Factor Models

by Hedibert Lopes
hedibert@im.ufrj.br

Factor models is certainly one of the most used (useful) statistical techniques. Factor models are mainly applied in two major situations: (i) data reduction and (ii) identifying underlying structures.

I would like to start this section by quoting Bartholomew (1995) (Spearman and the origin and development of factor analysis, *British Journal of Mathematical and Statistical Psychology, 48, 211-220*), who starts his paper's abstract by saying that

> Spearman [Charles Edward Spearman F.R.S. 1863-1945] invented factor analysis but his almost exclusive concern with the notion of a general factor prevented him from realizing its full potential.

Fortunately, factor models potentials have been discovered and are still being discovered, even after almost a century has passed since Spearman wrote his seminal paper ('General Inteligente' objectively determined and measured, *American Journal of Psychology*, 5, 201-293, 1904.)

I organized this annotated bibliography with the idea of providing the reader with a modest (and subjective) set of papers and books that would lead him/her to the realm of (latent) factor models.

Factor analysis: estimation

1. Lawley (1940) The estimation of factor loadings by the method of maximum likelihood, Proceedings of the Royal Society of Edinburgh, 60, 64-82. Lawley (1941) Further investigations in factor estimation, Proceedings of the Royal Society of Edinburgh, 61, 176-185. Introduces maximum likelihood factor model.

2. Jöreskog, K.G. (1967) Some contributions to maximum likelihood factor analysis, Psychometrika, 32, 443-382. Jöreskog, K.G. (1969) A general approach to confirmatory maximum likelihood factor analysis, Psychometrika, 34, 183-220. Maximum likelihood estimation is made feasible.

3. Martin, J.K. and McDonald, R.P. (1975) Bayesian estimation in unrestricted factor analysis: a treatment for Heywood cases, Psychometrika, 40, 505-517. Treatment of the Heywood case (zero variances) by proper specification of the prior distributions.

4. Geweke, J.F. and Singleton, K.J. (1980) Interpreting the likelihood ratio statistic in factor models when sample size is small, Journal of the American Statistical Association, 75, 133-137.
From Monte Carlo simulations and under certain regularity conditions, asymptotic theory is appropriate when sample size is greater than 30. The same is not true when the regularity conditions fail.

5. Bartholomew, D.J. (1981) Posterior analysis of the factor model, British Journal of Mathematical and Statistical Psychology, 34, 93-99. The posterior analysis is restricted to the common factors upon previous estimation of the model parameters (loadings and idiosyncrasies).

6. Lee, S-Y (1981) A Bayesian approach to confirmatory factor analysis, Psychometrika, 46, 153-160. A Newton-Raphson algorithm is implemented to find the posterior mode for four different prior specifications.

7. Rubin, D.B. and Thayer, D.T. (1982) EM algorithms for ML factor analysis, Psychometrika, 47, 69-76. Bentler, P.M. and Tanaka, J.S. (1983) Problems with EM algorithms for ML factor analysis, Psychometrika, 48, 247-251. Rubin, D.B. and Thayer D.T. (1983) More on EM for factor analysis, Psychometrika, 48, 253-257. The EM algorithm is introduced as an alternative optimization algorithm to Jöreskog's (1967,1969) maximum likelihood scheme.

8. Bartholomew, D.J. (1984) The foundations of factor analysis, Biometrika, 71, 221-232.

9. Bartholomew, D.J. (1985) Foundations of factor analysis: some practical implications, British Journal of Mathematical and Statistical Psychology, 38, 1-10. With discussion on the British Journal of Mathematical and Statistical Psychology, 38, 127-229.

10. Anderson, T.W. and Amemiya, Y. (1988) The asymptotic normal distribution of estimators in factor analysis under general conditions. The Annals of Statistics, 16, 759-771. Amemiya, Y. and Anderson, T.W. (1990) Asymptotic chi-square tests for a large class of factor analysis models. The Annals of Statistics, 18, 1453-1463. Asymptotic estimation and hypothesis testing in factor analysis models.

11. Press, S.J. and Shigemasu, K. (1989) Bayesian inference in factor analysis, in Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin, L.J. Gleser, M.D. Perlman, S.J.Press, A.R.Sampson (Eds.), New York: Springer-Verlag, 271-287. Posterior large sample interval estimators of common factors, factor loading and idiosyncratic variances. Three step procedured: (1) estimation of the common factors, (2) estimation of the factor loadings given the common factors estimate, and (3) estimation of the idiosyncratic variances given both the common factors and factor loadings estimates.

12. Bartholomew, D.J. (1995) Spearman and the origin and development of factor analysis, British Journal of Mathematical and Statistical Psychology, 48, 211-220. Historical account of the development of factor models.

13. Ihara, M. and Kano, Y. (1995) Identifiabiliy of full, marginal, and conditional factor analysis models, Statistics and Probability Letters, 23, 343-350. Conditions for full, marginal and conditional model identification are discussed.

14. Schneeweiss, H. and Mathes, H. (1995) Factor analysis and principal components, Journal of multivariate analysis, 55, 105-124. Similarities and differences between the factor analysis and principal component analysis are discussed.

15. Yung, Y.-F. (1997) Finite mixtures in confirmatory factor-analysis models. Psychometrika, 62, 297-330. Finite mixture of factor models for handling heterogeneity. Approximate-Scoring (AS) and Expectation-Maximization (EM) methods are developed.

16. Lee, S.E. and Press, S.J. (1998) Robustness of Bayesian factor5analysis estimates, Communications in Statistics, Theory and Methods, 27, 1871-1893. Posterior robustness of the loadings, common factors and idiosyncratic covariance.

17. Fokoué, E. and Titterington, D.M. (2000) Bayesian sampling for mixtures of factor analysers, Technical Report, Department of Statistics, University of Glasgow. Gibbs sampler in mixture of factor models where the number of components and common factors are fixed and known.

18. Lopes, H.F. (2003) Expected posterior priors in factor analysis. (Brazilian Journal of Probability and Statistics, to appear), Technical report, Department of Statistical Methods, Federal University of Rio de Janeiro.

Factor analysis: model selection

19. Akaike, H. (1987) Factor analysis and AIC, Psychometrika, 52, 317-332. Model selection through the Akaike Information Criterion.

20. Press, S.J. and Shigemasu, K. (1994) Posterior distribution for the number of factors, in American Satistical Association Proceedings of the Section on Bayesian Statistical Science, 75-77. Large sample is assumed to approximate the factor model's predictive density, which is used for model comparison.

21. Polasek, W. (1997) Factor analysis and outliers: A Bayesian Approach. Discussion paper, University of Basel. The number of factors is determined by computing marginal likelihoods through Chib's (1995,JASA, 1313-1321) algorithm.

22. Bozdogan, H. and Shigemasu, K. (1998) Bayesian factor analysis model and choosing the number of factors using a new informational complexity criterion. Technical report, Department of Statistics, University of Tennessee.

23. Fokoué, E. and Titterington, D.M. (2000) Stochastic model selection for Bayesian mixtures of factor analysers, Technical Report, Department of Statistics, University of Glasgow. A birth-death marked Markov point process in continuous time is used as a stochastic model selection algorithm for Bayesian mixture of factor models.

24. Lopes, H.F. and West, M. (2003) Model assessment in factor analysis. Statistica Sinica (to appear) ISDS-Discussion Paper 98-38. A reversible jump Markov chain Monte Carlo (RJMCMC) algorithm is developed to fully account the uncertainty on the number of common factors. Comparisons are made to several additional algorithms that approximate the predictive density.

25. West, M. (2002) Bayesian Factor Regression Models in the "Large p, Small n" Paradigm. Discussion paper #02-12, ISDS, Duke University. Factor models where the number of variables is extremely larger than the number of observations, a situation commonly present in studies of gene expression.

Factor analysis in time series

26. Peña, D. and Box, G.E.P. (1987) Identifying a simplifying structure in time series. Journal of the American Statistical Association, 82, 836-843. Factor models with common (independent/dependent) factors following ARMA processes.

27. Engle, R. (1987) Multivariate ARCH with factor structures – cointegration in variance, University of California, San Diego, Dept. of Economics Discussion Paper 87-27. One of the first papers to apply common factors to model covariances in time series.

28. Diebold, F.X. and Nerlove, M. (1989) The dynamics of exchange rate volatility: a multivariate latent ARCH model, Journal of Applied Econometrics, 4, 1-21. Multivariate GARCH structures through one latent factor.

30. Engle, R.F., NG, V.K. and Rothschild, M. (1990) Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills, Journal of Econometrics, 45, 213-238. Factor-ARCH to model conditional covariance matrix of asset returns.

31. Ng, V., Engle, R.F. and Rothschild, M. (1992) A multi-dynamic factor model for stock returns. Journal of Econometrics, 52, 245-266. Relates dynamic and static factors to porfolio allocation in financial markets.

32. Lin, W-L. (1992) Alternative estimators for factor GARCH models - a Monte Carlo comparison, Journal of Applied Econometrics, 7, 259-279. Compares four frequentist estimators for factor GARCH models: two-stage univariate GARCH (2SUE), two-stage quasi-maximum likelihoo (2SML), quasi-maximum likelihood with known factor weights (RMLE) and quasi-maximum likelihood with unknown factor weights (MLE).

33. Molenaar, P.C.M. and Gooijer, J.G.D. and Schmitz, B. (1992) Dynamic factor analysis of nonstationary multivariate time series. Factor models with lagged common factors to account for the persistance in time series trends.

34. Bollerslev, T. and Engle, R.F. (1993) Common persistence in conditional variances, Econometrica, 61, 167-186. K-factor generalized autoregressive conditional heteroscedasticity (GARCH) models are discussed and conditions are given for covariance stationarity. They also study co-persistence in multivariate integrated GARCH models.

35. Harvey, A., Ruiz, E. and Shephard, N. (1994) Multivariate stochastic variance models, Review of Economic Studies, 61, 247-264. Common factors, as multivariate random walk, are used to model persistent movements in stochastic volatility models.

36. Escribano, A. and Peña, D. (1994) Cointegration and common factors. Journal of time series analysis, 15, 577-586. Cointegrated vectors are viewed as Peña and Box's (1987) dynamic factor models.

37. Geweke, J. and Zhou, G. (1996) Measuring the pricing error of the arbitrage pricing theory. The review of financial studies, 9, 557-587. First paper to implement the Gibbs sampler for exact Bayesian inference in (static) factor models.

38. Demos, A. and Sentana, E. (1998) An EM algorithm for conditionall heterescedastic factor models, 16, 357-361. Application of the EM algorithm to factor models with dynamic heterocedasticity in the common factors.

39. Sentana, E. (1998) The relation between conditionally heteroskedastic factor models and factor GARCH models, Econometrics Journal, 1, 1-9. Investigation of the similarities and differences of Engle's (1987) factor GARCH model and Diebold and Nerlove's (1989) latent factor ARCH model.

40. Pitt, M.K. and Shephard, N. (1999) Time varying covariances: A factor stochastic volatility approach (with discussion). In Bayesian statistics 6, Ed. Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., 547–570. London: Oxford University Press. Univariate stochastic volatility structures is used to model the common factor variances through time. Sequential portfolio allocation is made possible by particle filters.

41. Aguilar, O. and West, M. (2000) Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. Journal of Business and Economic Statistics, 18, 338–357. Multivariate stochastic volatility structure is used to model the common factor variances through time.

42. Chib, S., Nardari, F. and Shephard, N (2003) Analysis of high dimensional multivariate stochastic volatility models. Technical Report, Nuffield College, University of Oxford. Feasible Bayesian inference (through MCMC algorithms) for multivarite stochastic volatility models in highly dimensional settings.
43. Vrontos, I.D., Dellaportas, P. and Politis, D.N. (2002) A full-factor multivariate GARCH model (2003). Classical and Bayesian estimation of a variant of the multivariate GARCH model with as many factors as variables. The order of variable problem is dealt with Bayesian model averaging.
44. Fiorentini, G., Sentana, E. and Shephard, N. (2003) Likelihood-based estimation of latent generalised ARCH structures. Technical Report, Nuffield College, University of Oxford. Fast MCMC and simulated EM algorithms for latent factor GARCH model.

Books

42. Lawley, D.N. an Maxwell, A.E. (1971) Factor analysis as a statistical method (2nd edition). London: Butterworths.

43. Anderson, T.W. (1984) An introduction to multivariate statistical analysis (2nd edition). New York: John Wiley & Sons, Inc.

44. Krzanowski, W.J. and Marriott, F.H.C. (1995) Multivariate analysis - Part 2: classification, covariance structures and repeated measurements. Kendall's Library of Statistics 2. London: Arnold.

45. Bartholomew, D.J. and Knott, M. (1999) Latent variable models and factor analysis (2nd edition). Kendall's Library of Statistics 7. London: Arnold.

46. Basilevsky, A. (19??) Statistical factor analysis and related methods: theory and applications. New York: John Wiley & Sons, Inc.

47. Press, S.J. (1982) Applied multivariate analysis: using Bayesian and frequentist methods of inference (2nd edition). New York: Krieger

48. Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) Multivariate Analysis. London: Academic Press.

STUDENT'S CORNER

by Lilla Di Scala and Luca La Rocca
lilla@dimat.unipv.it luca@dimat.unipv.it

Dear Members of ISBA, we would like to use this Student's Corner as an opportunity to send out a request to all of you. As some may already be aware of, we are running a survey on the following topic:

"Which groundbreaking/essential papers do you believe that a graduate student with a serious interest in Bayesian statistics should not miss reading?"

In other words, if you had to decide on next years' choices for a Journal Club, which Bayesian papers would you include? We would very much appreciate if you could spare five minutes of your time and make a list of 3 or 4 such papers, whether methodological or computational, whether your own or the product of other distinguished authors. Please keep in mind that it would be very useful for students to have such a hit-list of papers which are considered a must from those within the field. We welcome all replies and hope these will be copious, looking forward to featuring the results in the next issues of the Bulletin. Finally, a note of thanks goes to all of you who have already responded!

➤ Phd Thesis

Let us now introduce this issue's two doctorate thesis abstracts.

The first one is from the Department of Mathematics at the University of Pavia, Italy. The PhD in Mathematics and Statistics (which actually foresees two distinct curricula) is a three-year program with an extensive course offering during its first half. The Mathematical Statistics graduate program is relatively new and is currently going into its seventh year. On the other hand, the University of Pavia is one of the oldest places of study in Italy. The town very much resembles an over-seas campus as the university and its population of students are the center of most of the town's activity.

Igor Pruenster - igor@eco.unipv.it
Department of Political Economics and Quantitative Methods University of Pavia, Italy
Thesis title: *Random probability measures derived from increasing additive processes and their application to Bayesian statistics*
Advisor: Eugenio Regazzini.