## General Strategy (Section 9.6)

(Step 1) Principal Component Analysis

- Look for "suspicious" observations
- examine eigenvalues to get an idea about m, the number of common factors
  - SCREE plot
  - gaps
- identify correlation patterns
- try a varimax rotation



(Step 2) Maximum likelihood factor analysis

- Repeat for several values of m
  Base final choice of m on
  - Communalities, proportion of variation explained
  - Other measures or tests, MSA, Reliability coefficients, AIC, etc.
  - Reasonable interpretations
  - Subject matter knowledge/ instrument construction
- Repeat for several choices of prior communalities

- Try some rotations
  - VARIMAX
  - PROMAX
  - QUARTIMAX

(step 3) Verification of stability based on data splitting

- Randomly split data into two halves
- Crossvalidation



## Structural Equations Models

- Construct linear models for cause-and-effect relationships
- Latent variables:
  cannot be directly observed
- Exogenous variables:
  not influenced by other variables is the model
- Endogenous variables:
  are affected by other variables in the model

## LISREL Model:

(Jöreskog + Sörbom, 1970's)

Observed data: $\begin{bmatrix} \underset{\sim}{Y}_{p\times 1} \\ \hdashline \underset{\sim}{X}_{q\times 1} \end{bmatrix}$

$$\underset{\sim}{Y}_{p\times 1} = \Lambda_y \underset{\sim}{\eta}_{m\times 1} + \underset{\sim}{\epsilon}_{p\times 1}$$

$\underset{\sim}{\eta}$: latent endogenous variables

$\underset{\sim}{\epsilon}$ is uncorrelated with $\underset{\sim}{\eta}$

$$\underset{\sim}{X}_{q\times 1} = \Lambda_x \underset{\sim}{\xi}_{n\times 1} + \underset{\sim}{\delta}_{q\times 1}$$

$\underset{\sim}{\xi}$: latent exogeneous variables

$\underset{\sim}{\delta}$ is uncorrelated with $\underset{\sim}{\xi}$



$$\underset{\sim}{\eta}_{m\times 1} = B \underset{\sim}{\eta}_{m\times 1} + \Gamma \underset{\sim}{\xi}_{n\times 1} + \underset{\sim}{\zeta}_{m\times 1}$$

$\underset{\sim}{\zeta}$ is uncorrelated with $\underset{\sim}{\xi}_{n\times 1}$

$B$ has zeros on the diagonal (no endogeneous variable can be used to predict itself)

$I - B$ is nonsingular

$E(\underset{\sim}{\zeta}) = \underset{\sim}{0}$ $\quad$ $Cov(\underset{\sim}{\zeta}) = \Psi$

$E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$ $\quad$ $Cov(\underset{\sim}{\epsilon}) = \Theta_{\underset{\sim}{\epsilon}}$

$E(\underset{\sim}{\delta}) = \underset{\sim}{0}$ $\quad$ $Cov(\underset{\sim}{\delta}) = \Theta_{\underset{\sim}{\delta}}$

$\underset{\sim}{\zeta}, \underset{\sim}{\epsilon}, \underset{\sim}{\delta}$ are uncorrelated

$E(\underset{\sim}{\xi}) = \underset{\sim}{0} \qquad \mathrm{cov}(\underset{\sim}{\xi}) = \Phi$

$E(\underset{\sim}{\eta}) = \underset{\sim}{0}$

Note that

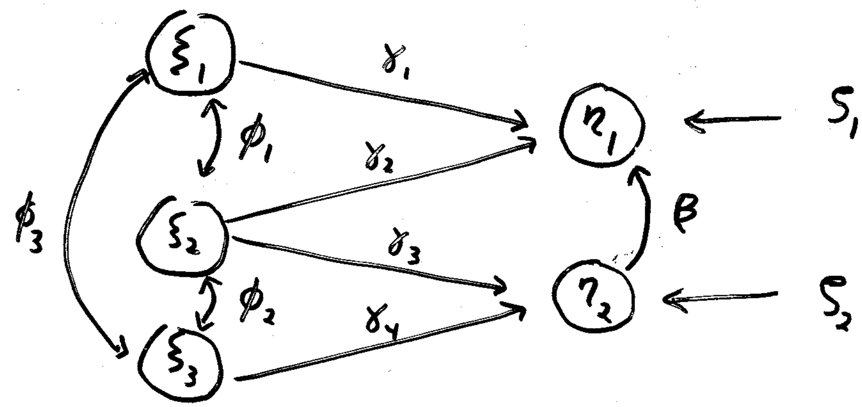
$$(I-B)\underset{\sim}{\eta} = \Gamma \underset{\sim}{\xi} + \underset{\sim}{\zeta}$$

$$\Rightarrow \quad \underset{\sim}{\eta} = (I-B)^{-1}(\Gamma \underset{\sim}{\xi} + \underset{\sim}{\zeta})$$

$$\Rightarrow \quad \mathrm{cov}(\underset{\sim}{\eta}) = (I-B)^{-1}\left[\Gamma \Phi \Gamma' + \Psi\right]\left[(I-B)^{-1}\right]'$$



## Path Diagram




$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_1 & \gamma_2 & 0 \\ 0 & \gamma_3 & \gamma_4 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

$\mathrm{cov}(\xi_1, \xi_2) = \phi_1 \qquad \mathrm{cov}(\xi_1, \xi_3) = \phi_3 \qquad \mathrm{cov}(\xi_2, \xi_3) = \phi_2$

## Path diagram

(1) A straight arrow is drawn to each dependent variable from each of its sources

(2) A straight arrow is drawn to each dependent variable from its residual

(3) A curved, double-headed arrow indicates correlation between two exogenous variables

(4) Single-headed curved arrows indicate directional relationships



## Comments

- Since only $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are observed, the LISREL model cannot be verified directly.
- Fitting a particular LISREL model does <u>not</u> validate cause-and-effect relationships imposed by the model.

Covariance structure imposed on $\begin{bmatrix} \underset{\sim}{Y} \\ \underset{\sim}{X} \end{bmatrix}$

$$Cov\begin{bmatrix} \underset{\sim}{Y} \\ \underset{\sim}{X} \end{bmatrix} = \Sigma = \left[\begin{array}{c:c} \Sigma_{11} & \Sigma_{12} \\ \hdashline \Sigma_{21} & \Sigma_{22} \end{array}\right]$$

where

$$\Sigma_{11} = \Lambda_y \, Cov(\underset{\sim}{\eta}) \Lambda_y' + \Theta_{\underset{\sim}{\epsilon}}$$

$$= \Lambda_y \left[ (I-B)^{-1} \left[ \Gamma \Phi \Gamma' + \Psi \right] \left( (I-B)^{-1} \right)' \right] \Lambda_y' + \Theta_{\underset{\sim}{\epsilon}}$$

$$\Sigma_{22} = \Lambda_x \, Cov(\underset{\sim}{\xi}) \Lambda_x' + \Theta_{\underset{\sim}{\delta}}$$

$$= \Lambda_x \Phi \Lambda_x' + \Theta_{\underset{\sim}{\delta}}$$

$$\Sigma_{12} = \Sigma_{21}' = Cov\left( \Lambda_y (I-B)^{-1} (\Gamma \underset{\sim}{\xi} + \underset{\sim}{\zeta}), \Lambda_x \underset{\sim}{\xi} + \underset{\sim}{\delta} \right)$$

$$= \Lambda_y (I-B)^{-1} \Gamma \Phi \Lambda_x'$$



Data

$$\begin{bmatrix} \underset{\sim}{Y}_1 \\ \underset{\sim}{X}_1 \end{bmatrix}, \ldots, \begin{bmatrix} \underset{\sim}{Y}_N \\ \underset{\sim}{X}_N \end{bmatrix}$$

Compute

$$S = \frac{1}{N-1} \sum_{j=1}^{N} \left[ \begin{pmatrix} \underset{\sim}{Y}_j \\ \underset{\sim}{X}_j \end{pmatrix} - \begin{pmatrix} \bar{\underset{\sim}{Y}} \\ \bar{\underset{\sim}{X}} \end{pmatrix} \right] \left[ \begin{pmatrix} \underset{\sim}{Y}_j \\ \underset{\sim}{X}_j \end{pmatrix} - \begin{pmatrix} \bar{\underset{\sim}{Y}} \\ \bar{\underset{\sim}{X}} \end{pmatrix} \right]'$$

$$= \left[\begin{array}{c:c} S_{11} & S_{12} \\ \hdashline S_{21} & S_{22} \end{array}\right]$$

Estimate $\Sigma$ by finding the "closest" match $\hat{\Sigma}$ to $S$.

- Least squares
- Maximum likelihood estimation

## Software:

SAS: CALIS

SPSS: LISREL

## Comments

- There are limitations to the observed data.
  - observational studies
  - latent variables
- The observed data may be consistent with many models or theories



## Artificial Example:

$$m = n = 1$$
$$p = 2$$
$$q = 2$$
$$B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\eta = \gamma \xi + \zeta$$

$$\begin{array}{ll} \text{profit} & \\ \text{stock price} & \end{array} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} \underset{\substack{\uparrow \\ \text{company performance}}}{\eta} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$\begin{array}{l} \text{CEO Years of experience} \\ \text{boards of directors} \end{array} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} \underset{\substack{\uparrow \\ \text{managerial talent}}}{\xi} + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$$

$$\xi \xrightarrow{\gamma} \eta \longleftarrow \zeta$$

managerial talent — performance — residual

Let

$$\mathrm{Var}(\xi) = \phi$$
$$\mathrm{Var}(\zeta) = \psi \quad \Rightarrow \mathrm{Var}(\eta) = \gamma^2\phi + \psi$$

$$\mathrm{Cov}\begin{pmatrix}\epsilon_1\\ \epsilon_2\end{pmatrix} = \begin{bmatrix}\theta_1 & 0\\ 0 & \theta_2\end{bmatrix}$$

$$\mathrm{Cov}\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix} = \begin{bmatrix}\theta_3 & 0\\ 0 & \theta_4\end{bmatrix}$$



Then

$$\mathrm{Cov}\begin{pmatrix}y_1\\ y_2\end{pmatrix} = \begin{bmatrix}\gamma^2\phi + \psi + \theta_1 & \lambda_1(\gamma^2\phi+\psi)\\ \lambda_1(\gamma^2\phi+\psi) & \lambda_1^2(\gamma^2\phi+\psi)+\theta_2\end{bmatrix}$$
$$= \Sigma_{11}$$

$$\mathrm{Cov}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \Sigma_{22} = \begin{bmatrix}\phi\lambda_2^2 + \theta_3 & \phi\lambda_2\\ \phi\lambda_2 & \phi + \theta_4\end{bmatrix}$$

$$\mathrm{Cov}(\underset{\sim}{y}, \underset{\sim}{x}) = \Sigma_{12} = \begin{bmatrix}\lambda_2\gamma\phi & \gamma\phi\\ \lambda_1\lambda_2\gamma\phi & \lambda_1\gamma\phi\end{bmatrix}$$

<u>Data provides</u>

$$\left[\begin{array}{c:c} S_{11} & S_{12}\\ \hdashline S_{21} & S_{22}\end{array}\right] = \left[\begin{array}{cc:cc} 14.3 & -27.6 & 6.4 & 3.2\\ -27.6 & 55.4 & -12.8 & -6.4\\ \hdashline 6.4 & -12.8 & 3.7 & 1.6\\ 3.2 & -6.4 & 1.6 & 1.1\end{array}\right]$$

Here you can get an "exact" match of $\hat{\Sigma}$ with $S$

$$\hat{\gamma} = 4 \qquad \hat{\lambda}_y = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \qquad \hat{\lambda}_x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\hat{\phi} = .8 \qquad \hat{\psi} = 1$$

$$\hat{\Theta}_\epsilon = \begin{bmatrix} .5 & 0 \\ 0 & .2 \end{bmatrix} \qquad \hat{\Theta}_\delta = \begin{bmatrix} .5 & 0 \\ 0 & .3 \end{bmatrix}$$

The model implies

boards of directors units

$$\text{company performance} = 4 \begin{pmatrix} \text{managerial} \\ \text{talent} \end{pmatrix} + \text{error } (0,1)$$

profit units

"cause-and-effect" relationship



**Comments:**

(1)
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} \eta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} \xi + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$$

The one's fix the scales for $\eta$ and $\xi$

(2) Need more equations than unknowns

$$\begin{pmatrix} \text{number of parameters} \\ \text{in } \Sigma \end{pmatrix} \leq \frac{(p+q)(p+q+1)}{2}$$

This does not guarantee unique estimates.

(3) Look for reasonable estimates
- negative variances?
- correlations with wrong signs?