Incorporating Expert Knowledge in Structural Equation Models: Applications in Psychological Research

Integrare il Parere degli Esperti con i Modelli di Equazioni Strutturali: Applicazioni nella Ricerca Psicologica

Gianmarco Altoè, Claudio Zandonella Callegher, Enrico Toffalini and Massimiliano Pastore

Abstract Structural Equation Modeling (SEM) is useful in psychological research, where complex structures of data must be modelled considering several sources of variability. Sample sizes in Psychology cannot be as large as ideal for SEM in most cases, and real effect sizes are generally small, leading to a widespread problem of insufficient power. Bayesian estimation with informed priors can be beneficial in this context. Our simulation study examines this issue over the real case of a mediation model. Parameter recovery, power and coverage were considered. The advantage of a Bayesian approach was evident for the smallest effects. Expert knowledge elicitation, i.e., the correct formalization of the theoretical expectations, is crucial. It requires increased collaboration among researchers in Psychology and Statistics.
Abstract *Abstract in Italian*

Key words: Expert elicitation, Informative Priors, Structural Equation Models (SEM), Small sample sizes, Psychological research

---

Gianmarco Altoè
Department of Developmental Psychology and Socialisation, University of Padova
e-mail: gianmarco.altoe@unipd.it

Claudio Zandonella Callegher
Department of Developmental Psychology and Socialisation, University of Padova
e-mail: claudio.zandonellacallegher@phd.unipd.it

Enrico Toffalini
Department of General Psychology, University of Padova
e-mail: enrico.toffalini@yahoo.it

Massimiliano Pastore
Department of Developmental Psychology and Socialisation, University of Padova
e-mail: massimiliano.pastore@unipd.it

1 Introduction

Structural Equation Modeling (SEM) encompasses a range of multivariate statistical techniques as, for example, confirmatory factor analysis, path analysis, or latent growth modeling. Generally, SEMs are composed of two parts: a *measurement model* and a *structural model* (see Fig. 1). The measurement model defines unobserved constructs (*latent variables*, represented in Fig. 1 as circles) according to a set of measured outcomes (*observed variables*, represented in Fig. 1 as squares), whereas the structural model describes the relationships between latent variables.

In SEM, parameters are estimated by minimizing a discrepancy function between the sample covariance matrix and the covariance matrix implied by the model. Considering the covariance structure (instead of modeling all the observed data) allows to model complex relations between variables taking into account also the measurement error.

Given their great flexibility, SEMs are widely used in psychology for various purposes such as validating psychological tests and questionnaires or evaluating hypotheses and theoretical models that involve complex relations between different latent psychological constructs.

However, as the complexity of the model increases, more data are required to obtain accurate parameter estimates and model fit statistics [15]. Recent concerns about the replicability of psychological results have raised awareness on the importance of properly define an adequate sample size [2, 8]. Nevertheless, in many research settings, the number of participants may be limited due to financial restriction, strict inclusion criteria, or clinical samples. In the case of small sample sizes, appropriate statistical techniques are required to enhance the reliability of the results.

Often in the literature, the Bayesian approach is suggested over frequentist estimation when limited data are available [3]. In the context of small sample, the inclusion of prior information can help in the parameter estimation, but researchers have to carefully consider priors choice, as estimates are highly sensitive to the prior specification (or misspecification).

The use of Bayesian statistical approach is increasing and availability of softwares, such as R-package `blavaan` [4], provide researchers with flexible tools for estimating even Bayesian structural equation models. However, most of the stud-

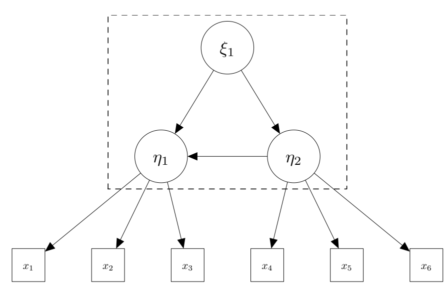

Fig. 1 A typical structural equation model with a structural part (within the dashed box) and multiple measurement models. Circles represent latent variables; rectangles represent observed variables.

ies are unlikely to carefully consider priors choices, and often they rely on default software prior settings. A recent review, considering the performance of Bayesian estimation for structural equation models with small sample sizes, underlined that the use of diffuse default priors can result in severely biased estimates, and this bias can be decreased only by incorporating informative priors [13]. Thus, authors warn against the use of *naive prior* (i.e., diffuse default priors) when samples are small, and encourage researchers to incorporate *thoughtful priors* (i.e., informative priors).

Informative priors allow researchers to include in the analysis relevant knowledge in the field, such as previous studies results, meta-analyses or expert opinions. Priors choice should be clearly discussed and researchers are required to evaluate the generalizability of external information to the specific characteristics of their study. When the number of available studies is limited or their quality is judged not adequate, relying only on previous results in the literature could be misleading. In these cases, researchers could consider to include opinions of experts as well. On the base of their experience in the field, experts can evaluate relevant information and help researchers in the definition of a plausible range of values and prior choices.

The remainder of this article is structured as follows. In Sec. 2, we briefly consider how informative priors can be defined according to experts' opinions. In Sec. 3, we present a simulation study to evaluate the influence of different prior specifications in the case of structural equation models with small sample size, considering an applied example of a mediation model in psychology. Finally, in Sec. 4, we discuss the obtained results highlighting limits and future developments of using expert knowledge with structural equation models in the case of small sample sizes.

2 Expert knowledge elicitation

Elicitation is a structured procedure that allows experts to express their knowledge and uncertainty about quantities of interest in the form of probability distributions [7]. Elicitation can be used to define informative priors according to experts' judgement.

Different elicitation methods have been proposed in the literature such as the *Cooke protocol* [1], the *Delphi method* [11] or the *SHELF protocol* [5]. The common aim of these methods is to make a subjective judgement as much objective as possible by limiting potential sources of bias, forcing the experts to carefully reason about the answer, and by documenting and transparently reporting all the procedure. In general, elicitation is composed of three phases [6]:

1. *Preparation and training* - experts are informed about the aim of the elicitation and the parameters and quantities to elicit are clearly defined. All the relevant information about the topic of interest is collected and made available to experts. Next, experts are trained to make probabilistic judgements and they are familiarize with the elicitation process in a practice example to avoid misunderstandings.
2. *Individual judgements elicitation* - experts express their own judgement for each parameter or quantity of interest according to the elicitation technique they were

trained before. Two of the main elicitation technique are the *quartile method* and the *roulette method*. In the former, experts provide values for the medians and the quartile of the distribution according to their expectation. In the latter, the range of plausible values is divided into different intervals and the experts can place a given number of tokens to allocate probabilities according to their expectation.

3. *Aggregation of individual judgements* - the individual judgements of the different experts are aggregate to obtain a unique final distribution. The two principal approaches are *mathematical aggregation*, where distributions are combined mathematically according to a pooling rule, and *behavioral aggregation*, where experts are required to discuss together their opinions and reach a consensus judgement from which the aggregate distribution is obtained.

Elicitation methods differ in the solutions adopted during each phase and in the emphasis given to the different aspects of the elicitation. For example, in the Cooke protocol experts are not required to meet each other. Experts return a form with their individual answer and mathematical aggregation is used to summarize the results. In the SHELF protocol, instead, much more emphasis is given to the collaboration between experts. After the individual judgments, experts discuss and share their opinions to then reach a common final answer.

Researchers willing to conduct an elicitation process should evaluate the pros and cons of each methods considering their own needs and constraints such as availability of experts, the possibility to group experts together, number of quantities to elicit or financial and time constraints.

3 Simulation

To evaluate the influence of different prior specifications in the case of structural equation models with small sample size, we considered a mediation model from [12] presented in Fig. 2. The study evaluated the relationship between participants' self-reported sleep quality, personality characteristics, and negative beliefs about sleeping problems. In particular, the study found that the association between participants' tendency to experience distress and become anxious (*Neuroticism*) and self-reported sleep quality (*Sleep quality*) is small ($\beta_3 = -.13$). A stronger association is given by the mediation role of the participants tendency to have negative thoughts about sleeping difficulties (*Metacognitive thoughts*). In other words, people with higher levels of distress and anxiety tend to have dysfunctional beliefs and

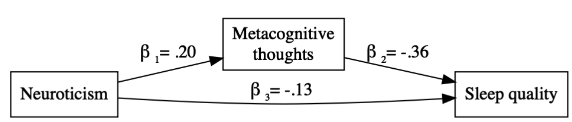

Fig. 2 Mediation model from [12]. Neuroticism is not directly associated to sleep quality but is mediated by metacognitive thoughts.

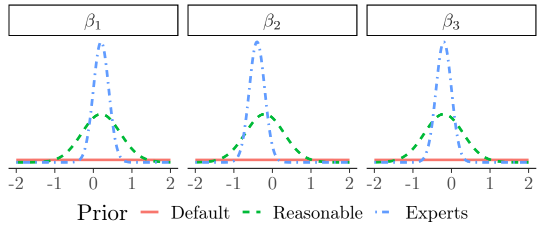

Fig. 3 Prior distribution in the three different settings. Default prior are intended to be non-informative. Reasonable prior are intended to exclude unplausible values. Experts priror represent experts judgement.

attitudes about sleep ($\beta_1 = .20$) that, in turns, induce them to perceive and report a worse-quality sleep ($\beta_2 = -.36$).

3.1 Simulation details

The simulation was carried in `R` version 3.X.X [9] using R-packages `lavaan` [10] and `blavaan` [4]. In the simulation, we considered as parameters of interest the regression coefficients (β_1, β_2, β_3) of the model presented above. We compared the performance of Maximum Likelihood (ML) estimation and Bayesian estimation under four different sample size conditions (i.e., 20, 50, 100, 500). In the current case, sample sizes below 100 can be considered small, whereas 500 is a more appropriate sample size and is considered as a reference benchmark.

Three different prior distribution specifications were used in the Bayesian estimation case (see Fig. 3):

1. *Default prior* - in `blavaan`, default priors for regression coefficients are $N(0,10)$. These priors are diffuse over a wide range of values (95% CI [-19.6; 19.6]) and are intended to be non-informative.
2. *Reasonable prior* - the priors are $\beta_1 \sim N(.20,\ .50)$, and $\beta_{2;3} \sim N(-.20,\ .50)$, that correspond, respectively, to a 95% CI of [-.78; 1.18] and [-1.18; .78]. These priors are moderately informative, they are intended to exclude excessively large values that are not reasonable within psychology research. Moreover, the mean of each prior distribution is slightly moved above or below zero to reflect the direction of the main results in the literature for each relation.
3. *Experts prior* - the priors are $\beta_1 \sim N(.20,\ .20)$, $\beta_2 \sim N(-.40,\ .20)$, and $\beta_3 \sim N(-.20,\ .20)$, that correspond, respectively, to a 95% CI of [-0.19; .59], [-0.79; .01], and [-.59; .19]. These priors are intended to be highly informative, they are intended to represent experts judgement of the quantities of interest according to previous results in the literature and their experience in the field. Confidence

To evaluate the estimation methods we used the same criteria considered by [14]: relative mean bias, relative median bias, mean square error (MSE), coverage and power. The relative mean bias (or median bias) evaluates the relative difference between mean estimate ($\bar{\theta}$; or median estimate $\widetilde{\theta}$) across replications and the population value (θ):

$$Relative\ mean\ bias = (\bar{\theta} - \theta)/\theta, \tag{1}$$

$$Relative\ median\ bias = (\widetilde{\theta} - \theta)/\theta. \tag{2}$$

Relative bias included between -.10 and .10 are considered acceptable [14]. MSE takes into account variability as well as bias of the estimates: $MSE = \sigma^2 + (\bar{\theta} - \theta)^2$, where σ is the standard deviation of the estimates across replications and $\bar{\theta}$ is the mean. Coverage is the proportion of replications in which the population value is included in the 95% confidence interval (CI; for the ML estimation) or 95% highest posterior density interval (HPD; for the Bayesian estimation). Instead, power is the proportion of replications in which the value zero is not included in the 95% CI or 95% HPD. Analyses were conducted considering the standardized parameters.

3.2 Results

The tables with detailed results for each parameter and condition are reported in the Appendix. To interpret the results of relative mean and median bias, we considered the distribution of the estimated parameters (see Fig. 4). Only with very small sample sizes ($n = 30$) it is possible to observe some differences between estimation methods: Maximum likelihood approach produces the widest distributions, whereas Bayesian approach with experts priors has narrower distributions. However, differences between methods are noticeable for the parameter β_3 (i.e., the parameter with the smallest population value) but are less evident for the other parameters and, as the sample size increases, estimation methods perform similar to each other.

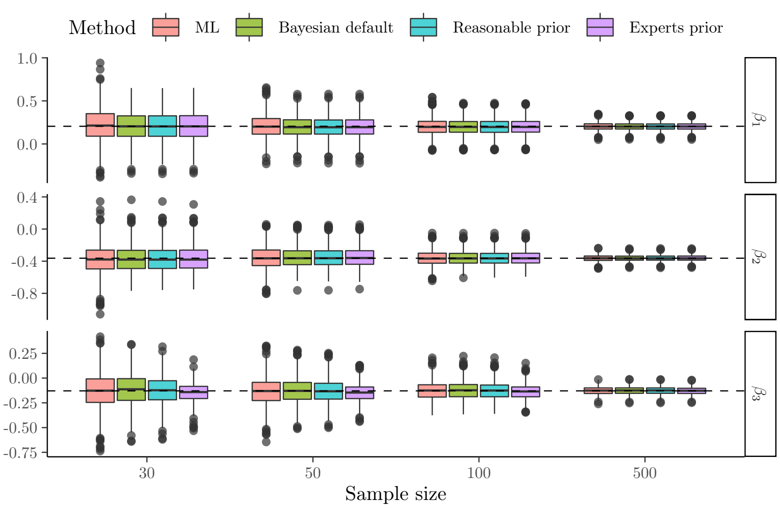

Fig. 4 Estimates distribution.

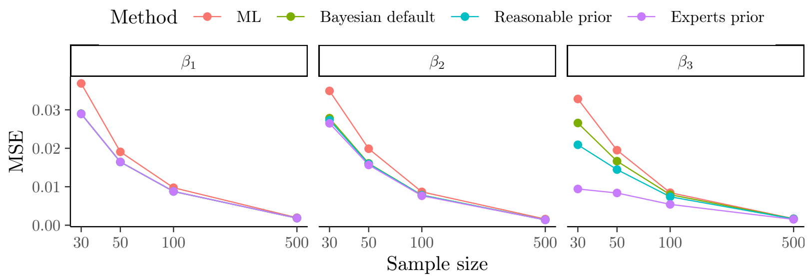

Fig. 5 MSE values.

Considering the MSE (see Fig.5), we have the same results pattern. Difference between methods are bigger in the case of very small samples ($n = 30$), with ML estimation presenting higher MSE values and Bayesian approach with experts prior performing better. However, differences between prior specification are noticeable only for the parameter β_3, whereas in the other cases all types of priors perform similarly.

Finally, the result of coverage and power are presented in Fig. 6. Coverage reaches adequate levels in all condition with sample size equal or greater of 100. With smaller sample sizes, Bayesian approach with experts prior showed excessive coverage in the case of the parameter β_3. Power is extremely low when sample sizes are small. ML estimation performs slightly better in terms of power across all conditions, except for the parameter β_3 where is outperformed by Bayesian approach with experts prior. However, adequate levels of power are reached for all parameters only with large sample sizes ($n = 500$).

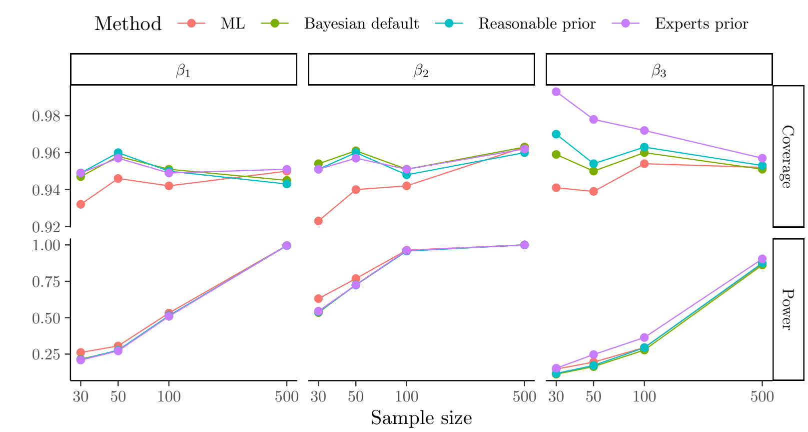

Fig. 6 Coverage and power values.

4 Discussion and conclusions

Overall, as expected , results indicate that the formalisation of prior according to expert judgement is particularly useful in the case of limited sample sizes. When data are limited

However it would be important to understand the different behaviour of the different parameter.

Limit on specification of prior how much is informative.

Elicitation does not inform only about the prior but the collaboration with experts in the field would help also in defining more reasonable models along all the research process. (incentivare collaborazione)

Appendix

Table 1 Summary of the simulation results for the parameter β_1

Sample size	Estimation method	Rel. mean bias	Rel. median bias	MSE	Coverage	Power
30	Maximum likelihood	0.066	0.040	0.037	0.932	0.261
30	Bayes default prior	-0.011	-0.005	0.029	0.947	0.215
30	Bayes reasonable prior	-0.010	-0.009	0.029	0.949	0.211
30	Bayes experts prior	-0.011	-0.005	0.029	0.949	0.209
50	Maximum likelihood	-0.015	-0.022	0.019	0.946	0.306
50	Bayes default prior	-0.053	-0.053	0.016	0.958	0.274
50	Bayes reasonable prior	-0.053	-0.056	0.016	0.960	0.277
50	Bayes experts prior	-0.054	-0.052	0.016	0.957	0.272
100	Maximum likelihood	-0.023	-0.038	0.010	0.942	0.532
100	Bayes default prior	-0.044	-0.038	0.009	0.951	0.512
100	Bayes reasonable prior	-0.044	-0.044	0.009	0.950	0.514
100	Bayes experts prior	-0.044	-0.040	0.009	0.949	0.509
500	Maximum likelihood	-0.012	-0.012	0.002	0.950	0.996
500	Bayes default prior	-0.016	-0.016	0.002	0.945	0.996
500	Bayes reasonable prior	-0.016	-0.012	0.002	0.943	0.995
500	Bayes experts prior	-0.016	-0.012	0.002	0.951	0.996

Table 2 Summary of the simulation results for the parameter β_2

Sample size	Estimation method	Rel. mean bias	Rel. median bias	MSE	Coverage	Power
30	Maximum likelihood	0.038	0.043	0.035	0.923	0.631
30	Bayes default prior	0.014	0.047	0.028	0.954	0.535
30	Bayes reasonable prior	0.013	0.047	0.027	0.951	0.540
30	Bayes experts prior	0.005	0.046	0.026	0.951	0.545
50	Maximum likelihood	-0.006	0.004	0.020	0.940	0.769
50	Bayes default prior	-0.030	0.002	0.016	0.961	0.727
50	Bayes reasonable prior	-0.031	0.000	0.016	0.960	0.724
50	Bayes experts prior	-0.037	-0.008	0.016	0.957	0.725
100	Maximum likelihood	-0.003	0.012	0.009	0.942	0.964
100	Bayes default prior	-0.004	0.007	0.008	0.951	0.959
100	Bayes reasonable prior	-0.005	0.007	0.008	0.948	0.958
100	Bayes experts prior	-0.010	0.003	0.008	0.951	0.962
500	Maximum likelihood	0.000	0.001	0.002	0.963	1.000
500	Bayes default prior	-0.002	0.004	0.001	0.963	1.000
500	Bayes reasonable prior	-0.002	0.005	0.001	0.960	1.000
500	Bayes experts prior	-0.004	0.004	0.001	0.962	1.000

Table 3 Summary of the simulation results for the parameter β_3

Sample size	Estimation method	Rel. mean bias	Rel. median bias	MSE	Coverage	Power
30	Maximum likelihood	-0.034	-0.015	0.033	0.941	0.147
30	Bayes default prior	-0.108	-0.124	0.027	0.959	0.112
30	Bayes reasonable prior	-0.037	-0.050	0.021	0.970	0.117
30	Bayes experts prior	0.148	0.096	0.009	0.993	0.153
50	Maximum likelihood	0.030	0.014	0.019	0.939	0.195
50	Bayes default prior	-0.025	-0.001	0.017	0.950	0.164
50	Bayes reasonable prior	0.013	0.034	0.014	0.954	0.172
50	Bayes experts prior	0.140	0.132	0.008	0.978	0.247
100	Maximum likelihood	-0.009	-0.028	0.008	0.954	0.295
100	Bayes default prior	-0.030	-0.049	0.008	0.960	0.278
100	Bayes reasonable prior	-0.012	-0.025	0.007	0.963	0.294
100	Bayes experts prior	0.070	0.049	0.005	0.972	0.364
500	Maximum likelihood	-0.016	-0.025	0.002	0.952	0.871
500	Bayes default prior	-0.021	-0.033	0.002	0.951	0.862
500	Bayes reasonable prior	-0.017	-0.034	0.002	0.953	0.875
500	Bayes experts prior	0.003	-0.011	0.002	0.957	0.904

References

1. Cooke, R., et al.: Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press on Demand, Oxford, UK (1991)
2. Ioannidis, J.P.A.: Why Most Published Research Findings Are False. PLOS Medicine **2**(8), e124 (2005). DOI 10.1371/journal.pmed.0020124
3. McNeish, D.: On Using Bayesian Methods to Address Small Sample Problems. Structural Equation Modeling: A Multidisciplinary Journal **23**(5), 750–773 (2016). DOI 10.1080/10705511.2016.1186549
4. Merkle, E.C., Rosseel, Y.: *Blavaan* : Bayesian Structural Equation Models via Parameter Expansion. Journal of Statistical Software **85**(4) (2018). DOI 10.18637/jss.v085.i04
5. Oakley, J.E., O'Hagan, A.: SHELF: The Sheffield Elicitation Framework. School of Mathematics and Statistics, University of Sheffield (2016)
6. O'Hagan, A. (ed.): Uncertain Judgements: Eliciting Experts' Probabilities. Statistics in Practice. John Wiley & Sons, London ; Hoboken, NJ (2006)
7. O'Hagan, A.: Expert Knowledge Elicitation: Subjective but Scientific. The American Statistician **73**(sup1), 69–81 (2019). DOI 10.1080/00031305.2018.1518265
8. Open Science Collaboration: Estimating the reproducibility of psychological science. Science **349**(6251), aac4716–aac4716 (2015). DOI 10.1126/science.aac4716
9. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing (2018). Https://www.R-project.org/
10. Rosseel, Y.: *Lavaan* : An *R* Package for Structural Equation Modeling. Journal of Statistical Software **48**(2) (2012). DOI 10.18637/jss.v048.i02
11. Rowe, G., Wright, G.: The Delphi technique as a forecasting tool: Issues and analysis. International journal of forecasting **15**(4), 353–375 (1999)
12. Sella, E., Carbone, E., Toffalini, E., Borella, E.: Personality traits and sleep quality: The role of sleep-related beliefs. Personality and Individual Differences **156**, 109,770 (2020). DOI 10.1016/j.paid.2019.109770
13. Smid, S.C., McNeish, D., Miočević, M., van de Schoot, R.: Bayesian Versus Frequentist Estimation for Structural Equation Models in Small Sample Contexts: A Systematic Review. Structural Equation Modeling: A Multidisciplinary Journal **27**(1), 131–161 (2020). DOI 10.1080/10705511.2019.1577140
14. Smid, S.C., Rosseel, Y.: Sem with Small Samples. In: R. van de Schoot, M. Miočević (eds.) Small Sample Size Solutions, 1 edn., pp. 239–254. Routledge (2020). DOI 10.4324/9780429273872-20
15. Wolf, E.J., Harrington, K.M., Clark, S.L., Miller, M.W.: Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. Educational and Psychological Measurement **73**(6), 913–934 (2013). DOI 10.1177/0013164413495237