Homework 1 of STAT 5020

WANG Lijun

March 20, 2019

Q1

(a)

The measurement equations are

$$
\begin{aligned}
y_1 &= \lambda_{11}\eta + \varepsilon_1 \\
y_2 &= \lambda_{21}\eta + \varepsilon_2 \\
y_3 &= \lambda_{32}\xi_1 + \varepsilon_3 \\
y_4 &= \lambda_{42}\xi_1 + \varepsilon_4 \\
y_5 &= \lambda_{53}\xi_2 + \varepsilon_5 \\
y_6 &= \lambda_{63}\xi_2 + \varepsilon_6 \\
y_7 &= \lambda_{74}\xi_3 + \varepsilon_7 \\
y_8 &= \lambda_{84}\xi_3 + \varepsilon_8 \\
y_9 &= \lambda_{95}\xi_4 + \varepsilon_9 \\
y_{10} &= \lambda_{10,5}\xi_4 + \varepsilon_{10} \\
y_{11} &= \lambda_{11,5}\xi_4 + \varepsilon_{11} \\
y_{12} &= \lambda_{12,5}\xi_4 + \varepsilon_{12} \\
y_{13} &= \lambda_{13,5}\xi_4 + \varepsilon_{13} \\
y_{14} &= \lambda_{14,6}\xi_5 + \varepsilon_{14} \\
y_{15} &= \lambda_{15,6}\xi_5 + \varepsilon_{15} \\
y_{16} &= \lambda_{16,7}\xi_6 + \varepsilon_{16} \\
y_{17} &= \lambda_{17,7}\xi_6 + \varepsilon_{17} \\
y_{18} &= \lambda_{18,7}\xi_6 + \varepsilon_{18}\,,
\end{aligned}
$$

write out in matrix form,

$$
\begin{aligned}
\boldsymbol{y} &= \begin{bmatrix}
\lambda_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\
\lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda_{32} & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda_{42} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \lambda_{53} & 0 & 0 & 0 & 0 \\
0 & 0 & \lambda_{63} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \lambda_{74} & 0 & 0 & 0 \\
0 & 0 & 0 & \lambda_{84} & 0 & 0 & 0 \\
0 & 0 & 0 & \lambda_{94} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_{10,5} & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_{11,5} & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_{12,5} & 0 & 0 \\
0 & 0 & 0 & 0 & \lambda_{13,5} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_{14,6} & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_{15,6} & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_{16,6} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \lambda_{17,7} \\
0 & 0 & 0 & 0 & 0 & 0 & \lambda_{18,7}
\end{bmatrix}
\begin{bmatrix} \eta \\ \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \end{bmatrix} + \boldsymbol{\varepsilon} \\
&= \Lambda \boldsymbol{\omega} + \boldsymbol{\varepsilon}
\end{aligned}
$$

the structural equation is

$$
\begin{aligned}
\eta &= \sum_{i=1}^{3} b_i d_i + \sum_{i=1}^{6} \gamma_i \xi_i + \delta \\
&= \mathbf{B}\mathbf{d} + \Gamma \boldsymbol{\xi} + \delta\,.
\end{aligned}
$$

(b)

- Non-debt tax shields contribute positively on DEPR and NDTS, and it contributes slightly more in DEPR (1) than in NDTS (0.708).
- Growth contributes positively on PCTA and PCPOR, and it contributes much more in PCTA (1) than in PCPOR (0.180).
- Size contributes positively on LTA and LPOR approximately equally (1 vs. 1.028).
- Profitability contributes positively on five response variables, ROE, ROA, NPMS, SPR and EPS, where it has much more influence on ROA and EPS (4.289 and 3.776), moderate influence on ROE and SPR (1 and 1.216), and less influence on NPMS (3.776).
- Liquidity contributes positively on QR and LR approximately equally (1 vs. 0.995).
- Ownership has contribution on three response variables, positive influence on CIRCS (1), negative influence on CORPS and STATS (-0.874 and -0.229).

- Debt contributes on SDR and TDR approximately equally (1 vs. 1.182).
- In terms of the structure, Liquidity and Profitability contribute negatively on Debt approximately equally, while Size has a slight less positive influence on Debt (0.101).

(c)

Because the fixed value 1 for $\lambda_{1i}, i = 1, \ldots, 7$ introduce a scale to latent variable, and hence identifiable by avoiding various loading matrix Λ.

(d)

d_1, d_2, d_3 is the fixed covariates, whose coefficients $\mathbf{B}$ is unknown. Then we can study the explanatory effects of Tangible assets, Signal, and Operational risk on the key outcome latent variable Debt.

(e)

SEMs can reduce the number of variables in the key regression equation in the conventional regression models by incorporating latent variables.

The conventional regression models often suffer from multicollinearity, but SEMs can group the highly correlated observed variables into latent variables to avoid this problem.

SEMs can assess the interrelationships of latent constructs, but conventional regression model would fail.

Q2

(a)

We often choose conjugate type prior distribution. Firstly, many conjugate prior distributions are good approximations of actual knowledge. Secondly, conjugate priors can let the calculation of Bayesian updating more easy.

Let $\mathbf{\Lambda}_k'$ be the k-th row of $\mathbf{\Lambda}$, and $\psi_{\epsilon k}$ be the k-th diagonal element of $\mathbf{\Psi}_\epsilon$, the conjugate type prior distributions of $\boldsymbol{\mu}$ and $(\mathbf{\Lambda}_k, \psi_{\epsilon k})$ are

$$\begin{aligned} \boldsymbol{\mu} &\sim N(\boldsymbol{\mu}_0, \mathbf{\Sigma}_0) \\ \psi_{\epsilon k} &\sim \text{inverse-Gamma}(\alpha_{0\epsilon k}, \beta_{0\epsilon k}) \\ \mathbf{\Lambda}_k \mid \psi_{\epsilon k} &\sim N(\mathbf{\Lambda}_{0k}, \psi_{\epsilon k}\mathbf{H}_{0yk}) \,. \end{aligned} \tag{1}$$

Rewrite the structural equation as

$$\boldsymbol{\eta}_i = \mathbf{\Lambda}_\omega \mathbf{G}(\boldsymbol{\omega}_i) + \boldsymbol{\delta}_i \,,$$

where $\mathbf{\Lambda}_\omega = (\mathbf{\Pi}, \mathbf{\Lambda})$ and $\mathbf{G}(\boldsymbol{\omega}_i) = (\boldsymbol{\eta}_i^T, \mathbf{F}(\boldsymbol{\xi}_i)^T)^T$. Let $\mathbf{\Lambda}_{wk}'$ be the k-th row of $\mathbf{\Lambda}_k$, and $\psi_{\delta k}$ be the k-th diagonal element of $\mathbf{\Psi}_\delta$. The conjugate type prior distribution of $\mathbf{\Phi}$ and

$(\mathbf{\Lambda}_{\omega k}, \psi_{\delta k})$ are:

$$\begin{aligned}\mathbf{\Phi} &\sim \text{inverse-Wishart}(\mathbf{R}_0^{-1}, \rho_0)\\ \psi_{\delta k} &\sim \text{inverse-Gamma}(\alpha_{0\delta k}, \beta_{0\delta k})\\ \mathbf{\Lambda}_k \mid \psi_{\delta k} &\sim N(\mathbf{\Lambda}_{0wk}, \psi_{\delta k}\mathbf{H}_{0wk})\end{aligned} \tag{2}$$

(b)

Let $\mathbf{\Omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n)$ be matrix of latent vectors, and $\boldsymbol{\theta}$ be the structural parameter vector. Then the conditional distribution $(\mathbf{\Omega} \mid \mathbf{Y}, \boldsymbol{\theta})$ is

$$\begin{aligned}&p(\mathbf{\Omega} \mid \mathbf{Y}, \boldsymbol{\theta})\\ =&\prod_{i=1}^{n} p(\boldsymbol{\omega}_i \mid \boldsymbol{y}_i, \boldsymbol{\theta})\\ \propto&\prod_{i=1}^{n} p(\boldsymbol{y}_i \mid \boldsymbol{\omega}_i, \boldsymbol{\theta}) p(\boldsymbol{\eta}_i \mid \boldsymbol{\xi}, \boldsymbol{\theta}) p(\boldsymbol{\xi}_i \mid \boldsymbol{\theta})\\ \propto&\exp\Big\{ -\frac{1}{2}(\boldsymbol{y}_i - \mathbf{\Lambda}\boldsymbol{\omega}_i)^T \mathbf{\Psi}_\epsilon^{-1}(\boldsymbol{y}_i - \mathbf{\Lambda}\boldsymbol{\omega}_i) - \frac{1}{2}(\boldsymbol{\eta}_i - \mathbf{\Lambda}_w \mathbf{G}(\boldsymbol{\omega}_i))^T \mathbf{\Psi}_\delta^{-1}(\boldsymbol{\eta}_i - \mathbf{\Lambda}_\omega \mathbf{G}(\boldsymbol{\omega}_i)) - \frac{1}{2}\boldsymbol{\xi}_i^T \mathbf{\Phi} \boldsymbol{\xi}_i \Big\},\end{aligned} \tag{3}$$

which is a non-standard and complex distribution.

Now consider the conditional distribution $(\boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{\Omega})$. Let $\boldsymbol{\theta}_y$ be the unknown parameters in the measurement equation, and $\boldsymbol{\theta}_\omega$ be the unknown parameters in the structural equation. We have

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_y) p(\boldsymbol{\theta}_\omega)$$

since the assumption that the prior distribution of $\boldsymbol{\theta}_y$ is independent of the prior distribution of $\boldsymbol{\theta}_\omega$. Moreover,

$$p(\mathbf{Y} \mid \mathbf{\Omega}, \boldsymbol{\theta}) = p(\mathbf{Y} \mid \mathbf{\Omega}, \boldsymbol{\theta}_y) \qquad p(\mathbf{\Omega}) = p(\mathbf{\Omega} \mid \boldsymbol{\theta}_\omega),$$

then

$$\begin{aligned}p(\boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{\Omega}) &\propto p(\mathbf{Y}, \mathbf{\Omega}, \boldsymbol{\theta})\\ &= p(\mathbf{Y} \mid \mathbf{\Omega}, \boldsymbol{\theta}) p(\mathbf{\Omega} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})\\ &= p(\mathbf{Y} \mid \mathbf{\Omega}, \boldsymbol{\theta}_y) p(\boldsymbol{\theta}_y) \cdot p(\mathbf{\Omega} \mid \boldsymbol{\theta}_\omega) p(\boldsymbol{\theta}_\omega),\end{aligned}$$

which implies that the conditional densities $p(\boldsymbol{\theta}_y \mid \mathbf{Y}, \mathbf{\Omega}) \propto p(\mathbf{Y} \mid \mathbf{\Omega}, \boldsymbol{\theta}_y) p(\boldsymbol{\theta}_y)$ and $p(\boldsymbol{\theta}_\omega \mid \mathbf{Y}, \mathbf{\Omega}) \propto p(\mathbf{\Omega} \mid \boldsymbol{\theta}_\omega) p(\boldsymbol{\theta}_\omega)$ can be treated separately.

Consider first the marginal conditional distribution of $\boldsymbol{\theta}_y$. Let $\mathbf{\Lambda}_y = (\boldsymbol{\mu}, \mathbf{\Lambda})$ with general elements $\lambda_{ykj}, j = 1, \ldots, 1+q; k = 1, \ldots, p$, and $\mathbf{u}_i = (1, \boldsymbol{\omega}_i^T)^T$. It follows that $\boldsymbol{y}_i = \mathbf{\Lambda}_y \mathbf{u}_i + \boldsymbol{\epsilon}_i$.

The positions of the fixed elements in $\mathbf{\Lambda}_y$ are identified via an index matrix $\mathbf{L}_y$ with the following elements:

$$l_{ykj} = \begin{cases} 0 & \text{if } \lambda_{ykj} \text{ is fixed} \\ 1 & \text{if } \lambda_{ykj} \text{ is free} \end{cases} \qquad \text{for } j = 1, \ldots, 1+q, \text{ and } k = 1, \ldots, p\,.$$

Let $\mathbf{\Lambda}_{yk}^T$ be the row vector that contains the unknown parameters in the k-th row of $\mathbf{\Lambda}_y$. For convenience, combine the prior of $\boldsymbol{\mu}$ and $\mathbf{\Lambda}_k \mid \psi_{\epsilon k}$ given in (1) as

$$\mathbf{\Lambda}_{yk} \mid \psi_{\varepsilon k} \sim N(\mathbf{\Lambda}_{0yk}, \psi_{\varepsilon k}\mathbf{H}_{0yk})\,.$$

Let $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ and $\mathbf{u}_k$ be the submatrix of $\mathbf{U}$ such that all the rows corresponding to $l_{ykj} = 0$ are deleted; and let $\mathbf{Y}_k^{*T} = (y_{1k}^*, \ldots, y_{nk}^*)$ with

$$y_{ik}^* = y_{ik} - \sum_{j=1}^{1+q} \lambda_{ykj} u_{ij} (1 - l_{ykj})\,.$$

We can get

$$\psi_{\epsilon k}^{-1} \mid \mathbf{Y}, \mathbf{\Omega} \sim \text{Gamma}(n/2 + \alpha_{0\epsilon k}, \beta_{\epsilon k}) \quad \mathbf{\Lambda}_{yk} \mid \mathbf{Y}, \mathbf{\Omega}, \psi_{\epsilon k}^{-1} \sim N(\mathbf{a}_{yk}, \psi_{\epsilon k}\mathbf{A}_{yk})\,. \tag{4}$$

Now, consider the conditional distribution of $\boldsymbol{\theta}_\omega$. Let $\mathbf{\Omega}_1 = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)$ and $\mathbf{\Omega}_2 = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)$. Then

$$\begin{aligned} p(\boldsymbol{\theta}_\omega \mid \mathbf{Y}, \mathbf{\Omega}) &\propto p(\mathbf{\Omega} \mid \boldsymbol{\theta}_\omega) p(\boldsymbol{\theta}_\omega) \\ &= p(\mathbf{\Omega}_1 \mid \mathbf{\Omega}_2, \boldsymbol{\theta}_\omega) p(\mathbf{\Omega}_2 \mid \boldsymbol{\theta}_\omega) \\ &= p(\mathbf{\Omega}_1 \mid \mathbf{\Omega}_2, \mathbf{\Pi}, \mathbf{\Gamma}, \mathbf{\Psi}_\delta) p(\mathbf{\Pi}, \mathbf{\Lambda}, \mathbf{\Psi}_\delta) p(\mathbf{\Omega}_2 \mid \mathbf{\Phi}) p(\mathbf{\Phi})\,, \end{aligned}$$

which implies that we can treat the marginal conditional densities of $(\mathbf{\Pi}, \mathbf{\Gamma}, \mathbf{\Psi}_\delta)$ and $\mathbf{\Phi}$ can be treated separately.

Note that

$$\begin{aligned} p(\mathbf{\Phi} \mid \mathbf{\Omega}_2) &\propto p(\mathbf{\Phi}) \prod_{i=1}^n p(\boldsymbol{\xi}_i \mid \boldsymbol{\theta}) \\ &\propto \Big[|\mathbf{\Phi}|^{-(\rho_0+q_2+1)/2} \exp\big\{-\frac{1}{2}\text{tr}[\mathbf{R}_0^{-1}\mathbf{\Phi}^{-1}]\big\}\Big] \Big[|\mathbf{\Phi}|^{-n/2} \exp\big\{-\frac{1}{2}\sum_{i=1}^n \boldsymbol{\xi}_i^T \mathbf{\Phi}^{-1} \boldsymbol{\xi}_i\big\}\Big] \\ &= |\mathbf{\Phi}|^{-(n+\rho_0+q_2+1)/2} \exp\big\{-\frac{1}{2}\text{tr}[\mathbf{\Phi}^{-1}(\mathbf{\Omega}_2\mathbf{\Omega}_2^T + \mathbf{R}_0^{-1})]\big\}\,, \end{aligned}$$

which implies that

$$\mathbf{\Phi} \mid \mathbf{\Omega}_2 \sim \text{inverse-Wishart}(\mathbf{\Omega}_2\mathbf{\Omega}_2^T + \mathbf{R}_0^{-1}, n + \rho_0)\,.$$

Note that $\boldsymbol{\eta}_i = \mathbf{\Lambda}_\omega \mathbf{v}_i + \boldsymbol{\delta}_i$, where $\mathbf{\Lambda} = (\mathbf{\Pi}, \mathbf{\Gamma})$ with general elements $\lambda_{\omega kj}$ for $k = 1, \ldots, q_1$, and $\mathbf{v}_i = \mathbf{G}(\omega_i) = (\boldsymbol{\eta}_i^T, \mathbf{F}(\boldsymbol{\xi}_i)^T)^T$ be an $(q_1+t)\times 1$ vector. The model $\boldsymbol{\eta}_i = \mathbf{\Lambda}_\omega \mathbf{v}_i + \boldsymbol{\delta}_i$ is similar to $\boldsymbol{y}_i = \mathbf{\Lambda}_y \mathbf{u}_i + \boldsymbol{\epsilon}_i$ considered before. Let $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$, $\mathbf{L}_\omega$ be the index matrix with general elements $l_{\omega kj}$ that similarly defined as $\mathbf{L}_y$ to indicate the fixed known parameters in $\mathbf{\Lambda}_\omega$. Let $\mathbf{V}_k$ be the submatrix of $\mathbf{V}$ such that all the rows corresponding to $l_{\omega kj} = 0$ are deleted; and let $\mathbf{\Xi}_k^T = (\eta_{1k}^*, \ldots, \eta_{nk}^*)$ where

$$\eta_{ik}^* = \eta_{ik} - \sum_{j=1}^{q_2+t} \lambda_{\omega kj} v_{ij} (1 - l_{\omega kj}) \, .$$

Then we can get

$$\psi_{\delta k}^{-1} \mid \mathbf{\Omega} \sim \text{Gamma}(n/2 + \alpha_{0\delta k}, \beta_{\delta k}) \qquad \mathbf{\Lambda}_{\omega k} \mid \mathbf{\Omega}, \psi_{\delta k}^{-1} \sim N(\mathbf{a}_{\omega k}, \psi_{\delta k} \mathbf{A}_{\omega k}) \, , \tag{5}$$

where $\mathbf{A}_{\omega k} = (\mathbf{H}_{0\omega k} + \mathbf{V}_k \mathbf{V}_k^T)^{-1}$, $\mathbf{a}_{\omega k} = \mathbf{A}_{\omega k}(\mathbf{H}_{0\omega k}^{-1} \mathbf{\Lambda}_{0\omega k} + \mathbf{V}_k \mathbf{\Xi}_k)$, and

$$\beta_{\delta k} = \beta_{0\delta k} + \frac{1}{2}(\mathbf{\Xi}_k^T \mathbf{\Xi}_k - \mathbf{a}_{\omega k}^T \mathbf{A}_{\omega k}^{-1} \mathbf{a}_{\omega k} + \mathbf{\Lambda}_{0\omega k}^T \mathbf{H}_{0\omega k}^{-1} \mathbf{\Lambda}_{0\omega k}) \, .$$

In summary, the full conditional distributions are given by (3), (4) and (5).