EPJ manuscript No.
(will be inserted by the editor)

Extracting Factors for Interest Rate Scenarios

Lutz Molgedey[1 a] and Elvis Galic[2]

[1] Institute of Physics, Humboldt–University, Invalidenstr. 110, 10115 Berlin, Germany, lutz@molgedey.de
[2] Allfonds BKG Asset Management, Arabellastr. 27, 81925 München, Germany, elvis.galic@allfonds-bkg.de

Received: date / Revised version: date

Abstract. Factor based interest rate models are widely used for risk managing purposes, for option pricing and for identifying and capturing yield curve anomalies. The movements of a term structure of interest rates are commonly assumed to be driven by a small number of orthogonal factors such as SHIFT, TWIST and BUTTERFLY (BOW). These factors are usually obtained by a Principal Component Analysis (PCA) of historical bond prices (interest rates). Although PCA diagonalizes the covariance matrix of either the interest rates or the interest rate changes, it does not use both covariance matrices simultaneously. Furthermore higher linear and nonlinear correlations are neglected. These correlations as well as the mean reverting properties of the interest rates become crucial, if one is interested in a longer time horizon (infrequent hedging or trading). We will show that Independent Component Analysis (ICA) is a more appropriate tool than PCA, since ICA uses the covariance matrix of the interest rates as well as the covariance matrix of the interest rate changes simultaneously. Additionally higher linear and nonlinear correlations may be easily incorporated. The resulting factors are uncorrelated for various time delays, approximately independent but nonorthogonal. This is in contrast to the factors obtained from the PCA, which are orthogonal and uncorrelated for identical times only. Although factors from the ICA are nonorthogonal, it is sufficient to consider only a few factors in order to explain most of the variation in the original data. Finally we will present examples that ICA based hedges outperforms PCA based hedges specifically if the portfolio is sensitive to structural changes of the yield curve.

PACS. 05.45.Tp Time series analysis – 02.50.Ey Stochastic processes

1 Introduction

Yield Curve Dynamics is commonly assumed to be driven by a small number of orthogonal factors such as parallel SHIFTS, changes in slope (TWIST) and changes in curvature (BUTTERFLY or BOW) [1–3]. However, in the present paper, we will argue that there is no reason to assume orthogonality of the factors.

Our study is based on the n–year zero interest rates $r_n(t)$ calculated from the REX–subindices for maturities of 1 to 10 years on a daily timescale from 1/1988 to 6/2000.

Further we make use of some empirical stylized facts about interest rate:

1. Interest rates are mean reverting.
2. Autocorrelation functions of interest rate changes are fast decaying – daily changes can be assumed to be δ–correlated.
3. Interest rate changes have leptokurtic distributions (*fat tails*).
4. Autocorrelation functions of squared and absolute changes are slow decaying (*volatility clustering* and *leverage effects*).

[a] corresponding author

2 Linear Factor Decomposition

Linear factor decomposition is based on a linear superposition of N independent factors $s_i(t)$

$$r_i(t) = \bar{r}_i + \sum_{j=1}^{N} W_{ij} s_j(t) \tag{1}$$

where $\bar{r}_i = \langle r_i(t) \rangle$, $\langle s_i(t) \rangle = 0$ and $\langle \cdots \rangle$ denotes time average. In the financial context the elements of the mixing matrix $\mathbf{W}$ are called factor loadings. By the convention

$$\sum_{i=1}^{N} W_{ij}^2 = 1 \tag{2}$$

the total variance is simply the sum of the factor variances

$$\sum_{i=1}^{N} \langle (r_i(t) - \bar{r}_i)^2 \rangle = \sum_{i=1}^{N} \langle s_i(t)^2 \rangle \quad . \tag{3}$$

In order to specify the model (1) one has to estimate N^2 parameters from the time series.

2.1 Principal Component Analysis

Principal Component Analysis (PCA) is traditionally used to obtain the linear mixing matrix $\mathbf{W}$ and the factor variances $\langle s_i(t)^2\rangle$ from a measured time series.

Due to the mean reverting properties of the interest rates one can assume stationarity of time series. Therefore PCA can be applied to the time series of either the interest rates or the interest rate changes.

PCA constructs an orthogonal superposition by an eigenvalue decomposition of the covariance matrix of the interest rates (Fig. 1 upper)

$$Q_{ij} = \langle (\mathbf{r_i} - \bar{\mathbf{r}}_\mathbf{i})\,(\mathbf{r_j} - \bar{\mathbf{r}}_\mathbf{j})^T \rangle \tag{4}$$

or of the interest rate changes $x_i(t) = r_i(t) - r_i(t-1)$ (Fig. 1 lower)

$$\hat{Q}_{ij} = \langle \mathbf{x_i}\mathbf{x_j}^T \rangle \quad . \tag{5}$$

Due to the orthogonality assumption, one only has to estimate $N(N+1)/2$ parameters for the orthogonal and normalized mixing matrix $\mathbf{W}^{-1} = \mathbf{W}^T$ ($N(N-1)/2$ parameters) and the factor variances (N parameters). This is in correspondence to the $N(N+1)/2$ elements of the measured symmetric covariance matrix.

2.2 Time Delayed Decorrelation

There is no reason to assume an orthogonal mixing, beside that the number of parameters which are to be estimated from a noisy time series is reduced by a factor of two. For the estimation of the full, not necessary orthogonal, mixing matrix one needs more information from the time series.

The simplest idea is to use both covariance matrices. Hence, $N(N+1)/2$ elements of the interest rate covariance matrix and $N(N+1)/2$ elements of the interest rate changes covariance matrix are known. From this knowledge the unknown $N(N-1)$ parameters of the mixing matrix, the N factor variances and the N factor changes variances can be obtained in a unique way.

The solution is given by the Time Delayed Decorrelation (TDD) [4] algorithm, adapted to the interest rate problem. The TDD algorithm diagonalizes both correlation matrices simultaneously (Fig. 1 dashed coordinates)

$$\mathbf{W}^{-1}\mathbf{Q}\left(\mathbf{W}^T\right) = \mathbf{\Lambda} \tag{6}$$

$$\mathbf{W}^{-1}\hat{\mathbf{Q}}\left(\mathbf{W}^T\right) = \hat{\mathbf{\Lambda}} \tag{7}$$

where $\mathbf{\Lambda}$ and $\hat{\mathbf{\Lambda}}$ are the diagonal matrices of the factor variances and factor changes variances respectively, i.e.

$$\lambda_i = \langle s_i(t)^2 \rangle \tag{8}$$

$$\hat{\lambda}_i = \langle \Delta_i^{(1)}(t)^2 \rangle \quad . \tag{9}$$

The τ-day factor changes are defined as

$$\Delta_i^{(\tau)}(t) = s_i(t+\tau) - s_i(t) \quad . \tag{10}$$

Since generally $\mathbf{W}$ is not an orthogonal matrix this leads to the eigenvalue problem

$$\left(\mathbf{Q}\hat{\mathbf{Q}}^{-1}\right)\mathbf{W} = \mathbf{W}\left(\mathbf{\Lambda}\hat{\mathbf{\Lambda}}^{-1}\right) \tag{11}$$

which may easily be solved by standard linear algebra packages.

2.3 Factor Extraction as an Optimization Problem

An equivalent formulation to that of the eigenvalue problem (6,7,11) is the following optimization problem

$$V = \frac{1}{2}\sum_{i\neq j} \langle s_i(t) s_j(t)\rangle^2 + \langle \Delta_i^{(1)}(t)\Delta_j^{(1)}(t)\rangle^2 \tag{12}$$

where V should be minimal. In the case of the TDD $V \equiv 0$ is the optimal solution. The advantage of this formulation as an optimization problem is the extendability. Larger time lags and nonlinear correlations may be easily incorporated, i.e.

$$\begin{aligned} V = \frac{1}{2}\sum_{i\neq j} \quad & \sum_{\tau=0} \alpha_\tau \langle s_i(t) s_j(t+\tau)\rangle^2 + \\ & \sum_{\tau=1} \beta_\tau \langle \Delta_i^{(\tau)}(t)\Delta_j^{(\tau)}(t)\rangle^2 + \\ & \gamma\left(\langle \Delta_i^{(1)}(t)^2 \Delta_j^{(1)}(t)^2\rangle - \langle \Delta_i^{(1)}(t)^2\rangle\langle \Delta_j^{(1)}(t)^2\rangle\right)^2 \end{aligned} \tag{13}$$

where $\alpha_\tau \geq 0$, $\beta_\tau \geq 0$, $\gamma \geq 0$ are free parameters. Noteably, this optimization problem is still solvable by iterated linear algebra in a highly efficient way - namely by the Joint Approximate Diagonalization of Eigen-matrices (JADE) algorithm [5].

A further extension of the cost function concept leads to the large class of Independent Component Analysis (ICA) algorithms which are based on different measures of the statistical independence of the factors [6–9].

2.4 Mean Reversion of Interest Rates

The mean reverting properties of interest rates are of importance if one is interested in a longer time horizon. According to Eqn. (1) the factors $s_i(t)$ have to be mean reverting. Another possibility would be to modify Eqn. (1). The mean reversion would then be treated outside the superposition of the factors. However, mean reverting factors seems to be natural. The simplest mean (zero) reverting model for the factors is

$$s_i(t) = (1-\alpha_i)s_i(t-1) + \sigma_i\xi_i(t) \tag{14}$$

where the mean reversion parameter $\alpha_i \geq 0$ and σ_i^2 is the variance of the factor fluctuations $\langle \xi_i(t)\rangle = 0$, $\langle \xi_i(t)^2\rangle = 1$.

2.5 Comparison of PCA and ICA applied to Interest Rates

Applying PCA to the interest rates or interest rate changes always results in similar factor loadings as depicted in Fig. 2.

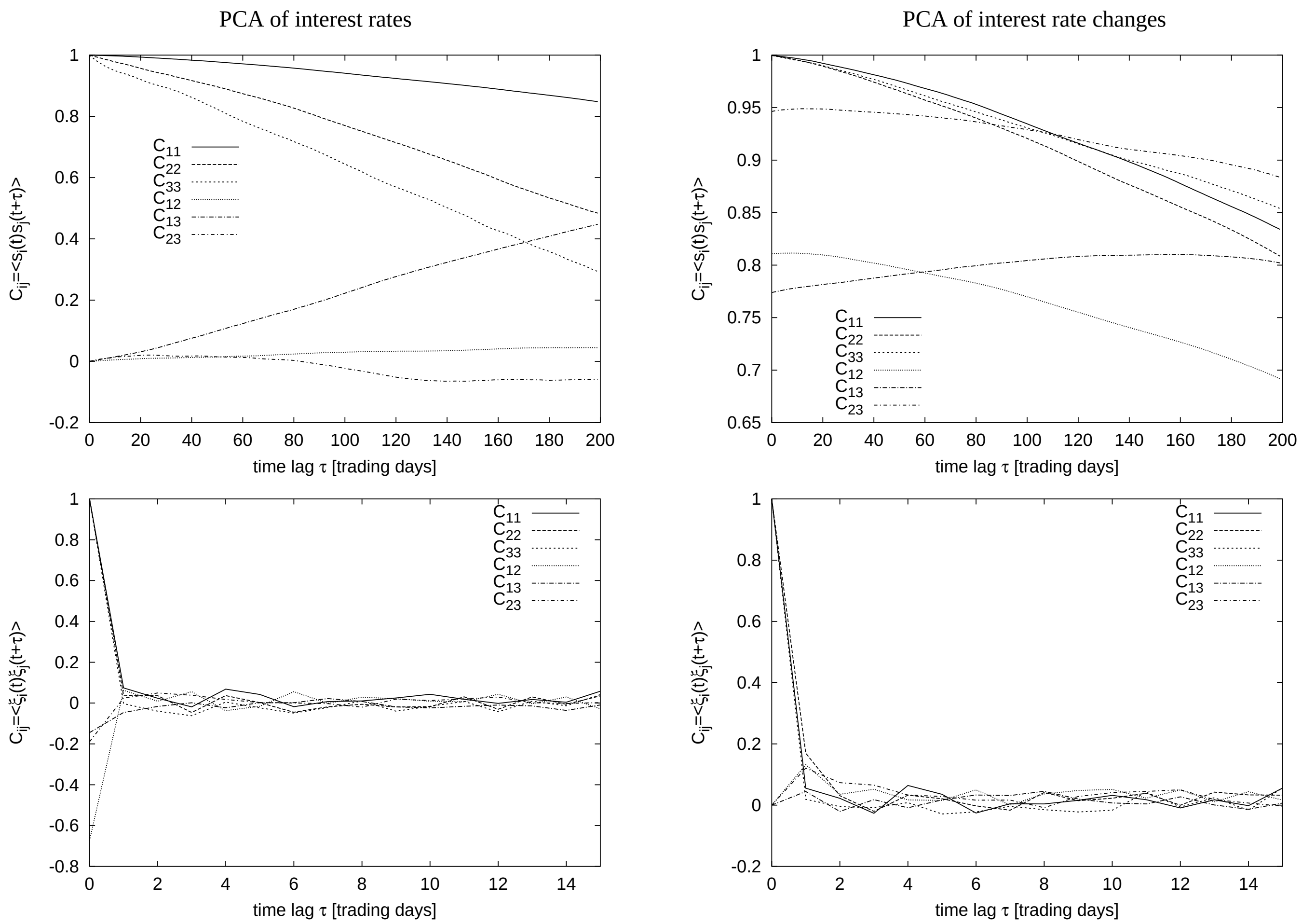

Fig. 5. *The PCA decorrelates either the slow factors $s_i(t)$ (left) or the fast factor fluctuations $\xi_i(t)$ (right).*

The benefit of the PCA is the natural interpretation of the factors [1]. However, due to the orthogonality of the factors and smoothness of the yield curve, there is not much freedom of choice for the qualitative picture. A further advantage of the PCA is the dimension reduction: only a few factors are sufficient to explain the total variation in the original data (see Table 1). Less important factors are faster mean reverting and may be interpreted as noise.

By construction, PCA decorrelates either the factors or the factor changes as shown in Fig. 5. The factors are in both cases far from independent. Particularly in the application on interest rate changes the factors are strongly correlated. The situation for the interest rates is slightly better. This is due to the fact that the autocorrelation functions of the interest rate changes are fast decaying. Hence factor changes are correlated for identical times only. Therefore one can expect better results in the application of the PCA on interest rates, if one is interested in a longer time horizon.

Remarkably, all tested ICA algorithms result in qualitatively similar factor loadings as depicted in Fig. 3. Since the factors are no longer orthogonal the interpretation is less natural. However, as for the PCA, only a few factors contribute significantly to the total variance of the data (see Table 1) and the less important factors are faster mean reverting and may be interpreted as noise. In contrast to the PCA, TDD decorrelates the factors and the factor changes simultaneously as shown in Fig. 6. Furthermore the crosscorrelations of the factors for larger time lags are smaller.

According to Eqn. (1) the inverse mixing matrix $\mathbf{W}^{-1}$ is the filter which has to applied in order to obtain the factors from the original time series of interest rates. Since for PCA the matrix is orthogonal, Fig. 2 shows also this filter. The filter of the TDD (Fig. 4) looks totally different and seems to extract more local yield curve anomalies. This may also explain the larger leptokurtic properties of the factor fluctuations $\xi_i(t)$ expressed by the kurtosis (Table 1) compared to those of the PCA.

3 Performance Analysis and Hedging

In order to test the advantage of ICA based factor extraction we have hedged two portfolios against each other and calculated the profit/loss (P&L) time series.

Both portfolios consist only of zero coupon bonds with the cash flow (from 1 to 10 years):

```
portf1 = [1000,0,0,0,0,0,0,0,0,-100]
portf2 = [0,0,0,0,200,0,0,0,0]   .
```

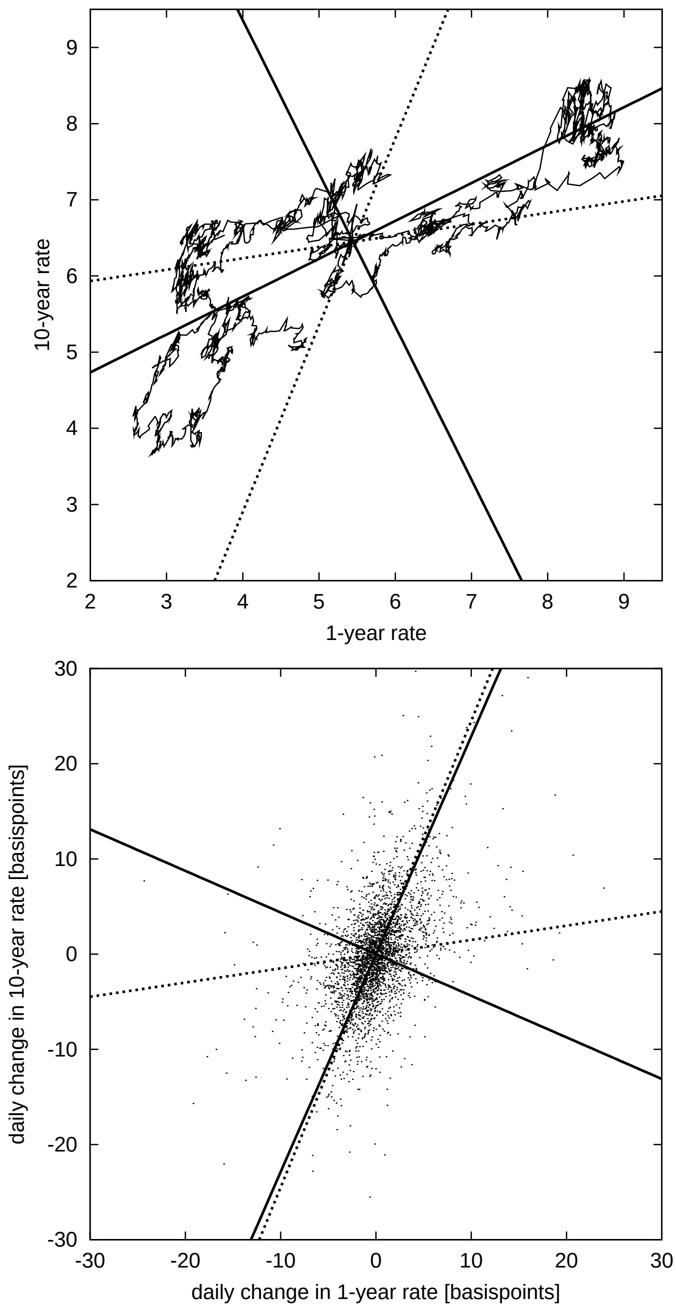

Fig. 1. *Illustration on a two dimensional dataset: The principal component transformation restates the data with reference to the rotated set of orthogonal axes (solid lines). The set of axes is noteably different for the interest rates (upper curve) and for the interest rate changes (lower curve). In contrast, the independent component transformation (by TDD) results in a unique set of axes (dashed lines) for the interest rates and interest rate changes.*

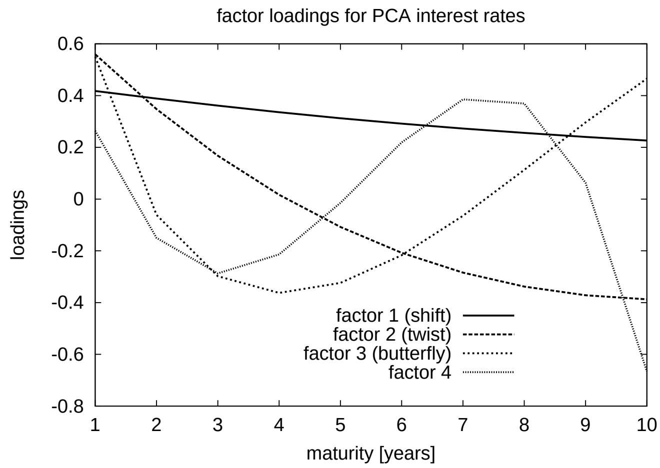

Fig. 2. *The loadings of the first four principal factors obtained from the interest rates have natural interpretation as common movements of the yield curve:* SHIFT, TWIST *and* BUTTERFLY. *The sum of the squares of the loadings of each factor is according to Eqn. (2) one.*

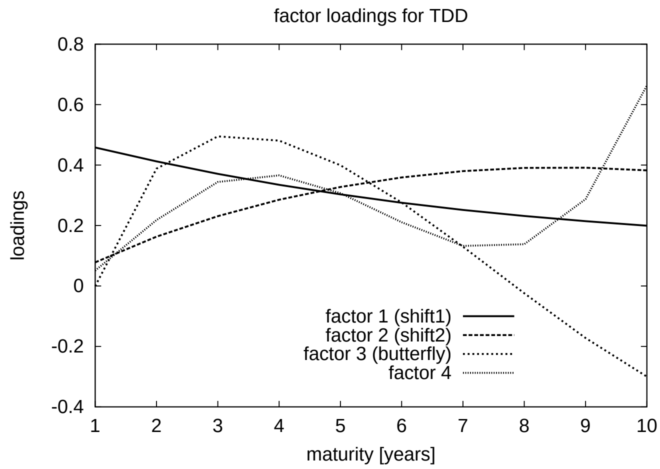

Fig. 3. *The loadings of the first four independent factors obtained by TDD have less natural interpretation compared to those of the PCA. The first two factors have all positive loadings and represent two different* SHIFTS *of the yield curve. The first factor is stronger for the short maturities, whereas the second factor emphasizes the long maturities.*

With this choice the first portfolio has a zero duration and is highly correlated with the TWIST. The second portfolio is highly correlated with the SHIFT.

Further the "rolling down the term structure" effects were neglected, i.e. the portfolios remain always the same.

To estimate the parameters of the factor model 1000-day rolling windows were used. Based on the obtained models monte carlo scenario simulations for the next 100 days (infrequent hedging) assuming Gaussian uncorrelated factor fluctuations were performed. In the scenario simulation only the first three most important factors were used to model the term structure movements. From the distribution of portfolio prices the minimal variance hedges neglecting mean reversion were calcu-

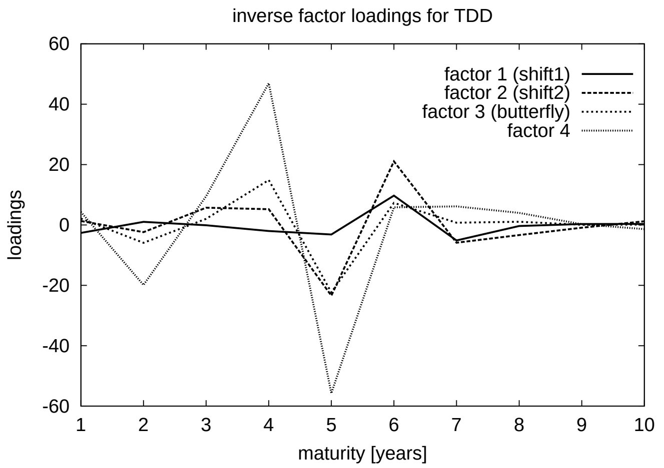

Fig. 4. *The demixing matrix $\mathbf{W}^{-1}$ obtained by TDD is sensitive to local yield curve anomalies.*

PCA of interest rates:

i	stddev(s_i)	α_i	σ_i	kurtosis(ξ_i)
1	4.678	0.0001	0.123	4.1
2	0.954	0.0020	0.055	7.1
3	0.225	0.0068	0.027	4.8
4	0.088	0.0314	0.022	10.2
5	0.040	0.0442	0.012	26.5

PCA of interest rate changes:

i	stddev(s_i)	α_i	σ_i	kurtosis(ξ_i)
1	4.441	0.0002	0.129	4.4
2	1.574	0.0003	0.041	6.8
3	0.690	0.0005	0.024	5.4
4	0.407	0.0010	0.020	38.7
5	0.084	0.0098	0.012	46.3

TDD-ICA:

i	stddev(s_i)	α_i	σ_i	kurtosis(ξ_i)
1	4.418	0.0000	0.074	16.3
2	1.791	0.0019	0.102	7.1
3	0.303	0.0065	0.035	10.5
4	0.169	0.0263	0.039	11.8
5	0.073	0.0408	0.021	33.9

Table 1. *Estimated parameters according to Eqn. (1) and (14) for the first five most important factors obtained by different algorithms: standard deviation of the factors, mean reversion parameter α_i, standard deviation of the factor fluctuations σ_i and kurtosis of the factor fluctuations.*

algorithm	P&L [bp]	variance
PCA-ir	0.38	0.65
PCA-irc	0.34	1.61
TDD	0.26	0.57
eTDD	0.24	0.41
JADE-ir	0.25	0.37
JADE-irc	0.16	0.81

Table 2. *Mean and variance of the P&L distribution for different hedging strategies in basis points (see Fig. 7).*

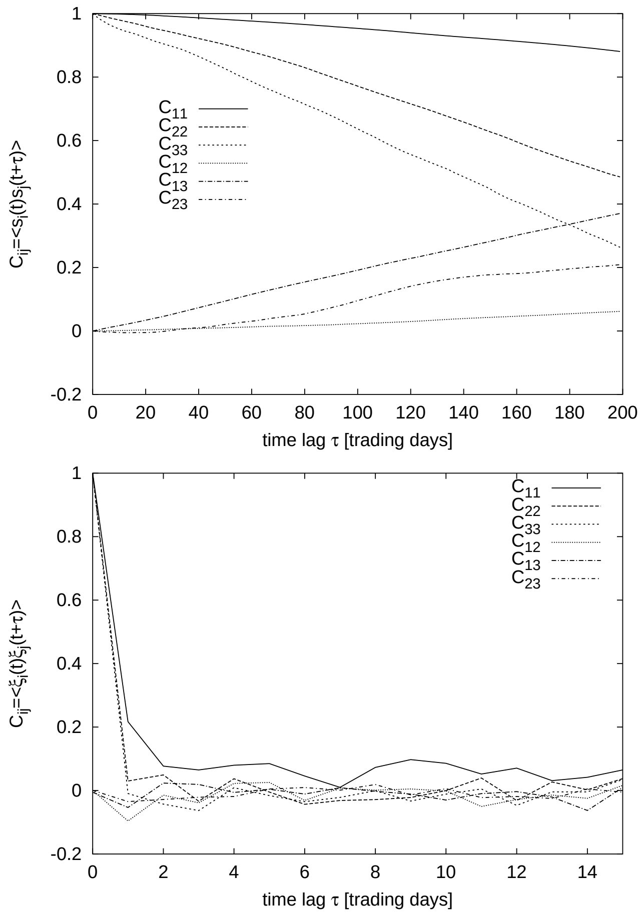

Fig. 6. *The TDD decorrelates the slow factors $s_i(t)$ and the fast factor fluctuations $\xi_i(t)$ simultaneously.*

lated. Finally using the historical interest rates the realized P&L were calculated.

The results for different algorithms are shown in Fig. 7 and Table 2.

First, the algorithms using the interest rate changes only show the largest P&L variances. This was already expected from the factor correlation functions.

Second, all ICA algorithms result in smaller variances compared to that of the PCA. The enhanced TDD and JADE have significantly smaller variances.

4 Conclusions

We have shown that applying ICA to interest rate time series is more appropriate than using PCA. For a long time horizon ICA based hedges outperforms those obtained by a PCA, i.e. have less P&L variance. Further, in that situation the long time correlations between the factors are of importance. Therefore it is more appropriate to use the interest rates itself, instead of the interest rate changes, as basis for the construction of the factor model.

The factor fluctuations obtained by an ICA have generally larger kurtosis compared to those obtained from a PCA. We expect that in some circumstances the estimation of the risk

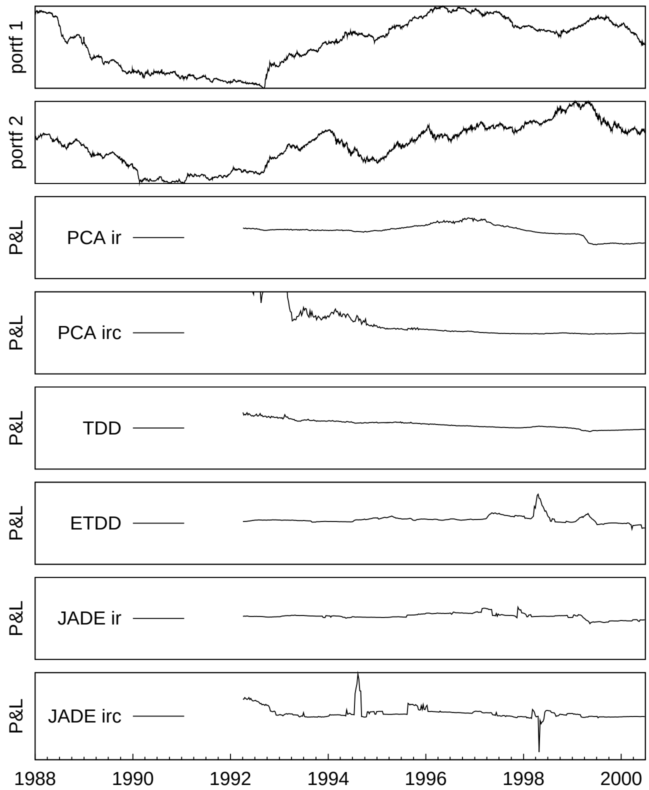

Fig. 7. *In the upper two curves the price movements of the two portfolios are shown. The lower curves show the Profit & Loss function of the hedged portfolios where the scales are identical. The hedges are based on (from upper to lower curve): PCA of interest rates, PCA of interest rate changes, TDD, enhanced TDD minimizing the factor crosscorrelations for the first 20 time lags, JADE [6] algorithm for the interest rates minimizing higher order factor crosscorrelations (cumulants) and JADE [6] algorithm for the interest rates changes.*

from a PCA may be significantly smaller than the true risk. However this conjecture is still not proven.

Since the real world is nonlinear it makes sense to design the cost function for the construction of the linear model for certain purposes. In the case of a long time horizon one can minimize the long time factor crosscorrelations. A different cost function may be optimal for option pricing purposes on a short time horizon.

Although we have concentrated here to interest rates only, this method is also applicable to stock prices or movements of the volatility term structure of option prices. In case of stock prices the time series in nonstationary. Therefore one has to apply the methods to the the price changes (returns). Due to the fast decaying autocorrelation functions of the returns the TDD is not applicable. However we suppose that higher order ICA will give better results compared to those from a PCA.

References

1. R. Litterman and J. Scheinkman, Journal of Fixed Income **1**, 54–61 (1991).
2. J. Frye, *Principals of Risk: Finding Value–at–Risk Through Factor–Based Interest Rate Scenarios*. In *VAR: Understanding and Applying Value at Risk*. London: Risk Publications, 275–288 (1997).
3. E. Galic and W. Kirschner, Solutions **2**, 43–54 (1998).
4. L. Molgedey and H. G. Schuster, Phys. Rev. Lett. **72**, 3634–3637 (1994).
5. J.-F. Cardoso and A. Souloumiac, SIAM J. Mat. Anal. Appl. **17**, 61–164 (1996).
6. J.-F. Cardoso, Neural Computation **11**, 157–192 (1999).
7. T. W. Lee, M. Girolami and T. J. Sejnowski, Neural Computation **11**, 409–433 (1999).
8. S. Amari, A. Cichocki and H. Yang, *A new learning algorithm for blind signal separationa*. In G. Tesauro, D Touretzky and T. Leen (eds), *Advances in Neural Information Processing Systems 8 (NIPS 95)*, The MIT Press, Cambridge, 757–763 (1996).
9. A. Bell and T. Sejnowski, Neural Computation **7**, 1129–1159 (1995).