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

Soumya Ghatak

May 23, 2023

University of Hyderabad

Introduction

- In many cases of macroeconomic or financial area we have large number of correlated time series.
- Analyzing those large number of data can be computationally expensive and often difficult to interpret. So we need a lower dimensional representation.
- Common dimensionality reduction techniques is not a suitable method here as the observations are not independent.
- Dynamic Factor Models are introduced (**Geweke (1977)**) for large dimensional time series data for getting a lower dimensional representation without loosing much of information of the original data, by allowing the some dynamic nature of the underlying factors.

Mathematical Formulation

- $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{nt})$ be a n-dimensional stationary multivariate time series process with mean 0 and unit variance at time t, $t = 1, 2, \ldots, T$.
- We can write the observations as $x_{it} = \chi_{it} + \varepsilon_{it}$, where the common component χ_{it} and the error component ε_{it} are assumed to be independent of each other for all i and t.

Mathematical Formulation

- Depending upon the structure of common component we have two representations, static and dynamic representation of DFM.
- In static representation, $\chi_{it} = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ir})\ \boldsymbol{F_t} = \lambda_i^{'} \boldsymbol{F}_t$, where $\boldsymbol{F}_t$ is the $r \times 1$ vector of unobserved (static) factors. The common component at time t depend only upon the factor at time t.
- In dynamic representation $\chi_{it} = \boldsymbol{\lambda}_{i0}^{'} \boldsymbol{f}_t + \boldsymbol{\lambda}_{i1}^{'} \boldsymbol{f}_{t-1} + \cdots + \boldsymbol{\lambda}_{is}^{'} \boldsymbol{f}_{t-s}$, $\boldsymbol{f}_t$ is the dynamic factor of size $q \times 1$ 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.
- In both static and dynamic representation the underlying factors $\boldsymbol{F}_t$or $\boldsymbol{f}_t$follow some dynamic relationships (e.g. VAR).

Mathematical Formulation

- From the dynamic representation of the model we can have the static representation of the model.
- By stacking the dynamic factors and the factor loading as, $\boldsymbol{\Lambda}_i = \begin{bmatrix} \boldsymbol{\lambda}_{i0} \\ \boldsymbol{\lambda}_{i1} \\ \vdots \\ \boldsymbol{\lambda}_{is} \end{bmatrix}$ 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.

Mathematical Formulation

- Assume that the dynamic factor $\boldsymbol{f}_t$ follow some auto regressive process of order $h < \infty$, then $\boldsymbol{F}_t$follows VAR(p), that is $\boldsymbol{F}_t = \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i} + \boldsymbol{u}_t$, with $p = max(1, h - s)$.
- 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.
- For $k = max(h, s) = h$, that is $s < h$, it can be shown that $\boldsymbol{F}_t$ follows a VAR of order $h - s$.

Mathematical Formulation

Examples

Consider the case with $s = 2$ and $h = 4$, then $p = max(1, 4 - 2) = 2$.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, $\boldsymbol{F}_t = \begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \end{bmatrix} = \begin{bmatrix} B_1 & B_2 & B_3 \\ I_q & 0 & 0 \\ 0 & I_q & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-1} \\ \boldsymbol{f}_{t-2} \\ \boldsymbol{f}_{t-3} \end{bmatrix} +$ $\begin{bmatrix} 0 & 0 & B_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_{t-2} \\ \boldsymbol{f}_{t-3} \\ \boldsymbol{f}_{t-4} \end{bmatrix} + \begin{bmatrix} I_q \\ 0 \\ 0 \end{bmatrix} \epsilon_t = A_1 \boldsymbol{F}_{t-1} + A_2 \boldsymbol{F}_{t-2} + \boldsymbol{u}_t$

Mathematical Formulation

- Sstatic representation is, $\boldsymbol{x}_t = \Lambda \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t \quad , \quad t = 1, 2, \ldots, T$, where $\boldsymbol{F}_t$is the $r \times 1$ underlying static factor at time t, Λ is a $n \times r$ matrix of factor loading and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})$ is the error or noise component which is uncorrelated with the underlying factor at all lags i.e. $cov(\varepsilon_t, \boldsymbol{F}_{t-h}) = 0_{n \times t}$, $h = 0, \pm 1, \pm 2, \ldots$.
- Factors follow a VAR(p) process: $\boldsymbol{F}_t = \sum_{i=1}^{p} A_i \boldsymbol{F}_{t-i} + \boldsymbol{u}_t \quad , \quad t = 1, 2, \ldots, T$.
- $\boldsymbol{\varepsilon}_t$'s can be assumed to be serially uncorrelated that is, $cov(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t+h}) = \begin{cases} R_{n \times n} & h = 0 \\ 0_{n \times n} & h = \pm 1, \pm 2, \ldots \end{cases}$, in most cases R is assumed diagonal.
- In some cases the elements of $\boldsymbol{\varepsilon}_t$are assumed to follow AR(1) process, that is $\varepsilon_{it} = a\varepsilon_{it-1} + e_i$
- $\boldsymbol{u}_t$, $t = 1, 2, \ldots, T$ is serially uncorrelated n-dimensional normal with mean 0 and variance covariane matrix $Q_{r \times r}$

Estimation of parameters

- Here we discussed the estimation for model with $p = 1$. That is, $\boldsymbol{x}_t = \Lambda \boldsymbol{F}_t + \varepsilon_t$, $\boldsymbol{F}_t = A\boldsymbol{F}_{t-1} + \boldsymbol{u}_t$, $t = 1, 2, \ldots, T$.
- $\Theta = \{\Lambda, A, R, Q\}$ set of unknown parameters. In most cases R is assumed to be diagonal.
- Also let, $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$ and $\mathscr{F} = \{\boldsymbol{F}_0, \boldsymbol{F}_1, \ldots, \boldsymbol{F}_T\}$
- $\boldsymbol{x}_1|\boldsymbol{F}_1, \boldsymbol{x}_2|\boldsymbol{F}_2, \ldots, \boldsymbol{x}_T|\boldsymbol{F}_T$are assumed to be normally distributed and independent.
- $\boldsymbol{F}_1|\boldsymbol{F}_0, \boldsymbol{F}_2|\boldsymbol{F}_1, \ldots, \boldsymbol{F}_T|\boldsymbol{F}_{T-1}$are assumed to be normally distributed and independent. Also $\boldsymbol{F}_0 \sim N_r(\boldsymbol{\mu}, \Sigma)$ is assumed.

Estimation of parameters

- The log likelihood can be written as

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

Estimation of parameters

- The factors $\boldsymbol{F}_t$are unobserved, so we can not directly maximize the log likelihood. Instead iteratively E-M algorithm is used.
- In the first step an initial value for Θ is given.
- 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 :

$$\Theta(j+1) = arg\, \underset{\Theta}{max} e(\Theta|\Theta(j))$$

Estimation of parameters

- $\Lambda(j+1) = (\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{x}_t \boldsymbol{F}_t' | X])(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{F}_t' | X])^{-1}$
- $A(j+1) =$
  $(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{F}_{t-1}' | X])(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_{t-1} \boldsymbol{F}_{t-1}' | X])^{-1}$
- For $R(j+1)$ and $Q(j+1)$ we use expected sum of squared residuals as a plug-in estimator
- $R(j+1) = \frac{1}{T}\sum_{t=1}^{T} E_{\Theta(j)}[(\boldsymbol{x}_t - \Lambda(j+1)\boldsymbol{F}_t)(\boldsymbol{x}_t - \Lambda(j+1)\boldsymbol{F}_t)' | X]$
  $= \frac{1}{T}(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{x}_t \boldsymbol{x}_t' | X] - \Lambda(j+1)\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{x}_t' | X])$
- $Q(j+1) =$
  $\frac{1}{T}\sum_{t=1}^{T} E_{\Theta(j)}[(\boldsymbol{F}_t - A(j+1)\boldsymbol{F}_{t-1})(\boldsymbol{F}_t - A(j+1)\boldsymbol{F}_{t-1})' | X]$
  $= \frac{1}{T}(\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{F}_t' | X] - A(j+1)\sum_{t=1}^{T} E_{\Theta(j)}[\boldsymbol{F}_{t-1} \boldsymbol{F}_t' | X])$

Estimation of parameters

- $E_{\Theta(j)}[\boldsymbol{x}_t \boldsymbol{x}_t^{'} | X] = \boldsymbol{x}_t \boldsymbol{x}_t^{'}$ and $E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{x}_t^{'} | X] = E_{\Theta(j)}[\boldsymbol{F}_t | X] \boldsymbol{x}_t^{'}$, given we have all the data.
- 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 .

Estimation of parameters

- Define $\boldsymbol{F}^j_{t|s} = E_{\Theta(j)}(\boldsymbol{F}_t | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_s)$ as the estimate of the unobserved factor given the observed data $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s\}$, and the error covariance $P^j_{t_1,t_2|s} = E_{\Theta(j)}((\boldsymbol{F}_{t_1} - \boldsymbol{F}^j_{t_1|s})(\boldsymbol{F}_{t_2} - \boldsymbol{F}^j_{t_2|s})')$ for $s, t, t_1, t_2 = 1, 2, \ldots, T$ (we write $P^j_{t|s}$ if $t_1 = t_2 = t$).
- The conditional moments of the unobserved factors can be written as,
$E_{\Theta(j)}[\boldsymbol{F}_t \boldsymbol{F}'_t | X] = \boldsymbol{F}^j_{t|T}(\boldsymbol{F}^j_{t|T})' + P^j_{t|T}$
$E_{\Theta(j)}[\boldsymbol{F}_{t-1} \boldsymbol{F}'_{t-1} | X] = \boldsymbol{F}^j_{t-1|T}(\boldsymbol{F}^j_{t-1|T})' + P^j_{t-1|T}$
$E_{\Theta(j)}[\boldsymbol{F}_{t-1} \boldsymbol{F}'_t | X] = \boldsymbol{F}^j_{t-1|T}(\boldsymbol{F}^j_{t|T})' + P^j_{t-1,t|T}$

Estimation of parameters

- $\boldsymbol{F}^j_{t|T}$, $P^j_{t|T}$,$\boldsymbol{F}^j_{t-1|T}$ and $P^j_{t-1,t|T}$ are estimated in iterative manner as
  $\boldsymbol{F}_1 = \boldsymbol{F}^j_{1|0} \overset{update}{\rightarrow} \boldsymbol{F}^j_{1|1} \overset{predict}{\rightarrow} \boldsymbol{F}^j_{2|1} \overset{update}{\rightarrow} \boldsymbol{F}^j_{2|2} \dashrightarrow \dashrightarrow$
  $\boldsymbol{F}^j_{T|T-1} \overset{update}{\rightarrow} \boldsymbol{F}^j_{T|T}$
  and, $P_1 = P^j_{1|0} \overset{update}{\rightarrow} P^j_{1|1} \overset{predict}{\rightarrow} P^j_{2|1} \overset{update}{\rightarrow} P^j_{2|2} \dashrightarrow \dashrightarrow$
  $P^j_{T|T-1} \overset{update}{\rightarrow} P^j_{T|T}$

Determination of r and p

- Rough estimate of r can be estimated from PCA and scree-plot.
- Information criteria based methods **Bai and Ng(2002),**
$IC_{r_1} = ln(V_r(\hat{\Lambda}, \hat{F})) + r.(\frac{n+T}{nT}) + ln(\frac{nT}{n+T})$
$IC_{r_2} = ln(V_r(\hat{\Lambda}, \hat{F})) + r.(\frac{n+T}{nT}) + ln(min(n, T))$
$IC_{r_3} = ln(V_r(\hat{\Lambda}, \hat{F})) + r\frac{ln(min(n,T))}{min(n,T)}$, where
$ln(V_r(\hat{\Lambda}, \hat{F})) = \frac{1}{nT}\sum_{t=1}^{T}(\boldsymbol{x}_t - \hat{\Lambda}\hat{\boldsymbol{F}}_t)'(\boldsymbol{x}_t - \hat{\Lambda}\hat{\boldsymbol{F}}_t)$
- The estimate $\hat{r}$ we get using this information based method is an consistent estimator for the true value of r , given some assumptions (**Bai and Ng ,2000**).

Determination of r and p

- Information criteria based measures for estimating p,
  $AIC(p) = ln(|\hat{\Sigma}_u(p)|) + \frac{2}{T} p\hat{r}^2$
  $BIC(p) = ln(|\hat{\Sigma}_u(p)|) + \frac{ln(T)}{T} p\hat{r}^2$
  $HQ(p) = ln(|\hat{\Sigma}_u(p)|) + \frac{2ln(ln(T))}{T} p\hat{r}^2$
  $SC(p) = ln(|\hat{\Sigma}_u(p)|) + \frac{ln(ln(T))}{T} p\hat{r}^2$
- Where, $\hat{\Sigma}_u(p) = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ and $\hat{u}_t = \hat{\boldsymbol{F}}_t - \sum_{i=1}^{p} \hat{A}_i \hat{\boldsymbol{F}}_{t-i}$
- In practice in many cases r and p 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.

Simulation Study: Data Generation

- We simulated data for various n $(10, 20, 50, 100)$ and T $(20, 50, 100, 200)$ for three Dynamic Factor Model with $r = 4, p = 1,\ r = 3, p = 1$and $r = 5, p = 1$.
- We have considered 200 replications for each combinations of (n, T, r, p)
- Then the data is fit to a model with $\widehat{p} = 1$ and $\widehat{r} = r$ and $r + 1$or $r - 1$.

Simulation Study: Fitting of the original series

- Here we measures how accurately the model fits the common component χ_t.
- $\mathscr{R}(\chi_t, \widehat{\chi}) = \frac{\sum_{i,t}(\chi_{i,t} - \widehat{\chi_{i,t}})^2}{\sum_{i,t}(\chi_{i,t})^2}$, this measure is used as a goodness of fit of the model.

Simulation Study: Estimation of the unobserved factors

- Here we measures how accurately the fitted model estimates the unobserved factors.
- $(\mathscr{T}(F, \widehat{F}) = \frac{Trace(F'\widehat{F}(\widehat{F'}\widehat{F})^{-1}\widehat{F'}F)}{Trace(F'F)}$, this measure is used as a measure of how accurately the factors are estimated. Where F is the $T \times r$ matrix of the factors. 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.

Simulation Study: Conclusion

- Both of the performance accuracy of the fit and accuracy of the estimation of unobserved factor increases as both (n, T) increases.
- It seems from values of the performance $\mathscr{R}(\chi_t, \widehat{\chi})$, the effect of number of series n is more dominant than the number of observations T.
- 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.

Application to Indian Economy:Data

- Here will will apply the Dynamic Factor Model to model Indian GDP growth.
- We have chosen 30 quarterly variables apart from GDP from different sectors, which are related to/ believed to be responsible for GDP growth.(**Stock and Watson (1989)**)
- GDP growth rate at time t is calculated by $y_t = \frac{GDP_{t+1} - GDP_t}{GDP_t}$.
- The data from 1996 (for some series 2000) to 2017 is used for estimation purpose and data for year 2018 and 2019 used for prediction purpose.
- Augmented Dickey–Fuller test is used to check for stationarity of the series and first and second differences ($\Delta x_t = x_t - x_{t-1}$, $\Delta^2(x_t) = \Delta(\Delta(x_t))$) are taken to make those series stationary.
- Among the initially selected 30 series/variables 9 series are selected using bidirectional step wise regression method.

Application to Indian Economy:Analysis

- Here we first estimated the factors and then using those factors model the GDP growth using multiple linear regression.
- We have considered two approaches. In Approach 1 we have used only those nine selected variables to get the estimates of the underlying factors. In Approach 2 along with those nine variables we also used the series of GDP growth rate to estimate the underlying factors.
- The approximate number of factors is decided using PCA, and for both the approaches $\widehat{r} = 5$. For estimating p information criteria based measures are used and $\widehat{p} = 1$.
- Then the underlying factors are estimated, say $\widehat{F}_i, i = 1, 2, 3, 4, 5$.

Application to Indian Economy:Analysis

Summary of the models from Approach 1 are given below.

- Model 1 :
  $y_t = \beta_0 + \beta_1 \widehat{F_{1t}} + \beta_2 \widehat{F_{2t}} + \beta_3 \widehat{F_{3t}} + \beta_4 \widehat{F_{4t}} + \beta_5 \widehat{F_{5t}} + \epsilon_t$
  (Multiple linear regression using the estimated factors)
- Model 2:
  $y_t = \beta_0 + \phi y_{t-1} + \beta_1 \widehat{F_{1t}} + \beta_2 \widehat{F_{2t}} + \beta_3 \widehat{F_{3t}} + \beta_4 \widehat{F_{4t}} + \beta_5 \widehat{F_{5t}} + \epsilon_t$
  (AR(1) model with covariates as the estimated factors)
- Model 3: $(y_t, \widehat{F_{1t}}, \widehat{F_{2t}}, \widehat{F_{3t}}, \widehat{F_{4t}}, \widehat{F_{5t}})' =$
  $A(1, y_{t-1}, \widehat{F_{1t-1}}, \widehat{F_{2t-1}}, \widehat{F_{3t-1}}, \widehat{F_{4t-1}}, \widehat{F_{5t-1}})' + \epsilon_t$ (VAR(p=1))
- Model 4:
  $y_t = \beta_0 + \beta_1 \widehat{F_{1t}} + \beta_2 \widehat{F_{2t}} + \beta_3 \widehat{F_{3t}} + \beta_4 \widehat{F_{4t}} + \gamma_5 \widehat{F_{5t}} + \gamma_6 w_1 + \beta_7 w_2 + \epsilon_t$
  (Multiple linear regression using the estimated factors along with two additional variables selected using step wise regression)

Application to Indian Economy:Analysis

Summary of the models from Approach 1 are given below.

Approach 1	R^2	Adjusted R^2	AIC
Model1	0.4249	0.3889	-520.6346
Model2	0.4268	0.3833	-518.92
Model3	0.1678 *	0.1038 *	
Model4	0.5885	0.542	-434.4233

* For GDP growth part only

Application to Indian Economy:Analysis

Summary of the models from Approach 2 are given below.

- Model 1 :
  $y_t = \beta_0 + \beta_1 \widehat{F_{1t}} + \beta_2 \widehat{F_{2t}} + \beta_3 \widehat{F_{3t}} + \beta_4 \widehat{F_{4t}} + \beta_5 \widehat{F_{5t}} + \epsilon_t$
  (Multiple linear regression using the estimated factors)
- Model 2: $y_t = \beta_0 + \beta_1 \widehat{F_{1t}} + \beta_2 \widehat{F_{2t}} + \beta_3 \widehat{F_{3t}} + \beta_4 \widehat{F_{4t}} + \gamma_5 \widehat{F_{5t}} + \gamma_6 w_1 + \beta_7 w_2 + \gamma_8 w_3 + \beta_9 w_4 + \gamma_{10} w_5 + \epsilon_t$ (Multiple linear regression using the estimated factors along with five additional variables selected using step wise regression)
- The main difference from the Approach 1 models are the estimated underlying factors.

Application to Indian Economy:Analysis

Summary of the models from Approach 2 are given below.

Approach 2	R^2	Adjusted R^2	AIC
Model1	0.6489	0.627	-563.0881
Model2	0.8712	0.8494	-509.7426

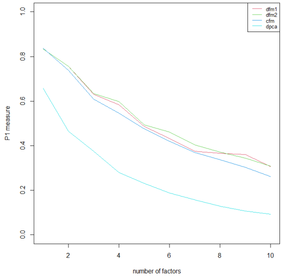
Application to Indian Economy:Comparisons of the Forecasts

- Used the estimated models to forecast the GDP growth rate for year 2018 and 2019.
Model	in sample RMSE	out of sample RMSE	in sample MAE	out of sample MAE
A1:Model 1	0.0108	0.0133	0.0086	0.0117
A1:Model 2	0.0107	0.0133	0.0087	0.012
A1:Model 3	0.0130	0.0137	0.0090	0.0128
A1:Model 4	0.0095	0.0136	0.0079	0.0127
A2:Model 1	0.0084	0.0134	0.0069	0.0122
A2:Model 2	0.0053	0.0131	0.0042	0.0118

Comparison with Other Dimensionality Reduction models:Overall comparison

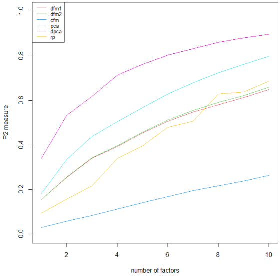
- We compared the performance of Dynamic Factor Model with other dimensionality reduction techniques like Classical factor model (non dynamic), Principal component analysis, Dynamic principal component analysis and Random projection method using the Indian economic data.
- We fitted the models for number of factors $r = 1, ..., 10$ and for DFM we used $p = 1, 2$.
- Here we compare how well the original series are estimated using those methods. For that we used the measure, $P_1 = \frac{\sum_{i,t}(x_{i,t} - \widehat{x_{i,t}})^2}{\sum_{i,t}(x_{i,t})^2}$. For a better fit the value of P_1should be smaller.
- Next we compare the proportion of variability explained by those methods. For this $P_2 = \frac{\sum_{i=1}^{r} var(F_i)}{\sum_{i=}^{n} var(x_i)}$ is used.

Comparison with Other Dimensionality Reduction models:Overall comparison



-
- 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.

Comparison with Other Dimensionality Reduction models:Overall comparison

- 

- Cumulative proportion of total variance explained by the factors is more or less same for two dynamic factor models. The cumulative variance explained by the classical factor model is very low. Here also the Dynamic PCA explains the most amount of variability of the data for each r.

Comparison with Other Dimensionality Reduction models:Model specific performance

- Using the estimated factors from the other methods, GDP growth is modeled using similar techniques as Model4 of Approach 1.
-

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

- We can see that here in GDP growth modeling and forecasting the DFM based method is performing better than other methods.(is=in sample, os=out of sample)

Thank You.