Forecasting Financial Time Series

Manfred Deistler and Christiane Zinner

Canberra, February, 2007

Contents

Forecasting Financial Time Series: Problems and Approaches

Time series:
(Relative) returns

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \quad , \quad t = 1 \dots T$$

p_t ... prices of financial assets such as shares, indices, exchange rates.
Different time scales: Intraday (high frequency), daily, weekly, monthly data.

Aims

In (quantitative) portfolio management it is of interest to forecast

- Conditional expectations
- Conditional variances

Here we only consider prediction of conditional expectations.

Stylized facts
Return series are not i.i.d., but show little serial correlation.
Support by ”classical“ theory:

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

or

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

where $I_t \ldots$ publically available information at time t
$\Rightarrow$ Problem: Can we find better forecasting models from data - Can we beat the market?

Main issues in this context:

- Input selection
- Modeling of dynamics
- Nonlinearities
- Structural changes and adaption
- Outliers
- Problem dependent criteria for forecast quality and forecast evaluation

- Input selection
  Large number (several thousands) of candidates, and large number of model classes: Overfitting and computation time.
  What are appropriate model selection criteria and strategies? *AIC* and *BIC* do not seem to be appropriate in cases where the number of model classes is of the same order as sample size. Intelligent search algorithms (An algorithm, Genetic programming); The role of prior knowledge.
- Modelling of dynamics
  Similar to input selection
- Nonlinearities
  NN as benchmarks

- Structural changes and adaptation
  Time varying parameters, adaptive estimation procedures, detection of local trends.
- Criteria for forecast quality
  Ideally depending on portfolio optimization criteria. Out of sample R^2 or hitrate are only easy-to-calculate-substitutes. Long run target: Portfolio optimization criteria based identification procedures.

Factor Models

Here we consider factor type models for forecasting return series:

- linear
- multivariate

The Basic Frame Work

Factor models are used to condense high dimensional data consisting of many variables into a much smaller number of factors. The factors represent the comovement between the single time series or underlying nonobserved variables influencing the observations. Here we consider (static and dynamic)

- Principal component models
- Frisch or idiosyncratic noise models
- Generalized linear factor models

for forecasting return series.

The model

The basic, common equation for all different kinds of factor models considered here is of the form

$$y_t = \Lambda(z)\xi_t + u_t, \tag{1}$$

where

y_t	...	observations (n-dim).
ξ_t	...	factors (unobserved) ($r << n$-dim).
$\Lambda(z) = \sum_{j=-\infty}^{\infty} \Lambda_j z^j$, $\Lambda_j \in \mathbb{R}^{n \times r}$	...	factor loadings
$\hat{y}_t = \Lambda(z)\xi_t$	...	latent variables
$\Lambda = \Lambda_0$	...	(quasi-)static case.

Assumptions

Throughout we assume the following:

- $\mathrm{E}\xi_t = 0,\ \mathrm{E}u_t = 0$ for all $t \in \mathbb{Z}$.
- $\mathrm{E}\xi_t u_s' = 0$ for all $s,\ t \in \mathbb{Z}$.
- (ξ_t) and (u_t) are wide sense stationary and (linearly) regular with covariances $\gamma_\xi(s) = \mathrm{E}\xi_t\xi_{t+s}'$ and $\gamma_u(s) = \mathrm{E}u_t u_{t+s}'$ satisfying

$$\sum_{s=-\infty}^{\infty} |s|\|\gamma_\xi(s)\| < \infty, \sum_{s=-\infty}^{\infty} |s|\|\gamma_u(s)\| < \infty \qquad (2)$$

- The spectral density $f_{\hat{y}}$ of $\hat{y}_t$ has rank r for all $\lambda \in [-\pi, \pi]$.

Then the spectral density f_y of y_t exists:

$$f_y(\lambda) = \Lambda(e^{-i\lambda}) f_\xi(\lambda) \Lambda(e^{-i\lambda})^* + f_u(\lambda). \tag{3}$$

And for the static case we obtain

$$\Sigma_y = \Lambda \Sigma_\xi \Lambda^* + \Sigma_u \quad \text{where e.g. } \Sigma_y = \mathbb{E} y_t y_t' \tag{4}$$

We will be concerned with the following questions.

- Identifiability questions:
  - Identifiability of $f_{\hat{y}} = \Lambda f_{\xi} \Lambda^{*}$ and f_u
  - Identifiability of Λ and f_{ξ}
- Estimation of integers and real-valued parameters:
  - Estimation of r
  - Estimation of the free parameters in Λ, f_{ξ}, f_u
  - Estimation of ξ_t
- Forecasting

Principal Component Analysis

The quasi-static case:

We commence from the eigenvalue decomposition of Σ_y:

$$\Sigma_y = O \Omega O' = \underbrace{O_1 \Omega_1 O_1'}_{\Sigma_{\hat{y}}} + \underbrace{O_2 \Omega_2 O_2'}_{\Sigma_u},$$

where Ω_1 is the $r \times r$-dim. diagonal matrix containing the r largest eigenvalues of Σ_y. This decomposition is unique for $\omega_r > \omega_{r+1}$.

A special choice for the factor loading matrix is $\Lambda = O_1$, then $y_t = O_1 \xi_t + u_t$, $\xi_t = O_1' y_t$ and $u_t = y_t - O_1 O_1' y_t = O_2 O_2' y_t$

Note: Here, factors are linear functions of y_t.

Estimation:

Determine r from $\omega_1, \ldots, \omega_n$
Estimate $\Lambda, \Sigma_\xi, \Sigma_u, \xi_t$ from the eigenvalue decomposition of
$\hat{\Sigma}_y = \frac{1}{T} \sum_{t=1}^{T} y_t y_t'$

The dynamic case:
We commence from the canonical representation of the spectral density f_y:

$$f_y(\lambda) = \underbrace{O_1(e^{-i\lambda})\Omega_1(\lambda)O_1(e^{-i\lambda})^*}_{f_{\hat{y}}(\lambda)} + \underbrace{O_2(e^{-i\lambda})\Omega_2(\lambda)O_2(e^{-i\lambda})^*}_{f_u(\lambda)},$$

then

$$y_t = O_1(z)\xi_t + u_t \text{ and } \xi_t = O_1^*(z)y_t$$

Note: Here $\mathbb{E}u_t'u_t$ is minimal among all decompositions where $\mathrm{rk}(f_{\hat{y}}(\lambda)) = r$ a.e.
Whereas for given r the decomposition above is unique and thus $(\hat{y}_t)$ and (u_t) are identifiable, the transfer function $\Lambda(z)$ and the factors ξ_t are only identifiable up to regular dynamic linear transformations.

Again $\xi_t = O_1^*(z) y_t$, i.e. factors are linear transformations of (y_t)
Problem: In general, the filter $O_1^*(z)$ will be non-causal and non-rational. Thus, naive forecasting may lead to infeasible forecasts for y_t. Restriction to causal filters is required.
In estimation, we commence from a spectral estimate.

The Frisch Model

Here the additional assumption f_u is diagonal is imposed in (1).
Interpretation: Factors describe the common effects, the noise u_t takes into account the individual effects, e.g. factors describe markets and sector specific movements and the noise the firm specific movements of stock returns.
For given $\hat{y}_t$ the components of y_t are conditionally uncorrelated.

The quasi-static case:

Identifiability: More demanding compared to PCA

$$\Sigma_y = \underbrace{\Lambda \Sigma_\xi \Lambda'}_{\Sigma_{\hat{y}}} + \Sigma_u, \quad (\Sigma_u) \text{ diagonal} \tag{5}$$

Identifiability of $\Sigma_{\hat{y}}$: Uniqueness of solution of (5)
for given n and r, the number of equations (i.e. the number of free elements in Σ_y) is $\frac{n(n+1)}{2}$. The number of free parameters on the r.h.s. is $nr - \frac{r(r-1)}{2} + n$.

Now let

$$B(r) = \frac{n(n+1)}{2} - (nr - \frac{r(r-1)}{2} + n) = \frac{1}{2}((n-r)^2 - n - r)$$

then the following cases may occur:

- $B(r) < 0$: In this case we might expect non-uniqueness of the decomposition
- $B(r) \geq 0$: In this case we might expect uniqueness of the decomposition

The argument can be made more precise, in particular, for $B(r) > 0$ generic uniqueness can be shown. Given Σ_y, if $\Sigma_\xi = I_r$ is assumed, then Λ is unique up to postmultiplication by orthogonal matrices (rotation).

Note that, as opposed to PCA, here the factors ξ_t, in general, cannot be obtained as a function of the observations y_t. Thus, the factors have to be approximated by a linear function of y_t.

Estimation:
If ξ_t and u_t were Gaussian white noise, then the (negative logarithm of the) likelihood function has the form

$$
\begin{aligned}
L_T(\Lambda, \Sigma_u) &= \frac{1}{2} T \log(\det(\Lambda\Lambda' + \Sigma_u)) + \frac{1}{2} \sum_{t=1}^{T} y_t'(\Lambda\Lambda' + \Sigma_u)^{-1} y_t = \\
&= \frac{1}{2} T \log(\det(\Lambda\Lambda' + \Sigma_u)) + \frac{1}{2} T \operatorname{tr}((\Lambda\Lambda' + \Sigma_u)^{-1} \hat{\Sigma}_y). (6)
\end{aligned}
$$

The dynamic case:
Here Equation (1) together with the assumption

$$f_u \text{ is diagonal.}$$

is considered. Again u_t represents the individual influences and ξ_t the comovements. The only difference to the previous section is that Λ is now a dynamic filter and the components of u_t are orthogonal to each other for all leads and lags.
There are still many unsolved problems.

Generalized Linear Dynamic Factor Model

- Motivation:
  - In a number of applications the cross-sectional dimension is high, possibly exceeding sample size.
  - Weak dependence (“local” correlation) between noise components should be allowed.
- Examples:
  - cross-country business cycle analysis
  - asset pricing
  - forecasting returns of financial instruments
  - monitoring and forecasting economic activity by estimation of common factors (“diffusion indexes”)

- The model
  For the analysis n is not regarded as fixed. Thus we consider a double sequence $(y_{it} | i \in \mathbb{N}, t \in \mathbb{Z})$, where our general assumptions hold true for every vector $y_t^n = (y_{1,t}, y_{2,t}, \ldots, y_{n,t})'$ with $n \in \mathbb{N}$.
  $\Rightarrow$ Sequence of factor models:

$$y_t^n = \Lambda^n(z)\xi_t + u_t^n, \; t \in \mathbb{Z}, \; n \in \mathbb{N}, \tag{7}$$

  where u_t^n and $\Lambda^n(z)$ are nested.

- Additional Assumptions
  - (Weak Dependence) The largest eigenvalue of f_u^n, $\omega_{u,1}^n : [-\pi, \pi] \to \mathbb{R}$ say, is uniformly bounded for all $n \in \mathbb{N}$, i.e. there exists a $\overline{\omega} \in \mathbb{R}$, such that $\omega_{u,1}^n(\lambda) \leq \overline{\omega}$ for all $\lambda \in [-\pi, \pi]$ and for all $n \in \mathbb{N}$.
    $\Rightarrow$ the variance of (u_t^n) "averages-out" for $n \to \infty$.
  - The first r eigenvalues of $f_{\hat{y}}^n$, $\omega_{\hat{y},j}^n$ say, $j = 1, \ldots, r$ diverge a.e. in $[-\pi, \pi]$ as $n \to \infty$.
    $\Rightarrow$ minimum amount of correlation between latent variables.

Representation, Forni and Lippi, 2001
The double sequence $\{y_{it} | i \in \mathbb{N}, t \in \mathbb{Z}\}$ can be represented by a generalized dynamic factor model, if and only if,

1. the first r eigenvalues of f_y^n, $\omega_{y,j}^n$ say, $j = 1, \ldots, r$, diverge a.e. in $[-\pi, \pi]$ as $n \to \infty$, whereas
2. the $(r+1)$-th eigenvalue of f_y^n, $\omega_{y,r+1}^n$ say, is uniformly bounded for all $\lambda \in [-\pi, \pi]$ and for all $n \in \mathbb{N}$.

The last result indicates that the spectral densities of $\hat{y}_t$ and u_t are asymptotically (for $n \to \infty$) identified by dynamic PCA, i.e. the canonical representation of f_y^n, decomposed as

$$f_y^n(\lambda) = O_1^n(e^{-i\lambda})\Omega_1^n(\lambda)O_1^n(e^{-i\lambda})^* + O_2^n(e^{-i\lambda})\Omega_2^n(\lambda)O_2^n(e^{-i\lambda})^*, \tag{8}$$

Factor Space, Forni and Lippi, 2001
The space spanned by the first r dynamic principal components of (y_t^n), i.e. $O_1^n(z)^* y_t^n$, converges to the space spanned by the factors ξ_t, where space is short for Hilbert space and where convergence of spaces is understood in the sense that the residuals of a projection from one space to the other converge to 0 in mean square.

Latent variables, Forni et al., 2000
The latent variables from the PCA-model converge to the corresponding generalized dynamic factor model variables as $n \to \infty$. Let $\hat{y}_{it}^n = O_{1,i}^n(z) O_1^n(z)^* y_t^n$ denote the projection of y_{it}^n onto $O_1^n(z)^* y_t^n$, i.e. the i-th element of the latent variable of the corresponding PCA-model at t, then

$$\lim_{n \to \infty} \hat{y}_{it}^n = \hat{y}_{it}, \text{ for all } i \in \mathbb{N} \text{ and for all } t \in \mathbb{Z}, \tag{9}$$

where $\hat{y}_{it}$ denotes the i-th element of $\Lambda^n(z)\xi_t$ (for $n \geq i$), hence the corresponding “true” latent variable.

Identifiability

Concerning identifiability, the results from dynamic PCA are adopted asymptotically, i.e. asymptotically (as n tends to infinity) the latent variables as well as the idiosyncratic components are identifiable, whereas the transfer function $\Lambda(z)$ and the factors ξ_t are only identifiable up to regular dynamic linear transformations. If the factors are assumed to be white noise with $\mathrm{E}\xi_t\xi_t' = I_r$, which is no further restriction, since our assumptions always allow this transformation, they are identifiable up to static rotations.

For estimation the dynamic PCA estimators are employed, thus e.g. $\hat{\hat{y}}_t^n = \left[\hat{O}_1^n(z)\hat{O}_1^n(z)^*\right]_t y_t^n$, where $\hat{O}_1^n(e^{-i\lambda})$ denotes the matrix consisting of the first r eigenvectors of a consistent estimator of $f_y^n(\lambda)$, $\hat{f}_x^n(\lambda)$ say.

The filter $\hat{O}_1^n(z)\hat{O}_1^n(z)^*$ is in general two-sided and of infinite order and has to be truncated at lag $t-1$ and lead $T-t$ as for $t \leq 0$ and $t > T$ y_t^n is not available. As a consequence of this truncation convergence of the estimators of $\hat{y}_t$ and u_t, as n and T tend to infinity, can only be granted for a “central” part of the observed series, whereas for fixed t the estimators are never consistent.

Determination of r

So far the number of factors r was considered fixed and known, whereas in practice it has to be determined from the data. The above discussion indicates that the eigenvalues of $\hat{f}_x^n$ could be used for determining the number of factors and Forni et al. propose a heuristic rule, but indeed, no formal testing procedure has been developed yet.

One-sided model

The fact that the filters occurring in dynamic PCA are in general two-sided and thus non causal yields infeasible forecasts. One way to overcome this problem is to assume that

- $\Lambda^n(z)$ is of the form $\Lambda^n(z) = \sum_{j=0}^{p} \Lambda_j^n z^j$

and that

- (ξ_t) is of the form $\xi_t = A(z)^{-1}\varepsilon_t$, $A(z) = I - A_1 z - \dots A_s z^s$ with $A_j \in \mathbb{R}^{r \times r}$, $\det A(z) \neq 0$ for $|z| \leq 1$, $s \leq p + 1$ and the innovations ε_t are r-dimensional white noise with $\mathrm{E}\varepsilon_t \varepsilon_t' = I$ (and are orthogonal to u_t at any leads and lags.)

The model then can be written in a quasi static form, on the cost of higher dimensional factors,

$$y_t = \Lambda F_t + u_t = \hat{y}_t + u_t,\ t \in \mathbb{Z}, \tag{10}$$

where $F_t = (\xi_t', \dots, \xi_{t-p}')'$ is the $q = r(p+1)$-dimensional vector of stacked factors and $\Lambda^n = (\Lambda_0^n, \dots, \Lambda_p^n)$ is the $(n \times q)$-dimensional static factor loading matrix. – Note that $(y_t^n), \Lambda^n, (\hat{y}_t^n), (u_t^n)$ and their variances and spectra still depend on n, but for simplicity we will drop the superscript from now. Under the assumptions imposed F_t and u_t remain orthogonal at any leads and lags and thus f_y is of the form

$$f_y(\lambda) = \Lambda f_F(\lambda)\Lambda^* + f_u(\lambda) \tag{11}$$

.

For estimation, Stock and Watson propose the static PCA procedure with q factors and they prove consistency for the factor estimates (i.e. the first q sample principal components of y_t) up to premultiplication with a nonsingular matrix as n and T tend to infinity. In other words, the space spanned by the true GDFM-factors can be consistently estimated.

An alternative two-stage "generalized PCA" estimation method has been proposed by Forni et al. It differs from classical PCA in two respects: firstly in the determination of the covariance and secondly in the determination of the weighting scheme. While classical PCA is based on the sample covariance $\hat{\Sigma}_y$ of (y_t), this approach commences from the estimated spectral density $\hat{f}_y$ decomposed according to the dynamic model. Then the covariance matrices $\Sigma_{\hat{y}}$ and Σ_u of the common component and the noise respectively are estimated as

$$\hat{\Sigma}_{\hat{y}} = \int_{\pi}^{-\pi} \hat{f}_{\hat{y}}(\lambda) d\lambda \quad \text{and} \tag{12}$$

$$\hat{\Sigma}_u = \int_{\pi}^{-\pi} \hat{f}_u(\lambda) d\lambda. \tag{13}$$

In the second step, the factors are estimated as $\hat{F}_t = C' y_t$, where the $(n \times q)$-dimensional weighting matrix C is determined as the first q generalized eigenvectors of the matrices $(\hat{\Sigma}_{\hat{y}}, \hat{\Sigma}_u)$, i.e. $\hat{\Sigma}_{\hat{y}} C = \hat{\Sigma}_u C \Omega_1$, where Ω_1 denotes the diagonal matrix containing the q largest generalized eigenvalues. C is then the matrix that maximizes

$$\begin{aligned} & C' \hat{\Sigma}_{\hat{y}} C \\ \text{s.t.} \quad & C' \hat{\Sigma}_u C = I_q, \end{aligned} \tag{14}$$

(whereas in static PCA the corresponding weights C maximize $C' \hat{\Sigma}_y C$, s.t. $C' C = I_q$).

The matrix Λ is then estimated by projecting y_t onto $C'y_t$, i.e. $\hat{\Lambda} = \hat{\Sigma}_y C(C'\hat{\Sigma}_y C)^{-1}$. The common component estimator $\hat{\hat{y}}_t$ defined this way can be shown to be consistent (for n and T tending to infinity). The argument for this procedure is that variables with higher noise variance or lower common component variance respectively get smaller weights and vice versa, thus the common-to-idiosyncratic variance ratio in the resulting latent variables is maximized, which could possibly improve efficiency upon static PCA.

Boivin and Ng propose a third estimation method called "weighted PCA" which is related to classical PCA as generalized least squares (GLS) is related to ordinary least squares (OLS) in the regression context. Remember, that, if the regression residuals are non-spherical, but their variance matrix is known, GLS weights the residuals with the inverse of the square root of their variance matrix and yields efficient estimators. Assume for a moment that the noise variance Σ_u were known, then the same principle could be applied to PCA by transforming the least squares problem $\min \mathrm{E} u_t' u_t$ into the generalized least squares problem $\min \mathrm{E} u_t' \Sigma_u^{-1} u_t$.
Since the residual matrix of the unweighted PCA-model is singular, it cannot be used. A feasible alternative is e.g. the estimator resulting ¿from dynamic PCA. Again the factor space is consistently estimated as n and T tend to infinity.

Forecasting

Concerning forecasting, at least two different approaches have been proposed.

- First, as suggested by Forni et al., the problem of forecasting $y_{i,t}$ say can be split into forecasting the common component $\hat{y}_{i,t}$ and forecasting the idiosyncratic $u_{i,t}$ separately. To forecast the common component, $y_{i,t+h}$ is projected onto the space spanned by the factor estimates $\hat{F}_t$:

$$\hat{y}_{i,t+h|t} = \mathrm{proj}(\hat{y}_{i,t+h}|\hat{F}_t) = \mathrm{proj}(y_{i,t+h}|\hat{F}_t).$$

  The idiosyncratic $u_{i,t}$ may be forecast by a univariate AR-process.

- Second, as proposed by Stock and Watson, forecasting the common component $\hat{y}_{i,t}$ and the idiosyncratic $u_{i,t}$ may be performed simultaneously. In this case $u_{i,t}$ is supposed to follow the AR(S)-process $b(z)u_{i,t} = \nu_{i,t}$ with $\nu_{i,t}$ being white noise. In the case of one-step ahead forecasts, the prediction equation

$$y_{i,t+1} = \Gamma F_t + u_{i,t+1} \tag{15}$$

can then be rewritten as

$$y_{i,t+1} = \tilde{\Gamma}\tilde{F}_t + \gamma(z)y_{i,t} + \nu_{i,t+1}, \tag{16}$$

where $\gamma(z) = (1 - b(z))z^{-1}$, and hence $y_{i,t+1}$ is estimated by a “factor augmented AR model” (FAAR).

Forecasting stock index returns

Data and sample period

- Targets (i.e. time series to be forecast)
  Weekly returns of 5 investable stock indices: Euro STOXX 50, Nikkei 225, S&P 500, Nasdaq and FTSE (data: Monday close prices)
- Additional information
  additionally used to estimate the factor space: relative differences (weekly) of sector indices of Euro STOXX 50, Nikkei 225 and S&P 500

$\Rightarrow$ 54 time series with 421 observations from 01/04/1999 until 01/22/2007.

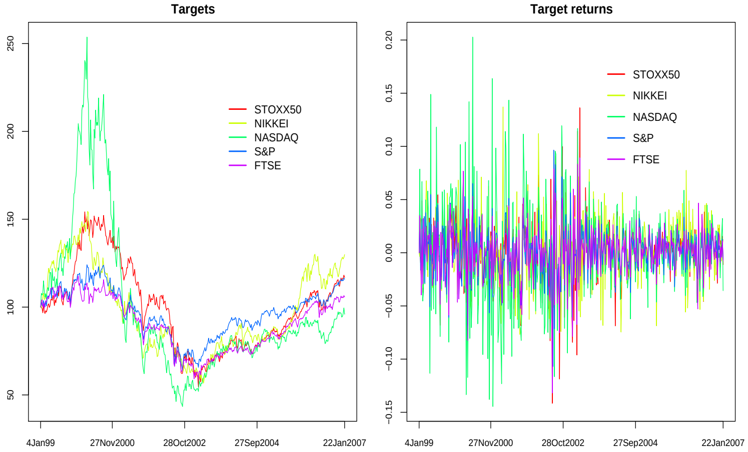

Figure: Target time series: absolute prices (indexed) and returns.

Estimation details

We calculate 1-week ahead forecasts recursively based on a rolling 250-weeks estimation window.

$\Rightarrow$ out of sample forecasting period: 171 weeks (10/20/2003 - 01/22/2007)

- Two benchmarks
  - Univariate AR(p)-models (lag selection: recurs. computed AIC, BIC with $0 \leq p \leq 15$
  - Univariate full ARX-models : all 54 time series are used as regressors (with lag 1), the target time series may enter with more lags (selection as for AR(p)-models).

Estimation details

- Generalized dynamic factor models
  - FAAR : simultaneous selection of q (no. of static factors) and p (AR-lags) by rec. computed AIC, BIC applied to the (univariate) forecasting equation.
  - PCA : selection of q by rec. computed AIC, BIC and modified AIC, BIC (as proposed by Bai, Ng, 2002) applied to the (multivariate) factor model equation of the targets. (idiosyncratic model, see AR(p)-model)
  - GPCA (Generalized PCA) : selection of r (no. of dynamic factors) based on dynamic eigenvalues, selection of q and p: see PCA.

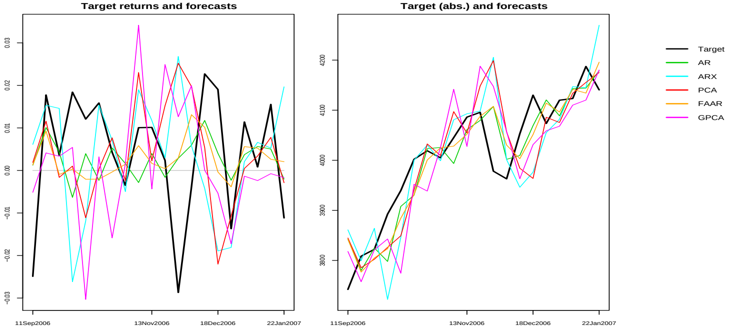

Figure: Example of target and forecast time series for Euro STOXX 50.

Results - Forecasting quality

With respect to o.o.s. R^2 and hitrate the GDFM-methods dominate the 2 benchmarks across different series and selection methods; in the case of forecasting root mean square error (RMSE) this does not hold true.

Average o.o.s. statistics over all target time series:

	AR	ARX	FAAR	PCA	GPCA
RMSE	0.0248	0.0302	0.0247	0.0255	0.0258
R2	0.0018	0.0034	0.0080	0.0054	0.0041
Hitrate	0.51	0.51	0.52	0.53	0.53

Portfolio simulations

We will consider

- One asset long/short portfolios : according to the forecast's direction the whole capital is invested either long or short into the asset.
- Equally weighted long short portfolios : according to the forecast's direction a constant proportion of the capital is invested either long or short into each asset (hence the weights can only be 0.2 or -0.2, as we have 5 assets to be traded).

For these portfolios we can compare returns, volatilities, Sharpe-ratios.

Portfolio simulations

Portfolio statistics for one-asset long/short strategy (average across series):

	AR	ARX	FAAR	PCA	GPCA
Return p.a.	0.0383	0.0137	0.0228	0.1404	0.0665
Volatility p.a.	0.1390	0.1394	0.1391	0.1379	0.1389
Sharpe-R. p.a.	0.2814	0.0940	0.1959	1.0188	0.5291

The two GDFM-methods that split the forecasting problem into 2 separated problems for the latent component and the noise, PCA and GPCA, clearly outperform all other models; FAAR-forecasts are very similar to AR-forecasts but slightly worse, and ARX-forecasts (without input selection!) perform worst not generating any significant return.

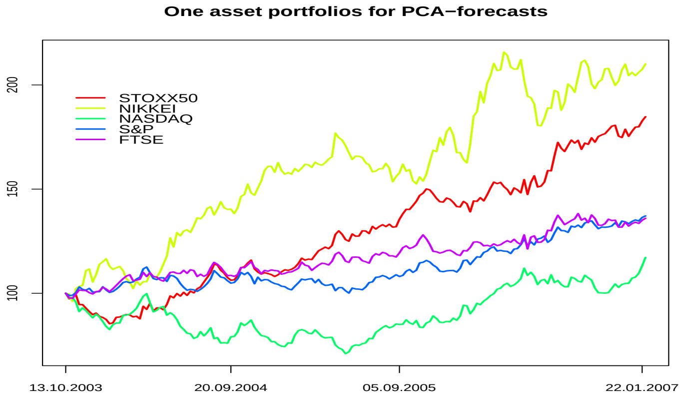

Figure: Example for one-asset portfolios according to PCA-forecasts.

Portfolio simulations

Portfolio statistics for equally weighted long/short strategy:

	AR	ARX	FAAR	PCA	GPCA
Return p.a.	0.0421	0.0202	0.0267	0.1450	0.1149
Volatility p.a.	0.0905	0.0869	0.0936	0.0873	0.0797
Sharpe-R. p.a.	0.4656	0.2328	0.2857	1.6614	0.9162

Once again, the PCA- and GPCA-forecasts, clearly outperform all other models; FAAR-forecasts behave very similar to AR-forecasts but slightly worse, and ARX-forecasts (without input selection!) perform worst not generating any significant return.

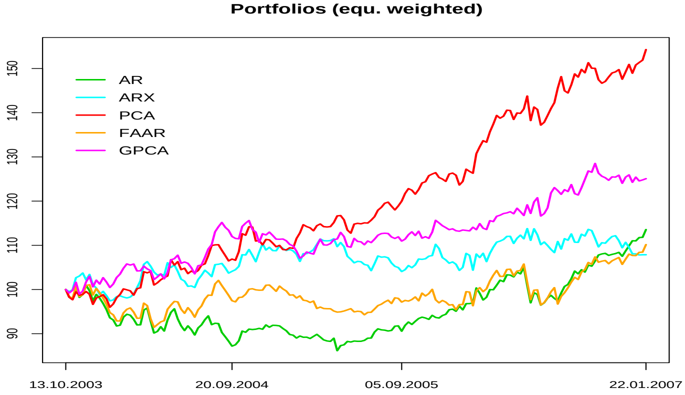

Figure: Equally weighted long short portfolios according to different forecasting models.

Conclusions

- Obviously, forecasting financial time series is a very difficult problem.
- A large number of candidate inputs and model classes highly increases the risk of overfitting.
- Forecasting models for high-dimensional time series are needed.
- Factor models solve, at least in part, the input selection problem as they condense the information contained in the data into a small number of factors. They allow for modelling dynamics and for forecasting high-dimensional time series.
- However, in this area there is still a substantial number of unsolved problems.