Tutorial lecture 4

Identification of Factor Models for Forecasting Returns

M. Deistler, E. Hamann

Econometrics, Operations Research
and Systems Theory

Research unit for Econometrics and System Theory

Institute for Mathematical Methods in Economics

University of Technology, Vienna

manfred.deistler@tuwien.ac.at

eva.hamann@tuwien.ac.at

http:\\www.eos.tuwien.ac.at

OUTLINE

1. INTRODUCTION
2. THE MODELS
3. MODEL SPECIFICATION
4. VALIDATION
5. EMPIRICAL ANALYSIS FOR DAILY SHARE PRICES
6. CONCLUSION

1. INTRODUCTION

Problem: Forecasting of returns

$r_t = \frac{p_t - p_{t-1}}{p_{t-1}}, \qquad t = 1, \ldots, T$

p_t . . . prices of financial assets, in particular, prices of shares.

This is important for ‘active’ portfolio management

‘Classical theory’:

$\mathbb{E}(p_{t+1}|p_t, p_{t-1}, \ldots) = p_t$ Weak form efficiency

or

$\mathbb{E}(p_{t+1}|I_t) = p_t$ Semi-strong efficiency,

I_t publically available information at time t

Question: Can we find ‘better’ forecasting models from data?

(Can we beat the market?)

Main issues and problems in statistical modeling:

1. Input selection
2. Modeling of dynamics
3. Possible nonlinearities
4. Structural changes
5. Outliers
6. Forecast evaluation

2. THE MODELS

In order to forecast the n–dimensional vector process (y_t) we make use of several kinds of factor models.

The main feature of factor models is that the vector of observed variables is explained by linear combinations of a small, say $r << n$, number of factors.

$$\boxed{\mathbf{y_{t+1}} = \boldsymbol{\Lambda}\xi_{\mathbf{t+1}} + \mathbf{u_{t+1}}} \tag{1}$$

(y_t) . . . n–dim. process of observed variables

(ξ_t) . . . r–dim. factor process

(u_t) . . . n–dim. noise process

$\Lambda \in \mathbb{R}^{n \times r}$ . . . matrix of factor loadings

General Assumptions:

- (ξ_t) and (u_t) are linearly regular, jointly stationary and ergodic processes with means zero.
- $\Sigma_\xi = \mathbb{E}\xi_t\xi_t' > 0.$
- $\mathbb{E}\xi_t u_s' = 0$ for all $t \leq s$.

Then the variance-covariance matrix of y_t, Σ_y, may be written as

$$\boxed{\mathbf{\Sigma_y} = \mathbf{\Lambda\Sigma_\xi\Lambda'} + \mathbf{\Sigma_u}.} \tag{2}$$

The assumptions imposed so far, however, do not determine a reasonable model class. In order to obtain reasonable model classes further assumptions have to be imposed.

Quasi static principal components (PCA) model:

Additional Assumption (PCA):

- The factor ξ_t and the loading matrix Λ are obtained by minimizing $\mathbb{E}u_t'u_t = \text{tr}(\Sigma_u)$ over all rank r matrices Λ and all Σ_ξ of rank r for given Σ_y.

An optimal solution to this problem is given by an eigenvalue decomposition of Σ_y,

$$\Sigma_y = O\Omega O' = \underbrace{O_1\Omega_1 O_1'}_{=\Lambda\Sigma_\xi\Lambda'} + \underbrace{O_2\Omega_2 O_2'}_{=\Sigma_u}, \tag{3}$$

where $\Omega = \text{diag}(\omega_1, \ldots, \omega_n)$ is the diagonal matrix of eigenvalues, with $\omega_i > \omega_j$ for all $j > i$, $O = [O_1, O_2]$ is an orthogonal matrix of eigenvectors and O' its transpose.

Implying $\Lambda = O_1$, $\xi_t = O_1' y_t$ and $u_t = O_2 O_2' y_t$.

Estimation of the factors and loadings:

Estimates for Λ and ξ_t are obtained by substituting the population variance-covariance matrix Σ_y by the sample variance-covariance matrix $\hat{\Sigma}_y^T = \frac{1}{T}\sum_{t=1}^T y_t y_t' = \hat{O}_T \hat{\Omega}_T \hat{O}_T'$, where T denotes sample size.

Due to the ergodicity of (y_t) and the fact that the eigenvalues as well as the normalized eigenvectors are continuous functions of $\hat{\Sigma}_y^T$ it follows that the estimates for the factors and the loadings are given by $\hat{\xi}_t = \hat{O}_{1,T}' y_t$ and $\hat{\Lambda} = \hat{O}_{1,T}$.

Forecasting the factors:

For forecasting the factor process (ξ_t), here, we use an ARX model of the form

$$\boxed{\boldsymbol{\xi_{t+1} = \mathbf{A(z)}\xi_t + \mathbf{D(z)x_t} + \epsilon_{t+1},}} \tag{4}$$

where $A(z)$ and $D(z)$ are polynomial matrices in the backward shift operator z of order p and q, resp., and the stability condition

$$\det(I - z(A(z))) \neq 0 \text{ for all } |z| \leq 1 \tag{5}$$

holds.

Assumptions (Forecasting Model):

- (ϵ_t) is white noise.
- (x_t) is an m–dimensional linearly regular, stationary and ergodic process with nonsingular spectral density and mean zero and $\mathbb{E}x_t\epsilon_s' = 0$ for all t, $s \in \mathbb{Z}$.
- Furthermore we assume, that $\mathbb{E}x_t u_s' = 0$ for all t, $s \in \mathbb{Z}$.

Forecasting the n–dim. process (y_t):

The one-step ahead forecasts of y_{t+1} are then obtained as $\hat{y}_{t+1|t} = \hat{\Lambda}\hat{\xi}_{t+1|t}$, where $\hat{\xi}_{t+1|t}$ are the one-step ahead forecasts of ξ_{t+1} based on Equation (4).

Thus, $\hat{\xi}_{t+1|t} = \hat{F}\hat{\eta}_t$, where
$F = [A_0\ A_1\ \ldots A_p\ D_0\ D_1\ \ldots\ D_q] \in \mathbb{R}^{r\times[(p+1)r+(q+1)m]}$,
$\eta_t = (\xi_t', \xi_{t-1}', \ldots, \xi_{t-p}', x_t', x_{t-1}', \ldots, x_{t-q})'$,
$\hat{F} = \sum_{s=1}^{t-1}(\hat{\xi}_{s+1}\hat{\eta}_s')(\sum_{s=1}^{t-1}(\hat{\eta}_s\hat{\eta}_s'))^{-1}$ and $\hat{\eta}_t$ is defined analogously to $\hat{\xi}_t$.

Note, that this predictor is not the (linearly least squares) optimal forecast of y_{t+1} given its past, because u_t may contain further forecasting information.

Quasi static factor model with idiosyncratic noise (IN):

Additional Assumptions (IN):

- Σ_u is diagonal.
- $\Sigma_\xi = I_r$, the identity.

The idea behind the first assumption is to attribute the joint effects to the factors and the individual effects to the noise.

In spite of the second assumption two identifiability problems arise in estimating the loading matrix Λ:

(i) Obtain $\Lambda\Lambda'$ from Σ_y.

(ii) Obtain Λ from $\Lambda\Lambda'$.

Ad (i): one can show that if $r \leq \frac{2n+1}{2} - \sqrt{(\frac{2n+1}{2})^2 - (n^2 - n)}$, the so-called Ledermann bound, then $\Lambda\Lambda'$ is generically unique from Σ_y.

Ad (ii): If a partition of Σ_y has been found, then (under the assumption that Λ has rank r) Λ is uniquely determined from $\Lambda\Lambda'$ up to postmultiplication by an arbitrary orthogonal matrix M:

- The normalization $\Lambda'\Sigma_u^{-1}\Lambda$ *is diagonal, its elements are larger than zero, ordered in size and distinct* ensures that Λ, given by $\Lambda = \Sigma_u^{\frac{1}{2}} P Q^{\frac{1}{2}}$, is unique up to sign changes in the columns of P, where $Q = \text{diag}(q_1, \ldots, q_r)$ is the $(r \times r)$–dimensional matrix of non-zero eigenvalues of $\Sigma_u^{-\frac{1}{2}}(\Sigma_y - \Sigma_u)\Sigma_u^{-\frac{1}{2}} = \Sigma_u^{-\frac{1}{2}}\Lambda\Lambda'\Sigma_u^{-\frac{1}{2}} \geq 0$ and P is the $(n \times r)$–dimensional matrix of the corresponding eigenvectors, thus $P'P = I_r$.
- Other choices for Λ may be obtained from methods like the *varimax* and the *promax* method.

Estimation of Λ and Σ_u:

Estimates for Λ and Σ_u are obtained by iteratively maximizing the function,

$$\boxed{\mathbf{L_T}(\mathbf{\Lambda}, \mathbf{\Sigma_u} | \hat{\mathbf{\Sigma}}_{\mathbf{y}}^{\mathbf{T}}) = -\frac{\mathbf{T}}{\mathbf{2}} \log \det \mathbf{\Sigma_y} - \frac{\mathbf{T}}{\mathbf{2}} \text{trace}(\mathbf{\Sigma_y^{-1}} \hat{\mathbf{\Sigma}}_{\mathbf{y}}^{\mathbf{T}}),} \tag{6}$$

subject to rank$(\Lambda) = r$, $\Sigma_u > 0$ and the default normalization condition from above.

Note, that in case of independent identically normally distributed noise and factors, the function given in (6) is (up to a constant) the loglikelihood function of y_t.

In case of autoregressive factors and noise considered here, (6) however, is not the likelihood function.

Nevertheless, the estimates $\hat{\Lambda}$ and $\hat{\Sigma}_u$ obtained from maximizing (6) can be shown to be consistent estimates for Λ and Σ_u, if $\hat{\Sigma}_y^T$ is a consistent estimate of Σ_y.

Estimation of the unobserved factors ξ_t:

In contrast to the PCA model, here, the factors, in general, cannot be obtained directly as a function of the observed y_t and, hence, have to be approximated by some (linear) function of y_t:

1. *Regression method by Thomson:* approximates the factor process in least squares sense by some linear combination of y_t, obtained from

$$\boxed{\min_{\mathbf{A}\in\mathbb{R}^{\mathbf{n}\times\mathbf{r}}} \mathbb{E}(\xi_{\mathbf{t}} - \mathbf{A}'\mathbf{y}_{\mathbf{t}})(\xi_{\mathbf{t}} - \mathbf{A}'\mathbf{y}_{\mathbf{t}})'.} \tag{7}$$

   From the assumptions on p. 5 we obtain $A' = \Lambda'\Sigma_y^{-1}$ and, therefore, substituting Λ by $\hat{\Lambda}$ and Σ_y by $\hat{\Sigma}y$ yields $\hat{\xi}_t = \hat{\Lambda}'\hat{\Sigma}_y^{-1}y_t$.

2. *Bartlett's method:* Bartlett's idea was to minimize the sum of the squared standardized residuals with respect to the r–dimensional factor process,

$$\boxed{\min_{\xi_t}(\mathbf{y_t} - \hat{\mathbf{\Lambda}}\xi_\mathbf{t})'\hat{\mathbf{\Sigma}}_\mathbf{u}^{-1}(\mathbf{y_t} - \hat{\mathbf{\Lambda}}\xi_\mathbf{t}),} \tag{8}$$

giving $\hat{\xi}_t = (\hat{\Lambda}'\hat{\Sigma}_u^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Sigma}_u^{-1}y_t$.

Which method should be chosen?

There is no general rule.

Decision might be based on the properties the estimates of the factor should possess:

smallest variance $\longrightarrow$ Regression method

some kind of unbiasedness $\longrightarrow$ Bartlett's method

Forecasting the factors:

This is done by an ARX scheme analogously to the PCA case on p. 8.

Forecasting the noise component:

Since the noise is assumed to be idiosyncratic and may be interpreted as asset specific component, we additionally consider univariate ARX models in order to predict the noise component,

$$\boxed{\mathbf{u}_{t+1}^{(i)} = \mathbf{a}_i(\mathbf{z})\mathbf{u}_t^{(i)} + \mathbf{D}_i(\mathbf{z})\mathbf{z}_t^{(i)} + \nu_{t+1}^{(i)}, \quad \mathbf{i} = 1, \ldots, \mathbf{n}.} \tag{9}$$

Assumptions (Forecasting the noise component):

- (ν_t) is white noise.
- $(z_t^{(i)})$ is an m_i–dimensional linearly regular, stationary and ergodic process with nonsingular spectral density and mean zero and $\mathbb{E} z_t^{(i)} \nu_s^{(i)'} = 0$ for all t, $s \in \mathbb{Z}$, and $i = 1, \ldots, n$.
- $\mathbb{E} z_t^{(i)} x_s' = 0$ for all t, $s \in \mathbb{Z}$ and $i = 1, \ldots, n$.
- $|1 - a_i(z)| \neq 0$ for all $|z| \leq 1$, $i = 1, \ldots, n$.

The one-step ahead forecasts for $u_{t+1}^{(i)}$ are given by $\hat{u}_{t+1|t}^{(i)} = \hat{F}_i \hat{\eta}_t^{(i)}$, where
$\hat{\eta}_t^{(i)} = (\hat{u}_t^{(i)'}, \ldots, \hat{u}_{t-p_i}^{(i)'}, z_t^{(i)'}, \ldots, z_{t-q_i}^{(i)'})$,
$\hat{F}_i = \sum_{s=1}^{t-1} (\hat{u}_{s+1}^{(i)} \hat{\eta}_s^{(i)'}) (\sum_{s=1}^{t-1} (\hat{\eta}_s^{(i)} \hat{\eta}_s^{(i)'}))^{-1}$ and $\hat{u}_t = y_t - \hat{\Lambda}\hat{\xi}_t$.

Finally, one may compute three different types of one-step ahead forecasts for y_{t+1}:

1. $\hat{y}_{t+1|t}^{I} = \hat{\Lambda}\hat{\xi}_{t+1|t}$,
2. $\hat{y}_{t+1|t}^{II} = \hat{\Lambda}\hat{\xi}_{t+1|t} + \hat{u}_{t+1|t}$
3. $\hat{y}_{t+1|t}^{III} = \hat{u}_{t+1|t}$.

Reduced rank regression (RR):

The model is of the form:

$$\mathbf{y_{t+1} = FG}\chi_\mathbf{t} + \mathbf{u_{t+1}}, \qquad \mathbf{t} \in \mathbb{Z} \tag{10}$$

where $\chi_t = (y'_t, y'_{t-1}, \ldots, y'_{t-p}, x'_t, x'_{t-1}, \ldots, x'_{t-q})'$,
(x_t) is the m–dimensional process of exogenous variables described above,
$\xi_{t+1} = G\chi_t$ is the r–dimensional factor process, $(r = \text{rank}(FG) \leq n)$ and
$F \in \mathbb{R}^{n\times r}$, $G \in \mathbb{R}^{r\times(n(p+1)+m(q+1))}$.

Additional Assumption (RR):

- u_t is white noise

Hence (10) is an ARX system with retricted parameters.

Estimation of the product FG:

The product $\beta = FG$ is given by

$$\boxed{\mathbf{FG} = \mathbf{\Sigma_{y\chi}\Sigma_{\chi}^{-1}},} \tag{11}$$

where $\Sigma_{y\chi} = \mathbb{E}y_t\chi_t{}'$ and $\Sigma_{\chi} = \mathbb{E}\chi_t\chi_t{}'$.

By replacing $\Sigma_{y\chi}$ and Σ_{χ} by their sample counterparts $\hat{\Sigma}_{y\chi} = \frac{1}{T-1}\sum_{t=1}^{T-1} y_{t+1}\chi_t{}'$ and $\hat{\Sigma}_{\chi} = \frac{1}{T-1}\sum_{t=1}^{T-1} \chi_t\chi_t{}'$ we obtain the least squares estimate $\hat{\beta} = \hat{\Sigma}_{y\chi}\hat{\Sigma}_{\chi}^{-1}$.

Typically rank$(\hat{\beta}) = n$, even if $r < n$ holds.

To obtain a rank reduced estimator of β we perform a Singular Value Decomposition (SVD) of $\hat{\beta}$:

$$\hat{\beta} = U\Sigma V',$$

where U and V are orthogonal matrices and $\Sigma \in \mathbb{R}^{n\times m}$ is the diagonal matrix of singular values, σ_i, $i = 1, \ldots, n$, arranged in decreasing order.

We consider two methods:

1. *Direct* estimator:

$$\boxed{\hat{\hat{\beta}}_{\mathbf{D}} = \mathbf{U_1 \Sigma_1 V_1}',} \tag{12}$$

where $\Sigma_1 \in \mathbb{R}^{r\times r}$ is the matrix formed from the r largest singular values of Σ and U_1 and V_1, respectively, are formed from the first r columns of U and V, respectively.

2. *Indirect* estimator: Here we form the SVD of the weigthed matrix

$$\hat{\Sigma}_y^{-\frac{1}{2}}\hat{\beta}\hat{\Sigma}_\chi^{\frac{1}{2}} = U\Sigma V',$$

where $\hat{\beta}$ is the least squares estimator $\hat{\Sigma}_{y\chi}\hat{\Sigma}_\chi^{-1}$. Retaining only the r largest singular values we obtain (using an obvious notation)

$$\boxed{\hat{\hat{\beta}}_{\mathbf{I}} = \hat{\mathbf{\Sigma}}_{\mathbf{y}}^{\frac{1}{2}}\mathbf{U_1 \Sigma_1 V_1}'\hat{\mathbf{\Sigma}}_\chi^{-\frac{1}{2}}.} \tag{13}$$

Forecasting the n–dim. process (y_t):

Forecasts of y_{t+1} are obtained from $\hat{y}_{t+1|t} = \hat{\beta}\chi_t$.

Analogously to the IN case one might consider forecasting models for the noise part.

3. MODEL SPECIFICATION

Specification, here, is completely data driven and basically consists of two parts:

- Estimation of the factor dimension r.
- Selection of the explanatory variables from a set of a priori defined candidates.

Due to the high number (compared to sample size) of different model classes that are compared with respect to fit and complexity one is confronted with two problems in practice:

- The computational costs may be very high.
- What may be even more serious is that severe overfitting may occur.

Quasi static principal components (PCA) model:

Determination of r:

- Rule of thumb: r equals the number of eigenvalues of the correlation matrix of y_t that are larger than one.
- Fix r at some small number, e.g. 2 or 3.

Given $\hat{r}$, estimates for loadings and factors are computed as described on p. 6.

Input selection and specification of the dynamics:

In practice, the number of candidates for explanatory variables, say k, is large.
For model selection we use information criteria (IC) of AIC or BIC type.
In many cases it is not reasonable to search over all possible subsets of the set of explanatory variables (denoted by S_k). Our model selection procedures are stepwise procedures based on the idea of An and Gu for single equation models.

Fast Step Procedure (FSP) for single equation models:

1. Perform a *forward procedure* to find an IC optimal initial set:
   - Search for the IC optimal singleton, denote it S_1.
   - Search for the IC optimal explanatory variable to be added to S_1, giving S_2.
   - Continue adding explanatory variables to S_{i-1}, until $S_i = S_k$.
   - Out of the sets $S_1, \ldots, S_k$, the one with the lowest criterion value is chosen to be the initial set.
2. Given the initial set, add and drop one variable in each step as long as the criterion value can be decreased, otherwise stop.

Fast Step Procedure (FSP) for a system of equations:

1. Find some initial set:

   - *mva method:* Apply a forward algorithm to all equations at once, i.e. search for the IC optimal explanatory variable to be added to all equations until the criterion value computed for the whole system cannot be reduced anymore.
   - *univ method:* Apply the forward algorithm from above to each single equation and take the union of the n sets obtained in this way as initial set.

   The selection of such an initial set imposes zero restrictions on the elements of matrix F from p. 9.

2. Add one variable or drop one variable at a time, meaning that in each step of the procedure zero restrictions on the parameters are canceled or added, as long as the criterion value of the system can be improved.

Quasi static factor model with idiosyncratic noise (IN):

Determination of r: The selection of r and the explanatory variables are related:

1. For each r smaller or equal the Ledermann bound the ARX models for the factor process are specified, analogous to the PCA case.
2. r is chosen to give an optimal trade-off between in-sample explanatory power for y_t and model complexity.

Note:

- In this procedure input selection and dynamic specification for the ARX model is based on a goodness-of-fit measure for the factor process, whereas determination of the number of factors is based on a goodness-of-fit measure of y_t.
- The procedure described here could also be applied to the principal components case.

Reduced rank regression (RR):

$$\left.\begin{array}{l}\text{Specification of the number of factors}\\[1em]\text{Inputselection}\\[1em]\text{Specification of dynamics}\end{array}\right\}\text{ is performed simultaneously.}$$

The initial number of factors is chosen to be n.

The initial set of explanatory variables is found by applying one of the forward procedures for systems of equations described above.

A refinement of the second step of the FSP from p. 25 for this model class is as follows:

It is not only allowed to add or drop variables from the set, but also to let the number of factors variate from 1 up to n and to weight observations at time t with some weighting factor λ^{T-t}, $\lambda \in (0, 1]$.

Thus, in each iteration step the IC optimal variable to be added to the set, dropped from the set, the IC optimal number of factors and the IC optimal weighting factor λ are determined, giving four 'optimal' criterion values. From these four 'optimal' specifications again the best is selected. The procedure stops, when the IC value cannot be improved anymore by any of the four possibilities mentioned.

4. VALIDATION

Our forecast procedures are 'honest', in the sense of being strictly out-of-sample, i.e. for forecasting y_{t+1} only data up to time t are used, both for estimation of real valued parameters and for model specification.

For each dependent variable, $y_{t+1}^{(i)}$, the one-step (here one day) ahead predictors, $\hat{y}_{t+1|t}^{(i)}$, and the corresponding prediction errors, $\hat{u}_{t+1}^{(i)} = y_{t+1}^{(i)} - \hat{y}_{t+1|t}^{(i)}$, are calculated from a model identified from data up to time t, using both an extending and a moving window, respectively.

The estimators of the real valued parameters are updated at every time instance.

The specification is updated every five or every ten days.

The sample is divided into two parts, $1, \ldots, T_1$ and $T_1 + 1, \ldots, T_2$. Only the latter part is used for evaluating the out-of-sample forecasts.

The evaluation sample, $T_1 + 1, \ldots, T_2$, consists of the last 30% of the whole sample.

We consider two measures for the quality of the forecasts:

- The out-of-sample coefficient of determination
$R^2_{(i)} = 1 - \frac{(\hat{u}^{(i)} - \bar{\hat{u}}^{(i)})'(\hat{u}^{(i)} - \bar{\hat{u}}^{(i)})}{(y^{(i)} - \bar{y}^{(i)})'(y^{(i)} - \bar{y}^{(i)})}$, where $\hat{u}^{(i)}$ and $y^{(i)}$, respectively, are the vectors consisting of the components $\hat{u}_t^{(i)}$ and $y_t^{(i)}$ from the validation sample, $t = T_1 + 1, \ldots, T_2$, $\bar{\hat{u}}^{(i)}$ and $\bar{y}^{(i)}$ denote the respective sample means.

- The out-of-sample hitrate given by,
$h_{(i)} = \frac{1}{T_2 - T_1} \sum_{t=T_1+1}^{T_2} sign(y_t^{(i)} \hat{y}_{t|t-1}^{(i)})$.

These measures should be interpreted with care. A real test of the forecasting quality in our context would be the evaluation via the profits made from portfolio optimization.

5. EMPIRICAL ANALYSIS FOR DAILY SHARE PRICES

Data: Daily close return data for shares of the banking sector in STOXX50E together with corresponding input data.

The *banks* are: ABN AMRO (H.AAB), Banco Bilbao (U.BBV), Banco Santander (E.SCH), HypoVereinsbank (D.HVM), Deutsche Bank (D.DBK), BNP Paribas (F.BNP), UniCredito (I.UC)

A set of 19 input candidates for the *factors* is given a priori, containing, among others, indices for the banking sector, interest rates and futures for indices. The list of explanatory variables for the factors is composed of present inputs and outputs together with their lags of order one and five.

The explanatory variables for the *errors* are chosen in order to reflect the influences on the main markets for the corresponding banks, e.g. a bank and financial index for Eastern Europe for the HypoVereinsbank.

These data are available for us from 16/06/00 to 13/11/02.

For each model class (PCA, IN and RR resp.) considered we allow for certain design specifications.

Coding system:

- PCA: “pca[number of principal components].[re-specification period].[initial method (mva/univ)].[criteria used in the two steps of the FSP].[logical for moving (TRUE) or expanding (FALSE) window]”
- IN: “in[re-specification period].[factor estimates (regression/Bartlett)]. [rotation (varimax/promax/default)].[initial method (mva/univ)]. [criteria used in the two steps of the FSP].[criterion used to specify the number of factors (AIC/BIC)] .[logical for moving (TRUE) or expanding (FALSE) window]. [orthogonal projection of the explanatory variables of the error models on the orthocomplement of the inputs of the factor models (TRUE/FALSE)]”
- RR: “rr[re-specification period].[estimation of β (direct/indirect)]. [initial method (mva/univ)].[criteria used in the two steps of the FSP]. [logical for moving (TRUE) or expanding (FALSE) window]. [orthogonal projection of the input variables of the error models on the orthocomplement of the inputs of the factor models (TRUE/FALSE)]”

Results:

- As far as the forecasting qualities are concerned there is a clear ranking: PCA is worst and RR is best, see Table 1 for an example.
- The performance of PCA is far from satisfactory in most cases.
- Choice of the IC is crucial: For IN BICF-BIC.AIC gives the best results and for RR, BICF-BIC.
- Adding the noise forecasts is helpful in some cases, but may deteriorate the forecasts in other cases, see Table 2.
- For IN *regression* gives somewhat better results than *Bartlett*.
- For RR *indirect* estimation gives somewhat better results than *direct* estimation, see Table 3.

	PCA		IN		RR	
banks	R^2	hitrate	R^2	hitrate	R^2	hitrate
H.AAB	0.03243	0.48947	0.01437	0.54211	0.06873	0.61579
U.BBV	-0.07921	0.46842	-0.04040	0.47895	0.00042	0.48421
E.SCH	-0.00847	0.50526	-0.02160	0.50526	0.01655	0.57895
D.HVM	-0.03152	0.46316	0.02917	0.47895	0.02631	0.53158
D.DBK	-0.00721	0.52105	-0.00377	0.53158	0.05968	0.58947
F.BNP	0.03233	0.52632	0.01342	0.52105	0.10100	0.57895
I.UC	-0.03411	0.45263	-0.04453	0.48947	0.00924	0.54211

Table 1: The out-of-sample results for pca2.5.mva.BICF.BIC.TRUE, in10.Bartlett.none.mva.BICF-BIC.BIC.TRUE.FALSE and rr5.direct.BICF-BIC.TRUE.FALSE considering only forecasts based on the factor part of the models. Note furthermore, that pca2.5.mva.BICF.BIC.TRUE is one of the best models of all computed PCA models.

	Forecasting with the factor part				Forecasting with the factor and noise part			
	IN		RR		IN		RR	
banks	R^2	hitrate	R^2	hitrate	R^2	hitrate	R^2	hitrate
H.AAB	0.05187	0.56842	0.10062	0.57368	0.04334	0.55789	0.06342	0.60000
U.BBV	-0.01701	0.47368	-0.00489	0.44737	0.00325	0.50000	-0.02530	0.44211
E.SCH	0.00976	0.52632	0.04256	0.54211	0.03343	0.53684	0.06387	0.53158
D.HVM	0.06070	0.51579	0.03325	0.48947	0.06724	0.50000	0.04214	0.49474
D.DBK	0.05627	0.54211	0.05914	0.53158	0.05384	0.54211	0.07514	0.54737
F.BNP	0.05906	0.54211	0.11594	0.52105	0.04201	0.56316	0.10915	0.53158
I.UC	-0.01234	0.50526	0.02512	0.50526	-0.02815	0.50000	-0.03518	0.50526

Table 2: The out-of-sample results of models in10.regression.default.univ.BICF-BIC.AIC.TRUE.TRUE and rr10.indirect.BICF-BIC.TRUE.TRUE.

	direct				*indirect*			
	moving window		expanding window		moving window		expanding window	
banks	R^2	hitrate	R^2	hitrate	R^2	hitrate	R^2	hitrate
H.AAB	0.06873	0.61579	0.06899	0.56842	0.09528	0.57368	0.08564	0.58421
U.BBV	0.00042	0.48421	-0.00571	0.43158	-0.00500	0.45263	-0.00396	0.47368
E.SCH	0.01655	0.57895	0.03012	0.55789	0.03508	0.53684	0.04869	0.53684
D.HVM	0.02631	0.53158	0.03380	0.49474	0.04181	0.50000	0.03948	0.53158
D.DBK	0.05968	0.58947	0.07036	0.57368	0.07339	0.54211	0.08297	0.57895
F.BNP	0.10100	0.57895	0.09290	0.52105	0.11356	0.52105	0.12012	0.55263
I.UC	0.00924	0.54211	0.04310	0.52632	0.01374	0.47368	0.05237	0.54211

Table 3: The out-of-sample results of the models rr5.direct.BICF-BIC.TRUE.TRUE, rr5.direct.BICF-BIC.FALSE.TRUE, rr5.indirect.BICF-BIC.TRUE.TRUE, rr5.indirect.BICF-BIC.FALSE.TRUE based of the factor part only.

6. CONCLUSION

We have investigated three different types of factor models for forecasting daily close return data:

- Quasi static principal components
- Quasi static factor models with idiosyncratic noise
- Reduced rank regression

The reduced rank models showed the best performance.
Input selection and dynamic specification is an important issue and here has been performed by modified An-algorithms, where BIC gives the best results. The best models obtained seem to be of reasonable quality.