Bayesian Hypothesis Testing in Social Science Research

Joris Mulder

Department of Methodology & Statistics
Tilburg University, the Netherlands

CRiSM Workshop: Contemporary Issues in Hypothesis Testing, University of Warwick

September 16, 2016

Outline

1. Testing constrained hypotheses on measures of association
   - Bayes factors
   - Prior specification
   - Numerical example
   - BCT Software
   - Empirical application
   - Summary
2. Information inconsistency when testing regression parameters
   - Testing precise hypotheses
   - Testing one-sided hypotheses
   - Summary

Outline

1. Testing constrained hypotheses on measures of association
   - Bayes factors
   - Prior specification
   - Numerical example
   - BCT Software
   - Empirical application
   - Summary
2. Information inconsistency when testing regression parameters
   - Testing precise hypotheses
   - Testing one-sided hypotheses
   - Summary

Testing constrained hypotheses on measures of association

- This work is based on Mulder (2016) (published in a special issue "Bayes Factors for Testing Hypotheses in Psychological Research: Practical Relevance and New Developments" in the *Journal of Mathematical Psychology*, guest edited by me and Eric-Jan Wagenmakers.
- and Mulder and Gelissen (in preparation).

Testing constrained hypotheses on measures of association

Application in social research

$$\mathbf{C} = \begin{array}{r} \\ Happy \\ Health \\ Education \\ Sex\ Perm. \end{array} \begin{array}{c} \begin{array}{cccc} Happy & Health & Educ. & S.Perm. \end{array} \\ \begin{bmatrix} 1 & & & \\ \rho_{21} & 1 & & \\ \rho_{31} & \rho_{32} & 1 & \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{bmatrix} \end{array}$$

Problem. Different hypotheses about the best single predictor of someone's *Happiness* out of *Health*, *Educational Level*, and *Personal-Sexual Permissiveness* when controlling for *Gender* and *Age*.

$$\begin{aligned} H_1 &: \rho_{21} = \rho_{31} = \rho_{41} \\ H_2 &: \rho_{21} > \rho_{31} > \rho_{41} > 0 \\ H_3 &: \rho_{21} > 0,\ \rho_{31} = \rho_{41} = 0 \\ H_4 &: \rho_{21} = \rho_{31} = \rho_{41} = 0 \\ H_5 &: \text{not } H_1, H_2, H_3, H_4. \end{aligned}$$

Generalized Multivariate Probit Model

The model for combinations of continuous and ordinal dependent variables

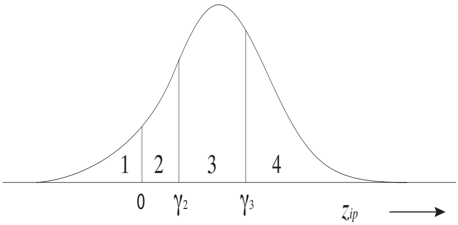
Let the first p_1 DVs be continuous, and the last p_2 DVs be ordinal, $P = p_1 + p_2$ (e.g., Boscardin et al., 2008). A multivariate probit model is used for the ordinal DVs using multivariate normal latent variables $z_{i,1}, \ldots, z_{i,p_2}$ with standard deviations of 1 to ensure identification, i.e.,

$$
\begin{aligned}
(y_{i,1}, \ldots, y_{i,p_1}, z_{i,1}, \ldots, z_{i,p_2})' &\sim N(\mathbf{B}\mathbf{x}_i, \boldsymbol{\Sigma}) \\
\text{with } \boldsymbol{\Sigma} &= \mathbf{DCD} \\
\mathbf{C} &= \begin{bmatrix} 1 & & & \\ \rho_{21} & 1 & & \\ \vdots & \ddots & \ddots & \\ \rho_{P1} & \ldots & \rho_{P(P-1)} & 1 \end{bmatrix} \\
\mathbf{D} &= \text{diag}(\sigma_1, \ldots, \sigma_{p_1}, 1, \ldots, 1)
\end{aligned}
$$

Generalized Multivariate Probit Model

Modeling ordinal variables

Cut-points γ determine the link between each latent z and the respective ordinal dependent variable. E.g., for an ordinal variable with 4 categories this implies the following:



Formulation of the testing problem

Testing problem

- Test hypotheses $H_1, \ldots, H_T$ of the form

$$H_t : \mathbf{R}_t^I \boldsymbol{\rho} > \mathbf{r}_t^I \ \& \ \mathbf{R}_t^E \boldsymbol{\rho} = \mathbf{r}_t^E,$$

such that constraints are of the form $\rho_{gh} \gtreqless \rho_{g'h'}$ or $\rho_{gh} \gtreqless r_{gh}$.

- The free parameters under H_t will be denoted by $\boldsymbol{\rho}_t$ with allowed subspace $\mathcal{C}_t$ resulting in a positive definite correlation matrices.

Challenges

- Prior specification of $\boldsymbol{\rho}$ under all constrained hypotheses while maintaining positive definiteness of the correlation matrix.
- Bayesian computation of marginal likelihoods.

Related topics

Pattern hypotheses

Considerable amount of literature on testing pattern hypotheses, i.e., $H_0 : \mathbf{R}_0\boldsymbol{\rho} = \mathbf{r}_0$ against the unconstrained alternative (e.g., Steiger, 1980).

Comparing correlational structures

$$\mathbf{R}^{AR(1)} = \begin{bmatrix} 1 & & \\ \rho & 1 & \\ \rho^2 & \rho & 1 \end{bmatrix}, \qquad \mathbf{R}^{Toeplitz} = \begin{bmatrix} 1 & & \\ \rho_1 & 1 & \\ \rho_2 & \rho_1 & 1 \end{bmatrix}$$

Classical methods

- The number of free parameters is undefined under an inequality constrained hypothesis which complicates the use of information criteria (e.g., AIC, BIC).
- Classical p-values are not suited for testing multiple hypotheses simultaneously or for testing hypotheses with combinations with equality and inequality constraints.

Bayes Factor Testing

Bayes factor test

- Definition. The Bayes factor is defined by

$$B_{12} = \frac{\iiint_{\mathcal{C}_1} f_1(\mathbf{Y}|\mathbf{B}, \boldsymbol{\sigma}, \boldsymbol{\rho}_1)\pi_1(\mathbf{B}, \boldsymbol{\sigma}, \boldsymbol{\rho}_1)d\mathbf{B}d\boldsymbol{\sigma}d\boldsymbol{\rho}_1}{\iiint_{\mathcal{C}_2} f_2(\mathbf{Y}|\mathbf{B}, \boldsymbol{\sigma}, \boldsymbol{\rho}_2)\pi_2(\mathbf{B}, \boldsymbol{\sigma}, \boldsymbol{\rho}_2)d\mathbf{B}d\boldsymbol{\sigma}d\boldsymbol{\rho}_2},$$

- Interpretation. The Bayes factor quantifies the relative predictive adequacy of the hypotheses and priors under consideration. Therefore, the outcome of the Bayes factor can be seen as a relative measure of support in the data between two hypotheses.
- Setting. Prior information for parameters is absent.
- Orthogonal parameters. $\mathbf{B}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\rho}$ are orthogonal parameters. Therefore, use independent priors, and use noninformative improper priors for the common nuisance parameters:

$$\pi_t(\mathbf{B}, \boldsymbol{\sigma}, \boldsymbol{\rho}_t) = \sigma_1^{-1} \times \ldots \times \sigma_{p_1}^{-1} \times \pi_t(\boldsymbol{\rho}_t).$$

Prior choice 1: Uniform constrained priors

Specification

- Uniform prior under constrained hypotheses:

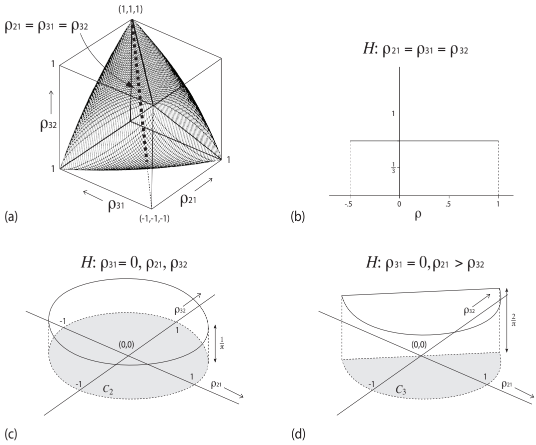
$$\pi_t^U(\boldsymbol{\rho}_t) = V_t^{-1} \times 1(\boldsymbol{\rho}_t \in \mathcal{C}_t)$$

  with

$$V_t^{-1} = \int_{\mathcal{C}_t} 1 d\boldsymbol{\rho}_t.$$

- For an order constrained hypothesis, V_t can be seen as the "volume" of the subspace.

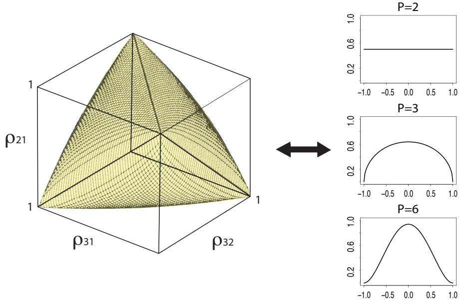
Prior choice 1: Uniform constrained priors



(upper left figure from Rousseeuw & Molenberghs, 1994)

Prior choice 1: Uniform constrained priors

Marginal prior: $\pi_u^U(\rho_{gh}) = beta(\frac{P}{2}, \frac{P}{2})$ on $(-1, 1)$ for a $P \times P$ correlation matrix $\mathbf{C}$ (Joe, 2006).



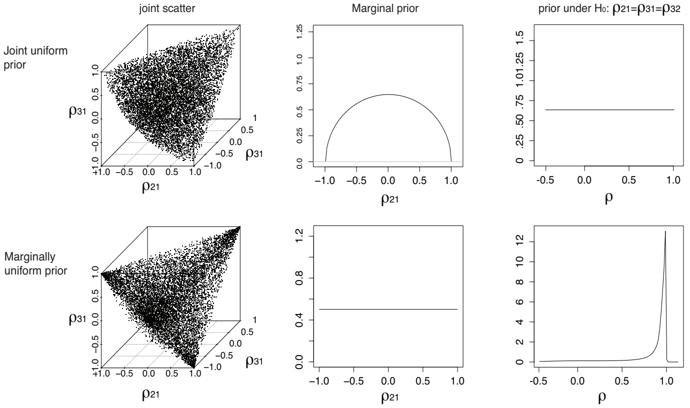
Prior choice 2: Marginal uniform prior

Marginal uniform prior

- Inverse-Wishart distribution for $\boldsymbol{\Sigma} \sim W^{-1}(\mathbf{I}_p, P+1)$.
- Transform $\boldsymbol{\Sigma} \rightarrow (\mathbf{R}, \boldsymbol{\sigma})$.
- Integrate out $\boldsymbol{\sigma}$ results in the marginal uniform prior π^{MU}, with $\pi_u^{MU}(\rho_{gh}) = U(-1, 1)$ (Barnard et al., 2000).
- Use truncations to obtain priors π_t^{MU} under the constrained hypotheses (encompassing prior approach; e.g., Berger & Mortera, 1999; Klugkist Laudy, & Hoijtink, 2005).
- May result in unreasonable priors under constrained hypotheses with equality constraints (Böing-Messing & Mulder, in prep.)

Prior comparison

Prior comparison



Conclusion: Constrained uniform priors seem preferable.

Derivation of the Bayes factor

Relation between prior under H_t and prior under H_u

- The uniform prior under a constrained hypothesis H_t is a truncation of the unconstrained uniform prior under H_u in $\mathcal{C}_t$.
- Parameter transformation under H_u by $\boldsymbol{\rho}_u \to (\boldsymbol{\rho}_t^{E'}, \boldsymbol{\rho}_t')'$:

$$\begin{bmatrix} \boldsymbol{\rho}_t^E \\ \boldsymbol{\rho}_t \end{bmatrix} = \begin{bmatrix} \mathbf{R}_t^E \\ \mathbf{R}^* \end{bmatrix} \boldsymbol{\rho}_u.$$

- Hypothesis $H_t : \mathbf{R}_t^E \boldsymbol{\rho}_u = \mathbf{r}_t^E,\ \mathbf{R}_t^I \boldsymbol{\rho}_u > \mathbf{r}_t^I$ becomes

$$H_t : \boldsymbol{\rho}_t^E = \mathbf{r}_t^E,\ \tilde{\mathbf{R}}_t^I \boldsymbol{\rho}_t > \tilde{\mathbf{r}}_t^I$$

- The truncated prior under H_t can be written as

$$\pi_t^U(\boldsymbol{\rho}_t) = \pi_u^U(\boldsymbol{\rho}_t | \boldsymbol{\rho}_t^E = \mathbf{r}_t^E) \times Pr\left(\tilde{\mathbf{R}}_t^I \boldsymbol{\rho}_t > \tilde{\mathbf{r}}_t^I | \boldsymbol{\rho}_t^E = \mathbf{r}_t^E, H_u\right)^{-1} \times I(\boldsymbol{\rho}_t \in \mathcal{C}_t).$$

Derivation of the Bayes factor

- Computation of the marginal likelihoods can be avoided because

$$
\begin{aligned}
B_{tu} &= \frac{\iiint_{\mathcal{C}_t} f_t(\mathbf{Y}|\mathbf{B},\boldsymbol{\sigma},\boldsymbol{\rho}_t)\pi_t^U(\mathbf{B},\boldsymbol{\sigma},\boldsymbol{\rho}_t)d\mathbf{B}d\boldsymbol{\sigma}d\boldsymbol{\rho}_t}{\iiint_{\mathcal{C}_u} f_u(\mathbf{Y}|\mathbf{B},\boldsymbol{\sigma},\boldsymbol{\rho}_u)\pi_u^U(\mathbf{B},\boldsymbol{\sigma},\boldsymbol{\rho}_u)d\mathbf{B}d\boldsymbol{\sigma}d\boldsymbol{\rho}_u} \\
&= \frac{\pi_u^U(\boldsymbol{\rho}_t^E = \mathbf{r}_t^E|\mathbf{Y})}{\pi_u^U(\boldsymbol{\rho}_t^E = \mathbf{r}_t^E)} \times \frac{Pr(\tilde{\mathbf{R}}_t^I\boldsymbol{\rho}_t > \tilde{\mathbf{r}}_t^I|\boldsymbol{\rho}_t^E = \mathbf{r}_t^E, \mathbf{Y}, H_u)}{Pr(\tilde{\mathbf{R}}_t^I\boldsymbol{\rho}_t > \tilde{\mathbf{r}}_t^I|\boldsymbol{\rho}_t^E = \mathbf{r}_t^E, H_u)}
\end{aligned}
$$

  (using results of Dickey, 1971; Klugkist et al., 2005: Pericchi et al., 2008; Wetzels et al., 2010; Mulder, 2014; among others).

- Interpretation. The posterior parts in the numerator can be seen as measures of **relative fit** of H_t relative to H_u. The prior parts in the denominator can be seen as measures of **relative complexity** of H_t relative to H_u.

Computation of the Bayes factor

Posterior parts $\pi_u(\boldsymbol{\rho}_t^E = \mathbf{r}_t^E|\mathbf{Y})$ and $Pr(\tilde{\mathbf{R}}_t^I \boldsymbol{\rho}_t > \tilde{\mathbf{r}}_t^I | \boldsymbol{\rho}_t^E = \mathbf{r}_t^E, \mathbf{Y}, H_u)$

- Posterior sample. Use an MCMC algorithm to obtain an unconstrained posterior sample under H_u (Chib & Greenberg, 1998; Liu & Daniels, 2006; Boscardin et al., 2008).
- Fisher transformation.
  - The Fisher Z transformed bivariate sample correlation r given ρ is approximately normal.
  - In the integrated likelihood r and ρ have a similar role because

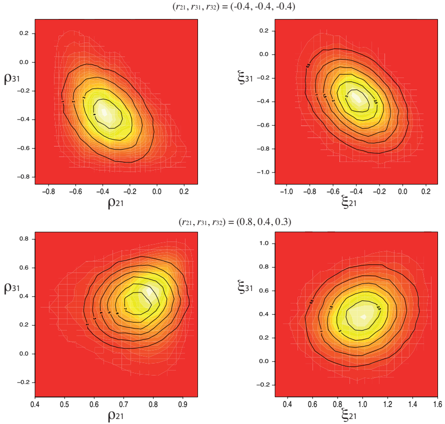
$$f(r|\rho) \propto \frac{(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}}{(1-\rho r)^{n-\frac{3}{2}}} F_{1,2}\left(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{2n-1}{2}; \tfrac{\rho r+1}{2}\right)$$

  - This implies that the Fisher transformed posterior of $\boldsymbol{\rho}$, denoted by $\boldsymbol{\xi}$, is approximately normal when using a vague prior, i.e.,

$$\pi(\boldsymbol{\xi}|\mathbf{Y}) \approx N(\mathbf{m}_{\xi}, \mathbf{S}_{\xi}).$$

Computation of the Bayes factor

- Fisher transformation.



Computation of the Bayes factor

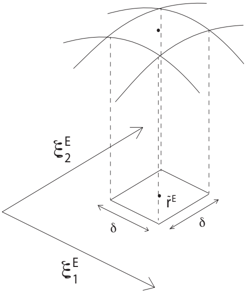
Posterior parts

- Posterior density. $\pi_u(\boldsymbol{\xi}_t^E = \mathbf{r}_t^E | \mathbf{Y})$ follows directly from the approximated normal distribution.
- Posterior probability. Estimation of $Pr(\tilde{\mathbf{R}}_t^I \boldsymbol{\xi}_t > \tilde{\mathbf{r}}_t^I | \boldsymbol{\xi}_t^E = \mathbf{r}_t^E, \mathbf{Y}, H_u)$ as proportion of draws satisfying the constraints can be inefficient. Instead use ideas from Mulder (2016) and Morey et al. (2010).
  - Transform parameters: E.g., if $Pr(\xi_{12} < \xi_{31} < \xi_{32} | \mathbf{Y})$, then
  $$(\theta_1, \theta_2) = (\xi_{31} - \xi_{21}, \xi_{32} - \xi_{31}).$$
  - Split constraints:
  $$\Pr(\theta_1 > 0, \theta_2 > 0 | \mathbf{Y}) = \Pr(\theta_1 > 0 | \mathbf{Y}) \times \Pr(\theta_2 > 0 | \theta_1 > 0, \mathbf{Y})$$
  - Use Monte Carlo estimation:
  $$\Pr(\theta_2 > 0 | \theta_1 > 0, \mathbf{Y}) \approx \frac{1}{S} \sum_s \Pr(\theta_2 > 0 | \theta_1^{(s)}, \mathbf{Y}).$$

Computation of the Bayes factor

Prior density $\pi_u(\boldsymbol{\xi}_t^E = \mathbf{r}_t^E)$

- Prior sample. Draw $\boldsymbol{\rho} \sim \pi_u^U(\boldsymbol{\rho})$ using algorithm of Joe (2006).
- Fisher transformation must also be applied on the prior draws, resulting in transformed prior draws for $\boldsymbol{\xi}$ which are not approximately normal.
- Estimation of the prior density using $\Pr(|\xi_1^E - \tilde{r}_1^E| < \frac{\delta}{2}, |\xi_2^E - \tilde{r}_2^E| < \frac{\delta}{2}) \approx \delta^2 \pi_u(\boldsymbol{\xi}^E = \tilde{\mathbf{r}}^E)$, with δ small.



Computation of the Bayes factor

Prior probability $Pr(\tilde{\mathbf{R}}_t^I \boldsymbol{\xi}_t > \tilde{\mathbf{r}}_t^I | \boldsymbol{\rho}_t^E = \mathbf{r}_t^E, H_u)$

- For few inequality constraints, e.g., $\rho_{21} > \rho_{31} > \rho_{32}$, the prior probability will be estimated as the proportion of prior draws satisfying the constraints using, say, 100,000 prior draws.
- For many inequality constraints, e.g., $\rho_{21} > \rho_{32} > \ldots > \rho_{98}$, the prior is approximated with a normal distribution and the same methodology is used to estimate the prior probability as was suggested for the posterior probability.

Numerical example

Behavior of the criterion

- Consider a 3 $\times$ 3 correlation matrix for 3 dependent variables of which the first two are measured on a continuous scale, and the third is measured on an ordinal scale with 3 level.
- Multiple hypothesis test:
  1. $H_1 : \rho_{21} = \rho_{31} = \rho_{32}$.
  2. $H_2 : \rho_{21} > \rho_{31} > \rho_{32}$.
  3. $H_3 :$ not H_1, H_2,
- Data generated under

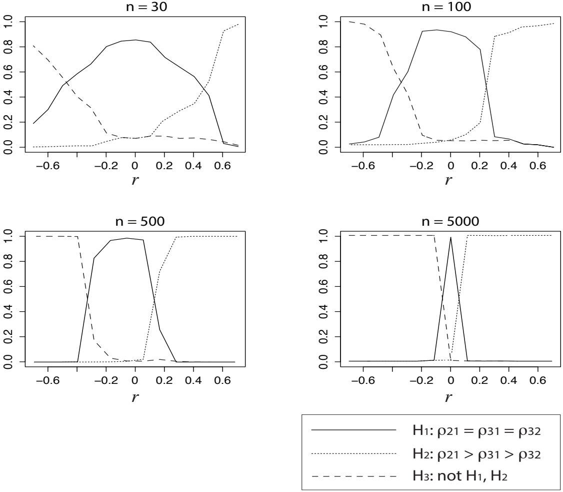
$$\mathbf{C} = \begin{bmatrix} 1 & & \\ \rho_{21} & 1 & \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ r & 1 & \\ r/2 & 0 & 1 \end{bmatrix}.$$

  for $r = -.7, \ldots, .7$.
- Equal prior probabilities for the hypotheses:

$$P(H_1) = P(H_2) = P(H_3) = \tfrac{1}{3}.$$

Numerical example

Behavior of the criterion



Software

BCT

- The methodology is written in a Fortran program called BCT (Bayesian Correlation Testing).
- An R-package will be available soon.
- Implementation in JASP (JASP Team, 2016) is planned for the future as well.

Empirical Application

Data and results

A sample of size $N = 1286$ was collected in the Netherlands.

- Self reported happiness (DV1) was measured on a 4-point scale.
- Health (DV2) was measured on a 5-point scale.
- Educational level (DV3) was measured on a 7-point scale.
- Personal-sexual permissiveness (DV4) was measured on a continuous scale.

Research question. Which hypothesis about predicting someone's happiness receives most support?

$$
\begin{aligned}
H_1 &: \rho_{21} = \rho_{31} = \rho_{41} \\
H_2 &: \rho_{21} > \rho_{31} > \rho_{41} > 0 \\
H_3 &: \rho_{21} > 0,\ \rho_{31} = \rho_{41} = 0 \\
H_4 &: \rho_{21} = \rho_{31} = \rho_{41} = 0 \\
H_5 &: \text{not } H_1, H_2, H_3, H_4.
\end{aligned}
$$

Empirical Application

Data and results

A sample of size $N = 1286$ was collected in the Netherlands.

- Self reported happiness (DV1) was measured on a 4-point scale.
- Health (DV2) was measured on a 5-point scale.
- Educational level (DV3) was measured on a 7-point scale.
- Personal-sexual permissiveness (DV4) was measured on a continuous scale.

Research question. Which hypothesis about predicting someone's happiness receives most support?

$$
\begin{array}{lcll}
H_1 & : & \rho_{21} = \rho_{31} = \rho_{41} & Pr(H_1|\mathbf{Y}) = .000 \\
H_2 & : & \rho_{21} > \rho_{31} > \rho_{41} > 0 & Pr(H_2|\mathbf{Y}) = .054 \\
H_3 & : & \rho_{21} > 0,\ \rho_{31} = \rho_{41} = 0 & Pr(H_3|\mathbf{Y}) = .902 \\
H_4 & : & \rho_{21} = \rho_{31} = \rho_{41} = 0 & Pr(H_4|\mathbf{Y}) = .000 \\
H_5 & : & \text{not } H_1, H_2, H_3, H_4. & Pr(H_5|\mathbf{Y}) = .044.
\end{array}
$$

Summary

- Researchers often have expectations about the degree of association between certain variables of interest. These expectations can be translated in statistical hypotheses with equality and inequality constraints on bivariate, partial, and ordinal correlations.
- Bayes factors are useful to test such hypotheses in a direct manner.
- Uniform constrained priors seem reasonable as a default setting.
- The methodology is implemented in the software program BCT.

Outline

Information inconsistency

Bayesian t testing example using Zellner's (1986) g prior

- Data: $y_i \sim N(\theta, \sigma^2)$, for $i = 1, \ldots, n$, for unknown θ and σ^2.
- Hypothesis test: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.
- Sufficient statistics: $\bar{y}$ and $s_y^2 = \sum_i (y_i - \bar{y})^2$.
- g prior choice: under $H_0 : \pi_0(\sigma^2) = \sigma^{-2}$; under $H_1 : \pi_1(\theta, \sigma^2) = N_{\theta|\sigma^2}(\theta_0, g\sigma^2/n) \times \sigma^{-2}$.
- Bayes factor of H_1 versus H_0 yields

$$B_{10} = (1+ng)^{-\frac{1}{2}} \left(1 - \frac{t^2}{(n-1)+t^2} \times \frac{ng}{1+ng}\right)^{-\frac{n}{2}},$$

  where $t = \frac{\bar{y}-\theta_0}{s_y/\sqrt{n-1}}$ is the usual test statistic.
- As $t \to \infty$, then $B_{10} \to (1+ng)^{(n-1)/2} < \infty$. This is called information inconsistency.

Information inconsistency

Goals

1. Investigate information inconsistency for the most commonly used priors: conjugate priors (proper or improper), independent priors, adaptive priors.
2. Investigate information inconsistency for precise hypothesis testing and one-sided hypothesis testing.
3. If information inconsistency occurs, investigate the severity of the problem from a practical point if view.

Joint work

This is joint work with Jim Berger and Susie Bayarri: Mulder, Berger, & Bayarri (in preparation), Mulder (2014).

Notation

Linear regression model

The response vector $\mathbf{y}$ of length n is modeled according to

$$\begin{aligned} \text{Model} &: \quad \mathbf{y} = \mathbf{X}\beta + \epsilon, \text{ with } \epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{\Sigma}) \\ \text{Precise hypothesis test} &: \quad H_0 : \mathbf{R}\beta = \mathbf{0} \text{ vs } H_1 : \mathbf{R}\beta \neq \mathbf{0}, \\ \text{One-sided hypothesis test} & \quad\quad H_0 : \mathbf{R}\beta \leq \mathbf{0} \text{ vs } H_1 : \mathbf{R}\beta \not\leq \mathbf{0}, \end{aligned}$$

where $\mathbf{\Sigma}$ is a known correlation matrix.

Reparametrization

Set $\boldsymbol{\theta} = \mathbf{R}\beta$. Then,

$$\begin{aligned} \text{Model} &: \quad \mathbf{y} = \mathbf{Z}_1\boldsymbol{\theta} + \mathbf{Z}_2\gamma + \epsilon \\ \text{Precise hypothesis test} &: \quad H_0 : \boldsymbol{\theta} = \mathbf{0} \text{ vs } H_1 : \boldsymbol{\theta} \neq \mathbf{0}, \\ \text{One-sided hypothesis test} & \quad\quad H_0 : \boldsymbol{\theta} \leq \mathbf{0} \text{ vs } H_1 : \boldsymbol{\theta} \not\leq \mathbf{0}, \end{aligned}$$

Conjugate Priors – improper and proper

Conjugate priors

$$\begin{aligned}
\pi_0(\gamma_0, \sigma_0^2) &\propto \pi_0(\sigma_0^2) = \text{inv-}\chi^2_{\sigma_0^2}(\nu_0, s_0^2) \\
\pi_1(\boldsymbol{\theta}, \gamma_1, \sigma_1^2) &\propto \pi_1(\boldsymbol{\theta} \mid \sigma_1^2)\, \pi_1(\sigma_1^2) \\
&= N_{\boldsymbol{\theta}|\sigma_1^2}(\mathbf{0}, \sigma_1^2\boldsymbol{\Omega})\ \text{inv-}\chi^2_{\sigma_1^2}(\nu_1, s_1^2).
\end{aligned}$$

Bayes factor for precise hypothesis test

The Bayes factor of H_1 to H_0 is then, with $\hat{\boldsymbol{\theta}} = \left(\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_1\right)^{-1}\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}\mathbf{y}$,

$$B_{10} = C_1 \times \frac{\left(s_{\mathbf{y}}^2 + s_1^2\nu_1 + \hat{\boldsymbol{\theta}}'\left(\left(\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_1\right)^{-1} + \boldsymbol{\Omega}\right)^{-1}\hat{\boldsymbol{\theta}}\right)^{-(n+\nu_1-k_2)/2}}{\left(s_{\mathbf{y}}^2 + s_0^2\nu_0 + \hat{\boldsymbol{\theta}}'\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_1\hat{\boldsymbol{\theta}}\right)^{-(n+\nu_0-k_2)/2}},$$

where C_1 is a contant that does not depend on $\hat{\boldsymbol{\theta}}$.

Conjugate Priors – improper and proper

Lemma

As $|\hat{\boldsymbol{\theta}}| \to \infty$,

$$B_{10} \quad \to \quad \begin{cases} 0 & \text{if } \nu_0 < \nu_1; \\ C_1 \left(\lim_{|\hat{\boldsymbol{\theta}}| \to \infty} \frac{\hat{\boldsymbol{\theta}}' \mathbf{Z}_1' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_1 \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}' \left(\left(\mathbf{Z}_1' \boldsymbol{\Sigma}^{-1} \mathbf{Z}_1 \right)^{-1} + \boldsymbol{\Omega} \right)^{-1} \hat{\boldsymbol{\theta}}} \right)^{\frac{(n+\nu-k_2)}{2}} < \infty & \text{if } \nu_0 = \nu_1; \\ \infty & \text{if } \nu_0 > \nu_1. \end{cases}$$

- $\nu_0 > \nu_1$ results in information consistency. This is not a logical prior because it implies that the distribution of σ_0^2 is more concentrated than the distribution of σ_1^2. The prior that results in the Savage-Dickey Bayes factors is a special case.
- $\nu_0 = \nu_1$ is the usual choice (in the objective Bayesian approach these would both be 0) and results in information inconsistency.
- $\nu_0 < \nu_1$ might seem logical, but is disastrously information inconsistent.

Independence Priors – improper and proper

Independence priors

Under this choice, the parameters $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$, and σ^2 are modeled independently a priori. We consider the conditionally conjugate case.

$$\begin{aligned}
\pi_0(\boldsymbol{\gamma}_0, \sigma_0^2) &= \pi_0(\boldsymbol{\gamma}_0) \times \pi_0(\sigma_0^2) \\
&= 1 \times \text{inv-}\chi^2_{\sigma_0^2}(\nu_0, s_0^2) \\
\pi_1(\boldsymbol{\theta}, \boldsymbol{\gamma}_1, \sigma_1^2) &= \pi_1(\boldsymbol{\theta}) \times \pi_1(\boldsymbol{\gamma}) \times \pi_1(\sigma_1^2) \\
&= N_{\boldsymbol{\theta}}(\mathbf{0}, \boldsymbol{\Omega}) \times 1 \times \text{inv-}\chi^2_{\sigma_1^2}(\nu_1, s_1^2).
\end{aligned}$$

Independence Priors – improper and proper

Using the result of Dawid (1972) we get the following

$$
\begin{aligned}
B_{10} &= C_2 \frac{\int \left(\nu_1 s_1^2 + s_{\mathbf{y}}^2 + (\theta - \hat{\theta})' \mathbf{Z}_1' \mathbf{\Sigma}^{-1} \mathbf{Z}_1 (\theta - \hat{\theta})\right)^{-\frac{n-k_2+\nu_1}{2}} N_{\theta}(\mathbf{0}, \mathbf{\Omega}) d\theta}{\left(\nu_0 s_0^2 + s_{\mathbf{y}}^2 + \hat{\theta}' \mathbf{Z}_1' \mathbf{\Sigma}^{-1} \mathbf{Z}_1 \hat{\theta}\right)^{-\frac{n-k_2+\nu_0}{2}}} \\
&\to C_2 \frac{\left(\nu_1 s_1^2 + s_{\mathbf{y}}^2 + \hat{\theta}' \mathbf{Z}_1' \mathbf{\Sigma}^{-1} \mathbf{Z}_1 \hat{\theta}\right)^{-\frac{n-r_2+\nu_1}{2}}}{\left(\nu_0 s_0^2 + s_{\mathbf{y}}^2 + \hat{\theta}' \mathbf{Z}_1' \mathbf{\Sigma}^{-1} \mathbf{Z}_1 \hat{\theta}\right)^{-\frac{n-r_2+\nu_0}{2}}} \\
&\to \begin{cases} \infty & \text{if } \nu_0 > \nu_1; \\ 1 & \text{if } \nu_0 = \nu_1; \\ 0 & \text{if } \nu_0 < \nu_1, \end{cases}
\end{aligned}
$$

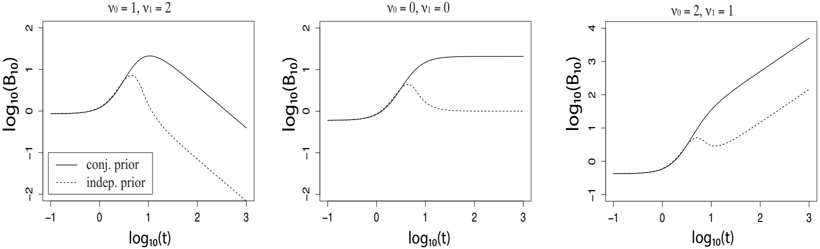
as $|\hat{\theta}| \to \infty$.

Numerical examples

Explanation of result when $\nu_0 = \nu_1$

Independence priors results in an even worse form of information inconsistency than in the natural conjugate priors. The reason is that extremely large effects of are equally unlikely under H_0 as under H_1 with the light-tailed normal prior resulting in equal evidence between the hypotheses in the limit.

Numerical Example ($n = 7$, $\rho = .5$, $s_0^2 = s_1^2 = 1$, $s_{\mathbf{y}}^2 = 6$)



Independence Priors (thick-tailed)

Based on the result of Dawid (1972) it implies that

$$B_{10} = C_2 \frac{\int \left(\nu_1 s_1^2 + s_{\mathbf{y}}^2 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_1(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right)^{-\frac{n-k_2+\nu_1}{2}} \pi_1(\boldsymbol{\theta})d\boldsymbol{\theta}}{\left(\nu_0 s_0^2 + s_{\mathbf{y}}^2 + \hat{\boldsymbol{\theta}}'\mathbf{Z}_1'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_1\hat{\boldsymbol{\theta}}\right)^{-\frac{n-k_2+\nu_0}{2}}}$$

$$\rightarrow \infty, \text{ as } |\hat{\boldsymbol{\theta}}| \rightarrow \infty,$$

only if

$$|\boldsymbol{\theta}|^{n-k_2}\pi_1(\boldsymbol{\theta}) \rightarrow \infty.$$

For this to hold even for a minimal sample size where $n = k_1 + k_2 + 1$, the tails of $\pi_1(\boldsymbol{\theta})$ must be of order smaller than $|\boldsymbol{\theta}|^{-(n-k_2)} = |\boldsymbol{\theta}|^{-(k_1+1)}$, which implies thicker tails than a (multivariate) Cauchy distribution.

Adaptive priors

Adaptive prior specification

Another possibility is to let the prior adapt to the likelihood.

- An adaptive independence prior of the form $N_{\boldsymbol{\theta}}(\mathbf{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega}$ is chosen such that it maximizes the marginal likelihood under H_1. This results in an information consistent Bayes factor.
- Similarly by choosing g in the g-prior such that it maximizes the marginal likelihood under H_1 also results in an information consistent Bayes factor.

One-sided hypothesis test

Test $H_0 : \boldsymbol{\theta} \leq \mathbf{0}$ versus $H_1 : \boldsymbol{\theta} \not\leq \mathbf{0}$

Definition: A Bayes factor is called **information consistent** for a one-sided hypothesis test if and only if $B_{10} \rightarrow \infty$ if for all limits where at least one element of $\hat{\boldsymbol{\theta}}$ goes to $+\infty$, and $B_{10} \rightarrow 0$, as all elements of $\hat{\boldsymbol{\theta}}$ go to $-\infty$. Otherwise, a Bayes factor is called information inconsistent.

Encompassing prior approach

It is natural to set an encompassing prior under $\boldsymbol{\Theta}_u = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$, and set truncated priors under H_0 and H_1 (e.g., Berger & Mortera, 1999), i.e.,

$$p_q(\boldsymbol{\theta}, \gamma, \sigma^2) = p_u(\boldsymbol{\theta}, \gamma, \sigma^2) I_{\boldsymbol{\Theta}_q}(\boldsymbol{\theta}) / \Pr(\boldsymbol{\theta} \in \boldsymbol{\Theta}_q | H_u).$$

Then

$$B_{10} = \frac{1 - Pr(\boldsymbol{\theta} \leq \mathbf{0} | \mathbf{y}, H_u)}{Pr(\boldsymbol{\theta} \leq \mathbf{0} | \mathbf{y}, H_u)} \times \frac{Pr(\boldsymbol{\theta} \leq \mathbf{0} | H_u)}{1 - Pr(\boldsymbol{\theta} \leq \mathbf{0} | H_u)}$$

One-sided hypothesis test

Special choice of prior odds of the hypotheses

It is well-known that

$$\frac{Pr(H_1|\mathbf{y})}{Pr(H_0|\mathbf{y})} = B_{10} \times \frac{Pr(H_1)}{Pr(H_0)}.$$

For the special case where the prior odds of the hypotheses are based on the probability of the subspaces of the parameters under the encompassing model, i.e.,

$$\frac{Pr(H_1)}{Pr(H_0)} = \frac{1 - Pr(\boldsymbol{\theta} \leq \mathbf{0}|H_u)}{Pr(\boldsymbol{\theta} \leq \mathbf{0}|H_u)}.$$

then the posterior odds corresponds to the posterior probabilities of the subspaces under $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_0$ under H_u,

$$\frac{Pr(H_1|\mathbf{y})}{Pr(H_0|\mathbf{y})} = \frac{1 - Pr(\boldsymbol{\theta} \leq \mathbf{0}|\mathbf{y}, H_u)}{Pr(\boldsymbol{\theta} \leq \mathbf{0}|\mathbf{y}, H_u)}.$$

One-sided hypothesis test

Conjugate encompassing prior

$$
\begin{aligned}
p(\boldsymbol{\theta}, \gamma, \sigma^2) &= p(\boldsymbol{\theta}|\sigma^2) \times p(\gamma) \times p(\sigma^2) \\
&\propto N_{\boldsymbol{\theta}|\sigma^2}(\mathbf{0}, \sigma^2 \boldsymbol{\Omega}) \times 1 \times \text{inv-}\chi^2_{\sigma^2}(\nu, s^2),
\end{aligned}
$$

and set truncated priors under H_0 and H_1, i.e.,

$$
p_q(\boldsymbol{\theta}|\sigma^2) = p(\boldsymbol{\theta}|\sigma^2) I_{\boldsymbol{\Theta}_q}(\boldsymbol{\theta}) / \Pr_\pi(\boldsymbol{\theta} \in \boldsymbol{\Theta}_q|\sigma^2),
$$

and $p_q(\gamma) = p(\gamma)$ and $p_q(\sigma^2) = p(\sigma^2)$, for $q = 0$ or 1.

Independence encompassing prior

$$
\begin{aligned}
p(\boldsymbol{\theta}, \gamma, \sigma^2) &= p(\boldsymbol{\theta}) \times p(\gamma) \times p(\sigma^2) \\
&\propto N_{\boldsymbol{\theta}}(\mathbf{0}, \boldsymbol{\Omega}) \times 1 \times \text{inv-}\chi^2_{\sigma^2}(\nu, s^2).
\end{aligned}
$$

One-sided hypothesis test

Lemma

When at least one element of $\hat{\boldsymbol{\theta}}$ goes to $+\infty$, the Bayes factor for the one-sided test based on the conjugate encompassing prior goes to

$$B_{10} \quad \rightarrow \quad \frac{1 - Pr^*(\boldsymbol{\theta} \leq \mathbf{0}|\mathbf{y}, H_u)}{Pr^*(\boldsymbol{\theta} \leq \mathbf{0}|\mathbf{y}, H_u)} \times \frac{Pr(\boldsymbol{\theta} \leq \mathbf{0}|H_u)}{1 - Pr(\boldsymbol{\theta} \leq \mathbf{0}|H_u)} < \infty,$$

where the limiting posterior probability Pr^* depends on the direction of the limit.

Lemma

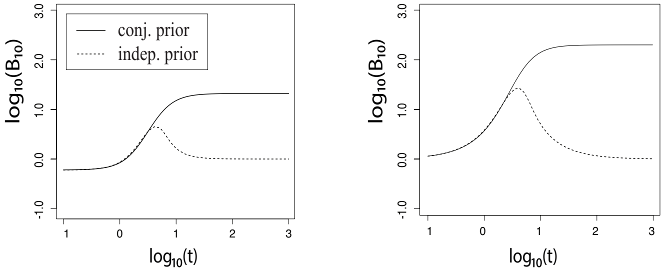
When at least one element of $\hat{\boldsymbol{\theta}}$ goes to $+\infty$, the Bayes factor for the one-sided test based on the independence encompassing prior yields

$$B_{10} \quad \rightarrow \quad \frac{Pr(\boldsymbol{\theta} \leq \mathbf{0}|H_u)}{1 - Pr(\boldsymbol{\theta} \leq \mathbf{0}|H_u)}.$$

One-sided hypothesis test

Numerical example ($n = 7$, $s_{\mathbf{y}}^2 = 6$, $\rho = .5$, and $\nu = 0$)

$H_0 : \theta = 0 \;\; vs \;\; H_1 : \theta \neq 0$ $\qquad$ $H_0 : \theta \leq 0 \;\; vs \;\; H_1 : \theta > 0$



One-sided hypothesis test

Possible solutions (among others)

- An independence encompassing prior with a (multivariate) Cauchy distribution (nonadaptive). This follows from the result of Dawid (1973).
- An encompassing conjugate g-prior where g goes to ∞ (nonadaptive). Note that the marginal likelihoods under H_0 and H_1 go to zero in the limit, but the ratio of marginal likelihoods (i.e., the Bayes factor) is well-defined.
- An adaptive encompassing g-prior where g maximizes the marginal likelihood of H_1 (H_0) if $\hat{\boldsymbol{\theta}} \not\leq \mathbf{0}$ ($\hat{\boldsymbol{\theta}} \leq \mathbf{0}$).

Multiple hypothesis test

$H_0 : \theta = 0$ versus $H_1 : \theta < 0$ versus $H_2 : \theta > 0$

- **Conjugate prior with $\nu_0 > \nu$.** Although this (nonlogical) setting resulted in an information consistent in the precise test, the limiting posterior probabilities satisfy

$$\Pr^*(H_2|\mathbf{y}) > \Pr^*(H_1|\mathbf{y}) > \Pr^*(H_0|\mathbf{y}) = 0,$$

as $t \to +\infty$ when $\Pr(H_0) = \Pr(H_1) = \Pr(H_2) = \frac{1}{3}$. Thus, a 'negative effect' receives more support than 'no effect' in the limit.

- **Independence encompassing prior.** Let $\nu_0 = \nu$. As $t \to \pm\infty$, the limiting posterior probabilities for the hypotheses satisfy

$$\Pr^*(H_2|\mathbf{y}) = \Pr^*(H_1|\mathbf{y}) = \Pr^*(H_0|\mathbf{y}) = \frac{1}{3}.$$

Summary

- Information inconsistency is ubiquitous.
  - It happens with conjugate and g-priors.
  - It happens with independence priors almost always; tails flatter than Cauchy are required to ensure information consistency.
  - It happens when testing of precise hypotheses, one-sided hypotheses, and combinations.
- It can be a practical issue when testing precise hypotheses in the case of small samples with dependent data. For one-sided hypotheses, the problem is less severe.
- If one adopts information consistency as a criterion, the class of priors to consider is drastically reduced.

Practical consequences of information inconsistency

Testing a precise hypothesis: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$

- For $\rho = 0$, the limit is $(1+n)^{(n-1)/2}$ is usually large; and the Bayes factors for, say, $t = 4$ are reasonable compared with the generic calibration $B_{10} \approx 1/[-ep \log p]$ (actually an upper bound, Sellke, Bayarri and Berger, 2001).
- But the limits and Bayes factors for $t = 4$ seem too small for larger correlations.

	n	2	5	7	10	20
$\rho = 0$	limit	1.73	36	512	$4.85 \cdot 10^4$	$1.79 \cdot 10^{11}$
	B_{10} for $t = 4$	1.55	6.36	12.21	23.61	66.20
$\rho = 0.5$	limit	1.53	7.10	20.8	106	$2.01 \cdot 10^4$
	B_{10} for $t = 4$	1.42	3.46	5.31	8.54	20.71
$\rho \approx 1$	limit	1.41	4	8	22.6	724
	B_{10} for $t = 4$	1.34	2.76	3.44	4.86	9.47
	p-value for $t = 4$	0.156	0.016	0.0071	0.0031	0.00077
	$B_{10} \approx 1/[-ep \log p]$	2.25	7.81	13.47	24.40	72.01

Testing constrained hypotheses on measures of association

Dependent overlapping correlations: Multimethod-multitrait example

$$
\begin{array}{rl}
 & \begin{array}{cc} \quad\quad\quad\quad\quad\quad \textit{Superiors} & \quad\quad\quad \textit{Peers} \end{array} \\
\mathbf{R} = \begin{array}{c} \textit{Quality} \\ \textit{Ability} \\ \textit{Quality} \\ \textit{Ability} \end{array} &
\left[\begin{array}{cc|cc}
1 & & & \\
\rho_{AS.QS} & 1 & & \\
\hline
\rho_{QP.QS} & \rho_{QP.AS} & 1 & \\
\rho_{AP.QS} & \rho_{AP.AS} & \rho_{AP.QP} & 1
\end{array}\right]
\end{array}
$$

Monotrait-heteromethod correlations > heterotrait-monomethod correlations > heterotrait-heteromethod correlations (Campbell & Fiske, 1959):

$$
\begin{aligned}
H_1 &: (\rho_{QP.QS}, \rho_{AP.AS}) > (\rho_{AS.QS}, \rho_{AP.QP}) > (\rho_{AP.QS}, \rho_{QP.AS}) \\
H_2 &: \text{not } H_1
\end{aligned}
$$

Testing constrained hypotheses on measures of association

Dependent nonoverlapping correlations: Repeated measurements

$$
\mathbf{C} = \begin{array}{c} \\ M \\ R \\ C \\ M \\ R \\ C \\ M \\ R \\ C \end{array}
\begin{array}{c}
\begin{array}{ccc} 1986 & 1988 & 1990 \end{array} \\
\left[\begin{array}{ccc|ccc|ccc}
1 & & & & & & & & \\
\rho_{RM86} & 1 & & & & & & & \\
\rho_{CM86} & \rho_{CR86} & 1 & & & & & & \\
\hline
 & & & 1 & & & & & \\
 & & & \rho_{RM88} & 1 & & & & \\
 & & & \rho_{CM88} & \rho_{CR88} & 1 & & & \\
\hline
 & & & & & & 1.000 & & \\
 & & & & & & \rho_{RM90} & 1.000 & \\
 & & & & & & \rho_{CM90} & \rho_{CR90} & 1.000
\end{array}\right]
\end{array},
$$

- Correlations between Mathematics (M), Reading Recognition (R), and Reading Comprehension (C) scores of children.
- Interest how the correlations between abilities change over time.

Testing constrained hypotheses on measures of association

Dependent nonoverlapping correlations: Repeated measurements

$$
\begin{aligned}
H_0 &: \begin{cases} \rho_{RM86} = \rho_{RM88} = \rho_{RM90} \\ \rho_{CM86} = \rho_{CM88} = \rho_{CM90} \\ \rho_{RC86} = \rho_{RC88} = \rho_{RC90} \end{cases} \\
H_1 &: \begin{cases} \rho_{RM86} < \rho_{RM88} < \rho_{RM90} \\ \rho_{CM86} < \rho_{CM88} < \rho_{CM90} \\ \rho_{RC86} < \rho_{RC88} < \rho_{RC90} \end{cases} \\
H_2 &: \begin{cases} \rho_{RM86} < \rho_{RM88} = \rho_{RM90} \\ \rho_{CM86} < \rho_{CM88} = \rho_{CM90} \\ \rho_{RC86} < \rho_{RC88} = \rho_{RC90} \end{cases} \\
H_3 &: \text{not } H_0, H_1, H_2
\end{aligned}
$$

Testing constrained hypotheses on measures of association

Independent correlations: Testing an ordered moderator effect

$$\mathbf{R} = \begin{array}{c c} & \begin{array}{cc} LR & WSS \end{array} \\ \begin{array}{c} LR \\ WSS \end{array} & \begin{bmatrix} 1 & \\ \rho_{LR.WSS|country} & 1 \end{bmatrix} \end{array}$$

Correlations between political left-right self-placement (LR) and welfare state support (WSS) depends on the institutional arrangements of welfare states (i.e., Social-Democratic countries (SD), liberal countries (L), and Mediterranean countries (M)).

$$\begin{aligned}
H_0 &: \rho_{LR.WSS|SD} = \rho_{LR.WSS|L} = \rho_{LR.WSS|M} \\
H_1 &: \rho_{LR.WSS|SD} > \rho_{LR.WSS|L} > \rho_{LR.WSS|M} \\
H_2 &: \rho_{LR.WSS|SD} = \rho_{LR.WSS|L} > \rho_{LR.WSS|M} \\
H_3 &: \text{not } H_0, H_1, H_2
\end{aligned}$$