Dynamic Factor Models (DFMs)

- Basics of DFMs: formulation, identification
- Estimation in low-dimension; EM
- Estimation in high-dimension; PCA
- Applications

DFM I

Definition

A time series $\{\mathbf{X}_t\}_{t\in\mathbb{Z}} = \{(X_{j,t}), j = 1, \ldots, d\}_{t\in\mathbb{Z}}$ *is said to follow a dynamic factor model (DFM, in short) if*

$$\mathbf{X}_t = \Lambda \mathbf{F}_t + \boldsymbol{e}_t, \quad t \in \mathbb{Z}$$

where Λ *is a* $d \times r$ *matrix with* $r < d$ *and* $\mathbf{F}_t = (F_{j,t})$ *is a* r*-vector time-series following a VAR(*p*) model*

$$\mathbf{F}_t = \Phi_1 \mathbf{F}_{t-1} + \Phi_2 \mathbf{F}_{t-2} + \cdots + \Phi_p \mathbf{F}_{t-p} + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z}$$

- Λ: loading matrix
- $\{\mathbf{F}_t\}$: latent factors

DFM II

- For a fixed dimension j, it states that

$$X_{j,t} = \Lambda_{j,1} F_{1,t} + \cdots + \Lambda_{j,r} F_{r,t} + e_{j,t}$$

  Hence, $\{X_t\}$ is driven by common r factors.
- We assume $\mathbb{E}\mathbf{X}_t = \mathbf{0}$ for simplicity.
- $\chi_t = \Lambda F_t$ is called "common component" and $\{e_t\}$ are idiosyncratic errors. We always assume common and idiosyncratic components are *uncorrelated*, i.e.

$$\mathrm{Cov}(\chi_{i,t}, e_{j,s}) = 0, \quad t, s \in \mathbb{Z},\ i, j = 1, \ldots, d.$$

DFM III

- *Exact factor model* if idiosyncratic errors has no cross-sectional dependence.

$$\mathrm{Cov}(\mathbf{e}_t) = \mathrm{diag}(\sigma^2_{\mathbf{e},1}, \ldots, \sigma^2_{\mathbf{e},d})$$

- *Approximate factor model* if cross-sectional dependence is allowed.

$$\mathrm{Cov}(\mathbf{e}_t) = \mathbf{\Sigma}_{\mathbf{e}}$$

DFM IV

- Parameter identifiability

$$\begin{aligned}\mathbf{X}_t &= \Lambda \mathbf{F}_t + \boldsymbol{\epsilon}_t \\ &= \Lambda C C^{-1} \mathbf{F}_t + \boldsymbol{\epsilon}_t =: \tilde{\Lambda} \tilde{\mathbf{F}}_t + \boldsymbol{\epsilon}_t\end{aligned}$$

  for a non-singular $r \times r$ matrix C and $\{\mathbf{F}_t\}$ still satisfy VAR(p) equation. Indeed, e.g. VAR(1)

$$\begin{aligned}C^{-1}\mathbf{F}_t &= C^{-1}\Phi_1 C C^{-1}\mathbf{F}_{t-1} + C^{-1}\boldsymbol{\epsilon}_t \\ \tilde{\mathbf{F}}_t &= \tilde{\Phi}_1 \tilde{\mathbf{F}}_{t-1} + \tilde{\boldsymbol{\epsilon}}_t\end{aligned}$$

- Hence, it is often required to satisfy

$$\mathbb{E}\tilde{\mathbf{F}}_t \tilde{\mathbf{F}}_t' = I_r$$

  but still loading matrix is only identifiable up to an orthogonal matrix C ($CC' = C'C = I_r$)

DFM V

- Bai and Wang (2015) suggested to achieve identifiability by writing
$$\Lambda = \begin{pmatrix} I_r \\ B_{(d-r)\times r} \end{pmatrix}$$
that is, take C such that
$$\Lambda C = \begin{pmatrix} \Lambda^{(1)}_{r\times r} \\ \Lambda^{(2)} \end{pmatrix} C = \begin{pmatrix} \Lambda^{(1)}_{r\times r} C \\ \Lambda^{(2)} C \end{pmatrix} = \begin{pmatrix} I_r \\ B_{(d-r)\times r} \end{pmatrix}$$
- Identifiability explains why we call it as latent factors. We can estimate "common components" $\{\mathbf{X}_t\}$ consistently, but cannot separate loadings/factors in a unique way.
- Stationarity: We assume that the VAR(p) model is causal and stationary.

DFM VI

- Dynamic form of DFM assumes time-dependence in loadings,

$$\begin{aligned}\mathbf{X}_t &= \Lambda_0 \mathbf{F}_t + \Lambda_1 \mathbf{F}_{t-1} + \cdots + \Lambda_s \mathbf{F}_{t-s} + \mathbf{e}_t \\ &=: \Lambda(L)\mathbf{F}_t + \mathbf{e}_t\end{aligned}$$

- Without time dependence, this is a usual factor analysis and widely used in statistics, psychometrics (IQ!) for more than 50 years.
- DFM for a "high-dimensional" time-series is characterized by
  (1) Dimension reduction; few latent factors represents components.
  (2) Temporal/cross correlations; idiosyncratic errors & VAR(p) modeling of factors
- We will focus more on HDTS setting

Estimation in Low-Dimension

- DFM can be considered as "state-space" model

$$\begin{cases} \mathbf{X}_t^{(n)} = \lambda^{(n)}\mathbf{F}_t + \mathbf{e}_t^{(n)} & \text{observation equation} \\ \mathbf{F}_t = \Phi\mathbf{F}_{t-1} + \epsilon_t & \text{state equation} \end{cases}$$

- Using Kalmen fiter & EM algorithm, the maximum likelihood estimator can be computed.
- EM algorithm
  Expectation step: calculate

$$\mathbf{Q}(\Theta|\Theta^{j-1}) = \mathbb{E}_{\mathbf{X}|\mathbf{F},\Theta^{j-1}}(-2\log L_{\mathbf{X},\mathbf{F}}(\Theta))$$

  Maximize step: $\Theta^{(j)} = \arg\max_{\Theta} \mathbf{Q}(\Theta|\Theta^{j-1})$
- Kalman filter is heavily involved in calculating likelihood
- See Sumway and Stoeffer (2011) for details

Estimation in High-Dimension I

- Take dimension $d \to \infty$, and possibly $T \to \infty$
- Key references: Bai and Ng (2002, 2008), Stock and watson (2002, 2010), Doz et al. (2011,2012)
- Why is it useful in practice?
- VARMA viewpoint : Consider VAR(1)

$$
\begin{aligned}
\mathbf{X}_t &= \Lambda \mathbf{F}_t + \mathbf{e}_t = \Lambda(\Phi_1 \mathbf{F}_{t-1} + \boldsymbol{\epsilon}_t) + \mathbf{e}_t \\
&= \Lambda \left(\Phi_1 \Lambda^{-1}(\mathbf{X}_{t-1} - \mathbf{e}_{t-1}) + \boldsymbol{\epsilon}_t\right) + \mathbf{e}_t \\
&= \Lambda \Phi_1 \Lambda^{-1} \mathbf{X}_{t-1} + (\Lambda \boldsymbol{\epsilon}_t + \mathbf{e}_t - \Lambda \Phi_1 \Lambda^{-1} \mathbf{e}_{t-1}) \\
&=: \tilde{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\eta}_t
\end{aligned}
$$

Estimation in High-Dimension II

$\tilde{\Phi}_1$ has rank at most r, $\{\eta_t\}$ is VARMA error.

Sparse VAR model focuses on the sparse estimation of $\tilde{\Phi}$, but in DFM it is low-rank Φ_1 (dimension reduction).

- **Covariance viewpoint**

$$\mathrm{Var}(\mathbf{X}_t) := \mathbf{\Sigma_x} = \mathbb{E}\mathbf{X}_t\mathbf{X_t}^{'} = \Lambda(\mathbb{E}\mathbf{F}_t\mathbf{F_t}^{'})\Lambda^{'} + \mathbb{E}\mathbf{e}_t\mathbf{e_t}^{'}$$

$$= \Lambda\Lambda^{'} + \mathbf{\Sigma_e} \quad (\because \mathbb{E}\mathbf{F}_t\mathbf{F_t}^{'} = \mathbf{I}_r)$$

Under suitable assumptions

$$\mathbf{\Sigma_X} \approx \Lambda\Lambda^{'}$$

Since rank $(\Lambda\Lambda^{'}) =$ rank$(\Lambda) = r$, we also observe dimension reduction here.

Estimation in High-Dimension III

- Consider simple DFM model with

$$r = 2, \quad \Lambda = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \quad \mathbf{\Sigma_X} \approx \left(\begin{array}{c|c} I & 0 \\ \hline 0 & I \end{array} \right)$$

- Therefore, in practice, if the covariance (correlation) matrix is clustered, then factor models are plausible.

Estimation in High-Dimension IV

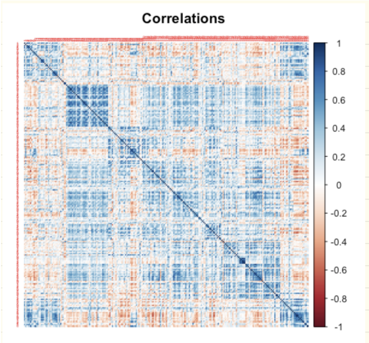

Figure: Correlaltion matrix from fmri series

Blessing of Dimensionality

- Blessing of dimensionality means that DFM becomes easier as $d \to \infty$.
- Consider cross-sectional average of DFM model ($r = 1$)

$$\bar{\mathbf{x}}_t = \frac{1}{d}\sum_{i=1}^{d} \mathbf{x}_{it} = \frac{1}{d}\sum_{i=1}^{d} \Lambda_i \mathbf{f}_t + \frac{1}{d}\sum_{i=1}^{d} \mathbf{e}_{it}$$

$$\mathrm{Var}(\mathbf{x}_t) = \Big(\frac{1}{d}\sum_{i=1}^{d} \Lambda_i\Big)^2 \mathrm{Var}(\mathbf{f}_t) + \frac{1}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d} \mathrm{Cov}(\mathbf{e}_{it}, \mathbf{e}_{jt})$$

$$= \Big(\frac{1}{d}\sum_{i=1}^{d} \Lambda_i\Big)^2 \cdot 1 + \frac{1}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d} \mathrm{Cov}(\mathbf{e}_{it}, \mathbf{e}_{jt}) \to \overline{\lambda}^2$$

  under suitable conditions on $\{\mathbf{e}_t\}$.
- $\mathrm{Var}(\bar{\mathbf{x}}_t) \longrightarrow \overline{\lambda}^2 = \mathrm{Var}(\bar{\chi}_t)$ as $d \to \infty$. This means that all information contained in the factors as measured by 2nd moment can be captured by aggregation of sample data.

PCA Estimation

- The principal component analysis (PCA) estimators of loadings and factors. Suppose the number of factors r is given and let Sample covariance matrix

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{X}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X_t}'$$

Step1 Diagonalize $\hat{\boldsymbol{\Sigma}}_{\mathbf{X}} = \hat{\mathbf{U}}\hat{\boldsymbol{\Pi}}\hat{\mathbf{U}}'$

$$\begin{array}{ll} \boldsymbol{\Pi} = \mathrm{diag}(\hat{\pi}_1, \ldots, \hat{\pi}_d), & \hat{\pi}_1 \geq \hat{\pi}_2 \geq \ldots \geq \hat{\pi}_d \text{ eigenvalues} \\ \hat{\mathbf{U}} = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_d), & \text{orthonormal eigenvectors} \end{array}$$

Step2 Then, take $r < d$ largest eigenvalues/eigenvectors

$$\widehat{\Lambda} = \sqrt{d}(\widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_r), \quad \widehat{\mathbf{F}}_t = \frac{1}{d}\hat{\Lambda}'\mathbf{X}_t$$

PCA Estimation

- PCA factors are estimated by the cross-sectional weighted average of $\{\mathbf{X}_t\}$.
- PCA estimator can be viewed as

$$\underset{F_1,\ldots,F_T,\Lambda}{\operatorname{argmin}} \frac{1}{dT}\sum_{t=1}^{T}(\mathbf{X}_t - \Lambda\mathbf{F}_t)'(\mathbf{X}_t - \Lambda\mathbf{F}_t)$$

subject to normalization

$$\frac{1}{d}\Lambda'\Lambda = I_r.$$

- First minimize over $\mathbf{F}_t$ given Λ gives

$$\widehat{\mathbf{F}}_t = (\Lambda'\Lambda)^{-1}\Lambda'\mathbf{X}_t$$

PCA Estimation

- Plug-in $\hat{\mathbf{F}}_t$ gives

$$
\begin{aligned}
& \underset{\Lambda}{\operatorname{argmin}} \ \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\left(I-\Lambda(\Lambda'\Lambda)^{-1}\Lambda'\right)\mathbf{X}_t \\
\Longleftrightarrow\ & \underset{\Lambda}{\operatorname{argmax}} \ \operatorname{tr}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\Lambda(\Lambda'\Lambda)^{-1/2}(\Lambda'\Lambda)^{-1/2}\Lambda'\mathbf{X}_t\right) \\
\Longleftrightarrow\ & \underset{\Lambda}{\operatorname{argmax}} \ \operatorname{tr}\left((\Lambda'\Lambda)^{-1/2}\Lambda' \quad \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t' \quad \Lambda(\Lambda'\Lambda)^{-1/2}\right) \\
\Longleftrightarrow\ & \underset{\Lambda}{\operatorname{argmax}} \ \Lambda'\mathbf{\Sigma}_X\Lambda
\end{aligned}
$$

PCA Estimator

- Consistency is proved under various conditions. See Bai and Ng (2008), e.g. $d \to \infty, T \to \infty$ and $d^2/T \to \infty$.
- Generalized PCA estimation. Take idiosyncratic error $\mathbf{\Sigma_e}$ into account

$$\underset{\mathbf{F}_1,\ldots,\mathbf{F}_T,\Lambda}{\operatorname{argmin}} \quad \frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t - \Lambda\mathbf{F}_t)'\mathbf{\Sigma_e}^{-1}(\mathbf{X}_t - \Lambda\mathbf{F}_t)$$

subject to normalization $d^{-1}\Lambda'\Lambda = I_r$. Then, $\widetilde{\Lambda}$ is the r largest eigenvectors from $\mathbf{\Sigma_e}^{-1/2}\hat{\mathbf{\Sigma}}_X\mathbf{\Sigma_e}^{-1/2}$.

Estimation of the Number of Factors r

- Information criteria is widely used
- Define sum of squared residuals using r-factors

$$S(r) = \frac{1}{dT}\sum_{t=1}^{T}\|\mathbf{X}_t - \hat{\Lambda}^r\hat{\mathbf{F}}_t^r\|^2$$

  Then, IC criteria is given by

$$\widehat{r} = \operatorname*{argmin}_r \mathrm{IC}(r),$$

  where

$$\mathrm{IC}(r) = \log S(r) + r \cdot g(d, T).$$

- Penalty function $g(d, T)$ need some theoretical properties to achieve consistency,

$$P(\widehat{r} \to r) \to 1,$$

Number of Factors

- For example, $d \wedge T = \min(d, T)$,

$$g_1(d,T) = \frac{d+T}{dT}\log\left(\frac{dT}{d+T}\right), \quad g_2(d,T) = \frac{d+T}{dT}\log\,(d \wedge T)^2$$

$$g_3(d,T) = \frac{\log\,(d \wedge T)^2}{(d \wedge T)^2}$$

- In practice, we also use scree plot of eigenvalues

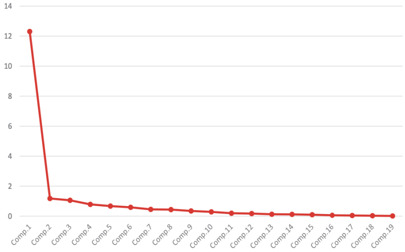

Figure 1. Scree plot for 19 RVs except the KOSPI. Elbow point 2 is selected for the number of factors.

Forecasting

- DFMs seem to work well in forecasting. Since $\{\mathbf{e}_t\}$ is independent with $\{\mathbf{F}_t\}$ across time. Note that

$$\mathbb{E}(\mathbf{X}_{t+1}|\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}, \ldots)$$

$$= \Lambda\mathbb{E}(\mathbf{F}_{t+1}|\mathbf{X}_t, \mathbf{F}_t, \mathbf{X}_{t-1}, \mathbf{F}_{t-1}, \ldots)$$

$$= \Lambda(\Phi_1\mathbf{F}_t + \Phi_2\mathbf{F}_{t-1} + \ldots + \mathbf{\Phi_p F_{t-p}})$$

- That is, plug-in factor estimates from VAR(p) gives you a factor forecasts. Finally, plug-into DFM equation gives you the final estimate

$$\widehat{\mathbf{X}}_{t+1} = \widehat{\Lambda}\widehat{\mathbf{F}}_t.$$

Applications: FAVAR

- Factor-augmented vector autoregression (FAVAR) by Bernanke, Boivin and Eliasz (2005).
- Use factors to improve VAR forecasting.

$$\begin{pmatrix} \mathbf{F}_t \\ \mathbf{X}_t \end{pmatrix} = \mathbf{\Phi}(\mathbf{L}) \begin{pmatrix} \mathbf{F}_{t-1} \\ \mathbf{X}_{t-1} \end{pmatrix} + \mathbf{e}_t$$

- Interested in how economic shocks affect monetary policy.
- For example, factors $\{\mathbf{F}_t\}$ is used to provide summarized information on various macroeconomic time series information.
- $\mathbf{X}_t$; monetary policy measured by federal funds rates.

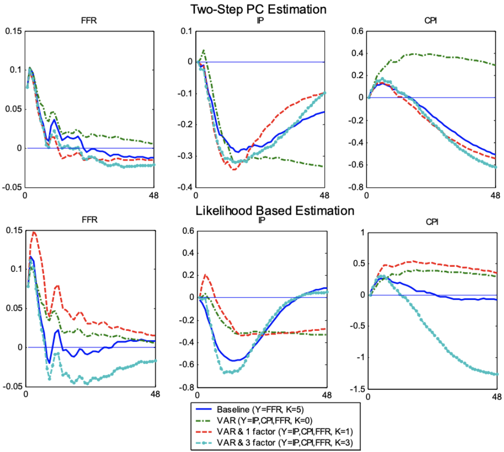

Figure 5. VAR – FAVAR comparison. The top panel displays estimated responses for the two-step principal component estimation and the bottom panel for the likelihood based estimation.

References

- Bai, J. and Wang, P. (2015), 'Identification and Bayesian Estimation of Dynamic Factor Models', Journal of Business & Economic Statistics 33(2), 221–240.
- J. Bai, S. Ng, et al. Large dimensional factor analysis. Foundations and Trends in Econometrics, 3(2):89–163, 2008.
- B. S. Bernanke, J. Boivin, and P. Eliasz. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. The Quarterly journal of economics, 120(1):387–422, 2005.
- C. Doz, D. Giannone, and L. Reichlin. A two-step estimator for large approximate dynamic factor models based on kalman filtering. Journal of Econometrics, 164(1):188–205, 2011.
- C. Doz, D. Giannone, and L. Reichlin. A quasi–maximum likelihood approach for large, approximate dynamic factor models. Review of economics and statistics, 94(4):1014–1024, 2012.
- Shumway, R. H. and Stoffer, D. S. (2006), Time Series Analysis and Its Applications: With R examples, Springer.
- Stock, J. H. and Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', Journal of the American Statistical Association 97(460), 1167–1179.
- J. H. Stock and M. Watson. Dynamic factor models. Oxford Handbooks Online, 2011.