor Analysis of Economic Time Series". He gave the dynamic nature in the classical factor models by introducing a dynamic nature of the unobserved factors. It was assumed that the co-movement of a large number of economic time series variables is determined by a few unobserved factors. In that paper the model is estimated using maximum likelihood method. **Shumway and Stoffer (1982)[25]** introduced EM algorithm for calculations of maximum likelihood estimators of the dynamic factor models in state space setup. In many cases specially in economic areas, there is a problem of missing value in time series data. **Bańbura and Modugno (2014)[5]** in the paper "Maximum Likelihood Estimation of Factor Models on Data Sets with Arbitrary Pattern of Missing Data" modified the estimation of dynamic factor model using maximum likelihood methods such that it can handle missing values in the data sets. In the paper "Determining the Number of Factors in Approximate Factor Models", **Bai and Ng(2002)[3]** discussed about choosing the appropriate number of factors. They proposed three measures to do so based on information criteria for estimating the number of factors. They also proposed some criteria such that the the estimated number of factors are asymptotically consistent. Another similar method for dimensionality reduction is principal component analysis. **Brillinger (1969)**[10], generalized the idea for principal component analysis for dependent data such as time series and introduced dynamic principal component analysis (DPCA). **Sumway and Stoffer (2000)[26]** discussed the concept of principal component analysis for time series data in the frequency domain using eigen value decomposition of spectral density matrix. Along with its use in dimensionality reduction in time series data it is also used to estimate the factors in dynamic factor model. Although maximum likelihood method works well for small number of series, for large number of series or when number of series is greater than number of observations this method does not work well, also some assumptions become very strict for large number of series. **Forni et al. (2000, 2004)**[18, 19], **Forni and Lippi (1997)**[16], **Forni and Reichlin (1998)[17]** introduced the idea for large approximate dynamic factor model for this kind of situation by relaxing some of the assumptions. In this situations non-parametric estimation methods of dynamic factor models using principal component analysis is proposed, **Chamberlain and Rothschild (1983)**[12], **Stock and Watson (2002)**[28]. **Doz, Giannone, Reichlin, (2011,2012)[14, 15]** proposed a two-step estimation method and also quasi likelihood estimation method for the large approximate dynamic factor model.

2.2 Applications

Early applications can be found in **Sargent and Sims (1977)**[23], where they showed that very few dynamic factors can explain the co movements of various macroeconomic variables in US. **Stock and Watson (1989)[27]** introduces dynamic factor models to make new economic indicators. In their paper "New Indexes of Coincident and Leading Economic Indicator ", they used one factor dynamic factor model with stationary macro economic variables for coincident economic indicator, the estimated factor acts as a coincident economic indicator. For leading indicator they proposed using an vector auto regression with a leading variable and the estimated factors from the dynamic factor model. In that paper they also discussed about the choice of variables choosing variables from different relevant groups using bivariate relationship with response variable and step-wise regression before using them in dynamic factor model. In the paper "Forecasting with many predictors", **Stock and Watson(2006)[29]** discusses about using dynamic factor model for accurate forecasting of economic variables. The main idea is to forecast the unobserved factors which follow some dynamic relation (e.g. VAR), and then using those factor forecast to forecast the desired variable using the factor model.

One of the most common application of dynamic factor model is to model and forecast/now cast GDP growth or business cycles. **Cuevas and Quilis (2011)[13]** in their paper "A factor analysis for the Spanish economy" used one factor dynamic factor model to forecast (now cast) performance of Spanish economy in very short term. They used DFM to construct a coincident indicator of Spanish economy using the single estimated factor and then using that they model and forecast the Spanish GDP growth. In the paper "On the design of data sets for forecasting with dynamic factor models" by **Gerhard Rünstler (2016)**[20], they modeled the GDP growth of euro area, France and Germany. They first made a variable selection using step-wise regression and then using those selected variables are used in a two factor dynamic factor model. Those estimated factors are then used to model GDP growth using a ols regression. **Chernis and Sekkel (2017)[11]** used Dynamic Factor Model for now-casting Canadian GDP and also showed that it outperformed the univariate benchmark methods as well as other competing methods. **Sezgin & Kinay (2010)[24]** used the dynamic factor model to study the effect of global economic crisis of 2008-2009 in Turkey. In context of Indian economy dynamic factor models also have been used , such as **Bhadury, Ghosh, Kumar (2020)[8]** used dynamic factor model to now cast Indian GDP growth, **Iyer, Sen Gupta (2019**)[22] used dynamic factor model to now cast Indian economy and also analyzed the effect of rainfall in Indian economy.

2.3 Competing Methods

Apart from those popular methods, **Claudia Becker and Roland Fried (2001)[7]** generalized the Sliced Inverse Regression (SIR) for dimensionality reduction problem for high dimensional time series data in regression setup. **Babikir , Mwambi (2016)** combined Dynamic Factor Model (DFM) and Artificial Neural Network (ANN) and introduced Factor Augmented Artificial Neural Network (FAANN) for more accurate forecast. They applied the model to model deposit rate, gold mining share prices and Long term interest rate using of South Africa using large monthly data, and they observed the out of sample forecast accuracy of this model is better than using only DFM or just auto regressive methods. Another dimensionality reduction technique is Random Projection method which is formulated on the fact that the distance between the data points in the reduced space is more or less similar to the distance between points in the original space (Johnson-Lindenstrauss theorem). This method is relatively new and used in the fields of image and text data , **Ella Bingham , Heikki Mannila (2001)**[9]. But its application in time series or data sets with dependency is not found in the literature.

3 Methodology

3.1 Mathematical Formulation

Let $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{nt})$ be a n-dimensional stationary multivariate time series process with mean 0 and unit variance at time t. We observe that variable for a finite period of time $t = 1, 2, \ldots, T$. In factor models (classical or static/dynamic) we write the observations as,

$$x_{it} = \chi_{it} + \varepsilon_{it}$$

for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. Also χ_{it} and ε_{it} independent of each other for all i and t . The term χ_{it} is called the common component and ε_{it} is the error component. Further the common component is written as $\chi_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \ldots + \lambda_{ir}F_{rt} = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ir})\ \boldsymbol{F_t} = \lambda_i^{'}\boldsymbol{F}_t$. So we can write $\boldsymbol{x}_t = \Lambda\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$, for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, where $\boldsymbol{F}_t$ is the $r \times 1$ vector of unobserved factors, $\Lambda = \begin{bmatrix} \lambda_1^{'} \\ \lambda_2^{'} \\ \vdots \\ \lambda_n^{'} \end{bmatrix}$ is a $n \times r$ matrix of factor loadings and $\boldsymbol{\varepsilon}_t$ is the $n \times 1$ error vector. Now this version of factor models are termed as static factor model or static representation of factor model. Although the factors $\boldsymbol{F}_t$ may be static or may follow some dynamic relationship that is a dynamic process, the term "static" is used because of the static relationship between $\boldsymbol{x}_t$ and the unobserved factor $\boldsymbol{F}_t$ (i.e. at time point t, $\boldsymbol{x}_t$ depends only on the factor $\boldsymbol{F}_t$ at time t) (**Bai and Ng 2007**)[4].

Now we will discuss the dynamic representation of the model. Here also the the we write, $x_{it} = \chi_{it} + \varepsilon_{it}$, but the common term χ_{it} is assumed to be sum of common shocks of dimension q of time t and also its lags up to s. That is the common component is written as (**Barhoumi, Darne and Ferrara ,2013**[6]),

$$
\begin{aligned}
\chi_{it} = {} & b_{i1}^0 u_{1,t} + \cdots\cdots + b_{i1}^s u_{1,t-s} \\
& + b_{i2}^0 u_{1,2} + \cdots\cdots + b_{i2}^s u_{2,t-s} \\
& \vdots \\
& + b_{iq}^0 u_{1,q} + \cdots\cdots + b_{i2}^s u_{2,t-q}
\end{aligned}
$$

$$
\begin{aligned}
& = (b_{i1}^0 L^0 + \cdots\cdots + b_{i1}^s L^s) u_{1,t} \\
& + \cdots + (b_{iq}^0 L^0 + \cdots\cdots + b_{iq}^s L^s) u_{q,t} \\
& = b_{i1}(L) u_{1,t} + b_{i2}(L) u_{2,t} + \ldots + b_{iq}(L) u_{q,t} \\
& \begin{bmatrix} b_{i1}(L) & b_{i2}(L) & \cdots & b_{iq}(L) \end{bmatrix} \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{q,t} \end{bmatrix} \\
& = \boldsymbol{B}_i(L) \boldsymbol{u}_t
\end{aligned}
$$

Where, $b_{ij}(L) = b_{ij}^0 L^0 + \cdots\cdots + b_{ij}^s L^s$ is the lag polynomial of order s, $\boldsymbol{B}_i(L) = \begin{bmatrix} b_{i1}(L) & b_{i2}(L) & \cdots & b_{iq}(L) \end{bmatrix}$ is the $1 \times q$ vector of lag polynomials of order s and $\boldsymbol{u}_t$ is the shock at time t. **Bai and Ng, 2007[4]** used a decomposition $\boldsymbol{B}_i(L) = \boldsymbol{\lambda}_i^{'}(L)\boldsymbol{C}(L)$, where $\boldsymbol{\lambda}_i^{'}(L)$ is a $1 \times q$ vector of lag polynomial of degree s, and $\boldsymbol{C}(L)$ is a $q \times q$ matrix. Then we can write the common component as $\chi_{it} = \boldsymbol{\lambda}_i^{'}(L)\boldsymbol{C}(L)\boldsymbol{u}_t = \boldsymbol{\lambda}_i^{'}(L)\boldsymbol{f}_t$, or this could be written as,

$$\chi_{it} = \boldsymbol{\lambda}_{i0}^{'}\boldsymbol{f}_t + \boldsymbol{\lambda}_{i1}^{'}\boldsymbol{f}_{t-1} + \cdots + \boldsymbol{\lambda}_{is}^{'}\boldsymbol{f}_{t-s}$$

or

$$x_{it} = \boldsymbol{\lambda}_{i0}^{'}\boldsymbol{f}_t + \boldsymbol{\lambda}_{i1}^{'}\boldsymbol{f}_{t-1} + \cdots + \boldsymbol{\lambda}_{is}^{'}\boldsymbol{f}_{t-s} + \varepsilon_{it}$$

with $\boldsymbol{\lambda}_i(L) = \boldsymbol{\lambda}_{i0} + \boldsymbol{\lambda}_{i1}L + \cdots + \boldsymbol{\lambda}_{is}L^s$. Here $\boldsymbol{f}_t$ is the dynamic factor and the model is dynamic factor model as at time t, x_{it} does depend not only on $\boldsymbol{f}_t$but also it's lags up to s. **Bai and Ng (2007)[4]** showed that for every dynamic factor model there is a static representation of that factor model. Next we will discuss that. Write $\boldsymbol{\Lambda}_i = \begin{bmatrix} \boldsymbol{\lambda}_{i0} \\ \boldsymbol{\lambda}_{i1} \\ \vdots \\ \boldsymbol{\lambda}_{is} \end{bmatrix}$ and $\boldsymbol{F}_t = \begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \vdots \\ \boldsymbol{f}_{t-s} \end{bmatrix}$ both are $q(s+1) \times 1$, and we can write $x_{it} = \boldsymbol{\Lambda}_i^{'}\boldsymbol{F}_t + \varepsilon_{it}$, as a static representation of the dynamic factor model. $r = q(s+1)$ is the number of static factor, unless number of lags in dynamic model in not 0 ($s = 0$), number of static factor (r) is always larger than number of dynamic factor (q). In addition to this the dynamic factors $\boldsymbol{f}_t$ also follow some dynamic process, here we will consider that $\boldsymbol{f}_t$follows an auto regressive process of order $h < \infty$. That is $(I_q - B_1L - \cdots - B_hL^h)\boldsymbol{f}_t = \boldsymbol{\epsilon}_t$, where $B_j$is the $q \times q$ coefficient matrix of the jth auto regressive model. Now define $k = max(h, s)$, and also define $B_{h+1} = B_{h+2} = \cdots = B_k = 0_{q\times q}$. Then we can write

$$\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \vdots \\ \boldsymbol{f}_{t-k} \end{bmatrix} = \begin{bmatrix} B_1 & B_2 & B_3 & \cdots & \cdots & \cdots & B_{k+1} \\ I_q & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_q & \cdots & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \vdots \\ \boldsymbol{f}_{t-k-1} \end{bmatrix} + \begin{bmatrix} I_q \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{\epsilon}_t$$

$$\boldsymbol{F}_t^* = A\boldsymbol{F}_{t-1}^* + \boldsymbol{u}_t$$

Where, $\boldsymbol{F}_t^* = \begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \vdots \\ \boldsymbol{f}_{t-k} \end{bmatrix}$, $A = \begin{bmatrix} B_1 & B_2 & B_3 & \cdots & \cdots & \cdots & B_{k+1} \\ I_q & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_q & \cdots & 0 \end{bmatrix}$, and $\boldsymbol{u}_t = \begin{bmatrix} I_q \\ 0 \\ \vdots \\ 0 \end{bmatrix} \boldsymbol{\epsilon}_t$.

If $k = max(h, s) = s$, that is $s > h$, then $\boldsymbol{F}_t^* = \boldsymbol{F}_t$. And then the static factor $\boldsymbol{F}_t$ follows a VAR(1) dynamic process.

Also for $k = max(h, s) = h$, that is $s < h$, it can be shown that $\boldsymbol{F}_t$ follows a VAR. In general (**Bai and Ng 2007**)[4] $\boldsymbol{F}_t$follows VAR(p), that is $\boldsymbol{F}_t = \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i} + \boldsymbol{u}_t$, with $p = max(1, h - s)$. For $s < h$ case we will illustrate the case with $s = 2$ and $h = 4$. Then $\boldsymbol{F}_t = \begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \end{bmatrix}$, and $\boldsymbol{f}_t = B_1\boldsymbol{f}_{t-1} + B_2\boldsymbol{f}_{t-2} + B_3\boldsymbol{f}_{t-3} + B_4\boldsymbol{f}_{t-4} + \boldsymbol{\epsilon}_t$. Then it can be written,

$$\boldsymbol{F}_t = \begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \end{bmatrix} = \begin{bmatrix} B_1 & B_2 & B_3 \\ I_q & 0 & 0 \\ 0 & I_q & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \boldsymbol{f}_{t-3} \end{bmatrix} + \begin{bmatrix} 0 & 0 & B_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-2} \\ \boldsymbol{f}_{t-3} \\ \boldsymbol{f}_{t-4} \end{bmatrix} + \begin{bmatrix} I_q \\ 0 \\ 0 \end{bmatrix} \boldsymbol{\epsilon}_t$$

$$= A_1\boldsymbol{F}_{t-1} + A_2\boldsymbol{F}_{t-2} + \boldsymbol{u}_t$$

So finally we can write,

$$\boldsymbol{x}_t = \Lambda\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \dots, T$$

,where $\Lambda = \begin{bmatrix} \boldsymbol{\Lambda}_1^{'} \\ \boldsymbol{\Lambda}_2^{'} \\ \vdots \\ \boldsymbol{\Lambda}_n^{'} \end{bmatrix}$, $\boldsymbol{F}_t = (F_{1t}, F_{2t}, \dots, F_{rt})$, $t = 1, 2, \dots, T$ is the unobserved common static factors at time t. Λ is a $n \times r$ matrix of factor loadings and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{nt})$ is the error or noise component which is uncorrelated with the underlying factor at all lags i.e. $cov(\boldsymbol{\varepsilon}_t, \boldsymbol{F}_{t-h}) = \boldsymbol{0}_{n\times t}$, $h = 0, \pm1, \pm2, \dots$. We also assume the error $\boldsymbol{\varepsilon}_t$, $t = 1, 2, \dots, T$ is n-dimensional normal with mean $\boldsymbol{0}$ and variance covariance matrix $R_{n\times n}$ that is, at first we will consider $R_{n\times n}$ as a diagonal matrix. That is the error representing different series are uncorrelated, but this is a very strict assumption. Later we will discuss that in approximate factor model we relax that assumption and we allow small correlation between the error components. $\boldsymbol{\varepsilon}_t$'s can be assumed to be serially uncorrelated that is,

$$cov(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t+h}) = \begin{cases} R_{n\times n} & h = 0 \\ \\ 0_{n\times n} & h = \pm 1, \pm 2, \dots \end{cases}$$

but, it also may be assumed that there is a serial dependency of the error components with its lags, for example the error components may follow an AR(1) process, that is $\varepsilon_{it} = a\varepsilon_{it-1} + e_i$, we will discuss this case separately in the estimation portion.

Also as shown before, the static factors follow a VAR(p) process:

$$\boldsymbol{F}_t = \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \dots, T$$

where, A_i's are the $r \times r$ coefficient matrices of that VAR process. Here also the assume the noise $\boldsymbol{u}_t$, $t = 1, 2, \dots, T$ is serially uncorrelated n-dimensional normal with mean $\mathbf{0}$ and variance covariane matrix $Q_{r\times r}$. The first equation is called measurement or observation equation and the second equation is called state or process equation. We cannot observe the process described by the underlying factors $\boldsymbol{F}_t$ directly which is our primary interest, but as $\boldsymbol{x}_t$.

3.2 Estimation

3.2.1 For no missing values and no serial correlation

Here for simplicity we will describe the estimation process for $p = 1$. That is the unobserved common factor follows a VAR(1) process . The state or process equation for that , is $\boldsymbol{F}_t = A\boldsymbol{F}_{t-1} + \boldsymbol{u}_t$, $t = 1, 2, \ldots, T$ .In that defined model we have four unknown parameters to be estimated . Define $\Theta = \{\Lambda, A, R, Q\}$as the set of unknown parameters . We will use maximum likelihood methods to estimate those parameters , but we need to first calculate the likelihood . Note that $\boldsymbol{x}_{1,}\boldsymbol{x}_2, \ldots, \boldsymbol{x}_T$ are not independent similarly , $\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_T$are also not independent . But the conditioning on the factor variables , $\boldsymbol{x}_1|\boldsymbol{F}_1, \boldsymbol{x}_2|\boldsymbol{F}_2, \ldots, \boldsymbol{x}_T|\boldsymbol{F}_T$are independent random variables . Similarly given that we know the distribution of initial factor variable $\boldsymbol{F}_0$, conditioning on the previous factor variable $\boldsymbol{F}_1|\boldsymbol{F}_0, \boldsymbol{F}_2|\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T|\boldsymbol{F}_{T-1}$ are independent random variables . Now define $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$ and $\mathscr{F} = \{\boldsymbol{F}_0, \boldsymbol{F}_1, \ldots, \boldsymbol{F}_T\}$.We also only consider the case where we do not have any missing data , that is we have every observations in X. Assume that $\boldsymbol{F}_0 \sim N_r(\boldsymbol{\mu}, \Sigma)$ the likelihood function can be written as :

$$L(\Theta; X, \mathscr{F}) = g(\boldsymbol{F}_0; \boldsymbol{\mu}, \Sigma) \prod_{t=1}^{T} g(\boldsymbol{x}_t|\boldsymbol{F}_t; \Lambda, R) g(\boldsymbol{F}_t|\boldsymbol{F}_{t-1}; A, Q)$$

Ignoring the constant terms the log-likelihood function can be written as :

$$
\begin{aligned}
l(\Theta; X, \mathscr{F}) = &-\frac{1}{2}log|\Sigma| - \frac{1}{2}(\boldsymbol{F}_0 - \boldsymbol{\mu})^{'}\Sigma^{-1}(\boldsymbol{F}_0 - \boldsymbol{\mu}) \\
&- \frac{T}{2}log|Q| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{F}_t - A\boldsymbol{F}_{t-1})^{'}Q^{-1}(\boldsymbol{F}_t - A\boldsymbol{F}_{t-1}) \\
&- \frac{T}{2}log|R| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{x}_t - \Lambda\boldsymbol{F}_t)^{'}R^{-1}(\boldsymbol{x}_t - \Lambda\boldsymbol{F}_t)
\end{aligned}
$$

$$
\begin{aligned}
= &-\frac{1}{2}log|\Sigma| - \frac{1}{2}(\boldsymbol{F}_0 - \boldsymbol{\mu})^{'}\Sigma^{-1}(\boldsymbol{F}_0 - \boldsymbol{\mu}) \\
&- \frac{T}{2}log|Q| - \frac{1}{2}tr(Q^{-1}\sum_{t=1}^{T}(\boldsymbol{F}_t - A\boldsymbol{F}_{t-1})(\boldsymbol{F}_t - A\boldsymbol{F}_{t-1})^{'}) \\
&- \frac{T}{2}log|R| - \frac{1}{2}tr(R^{-1}\sum_{t=1}^{T}(\boldsymbol{x}_t - \Lambda\boldsymbol{F}_t)(\boldsymbol{x}_t - \Lambda\boldsymbol{F}_t)^{'})
\end{aligned}
$$

Now we do not observe the the factor variables $\boldsymbol{F}_t$‘s . So we can not directly optimize the log-likelihood to estimate Θ. Here we will use the Expectation-Maximization algorithm , which given an initial value of estimate iteratively finds the optimal estimate using two steps in each iteration **Shumway and Stoffer (1982)**. The steps are Expectation step or E-step and Maximization step or M-step . Given that at jth iteration the coefficient are $\Theta(j)$, then in the $j+1$th step ,

E-step : we take expectation of the log-likelihood given $\Theta(j)$, i.e. we calculate :

$$e(\Theta|\Theta(j)) = E_{\Theta(j)}[l(\Theta; X, \mathscr{F})|Y]$$

M-step : Here we estimate the parameters for $j+1$th stage by maximizing $e(\Theta|\Theta(j))$ with respect to Θ :

$$\Theta(j+1) = arg\max_{\Theta} e(\Theta|\Theta(j))$$

In the E-step $E_{\Theta(j)}[.|X]$ means we are taking expectation given that we have all the available data (no missing data) .

To get an estimate of Λ at $j+1$th iteration we need to find derivative of $E_{\Theta(j)}[l(\Theta; X, \mathscr{F})|X]$ with respect to $\mathit{\Lambda}$.

$$\begin{aligned}
\frac{\partial}{\partial \mathit{\Lambda}} E_{\Theta(j)}[l(\Theta; X, \mathscr{F})|Y] &= -\frac{1}{2}\frac{\partial}{\partial \mathit{\Lambda}} tr\{R^{-1}\sum_{t=1}^{T} E_{\Theta(j)}[(\boldsymbol{x}_t - \Lambda \boldsymbol{F}_t)(\boldsymbol{x}_t - \Lambda \boldsymbol{F}_t)^{'}|X]\} \\
&= -\frac{1}{2}\frac{\partial}{\partial \mathit{\Lambda}} tr\{R^{-1}\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{x}_t\boldsymbol{x}_t^{'} - \boldsymbol{x}_t\boldsymbol{F}_t^{'}\Lambda^{'} - \Lambda\boldsymbol{F}_t\boldsymbol{x}_t^{'} + \Lambda\boldsymbol{F}_t\boldsymbol{F}_t^{'}\Lambda^{'}|X]\} \\
&= -R^{-1}\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{x}_t\boldsymbol{F}_t^{'}|X] + R^{-1}\Lambda\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t\boldsymbol{F}_t^{'}|X]
\end{aligned}$$

Equating that above form to 0 , we get

$$\Lambda(j+1) = (\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{x}_t\boldsymbol{F}_t^{'}|X])(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t\boldsymbol{F}_t^{'}|X])^{-1}$$

In similar manner we also get ,

$$A(j+1) = (\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{F}_{t-1}^{'}|X])(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_{t-1} \boldsymbol{F}_{t-1}^{'}|X])^{-1}$$

Now at $j+1$th stage we have two updated estimates $\Lambda(j+1)$ and $A(j+1)$, using these two we can find a plug-in estimate for the other two covariance parameters . Note that now the error/noise terms can be written as :

$$\boldsymbol{\varepsilon}_t = \boldsymbol{x}_t - \Lambda(j+1)\boldsymbol{F}_t + \quad , \quad t = 1, 2, \ldots, T$$

and , $$\boldsymbol{u}_t = \boldsymbol{F}_t - A(j+1)\boldsymbol{F}_{t-1} \, , \, t = 1, 2, \ldots, T \, .$$

So we use the expected sum of squared residuals as the plug-in estimate of covariance matrices R and Q :

$$\begin{aligned} R(j+1) &= \frac{1}{T} \sum_{t=1}^{T} E_{\Theta(j)}[(\boldsymbol{x}_t - \Lambda(j+1)\boldsymbol{F}_t)(\boldsymbol{x}_t - \Lambda(j+1)\boldsymbol{F}_t)^{'}|X] \\ &= \frac{1}{T}(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{x}_t \boldsymbol{x}_t^{'}|X] - \Lambda(j+1) \sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{x}_t^{'}|X]) \end{aligned}$$

and ,

$$\begin{aligned} Q(j+1) &= \frac{1}{T} \sum_{t=1}^{T} E_{\Theta(j)}[(\boldsymbol{F}_t - A(j+1)\boldsymbol{F}_{t-1})(\boldsymbol{F}_t - A(j+1)\boldsymbol{F}_{t-1})^{'}|X] \\ &= \frac{1}{T}(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{F}_t^{'}|X] - A(j+1) \sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_{t-1} \boldsymbol{F}_t^{'}|X]) \end{aligned}$$

Sometimes it is assumed that R is a diagonal matrix , that is we assume that components of $\boldsymbol{y}_t$is only related to each other through the unobserved factor $\boldsymbol{f}_t$. In that case the estimate of R at $j+1$th stage is

$$R(j+1) = diag(\frac{1}{T}(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{x}_t\boldsymbol{x}_t^{'}|X] - \Lambda(j+1)\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t\boldsymbol{x}_t^{'}|X]))$$

Given we have all observable data the we can estimate the conditional moments used in the above estimates by , $E_{\Theta(j)}[\boldsymbol{x}_t\boldsymbol{x}_t^{'}|X] = \boldsymbol{x}_t\boldsymbol{x}_t^{'}$ and $E_{\Theta(j)}[\boldsymbol{F}_t\boldsymbol{x}_t^{'}|X] = E_{\Theta(j)}[\boldsymbol{F}_t|X]\boldsymbol{x}_t^{'}$. But the conditional moments of the unobserved factor variables such as $E_{\Theta(j)}[\boldsymbol{F}_t|X]$, $E_{\Theta(j)}[\boldsymbol{F}_t\boldsymbol{F}_t^{'}|X]$,$E_{\Theta(j)}[\boldsymbol{F}_{t-1}\boldsymbol{F}_{t-1}^{'}|X]$ and $E_{\Theta(j)}[\boldsymbol{F}_{t-1}\boldsymbol{F}_t^{'}|X]$ can not be estimated directly as we do not observe those factors . To estimate those we will use Kalman filter and smoother . Below we discuss the results and the procedure for estimating those conditional estimates of the unobserved factor variables .

Define $\boldsymbol{F}_{t|s}^{j} = E_{\Theta(j)}(\boldsymbol{F}_t|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_s)$ as the estimate of the unobserved factor given the observed data $\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_s\}$, and the error covariance $P_{t_1,t_2|s}^{j} = E_{\Theta(j)}((\boldsymbol{F}_{t_1} - \boldsymbol{F}_{t_1|s}^{j})(\boldsymbol{F}_{t_2} - \boldsymbol{F}_{t_2|s}^{j})^{'})$ for $s,t,t_1,t_2 = 1,2,\ldots,T$ (we write $P_{t|s}^{j}$if $t_1 = t_2 = t$). Now we will write those conditional moments of the unobserved factors using the defined notation .

We can write $\boldsymbol{F}_t = \boldsymbol{F}_{t|T}^{j} + (\boldsymbol{F}_t - \boldsymbol{F}_{t|s}^{j})$ then ,

$$E_{\Theta(j)}[\boldsymbol{F}_t\boldsymbol{F}_t^{'}|X] = E_{\Theta(j)}[(\boldsymbol{F}^j_{t|T} + (\boldsymbol{F}_t - \boldsymbol{F}^j_{t|s}))(\boldsymbol{F}^j_{t|T} + (\boldsymbol{F}_t - \boldsymbol{F}^j_{t|s}))^{'}|X]$$
$$= \boldsymbol{F}^j_{t|T}(\boldsymbol{F}^j_{t|T})^{'} + E_{\Theta(j)}((\boldsymbol{F}_t - \boldsymbol{F}^j_{t|T})(\boldsymbol{F}_t - \boldsymbol{F}^j_{t|T})^{'})$$
$$= \boldsymbol{F}^j_{t|T}(\boldsymbol{F}^j_{t|T})^{'} + P^j_{t|T}$$

Similarly , the other conditional moments can be written as ,

$$E_{\Theta(j)}[\boldsymbol{F}_{t-1}\boldsymbol{F}_{t-1}^{'}|X] = \boldsymbol{F}^j_{t-1|T}(\boldsymbol{F}^j_{t-1|T})^{'} + P^j_{t-1|T}$$
$$E_{\Theta(j)}[\boldsymbol{F}_{t-1}\boldsymbol{F}_t^{'}|X] = \boldsymbol{F}^j_{t-1|T}(\boldsymbol{F}^j_{t|T})^{'} + P^j_{t-1,t|T}$$

Now we discuss how to estimate $\boldsymbol{F}^j_{t|T}$, $P^j_{t|T}$,$\boldsymbol{F}^j_{t-1|T}$ and $P^j_{t-1,t|T}$. Similar to least square estimate here also we estimate $\boldsymbol{F}^j_{t|T}$ such that the error covariance is minimized . At the beginning of $j+1$th iteration we have the estimates from the jth iteration $\Theta(j) = \{\Lambda(j), A(j), R(j), Q(j)\}$. The factor model at the beginning of $j+1$th iteration is ,

$$\boldsymbol{x}_t = \Lambda(j)\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \underset{\sim}{iid} N(\boldsymbol{0}, R(j))$$

$$\boldsymbol{F_t} = A(j)\boldsymbol{F}_{t-i} + \boldsymbol{u}_t \quad \boldsymbol{u}_t \underset{\sim}{iid} N(\boldsymbol{0}, Q(j))$$

for $t = 1, 2, \ldots, T$. Now we use the properties from **Shumway and Stroffer (2000)** , to get those estimates .

Property 1 (Filter)

Given initial estimate $\boldsymbol{F}_1 = \boldsymbol{F}^j_{1|0}$ and $P_0 = P^j_{1|0}$,

the prediction stage ,

$$\boldsymbol{F}^j_{t|t-1} = A(j)\boldsymbol{F}^j_{t-1|t-1}$$

$$P^j_{t|t-1} = A(j)P^j_{t-1|t-1}A(j)' + Q(j)$$

the update stage ,

$$\boldsymbol{F}^j_{t|t} = \boldsymbol{F}^j_{t|t-1} + K^j_t(\boldsymbol{y}_t - \Lambda(j)\boldsymbol{F}^j_{t|t-1})$$

$$P^j_{t|t} = [I - K^j_t\Lambda(j)]P^j_{t|t-1}$$

The term , $K^j_t = P^j_{t|t-1}\Lambda(j)'[\Lambda(j)P^j_{t|t-1}\Lambda(j)' + R(j)]^{-1}$is the Kalman gain ,$t = 1, 2, \ldots, T$.

Property 2 (Smoother)

Using the property one we can get $\boldsymbol{F}^j_{T|T}$ and $P^j_{T|T}$from , we can go backward and estimate ,

$$\boldsymbol{F}^j_{t-1|T} = \boldsymbol{F}^j_{t-1|t-1} + J^j_{t-1}(\boldsymbol{F}^j_{t|T} - \boldsymbol{F}^j_{t|t-1})$$

$$P^j_{t-1|T} = P^j_{t-1|t-1} + J^j_{t-1}(P^j_{t|T} - P^j_{t|t-1})[J^j_{t-1}]'$$

where $J^j_{t-1} = P^j_{t-1|t-1}A(j)[P^j_{t-1|t-1}]^{-1}$,$t = T, T-1, \ldots, 1$

Property 3

$$P^j_{T|,T-1T} = [I - K^j_T\Lambda(j)]A(j)P^j_{T-1|T-1}$$

$$P^j_{t-1,t-2|T} = P^j_{t-1|t-1}[J^j_{t-2}]' + J^j_{t-1}(P^j_{t,t-1|T} - \Lambda(j)P^j_{t-1|t-1})[J^j_{t-2}]'$$

where J^j_t and K^j_t are defined as similar to as in previous two properties . And $P^j_{T-1|T-1}$,$P^j_{t-1|t-1}$ can be computed previous properties .

Now , using prediction and update stage of Property 1 iteratively we can calculate of all one step predictions of the form $\boldsymbol{F}^j_{t|t-1}$ and estimates $\boldsymbol{F}^j_{t|t}$ in the following way .

$$\boldsymbol{F}_1 = \boldsymbol{F}^j_{1|0} \overset{update}{\rightarrow} \boldsymbol{F}^j_{1|1} \overset{predict}{\rightarrow} \boldsymbol{F}^j_{2|1} \overset{update}{\rightarrow} \boldsymbol{F}^j_{2|2} \dashrightarrow\dashrightarrow \boldsymbol{F}^j_{T|T-1} \overset{update}{\rightarrow} \boldsymbol{F}^j_{T|T}$$

similarly ,

$$P_1 = P^j_{1|0} \overset{update}{\rightarrow} P^j_{1|1} \overset{predict}{\rightarrow} P^j_{2|1} \overset{update}{\rightarrow} P^j_{2|2} \dashrightarrow\dashrightarrow P^j_{T|T-1} \overset{update}{\rightarrow} P^j_{T|T}$$

After getting those estimates we can calculate $\boldsymbol{F}^j_{t|T}$ and $P^j_{t|T}$ for $t = 1, 2, \cdots, T-$

1 using Property 2 . After that using property 3 calculate $P^{j}_{t-1,t|T}$ for $t = 1, 2, \cdots, T-1$. Now using those estimates the conditional expectations of the unobserved factors can be estimated . And finally the the estimates for the $j+1$ th stage can be calculated , $\Theta(j+1) = \{\Lambda(j+1), A(j+1), R(j+1), Q(j+1)\}$.

So in conclusion we first pass some initial estimates at the beginning of the EM algorithm . At each stage , using the estimated parameters from the previous stage , three levels of Kalman filter and smoother is applied itteratively to estimate the conditional moments of the unobserved factors and then estimate the parameters . This process continues until certain the log-likelihood converges or in notation it continues until $\frac{2|l(\Theta(j);X,\mathscr{F})-l(\Theta(j-1;X,\mathscr{F})|}{(l(\Theta(j);X,\mathscr{F})+l(\Theta(j-1;X,\mathscr{F}))} \leq tolarnce\,limit$ (this tolerance limit can be per-specified) .

3.2.2 In presence of arbitrary missing values

In the last section we discussed the estimation method for no missing observation case . But in practice we may have data with missing observations , this is specially noticeable in macroeconomic data sets . In the paper Maximum Likelihood Estimation of Factor Models on Data Sets With Arbitrary Pattern of Missing Data ,**Bańbura and Modugno (2010)** proposed a method for estimating dynamic factor model by maximum likelihood method in presence of missing observations . Let $\Omega_T$be the available data . In the last section we considered $\Omega_T = X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$, that is whole data is available . In presence of missing data $\Omega_T \subset X$.

In the last section we estimated the conditional moments required for the estimation using $E_{\Theta(j)}[\boldsymbol{x}_t\boldsymbol{x}_t^{'}|\Omega_T = X] = \boldsymbol{x}_t\boldsymbol{x}_t^{'}$, $E_{\Theta(j)}[\boldsymbol{F}_t\boldsymbol{x}_t^{'}|\Omega_T = X] = E_{\Theta(j)}[\boldsymbol{F}_t|\Omega_T = X]\boldsymbol{x}_t^{'}$, but here as $\Omega_T \subset X$ we can not use those as $\boldsymbol{x}_t$contain missing observations ,and so we can not use the estimates of the parameters from the previous section . To deal with that they introduced a $n \times n$ diagonal matrix $W_t = diag(w_{11,t}, w_{22,t}, \ldots w_{nn,t})$,(selection matrix) where $w_{ii,t} = \begin{cases} 0 & if\, x_{i,t}\, is\, missing \\ 1 & otherwise \end{cases}$ (Note that for no missing data case ,i.e, for $\Omega_T = X$, $W_t = I_n$) for $t = 1, 2, \ldots, T$. Also note that the estimates of A and Q does not involve the term $\boldsymbol{x}_t$, so they are not effected by the presence of missing value , only the estimates of Λ and R is effected by the presence of missing value on the data set .

The modified estimates of Λ and R as given by Bańbura and Modugno (2010) are

$$vec(\Lambda(j+1)) = (\sum_{t=1}^{T} E_{\theta(j)}[\boldsymbol{F}_t\boldsymbol{F}_t'|\Omega_T] \otimes W_t)^{-1} vec(\sum_{t=1}^{T} W_t\boldsymbol{x}_t E_{\theta(j)}[\boldsymbol{F}_t'|\Omega_T])$$

and ,

$$
\begin{aligned}
R(j+1) = diag(\frac{1}{T}\sum_{t=1}^{T}(W_t\boldsymbol{x}_t\boldsymbol{x}_t'W_t' - W_t\boldsymbol{x}_tE_{\theta(j)}[\boldsymbol{F}_t'|\Omega_t]\Lambda(j+1)'W_t \\
- W_t\Lambda(j+1)E_{\theta(j)}[\boldsymbol{F}_t'|\Omega_t]\boldsymbol{x}_t'W_t \\
+ W_t\Lambda(j+1)E_{\theta(j)}[\boldsymbol{F}_t\boldsymbol{F}_t'|\Omega_t]\Lambda(j+1)'W_t \\
+ (I-W_t)R(j)(I-W_t)))
\end{aligned}
$$

,where the notations are similar as previous , $\otimes$ denote Kronecker product and $vec(M)$ denote vectorized form $(pq \times 1)$ of the matrix $M_{p\times q}$.

3.2.3 In presence of serial correlation

We can extend the dynamic factor model mentioned above by considering serial correlation between the error components . Thus the model can be written as ,

$$
\begin{aligned}
\boldsymbol{x}_t &= \Lambda\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \\
\boldsymbol{F}_t &= A\boldsymbol{F}_{t-1} + \boldsymbol{u}_t \quad , \quad t = 1,2,\ldots,T
\end{aligned}
$$

where all the notations and the distributional assumptions are same as previously mentioned model , but here we consider the components of $\boldsymbol{\varepsilon}_t$,that is $\varepsilon_{i,t}$follow an AR(1) process . There are several ways to modify the estimation part such that it considers the serially correlated errors , but here we will describe the method proposed by **Bańbura and Modugno (2010)** which can also be used in presence of missing observations .

Here the error component $\varepsilon_{i,t}$ is decomposed in the way ,

$$\varepsilon_{i,t} = \tilde{\varepsilon}_{i,t} + \zeta_{i,t}$$

where $\zeta_{i,t} \underset{\sim}{iid} N(0, \kappa)$,(where κ is very small) and we consider $\tilde{\varepsilon}_{i,t}$ an AR(1) process ,

$$\tilde{\varepsilon}_{i,t} = \alpha_i \tilde{\varepsilon}_{i,t-1} + e_{i,t}$$

where $e_{i,t} \underset{\sim}{iid} N(0, \sigma_i^2)$, for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, and α_i is the auto regressive coefficient .

Define $\boldsymbol{\zeta}_t = (\zeta_{1,t}, \zeta_{2,t}, \ldots, \zeta_{n,t})'$ and $\widetilde{\boldsymbol{\varepsilon}}_t = (\tilde{\varepsilon}_{1,t}, \tilde{\varepsilon}_{2,t}, \ldots, \tilde{\varepsilon}_{n,t})'$. That is $\boldsymbol{\varepsilon}_t = \widetilde{\boldsymbol{\varepsilon}}_t + \boldsymbol{\zeta}_t$, where $\boldsymbol{\zeta}_t \sim N_n(0, \widetilde{R} = \begin{bmatrix} \kappa & 0 & \cdots & 0 \\ 0 & \kappa & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \kappa \end{bmatrix})$

Now using the previously defined factor model and the decomposition if the error terms we can write ,

$$
\begin{aligned}
\boldsymbol{x}_t &= \Lambda \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \\
&= \Lambda \boldsymbol{F}_t + \widetilde{\boldsymbol{\varepsilon}}_t + \boldsymbol{\zeta}_t \\
&= \begin{bmatrix} \Lambda & I_n \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_t \\ \widetilde{\boldsymbol{\varepsilon}}_t \end{bmatrix} + \boldsymbol{\zeta}_t \\
&= \widetilde{\Lambda} \widetilde{\boldsymbol{F}}_t + \boldsymbol{\zeta}_t
\end{aligned}
$$

And similarly ,

$$
\begin{aligned}
\widetilde{\boldsymbol{F}}_t = \begin{bmatrix} \boldsymbol{F}_t \\ \widetilde{\boldsymbol{\varepsilon}}_t \end{bmatrix} &= \begin{bmatrix} A\boldsymbol{F}_{t-1} + \boldsymbol{u}_t \\ \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{bmatrix} \widetilde{\boldsymbol{\varepsilon}}_{t-1} + \boldsymbol{e}_t \end{bmatrix} \\
&= \begin{bmatrix} A & 0 \\ 0 & diag(\alpha_1, \alpha_2, \ldots, \alpha_n) \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{t-1} \\ \widetilde{\boldsymbol{\varepsilon}}_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{e}_t \end{bmatrix} \\
&= \widetilde{A} \widetilde{\boldsymbol{F}}_{t-1} + \widetilde{\boldsymbol{u}}_t
\end{aligned}
$$

,Where $\boldsymbol{\zeta}_t \sim N_n(\mathbf{0}, \widetilde{R} = \begin{bmatrix} \kappa & 0 & \cdots & 0 \\ 0 & \kappa & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \kappa \end{bmatrix})$ and

$\widetilde{\boldsymbol{u}}_t \sim N_{r+n}(\mathbf{0}, \widetilde{Q} = \begin{bmatrix} Q & \mathbf{0}' \\ \mathbf{0} & diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2) \end{bmatrix})$.

Now the dynamic factor model with serially correlated errors is written in a similar form as before where errors are not serially correlated .Thus the parameters can be estimated using the previous likelihood based method with a little modification .

3.3 Determining r and p

3.3.1 Determining number of factors r

There are several ways to determine the number of factors r in the model . Here we will mention two methods . Both of them ignore the dynamic part of the factors that is the VAR part and approximately determines the number of factors .

First we could decide from plotting the scree-plot . Ignoring the dynamic nature we can calculate the sample variance covariance matrix $\hat{\Sigma_{\boldsymbol{x}}} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{x}_t - \bar{\boldsymbol{x}})(\boldsymbol{x}_t - \bar{\boldsymbol{x}})'$, where $\bar{\boldsymbol{x}}$ is the sample mean vector and then calculate the eigen values and plot as scree-plot (PCA approach) .

Another approach is based on information criteria . It is done by minimizing penalized likelihood . In the paper **Bai and Ng(2002)** they introduced the following information criteria , $IC(r) = ln(V_r(\hat{\Lambda}, \hat{F})) + r.g(n, T)$, where $ln(V_r(\hat{\Lambda}, \hat{F})) = \frac{1}{nT} \sum_{t=1}^{T} (\boldsymbol{x}_t - \hat{\Lambda}\hat{\boldsymbol{F}}_t)'(\boldsymbol{x}_t - \hat{\Lambda}\hat{\boldsymbol{F}}_t)$ is the least square objective function and $g(n, T)$ is the penalty factor such that $g(n, T) \rightarrow 0$and $min(n, T)g(n, T) \rightarrow 0$ as $n, T \rightarrow \infty$. Bai and Ng also shown that the approximate value $\hat{r}$ we get using this method is an consistent estimator for the true value of r , given some assumptions (**Bai and Ng ,2000**).

Some specific information criteria for different penalty factor given in that paper are :

$$IC_{r_1} = ln(V_r(\hat{\Lambda}, \hat{F})) + r.(\frac{n+T}{nT}) + ln(\frac{nT}{n+T})$$

$$IC_{r_2} = ln(V_r(\hat{\Lambda}, \hat{F})) + r.(\frac{n+T}{nT}) + ln(min(n, T))$$

$$IC_{r_3} = ln(V_r(\hat{\Lambda}, \hat{F})) + r\frac{ln(min(n, T))}{min(n, T)}$$

3.3.2 Determining the order of VAR part p

After determining an approximate estimate of number of factors we have to estimate the order of the dynamic VAR part . Let the estimated number of factors be $\hat{r}$, then we can get approximate estimate factors from simple principal component analysis with out considering the dynamic nature of the factors and then can fit the VAR model with order p , $(p = 1, 2, ...)$ using those approximately estimated factors .And we can choose appropriate choice of p using various information criteria based measures .Let the VAR equation be ,

$$\hat{\boldsymbol{F}}_t = \sum_{i=1}^{p} A_i \hat{\boldsymbol{F}}_{t-i} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \ldots, T$$

where $\hat{\boldsymbol{F}}$ is the $\hat{r} \times 1$ approximately estimated factors . After fitting the model define ,

$$\hat{u}_t = \hat{\boldsymbol{F}}_t - \sum_{i=1}^{p} \hat{A}_i \hat{\boldsymbol{F}}_{t-i}$$

and ,

$$\hat{\Sigma}_u(p) = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t^{'}$$

Then some of the information criteria based measures are ,

$$\begin{aligned}
AIC(p) &= ln(|\hat{\Sigma}_u(p)|) + \frac{2}{T} p\hat{r}^2 \\
BIC(p) &= ln(|\hat{\Sigma}_u(p)|) + \frac{ln(T)}{T} p\hat{r}^2 \\
HQ(p) &= ln(|\hat{\Sigma}_u(p)|) + \frac{2ln(ln(T))}{T} p\hat{r}^2 \\
SC(p) &= ln(|\hat{\Sigma}_u(p)|) + \frac{ln(ln(T))}{T} p\hat{r}^2
\end{aligned}$$

etc . Based on those we can estimate p . Another way is to after estimating $\hat{r}$, fit the dynamic factor model for various p , and choose the optimal p based on the desired performance of the dynamic factor model .

Although the above mentioned methods can be used to determine r and p , but in practice in many cases they are chosen on judgmental basis or on the basis of the application for which the dynamic factor model will be used . For example , while using for construction of some economic indicator based on many macroeconomic variables $r = 1$ is taken in many literature.

3.4 Large approximate dynamic factor models and alternative estimation methods

Until now , we have consider this model ,

$$\boldsymbol{x}_t = \Lambda \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$$

and ,

$$\boldsymbol{F}_t = \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \ldots, T$$

,where $\boldsymbol{x}_t$ is an $n \times$ 1vector of observations at time t , i.e. we have total n series or variables . The components of $\boldsymbol{\varepsilon}_t$ may or may not be auto correlated .The error terms $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ both serially uncorrelated with mean 0 and variance covariance matrix R and Q respectively , moreover we also assumed that those errors are normal random variable . One strong assumption we have considered is that R is a diagonal $n \times n$ matrix , that is the errors for different series or variables are uncorrelated . And we have discussed the maximum likelihood estimation using EM algorithm and Kalman filter for this model . These all assumptions and estimation methods are okay if the number of series , n is relatively small and $n < T$. In literature this model is called strict or exact dynamic factor model , due to its restricted assumptions .

But in many situations we have very large number of series / variables even more than the number of observations in the data $(n > T)$. In those cases the above assumptions and the estimation methods are not applicable . The diagonality assumption of the error variance covariance matrix R is too strong for this context when n is too large . So in this case we relax this assumption and allow weakly mutually correlated errors , that is we consider R to be non diagonal here . As the n is too large , (even $n > T$) the number of parameters in the model is also too high . In those cases (specially when $n > T$) maximum likelihood estimates are unachievable (**Bai and Ng, 2003**) . Also the the maximum likelihood estimates are asymptotically converges when T is very large , but it may not converge for very large n , when T is fixed . For these limitations of the strict dynamic factor model , the approximate dynamic factor model is introduced for large n.

As the maximum likelihood estimation method is not applicable here , we will discuss two different methods of estimation for approximate dynamic factor models . First is time domain method , “Two step approach” by **Doz, Giannone and Reichlin (2011)** , and the second is in frequency domain using dynamic PCA .

The two step approach

Consider the model of the form ,

$$\boldsymbol{x}_t = \Lambda \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$$

$$\boldsymbol{F}_t = \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \ldots, T$$

, with variance covariance , matrix of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ as as R and Q respectively , also assuming that R is not diagonal .

Step 1

Here usual principal component analysis is applied to estimate the factor loading matrix Λ , and the unobserved factors $\boldsymbol{F}_t$. Let $\hat{\Lambda}$and $\hat{\boldsymbol{F}}_t$be the estimates using pca . Then the estimates $\hat{\boldsymbol{F}}_t$ are fitted to a VAR(p) model to estimate the VAR coefficients A_i's , for $i = 1(1)p$. Then using the residuals $\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{x}_t - \widehat{\Lambda}\widehat{\boldsymbol{F}_t}$ we calculate the estimates of R .

Step 2

Here we assume that $Q = I_r$ and using the estimates , $\hat{\Lambda}, \widehat{R}, \widehat{A_i}$for $i = 1(1)p$, we update the estimate of factor from step 1 using Kalman filter (i.e. $\boldsymbol{F}_{t|T}$) ,as discussed previously in section 1.2.1 .

They also showed that the estimated factor generated by this method is $min(\sqrt{n}, \sqrt{T})$ consistent , under some assumptions about the eigen values of the variance covariance of $\boldsymbol{x}_t$ as a function of n (i.e. they defined $\boldsymbol{x}_t^n =$

$(x_{1t}, ..., x_{nt})$,and $cov(\boldsymbol{x}_t^n) = \Sigma_n$as a function of n . They stated three assumptions about the eigen values of Σ_n as $n \to \infty$,**Doz, Giannone and Reichlin (2011)**) .

Estimation in frequency domain

Brillinger (1981) first generalized the principal component analysis to frequency domain . He used spectral density matrix instead of variance covariance matrix , and using the eigen values of the spectral density matrix obtained the principal components . For a n dimensional stationary time series $\boldsymbol{x}_t$the $n \times n$ spectral density matrix for the frequency $\omega \in [-\pi, \pi]$ is defined as , $\Sigma_{\boldsymbol{x}}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\omega} \Gamma_{\boldsymbol{x}}(k)$, where $\Gamma_{\boldsymbol{x}}(k) = E(\boldsymbol{x}_t \boldsymbol{x}_{t-k}^{'})$ the cross covariance matrix of lag k . The spectral density matrix is obtained by doing a Fourier transformation of the cross correlation matrix (they are called Fourier pairs) , they contain the same information of the stationary time series process $\boldsymbol{x}_t$. The matrix $\Sigma_{\boldsymbol{x}}(\omega)$ is hermitian matrix , so the eigen values of this matrix are all real and the eigen vectors are orthogonal to each other .

Forni et al. (2000) proposed an procedure for estimating the dynamic principal components and by that the common components of the factor model .

Step 1 For the sample $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T$ (each is a n dimensional vector) , calculate the spectral density matrix $\widehat{\Sigma}_{\boldsymbol{x}}(\omega_h) = \frac{1}{2\pi} \sum_{k=-M}^{M} w_k e_w^{-ik\omega_h} \widehat{\Gamma}_{\boldsymbol{x}}(k)$, where $\omega_h = \frac{2\pi h}{2M+1}$,$h = -M, ..., M$, $w_k = 1 - \frac{|k|}{M+1}$ are the weights and M

is the window of the window . The cross covariance matrix is estimated by $\widehat{\Gamma}_{\boldsymbol{x}}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k}(\boldsymbol{x}_t - \bar{\boldsymbol{x}})(\boldsymbol{x}_{t+k} - \bar{\boldsymbol{x}})'$. This estimate of the spectral density matrix is consistent for $M \to \infty$ and $\frac{M}{T} \to 0$ as $T \to \infty$. For choice of M **Forni et al. (2000)** stated that $M = \frac{2}{3}T^{\frac{1}{3}}$. That is if $T = 1000$ then roughly $h = -7, -6, ..., 6, 7$.

Step 2 For $h = -M, ..., M$, for each $\widehat{\Sigma}_{\boldsymbol{x}}(\omega_h)$ compute the eigen vectors corresponding to the largest r eigen values , say $\boldsymbol{\lambda}_1(\omega_h), ..., \boldsymbol{\lambda}_r(\omega_h)$.

Step 3 The first r dynamic principal components are computed as $\widehat{pc}_{it} = \boldsymbol{\gamma}_i(L)'\boldsymbol{x}_t$, for $i = 1, 2..., r$. Where $\boldsymbol{\gamma}_i(L) = \sum_{k=-M}^{M}\boldsymbol{\lambda}_{ik}L^k$, where L is the lag operator and $\boldsymbol{\lambda}_{ik} = \frac{1}{2M+1}\sum_{h=-M}^{M}\boldsymbol{\lambda}_j(\omega_h)e^{-ik\omega_h}$. And $\widehat{\boldsymbol{pc}}_t = (\widehat{pc}_{1t}, \widehat{pc}_{2t}, ..., \widehat{pc}_{rt})'$

Step 4 Then we regress $\boldsymbol{x}_t$ on $\widehat{\boldsymbol{pc}}_{t-l}$, where $l = -q, ..., p$.

$$\boldsymbol{x}_t = \Psi_{-q}\widehat{\boldsymbol{pc}}_{t+q} + ... + \Psi_p\widehat{\boldsymbol{pc}}_{t-p}$$

and then estimated common component of $\boldsymbol{x}_t$ as the fitted value .

$$\widehat{\boldsymbol{\chi}_t} = \widehat{\Psi}_{-q}\widehat{\boldsymbol{pc}}_{t+q} + ... + \widehat{\Psi}_p\widehat{\boldsymbol{pc}}_{t-p}$$

Where the lags p, q have to be selected. The residuals can be estimated as $\widehat{\boldsymbol{\varepsilon}_t} = \boldsymbol{x}_t - \widehat{\boldsymbol{\chi}_t}$, and using the residuals we can estimate $\widehat{R}$.

The above frequency domain approach is basically the dynamic principal component analysis. **Shumway and Stoffer (2000)** also proposed an method in frequency domain which is analogous to the classical factor model for time series. For the model $\boldsymbol{x}_t = \Lambda \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$ (taking variance both side, $\Sigma_{\boldsymbol{x}} = \Lambda^{'} \mathit{\Lambda} + R$, assuming $E(\boldsymbol{F}_t) = \mathbf{0}$,and $E(\boldsymbol{F}_t \boldsymbol{F}_t^{'}) = I_r$in classical factor model notation), instead of taking variance covariance matrix, they considered the frequency density matrix i.e. $\Sigma_{\boldsymbol{x}}(\omega) = \Lambda(\omega)^{'} \mathit{\Lambda}(\omega) + R(\omega)$, where $\mathit{\Lambda}(\omega)$ is a $r \times n$ complex valued matrix and $R(\omega)$ is a real , positive definite diagonal matrix. After that similar to classical factor model they estimated the factor loadings by eigen values of the $\Sigma_{\boldsymbol{x}}(\omega)$. Apart from these methods there is also quasi maximum likelihood method for this type of large approximate dynamic factor model **Doz, Giannone and Reichlin (2012).**

4 Simulation Study

4.1 Data Generation

Here we did simulation studies to get an idea about the accuracy of the dynamic factor model with respect to the number of variables/series n and the number of observations available T. We will consider the following three models for simulation study.

Model1 with $r = 4$ and $p = 1$,

$$\boldsymbol{x}_t = \Lambda \begin{bmatrix} F_{1t} \\ F_{2t} \\ F_{3t} \\ F_{4t} \end{bmatrix} + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$$

$$\begin{bmatrix} F_{1t} \\ F_{2t} \\ F_{3t} \\ F_{4t} \end{bmatrix} = A_1 \begin{bmatrix} F_{1,t-1} \\ F_{2,t-1} \\ F_{3,t-1} \\ F_{4,t-1} \end{bmatrix} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \ldots, T$$

And the variance covariance matrix of $\boldsymbol{u}_t$being Q

Model2 $r = 3$ and $p = 1$

$$\boldsymbol{x}_t = \Lambda \begin{bmatrix} F_{1,t} \\ F_{2,t} \\ F_{3,t} \end{bmatrix} + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$$

$$\begin{bmatrix} F_{1,t} \\ F_{2,t} \\ F_{3,t} \end{bmatrix} = A_1 \begin{bmatrix} F_{1,t-1} \\ F_{2,t-1} \\ F_{3,t-1} \end{bmatrix} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \ldots, T$$

And the variance covariance matrix of $\boldsymbol{u}_t$being Q

Model3 $r = 5$ and $p = 1$

$$\boldsymbol{x}_t = \Lambda \begin{bmatrix} F_{1,t} \\ F_{2,t} \\ F_{3,t} \\ F_{4,t} \\ F_{5,t} \end{bmatrix} + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$$

$$\begin{bmatrix} F_{1,t} \\ F_{2,t} \\ F_{3,t} \\ F_{4,t} \\ F_{5,t} \end{bmatrix} = A_1 \begin{bmatrix} F_{1,t-1} \\ F_{2,t-1} \\ F_{3,t-1} \\ F_{4,t-1} \\ F_{5,t-1} \end{bmatrix} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \ldots, T$$

And the variance covariance matrix of $\boldsymbol{u}_t$being Q .

The elements of coefficient matrices of the VAR part (i.e. the A_1's) are simulated from normal distribution with mean 0 and sd 0.05 , the variance covariance matrix of the $\boldsymbol{u}_t$that is Q is kept diagonal and the diagonal elements are simulated from $Exp(2)$ distribution . This matrices for each model that is A_1's and Q are kept same throughout the simulation for each replications . The elements of the loading matrix Λ are simulated from $U(0,1)$ distribution , and the error variance covariance matrix R is also kept diagonal and the diagonal elements are simulated from $Exp(2)$ distribution . But these two matrices Λ and R for each models are not kept same throughout the simulations , they are simulated at each replication as these two matrices are responsible for data generation process . So for a model we first simulated the matrices A_i's and Q, using those at each replications a VAR(p) process with n observations are generated then using the simulated Λ and the VAR observation $\boldsymbol{F}_t$the common component $\boldsymbol{\chi}_t = \Lambda \boldsymbol{F}_t$ is generated , then the error vector $\boldsymbol{\varepsilon}_t$ is generated from a multivariate normal distribution with mean $\mathbf{0}$and variance covariance matrix R and the data for each replication is generated $\boldsymbol{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\varepsilon}_t$. In each replication we estimated the dynamic factor model considering we know the correct number of factors and we estimated Λ , $\boldsymbol{F}_t$ and also the common factor by $\widehat{\boldsymbol{\chi}} = \widehat{\Lambda}\widehat{\boldsymbol{F}}_t$. We generated data for different values of n and T. Here the values of n considered is $10, 20, 50, 100$ and for T $20, 50, 100, 200$ is considered. We have considered 200 replications for each combinations of (n, T, r, p) . While fitting the data data to a particular model we assumed that we know the order of the VAR that is for fitting we have considered $p = 1$. Regarding number of factors, we considered fitting the model for three values number of factors. For number of factor we considered three cases, first is correct model that is the number of factors used to fit the model is same as number of factors used in simulation $\widehat{r} = r$, also we considered two misspecified models as $\widehat{r} = r + 1$ and $\widehat{r} = r - 1$.

4.2 Fitting of the original series

Here we will measure accurately the model fit the data. Now for calculating accuracy we will use the measure $\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$, which is based on how accurately the the model estimates the common factor $\boldsymbol{\chi}_t$, based on the paper **Fomi, Hallin, Lippi, and Reichlin (2000)** the measure is defined as ,

$$\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}}) = \frac{\sum_{i,t}(\chi_{i,t} - \widehat{\chi_{i,t}})^2}{\sum_{i,t}(\chi_{i,t})^2}$$

A smaller value indicates comparatively a better fit for this measure .

Model1 ($r = 4$ and $p = 1$)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r$

n\T	20	50	100	200
10	0.685 (0.261)	0.492 (0.148)	0.43 (0.111)	0.424 (0.108)
20	0.553 (0.218)	0.327 (0.085)	0.236 (0.049)	0.21 (0.04)
50	0.452 (0.154)	0.236 (0.054)	0.167 (0.028)	0.143 (0.02)
100	0.356 (0.089)	0.202 (0.038)	0.145 (0.019)	0.123 (0.014)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r$-1

n\T	20	50	100	200
10	0.569 (0.230)	0.391 (0.122)	0.33 (0.09)	0.333 (0.087)
20	0.439 (0.149)	0.285 (0.076)	0.219 (0.05)	0.188 (0.038)
50	0.341 (0.1)	0.207 (0.040)	0.161 (0.027)	0.148 (0.02)
100	0.299 (0.073)	0.192 (0.033)	0.150 (0.019)	0.134 (0.015)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r$+1

n\T	20	50	100	200
10	0.726 (0.280)	0.574 (0.16)	0.558 (0.104)	0.531 (0.117)
20	0.622 (0.231)	0.369 (0.1)	0.28 (0.065)	0.24 (0.048)
50	0.486 (0.143)	0.266 (0.053)	0.191 (0.033)	0.15 (0.022)
100	0.434 (0.124)	0.235 (0.046)	0.166 (0.018)	0.133 (0.015)

Model2 ($r = 3$ and $p = 1$)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r$

n\T	20	50	100	200
10	1.156 (0.604)	0.792 (0.296)	0.635 (0.145)	0.67 (0.186)
20	0.946 (0.373)	0.51 (0.171)	0.366 (0.098)	0.293 (0.07)
50	0.754 (0.288)	0.404 (0.089)	0.290 (0.05)	0.218 (0.036)
100	0.663 (0.222)	0.347 (0.0777)	0.237 (0.042)	0.188 (0.026)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r$-1

n\T	20	50	100	200
10	2.715 (1.272)	1.54 (0.595)	1.089 (0.407)	0.894 (0.285)
20	2.15 (0.865)	1.224 (0.361)	0.845 (0.21)	0.675 (0.1447)
50	1.698 (0.590)	0.974 (0.213)	0.7 (0.134)	0.556 (0.106)
100	1.525 (0.476)	0.852 (0.178)	0.611 (0.096)	0.505 (0.081)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r+1$

n\T	20	50	100	200
10	4.59 (1.983)	2.806 (1.13)	1.895 (0.535)	1.640 (0.464)
20	3.783 (1.394)	2.063 (0.609)	1.356 (0.290)	0.956 (0.216)
50	3.056 (1.059)	1.599 (0.382)	1.091 (0.175)	0.776 (0.128)
100	2.787 (0.8)	1.431 (0.295)	0.952 (0.165)	0.674 (0.084)

Model3 ($r = 5$ and $p = 1$)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r$

n\T	20	50	100	200
10	0.45 (0.129)	0.365 (0.08)	0.339 (0.069)	0.341 (0.071)
20	0.366 (0.096)	0.256 (0.048)	0.226 (0.039)	0.207 (0.034)
50	0.306 (0.065)	0.201 (0.031)	0.171 (0.022)	0.156 (0.018)
100	0.268 (0.055)	0.182 (0.027)	0.159 (0.016)	0.146 (0.015)

Measure 1 ($\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$) considering $\widehat{r} = r$-1

n\T	20	50	100	200
10	0.406 (0.109)	0.328 (0.069)	0.305 (0.063)	0.311 (0.065)
20	0.342 (0.078)	0.247 (0.043)	0.22 (0.037)	0.207 (0.030)
50	0.28 (0.047)	0.204 (0.027)	0.185 (0.022)	0.175 (0.018)
100	0.257 (0.054)	0.197 (0.027)	0.178 (0.018)	0.167 (0.013)

From the above observations we can see that the performance of the fit improves as both (n, T) increases. Also for a fixed value of n (or T) when T (or n) increases the performance increases. Also it seems from values of the performance $\mathscr{R}(\boldsymbol{\chi}_t, \widehat{\boldsymbol{\chi}})$, the effect of number of series n is more dominant than the number of observations T.

4.3 Estimation of the unobserved factors

Here we measure how accurately the fitted model estimates the unobserved factors. The measure used here, $\mathscr{T}(F, \widehat{F})$ is based on how accurately the we are estimating the the unobserved dynamic factors (this calculation is only possible as we have simulated the underlying factors also , for real data this measure can not be calculated as the factors are unobserved.). Based on the Bańbura and Modugno (2010) paper the measure is defined as ,

$$(\mathscr{T}(F, \widehat{F}) = \frac{Trace(F'\widehat{F}(\widehat{F}'\widehat{F})^{-1}\widehat{F}'F)}{Trace(F'F)}$$

Where F is the $T \times r$ matrix of the factors. Notice that, if $\widehat{F} = F$, then the measure is 1 . So the measure lie between the interval 0 and 1 and a value closer to 1 indicates good fit.

Model1 ($r = 4$ and $p = 1$)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r$

n\T	20	50	100	200
10	0.542 (0.111)	0.507 (0.095)	0.535 (0.1)	0.513 (0.09)
20	0.636 (0.107)	0.654 (0.096)	0.703 (0.075)	0.713 (0.081)
50	0.765 (0.075)	0.834 (0.057)	0.869 (0.043)	0.881 (0.033)
100	0.857 (0.056)	0.914 (0.03)	0.937 (0.02)	0.945 (0.016)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r$-1

n\T	20	50	100	200
10	0.506 (0.121)	0.505 (0.099)	0.515 (0.107)	0.511 (0.097)
20	0.618 (0.111)	0.634 (0.096)	0.677 (0.079)	0.694 (0.071)
50	0.747 (0.085)	0.802 (0.053)	0.822 (0.036)	0.829 (0.030)
100	0.822 (0.059)	0.855 (0.029)	0.867 (0.023)	0.87 (0.02)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r$+1

n\T	20	50	100	200
10	0.595 (0.098)	0.541 (0.089)	0.533 (0.106)	0.532 (0.089)
20	0.679 (0.095)	0.684 (0.084)	0.7 (0.087)	0.726 (0.073)
50	0.792 (0.075)	0.837 (0.055)	0.871 (0.043)	0.887 (0.033)
100	0.866 (0.055)	0.916 (0.033)	0.937 (0.018)	0.948 (0.013)

Model2 ($r = 3$ and $p = 1$)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r$

n\T	20	50	100	200
10	0.688 (0.142)	0.693 (0.104)	0.703 (0.086)	0.697 (0.079)
20	0.754 (0.101)	0.779 (0.073)	0.788 (0.065)	0.807 (0.056)
50	0.821 (0.086)	0.864 (0.05)	0.884 (0.036)	0.895 (0.032)
100	0.867 (0.075)	0.917 (0.035)	0.934 (0.024)	0.944 (0.021)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r$-1

n\T	20	50	100	200
10	0.414 (0.165)	0.423 (0.144)	0.438 (0.142)	0.432 (0.124)
20	0.503 (0.154)	0.53 (00)	0.55 (0.105)	0.570 (0.099)
50	0.604 (0.126)	0.644 (0.085)	0.673 (0.078)	0.712 (0.075)
100	0.659 (0.108)	0.735 (0.087)	0.774 (0.061)	0.798 (0.045)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r+1$

n\T	20	50	100	200
10	0.497 (0.154)	0.442 (0.136)	0.450 (0.107)	0.457 (0.1115)
20	0.567 (0.133)	0.562 (0.108)	0.565 (0.094)	0.598 (0.090)
50	0.686 (0.109)	0.687 (0.084)	0.704 (0.082)	0.722 (0.082)
100	0.726 (0.093)	0.775 (0.076)	0.804 (0.054)	0.832 (0.045)

Model3 ($r = 5$ and $p = 1$)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r$

n\T	20	50	100	200
10	0.545 (0.084)	0.494 (0.086)	0.482 (0.086)	0.478 (0.09)
20	0.667 (0.091)	0.674 (0.078)	0.689 (0.078)	0.719 (0.07)
50	0.81 (0.064)	0.875 (0.047)	0.901 (0.078)	0.913 (0.024)
100	0.896 (0.039)	0.945 (0.018)	0.956 (0.011)	0.962 (0.01)

Measure 2 ($\mathscr{T}(F, \widehat{F})$) considering $\widehat{r} = r$-1

n\T	20	50	100	200
10	0.511 (0.088)	0.471 (0.089)	0.468 (0.083)	0.453 (0.094)
20	0.617 (0.091)	0.637 (0.08)	0.661 (0.079)	0.678 (0.063)
50	0.772 (0.057)	0.81 (0.038)	0.809 (0.035)	0.807 (0.029)
100	0.85 (0.044)	0.852 (0.028)	0.848 (0.024)	0.848 (0.021)

From the above observations we can see that the accuracy of estimated underlying factors increases as both (n, T) increases. Also for a fixed value of n (or T) when T (or n) increases the accuracy increases. Also it seems from values of the performance $\mathscr{T}(F, \widehat{F})$, the effect of number of observations T is more dominant than the number of series n.

5 Application to Indian Economy

Here will will apply the Dynamic Factor Model to model Indian GDP growth. In many cases the GDP of each quarter is published with some delay. But for economic policy making or economic decision making an estimate of the GDP or GDP growth is required. This is why there is a need for forecasting GDP/GDP growth of present or immediate future. This forecasting of present time or just immediate future is also termed as now-casting in literature. Dynamic factor models can be used here as we consider various other economic series for which there is no publication delay to model the GDP.

5.1 Data

Stock and Watson (1989) in their paper "New Indexes of Coincident and Leading Economic Indicator " had discussed the choice of variables. We tried follow that, but due to data unavailability we could not follow that properly. For this application we have chosen 30 quarterly variables apart from GDP from different sectors. The details of the chosen variables are given in the Appendix. After visually inspecting we found that many of the series have an trend component thus not stationary. But to apply dynamic factor model we need to convert them into stationary series. We used augmented Dickey–Fuller (ADF) test for this task. First ADF test is applied on all the series we considered and we checked for stationarity. The series which are not stationary we took first difference $\Delta x_t = x_t - x_{t-1}$, and again ADF test is applied on those series.

Still the series are not stationary second order difference $\Delta^2(x_t) = \Delta(\Delta(x_t))$ is taken and checked for stationarity. After second order difference all the series are stationary. In the table the last column indicates the order of the difference we needed to take to make them stationary. As we are interested in modeling the GDP growth rate we calculate GDP growth rate by $y_t = \frac{GDP_{t+1} - GDP_t}{GDP_t}$. We used the data from 1996 up to year 2017 for all the analysis, and we will use the data of year 2017 and 2018 (only the GDP growth) to check how good is the forecast done by the models. So we split the data accordingly. Also the series (other than GDP growth rate) is scaled and centered, this is done only on the estimation data set that is the data before 2017.

5.2 Analysis

Here we have taken three approaches. The first one is two step modeling approach. Here we first further selected some variables from the 30 series we initially have and then apply dynamic factor model on them. After that we model the GDP growth using the estimated factors. In the first approach we have taken a two step modeling approach. For the second and third approach we only used dynamic factor model . The details will be discussed below.For all the dynamic factor model here we will use, we will consider the restricted case. That is we will consider the error variance covariance matrix R is a diagonal matrix and there is no auto-correlation between the lags of components of error vector.

5.2.1 Initial Variable Selection

Here we further selected a subset of the all variables we considered for the model. To get an idea about the bivariate relationship between GDP growth and the series we considered, we calculated the correlation of the variables with GDP growth of the same quarter along with one quarter before and after (that is correlation with y_t,$y_{t-1}$and y_{t+1}). We found that most of the series except for the series M1, INR_EURO,Br_Eff_Ex_rate and INR_SDR. After doing that we selected a subset of 9 variable using step-wise regression. We used a bidirectional (forward and backward both direction) step-wise regression for that.

Table 1:Selected variables using step-wise regression

Sr No	Variables
1	IP_Total_ExMining
2	ConGP_Total
3	Exp
4	BTS_FinG_Stock
5	BTS_Employ
6	BMP_Total
7	CPI
8	WPI_Indus_Total
9	WPI_Food

5.2.2 **Approach 1** (Considering the GDP growth data to estimate the underlying factors)

After selecting those variable we will apply DFM to this set of variables(We have used **“dfms”[30]** package in R for this analyisis.). For that we need to estimate the number of factors r and the order of the VAR equation p. As for our case both n and T is not that large the measures for estimating r in section 3.3.1 might be inaccurate (for 9 variables those measures indicating that 8 is appropriate estimate for r). So we tried roughly estimating r with the help of principal component analysis. We observed that 5 principal components can explain more than 86% variability of the selected variables. Thus for an rough estimate of r we will consider $\widehat{r} = 5$ here. Now after fixing $r = 5$ we choose the value of p based on how good it fits the model. So we considered $p = 1, 2, 3, 4, 5, 6$ and $r = 5$ the fit DFM for each of the (r, p) combination and selected the value of p for which the model fit is best. Here as a measure of model fit we consider the metric $\frac{\sum_{i,t}(x_{i,t}-\widehat{x_{i,t}})^2}{\sum_{i,t}(x_{i,t})^2}$, where our model is $\boldsymbol{x}_t = \Lambda \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$ (along with the VAR equation) and $\widehat{\boldsymbol{x}_t} = \widehat{\Lambda}\widehat{\boldsymbol{F}_t}$. The chosen value for which the value of the metric is lowest is $p = 1$. Also we applied information based criteria to estimate p, and estimates based on information based measure $HQ(p)$ and $SC(p)$ is also $p = 1$ (section 3.3.2). So here we considered the DFM with $r = 5$ and $p = 1$.

Table 2:Estimated Factor Loadings (Approach 1)

	F_1	F_2	F_3	F_4	F_5
IP_Total_ExMining	0.32	0.51	0.22	0.07	-0.14
ConGP_Total	0.25	0.54	0.12	0.41	-0.49
Exp	0.2	-0.16	0.51	-0.74	-0.29
BTS_FinG_Stock	0.23	0.11	0.32	0.13	0.87
BTS_Employ	0.23	0.36	-0.15	-0.37	0.32
BMP_Total	0.03	-0.4	0.67	0.46	-0.08
CPI	0.48	-0.24	-0.34	0.02	-0.12
WPI_Indus_Total	0.5	-0.13	0.04	0.04	-0.02
WPI_Food	0.5	-0.36	-0.28	0.15	-0.01

For estimation here EM based method by **Banbura and Modugno (2010)** (section 3.2.1 and 3.2.2) is used as in the data there was some missing values. For the estimated factors we used the factor estimates by two-step algorithm by **Doz, Giannone and Reichlin (2011)** (section 3.4). Now in this approach we will use those estimated factors (denoted by $\widehat{F_i}, i = 1, 2, 3, 4, 5$)to model the GDP growth.

Model1

Here we considered multiple linear regression model of the form ,

$$y_t = \beta_0 + \beta_1\widehat{F_{1t}} + \beta_2\widehat{F_{2t}} + \beta_3\widehat{F_{3t}} + \beta_4\widehat{F_{4t}} + \beta_5\widehat{F_{5t}} + \epsilon_t$$

Where $y_t$denotes the GDP growth at time t.

Table 3: Estimate of the coefficients

Parameter	Estimate	s.e
β_0	0.0304	0.0012
β_1	0.0040	0.0007
β_2	0.0048	0.0010
β_3	0.0013	0.0011
β_4	-0.0027	0.0013
β_5	0.0023	0.0015

,with $\widehat{F_1}$, $\widehat{F_2}$and $\widehat{F_4}$highly significant at 95% confidence level.

Table 4: Performance of the fit

R^2	0.4249
Adjusted R^2	0.3889
AIC	-520.6346

Model2

Here we used an AR(1) model with covariates as the estimated factors. That is the model of the form,

$$y_t = \beta_0 + \phi y_{t-1} + \beta_1 \widehat{F_{1t}} + \beta_2 \widehat{F_{2t}} + \beta_3 \widehat{F_{3t}} + \beta_4 \widehat{F_{4t}} + \beta_5 \widehat{F_{5t}} + \epsilon_t$$

Table 5: Estimate of the coefficients

Parameter	Estimate	s.e
ϕ	-0.0622	0.1161
β_0	0.0304	0.0011
β_1	0.0041	0.0007
β_2	0.0049	0.0010
β_3	0.0014	0.0011
β_4	-0.0030	0.0014
β_5	0.0023	0.0016

Table 6: Performance of the fit

R^2	0.4268
Adjusted R^2	0.3833
AIC	-518.92

Model3

Here we applied VAR. We combined the GDP growth data with the factors, and applied VAR. The number of order of the VAR based on information criteria (considering $HQ(p), SC(p)$) based method suggested $p = 1$. So we considering fitting a VAR(1) model of the form,

$$
\begin{bmatrix} y_t \\ \widehat{F_{1t}} \\ \widehat{F_{2t}} \\ \widehat{F_{3t}} \\ \widehat{F_{4t}} \\ \widehat{F_{5t}} \end{bmatrix} = A \begin{bmatrix} 1 \\ y_{t-1} \\ \widehat{F_{1t-1}} \\ \widehat{F_{2t-1}} \\ \widehat{F_{3t-1}} \\ \widehat{F_{4t-1}} \\ \widehat{F_{5t-1}} \end{bmatrix} + \boldsymbol{\epsilon}_t
$$

or , the part for GDP growth is $y_t = A_{1,1} + A_{1,2}y_{t-1} + A_{1,3}\widehat{F_{1t-1}} + A_{1,4}\widehat{F_{2t-1}} + A_{1,5}\widehat{F_{3t-1}} + A_{1,6}\widehat{F_{4t-1}} + A_{1,7}\widehat{F_{5t-1}} + \epsilon_{1,t}$. This is similar to the Model2 but using 1 lagged previous estimated factors.

Table 7: Estimate of the coefficients

Parameter	Estimate	s.e
$A_{1,1}$	3.554e-02	4.469e-03
$A_{1,2}$	-1.646e-01	1.378e-01
$A_{1,3}$	2.772e-03	1.077e-03
$A_{1,4}$	3.301e-03	1.483e-03
$A_{1,5}$	3.406e-03	1.393e-03
$A_{1,6}$	-9.396e-05	1.718e-03
$A_{1,7}$	3.554e-02	4.469e-03

Table 8:Performance of the fit

R^2	0.1678 (for GDP growth part only)
Adjusted R^2	0.1038 (for GDP growth part only)
AIC	-

Model4

Here also we used a multiple linear regression model. But here we used all the five estimated factor, additionally we have chosen some of the variables from the 9 previously selected variables by a multi-directional step-wise regression. The additionally chosen variables by the step-wise regression are WPI_Indus_Total and CPI (denoted as w_1, w_2respectively), so the model is of the form,

$$y_t = \beta_0 + \beta_1\widehat{F_{1t}} + \beta_2\widehat{F_{2t}} + \beta_3\widehat{F_{3t}} + \beta_4\widehat{F_{4t}} + \gamma_5\widehat{F_{5t}} + \gamma_6 w_1 + \beta_7 w_2 + \epsilon_t$$

Table 9: Estimate of the coefficients

Parameter	Estimate	s.e
β_0	0.0301	0.0012
β_1	0.0122	0.0023
β_2	0.0021	0.0012
β_3	-0.0008	0.0014
β_4	-0.0019	0.0012
β_5	0.0013	0.0015
γ_6	-0.0077	0.002
γ_7	-0.0088	0.003

Table 10: Performance of the fit

R^2	0.5885
Adjusted R^2	0.542
AIC	-434.4233

We also tried fitting a similar model with all the five estimated factors, the other variables selected in a step wise manner not only from the nine selected variables but from the whole 30 variables which we considered. This way also we got the same model.

5.2.3 **Approach2** (Considering the GDP growth data to estimate the underlying factors)

Here also we will consider the initially nine selected variables. But along with these nine variables we combine the GDP growth data (y_t) and applied DFM to it. Same as previous approach we estimated the number of factors and the estimated values are $r = 5$ and $p = 1$.

Table 11: Estimated Factor Loadings (Approach 2)

	F_1	F_2	F_3	F_4	F_5
y_t	0.34	-0.3	0.00	-0.21	0.1
IP_Total_ExMining	0.35	-0.38	0.08	0.21	-0.24
ConGP_Total	0.24	-0.34	0.01	0.42	-0.26
Exp	0.19	0.10	0.52	-0.60	-0.42
BTS_FinG_Stock	0.24	-0.11	0.32	0.05	0.76
BTS_Employ	0.27	-0.29	-0.19	-0.27	0.07
BMP_Total	-0.01	0.25	0.5	0.24	0.03
CPI	0.37	0.32	-0.29	-0.05	-0.03
WPI_Indus_Total	0.42	0.24	0.1	0.17	-0.09
WPI_Food	0.4	0.47	-0.23	0.08	0.07

After that here also we will try fitting regression models

Model1

Here we considered multiple linear regression model of the form ,

$$y_t = \beta_0 + \beta_1 \widehat{F_{1t}} + \beta_2 \widehat{F_{2t}} + \beta_3 \widehat{F_{3t}} + \beta_4 \widehat{F_{4t}} + \beta_5 \widehat{F_{5t}} + \epsilon_t$$

Where $y_t$denotes the GDP growth at time t.

Table 12: Estimate of the coefficients

Parameter	Estimate	s.e
β_0	0.0304	0.0009
β_1	0.0055	0.0006
β_2	-0.0054	0.0008
β_3	-0.0007	0.0009
β_4	-0.0044	0.0012
β_5	0.0028	0.0013

,with $\widehat{F_1}, \widehat{F_2}, \widehat{F_4}$and $\widehat{F_5}$highly significant at 95% confidence level.

Table 13: Performance of the fit

R^2	0.6489
Adjusted R^2	0.627
AIC	-563.0881

Model2

Here we did the similar thing as of Model4 in approach1. The additionally chosen variables by the step-wise regression are BTS_FinG_Stock, BTS_Employ, WPI_Food, IP_Total_ExMining, and WPI_Indus_Total (denoted as w_1, w_2, w_3, w_4, w_5res so the model is of the form,

$$y_t = \beta_0+\beta_1\widehat{F_{1t}}+\beta_2\widehat{F_{2t}}+\beta_3\widehat{F_{3t}}+\beta_4\widehat{F_{4t}}+\gamma_5\widehat{F_{5t}}+\gamma_6 w_1+\beta_7 w_2+\gamma_8 w_3+\beta_9 w_4+\gamma_{10} w_5+\epsilon_t$$

Table 14: Estimates of the coefficients

Parameter	Estimate	s.e
β_0	0.0302	0.0007
β_1	0.0096	0.0067
β_2	-0.0251	0.007
β_3	0..0171	0.0034
β_4	-0.0051	0.0018
β_5	0.0317	0.0057
γ_6	-0.035	0.0055
γ_7	-0.0061	0.0012
γ_8	0.0235	0.014
γ_9	-0.0049	0.0016
γ_{10}	-0.0021	0.0016

Table 15: Performance of the model fit

R^2	0.8712
Adjusted R^2	0.8494
AIC	-509.7426

5.2.4 Comparisons of the Forecasts

As we considered the GDP growth data in the DFM to estimate the underlying factors in the approach2 it is expected to get a better model fit in that approach. But here we will check if that also give a better forecast or not. After fitting the models we now forecast the GDP growth for the next eight quarters (2018 and 2018). We get the forecast of the underlying factors from the dynamic part of the DFM. Using that and the estimated loading matrix we forecast the series we needed. For measuring the forecasting performance we used RMSE and MAE for both within sample and out of sample (2018,2018). The results are given in the following table.

Table 16: Forecasting performance of the models

Model	in sample RMSE	out of sample RMSE	in sample MAE	out of sample MAE
Approach1				
Model 1	0.0108	0.0133	0.0086	0.0117
Model 2	0.0107	0.0133	0.0087	0.012
Model 3	0.0130	0.0137	0.0090	0.0128
Model 4	0.0095	0.0136	0.0079	0.0127
Approach2				
Model 1	0.0084	0.0134	0.0069	0.0122
Model 2	0.0053	0.0131	0.0042	0.0118

What we observe is that in Approach2 the models have greater forecasting power than the models in Approach1.

6 Comparison With Other Dimensionality Reduction Methods

Here we will compare the performance of dynamic factor modeling on this data with other competing methods. More specifically we will consider non dynamic version of factor analysis (classical factor analysis), principal component analysis, and dynamic principal analysis (discussed in section 3.4) along with random projection methods. We compare these models based on two criteria. First is overall comparison, this is how well those model fit the data we consider and how efficiently they are reducing the dimension. The second criteria is based on modeling the GDP growth like we did on the previous part. Before that we will briefly describe the competing methods.

First is classical factor model (Orthogonal Factor model). Here also we consider the model $\boldsymbol{x}_t = \Lambda \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$. But here we do not consider any dynamic nature of the factors and also here the $\boldsymbol{x}_t$'s are independent and identically distributed unlike dynamic factor model, where $\boldsymbol{x}_t$'s are time series observations so there were some sort of dependencies and they also might not be identical. Also here we assume the following, $\boldsymbol{F}_t$and $\boldsymbol{\varepsilon}_t$, $E(\boldsymbol{F}_t) = \mathbf{0}$ $cov(\boldsymbol{F}_t) = \boldsymbol{I}_r$, $E(\boldsymbol{\varepsilon}_t) = \mathbf{0} cov(\boldsymbol{\varepsilon}_t) = \boldsymbol{\psi}$(a diagonal matrix). So the variance covariance structure of the model is $cov(\boldsymbol{x}_t) = \Lambda\Lambda^{'} + \boldsymbol{\psi}$. Now the factor loadings Λ and the variance covariance matrix $\boldsymbol{\psi}$, can be estimated using principal component methods or by using maximum likelihood methods assuming $(\boldsymbol{F}_t, \boldsymbol{\varepsilon}_t)$ jointly normal. Further the underlying random factors are estimated (factor scores) by regression method using estimates of $\Lambda$and $\boldsymbol{\psi}$. (**Johnson and Wichern, 2013**)

In principal component analysis, we construct principal components by taking appropriate linear combinations of the variables. The first principal component is the linear combination of the variables $\boldsymbol{pc}_1 = \sum_{i=1}^{n} a_{1i}\boldsymbol{x}_i$,that maximizes it's variance subject to $\boldsymbol{a}_1^{'}\boldsymbol{a}_1 = 1$. The second principal component is the linear combination $\boldsymbol{pc}_2 = \sum_{i=1}^{n} a_{2i}\boldsymbol{x}_i$,that maximizes it's variance subject to $\boldsymbol{a}_2^{'}\boldsymbol{a}_2 = 1$ and $\boldsymbol{pc}_2$is orthogonal to $\boldsymbol{pc}_1$. Similarly the kth principal component is the linear combination $\boldsymbol{pc}_k = \sum_{i=1}^{n} a_{ki}\boldsymbol{x}_i$, that maximizes it's variance subject to $\boldsymbol{a}_k^{'}\boldsymbol{a}_k = 1$ and it is orthogonal to previous $k-1$ principal components. In practice it can be shown that if S be the sample variance covariance matrix of the data and $(l_1, \boldsymbol{e}_1)$,..., $(l_n, \boldsymbol{e}_n)$ be the eigen value vector pair such that $l_1 \geq l_2 \cdots \geq l_n$ then $\boldsymbol{a}_i = \boldsymbol{e}_i$ and $var(\boldsymbol{pc}_i) = l_i$. (**Johnson and Wichern, 2013**)

Random Projection relatively new dimensionality reduction technique. Here a n dimensional data is projected r dimensional subspace through origin. It is done by using a $n \times r$ random matrix R whose rows have unit lengths. Mathematically $X_{T\times r}^{RP} = X_{T\times n}R_{n\times r}$. This methods if there are T data points in $n$dimensional space, if we reduce the dimension to r such that $r > \frac{log(T)}{\epsilon^2}$ then this method preserves the distance between the points in the range of $(1 \pm \epsilon)$ time of the distance in original space with a high probability, this is due to a result known as Johnson-Lindenstrauss theorem. No other dimensionality techniques guarantee this type of preservation of distance of two points. Moreover when number of dimensions is very high then this method is faster and less complex than other dimesionality reduction techniques. (**Ella Bingham , Heikki Mannila, 2001**)

6.1 Overall comparison

In overall comparison we compare the different models based on how well the model fits the data. For which we are using a measure which measures how close are the estimated and the actual values of the variables are. The measure is defined as $P_1 = \frac{\sum_{i,t}(x_{i,t}-\widehat{x_{i,t}})^2}{\sum_{i,t}(x_{i,t})^2}$. Note that this measure we also used in the previous section for for selecting the order of VAR p in the previous section. Another thing we measured here is that how efficiently, for which we used the cumulative proportion of variance explained by the method using the measure $P_2 = \frac{\sum_{i=1}^{r} var(F_i)}{\sum_{i=}^{n} var(x_i)}$. For each of the methods we have considered number of factors from $r = 1$to 10. For DFM we have considered both $p = 1$ and 2.

The following **Table 17** and the **Figure 1** shows the values of measure $P_1$for different number of factors for different methods. It can be observed that the fit for DFMs and classical factor model are more or less similar for each r, but the fit for dynamic pca is much better than these factor models.

Table 17: Performance regarding estimating the observed series

r	DFM($p=1$)	DFM($p=2$)	CFM	DPCA
1	0.833	0.833	0.838	0.658
2	0.757	0.756	0.738	0.464
3	0.631	0.634	0.609	0.374
4	0.583	0.597	0.546	0.279
5	0.487	0.494	0.476	0.23
6	0.433	0.461	0.419	0.188
7	0.374	0.404	0.369	0.158
8	0.366	0.372	0.337	0.129
9	0.361	0.344	0.303	0.107
10	0.307	0.308	0.261	0.092

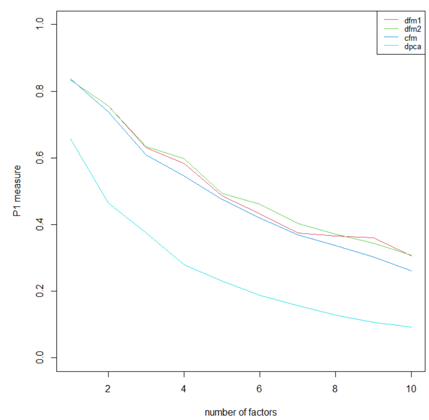

Figure 1:Performance regarding estimating the observed series

The **Table 18** and the **Figure 2** shows the values of measure P_2(the cumulative variance explained by the factors) for different number of factors for different methods. We can observe that the cumulative proportion of total variance explained by the factors is more or less same for two dynamic factor models, the performance of the random projection is also more or less similar to the dynamic factor models (although random projection is used for very high dimensions and there are some conditions about the number of factors, but here the number of dimensions $n = 30$ is not that high, also the resulting lower dimension r must be greater than $\frac{log(T)}{\epsilon^2}$,which also we can not consider for smaller ϵ, as here $T = 67$). The cumulative variance explained by the classical factor model is very low, for 10 factors also it only explains approximately 26% of the total variability of the series. The performance of both PCA and dynamic PCA is better than the dynamic factor models. Here also the Dynamic PCA explains the most amount of variability of the data for each r.

Table 18: Cumulative proportion of variation explained by different methods

r	DFM($p = 1$)	DFM($p = 2$)	CFM	PCA	DPCA	RP
1	0.155	0.155	0.029	0.182	0.34	0.093
2	0.254	0.256	0.057	0.336	0.532	0.157
3	0.34	0.343	0.083	0.438	0.618	0.217
4	0.393	0.396	0.113	0.504	0.714	0.339
5	0.454	0.458	0.141	0.567	0.76	0.396
6	0.506	0.512	0.168	0.628	0.803	0.479
7	0.548	0.554	0.195	0.67	0.832	0.506
8	0.581	0.590	0.216	0.723	0.860	0.629
9	0.612	0.622	0.237	0.762	0.881	0.637
10	0.649	0.659	0.263	0.798	0.896	0.686

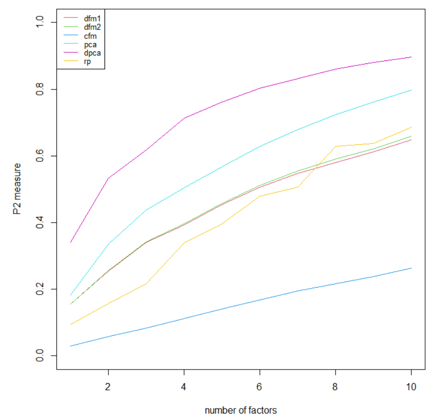

Figure 2: Cumulative proportion of variation explained by different methods

6.2 Model specific performance

Here we model the GDP growth using the using the estimated factors of the different methods. The main difference of these methods with DFM is that, in DFM we assume that the underlying factors follows a dynamic nature, which we considered here as VAR(p) and we estimate those factors considering these assumption. This helps us forecasting those factors in future times and also gives an advantage in forecasting the series we considered in the dynamic factor model. But the other methods such as classical factor model, PCA or dynamic PCA do not consider this assumption of dynamic nature of the underlying factors. But to compare the forecasting performance of these methods we need an forecast of those underlying factors and also forecast of some of the variables which will be used in the model. So here for forecasting the underlying factors we separately fitted a VAR model with $p = 1$ on the estimated factors and then made a forecast for the next eight quarters. For the original variables also we used VAR, but for classical factor model we have the loading matrix so we used that loading matrix and the foretasted factors to get the forecast of the variables. After that we using those estimated factors, we fit a model for GDP growth using the approach we used in Approach1 and Model4 in the previous section (as this model has one of the best in sample and out of sample forecasting accuracy). For comparison we used the usual model fit measures along with in sample and out of sample forecasting performance. The comparison is given in the following ("is" stands for in-sample and "os" stands for out-of-sample).

Table 20: Comparison of model performance

Approach	R^2	Adjusted R^2	AIC	is RMSE	os RMSE	is MAE	os MAE
A1M4	0.5885	0.542	-434.4233	0.0079	0.0136	0.0079	0.0127
A2M2	0.8712	0.8494	-509.7426	0.0053	0.0131	0.0042	0.0119
CFM	0.5945	0.5463	-413.3964	0.0097	0.0140	0.008	0.0128
PCA	0.518	0.473	-425.4469	0.0103	0.0180	0.0086	0.0140
DPCA	0.524	0.4615	-422.2276	0.0103	0.0156	0.0083	0.0140

What we can observe is that for forecasting the models based on DFM performs better than the models based on other methods.

7 Conclusion

In this report we have discussed the popular dimensionality reduction technique dynamic factor model for high dimensional time series data. After discussing the theory portion we did a simulation study where we found that the performance of the dynamic factor model increases as both (n, T) increases, both in model fit (estimating the original series) and also in estimating the underlying factors. In application we used a small-scale dynamic factor model to model and forecast Indian GDP growth with various other economic series. For that we used a step wise variable selection procedure followed by dynamic factor model to estimate the underlying factors and then model the GDP growth with the estimated factors using linear regression. In one approach we do not used the GDP growth data in the dynamic factor model but in the other we used that to estimate the underlying factors. It was obvious that where we used the GDP growth data to estimate the factors, the regression models have a better fit. We also compared the forecasting accuracy (RMSE and MAE) of the models, it is observed that in the approach where we have used the GDP growth in the dynamic factor model have a better forecast than the approach where we didn't. Finally we compare dynamic factor model with other competing methods like classical factor analysis, principal component analysis, and dynamic principal analysis along with random projection methods using the data we have. We first compare them first based on how good is the fit (on the series used) and also on how efficiently they reduce the dimension. We have found that for performance of fit of the model dynamic PCA is best and the performance of classical and dynamic factor models are more or less same. For the dimensionality reduction dynamic PCA and the classical PCA is better than the factor models, also the dynamic factor model performs much more efficiently than classical factor models. Finally we compare the performance of forecasting the GDP growth using the similar approach used previously. It was observed that the forecast based on dynamic factor model is better than the forecasts based on other methods.

References

[1] Babikir, A., Mwambi, H. Evaluating the combined forecasts of the dynamic factor model and the artificial neural network model using linear and nonlinear combining methods. Empir Econ 51, 1541–1556 (2016). https://doi.org/10.1007/s00181-015-1049-1

[2] Bai J. (2003). Inferential theory for factor models of large dimensions. Econometrica, 71, 1, 135- 171.

[3] Bai, J., and S. Ng (2002), "Determining the Number of Factors in Approximate Factor Models," Econometrica, 70, 191-221.

[4] Bai J. and Ng S. (2007). Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics, 25, 52-60.

[5] Bańbura, Marta, and Michele Modugno. "Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data." Journal of applied econometrics 29.1 (2014): 133-160.

[6] Barhoumi, Karim, Olivier Darné, and Laurent Ferrara. "Dynamic factor models: A review of the literature." Available at SSRN 2291459 (2013).

[7] Becker, C., & Fried, R. (2003). Sliced inverse regression for high-dimensional time series. In Exploratory Data Analysis in Empirical Research: Proceedings of the 25 th Annual Conference of the Gesellschaft für Klassifikation

eV, University of Munich, March 14–16, 2001 (pp. 3-11). Springer Berlin Heidelberg.

[8] Bhadury, Ghosh, Kumar (2020) Nowcasting Indian GDP growth using a Dynamic Factor Mode RBI Working Paper Series No. 03

[9] Bingham, E., & Mannila, H. (2001, August). Random projection in dimensionality reduction: applications to image and text data. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 245-250).

[10] Brillinger (1969) Time Series: Data Analysis and Theory. Holden-Day, San Francisco.

[11] Chernis, T., & Sekkel, R. (2017). A dynamic factor model for now-casting Canadian GDP growth. Empirical Economics, 53, 217-234.

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

[13] Cuevas, Á., Quilis, E.M. A factor analysis for the Spanish economy. SERIEs 3, 311–338 (2012). https://doi.org/10.1007/s13209-011-0060-9

[14] Doz C., Giannone D. and Reichlin L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. Journal of Econometrics, 164, 188-205.

[15] Doz C., Giannone D. and Reichlin L. (2012). A quasi maximum likelihood approach for large approximate dynamic factor models. Review of Economics and Statistics, 94, 4, 1014-1024.

[16] Forni M. and Lippi M. (1997). Aggregation and the Microfoundations of Dynamic Macroeconomics. Oxford University Press

[17] Forni M. and Reichlin L. (1998). Let's get real: A factor-analytic approach to disaggregated business cycle dynamics. Review of Economic Studies, 65, 453-473.

[18] Forni M., Hallin M., Lippi M. and Reichlin L. (2000). The generalized dynamic-factor model: Identification and estimation. Review of Economics and Statistics, 82, 540-554

[19] Forni M., Hallin M., Lippi M. and Reichlin L. (2004). The generalized factor model: Consistency and rates. Journal of Econometrics, 119, 231-255.

[20] Gerhard Rünstler (2016) On the design of data sets for forecasting with dynamic factor models ECB Working Paper 1893, April 2016

[21] Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in Latent Variables in Socio-Economic Models, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland

[22] Iyer, Sen Gupta (2019) NOWCASTING ECONOMIC GROWTH IN INDIA THE ROLE OF RAINFALL ADB ECONOMICS WORKING PAPER SERIES NO. 593 October 2019

[23] Sargent, T.J., and C.A. Sims (1977), "Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory," in New Methods in Business Cycle Research, ed. by C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.

[24] Sezgin, F. H., & Kınay, B. (2010). A dynamic factor model for evaluation of financial crises in turkey. Bulletin De la Societe des Sciences Medicales, (1/10), 109-117.

[25] Shumway, R., and D. Stoffer (1982): "An approach to time series smoothing and forecasting using the EM algorithm," Journal of Time Series Analysis, 3, 253–264.

[26] Shumway, R. H., Stoffer, D. S., & Stoffer, D. S. (2000). Time series analysis and its applications (Vol. 3). New York: springer.

[27] Stock, J.H., and M.W. Watson (1989), "New Indexes of Coincident and Leading Economic Indicators," NBER Macroeconomics Annual 1989, 351-393.

[28] 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, 1167-1179.

[29] Stock, J.H., and M.W. Watson (2006), "Forecasting with Many Predictors," ch. 6 in Handbook of Economic Forecasting, ed. by Graham Elliott, Clive W.J. Granger, and Allan Timmermann, Elsevier, 515-554.

[30] Krantz S, Bagdziunas R (2023). _dfms: Dynamic Factor Models_. R package version 0.1.4, <https://CRAN.R-project.org/package=dfms>.

8 Appendix

8.1 Appendix 1: Data description

Table 20: Data Description

Variable Name	Description	Type	From	To	Source	Difference
MP_Total	Total Manufacturing Production	growth rate	1996-04	2019-10	FRED	1
EP_Total	Total Electricity Profuction	Index , 2015=100	1996-04	2018-10	FRED	1
IP_Mining	Total Mining Production	Index , 2015=100	2000-01	2019-10	OECD	1
IP_Total_ExMining	Total Industrial Production Excluding Mining	Index , 2015=100	2000-01	2019-10	OECD	1
ConGP_Total	Total Consumar Goods Production	Index , 2015=100	2000-01	2017-01	OECD	1
DurGP_Total	Total Durable Goods Production	Index , 2015=100	2000-01	2019-10	OECD	1
NDurGP_Total	Total Nondurable Goods Production	Index , 2015=100	2000-01	2018-10	OECD	2
M1	Narrow Money Supply	INR	1996-04	2019-10	FRED	1
M3	Broad Money Supply	INR	1996-04	2018-10	FRED	2
Share_All	Share price all share broad	index	2000-01	2019-10	OECD	1
S&P BSE500	Stocks index	index	1999-04	2019-10	BSE	1
S&P Sensex	Stocks index	index	1996-01	2019-10	BSE	1
CPI	Consumer Price Index	index 2015=100	1996-01	2019-10	FRED	2
WPI_Indus_Total	Wholesale Price Index Industrial Total	index 2015=100	1996-01	2019-10	FRED	2
WPI_Food	Wholesale Price Index Food	index 2015=100	1996-01	2019-10	FRED	2
WPI_Indus_M	Wholesale Price Index Manufacturing	index 2015=100	1996-01	2019-10	FRED	1
Exp_Imp_Ratio	Export Import Ratio	percentage	1996-01	2019-10	FRED	1
Exp	Export of Goods and Services In India	INR	1996-01	2019-10	FRED	1
Imp	Import of Goods and Services In India	INR	1996-01	2019-10	FRED	1
Net_trade	Net Trade	INR	1996-01	2019-10	FRED	1
INR_USD	INR USD exchange rate	INR	1996-01	2019-10	FRED	1
INR_EURO	INR EURO exchange rate	INR	1999-01	2019-10	IMF	1
Br_Eff_Ex_rate	Real Broad Effective exchange rate	index , 2020=100	1996-01	2019-10	FRED	1
INR_SDR	INR SDR exchange rate	INR	1996-01	2019-10	IMF	1
BTS_Production	Business Tendency Survay Production	percentage	2000-04	2019-10	OECD	0
BTS_FinG_Stock	Business Tendency Survay finished Goods	percentage	2000-04	2019-10	OECD	1
BTS_Employ	Business Tendency Survey Employment	percentage	2000-01	2019-10	OECD	1
Int_Rate_Immidiate	Interest rates immideate (<24hr)	percentage rate	2000-01	2019-10	OECD	1
BMP_Total	Balance of Payment BMP6 Total	percentage rate	1996-04		OECD	1
BMP_GnS	Balance of Payment BMP6 Goods and Services	percentage rate	1996-04		OECD	1
GDP	Gross Domastic Product	INR	1996-04	2019-10	FRED	