Dynamic Factor Models

Francis J. DiTraglia

University of Pennsylvania

Econ 722

Survey Articles on Dynamic Factor Models

Stock & Watson (2010)

Best general overview of dynamic factor models and applications.

Bai & Ng (2008)

Comprehensive review of large-sample results for high-dimensional factor models estimated via PCA.

Stock & Watson (2006)

Handbook chapter on forecasting with many predictors. One section is devoted to dynamic factor models.

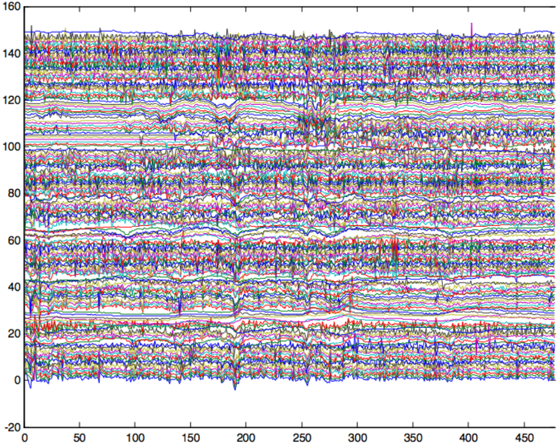
Breitung & Eickmeyer (2006)

Brief overview with an application to Euro-area business cycles.

The Basic Idea

We're interested in settings with a large number of time series N and a comparable number of time periods T.

Example: Stock and Watson Dataset



Monthly Macroeconomic Indicators: $N > 200$, $T > 400$

Why Factor Models?

1. Factors could be intrinsically interesting if they arise from a theoretical model (e.g. Financial Economics)
2. Many variables without running out of degrees of freedom
   - More information could improve forecasts/macro analysis
   - Mimic central banks "looking at everything"
3. Eliminate measurement error and idiosyncratic shocks to provide more reliable information for policy
4. "Remain Agnostic about the Structure of the Economy"
   - Advantages over SVARs: don't have to choose variables to control degrees of freedom, and can allow fewer underlying shocks than variables.

Last Time: Classical Factor Analysis Model

I've eliminated the mean μ and renamed the factor F_t

$$\underset{(N\times 1)}{X_t} = \Lambda \underset{(r\times 1)}{F_t} + \epsilon_t$$

$$\begin{bmatrix} F_t \\ \epsilon_t \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & 0 \\ 0 & \Psi \end{bmatrix} \right)$$

$\Lambda =$ matrix of factor loadings

$\Psi =$ diagonal matrix of idiosyncratic variances.

Adding Time-Dependence

$$\underset{(N\times 1)}{X_t} = \Lambda \underset{(r\times 1)}{F_t} + \epsilon_t$$

$$\underset{(r\times 1)}{F_t} = A_1 F_{t-1} + \ldots + A_p F_{t-p} + u_t$$

$$\begin{bmatrix} u_t \\ \epsilon_t \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & 0 \\ 0 & \Psi \end{bmatrix} \right)$$

Some (Confusing) Terminology

Caution: different authors use different conventions!

Static	X_t depends only on F_t
Dynamic	X_t depends on lags of F_t as well
Exact	Ψ is diagonal and ϵ_t independent over time
Approximate	Some cross-sectional & temporal dependence in ϵ_t

The model I wrote down on the previous slide is sometimes called an "exact, static factor model" even though it has dynamics! I'll still call it a dynamic factor model...

Editorial: This is all a bit confused...

1. The difference between “static” and “dynamic” is unclear
   - Can write dynamic model as a static one with more factors
   - Static representation involves “different” factors, but we may not care: are the factors “real” or just a data summary?
2. Not really possible to allow cross-sectional dependence in ϵ_t
   - Unless the off-diagonal elements of Ψ are close to zero we can’t tell them apart from the common factors
   - “Approximate” factor models basically assume conditions under which the off-diagonal elements of Ψ are negligible
   - Similarly, time series dependence in ϵ_t can’t be very strong (stationary ARMA is ok)

Methods of Estimation for Dynamic Factor Models

1. Bayesian Estimation
2. Maximum Likelihood: EM-Algorithm + Kalman Filter
   - Watson & Engle (1983)
   - Ghahramani & Hinton (1996)
   - Jungbacker & Koopman (2008)
   - Doz, Giannone & Reichlin (2012)
3. "Nonparametric" Estimation
   - Just carry out PCA on X and ignore the time-series element
   - The first r PCs are our estimates $\widehat{F}_t$
   - Essentially treats F_t as an r-dimensional *parameter* to be estimated from an N-dimensional observation X_t

Estimation by PCA

PCA Normalization

- $F'F/T = I_r$ where $F = (F_1, \ldots, F_T)'$
- $\Lambda'\Lambda = \text{diag}(\mu_1, \ldots, \mu_r)$ where $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r$

Assumption I

Factors are *pervasive*: $\Lambda'\Lambda/N \to D_\Lambda$ an $(r \times r)$ full rank matrix.

Assumption II

max e-value $E[\epsilon_t \epsilon_t'] \leq c \leq \infty$ for all N.

Upshot of the Assumptions

If we average over the cross-section, the contribution from the factors persists and the contribution from the idiosyncratic terms disappears as $N \to \infty$.

Key Result for PCA Estimation

Under the assumptions on the previous slide and some other technical conditions, the first r PCs of X consistently estimate the space spanned by the factors as $N, T \to \infty$.

Doz, Giannone & Reichlin (2012)

Typical Justifications for PCA Approach

- Consistent estimation of factors under very weak assumptions
- MLE is computationally infeasible for large N

But Neither is True!

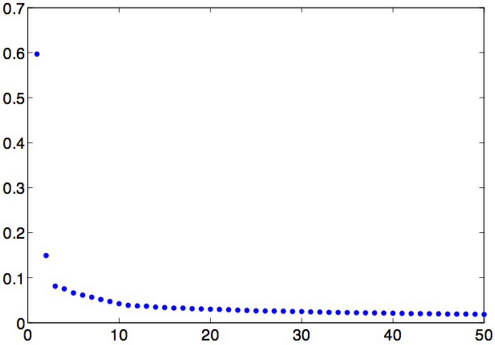
- EM-algorithm + Kalman Filter is *very efficient* – complexity depends on number of *factors*, not number of series
- Treat exact, static factor model (the one I wrote out) as a mis-specified *approximating model* (Quasi-MLE)
- Identical large-sample results as PC under similar assumptions, but better finite-sample properties and temporal smoothing

Choosing the Number of Factors

If we use Likelihood-based or Bayesian estimation, we could try to resort to the familiar tools from earlier in the semester.There are a lot of parameters in factor models, however, so the asymptotic approximations (I'm looking at you, AIC) could be poor.

Choosing the Number of Factors – Scree Plot

If we use PC estimation, we can look a something called a "scree plot" to help us decide how many PCs to include:



This figure depicts the eigenvalues for an $N = 1148$, $T = 252$ dataset of excess stock returns

Choosing the Number of Factors – Bai & Ng (2002)

Choose r to minimize an information criterion:

$$IC(r) = \log V_r(\widehat{\Lambda}, \widehat{F}) + r \cdot g(N, T)$$

where

$$V_r(\Lambda, F) = \frac{1}{NT} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t)$$

and g is a penalty function. The paper provides conditions on the penalty function that guarantee consistent estimation of the true number of factors.

Onatski (2013)

Nobody volunteered to present this paper! Unlike Bai & Ng (2002), the idea here is *not* to consistently estimate the number of factors. Instead, and in the spirit of AIC and C_p, we try to reconstruct the factors with minimum quadratic loss. The results in the paper apply both to "strong" and "weak" factor asymptotics.

What Can We Do with Factors?

Among other possibilities:

1. Use them as Instrumental Variables
2. Use them to construct Forecasts
3. Use them to “Augment” a VAR

Factors as Instruments – Bai & Ng (2010)

Endogenous Regressors x_t

$$y_t = x_t'\beta + \epsilon_t \qquad E[x_t\epsilon_t] \neq 0$$

Unobserved Variables F_t are Strong IVs

$$\underset{(k\times 1)}{x_t} = \Psi' \underset{(r\times 1)}{F_t} + u_t \qquad E[F_t\epsilon_t] = 0$$

Observe Large Panel $(z_{1t}, \ldots, z_{Nt})$

$$z_{it} = \lambda_i' F_t + e_{it}$$

Factors as Instruments – Bai & Ng (2010)

$$\boxed{y_t = x_t'\beta + \epsilon_t, \qquad x_t = \Psi' F_t + u_t, \qquad z_{it} = \lambda_i' F_t + e_{it}}$$

Procedure

1. Calculate the PCs of Z
2. Calculate $\widetilde{F}_t$ using the first r PCs of Z
3. Use $\widetilde{F}_t$ in place of F_t for IV estimation

Main Result

Under certain assumptions, as $(N, T) \to \infty$ "estimation and inference can proceed as though F_t were known." The resulting estimator is consistent and asymptotically normal.

Factors as Instruments – Bai & Ng (2010)

Why Might This be Helpful?

1. Avoid many instruments bias
2. Avoid bias from irrelevant instruments
3. Allow more observed instruments z_{it} than sample size T
4. Provided that $\sqrt{T}/N \to 0$, all of the observed instruments z_{it} can be *endogenous* as long as F_t is exogenous

Forecasting with Dynamic Factors

- Similar to Principal Components Regression (PCR)
- Estimate factors $\widehat{F}_t$ from a large number of regressors X_t
- Run a regression to forecast y_t using $\widehat{F}_t$ rather than X_t

We'll talk about this more next time and compare to other high-dimensional forecasting procedures.

FAVARs – Bernanke, Boivin & Eliasz (2005)

Two Problems with Structural VARs

1. Number of parameters is *quadratic* in the number of variables. Unrestricted VAR infeasible unless T is large relative to N.
   - You've studied one solution to this problem already this semester: Bayesian Estimation with informative priors
2. To keep estimation tractable we typically use a small number of variables, but then the VAR innovations "might not span the space of structural shocks."

FAVARs – Bernanke, Boivin & Eliasz (2005)

Factor-Augmented VAR Model

$$\begin{bmatrix} Y_t \\ F_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t$$

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t$$

$\underset{(M\times 1)}{Y_t}$ = observable variables that "drive dynamics of the economy"

$\underset{(K\times 1)}{F_t}$ = Small # of unobserved factors: "additional information"

$\underset{(N\times 1)}{X_t}$ = Large # of observed "informational time series"

FAVARs – Bernanke, Boivin & Eliasz (2005)

$$\boxed{\begin{bmatrix} Y_t \\ F_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t \qquad\qquad X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t}$$

Consider Two Estimation Procedures

1. Two-step Procedure:
   - Estimate space spanned by factors using first $K + M$ PCs of X
   - Estimate VAR with $\widehat{F}_t$ in place of F_t
2. Full Bayes (Gibbs Sampler)

Empirical Application

Additional information contained in FVAR is "important to properly identify the monetary transmission mechanism."