Bayesian SEM Perspectives
(*from the hills of asymptotia*)

Albert Satorra
Universitat Pompeu Fabra. Barcelona

May 20, 2014

2014 Modern Modeling Methods
University of Connecticut, USA, May 20-21, 2014

part of this work —the panel data models— is join work with
Joan Carles Bou,
Universitat Jaume I, Spain

Table of contents

Preliminars

Some key Assumptions in SEM

A toy-model

A fixed/random intercept panel data regression

A multiple indicator dynamic panel data model

A penalized minimum discrepancy estimator, PMD

Discussion

from the hills of asymptotia . . .

- . . . where sample size is large enough, always! Bayesian SEM may not be needed, since — I been told —any prior information will be washed out by the data if sample size is very large.
- SEM (asymptotia, frequentist) produce distributions of estimates which can be regarded as the posterior distribution from a flat prior.
- For large sample size, Bayesian analysis is insensititivy to the choice of the prior?

From the hills of asymptotia (cont.)

- The practice of SEM, however, is confronted with small or moderate sample size, large models, we encounter negative variances, . . . many deviations from "asymptotia"
- This talk is comparative of SEM methods in a non asymptotia context, with emphasis on the Bayesian versus frequentist approach. A penalized estimator will also enter the scenario.

Aassumptions in SEM

Frequentist, asymptotia Distribution & moment structure

1. data (conditional to μ and Σ):

$$y_1, \ldots, y_n \sim \text{ iid } \mathcal{N}(\mu, \Sigma)$$

   This is usually relaxed to a semi-parametric setting, beyond normality; the iid is also relaxed aggregate analysis . Robust s.e. and test statistics.
2. Moment structure: $\mu = \mu(\theta)$ and $\sigma = \sigma(\theta)$, where $\sigma = \text{vec}(\Sigma)$
   Relaxed to an approximative model
   i.e. $\sqrt{n}(\sigma - \sigma_0)$ does not blow up.

Bayesian . . . plus a *prior distribution* on parameters, e.g.,

1. a prior distribution on μ, the population mean of the y_is: $\mu \sim \mathcal{N}(\mu_0, \Lambda_0)$
2. a prior distribution on Σ, the covariance matrix of the y_is: $\Sigma^{-1} \sim \mathcal{W}(\nu_0, \Omega_0^{-1})$, inverse Wishart distribtion

Algebra of posterior distributions:

- $\{\mu \mid y_1, \ldots, y_n, \Sigma\} \sim \mathcal{N}(\mu_n, \Lambda_n)$
- $\{\Sigma \mid y_1, \ldots, y_n, \mu\} \sim \mathcal{IW}(\nu_n, \Omega_n^{-1})$
  where

$$\mu_u = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y})^{-1} \tag{1}$$

$$\Lambda_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} \tag{2}$$

$$\Omega_n = \Omega_0 + nS_\mu \tag{3}$$

and

$$\nu_n = \nu_0 + n \tag{4}$$

with

$$S_\mu = \frac{1}{n}\sum_1^n (y_i - \mu)(y_i - \mu)'$$

e.g., Hoff (2009) [1],

[1] Hoff, P.D. (2009), 'A first Course in Bayesian Statistical Methods" , Springer

Gibbs sampler

Given a starting matrix $\Sigma^{(0)}$, Gibbs sampler produces $(\mu^{(s+1)}, \Sigma^{(s+1)})$ from $(\mu^{(s)}, \Sigma^{(s)})$ as follows (see Hoff, 2009) :

1. Sample $\mu^{(s+1)}$ from its full conditional distribution:
   - 1.1 compute μ_n and Λ_n from $y_1, \ldots, y_n$ and $\Sigma^{(s)}$
   - 1.2 sample: $\mu^{(s+1)} \sim \mathcal{N}(\mu_n, \Lambda_n)$
2. Sample $\Sigma^{(s+1)}$ from its full conditional distribution
   - 2.1 compute Ω_n from from $y_1, \ldots, y_n$ and $\mu^{(s+1)}$
   - 2.2 sample: $\Sigma^{(s+1)} \sim \mathcal{IW}(\nu_0 + n, \Omega_n^{(-1)})$

At each simulated Σ, we fit the model $\Sigma = \Sigma(\theta)$, obtaining a simulated value of the posterior of θ. This is our Bayesian Gibbs sampling approach to the Toy-model example discussed below.

A Toy-model:

$$\begin{aligned} V_1 &= \lambda f + \epsilon_1 \\ V_2 &= \lambda f + \epsilon_1 \\ V_3 &= \lambda f + \epsilon_1 \end{aligned} \tag{5}$$

where f is of variance 1 and the ϵ_i's are independent with the a common variance ψ. We thus have the Toy-model simple covariance structure $\Sigma = \Sigma(\theta)$,

$$\Sigma(\theta) = \lambda^2 1_3 1_3' + \psi I_3$$

where $\theta = (\lambda, \psi)'$ and Σ is the covariance matrix of say $z = (V_1, V_2, V_3)'$.
With two sample sizes

moderate large: n = 200

small: n = 20

Toy-model: n=200

Table : Analyses of the Toy-Model, n=200

Par.	True	EQSml	MPLUSml	MPLUSbay[†]	Gibbs $\nu_0 = 1$	Gibbs $\nu_0 = 3$	Gibbs $\nu_0 = 30$
λ	1	1.148	1.146	1.143	1.162	1.122	2.362
		.206	*.206*	*.225*	*.214*	*.210*	*11.398*
ψ	9	8.822	8.778	8.912	9.069	8.979	7.742
		.625	*.621*	*.621*	*.672*	*.668*	*1.237*
CHI2		5.429	5.457				
df		4	4				
p		.246	.244	0.417 [‡]			

[‡] Posterior credibility value

Toy-model: n=200

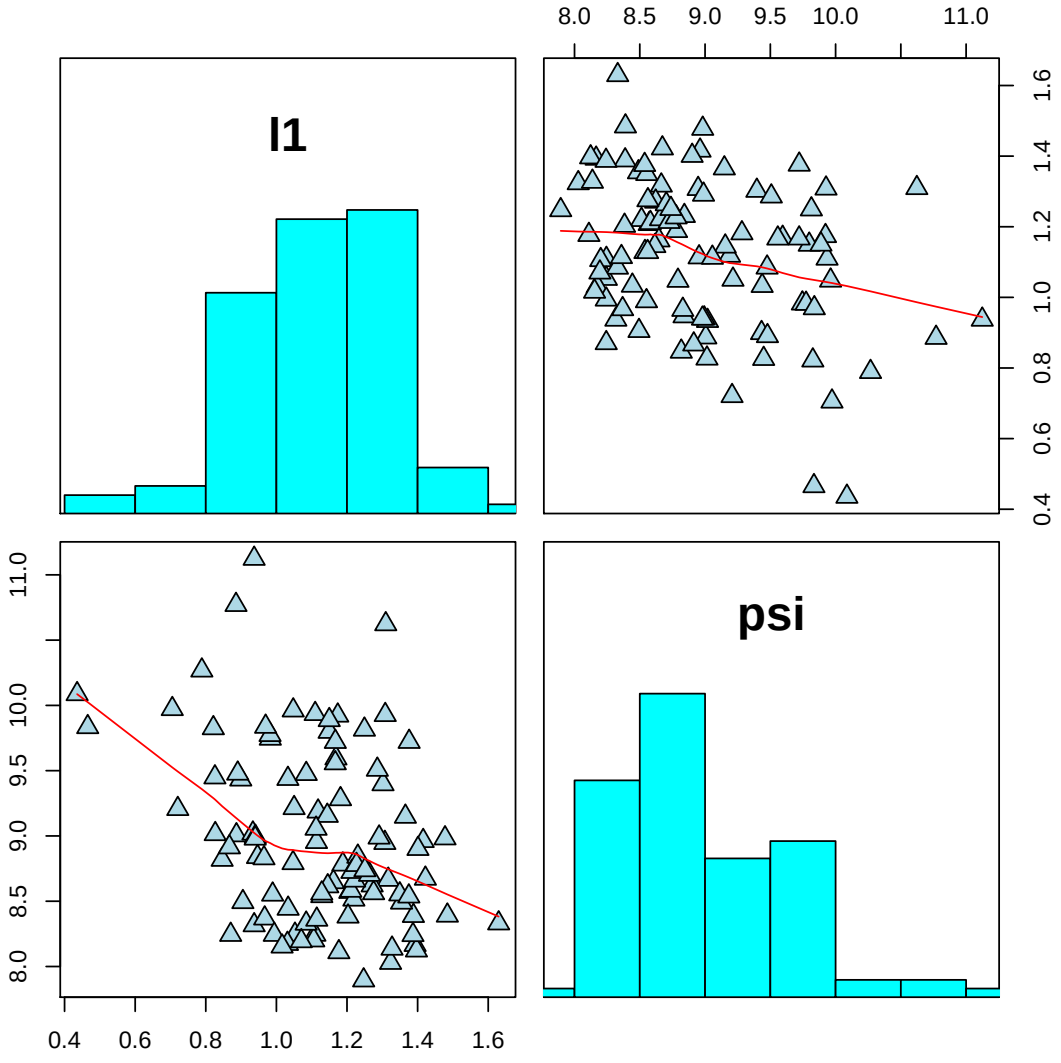

Figure : Gibbs distribution of the posterior

Toy-model: n = 20

Table : Frequentist and Bayesian analysis of the Toy-Model. Small sample size:, n =20

Par.	True	EQSml	MPLUSml	MPLUSbay †	Gibbs	Gibbs
					$\nu_0 = 3$	$\nu_0 = 10$
λ	1	.000	.000	.059	.281	.136
		4.206e+06	*.885*	*.744*	*.846*	*0.575*
ψ	9	9.044	8.592	9.173	9.649	6.872
		2.074	*2.345*	*1.775*	*2.345*	*1.663*
$CHI2_{df=4}$		5.441	5.728			
p-value		.245	.220	0.833		

Toy-model: n =20

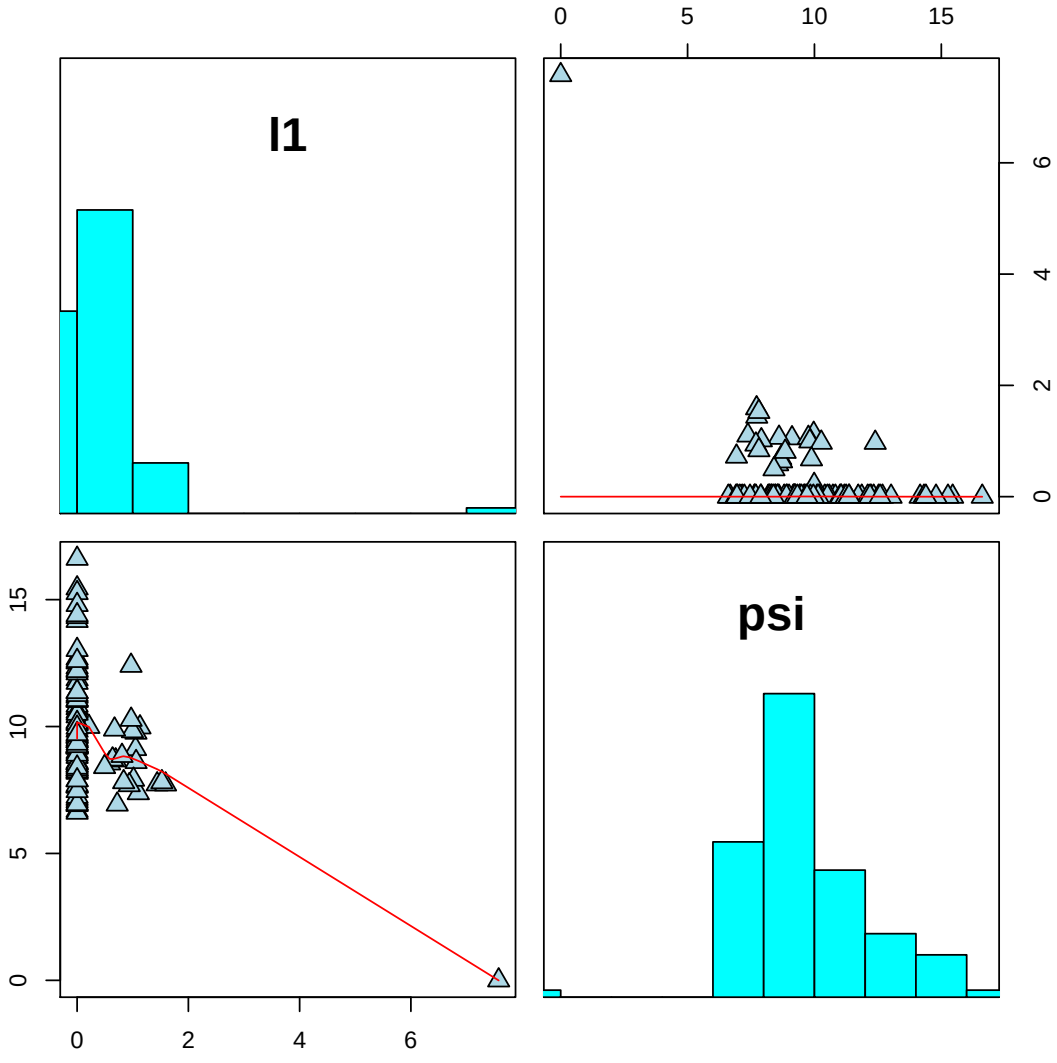

Figure : Gibbs distribution of the posterior

Panel data model I

The model is

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$$

Two sample sizes: large $n = 600$, and small $n = 40$. Tables 3 and 4 show the relevant results for simulated data. The Fixed Model approach uses `stata`. The other approaches use Mplus, frequentist and Bayesian.

Results for large sample size, n =600

Table : Analyses for simulated data, n=600

	Estimate	Mean	Variance
	Frequentist approach		
Fixed Model (n=600)			
alpha	5.320 (0.047)	–	–
beta	0.859 (0.019)	–	–
Residual variance	0.409 (0.049)	–	–
Mixed Model (n=600)			
alpha	–	5.315 (0.047)	0.927 (0.135)
beta	0.482 (0.166)	–	–
Residual variance	0.655 (0.162)	–	–
Random Model (n=600)			
alpha	–	5.309 (0.047)	1.032 (0.078)
beta	–	0.430 (0.047)	0.009 (0.006)
Residual variance	0.421 (0.029)	–	–
	Bayesian approach		
Fixed Model (n=600)			
alpha	5.321 (0.046)	–	–
beta	0.864 (0.020)	–	–
Residual variance	0.392 (0.055)	–	–
Mixed Model (n=600)			
alpha	–	5.318 (0.047)	0.818 (0.114)
beta	0.626 (0.061)	–	–
Residual variance	0.526 (0.042)	–	–
Random Model (n=600)			
alpha		–	–
beta	–	N/A	–
Residual variance	–	–	–

Analyses for mall sample size, n =40

Table : Analyses for simulated data, (n=40)

	Estimate	Mean	Variance
	Frequentist approach		
Fixed Model (n=40)			
alpha	5.425 (0.222)	–	–
beta	0.820 (0.092)	–	–
Residual variance	0.583 (0.336)	–	–
Mixed Model (n=40)			
alpha	–	5.427 (0.222)	1.200 (0.357)
beta	0.319 (0.265)	–	–
Residual variance	1.245 (0.864)	–	–
Random Model (n=40)			
alpha	–		
beta	–	No convergence	
Residual variance			
	Bayesian approach		
Fixed Model (n=40)			
alpha	5.431 (0.181)	–	–
beta	0.856 (0.047)	–	–
Residual variance	0.449 (0.131)	–	–
Mixed Model (n=40)			
alpha	–	5.442 (0.236)	1.105 (0.526)
beta	0.703 (0.137)	–	–
Residual variance	0.508 (0.159)	–	–
Random Model (n=40)			
alpha	–	–	–
beta	–	N/A	–
Residual variance	–	–	–

Multiple indicator dynamic panel data model

$y_t = (y_{1t}, \ldots, y_{Kt})'$, $j = 1, \ldots, n$, $k = 1, \ldots, K$, $t = 1, \ldots, T$.

$$
\begin{aligned}
y_{kt} &= \lambda_{kt} F_t + \epsilon_{kt} && (6)\\
F_t &= f_t + f_0 + \gamma X_t + w_t \\
f_t &= \rho f_{t-1} + v_t && (7)\\
f_0 &= \beta_1 Z_1 + \beta_2 Z_2 + v_0
\end{aligned}
$$

The f_t and ϵ_{kt} are common and unique factors; λ_{kt} are loading parameter; f_t is a dynamic component v_t is independent of f_{t-1} and v_0 is independent of Z_1 and Z_2. Note that only the vectors y_t and the variables Z_1 and Z_2, and X_1 to X_4 are observable. Correlation among time varying variables Xs and v_0 are allowed. Bou and Satorra (2007, 2010) [2] apply a similar model to profitability data of firms.

[2] Bou, J.C and A. Satorra (2009), Variation of Firm Profitability Across EU Countries: A Multi-group Structural Equation Approach, *Organizational Research Methods, 13* , 738-766;
Bou and A. Satorra (2007), The Persistence of Abnormal Returns at Industry and Firm Levels: Evidence from Spain, *Strategic Management Journal, 28(7)* 707-722.

Multiple indicator dynamic panel data model

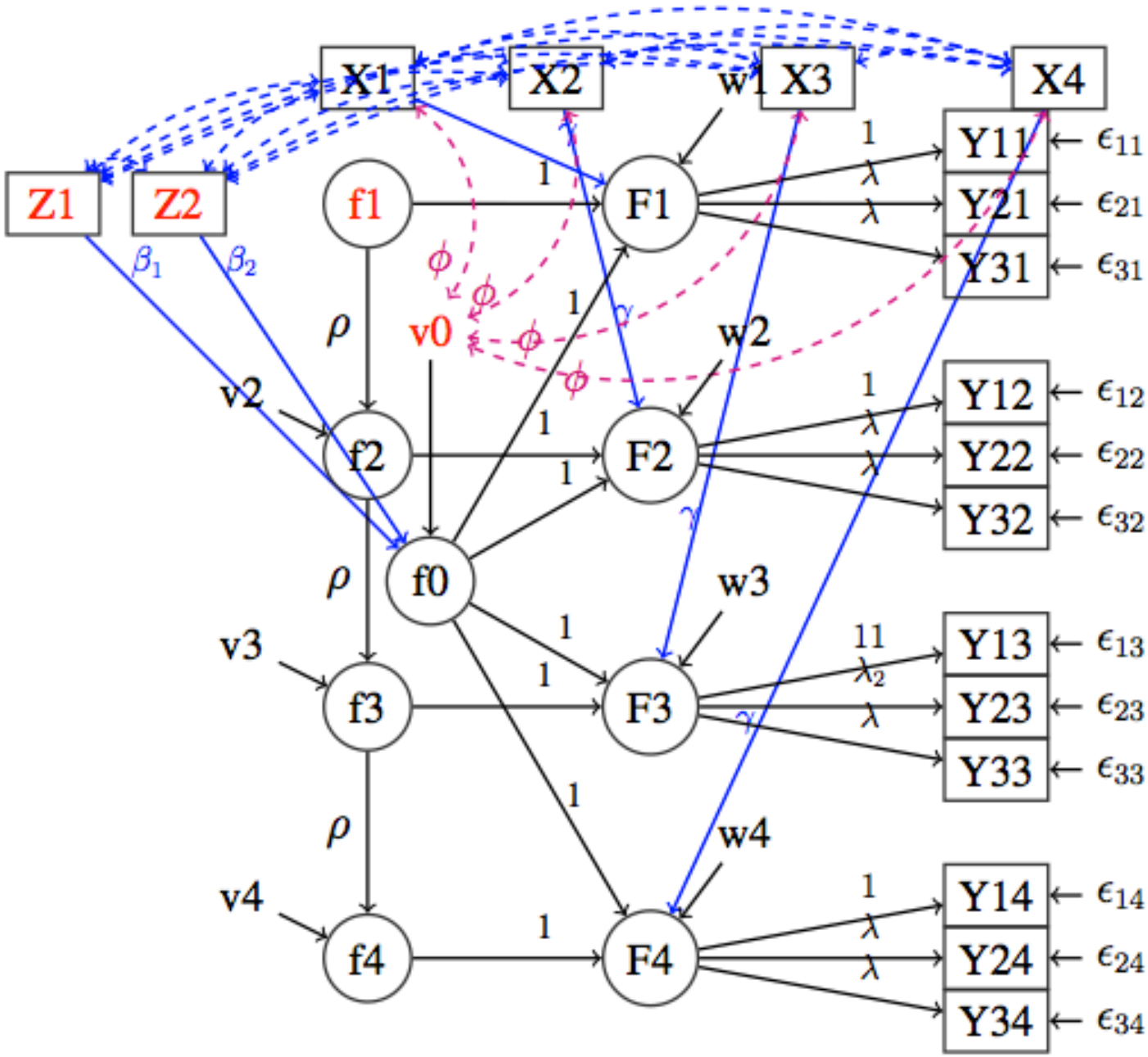

Table : ML, LS and Bayes estimates for panel data, large sample size: n = 20000 (s.e. in italic below)

Parameter	True	EQSml	EQSls	MPLUSml	MPLUSbay	PMD
λ_1	1	1	1	1	1	1
λ_2	0.800	.799	.798	.799	.799	.799
		.001[†]	*.001*	*.001*	*.001*	*.001*
ρ	0.700	.654	.789	.653	.661	.655
		.084	*.138*	*.087*	*.081*	*.083*
β_1	2.000	1.983	1.985	1.983	1.981	1.983
		.011	*.011*	*.010*	*.010*	*.010*
β_2	1.000	1.011	1.012	1.012	1.012	1.011
		.007	*.007*	*.007*	*.007*	*.007*
γ	2.000	2.008	1.994	2.009	2.009	2.008
		.006	*.011*	*.006*	*.006*	*.005*
σ^2_ϵ	0.300	.300	.299	.300	.298	.300
		.001	*.002*	*.001*	*.004*	*.001*
$\sigma^2_{f_1}$	1.000	1.001	1.180	.999	1.022	1.002
		.103	*.539*	*.098*	*.126*	*.100*
ϕ	0.000	.001	.0255	.001	.001	.001
		.011	*.017*	*.011*	*.011*	*.011*
$\sigma^2_{v_0}$	1.000	1.102	.784	1.104	1.087	1.100
		.147	*.585*	*.148*	*.171*	*.143*
$\sigma^2_{v_i}$	0.510	.531	.447	.532	.532	.531
		.056	*.067*	*.056*	*.054*	*.055*
σ^2_w	1.000	.982	1.080	.982	.984	.983
		.052	*.056*	*.053*	*.051*	*.051*
GOF Test [‡]		148.327	11.797	148.334	—	149.917
df		139	16	139	—	145
p-value		.278	0.757	.278	.347 [‡]	.372

NT for ML and Robust for LS (for ML, NT and R are practically identical)

Convergence and improper solutions

Table : Convergence and improper solutions for different sample sizes and estimation methods

	EQSml		EQSls		MPLUSml		MPLUSbay		PMD	
N	Conv.*	Prop. Sol. †	Conv.	Prop. Sol.	Conv. .	Prop. Sol.	Conv.	Prop. Sol.	Conv.	Prop. Sol.
20,000	ok	ok	ok	ok	ok	ok	ok	ok	ok	ok
1,000	ok	no	ok	no	ok	no	ok	ok	ok	ok
800	ok	no	ok	no	no	—	no	—	ok	ok
600	ok	no	ok	no	no	—	no	—	ok	ok
400	no	—	ok	no	no	—	no	—	ok	ok
200	no	—	ok	ok	ok	no	ok	ok	ok	ok

- Savalei, V. and Kolenikov, S. (2008), Constrained versus unconstrained estimation in structural equation modeling, *Psychological Methods, 13(2)*, pp. 150-170.
- Kolenikov, S. and K.A. Bollen (2012), Testing Negative Error Variances: Is a Heywood Case a Symptom of Misspecificaiton? , *Sociological Methods & Research, 41*, pp. 124-167
- Martin, J.K. and R.P. McDonald (1975), Bayesian Estimation in Unrestricted Factor Analysis: A Treatment for Heywood Cases, *Psychometrika, 40*, pp. 505-517
- Chung, Y., S. Rabe-Hesketh, V. Dorie, A. Gelma, and J. Liu (2013). A Nondegenerate Penalized Likelihood Estimator for Variance Parameters in Multilevel Models, *Psychometrika*

Table : ML, LS, Bayes and PMD estimates for panel data, small sample size: n = 200

Parameter	True	EQSIs	PMD	MPLUSml	MPLUSbay
λ_1	1	1	1	1	1
λ_2	.800	.811	.812	.811	.810
		.005	*.001*	*.005*	*.005*
ρ	.700	.284	.206	.075	.090
		.537	*.189*	*.385*	*.115*
β_1	2.000	2.038	2.048	2.048	2.038
		.072	*.079*	*.080*	*.088*
β_2	1.000	1.135	1.031	1.032	1.042
		.143	*.061*	*.061*	*.065*
γ	2.000	1.938	1.930	1.931	1.924
		.093	*.047*	*.047*	*.044*
σ^2_ϵ	0.300	0.274	.283	.281	.283
		.030	*.010*	*.010*	*.028*
$\sigma^2_{f_1}$	1.000	1.378	.826	2.789	2.022
		3.064	*.858*	*15.304*	*3.532*
$\sigma^2_{v_0}$	1.000	.045	.842	.854	.901
		.634	*.160*	*.157*	*.152*
$\sigma^2_{v_i}$	0.510	2.027	1.314	3.225	2.535
		2.958	*.875*	*15.248*	*3.507*
σ^2_w	1.000	.001	.423	-1.518	-.761
		2.850	*.847*	*15.241*	*3.472*
GOF Test ‡		5.015	127.228	121.501	–
df		12	140	140	–
p-value		.957**	0.777	0.778	0.789 ‡

Penalized estimator

Earlier work on penalization estimator in the psychometric literature is Martin and McDonald (1975). Recently, Chung, Rabe-Hesketh, Dorie, Gelman and Liu (2013) discuss penalized estimation to avoid negative estimates of variance in mixture regression models.

They used alternative penalty functions justified by specific priors on parameters. We use a simple penalization function.

Penalized estimator as a Bayesian approach (f or ML discrepancy function)

$$p(\theta \mid Y) \sim L(Y \mid \theta) p(\theta)$$

So, taking logarithms

$$\log p(\theta \mid Y) \sim L(Y \mid \theta) + \log p(\theta)$$

A penalized minimum discrepancy estimator PMD

We include in the evaluation the penalized minimum discrepancy (PMD) estimator defined simply as the minimizer of

$$F^* = F(S, \Sigma(\theta)) + g(n)(\psi - \psi_0)'(\psi - \psi_0)$$

where $F(S, \Sigma(\theta))$ is a discrepancy function, $\theta = (\theta_1, \psi')'$, with ψ being a vector of variances, and ψ_0 an a priory value for those parameter variances. We take $g(n)$ to be a decreasing function with $\lim_{n\to\infty} g(n) = 0$. In fact, in our example here, we used simply $\psi_0 = 0$ and $g(n) = A/n$.

A penalized minimum discrepancy estimator PMD (cont.)

ML case

For the ML case, we have

$$F^* = F_{ML}(S, \Sigma(\theta)) + g(n) \mid\mid (\psi - \psi_0) \mid\mid^2$$

where we assume a constant prior for the subset of parameters θ_1 and a gaussian prior with independent components for ψ. More general forms of the the penalty functions could be used. But for the model we considered here this simple penalty term does the job.

Penalized estimation (cont.)

We found high insensitivity of the estimator of the parameters defining the penalization term.

The tuning parameter A ($g(n)$) was chosen as the smallest to attaining convergence and non-negative variances.

We implemented PMD using a small trick with EQS software.

Classical diagnostic for goodness-of-fit and model modification can easily be adapted for this PMD analysis

Discussion

- When n is large, alternative methods of estimation perform similarly, on parameter estimates and s.e No convergence problems and improper solutions. arise in this set-up of a large sample (for correctly specified models)
- When samples is small, substantial differences arise among the methods, on convergence, improper solutions, and estimates. Some parameter estimates are more insensitive than others to this method variation. .
- Bayesian estimators *a la* Martin and McDonald's (1975) (adding a penalty term to the fitting function) offers a simple alternative to avoid offending (negative) variances and non-convergence.
- Fairly amount of insensitivity of parameter estimates and s.e. to the the choice of the penalty term

Discussion (cont.)

- Further investigation on the asymptotic properties of PMD is under way (asymptotic bias, asympt. variance, model modification statistics, ...)

Discussion (cont.)

- . . . so we may need to get refuge again to the *hills of asymptotia!*