

# FACTOR MODELS: AN ANNOTATED BIBLIOGRAPHY

by Hedibert Lopes

A version of this annotated bibliography appeared in the August 2003 Issue of the ISBA Bulletin, The official bulletin of the International Society for Bayesian Analysis

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.
2. Lawley (1941) Further investigations in factor estimation, *Proceedings of the Royal Society of Edinburgh*, 61, 176-185. Introduces maximum likelihood factor model.
3. Anderson, T.W. (1963) The Use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, 28, 1-25.
4. Jöreskog, K.G. (1967) Some contributions to maximum likelihood factor analysis, *Psychometrika*, 32, 443-382.



5. Jöreskog, K.G. (1969) A general approach to confirmatory maximum likelihood factor analysis, *Psychometrika*, 34, 183-220. Maximum likelihood estimation is made feasible.
6. 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.
7. 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.
8. 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).
9. 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.
10. Rubin, D.B. and Thayer, D.T. (1982) EM algorithms for ML factor analysis, *Psychometrika*, 47, 69-76.
11. Bentler, P.M. and Tanaka, J.S. (1983) Problems with EM algorithms for ML factor analysis, *Psychometrika*, 48, 247-251.
12. 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.
13. Bartholomew, D.J. (1984) The foundations of factor analysis, *Biometrika*, 71, 221-232.
14. 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.
15. Amemiya, Y., Fuller, W.A. and Pantula, S.G. (1987) The asymptotic distributions of some estimator sfor a factor analysis model. *Journal of Multivariate Analysis*, 22, 51-64.
16. 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.
17. 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 procedure: (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.

18. 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.
19. 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.
20. Ihara, M. and Kano, Y. (1995) Identifiability 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.
21. 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.
22. 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.
23. Lee, S.E. and Press, S.J. (1998) Robustness of Bayesian factor analysis estimates, *Communications in Statistics, Theory and Methods*, 27, 1871-1893. Posterior robustness of the loadings, common factors and idiosyncratic covariance.
24. Fokoué, E. and Titterton, 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.
25. Ghahramani, Z. and Beal, M. (2000) Variational inference for Bayesian mixture of factor analysers. In S.A. Solla, T.K. Leen and K.-R. Müller (Eds.) *Advanced in neural information processing systems 12*. Cambridge, MA: MIT Press.
26. Utsugi, A. and Kumagai, T. (2001) Bayesian analysis of mixture of factor analysers. *Neural Computation*, 13, 993-1002.
27. Bai, J. (2003) Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71, 135–171.



## Factor analysis: model selection

1. Akaike, H. (1987) Factor analysis and AIC, *Psychometrika*, 52, 317-332. Model selection through the Akaike Information Criterion.
2. Press, S.J. and Shigemasu, K. (1994) Posterior distribution for the number of factors, in *American Statistical 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.
3. Connor, G. and Korajczyk, R.A. (1996) A Test for the Number of Factors in an Approximate Factor Model. *The Journal of Finance*, 47, 1263–1291.
4. 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.
5. 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.
6. Bai, J. and Ng, S. (2002) Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70, 191-221.
7. Fokoué, E. and Titterington, D.M. (2003) Mixtures of Factor Analysers. *Bayesian Estimation and Inference by Stochastic Simulation. Machine Learning*, 50, 73-94. A birth-death marked Markov point process in continuous time is used as a stochastic model selection algorithm for Bayesian mixture of factor models.
8. West, M. (2003) Bayesian factor regression models in the "Large p, Small n" paradigm *Bayesian Statistics*, 7, 723-732. Bernardo, J.M, Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M and West, M. (Eds.). Oxford University Press. Factor models where the number of variables is extremely larger than the number of observations, a situation commonly present in studies of gene expression.
9. Lopes, H.F. (2003) Expected posterior priors in factor analysis., *Brazilian Journal of Probability and Statistics*, 17, 91-105.
10. Lopes, H.F. and West, M. (2004) Bayesian Model assessment in factor analysis. *Statistica Sinica*, 14, 41-67. 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.



## Spatial Factor models

1. Christensen, W.F. and Amemiya, Y. (2001) Generalized Shifted-Factor Analysis Method for Multivariate Geo-referenced data. *Mathematical Geology*, 33,801-824.
2. Wang, F. and Wall, M.M. (2001) Modeling multivariate data with a common spatial factor.
3. Christensen, W.F. and Amemiya, Y. (2002) Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association*, 97, 302-317.
4. Christensen, W.F. and Amemiya, Y. (2003) Modeling and prediction for multivariate spatial factor analysis. *Journal of Statistical Planning and Inference*, 115, 543-464.
5. Hogan, J.W. and Tchernis, R. (2003) Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data.
6. Wang, F. and Wall, M.M. (2003) Generalized common spatial factor model. *Biostatistics*, 4, 569-582.
7. Calder, C.A. (2004) Efficient posterior inference and prediction of space-time processes using dynamic process convolutions. Department of Statistics Preprint No. 734, The Ohio State University.

## Factor analysis in finance and econometrics

1. Ross, S. (1976) The Arbitrage Theory of Capital Asset Pricing. *Journal of Finance*, 13, 341-360.
2. Geweke, J. (1977) The dynamic factor analysis of economic time series. In: Aigner, D.J., Goldberger, A.S. (Eds.), *Latent Variables in Socio-Economic Models*. North-Holland, Amsterdam.
3. Engle, R.F. and Watson, M.F. (1981) A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76, 774-781.
4. Chamberlain, G., Rothschild, M. (1983) Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51, 1281-1304.
5. Chamberlain, G. (1983) Funds, factors, and diversification in arbitrage pricing models. *Econometrica*, 51, 1305-1324.
6. Molenaar, P.E.C. (1985) A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50, 181-202.
7. 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.



8. 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.
9. 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.
10. 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.
11. 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 portfolio allocation in financial markets.
12. 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 likelihood (2SML), quasi-maximum likelihood with known factor weights (RMLE) and quasi-maximum likelihood with unknown factor weights (MLE).
13. Molenaar, P.E.C. 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 persistence in time series trends.
14. 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.
15. 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.
16. 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.
17. Conner, G. (1995) The Three Types of Factor Models: A comparison of their explanatory power. *Financial Analysts Journal*, May-June, 42-46.
18. 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.[0.5ex]



19. Forni, M. and Reichlin, L. (1996) Dynamic common factors in large cross-sections. *Empirical Economics*, 21, 27-42.
20. Forni, M. and Reichlin, L. (1998) Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics, 65, 453-473.
21. Demos, A. and Sentana, E. (1998) An EM algorithm for conditional heteroscedastic factor models, 16, 357-361. Application of the EM algorithm to factor models with dynamic heteroscedasticity in the common factors.
22. 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.
23. 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.
24. Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000) The Generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics*, 82, 540-554.
25. 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.
26. Lopes, H.F., Aguilar, O. and West, M. (2000) Time-varying Covariance Structures in Currency Markets. In *Annals of the XXII Brazilian Meeting of Econometrics*, Campinas, Brazil, December 2000.
27. Fiorentini, G., and Sentana, E. (2001) Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models. *Journal of Econometrics*, 102, 143-164.
28. Van Der Weide, R. (2002) GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17, 549-564.
29. Han, Y. (2002) The Economic value of volatility modeling: Asset allocation with a high dimensional dynamic latent factor multivariate stochastic volatility model.
30. Stock, J.H. and Watson, M.W. (2002) Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 20, 147-162.
31. Stock, J.H. and Watson, M.W. (2002) Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 97, 1167-1179.



32. Lopes, H.F. and Migon, H. (2002) Comovements and contagion in emergent markets: stock indexes volatilities. *Case Studies in Bayesian Statistics, Volume VI*, 285-300. Springer-Verlag.
33. 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 multivariate stochastic volatility models in highly dimensional settings.
34. Vrontos, I.D., Dellaportas, P. and Politis, D.N. (2003) A full-factor multivariate GARCH model. *Econometrics Journal*, 6, 312-334. 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.
35. Fiorentini, G., Sentana, E. and Shephard, N. (2003) Likelihood-based estimation of latent generalised ARCH structures. *Econometrica* (forthcoming). Fast MCMC and simulated EM algorithms for latent factor GARCH model.
36. Nardari, F. and Scruggs, J.T. (2003) Analysis of linear factor models with multivariate stochastic volatility for stock and bond returns.
37. Quintana, J., Lourdes, V., Aguilar, O. and Liu, J. (2003) Global Gambling, *Bayesian Statistics 7*. Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith, A.F.M and West, M. (Eds.). Oxford University Press.
38. Connor, G., Korajczyk, R.A. and Linton, O. (2003) The Common and Specific Components of Dynamic Volatility.
39. Watanabe (2003) Measuring Business Cycle Turning Points in Japan with a Dynamic Markov Switching Factor model. *Monetary and Economic Studies*, 35-68.
40. Sentana, E. (2004) Factor representing portfolios in large asset markets. *Journal of Econometrics*, 119, 257-290.
41. Doz, C. and Renault, E. (2004) Conditionally Heteroskedastic Factor Models: Identification and Instrumental Variables Estimation. Manuscript, Université de Montréal.
42. Dovonon, P., Doz, C. and Renault, E. (2004) Conditionally Heteroskedastic Factor Models with Skewness and Leverage Effects. Manuscript, Université de Montréal.
43. Peña, D. and Poncela, P. (2004) Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, 119, 291-321.
44. Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004) The Generalized dynamic factor model consistency and rates. *Journal of Econometrics*, 119, 231-255.
45. Kapetanios, G. and Marcellino, M. (2004) A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions. Technical Report, Queen Mary University of London.



## Books

1. Lawley, D.N. and Maxwell, A.E. (1971) Factor analysis as a statistical method (2nd edition). London: Butterworths.
2. Anderson, T.W. (1984) An introduction to multivariate statistical analysis (2nd edition). New York: John Wiley & Sons, Inc.
3. 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.
4. Bartholomew, D.J. and Knott, M. (1999) Latent variable models and factor analysis (2nd edition). Kendall's Library of Statistics 7. London: Arnold.
5. Basilevsky, A. (19??) Statistical factor analysis and related methods: theory and applications. New York: John Wiley & Sons, Inc.
6. Press, S.J. (1982) Applied multivariate analysis: using Bayesian and frequentist methods of inference (2nd edition). New York: Krieger
7. Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) Multivariate Analysis. London: Academic Press.
8. Rowe, D.B. (2002) Multivariate Bayesian statistics: models for source separation and signal unmixing. Boca Raton: CRC Press.
9. Press, S.J. (2003) Subjective and Objective Bayesian Statistics: Principles, Models and Applications, 2nd Edition. Wiley Series in Probability and Statistics (Chapter 15).

