UNIVERSITY OF TEXAS ARLINGTON

Bayesian Hierarchical Dynamic Factor Models

Anthony M. Thomas, Jr.
anthony.thomas3@mavs.uta.edu

Department of Mathematics
The University of Texas at Arlington

September 1, 2020

Contents

1. Background
   - Bayesian Inference
   - Factor Analysis
   - Dynamic Factor Analysis
2. Hierarchical Dynamic Factor Analysis
3. Variational Bayesian Inference
   - Orthogonal Factor Model
4. Future Work
5. References

Section 1

1. Background
   - Bayesian Inference
   - Factor Analysis
   - Dynamic Factor Analysis
2. Hierarchical Dynamic Factor Analysis
3. Variational Bayesian Inference
   - Orthogonal Factor Model
4. Future Work
5. References

Bayesian Inference

Bayesian Inference can be described by two parts:

1. Build a model based on data $\boldsymbol{X}$ and parameters $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2\}$
   - Likelihood: $p(\boldsymbol{X}|\boldsymbol{\Theta})$
   - Prior: $p(\boldsymbol{\Theta})$
2. Compute the posterior
   - Posterior:
   $$p(\boldsymbol{\Theta}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\boldsymbol{X})}$$
   - Report summaries, e.g. posterior expectations
   $$\mathbb{E}[h(\boldsymbol{\Theta})|\boldsymbol{X}]$$
   or marginal posterior expectations
   $$\mathbb{E}[h(\boldsymbol{\Theta}_i)|\boldsymbol{X}]$$

Factor Analysis

- **Factor Analysis** (FA) is a method that assumes that the covariance structure of a set of observations can be described in terms of a linear combination of unobservable (latent) variables called factors

- Typical uses of FA:
  1. Dimension reduction: explains covariation between N variables using $K < N$ factors
  2. Data interpretation: finds factors that explain the covariation
  3. Theory testing: tests whether a hypothesized factor structure fits observed data

Time Series Factor Analysis

FA can be extended to time series data. For the remainder of the presentation we will assume that we are working with time series data.

For a sample of T time periods we have the following:

- $\boldsymbol{F}_t = (F_{1t}, \ldots, F_{Kt})^\top$ denotes the factors
- $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{Nt})^\top$ denotes the observations

Time Series Factor Analysis

The measurement equation that relates the observations to the factors is given by

$$\boldsymbol{X}_t = \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{e}_t \tag{1}$$

where

- $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{Nt})^\top$ denotes a vector of observations
- $\boldsymbol{\Lambda} = [\lambda_{nk}]_{N \times K}$ denotes a matrix of factor loadings that act as weights
- $\boldsymbol{F}_t = (F_{1t}, \ldots, F_{Kt})^\top$ denotes a vector of factors
- $\boldsymbol{e}_t$ is a vector of measurement errors, disturbances, and idiosyncratic (unique) factors

Dynamic Factor Analysis

Dynamic factor models allow the factors to not only affect the observations contemporaneously, but also through their lags:

$$\boldsymbol{X}_t = \boldsymbol{\Lambda}(L)\boldsymbol{F}_t + \boldsymbol{\epsilon}_t \tag{2}$$

where

$$\boldsymbol{\Lambda}(L) = \boldsymbol{\Lambda}_0 + \boldsymbol{\Lambda}_1 L + \cdots + \boldsymbol{\Lambda}_s L^s$$

is a distributed lag matrix polynomial of order s in the lag operator L and

$$L^s \boldsymbol{F}_t = \boldsymbol{F}_{t-s}.$$

The observation equation can be expressed instead as

$$\boldsymbol{X}_t = \boldsymbol{\Lambda}_0 \boldsymbol{F}_t + \boldsymbol{\Lambda}_1 \boldsymbol{F}_{t-1} + \cdots + \boldsymbol{\Lambda}_s \boldsymbol{F}_{t-s} + \boldsymbol{\epsilon}_t$$

Dynamic Factor Analysis

The factors are usually assumed to follow a vector autoregressive process:

$$\boldsymbol{\Phi}(L)\boldsymbol{F}_t = \boldsymbol{\varepsilon}_t \tag{3}$$

where

$$\boldsymbol{\Phi}(L) = \mathbf{I}_K - \boldsymbol{\Phi}_1 L - \cdots - \boldsymbol{\Phi}_p L^p$$

is a matrix polynomial of autoregressive coefficients in the lag operator L.

Section 2

1. Background
   - Bayesian Inference
   - Factor Analysis
   - Dynamic Factor Analysis
2. Hierarchical Dynamic Factor Analysis
3. Variational Bayesian Inference
   - Orthogonal Factor Model
4. Future Work
5. References

Hierarchical Dynamic Factor Model

In [1], the authors generalize the two-level dynamic factor model by positing that for each t, the nth series in a given block b, denoted by X_{bnt}, has three sources of variation:

1. idiosyncratic
2. block-specific
3. common

Hierarchical Dynamic Factor Model

A three-level representation of the data is given as

$$X_{bnt} = \boldsymbol{\lambda}^{n}_{G.b}(L)\boldsymbol{G}_{bt} + e_{Xbnt} \tag{4}$$

$$\boldsymbol{G}_{bt} = \boldsymbol{\Lambda}_{F.b}(L)\boldsymbol{F}_{t} + \boldsymbol{e}_{Gbt} \tag{5}$$

$$\boldsymbol{\Psi}_{F}(L)\boldsymbol{F}_{t} = \boldsymbol{\epsilon}_{Ft}, \tag{6}$$

where

- $\boldsymbol{\lambda}^{n}_{G.b}(L)$ denotes a distributed lag polynomial of block-level factor loadings
- $\boldsymbol{\Lambda}_{F.b}(L)$ denotes a distributed lag matrix polynomial of common factor loadings
- $\boldsymbol{G}_{bt} = (G_{b1t}, \ldots, G_{bK_{G}bt})^{\top}$ denotes the block-level factors
- $\boldsymbol{F}_{t} = (F_{1t}, \ldots, F_{K_{F}t})^{\top}$ denotes the common factors

Hierarchical Dynamic Factor Model

For some blocks, it may be appropriate to break up the data into subblocks, which adds another source of variation. Let Z_{bsnt} be the nth series in subblock s of block b. A four-level representation of the subblock data is given as

$$\begin{aligned} Z_{bsnt} &= \boldsymbol{\lambda}^{n}_{H.bs}(L)\boldsymbol{H}_{bst} + e_{Zbsnt} \\ \boldsymbol{H}_{bst} &= \boldsymbol{\Lambda}_{G.bs}(L)\boldsymbol{G}_{bt} + \boldsymbol{e}_{Hbst} \\ \boldsymbol{G}_{bt} &= \boldsymbol{\Lambda}_{F.b}(L)\boldsymbol{F}_{t} + \boldsymbol{e}_{Gbt} \\ \boldsymbol{\Psi}_{F}(L)\boldsymbol{F}_{t} &= \boldsymbol{\epsilon}_{Ft} \end{aligned}$$

where

- $\boldsymbol{\lambda}^{n}_{H.bs}(L)$ denotes a distributed lag polynomial of subblock-level factor loadings
- $\boldsymbol{\Lambda}_{G.bs}(L)$ denotes a distributed lag matrix polynomial of block-level factor loadings
- $\boldsymbol{H}_{bst} = (H_{bs1t}, \ldots, H_{bsK_{Hbs}t})^{\top}$ denotes the subblock-level factors

Hierarchical Dynamic Factor Model

The idiosyncratic components, the subblock-specific, block-specific, and common factors are assumed to be stationary, Gaussian autoregressive processes of orders q_{Zbsn}, q_{Xbn}, q_{Hbsi}, q_{Gbj}, and q_{Fk}, respectively, i.e.

$$\begin{aligned}
\psi_{Z.bsn}(L)e_{Zbsnt} &= \epsilon_{Zbsnt}, \quad \epsilon_{Zbsnt} \sim \mathcal{N}\left(0, \sigma^2_{Zbsn}\right) \quad n = 1, \ldots, N_{bs} \\
\psi_{X.bn}(L)e_{Xbnt} &= \epsilon_{Xbnt}, \quad \epsilon_{Xbnt} \sim \mathcal{N}\left(0, \sigma^2_{Xbn}\right) \quad n = 1, \ldots, N_{b} \\
\psi_{H.bsi}(L)e_{Hbsit} &= \epsilon_{Hbsit}, \quad \epsilon_{Hbsi} \sim \mathcal{N}\left(0, \sigma^2_{Hbsi}\right) \quad i = 1, \ldots, K_{Hbs} \\
\psi_{G.bj}(L)e_{Gbjt} &= \epsilon_{Gbjt}, \quad \epsilon_{Gbjt} \sim \mathcal{N}\left(0, \sigma^2_{Gbj}\right) \quad j = 1, \ldots, K_{Gb} \\
\psi_{F.k}(L)F_{kt} &= \epsilon_{Fkt}, \quad \epsilon_{Fkt} \sim \mathcal{N}\left(0, \sigma^2_{Fk}\right) \quad k = 1, \ldots, K_{F}
\end{aligned}$$

Hierarchical Dynamic Factor Model

- Not all series need to belong to blocks and subblocks, hence the data used in a four-level model are a mixture of Z_{bsnt}, X_{bnt}, and X_{nt}
- For brevity, we work with only the three-level model for which all data belongs to a block, therefore data will consist of X_{bnt} for $b = 1, \ldots, B$

Hierarchical Dynamic Factor Model

From [1], this posterior distribution of the factors and model parameters is estimated via a Gibbs sampler. The main steps are outlined:

Algorithm 1: Gibbs Sampler for a Hierarchical DFM

Input: Data $\{\boldsymbol{X}_t\}$

Output: Sample from posterior distribution over factors and model parameters

Initialize: $\{\boldsymbol{F}_t\}^{(0)}, \{\boldsymbol{G}_t\}^{(0)}, \{\boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}^{(0)}$

for $i = 1, \ldots, M$ **do**

$$\{\boldsymbol{G}_t\}^{(i)} \sim \{\boldsymbol{G}_t\} | \{\boldsymbol{F}_t\}^{(i-1)}, \{\boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}^{(i-1)}, \{\boldsymbol{X}_t\}$$

$$\{\boldsymbol{F}_t\}^{(i)} \sim \{\boldsymbol{F}_t\} | \{\boldsymbol{G}_t\}^{(i)}, \{\boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}^{(i-1)}, \{\boldsymbol{X}_t\}$$

$$\{\boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}^{(i)} \sim \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Sigma} | \{\boldsymbol{F}_t\}^{(i)}, \{\boldsymbol{G}_t\}^{(i)}, \{\boldsymbol{X}_t\}$$

end

return $\left\{ \{\boldsymbol{F}_t\}^{(i)}, \{\boldsymbol{G}_t\}^{(i)}, \{\boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}\}^{(i)} \right\}_{i=1}^{M}$

Section 3

1. Background
   - Bayesian Inference
   - Factor Analysis
   - Dynamic Factor Analysis
2. Hierarchical Dynamic Factor Analysis
3. Variational Bayesian Inference
   - Orthogonal Factor Model
4. Future Work
5. References

Why Variational Bayesian Inference?

Consider a model with data $\boldsymbol{X}$ and latent variables $\boldsymbol{Z} = \{\boldsymbol{Z}_1, \boldsymbol{Z}_2\}$. The goal is to compute the joint posterior of the latent variables given the data

$$p(\boldsymbol{Z}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{Z})p(\boldsymbol{Z})}{p(\boldsymbol{X})} \tag{7}$$

- Likelihood: $p(\boldsymbol{X}|\boldsymbol{Z})$
- Prior: $p(\boldsymbol{Z})$
- Evidence: $p(\boldsymbol{X})$

$$p(\boldsymbol{X}) = \int p(\boldsymbol{X}, \boldsymbol{Z})\, d\boldsymbol{Z} \tag{8}$$

Why Variational Bayesian Inference?

For complex models (7) and (8) typically have no closed-form. As a result, the joint and marginal posteriors have to be approximated. Markov chain Monte Carlo (MCMC) has been the gold standard to solve this problem.

1. Construct an ergodic Markov chain on $\boldsymbol{Z}$ whose stationary distribution is the joint posterior $p(\boldsymbol{Z}|\boldsymbol{X})$
2. Sample from the chain to collect samples from the stationary distribution
3. Approximate the posterior with an empirical estimate constructed from a subset of the collected samples
4. Use the subset of collected samples to estimate expectations of interest

Why Variational Bayesian Inference?

A major disadvantage for MCMC is that high levels of accuracy requires a great deal of time

1. MCMC is typically slow, but ultimately accurate
2. VB is typically much faster

Variational Bayesian Inference

Rather than use sampling, VB uses optimization to find an approximation to the posterior.

1. Posit a family of "nice" approximate densities $\mathcal{Q}$
2. Find a member of that family that is "closest" to the exact posterior, i.e.

$$q^\star(\boldsymbol{Z}) = \underset{q(\boldsymbol{Z})\in\mathcal{Q}}{\arg\min}\, \mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) \tag{9}$$

where

$$\mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) = \mathbb{E}_q[\log q(\boldsymbol{Z})] - \mathbb{E}_q[\log p(\boldsymbol{Z}|\boldsymbol{X})] \tag{10}$$

Variational Bayesian Inference

It follows that

$$\mathrm{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) = \mathbb{E}_q[\log q(\boldsymbol{Z})] - \mathbb{E}_q[\log p(\boldsymbol{X}, \boldsymbol{Z})] + \log p(\boldsymbol{X})$$

This reveals that the objective in (9) depends on the evidence, thus it cannot be computed directly.

Variational Bayesian Inference

Instead, we optimize an alternative objective that is equivalent to (9) up to an added constant called the **evidence lower bound** (ELBO)

$$\text{ELBO}(q) = \mathbb{E}_q[\log p(\boldsymbol{X}, \boldsymbol{Z})] - \mathbb{E}_q[\log q(\boldsymbol{Z})] \tag{11}$$

It can be shown that

$$\arg\min_{q(\boldsymbol{Z}) \in \mathcal{Q}} \text{KL}(q(\boldsymbol{Z}) \,||\, p(\boldsymbol{Z}|\boldsymbol{X})) = \arg\max_{q(\boldsymbol{Z}) \in \mathcal{Q}} \text{ELBO}(q)$$

Variational Bayesian Inference

The VB framework is now

1. Posit a family of "nice" approximate densities $\mathcal{Q}$
2. Find a member of that family that is "closest" to the exact posterior, i.e.

$$q^\star(\boldsymbol{Z}) = \underset{q(\boldsymbol{Z})\in\mathcal{Q}}{\arg\max}\, \mathrm{ELBO}(q) \tag{12}$$

Solving this optimization problem is still difficult in general

- Using the mean-field assumption can make it easier

Mean-Field Assumption

1. Partition the latent variables into M groups, say $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M$
2. Assume that the distributions in $\mathcal{Q}$ factorize across the groups, i.e.

$$\mathcal{Q} = \left\{ q : q(\boldsymbol{Z}) = \prod_{m=1}^{M} q_m(\boldsymbol{Z}_m) \right\}$$

3. Learning the optimal q now reduces to learning the optimal $q_1, \ldots, q_M$
4. Straightforward to optimize via coordinate ascent
5. This is **NOT** a modeling assumption

Mean-Field Assumption

Interestingly, under the mean-field assumption, the optimization problem for a single q_m has the solution:

$$q_m(\boldsymbol{Z}_m) = \frac{\exp\{\mathbb{E}_{-m}[\log p(\boldsymbol{X}, \boldsymbol{Z})]\}}{\int \exp\{\mathbb{E}_{-m}[\log p(\boldsymbol{X}, \boldsymbol{Z})]\}\, d\boldsymbol{Z}_m} \tag{13}$$

This establishes what is called the **coordinate ascent variational inference** algorithm.

Coordinate Ascent Variational Inference

Algorithm 2: Coordinate ascent variational inference

Input: Model $p(\boldsymbol{X}, \boldsymbol{Z})$, Data $\boldsymbol{X}$,
Output: Variational density $q(\boldsymbol{Z}) = \prod_{m=1}^{M} q_m(\boldsymbol{Z}_m)$
Initialize: Variational densities $q_m(\boldsymbol{Z}_m)$
while *the ELBO has not converged* **do**
  for $m \in \{1, 2, \ldots, M\}$ **do**
    Set $q_m(\boldsymbol{Z}_m) \propto \exp\{\mathbb{E}_{-m}[\log p(\boldsymbol{X}, \boldsymbol{Z})]\}$
  end
  Compute ELBO(q)
end
return $q(\boldsymbol{Z})$

The Orthogonal Factor Model

Here we present a VB framework for the orthogonal time series factor analysis model. The orthogonal FA model assumes the form

$$\boldsymbol{X}_t = \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{e}_t$$

with the following assumptions:

1. $\boldsymbol{F}_t \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_K, \mathbf{I}_\mathbf{K})$
2. $\boldsymbol{e}_t \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \text{diag}\big(\sigma_1^2, \ldots, \sigma_N^2\big)$
3. $\boldsymbol{F}_t$ and $\boldsymbol{e}_s$ are independent for every pair t, s

where $\mathbf{0}_K$ and $\mathbf{0}_N$ are zero-vectors of lengths K and N, respectively, and $\mathbf{I}_K$ is the $K \times K$ identity matrix.

Under the assumptions in the previous slide the likelihood is

$$p(\mathbf{X}_{1:T}|\mathbf{F}_{1:T},\mathbf{\Lambda},\mathbf{\Sigma}) \propto \det(\mathbf{\Sigma})^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(\mathbf{X}_t-\mathbf{\Lambda}\mathbf{F}_t)^\top\mathbf{\Sigma}^{-1}(\mathbf{X}_t-\mathbf{\Lambda}\mathbf{F}_t)\right\}$$

We assign the following prior distributions:

$$p(\mathbf{\Lambda}|\mathbf{\Sigma}) = \prod_{n=1}^{N}\mathcal{N}\left(\boldsymbol{\lambda}_n \,|\, \mathbf{0}_K, \sigma_n^2\mathbf{I}_K\right)$$

$$p(\mathbf{\Sigma}) = \prod_{n=1}^{N}\text{Scale-inv-}\chi^2\left(\sigma_n^2 \,|\, \nu_0, \tau_0^2\right)$$

The goal here is to find a variational approximation to the posterior distribution over using CAVI, i.e. to find

$$q^\star(\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \approx p(\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma} | \boldsymbol{X}_{1:T})$$

such that

$$q^\star(\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) = \underset{q(\boldsymbol{F}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) \in \mathcal{Q}}{\arg\max} \ \mathrm{ELBO}(q)$$

where

$$\begin{aligned}
\mathcal{Q} &= \{q : q(\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) = q(\boldsymbol{F}_{1:T}) q(\boldsymbol{\Lambda}) q(\boldsymbol{\Sigma})\} \\
&= \left\{ q : q(\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) = \prod_{t=1}^{T} q(\boldsymbol{F}_t) \prod_{n=1}^{N} q(\boldsymbol{\lambda}_n) q\big(\sigma_n^2\big) \right\}
\end{aligned}$$

- With the mean-field assumption we have

$$\begin{aligned}\log q(\boldsymbol{F}_{1:T}) &= \mathbb{E}_{-\boldsymbol{F}}[\log p(\boldsymbol{X}_{1:T}, \boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})] + \text{const} \\ \log q(\boldsymbol{\Lambda}) &= \mathbb{E}_{-\boldsymbol{\Lambda}}[\log p(\boldsymbol{X}_{1:T}, \boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})] + \text{const} \\ \log q(\boldsymbol{\Sigma}) &= \mathbb{E}_{-\boldsymbol{\Sigma}}[\log p(\boldsymbol{X}_{1:T}, \boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})] + \text{const}\end{aligned}$$

- In this case the log-joint $\log p(\boldsymbol{X}_{1:T}, \boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})$ expands as

$$\log p(\boldsymbol{X}_{1:T}|\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) + \log p(\boldsymbol{F}_{1:T}) + \log p(\boldsymbol{\Lambda}|\boldsymbol{\Sigma}) + \log p(\boldsymbol{\Sigma})$$

- Therefore,

$$\begin{aligned}\log q(\boldsymbol{F}_{1:T}) &= \mathbb{E}_{-\boldsymbol{F}}[\log p(\boldsymbol{X}_{1:T}|\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) + \log p(\boldsymbol{F}_{1:T})] + \text{const} \\ \log q(\boldsymbol{\Lambda}) &= \mathbb{E}_{-\boldsymbol{\Lambda}}[\log p(\boldsymbol{X}_{1:T}|\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) + \log p(\boldsymbol{\Lambda}|\boldsymbol{\Sigma})] + \text{const} \\ \log q(\boldsymbol{\Sigma}) &= \mathbb{E}_{-\boldsymbol{\Sigma}}[\log p(\boldsymbol{X}_{1:T}|\boldsymbol{F}_{1:T}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}) + \log p(\boldsymbol{\Lambda}|\boldsymbol{\Sigma}) + \log p(\boldsymbol{\Sigma})] + \text{const}\end{aligned}$$

- Substituting in the expressions for the likelihood and prior for $\boldsymbol{F}_{1:T}$ we get

$$\log q(\boldsymbol{F}_{1:T}) = -\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{F}_t^\top \left(\mathbf{I}_K + \mathbb{E}_{-\boldsymbol{F}}\left[\Lambda^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\right]\right)\boldsymbol{F}_t + \sum_{t=1}^{T} \mathbf{X}_t^\top \mathbb{E}_{-\boldsymbol{F}}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}\right]\boldsymbol{F}_t + \text{const}$$

- The equation above is Gaussian wrt $\boldsymbol{F}_t$. Thus,

$$q(\boldsymbol{F}_{1:T}) = \prod_{t=1}^{T} \mathcal{N}(\boldsymbol{F}_t \,|\, \mathbf{m}_{\boldsymbol{F}t}, \mathbf{P}_{\boldsymbol{F}})$$

where

$$\mathbf{P}_{\boldsymbol{F}}^{-1} = \mathbf{I}_K + \sum_{n=1}^{N} \mathbb{E}_{\boldsymbol{\Sigma}}\left[\frac{1}{\sigma_n^2}\right] \mathbb{E}_{\boldsymbol{\Lambda}}\left[\boldsymbol{\lambda}_n \boldsymbol{\lambda}_n^\top\right]$$

$$\mathbf{m}_{\boldsymbol{F}t} = \mathbf{P}_{\boldsymbol{F}}\, \mathbb{E}_{\boldsymbol{\Sigma}}\left[\boldsymbol{\Sigma}^{-1}\right] \mathbb{E}_{\boldsymbol{\Lambda}}[\boldsymbol{\Lambda}] \boldsymbol{X}_t$$

- Proceeding in a similar way, we have that

$$q(\boldsymbol{\Lambda}) = \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{\lambda}_n \,|\, \mathbf{m}_{\boldsymbol{\lambda} n}, \mathbf{P}_{\boldsymbol{\lambda} n})$$

$$q(\boldsymbol{\Sigma}) = \prod_{n=1}^{N} \text{Scale-inv-}\chi^2\big(\sigma_n^2 \,|\, \nu_\sigma, \tau_n^2\big)$$

where

$$\mathbf{P}_{\boldsymbol{\lambda} n}^{-1} = \mathbb{E}_{\boldsymbol{\Sigma}}\left[\frac{1}{\sigma_n^2}\right]\left(T\,\mathbf{I}_K + \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{F}}\left[\boldsymbol{F}_t \boldsymbol{F}_t^\top\right]\right)$$

$$\mathbf{m}_{\boldsymbol{\lambda} n} = \mathbf{P}_{\boldsymbol{\lambda} n} \mathbb{E}_{\boldsymbol{\Sigma}}\left[\frac{1}{\sigma_n^2}\right] \sum_{t=1}^{T} X_{nt} \mathbb{E}_{\boldsymbol{F}}[\boldsymbol{F}_t]$$

$$\nu_\sigma = T + \nu_0$$

$$\nu_\sigma \tau_n^2 = \nu_0 \tau_0^2 + \mathbb{E}_{\boldsymbol{\Lambda}}\left[\boldsymbol{\lambda}_n^\top \boldsymbol{\lambda}_n\right] + \sum_{t=1}^{T}\Big[X_{nt}^2 - 2X_{nt}\mathbb{E}_{\boldsymbol{\Lambda}}[\boldsymbol{\lambda}_n]^\top \mathbb{E}_{\boldsymbol{F}}[\boldsymbol{F}_t] + \text{tr}\big[\mathbb{E}_{\boldsymbol{\Lambda}}\left[\boldsymbol{\lambda}_n \boldsymbol{\lambda}_n^\top\right] \mathbb{E}_{\boldsymbol{F}}\left[\boldsymbol{F}_t \boldsymbol{F}_t^\top\right]\big]\Big]$$

Section 4

1. Background
   - Bayesian Inference
   - Factor Analysis
   - Dynamic Factor Analysis
2. Hierarchical Dynamic Factor Analysis
3. Variational Bayesian Inference
   - Orthogonal Factor Model
4. Future Work
5. References

Future Work

The main goal of the project is to develop a variational Bayesian framework to handle the four-level model from [1].

Section 5

1. Background
   - Bayesian Inference
   - Factor Analysis
   - Dynamic Factor Analysis
2. Hierarchical Dynamic Factor Analysis
3. Variational Bayesian Inference
   - Orthogonal Factor Model
4. Future Work
5. References

References I

Emanuel Moench, Serena Ng, and Simon Potter. "Dynamic Hierarchical Factor Models". en. In: *Review of Economics and Statistics* 95.5 (Dec. 2013), pp. 1811–1817. ISSN: 0034-6535, 1530-9142. DOI: 10.1162/REST_a_00359.