TI 2024-068/III
Tinbergen Institute Discussion Paper

Time-Varying Factor Model Components for Effective Momentum Strategy

Jamie Cross[1]
Lennart Hoogerheide[2]
Herman van Dijk[3]

1 Melbourne Business School

2 Vrije Universiteit Amsterdam

3 Erasmus School of Economics, Erasmus University Rotterdam, Tinbergen Institute

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at https://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Time-Varying Factor Model Components for Effective Momentum Strategy*

Jamie L. Cross[1,2], Lennart Hoogerheide[3,4], and Herman K. van Dijk[4,5]

[1] *Melbourne Business School, University of Melbourne*

[2] *BI Norwegian Business School*

[3] *Department of Econometrics, VU University Amsterdam*

[4] *Tinbergen Institute*

[5] *Erasmus University Rotterdam and Norges Bank*

October 28, 2024

Abstract

Determining a plausible number of components in a factor model is a nontrivial issue in case of weak data, sparse model restrictions and diffuse prior information. We discuss the issue of structural parametric identification in a static factor model and introduce orthogonal restrictions which imply that inference is independent of the order of the dependent variables. Given that financial and economic relations vary over time, we propose the use of predictive likelihoods in combination with moving window estimation in order to determine a plausible time-varying number of factor model components. Results are presented on a residual momentum strategy based on a time-varying latent factor model which outperforms a standard momentum strategy using a portfolio of US industrial stocks.

*This paper is an invited comment on Frühwirth-Schnatter et al. (2024), Sparse Bayesian Factor Analysis When the Number of Factors Is Unknown. It should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

1 Introduction

In their interesting and mathematically elegant paper Frühwirth-Schnatter et al. (2024) discuss the issue of determining a plausible number of latent components in a sparse factor model (technically: determining the rank of the factor space) which is a nontrivial issue in case of weak data, sparse model restrictions and diffuse prior information. In this context the authors focus on the connection between the theoretical issue of parametric identification restrictions including informative prior information and the operational issue of Bayesian Markov chain Monte Carlo estimation. Specifically, the authors achieve identification and inference which is independent from the ordering of the dependent variables by making use of the concept of Unordered Generalized Lower Triangular (UGLT) structure and for estimation they introduce a novel Markov chain Monte Carlo procedure which makes use of a reversible jump sampler. All this in order to learn about a plausible number of latent factors with substantial posterior probability.

We introduce two contributions to this research. We start to discuss the issue of identification restrictions within the authors' framework of a static factor model and present an operational alternative to the UGLT structure by introducing *orthogonal* parameter restrictions. Second, we propose the use of predictive likelihoods in combination with moving window estimation in order to determine a plausible time-varying number of factor model components. Our motivation stems from the observation that financial and economic relations vary over time. One of these time-varying relations is the increase in the correlations between equities during market downturns, that is, during equity market downturns fewer latent factors are assumed to be able to explain the same amount of variation in equity returns. We present empirical results on how a residual momentum strategy based on a time-varying latent factor model outperforms a standard momentum strategy using a portfolio of industrial stocks. This strategy has been popular among investors over a long time.

2 Identification restrictions in factor models

For expository purposes we start with a basic multivariate regression model:

$$\mathbf{y_t'} = \mathbf{x_t'}\mathbf{B}' + \epsilon_\mathbf{t}', \quad \epsilon_t \sim \mathcal{NID}(\mathbf{0}, \mathbf{\Sigma}), \tag{1}$$

where $\mathbf{y_t}$ is an m-vector, $\mathbf{x_t}$ is an r-vector and $\mathbf{B}$ an $m \times r$ matrix. It is well-known that a Bayesian analysis of this model using diffuse priors leads to a marginal posterior of $\mathbf{B}$ that is bell-shaped and belongs to the class of matrix Student-t distributions. Determining a plausible number of explanatory variables is a standard topic in an introductory Bayesian course. The connection between model structure and estimation is direct: analytical as well as simulation methods are used.

Next, consider a static factor model and adjust formula (1.1) of Frühwirth-Schnatter et al. (2024) as:

$$\mathbf{y_t'} = \mathbf{f_t'}\mathbf{\Lambda} + \epsilon_\mathbf{t}', \quad \mathbf{f_t} \sim \mathcal{NID}(\mathbf{0}, \mathbf{I_r}), \quad \epsilon_\mathbf{t} \sim \mathcal{NID}(\mathbf{0}, \mathbf{\Sigma_0}), \quad \mathbf{\Sigma_0} = \text{Diag}(\sigma_1^2, \ldots, \sigma_m^2), \tag{2}$$

where $\mathbf{f_t}$ is an r-vector and $\mathbf{\Lambda}$ an $r \times m$ matrix. The diagonal covariance matrix assumption with respect to the disturbances $\epsilon_\mathbf{t}$ implies that all cross-sectional correlation is captured by the factors $\mathbf{f_t}$, in addition, $\text{cor}(\mathbf{f_t}, \epsilon_\mathbf{s}) = 0$ for $\forall s, t$. The vector of observations $\mathbf{x_t}$ is replaced by a vector of *unobserved* random factors $\mathbf{f_t}$ and the matrix of coefficients $\mathbf{B}'$ by a matrix $\mathbf{\Lambda}$, labeled factor loadings.

Let $\mathbf{F}$ be the $T \times r$ matrix of factors. The identification problem of $\mathbf{F}$ and $\mathbf{\Lambda}$ can be seen from the equality $\mathbf{F\Lambda} = \mathbf{FRR^{-1}\Lambda}$ for an $r \times r$ invertible (or invertible rotation) matrix $\mathbf{R}$, which has r^2 free parameters. Hence, at least r^2 parameter restrictions are needed for the model to be identified. The identity covariance matrix of the $\mathbf{f_t}$ imposes $\frac{r(r+1)}{2}$ restrictions, so an additional $\frac{r(r-1)}{2}$ restrictions are required for identification. The transformed and/or rotated factors and loadings still provide the same likelihood value.

We note that the key feature of factor models is that the information in the m economic variables of interest $\mathbf{y_t}$ can be compressed to a much lower number of r unobserved random factors $\mathbf{f_t}$. Given model and data, we intend to have this information dominate prior information.

However, given the present likelihood of the model with a diffuse prior it is clear that there does not exist an operational estimation procedure for structural inference using Bayesian MCMC. Of course, structural identification is not a necessary condition for forecasting, see Geweke (2007).

Next, consider the static factor model with a triangular normalization on $\mathbf{\Lambda}$, given as:

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda_1^{(r\times r)}} & \mathbf{\Lambda_2^{(r\times (m-r))}} \end{pmatrix}, \ \mathbf{\Lambda_1} = \begin{pmatrix} \lambda_{11} & 0 & \dots & 0 \\ \lambda_{21} & \lambda_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \dots & \lambda_{rr} \end{pmatrix} \tag{3}$$

where $\mathbf{\Lambda_2}$ is unrestricted. The triangular normalization on $\mathbf{\Lambda_1}$ provides $\frac{r(r-1)}{2}$ restrictions. Together with the restrictions on the covariance of the $\mathbf{f_t}$ this gives parametric identification. Combining a diffuse prior with the likelihood yields a posterior which is unbounded (for $\mathbf{F}$ tending to 0), but integrable.[1] Given the posterior structure and given an *a priori* fixed number of factors r the corresponding MCMC method is a basic Gibbs sampler. However, the important disadvantage is that inference depends on the ordering of the dependent variables.

As mentioned, Frühwirth-Schnatter et al. (2024) achieve inference which is independent from the ordering of the dependent variables by making use of Unordered Generalized Lower Triangular structure. We propose to obtain this independence by making use of orthogonal normalization on the parameters of the model. The orthogonal normalization implies that in this case no preferred ordering of the variables is imposed and, conditionally upon a largest singular value, the region of integration of the factors and factor loadings is bounded. That is, the parametrization $\mathbf{F\Lambda}$ can be linked to the singular value decomposition $\mathbf{F\Lambda} = \mathbf{UKV}$, where the rectangular $T \times r$ matrix $\mathbf{U}$ is an element of the Stiefel manifold $\mathbf{U'U} = \mathbf{I_r}$ and the $r \times m$ matrix $\mathbf{V}$ is an element of the Stiefel manifold $\mathbf{V'V} = \mathbf{I_m}$. $\mathbf{K}$ is a diagonal $r \times r$ matrix with positive diagonal entries equal to the singular values of $\mathbf{F\Lambda}$, denoted by $\kappa = (\kappa_1, \dots, \kappa_r)'$. The manifolds on which $\mathbf{U}$ and $\mathbf{V}$ are defined have finite volume conditionally upon a largest singular value and the region of integration of the factors and factor loadings is then bounded.

In order to achieve this we propose an approach that directly uses the structure of the singular value decomposition and makes use of a lasso type shrinkage prior for regularization,

[1] See Bastürk et al. (2017), Section 3.3 for proofs.

see Tibshirani (1996). As it is specified above, the singular value decomposition is not uniquely defined. Any simultaneous permutation of the columns of $\mathbf{U}$, $\mathbf{K}$ and $\mathbf{V}$ also constitutes a singular value decomposition. A common way to avoid this ambiguity is by ordering the singular values that occur on the diagonal of $\mathbf{K}$ as $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_r \geq 0$, which is more straightforward than devising an ordering of the columns of $\mathbf{F}$ and $\mathbf{\Lambda}$. Because of this ordering each element κ_{i+1} for $i = 1, \ldots, r-1$ is bounded by κ_i. Only κ_1 remains unbounded towards $+\infty$. Integrability is thus determined by the behavior of κ_1.[2]

A natural choice for κ_1 that is consistent with the uniform prior on the simplex for $\kappa_2, \ldots, \kappa_r|\kappa_1$ is the exponential distribution. Conditional on κ_1, all model parameters (except the covariance matrix Σ) are bounded to finite areas.

We conclude that given the standard form of a static factor model and using a lasso type shrinkage prior under orthogonal normalization on the parameters of the matrix with reduced rank, the marginal posteriors of these parameters are proper with finite first and higher order moments and inference is independent of the ordering of the dependent variables. For a survey on alternative identification restrictions with corresponding MCMC algorithms, see Koop et al. (2006).

3 Learning on a plausible number of factor model components using predictive likelihoods

We propose a 'predictive likelihood' approach to assess the number of factors in a factor model. The basic idea of this approach is to split the data into two parts: a training set and a 'hold-out' set of data. In the first part a possible weak prior is transformed to an informative posterior which serves as a prior for the second part of the data. This also refrains from the Bartlett (1957) paradox occurring under diffuse priors.[3] The gain in the use of 'prior data points' is to obtain predictive likelihoods for the computation of reliable predictive model probabilities. It is important to note that predictive likelihoods evaluated at different times (using moving

[2] If a singular value occurs more than once, then the columns of U and V corresponding to these singular values are not uniquely defined. Any other orthonormal basis that spans the same space will also do. Although the transformation between the matrix $\mathbf{F\Lambda}$ and its singular value decomposition (U, K, V) is still not invertible everywhere, this is an event with zero measure.

[3] Another approach is to construct a so-called imaginary sample by introducing a set of dummy observations. It yields a pragmatic class of priors proposed by Christopher Sims (Sims, 2005).

estimation windows) provide time-varying model probabilities. Any policy based on these model weights will therefore be time-varying as well.

Let M_r denote the factor model with $r << m$ factors. A predictive likelihood for model M_r is computed by splitting the dataset as follows:

$$Y = \begin{pmatrix} Y_{t_0:t_1} \\ Y_{t_1+1:t_2} \end{pmatrix} = \begin{pmatrix} Y^\star \\ \tilde{Y} \end{pmatrix} \tag{4}$$

where observations from t_0 to t_1 are defined as the 'training sample' and observations from t_1+1 to t_2 are defined as the 'hold-out sample'. The predictive likelihood for the hold-out sample is then defined as:

$$p(\tilde{Y}|Y^\star, M_r) = \frac{p(\tilde{Y}, Y^\star|M_r)}{p(Y^\star|M_r)} = \frac{p(Y|M_r)}{p(Y^\star|M_r)} \tag{5}$$

Choosing the size of the hold-out samples is important. If the hold-out sample is very small, the qualities of the models may be hard to distinguish (with almost equal model probabilities), and the results may be sensitive to just a few hold-out observations. If the hold-out sample is very large, the results may be sensitive to the few observations in the small training sample, and the Bartlett (1957) paradox may imply that we choose a model with too small number of factors r. Naturally, a robustness check should be performed to see the effect of training and hold-out sample size selection.

A simple method to estimate model probabilities is the harmonic mean estimator (Newton and Raftery, 1994; Ardia et al., 2012), which has the advantage that it is easily estimated using a set of draws generated from the posterior distribution of parameters θ_r of model M_r. The computational steps are as follows. Calculate two marginal likelihoods for each model M_r, a marginal likelihood for the whole sample and the second for the training sample. The full sample marginal likelihood is given as:

$$p(Y|M_r) = \int_{\theta_r} p(Y|\theta_r, M_r)p(\theta_r|M_r)d\theta_r \approx \left(\frac{1}{N} \sum_{i=1}^{N} p(Y|\theta_r^{f,i}, M_r)^{-1} \right)^{-1}$$

with posterior draws $\theta_r^{f,i}$ $(i = 1, \ldots, N)$ based on the full data sample. The training sample marginal likelihood is given as:

$$p(Y^\star|M_r) = \int_{\theta_r} p(Y^\star|\theta_r, M_r)p(\theta_r|M_r)d\theta_r \approx \left(\frac{1}{N}\sum_{i=1}^{N} p(Y^\star|\theta_r^{\star,i}, M_r)^{-1}\right)^{-1}$$

with posterior draws $\theta_r^{\star,i}$ $(i = 1, \ldots, N)$ based on the training sample. Next calculate predictive likelihoods for each model M_r using (5) as:

$$p(\tilde{Y}|Y^\star, M_r) = \frac{p(Y|M_r)}{p(Y^\star|M_r)} \approx \frac{\sum_{i=1}^{N} p(Y^\star|\theta_r^{\star,n}, M_r)^{-1}}{\sum_{i=1}^{N} p(Y|\theta_r^{f,n}, M_r)^{-1}}. \tag{6}$$

From the predictive likelihoods for each model compute model probabilities for M_r for $r \in 1, \ldots, m-1$:

$$p(M_r|Y) = \frac{p(\tilde{Y}|Y^\star, M_r) \times p(M_r)}{\sum_{r'=1}^{m-1} p(\tilde{Y}|Y^\star, M_{r'}) \times p(M_{r'})},$$

where $p(M_r)$ is the prior model probability. An uninformative prior, such as $p(M_r) = \frac{1}{m-1}$ is easy to use in this setting. Based on the predictive likelihood calculation from a rolling window of predictive likelihoods, an optimal model $M^\star$ can be chosen or Bayesian Model Averaging can be applied.

Simulated data experiment For illustrative purposes we apply the predictive likelihood approach for the factor model to simulated data. We consider simulated datasets with $T = 100$ and $T = 250$ observations. In order to see the effect of number of dependent variables m and number of factors r on the predictive likelihood methodology, we consider $r = 1, 2$ common factors for $m = 2, 4, 10, 20$ data series. For each simulation experiment, we apply the predictive likelihood approach with different sizes of training samples, consisting of 5%, 10%, 20% and 50% of observations. We replicate each simulation experiment 100 times.

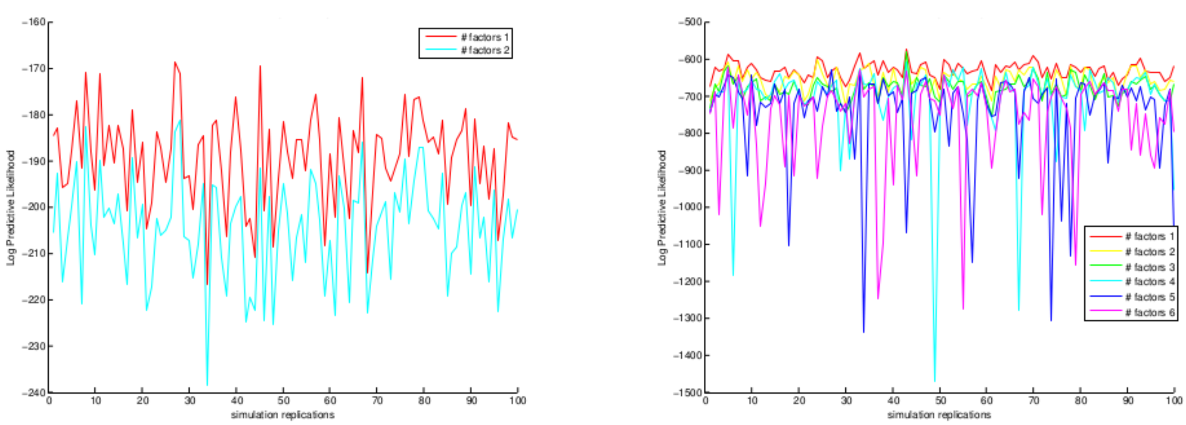
Table 1 presents the posterior probabilities from all simulation experiments, where we report the posterior model probabilities for different number of factors averaged over 100 simulation experiments for each simulation setting. Posterior results are based on 4000 posterior draws, where the first 2000 draws are burn-in draws.

The results in Table 1 indicate that the highest probabilities (indicated by boldface entries) for each simulation experiment, indicated in rows, correspond to the true number of factors.

Table 1: Average posterior probabilities from 100 simulation replications with T observations, m variables and r factors. Highest probabilities are indicated by **boldface** table entries.

T	m	r	pr($r=1$)	pr($r=2$)	pr($r=3$)	pr($r=4$)	pr($r=5$)	pr($r=6$)
5% training sample								
100	2	1	**1.00**	0.00	-	-	-	-
100	10	1	**0.96**	0.02	0.00	0.02	0.00	0.00
250	4	2	0.00	**1.00**	0.00	0.00	-	-
250	20	2	0.00	**1.00**	0.00	0.00	0.00	0.00
10% training sample								
100	2	1	**1.00**	0.00	-	-	-	-
100	10	1	**0.89**	0.07	0.00	0.03	0.01	0.00
250	4	2	0.00	**1.00**	0.00	0.00	-	-
250	20	2	0.00	**1.00**	0.00	0.00	0.00	0.00
20% training sample								
100	2	1	**1.00**	0.00	-	-	-	-
100	10	1	**0.77**	0.14	0.05	0.02	0.01	0.00
250	4	2	0.01	**0.98**	0.01	0.00	-	-
250	20	2	0.00	**0.95**	0.01	0.00	0.03	0.01
50% training sample								
100	2	1	**1.00**	0.00	-	-	-	-
100	10	1	**0.34**	0.13	0.02	0.08	0.18	0.24
250	4	2	0.00	**0.95**	0.05	0.00	-	-
250	20	2	0.00	**0.88**	0.02	0.02	0.01	0.07

Figure 1: Log-predictive likelihoods for different number of factors for two sets of simulated data with $T = 100$ observations. The left panel corresponds to $m = 2$ series and $r = 1$ common factor. The right panel corresponds to $m = 10$ series and $r = 1$ common factor. Each simulation experiment is repeated 100 times, as shown in the x-axes. Predictive likelihoods are calculated using 10 percent of the sample as the training sample.



In most simulation studies, the posterior probability is very close to 1 for the correct model specification. Hence the predictive likelihoods provide a clear choice of models. Comparing the bottom panel of Table 1 with the other panels, we conclude that the predictive likelihood approach with a smaller training sample than 50% provides more clear indications of the correct number of factors, with posterior probabilities being closer to 1 compared to the same simulation setting but a larger training sample (50%). Thus, the length of the training sample should not be chosen too large compared to the total length of the sample and a sensitivity analysis with respect to the length of the training sample will give more confidence in the results.

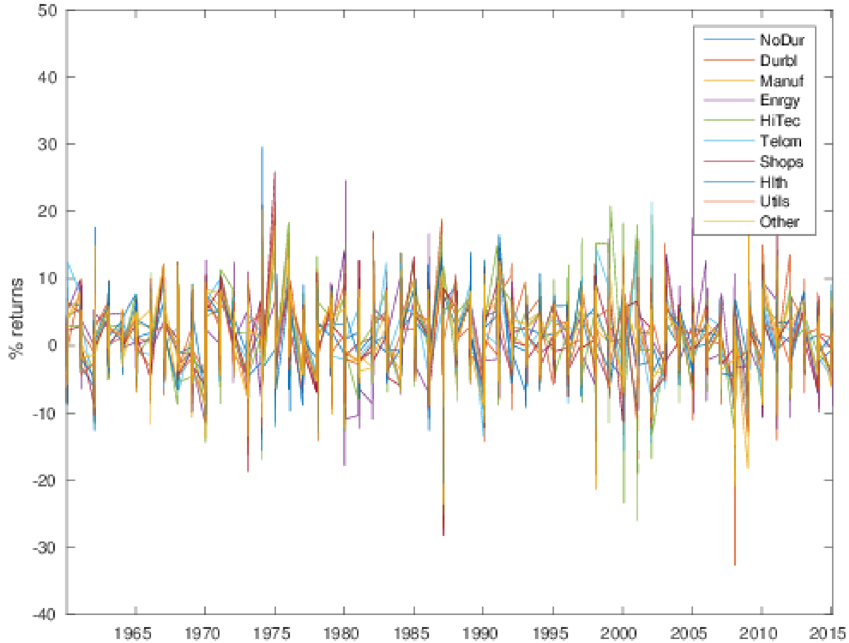
Figure 1 presents the details of predictive likelihoods for two sets of simulated data and for each simulation. These data correspond to $T = 100, p = 2, r = 1$ on the left panel of Figure 1 and $T = 100, p = 10, r = 1$ on the right panel of Figure 1. For both simulation specifications, the correct number of factors $r = 1$, shown by the red lines in the figure, has the highest posterior probability in almost all simulation replications. We therefore conclude that the predictive likelihood approach accurately detects the number of factors, even with a small sample size.

Equity momentum at work using a time-varying latent factor model Financial momentum strategies are based on the expectation that past stock winners will continue to be winners and past stock losers will continue to be losers. Standard equity momentum strategy ranks stocks on their recent returns, skips a short period to overcome short term reversals and then buys stocks in the top of the ranking and short-sells stocks in the bottom of the ranking. We emphasize that standard momentum's risk and return vary over time.

We compare the performance of residual momentum strategies based on the *residual returns* (returns in excess of what is to be expected based on the factors and factor loadings) with a standard momentum strategy. We use monthly return data on ten industry portfolios between 1960M7 and 2015M6, shown in Figure 2. The ten industries are labeled as 'non-durables', 'durables', 'manufacturing', 'energy', 'hi-tech', 'telecom', 'shops', 'health', 'utilities' and the final category 'others'. This residual industry momentum strategy is a combination of residual momentum (Blitz et al., 2011) and industry momentum (Moskowitz and Grinblatt, 1999).

Results are presented in Table 2. In order to allow for changes in the number of factors, we apply the moving window estimation method as follows. Within one estimation period, the

Figure 2: Monthly percentage returns for ten industry portfolios.

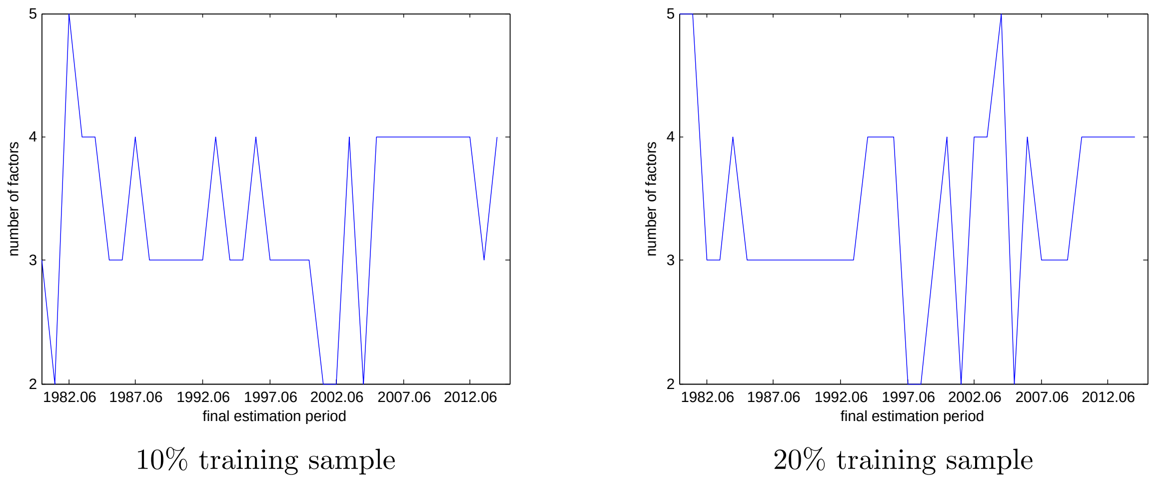


static factor model is estimated for $r = 2, \dots, 5$ factors, and the optimal number of factors is chosen based on the predictive likelihoods. The moving window estimation is based on a sample of $T = 240$ observations. We consider two cases for the predictive likelihood calculation with a 'small training sample' and a 'large training sample', consisting of 10% and 20% of the full moving window sample, respectively. We first analyze the evolution of the number of factors through the moving windows for the two training sample percentages. Figure 3 presents the number of factors with the highest predictive likelihood for each estimation sample under two training sample choices, where 10% and 20% of the total sample is used as training samples.

Figure 3 shows that the model with $r = 3$ factors is the most frequently chosen model using both training samples. Despite this high frequency, the optimal number of factors according to predictive likelihoods changes substantially over time. The obtained number of factors in the left and right panels of Figure 3 are different, in particular, at the end of the sample period. On the one hand, this difference indicates that the training sample choice should be made with care in order to find the appropriate number of factors. On the other hand, this variation in the number of factors may influence the gains from a trading strategy, like momentum. For the latter reason, we next report the gains from trading strategies using both training samples.

One would expect that the number of optimal factors varies with the performance of equity

Figure 3: Optimal number of factors for 10 industry portfolios. The figure presents the number of factors with highest predictive likelihood at each estimation window for two training sample choices (10% or 20% of the estimation window). X-axes correspond to the final month in the estimation window.



markets, in particular fewer factors are present in the model during market declines. We have two major market declines in our sample: 2000-2002 and 2008 where equity markets lost 56 and 38 percent, respectively. We indeed find that during the equity market losses in 2000 to 2002 the optimal number of factors was 2. For the 2008 crash we do not find a smaller number of optimal factors. This needs to be explored in further research.

Next, we report the performance of a standard momentum strategy and compare it with a residual momentum strategy based on the factor model. The standard strategy is a benchmark and by definition does not depend on an underlying model. The ten industry portfolios are sorted on their mean (residual) returns in the last 12 months. The strategy is long in the best industry and short in the worst industry, where this position is held for 12 months. The first investment month is July 1980, as we require 240 months since 1960 for the first estimation of the model parameters.

In Table 2, we report the following risk and return measures for the returns of each strategy: mean return, volatility, Sharpe ratio, largest loss encountered, maximum drawdown, all measured in percentages, and maximum recovery period measured in months. These values are based on the realized returns of each investment strategy.

The standard industry momentum strategy does not yield positive average returns; its average annual return is minus 1.1 percent. The 20 percent Bayesian Factor Model scores better in all six criteria compared to the standard industry momentum strategy.

Table 2: Risk and return characteristics of standard momentum compared with momentum from time-varying factors.

	Standard	Latent Factor Models	
		10%	20%
mean	-1.13	2.82	3.75
volatility	21.54	20.16	12.56
Sharpe ratio	-0.05	0.14	0.30
Largest loss	-68.52	-68.52	-20.95
Max. drawdown	135.3	68.52	35.19
Max. recovery	15	9	9

We conclude that a Bayesian latent factor model with a time-varying number of factors, moving window estimation and a training sample of 20 percent is able to outperform a standard momentum strategy for all criteria in this portfolio setting of ten industries. Apparently, the model adjusts quickly to big shocks and the number of optimal factors decreases when the equity market experiences large losses. However, more empirical work needs to be done to assess its properties adequately, which is outside the scope of the present paper.

4 Final comments

Since the early nineteen-seventies there has been a strong tradition in Bayesian econometrics of studying the shape and integrability of posteriors of parameters of multivariate regression models with a reduced rank using different normalization restrictions and so-called *regularization priors*. Apart from the factor model, the other models are a time series model with an unknown number of non-stationary components and a structural instrumental variable regression model where number and strength of instrumental variables is not known. Research in this field was started in econometrics by Anderson and Rubin (1949) in simultaneous equations models and in 1956 by the same authors in factor models (Anderson and Rubin, 1956). Johansen (1991) treated reduced rank in a time series model with possibly non-stationary components. A survey of the extensive recent frequentist literature is beyond the scope of this paper. There exists also an emerging Bayesian literature about reduced rank estimation, see Bastürk et al. (2017), Section 3 and Appendix 3.2 for details.

We end with emphasizing that the paper by Frühwirth-Schnatter et al. (2024) gave much *food for thought*. We look forward to more theoretical and empirical work on the topic by the authors. In this context, the dynamic nature of many models in economics is relevant. It is very natural to allow parameters of such models to move through time. The well-known Normal or Kalman Filter is a fundamental tool for this and it helps to give identification in factor models. Although dynamic factor models make the mathematics of identification and possible MCMC algorithms more complex, yet, this is a promising research field. We note that the application of static factor models using moving estimation windows and predictive likelihoods for time-varying posterior probabilities of numbers of factors is also able to yield profitable residual momentum strategies that outperform benchmark strategies as the standard momentum strategy.

A second topic is to extend the work by the authors to the field of forecast combinations. Some recent work is given in Billio et al. (2013), Casarin et al. (2023), Aastveit et al. (2023), and Aastveit et al. (2024). The predictive probabilities introduced in the present paper can be used that framework.

References

Aastveit KA, Cross J, Furlanetto F, Van Dijk HK. 2024. Taylor rules with endogenous regimes. Technical report, Tinbergen Institute Discussion Paper.

Aastveit KA, Cross JL, van Dijk HK. 2023. Quantifying time-varying forecast uncertainty and risk for the real price of oil. *Journal of Business & Economic Statistics* **41**: 523–537.

Anderson TW, Rubin H. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics* **20**: 46 —63.

Anderson TW, Rubin H. 1956. Statistical inference in factor analysis. In Neyman J (ed.) *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, Vol. 5*. Berkeley, 111–150.

Ardia D, Basturk N, Hoogerheide L, Van Dijk HK. 2012. A comparative study of monte carlo methods for efficient evaluation of marginal likelihood. *Computational Statistics and Data Analysis* **56**: 3398–3414.

Bartlett MS. 1957. A comment on D.V. Lindley's statistical paradox. *Biometrika* **44**: 533–534.

Bastürk N, Hoogerheide L, van Dijk HK. 2017. Bayesian analysis of boundary and near-boundary evidence in econometric models with reduced rank. *Bayesian Analysis* **12**: 879–917.

Billio M, Casarin R, Ravazzolo F, Van Dijk HK. 2013. Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics* **177**: 213–232.

Blitz D, Huij J, Martens M. 2011. Residual momentum. *Journal of Empirical Finance* **18**: 506–521.

Casarin R, Grassi S, Ravazzolo F, van Dijk HK. 2023. A flexible predictive density combination for large financial data sets in regular and crisis periods. *Journal of Econometrics* **237**: 105370.

Frühwirth-Schnatter S, Hosszejni D, Lopes HF. 2024. Sparse Bayesian factor analysis when the number of factors is unknown. *Bayesian Analysis* .

Geweke J. 2007. Interpretation and inference in mixture models: Simple mcmc works. *Computational Statistics & Data Analysis* **51**: 3529–3550.

Johansen S. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59**: 1551–1580.

Koop G, Strachan R, Van Dijk H, Villani M. 2006. Bayesian approaches to cointegration .

Moskowitz TJ, Grinblatt M. 1999. Do industries explain momentum. *The Journal of Finance* **54**: 1249–1290.

Newton MA, Raftery AE. 1994. Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society: Series B (Methodological)* **56**: 3–48.

Sims CA. 2005. Dummy observation priors revisited. *Manuscript, Princeton University* .

Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**: 267–288.