So You Want to Specify an Inverse-Wishart Prior Distribution

N. K. Schuurman

Mplus User Meeting 2016, Utrecht University

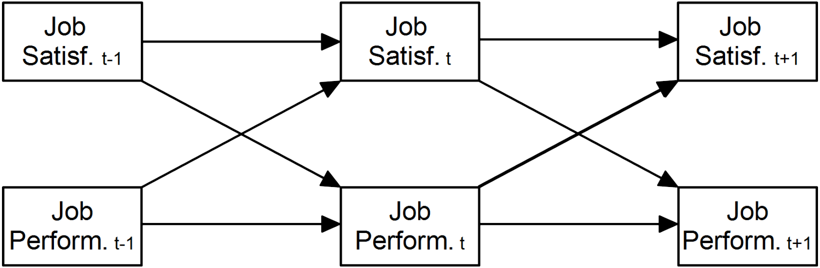
Januari 13 2016

Introduction

- Inverse-Wishart prior distribution for covariance matrices.
- Specification of uninformative prior can be difficult when variances may be small (see also Gelman 2006 on Inverse-Gamma distributions).
- Especially an issue for multilevel (autoregressive time series) models.

Introduction

- How do psychological variables affect each other over time?



Introduction

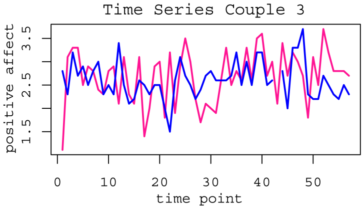
Cross-lagged Panel Models

- SEM
- few repeated measures, many persons
- ignores differences between persons

person	year	income	age	sex
1	2003	1500	27	1
1	2004	1700	28	1
1	2005	2000	29	1
2	2003	2100	41	2
2	2004	2100	42	2
2	2005	2200	43	2

Time Series Models

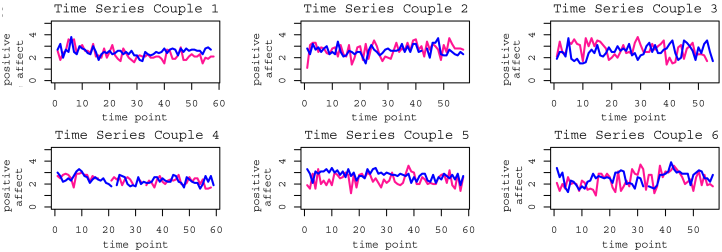
- many repeated measures, one person
- difficult to generalize



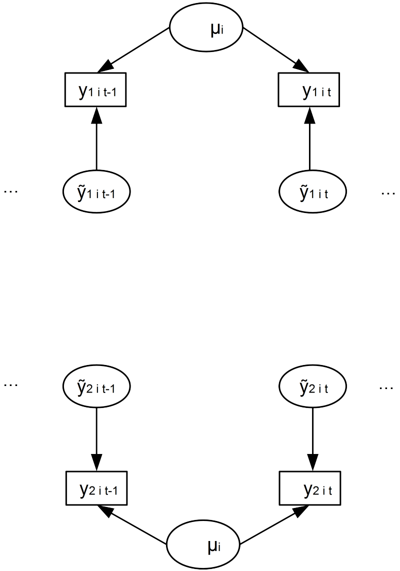
Introduction

Multilevel Autoregressive Models

- many repeated measures, many persons
- fit autoregressive model for all persons at once
- model parameters are allowed to vary over persons
- In the next version of Mplus!

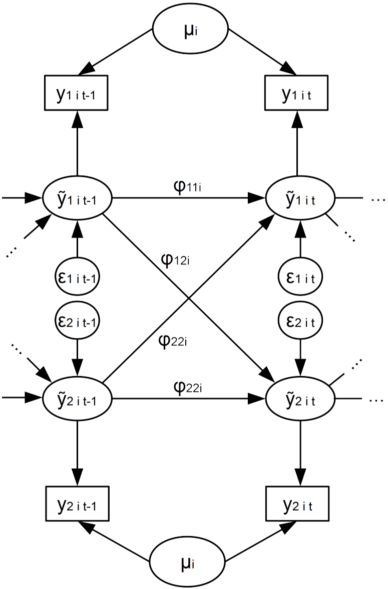


Bivariate multilevel autoregressive model



$$y_{it} = \mu_i + \tilde{y}_{it}$$

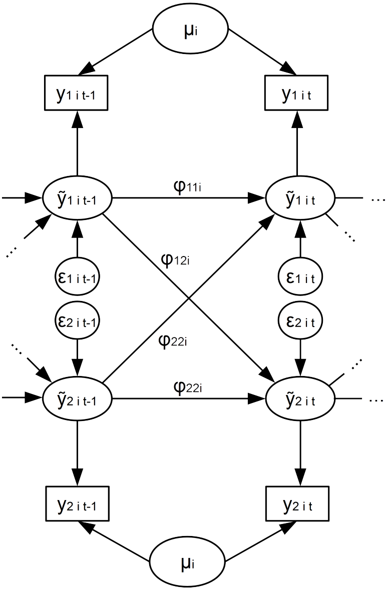
Bivariate multilevel autoregressive model



$$y_{it} = \mu_i + \tilde{y}_{it}$$
$$\tilde{y}_{it} = \Phi_i \tilde{y}_{it-1} + \epsilon_{it}$$

$$\epsilon_{it} \sim MvN\,(0, \Sigma)$$

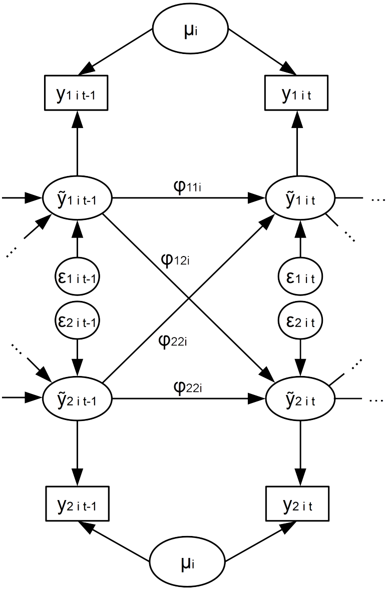
Bivariate multilevel autoregressive model



$$y_{it} = \mu_i + \tilde{y}_{it}$$
$$\tilde{y}_{it} = \Phi_i \tilde{y}_{it-1} + \epsilon_{it}$$

$$\epsilon_{it} \sim MvN\left(0, \Sigma\right)$$
$$\mu_i, \Phi_i \sim MvN\left(\gamma, \Psi\right)$$

Bivariate multilevel autoregressive model

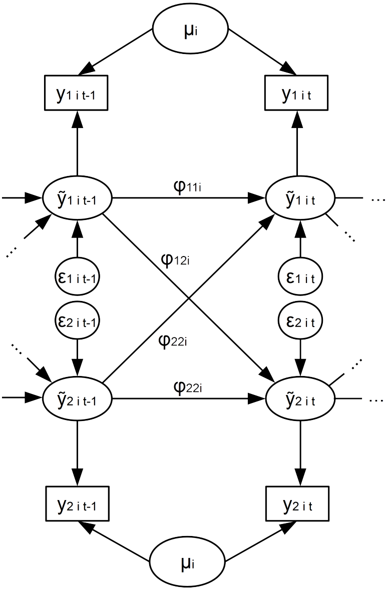


$$y_{it} = \mu_i + \tilde{y}_{it}$$
$$\tilde{y}_{it} = \Phi_i \tilde{y}_{it-1} + \epsilon_{it}$$

$$\epsilon_{it} \sim MvN\,(0, \Sigma)$$
$$\mu_i, \Phi_i \sim MvN\,(\gamma, \Psi)$$

Inverse-Wishart prior for Ψ.

Bivariate multilevel autoregressive model



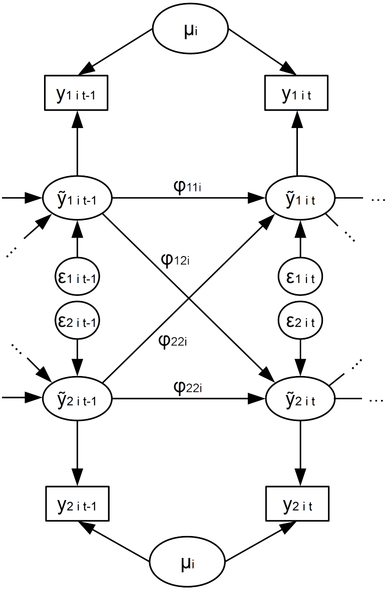
$$y_{it} = \mu_i + \tilde{y}_{it}$$
$$\tilde{y}_{it} = \Phi_i \tilde{y}_{it-1} + \epsilon_{it}$$

$$\epsilon_{it} \sim MvN\left(0, \Sigma\right)$$
$$\mu_i, \Phi_i \sim MvN\left(\gamma, \Psi\right)$$

Inverse-Wishart prior for Ψ.

The regression parameters are restricted in range for each person.

Bivariate multilevel autoregressive model



$$y_{it} = \mu_i + \tilde{y}_{it}$$
$$\tilde{y}_{it} = \Phi_i \tilde{y}_{it-1} + \epsilon_{it}$$

$$\epsilon_{it} \sim MvN\,(0, \Sigma)$$
$$\mu_i, \Phi_i \sim MvN\,(\gamma, \Psi)$$

Inverse-Wishart prior for Ψ.

The regression parameters are restricted in range for each person.

Variances for the regression parameters in Ψ will be small (e.g., .005 to .05).

Why care about (not miss-specifying) priors for the variances?

- The variances give use an impression of the range of parameters in the population.

Why care about (not miss-specifying) priors for the variances?

- The variances give use an impression of the range of parameters in the population.
- Bias in the variances will result in biases in the individual parameters.

Why care about (not miss-specifying) priors for the variances?

- The variances give use an impression of the range of parameters in the population.
- Bias in the variances will result in biases in the individual parameters.
- Severe bias in the variances will mess up estimates of the fixed effects.

Inverse-Wishart Prior Distribution

- Conjugate prior for covariance matrices of normal distributed variables

Inverse-Wishart Prior Distribution

- Conjugate prior for covariance matrices of normal distributed variables
- Multivariate extension of Inverse-Gamma distribution

Inverse-Wishart Prior Distribution

- Conjugate prior for covariance matrices of normal distributed variables
- Multivariate extension of Inverse-Gamma distribution
- Specified with a Scale matrix **S**, and degrees of freedom **df**

Inverse-Wishart Prior Distribution

- Conjugate prior for covariance matrices of normal distributed variables
- Multivariate extension of Inverse-Gamma distribution
- Specified with a Scale matrix $\mathbf{S}$, and degrees of freedom $\mathbf{df}$
- Ensures positive definite covariance matrix

Inverse-Wishart Prior Distribution

Scale and degrees of freedom

- **S** is used to position the IW distribution in parameter space
- **df** is used to set the certainty about the prior information in the scale matrix; *df* $>$r-1

Inverse-Wishart Prior Distribution

Actually Not That Simple

IW mean:

$$\frac{\mathbf{S}}{df - r - 1} \tag{1}$$

IW variances:

$$\frac{2s_{kk}^2}{(df - r - 1)^2(df - r - 3)}. \tag{2}$$

Inverse-Wishart becomes more informative when:

- degrees of freedom increase

Inverse-Wishart Prior Distribution

Actually Not That Simple

IW mean:

$$\frac{\mathbf{S}}{df - r - 1} \tag{1}$$

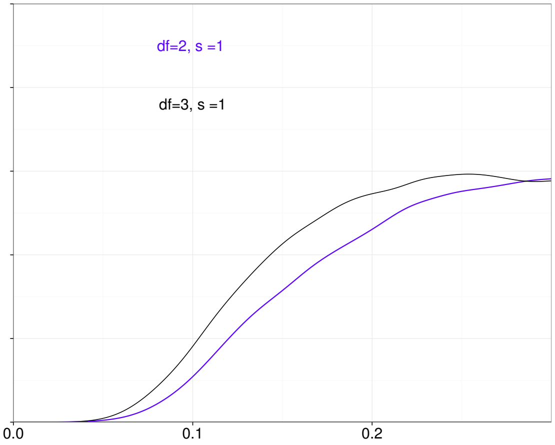
IW variances:

$$\frac{2s_{kk}^2}{(df - r - 1)^2(df - r - 3)}. \tag{2}$$

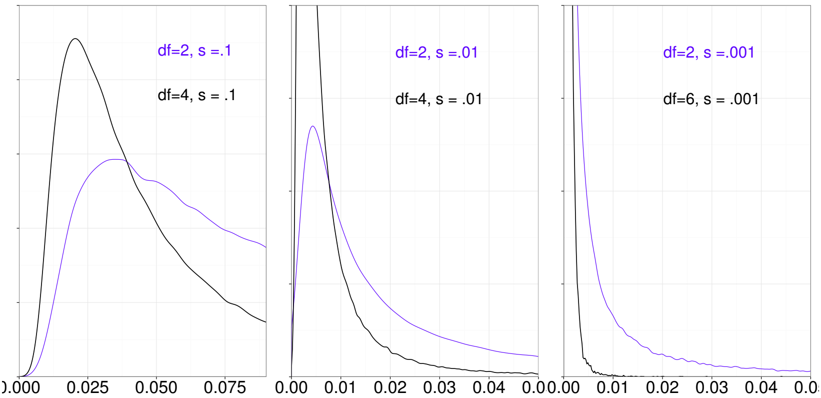
Inverse-Wishart becomes more informative when:

- degrees of freedom increase
- values in the scale matrix become smaller

Difficult to balance **S** and *df* when variances are small



Difficult to balance **S** and *df* when variances are small



Options that work relatively well

- Avoid specifying (Inverse) Gamma or Wishart distributions - use uniform instead (Mplus-friendly).
  * For univariate and bivariate covariance matrices.

Options that work relatively well

- Avoid specifying (Inverse) Gamma or Wishart distributions - use uniform instead (Mplus-friendly).
  * For univariate and bivariate covariance matrices.
- Use a data-based prior (Mplus-friendly).
  * Little bias, but we use the data twice: too small credible intervals. (cf. Schuurman et al., in press)

Options that work relatively well

- Avoid specifying (Inverse) Gamma or Wishart distributions - use uniform instead (Mplus-friendly).
  * For univariate and bivariate covariance matrices.
- Use a data-based prior (Mplus-friendly).
  * Little bias, but we use the data twice: too small credible intervals. (cf. Schuurman et al., in press)
- Use training data (Mplus-friendly).
  * Requires a certain amount of data.
- Use an informative prior based on previous studies. (Mplus-friendly)
  * May be difficult to obtain appropriate data.

Other options to try that may work

- Use improper priors. Default in Mplus (IW with negative df, Scale $= 0$).
  * Prior difficult to interpret. Still ensures positive definite matrix?

Other options to try that may work

- Use improper priors. Default in Mplus (IW with negative df, Scale $= 0$).
  * Prior difficult to interpret. Still ensures positive definite matrix?
- Transform the covariance matrix, put prior on transformed matrix. (Mplus friendly..?)
  * Convergence can be wonky, inconsistent results.

Other options to try that may work

- Use improper priors. Default in Mplus (IW with negative df, Scale = 0).
  * Prior difficult to interpret. Still ensures positive definite matrix?
- Transform the covariance matrix, put prior on transformed matrix. (Mplus friendly..?)
  * Convergence can be wonky, inconsistent results.
- Put Gamma priors on the diagonal elements in the IW-scale matrix. (Mplus friendly..?; cf. Huang & Wand, 2013)

Other options to try that may work

- Use improper priors. Default in Mplus (IW with negative df, Scale = 0).
  * Prior difficult to interpret. Still ensures positive definite matrix?
- Transform the covariance matrix, put prior on transformed matrix. (Mplus friendly..?)
  * Convergence can be wonky, inconsistent results.
- Put Gamma priors on the diagonal elements in the IW-scale matrix. (Mplus friendly..?; cf. Huang & Wand, 2013)
- Decompose the covariance matrix, specify priors on its parts.. (Mplus friendly..?; cf. Barnard, McCulloch & Meng)

So you want to specify an IW-prior...

So you want to specify an IW-prior...

- Collect lots of data.

So you want to specify an IW-prior...

- Collect lots of data.
- Do not automatically trust defaults.

So you want to specify an IW-prior...

- Collect lots of data.
- Do not automatically trust defaults.
- Try a couple of different priors and compare the results. (do a sensitivity analysis)
- Priors that are convenient to include for your sensitivity analysis: uniform priors on the variances. A data-based prior.

References

- Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica, 10(4), 1281-1312.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis, 1(3)*, 515-534.
- Huang, A., & Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. Bayesian Analysis, 8(2), 439-452.
- Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E.L. (in press). A Comparison of Inverse-Wishart Prior Specifications for Covariance Matrices in Multilevel Autoregressive Models. *Multivariate Behavioral Research*.