

Analysis of Implied Volatility Surfaces

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Marina Schnellen
aus
Korbach

Göttingen, 2007

D7

Referent: Prof. Dr. Axel Munk

Koreferent: Prof. Dr. Martin Schlather

Tag der mündlichen Prüfung: 04.05.2007

Contents

1	Introduction	4
2	Selected Topics of Functional Data Analysis	9
2.1	Local Polynomial Kernel Estimation	9
2.2	Local Polynomial Kernel Estimation of Functional Data	14
2.3	Principal Component Analysis	17
3	Functional Variance Processes	22
3.1	The Model	22
3.2	Estimation of model components	23
3.3	Asymptotics	26
4	Financial Options and Implied Volatilities	30
4.1	Basic Properties of Stock Options	30
4.2	Black-Scholes Option Pricing	31
4.3	Implied Volatility	34
5	Analysis of Implied Volatilities	37
5.1	Introduction of the Data Set	37
5.2	Functional Data Analysis of the IVS	38
5.2.1	First steps of data analysis	38
5.2.2	Principal component analysis	46
5.3	Application of Functional Variance Processes	59
5.3.1	Index IV example: DAX30X	60
5.3.2	Stock IV example: VOLKSWAGEN	68
5.3.3	Swap IV example: JPY	74
5.4	Time series of functional variance processes	80
6	Summary and Open Questions	86

1 Introduction

The volatility of financial assets is an important parameter in risk management, portfolio trading and option pricing. The variability in price changes displays the uncertainty of the market and thus volatility is mainly determined by trading itself. There is much literature on volatility models mainly on stochastic models but during the last years more and more work is done on implied volatilities resulting from the well known Black-Scholes (BS) formula for option pricing, when the option price is known.

The Black-Scholes model, introduced by Black and Scholes [4] and Merton [35] in 1973, gives us a clear and easy model to price options on financial assets like stocks and stock indexes under only few, well-interpretable assumptions, such as log normality of asset price returns, constant trading and a constant volatility parameter. Last one is the crucial point in the model, as the volatility of underlying asset prices cannot be directly observed at the market. The model assumes a constant volatility across time, strike prices and option maturity. This assumption turned out to be wrong in practice. When determining the implied volatility from observed option prices according to the BS formula, we get a volatility function that is curved in strike and maturity dimension, i.e. the volatility smile and term structure, respectively. Implied volatility time series vary in trend and variance. Consequently the log normal assumption for asset returns cannot be correct and in fact the distribution of asset prices turns out to have fatter tails than the log normal distribution in practice.

In the past two decades much work has been done on the improvement of the Black-Scholes option pricing model. Degrees of freedom in the model are increased by introducing stochastic volatility, additional stochastic diffusion coefficients or jump intensities and amplitudes, see for example Cox et al. [10], Hull and White [26], Stein and Stein [48], Heston [24], Rubinstein [43], Bates [2], Derman [11], Barndorff-Nielsen [1] and Dumas et al. [12].

Although it is defective, the Black-Scholes model has become very popular and is largely applied in the financial world. Therefore, we do not want to follow the approaches to improve the model but to regard the resulting implied volatility as a state variable that reflects current market situations and thus is interesting by itself.

Implied volatilities are analysed as functions of the strike price K and option maturity τ . Traders are interested in how the shape of a volatility smile changes among options with different maturities or how the term structure changes when options move in-the-money or out-of-the-money. Further, variation of implied volatilities in time is of particular interest, see Xu and Taylor [53], Tompkins [49], Hafner and Wallmeier [20], Mixon [36] and Lee [33].

Steaming from options on the same underlying asset that are linked by various arbitrage relations, implied volatilities form a highly correlated multivariate system and cannot evolve independently for different strike and maturity values. Therefore the implied volatility surface (IVS) should be analysed as a whole by simultaneous examination of several slices of the surface. Literature deals with the number and shape of shocks that move the surface

across space and time, see for example Skiadopoulos et al. [45], Fengler et al. in [16], [17], [18], and Cont and da Fonseca [7].

It is common to display IV as a function of option maturity and moneyness κ , which is the strike divided by the future asset price, $\kappa = K/(S_t \exp(r\tau))$. Let $IV_t(\kappa, \tau)$ the implied volatility of an option with maturity τ and moneyness κ observed at time t . The main idea is that the IVS is analysed in terms of principal components (PCs) that reflect the shocks influencing IV movements in time. In practice, for options at each point in time only a limited number of option maturities exist whereas in the strike dimension a large number of options is available. Hence, options naturally fall into groups of different maturities τ_i , $i = 1, \dots, n$, and we sample the IVS as a set of IV smile curves $IV_t(\kappa, \tau_i)$ at day t .

In general, when modeling functional data like IV curves, methods of interest are nonparametric smoothing, like kernel or spline estimation, and principal component analysis. A common model is a heteroscedastic variance model. Data is assumed to be composed of a mean function m and an additional variance term,

$$Y(x) = m(x) + \sigma(x)\epsilon, \quad (1.1)$$

where $E[Y(x)] = m(x)$ and $Var[Y(x)] = \sigma^2(x)$ and ϵ is a random noise with zero mean and unit variance. Both, the mean and the variance are assumed to be functions of the predictor x . $m(x)$ is estimated by nonparametric smoothing methods, the residuals variance $\sigma^2(x)$ is analysed in terms of principal components.

According to the early work of Fengler et al. [17], single smile or smirk curves for different option maturities are driven by the same modes of variation in time, thus identic PCs are assumed or, if not fully identic, at least some identic PCs and some specific ones. The covariances of groups of smiles are assumed to have a common structure among the maturity groups. I.e. the space spanned by eigenvectors of the covariance matrix is assumed identical across groups while eigenvalues may vary.

It is common to use log IVs as the data is less skewed and potential outliers are scaled down. Being interested in daily changes of IVs, let

$$\sigma_t(\kappa, \tau) = \log IV_t(\kappa, \tau) - \log IV_{t-1}(\kappa, \tau)$$

the implied volatility log returns for $t = 1, \dots, T$. Let $Y_{ij} = [\sigma_t(\kappa_j, \tau_i)]_{t=1, \dots, T}$ the observed implied volatilities, nonparametrically smoothed onto a grid of option maturities τ_i , $i \in 1, \dots, n$, and moneyness κ_j , $j = 1, \dots, m$. Let

$$G_i = [Cov(Y_{ij}, Y_{ij'})]_{j, j'=1, \dots, m}$$

the $m \times m$ covariance matrices corresponding to option maturity groups τ_i . One assumes that G_i can be decomposed into $G_i = \Phi_i^T P_i \Phi_i$, where Φ_i is the matrix of eigenvectors and P_i is the diagonal matrix of ordered eigenvalues of G_i . When there is a common structure among option maturity groups one may assume a common principal components model (CPC)

$$H_{CPC} : G_i = \Phi^T P_i \Phi,$$

where $\Phi = \Phi_i$, $i = 1, \dots, n$, is the matrix of common eigenvectors and P_i the diagonal matrix of specific eigenvalues. Assumptions can be relaxed by presuming only $p \leq m - 2$ common eigenvectors and $m - p$ specific ones, i.e. the partial common principal components model (PCPC)

$$H_{PCPC(p)} : G_i = \Phi_{(i)}^T P_i \Phi_{(i)}.$$

Here $\Phi_{(i)} = (\Phi^C, \Phi_i^S)$, where Φ^C is the $m \times p$ matrix of common eigenvectors and Φ_i^S the $m \times (m - p)$ matrix of specific eigenvectors of a maturity group i . Thus a series of hierarchical models $PCPC(p) < PCPC(p - 1)$ is created, meaning that a model with $p - 1$ common PCs may contain a model with p common PCs. Note in $PCPC(p)$ the number of common PCs is less than $m - 2$ as by orthogonality of principal components $PCPC(m - 1) = PCPC(m)$ and thus falls into model H_{CPC} .

In different model selection tests Fengler et al. found that models with common principal components are preferred to standard models where the covariance structure is arbitrary as well as to models that assume linearly dependent covariances $G_i = a_i G_1 \forall i$. The main modes of daily variation of IV smiles were found to be all over up-and-down shifts of the IV curve and changes of the slope or curvature.

Cont and da Fonseca [7] generalized the method of analysing the IVS dynamics by common principal components by performing a functional PCA. Implied volatility surfaces $IV_t(\kappa, \tau)$ are assumed to be realisations of a stationary random process. To get smooth surfaces observed market data is nonparametrically smoothed. Functional IV log returns are denoted $Y_t(\kappa, \tau) = \sigma_t(\kappa, \tau)$. The corresponding covariance function $\mathcal{G}(x, x') = Cov(Y(x), Y(x'))$, $x = (\kappa, \tau)$, is the kernel of the integral operator $\mathbf{G}y(x) = \int \mathcal{G}(x, x')y(x')dx'$. The eigenproblem $\mathbf{G}\phi_k(x) = \rho_k\phi_k(x)$ is numerically solved by expanding eigenfunctions ϕ_k by a finite number of orthogonal basis functions. Finally implied volatilities are modeled according to

$$IV_t(x) = IV_0(x) \exp \left\{ \sum_k \zeta_k(t) \phi_k(x) \right\}$$

for PC scores $\zeta_k(t) = \int Y_t(x) \phi_k(x) dx$. The PCs $\phi_k(x)$ are now two dimensional surfaces, that describe the modes of daily log variation of the whole IVS in time. Cont and da Fonseca found that two or three PCs are sufficient to describe variation. The first PC corresponds to an overall shift in the level of the IVS, the second PC reflects a changing slope of the IVS in moneyness direction and the third PC reflects changes in the convexity of the surface.

Fengler et al. [18] note that due to a degenerated data design, i.e. observed implied volatilities usually appear in smile strings that move through the observation space, modeling biases may occur when initially smoothing the IVS and then performing a PCA using a functional norm. Instead they propose a semiparametric factor model that performs directly on the observed data. Log implied volatilities $Y_{ij} = \log IV(t_i, x_{ij})$, $x_{ij} = (\kappa_{ij}, \tau_{ij})$, are intra-day data where t_i is the day and $j = 1, \dots, J_i$ is an intra-day numbering. Data is

modeled according to

$$\begin{aligned} Y_{ij} &= m_0(x_{ij}) + \sum_{l=1}^L \beta_{il} m_l(x_{ij}) \\ &= \sum_{l=0}^L \beta_{il} m_l(x_{ij}), \end{aligned}$$

$\beta_0 = 1$. Functions $m_l : \mathbb{R}^2 \rightarrow \mathbb{R}$ are arbitrary except that they are chosen orthogonal and ordered according to maximal variance in β_l . β_l and m_l are estimated by local kernel smoothing methods under orthogonality constraints, using a functional norm that depends on day t_i . Thus m_l are not eigenfunctions of an operator as in standard functional PCA, and different models according to their number L of included smoothing functions are not necessarily nested for increasing L .

In practice, Fengler et al. result in a model with $L = 3$ basis functions m_l . Main shocks determining the IVS in time are proven to be shift effects and slope effects in moneyness direction as well as in maturity direction.

In this work we follow another approach to analyse variation of IVs. We are interested in how daily changes of IV smiles vary within the surface, i.e. for different option maturities. We apply a new method that combines smoothing techniques and principal component analysis to find a functional variance process (FVP) that determines variation among IV log return curves $Y_i(\kappa) = \sigma_t(\kappa, \tau_i)$, $i = 1, \dots, n$ for a fixed day t . The curves are regarded as random trajectories of a common stochastic variance process and thus as random phenomena themselves. The analysis of functional variance processes was introduced in 2006 by Müller et al. [34]. The concept of a variance function according to model 1.1 is extended to a random variance process model that includes random components for variation, when the observed data scatter randomly around a fixed regression function, and there is the need to model locally changing variances of data structures of increasing complexity.

Our findings are based on a data sets of daily implied volatility surfaces on stock, index and swap options. Data was kindly offered by the COMMERZBANK AG in Frankfurt a.M., and originally steam from the externals REUTERS and BLOOMBERG. IVs with respect to swap options are an exception among usual implied volatilities, as they depend on option maturity and on swap maturity instead of moneyness. Thus a swap IVS can be seen as consisting of several IV curves with respect to swap maturity, different IV curves correspond to option maturity groups.

The data show classic features. Stock and index IV have the typical smile or smirk shape in moneyness direction and in time IVs resemble stochastic processes. A discrete PCA confirms previous results in literature. The strength of variation of IV curves differs among option maturity groups and increases for decreasing maturities and moneyness values. The main modes of variation in time are similar among groups. Most influential modes are regular up-and-down shifts of the curves and slope and curvature changes.

In the analysis we detected structural differences between options with short and long time to maturity. Thus we decided to separate stock, index and swap data sets into two groups

in each case and determine functional variance processes separately for the groups. We use functional variance processes to describe variation within the surface of daily IV log returns. According to our empirical results FVPs are decreasing, after a short increase at the left boundary in some cases. Hence variation in the surface is stronger in parts where moneyness is small or swap maturity is short, respectively. Trajectories of the process are mainly driven by up-and-down shifts. Comparing FVPs for short and long option maturities the processes differ in means of slope and curvature. But in general modes of variation within the IV log return surface are similar for short and long option maturities.

Moreover, we determine FVPs of stock index and swap options at different points in time to get an impression of time dependent development of these processes and thus of time dependent changes of daily variation of the IVS. The shape of a FVP remains in time, meaning that modes of variation remain and our findings at fixed days are valid in general.

The work is set out as follows. In section 2, we present the basic principles of functional data analysis, namely local polynomial kernel estimation and principal component analysis. These two techniques are of main importance to estimate single components of the functional variance processes model, that is introduced in section 3. Consistency results for each of these components are given in section 3.3. In section 4, we summarize main principals of financial options and introduce the Black-Scholes formula for stock, index and swap option pricing. Finally in section 5, we analyse implied volatility surface dynamics for our data sets by performing a discrete PCA, following Fengler et al. [17], and further by applying the concept of functional variance processes. In section 6, we summarize and open questions are given for future research.

Acknowledgments

I am very grateful to my adviser, Prof. Dr. Axel Munk, for proposing the topic of my dissertation, for helpful suggestions during the research work and constant supervision. I wish to thank him and Prof. Dr. Manfred Denker, who encouraged and supported me to recover from impasses. Moreover, I would like to thank Prof. Dr. Martin Schlather for being the co-referee of my thesis.

I gratefully acknowledge financial backing from the Center for Statistics and the Graduiertenkolleg “Identifikation in mathematischen Modellen: Synergie stochastischer und numerischer Methoden”.

I attended my Ph.D. studies within the Ph.D. program “Applied Statistics and Empirical Methods” that offered me an excellent working environment and familiar surrounding by my colleagues and enabled me insight into different statistical areas on the one hand, and insight into a crossfire of foreign cultures and nations on the other hand. Many thanks go to Janis, Leif, Mikhail and Andreas, my office neighbors, for the familiar atmosphere they gave to me during my time in Göttingen. Special thanks to Marcus who backed me all over the time and stands my caprices and to my parents, grandparents and sister for mental support and understanding.

2 Selected Topics of Functional Data Analysis

One main task in statistics is the analysis of functional data, i.e. when the data are random curves, as for example growth curves or price curves of financial assets. Observed curves have similar shapes and we aim in finding mean trends and information on variation among the data. To extract relevant information from sampled curves popular methods are smoothing methods like kernel smoothing, splines or wavelets, and principal component analysis for data reduction and extraction of modes of variation. A good introduction into functional data analysis and applications is given in [39] and [40].

In this section we will first resume local polynomial kernel estimation for a random regressor Y with predictor X , that can either take fixed values or is random itself. Then we expand theory for functional data, where Y is a stochastic process with trajectories Y_i being functions on X . Afterwards we introduce principal components analysis, to determine the covariance structure of functional data and to find simple functional forms that describe the data structure.

2.1 Local Polynomial Kernel Estimation

The idea of kernel smoothing was first developed in the 1950th in terms of density estimation. Most popular literature in this context is from Rosenblatt [42] and Parzen [38], who introduced the kernel density estimator

$$\hat{f}(x, h) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i),$$

the Rosenblatt-Parzen estimator. It estimates the density function f of the random variable X from observations $X_i, i = 1, \dots, n$, weighted by a kernel function $K_h(x) = \frac{1}{h} K(x/h)$ with smoothing parameter $h > 0$, where usually K is a symmetric probability density function. When determining the functional relationship of two variables Y and X , where Y is the random response and X is the predictor variable, we assume the model

$$Y = m(X) + \sigma(X)\epsilon,$$

which is said to be fixed design if we observe Y_1, \dots, Y_n at fixed and ordered predictor values X_1, \dots, X_n in a compact interval $I \subset \mathbb{R}$. The model is called random, if X is random by itself with continuous, univariate density function f and with realisations in I . In the random case, observations are pairs, i.e. $(Y_i, X_i), i = 1, \dots, n$, are realisations of the random pair (Y, X) . m and σ are smooth functions on X and ϵ is a random variable with zero expectation and unit variance. Thus the expectation of Y at X is $m(X)$ and its variance is $\sigma^2(X)$.

There are many situations when it is not preferable to impose a parametric model

$$m(x) \in \{\beta_0 + \beta_1 x + \dots + \beta_p x^p \mid \beta_1, \dots, \beta_p \in \mathbb{R}\} \quad (2.1)$$

for some number $p \geq 0$. For example when there is no clear structure visible in the observations (Y_i, X_i) or when the data analysis is part of a procedure or algorithm that

works on different data structures, it might be disadvantageous to restrict on the parametric family in (2.1). Therefore we tie up to the kernel estimation idea and, assumed X is random, estimate m at x from the conditional expectation of Y given $X = x$, which is

$$E[Y|X = x] = \frac{\int y f(y, x) dy}{f(x)}, \quad (2.2)$$

where $f(x, y)$ is the common density of (Y, X) . According to the Rosenblatt-Parzen kernel density estimator expression (2.2) is then estimated by

$$\hat{m}(x) = \frac{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i}{\frac{1}{n} \sum_{i=1}^n K_h(x - X_i)}. \quad (2.3)$$

This estimator was first introduced in 1964 independently by Nadaraya [37] and by Watson [52], and fits into the family of local polynomial kernel estimators, that are presented now.

In local polynomial kernel estimation we locally fit polynomials to the data using weighted least squares. The estimator is motivated by the Taylor expansion of the mean function m at point x , which is

$$\begin{aligned} m(x) &= m(X_i) + m'(X_i)(x - X_i) + \frac{1}{2}m''(X_i)(x - X_i)^2 + \dots \\ &\quad + \frac{1}{p!}m^{(p)}(X_i)(x - X_i)^p + o(\|x - X_i\|^p) \\ &=: \beta_0 + \beta_1(X_i - x) + \beta_2(X_i - x)^2 + \dots + \beta_p(X_i - x)^p + o(\|X_i - x\|^p), \end{aligned}$$

assuming that m has continuous derivatives $m^{(k)}$, $k = 1, \dots, p$, $p \geq 0$. Thus estimate $m(x)$ by its Taylor expansion up to p -th degree by minimizing the sum of weighted least squares

$$RSS(x, p, h) = \sum_{i=1}^n K_h(X_i - x) \left\{ Y_i - \sum_{k=0}^p \beta_k (X_i - x)^k \right\}^2, \quad (2.4)$$

with respect to $\beta = (\beta_0, \dots, \beta_p)^T \in \mathbb{R}^{p+1}$. Let $\hat{\beta}$ the solution of the least squares problem then we estimate the mean function m at x by $\hat{m}(x, p, h) = \hat{\beta}_0$.

To give an explicit expression of $\hat{\beta}_0$ we rewrite (2.4) in matrix notation

$$RSS(x, p, h) = (Y - X_x \beta)^T W_x (Y - X_x \beta)$$

where $Y = (Y_1, \dots, Y_n)^T$ is the vector of observations, $W_x = \text{diag}\{K_h(X_i - x) | i = 1, \dots, n\}$ is the diagonal weighting matrix and X_x is the design matrix

$$X_x = \begin{pmatrix} 1 & (X_1 - x) & \dots & (X_1 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x) & \dots & (X_n - x)^p \end{pmatrix}.$$

Under the assumption that $(X_x^T W_x X_x)$ is invertible the solution is

$$\hat{m}(x, p, h) = \hat{\beta}_0 = e_1 (X_x^T W_x X_x)^{-1} X_x^T W_x Y, \quad (2.5)$$

the local polynomial kernel estimator of $m(x)$, where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^{1 \times (p+1)}$.

$$S(h) = e_1 (X_x^T W_x X_x)^{-1} X_x^T W_x \quad (2.6)$$

is denoted the hat matrix as it maps Y onto the estimate $\hat{Y} = \hat{m}(x, p, h)$. In local polynomial kernel estimation $S(h)$ is just a “hat vector” in $\mathbb{R}^{1 \times n}$. When $p = 0$ the explicit expression of (2.5) is

$$\hat{m}(x, 0, h) = \frac{1}{n} \sum_{i=1}^n \frac{K_h(x - X_i)}{\frac{1}{n} \sum_{j=1}^n K_h(x - X_j)} Y_i,$$

the local constant estimator that coincides the Nadaraya-Watson estimator (2.3). For $p = 1$ we get the local linear estimator

$$\hat{m}(x, 1, h) = \frac{1}{n} \sum_{i=1}^n \frac{\{s_2(x, h) - s_1(x, h)(X_i - x)\} K_h(X_i - x)}{s_2(x, h)s_0(x, h) - s_1(x, h)^2} Y_i,$$

with moments

$$s_r(x, h) = \frac{1}{n} \sum_{i=1}^n (X_i - x)^r K_h(X_i - x), \quad r = 0, 1, 2. \quad (2.7)$$

The local linear kernel estimator has several nice properties. It is easy to understand locally being a linear (parametric) model estimator. Further it has a well interpretable MSE, shown below, and nice boundary properties, i.e. the estimation bias remains the same for interior points and boundary points. Only the estimation variance increases at boundaries as a consequence of sparse data. A comparison of different kernel estimators is given by Wand and Jones [51], who result in proposing local polynomial kernel estimators with polynomial degree $p = 1$ or $p = 3$.

The mean squared error (MSE) is a local measure of goodness-of-fit controlling the estimation bias and variance.

$$\begin{aligned} MSE(\hat{m}(x, p, h)) &= E[\hat{m}(x, p, h) - m(x)]^2 \\ &= [Bias(\hat{m}(x, p, h))]^2 + Var(\hat{m}(x, p, h)), \end{aligned}$$

where the bias is defined as $Bias(\hat{m}(x, p, h)) = E[\hat{m}(x, p, h) - m(x)]$. We now give asymptotic bias and variance approximations for the fixed design model and local linear estimation. Asymptotics can be easily extended to higher degree polynomial estimation and to the random design model, see [51], by lightly strengthen the conditions.

Here, without loss of generality let $X_j \in [0, 1]$ for all j . Further assume

- m'', σ are continuous on $[0, 1]$,
- K is bounded, symmetric about 0 and supported on $[-1, 1]$,
- $h = h_n \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$,
- the estimation point x is an interior point, i.e. $h < x < 1 - h$.

Then the estimation bias is

$$\text{Bias}(\hat{m}(x, 1, h)) = \frac{1}{2}h^2 m''(x)\mu_2(K) + o(h^2) + O\left(\frac{1}{n}\right) \quad (2.8)$$

and the estimation variance is

$$\text{Var}(\hat{m}(x, 1, h)) = \frac{1}{nh} \|K\|^2 \sigma^2(x) + o\left(\frac{1}{nh}\right), \quad (2.9)$$

where $\sigma^2(x)$ is the variance of Y in x , $\mu_2(K) = \int_{-\infty}^{\infty} u^2 K(u) du$ and $\|K\|^2 = \int_{-\infty}^{\infty} K^2(u) du$. Proofs can be found in [51].

These terms are easy to interpret. What we would expect when linearly fitting a curve is that the estimation bias is small when the true curve m is close to linearity. In fact the bias depends on m'' which is the curvature of m . Hence when m is close to linearity m'' is small and the bias decreases. Further the bias depends on the kernel bandwidth h that reflects the amount of smoothing. For increasing h the bias increases. Thus the larger is h the worse is the fit, obviously, as we would “cut off the hills and rise the valleys” of m by a large amount of smoothing.

Now regard the variance that depends on the term $\frac{\|K\|^2}{nh}$. In general for kernel functions we assume $\|K\|^2 < \infty$, thus the term is approximately the reciprocal of the local sample size. Hence the estimation variance increases when the data is locally sparse. When n is fixed, a large bandwidth will decrease the variance. Further the estimation variance is penalized by large variance σ^2 .

Summarizing, a small bandwidth h will decrease the estimation bias, but it will increase the variance. Thus to get a small mean squared error, and thus a good estimate, we have to find a bandwidth that is small enough to ensure a small bias and that is big enough to ensure a small variance.

Once the polynomial degree and the kernel function are fixed, the local estimate is said to be optimal for that $h > 0$ that minimizes the MSE. In general, one aims in globally choosing an optimal bandwidth, i.e. fitting Y over the whole range of X using the same bandwidth. While MSE is only a local measure, the mean integrated squared error (MISE) provides a global one and thus is better suited for choosing h ,

$$\begin{aligned} \text{MISE}(\hat{m}(x, p, h)) &= E\left[\int (\hat{m}(x, p, h) - m(x))^2 dx\right] \\ &= \int \text{MSE}(\hat{m}(x, p, h)) dx \end{aligned}$$

by changing the order of integration according to Fubini's Theorem, see [3]. Unfortunately MSE and MISE and their asymptotic approximations depend on the unknown mean function m and variance function σ^2 . In case of a random design additionally the predictor's density f is unknown. Therefore we cannot directly apply the above measures to choose optimal bandwidths. In practice it is often convenient to choose the bandwidth by visual inspection of the data and different estimators. But there are circumstances where we need

automatical bandwidth selection, for example when we do not have a feeling for the true mean function m or in situations when it is just too time consuming to choose h by eye. We then use approximations of the MSE or MISE to automatize bandwidth selection, as for example in cross validation. The cross validation term is

$$CV(h) = \sum_{i=1}^n (Y_i - \hat{m}^{(-i)}(X_i, p, h))^2$$

where $\hat{m}^{(-i)}(X_i, p, h)$ is the local polynomial kernel estimator obtained from observations Y_j , $j \neq i$. $CV(h)$ is thus the sum of squared residuals resulting from estimating $m(X_i)$ when the i -th observation Y_i is unknown. The optimal smoothing parameter h_{opt} is chosen to minimize $CV(h)$. Justification of cross validation bandwidth selection in case of functional data is given in section 2.2.

Some notes on kernel functions. In general to ensure asymptotic properties a kernel $K : \mathbf{R} \rightarrow \mathbf{R}$ is assumed to satisfy

1. $K(x) \geq 0$ and $K(x) = K(-x) \forall x \in \mathbf{R}$ (nonnegative and symmetric)
2. $\sup_{x \in \mathbf{R}} K(x) < \infty$ (bounded)
3. $\int K(x) dx = 1$ (normed)
4. $|x|K(x) \rightarrow 0$ for $|x| \rightarrow \infty$, $\int x^2 K(x) dx < \infty$

where \int means $\int_{-\infty}^{\infty}$. When choosing K to be a symmetric probability density function items 1, 2 and 3 are guaranteed. Item 1 supports the estimation idea, i.e. observations with the same distance to the value to estimate are equally weighted, and it ensures positive estimates when the data is positive. By symmetry of K and items 2 and 3 it follows

$$\int xK(x)dx = 0 \text{ and } \int K^2(x)dx < \infty,$$

which lead to attractive MSE properties as shortly presented above. Frequently used kernel functions are

$$\begin{aligned} K(x) &= \frac{1}{2}I_{(-1,1)}(x) && \text{Uniform Kernel} \\ K(x) &= (1 - |x|)_+ && \text{Triangular Kernel} \\ K(x) &= \frac{3}{4}(1 - x^2)_+ && \text{Epanechnikov Kernel} \\ K(x) &= \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} && \text{Gaussian Kernel} \end{aligned}$$

where $I_{(-1,1)}(x) = 1$ iff $x \in (-1, 1)$ and $I_{(-1,1)}(x) = 0$ elsewhere. $(\cdot)_+$ denotes the positive part.

Note that all these kernels except Gaussian have bounded support $[-1, 1]$. It is well-known, that the choice of kernel functions has only little impact onto the estimate, although theoretical results show that the Epanechnikov kernel has best asymptotic properties, see [51]. In practice, when data is locally sparse bounded support kernels may lead to gaps in the

estimated mean function $\hat{m}(x, p, h)$. When there is no observation in a h -neighborhood of estimation point x we cannot estimate $m(x)$ from (2.4) as weights are zero everywhere. We then locally need larger bandwidths, or we avoid the problem at the outset by choosing a kernel with support \mathbb{R} , like Gaussian kernel, that gives (at least little) weight to each observation. But note also, that bounded support kernels bind computational effort. Especially when sample size m is large, one may prefer a bounded support kernel which gives no weight to observations outside the range determined by h .

2.2 Local Polynomial Kernel Estimation of Functional Data

In general, when observed data are curves we sample the curves over a discrete grid in the prediction range. Let Y a stochastic process with trajectories Y_i , $i = 1, \dots, n$, observed at discrete prediction points $X_i = (X_{i1}, \dots, X_{im})$. Denote $Y_i(X_{ij}) = Y_{ij}$. We assume for each trajectory Y_i the number of observations m is equal and prediction points remain for different i , i.e. $(X_{i1}, \dots, X_{im}) = (X_1, \dots, X_m)$ for all i . If in practice curves are observed at different prediction points we obtain observations onto a fixed grid by linear interpolation. Note, that for good estimation results concerning asymptotic properties the discrete grid has to be sufficiently dense and equidistant. We assume the model

$$Y_{ij} = m(X_j) + \sigma(X_j)\epsilon_i, \quad (2.10)$$

$i = 1, \dots, n$, $j = 1, \dots, m$, where ϵ_i are iid random variables with zero mean and unit variance. The expectation and variance of Y_{ij} given X_j are

$$E[Y_{ij}|X_j] = m(X_j) \quad \text{and} \quad \text{Var}(Y_{ij}|X_j) = \sigma^2(X_j)$$

and the covariance is

$$\text{Cov}(Y_{ij}, Y_{kl}|X_j, X_l) = \delta_{ik}\gamma(X_j, X_l)$$

with $\delta_{ik} = 0$ if $i \neq k$ and $\delta_{ii} = 1$, resulting from independence of Y_i and Y_k . The mean function $m(x)$ and variance function $\sigma(x)$ are assumed to be smooth.

To estimate the mean function m we use local polynomial kernel estimation, that is analogous to estimation of random variables Y that depend on a prediction variable X . We estimate $m(x)$ by minimizing the weighted sum

$$RSS(x, p, h) = \sum_{i=1}^n \sum_{j=1}^m K_h(X_j - x) \{Y_{ij} - P_\beta(X_j - x)\}^2 \quad (2.11)$$

with respect to β , where $P_\beta(x) = \sum_{k=0}^p \beta_k x^k$ is a polynomial of degree $p \geq 0$, K is a symmetric kernel function and $h > 0$ is the smoothing parameter. Let $\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_p)$ the solution then $\hat{m}(x, p, h) = \hat{\beta}_0$. The bandwidth is chosen by one-curve-leave-out cross validation as introduced by Rice and Silverman [41], where instead of leaving out single observation points Y_{ij} we leave out whole curves Y_i in the cross validation term

$$CV(h) = \sum_{i=1}^n \sum_{j=1}^m (Y_{ij} - \hat{m}^{(-i)}(X_i, p, h))^2. \quad (2.12)$$

The estimate $\hat{m}^{(-i)}(X_i, p, h)$ results from minimizing

$$RSS^{(-i)}(X_i, p, h) = \sum_{k \neq i} \sum_j K_h(X_{kj} - X_j)(Y_{kj} - P_\beta(X_{kj} - X_j))^2$$

with respect to $h > 0$. Examples for usual and one-curve-leave-out cross validation when fitting functional data are given in figure 1. We fitted residuals of implied volatility log returns of DAX30X options by local linear estimation. (Data will be presented in detail in section 5.) Four different residual curves were discretely observed. In the upper four panels these curves are estimated using bandwidths that were independently chosen for each curve by usual CV. In the lower four panels we draw the estimated mean curve that is determined by local linear kernel estimation according to (2.11) using a bandwidth chosen by one-curve-leave-out CV.

As already mentioned in section 2.1, $CV(h)$ is an approximation of the MISE,

$$MISE(\hat{m}(x, p, h)) = \int E \left[(\hat{m}(x, p, h) - m(x))^2 \right] dx,$$

and thus is suited to choose an appropriate bandwidth h . Note therefore that for the functional model (2.10) the expected one-curve-leave-out CV term is

$$\begin{aligned} E \left[\frac{1}{n} CV(h) \right] &= E \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m (Y_{ij} - \hat{m}^{(-i)}(X_j, p, h))^2 \right] \\ &= \sum_{j=1}^m E \left[\left(Y_{nj} - \hat{m}^{(-n)}(X_j, p, h) \right)^2 \right] \end{aligned} \quad (2.13)$$

$$\begin{aligned} &= \sum_{j=1}^m E \left[(Y_{nj} - m(X_j))^2 \right] + \sum_{j=1}^m E \left[\left(\hat{m}^{(-n)}(X_j, p, h) - m(X_j) \right)^2 \right] \\ &\quad - 2 \sum_{j=1}^m E \left[(Y_{nj} - m(X_j)) \left(\hat{m}^{(-n)}(X_j, p, h) - m(X_j) \right) \right] \\ &= tr(G) + M_{n-1}(h). \end{aligned} \quad (2.14)$$

(2.13) holds as Y_i , $i = 1, \dots, n$, (and in random case X_j , $j = 1, \dots, m$) are identically distributed. The expectation in (2.14) is zero as Y_{nj} is independent of $\hat{m}^{(-n)}(X_j, p, h)$ and $E[Y_{nj} - m(X_j)] = 0$. Here $G = [G(X_{j_1}, X_{j_2})]_{j_1, j_2=1, \dots, m}$ is the covariance matrix of $Y_i = (Y_{i1}, \dots, Y_{im})^T$ thus the trace $tr(G)$ is the sum of variances $\sigma^2(X_j)$ and thus finite. Hence, minimizing the expected CV term with respect to h is equivalent to minimizing $M_{n-1}(h) = \sum_{j=1}^m E \left[\left(\hat{m}^{(-n)}(X_j, p, h) - m(X_j) \right)^2 \right]$. Let $h_{opt} = \arg \min_{h>0} M_{n-1}(h)$. If the sample size n is large enough, h_{opt} will approximately achieve the minimum of $M_n(h) = \sum_{j=1}^m E \left[\left(\hat{m}(X_j, p, h) - m(X_j) \right)^2 \right]$, which is the Rieman approximation of $MISE(\hat{m}(x, p, h))$ up to some constant factor. Thus h_{opt} approximately minimizes the MISE.

The cross validation procedure is computationally very expensive. Let \mathcal{A} the algorithm to determine a local polynomial estimator $\hat{m}(x, p, h)$. In (2.12) we have to run \mathcal{A} nm

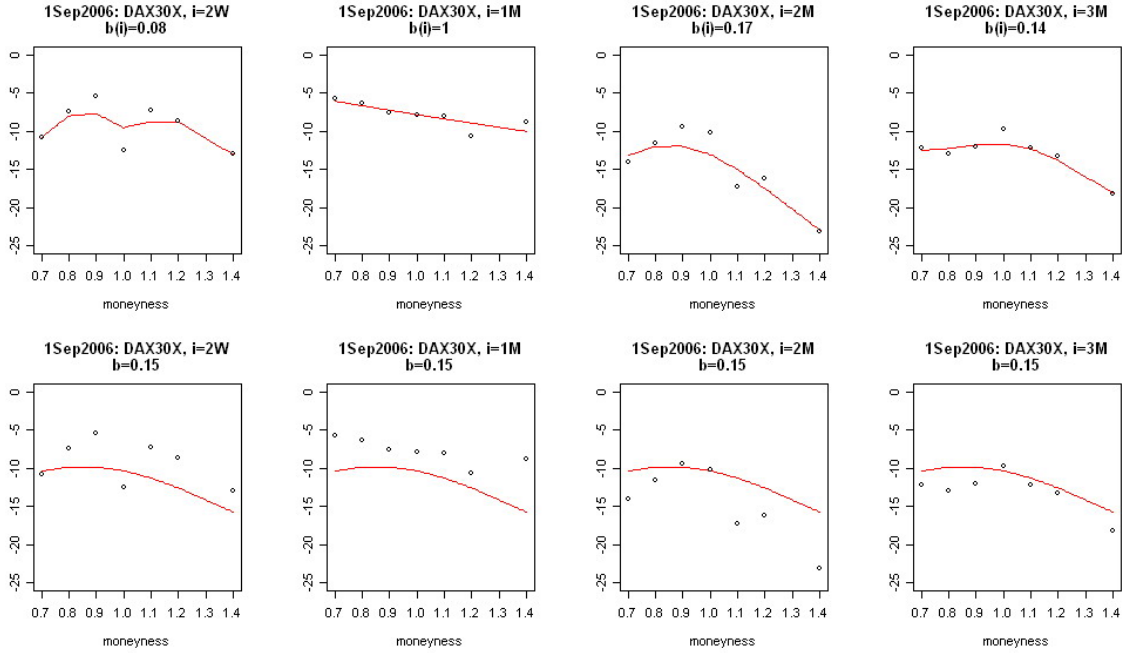


Figure 1: Local linear kernel estimation of residual curves of DAX30X IV log returns at 1 September 2006 according to option maturities $\tau_i \in \{2W, 1M, 2M, 3M\}$. Observed data and smooth estimates of single curves are shown in top panels. Bandwidths were chosen by cross validation. Observed data and estimated mean curve is shown in bottom panels. The bandwidth is chosen by one-curve-leave-out cross validation.

times. It is possible to improve computational efficiency to m runs of \mathcal{A} . Let therefore $\bar{Y}_j = \frac{1}{n} \sum_{i=1}^n Y_{ij}$ and $\bar{Y} = (\bar{Y}_1, \dots, \bar{Y}_m)$. Since

$$\begin{aligned}
 RSS(x, p, h) &= \sum_{i=1}^n \sum_{j=1}^m K_h(X_j - x) \{Y_{ij} - P_\beta(X_j - x)\}^2 \\
 &= n \sum_{j=1}^m K_h(X_j - x) (\bar{Y}_j - P_\beta(X_j - x))^2 \\
 &\quad - n \sum_{j=1}^m K_h(X_j - x) \bar{Y}_j^2 + \sum_{j=1}^m \sum_{i=1}^n K_h(X_j - x) Y_{ij}^2 \quad (2.15)
 \end{aligned}$$

minimizing $RSS(x, p, h)$ with respect to β is equivalent to minimizing

$$RSS^*(x, p, h) = \sum_{j=1}^m K_h(X_j - x) (\bar{Y}_j - P_\beta(X_j - x))^2,$$

as terms in (2.15) do not depend on β . Let now $S^*(h, j) = (S^*(h, j)_1, \dots, S^*(h, j)_m)$ the hat vector resulting from minimizing $RSS^*(X_j, p, h)$, that maps \bar{Y} onto

$$\hat{m}(X_j, p, h) = S^*(h, j) \bar{Y} = \sum_l S^*(h, j)_l \bar{Y}_l,$$

analogous to (2.6). The hat vector $S^*(h, j)$ remains when we exclude trajectory Y_i from the estimation procedure, thus $\hat{m}^{(-i)}(X_j, p, h) = S^*(h, j)\bar{Y}^{(-i)}$ and $\bar{Y}_j^{(-i)} = \frac{1}{n-1} \sum_{k \neq i} Y_{kj}$. We further have

$$\begin{aligned} \hat{m}^{(-i)}(X_j, p, h) &= \sum_l S^*(h, j)_l \bar{Y}_l^{(-i)} \\ &= \sum_l S^*(h, j)_l \left\{ \frac{1}{n-1} (n\bar{Y}_l - Y_{il}) \right\} \\ &= \sum_l S^*(h, j)_l \left\{ \bar{Y}_l + \frac{1}{n-1} (\bar{Y}_l - Y_{il}) \right\} \\ &= \hat{m}(X_j, p, h) + \frac{1}{n-1} \underbrace{\left\{ \sum_l S^*(h, j)_l Y_{il} - \hat{m}(X_j, p, h) \right\}}_{\hat{Y}^{(i)}}, \end{aligned} \quad (2.16)$$

where $\hat{Y}^{(i)}$ is obtained by smoothing just the i -th curve using the hat vector $\hat{S}^*(h, j)$. We use algorithm \mathcal{A} to compute the hat vector $S^*(h, j)$. Thus, we can determine $\hat{m}^{(-i)}(X_j, p, h)$ for all $i \in \{1, \dots, n\}$ running \mathcal{A} only once. Therefore, substituting equation (2.16) in formula (2.12), we can compute $CV(h)$ running \mathcal{A} only m times.

2.3 Principal Component Analysis

After extraction of the mean function in the functional model (2.10) our main purpose is the analysis of the covariance structure of the data. Therefore we determine principal components (PCs) of the data that provide a physical interpretation and simple description of main features of variability of the data. Besides feature extraction, PCs are used for data reduction and prediction and PC scores serve for data clustering and classification. For further reading one is referred to [29].

Let $\mathcal{G} : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$\mathcal{G}(x, x') = Cov(Y(x), Y(x')),$$

the covariance function of Y . We assume that there exists an orthogonal decomposition of \mathcal{G} according to

$$\mathcal{G}(x, x') = \sum_k \rho_k \phi_k(x) \phi_k(x'),$$

$k = 1, 2, \dots$, where $\rho_1 \geq \rho_2 \geq \dots \geq 0$ and $\rho_k \xrightarrow{k \rightarrow \infty} 0$, and $\phi_k : \mathbb{R} \rightarrow \mathbb{R}$ are smooth, orthogonal functions in terms of

$$\langle \phi_{k_1}, \phi_{k_2} \rangle = \int \phi_{k_1}(x) \phi_{k_2}(x) dx = 0, \quad k_1 \neq k_2$$

and $\langle \phi_k, \phi_k \rangle = \|\phi_k\|^2 \neq 0$. From the observed residuals $Y_{ij}^* = Y_{ij} - \hat{m}(X_j, p, h)$, $i = 1, \dots, n$, $j = 1, \dots, m$, we estimate \mathcal{G} for example discretely by the sample covariance matrix $G = [G(j_1, j_2)]_{j_1, j_2=1, \dots, m}$,

$$G(j_1, j_2) = \frac{1}{n} \sum_{i=1}^n Y_{ij_1}^* Y_{ij_2}^*.$$

For the discrete PCA we display functions $f : \mathbb{R} \rightarrow \mathbb{R}$ as data vectors $f = (f_1, \dots, f_m)^T = (f(x_1), \dots, f(x_m))^T \in \mathbb{R}^m$ and define $\langle f, g \rangle = \sum_{j=1}^m f_j g_j$ and $\|f\|^2 = \langle f, f \rangle$. The first principle component is the weight vector $\phi_1 = (\phi_{11}, \dots, \phi_{1m})^T$ for which

$$\zeta_{i1} = \langle \phi_1, Y_i^* \rangle = \sum_{j=1}^m \phi_{1j} Y_{ij}^*, \quad i = 1, \dots, n,$$

have largest possible mean square $\frac{1}{n} \sum_{i=1}^n \zeta_{i1}^2$ subject to $\|\phi_1\|^2 = 1$ (normality). ζ_{i1} are called PC scores. The normality assumption controls the scores variance, that without the assumption could take arbitrary values. We may change the sign of ϕ_1 without changing the result. Further PCs ϕ_k are defined to maximize $\frac{1}{n} \sum_{i=1}^n \zeta_{ik}^2$,

$$\zeta_{ik} = \langle \phi_k, Y_i^* \rangle = \sum_{j=1}^m \phi_{kj} Y_{ij}^*, \quad i = 1, \dots, n,$$

subject to normality constraint $\|\phi_k\|^2 = 1$ and orthogonality constraint

$$\langle \phi_k, \phi_{k'} \rangle = 0, \quad \text{for all } k' < k.$$

Note that $\frac{1}{n} \sum_{i=1}^n \zeta_{ik}^2$ decrease for increasing k .

We determine principal components from the eigendecomposition of the covariance matrix G . Note therefore that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \zeta_{ik}^2 &= \frac{1}{n} \sum_{i=1}^n \sum_{j_1, j_2=1}^m \phi_{kj_1} Y_{ij_1}^* Y_{ij_2}^* \phi_{kj_2} \\ &= \frac{1}{n} \phi_k^T (Y^*)^T Y^* \phi_k \\ &= \phi_k^T G \phi_k \end{aligned} \tag{2.17}$$

where $Y^* = [Y_{ij}^*] \in \mathbb{R}^{n \times m}$ is the residual matrix with rows $(Y_i^*)^T$. Maximizing the scores variance is thus equivalent to maximizing $\phi_k^T G \phi_k$ with respect to orthonormality of ϕ_k . This last maximization problem is known to be solved by finding ordered eigenvalues ρ_k and according eigenvectors ϕ_k of G , i.e. solve

$$G \phi_k = \rho_k \phi_k$$

simultaneously for all k , or equivalently solve

$$G = \Phi^T P \Phi, \tag{2.18}$$

where Φ the $(m \times K)$ -matrix with orthonormal columns ϕ_k , and $P = \text{diag}\{\rho_k, k = 1, \dots, K\}$ the diagonal matrix of ordered eigenvalues.

The number of nonzero eigenvalues and thus the number of PCs is bounded by the rank of

the covariance matrix G . Y_i^* are centered vectors, thus the rank of $Y^* \in \mathbb{R}^{n \times m}$ is at most $n - 1$ and further the rank of $G = \frac{1}{n}(Y^*)^T Y^* \in \mathbb{R}^{m \times m}$ is at most $\min(m, n - 1)$, i.e.

$$\text{rk}(G) \leq \min(m, n - 1).$$

Therefore the number of nonzero eigenvalues and according eigenvectors ρ_k and ϕ_k and thus the number of PCs is bounded by $\text{rk}(G)$.

By orthonormality of ϕ_k , $k = 1, 2, \dots$, the equation

$$\begin{aligned} \phi_k^T \sum_l \zeta_{il} \phi_l &= \phi_k^T \sum_l \langle Y_i^*, \phi_l \rangle \phi_l \\ &= \langle Y_i^*, \phi_k \rangle \\ &= \phi_k^T Y_i^*, \end{aligned}$$

holds for all k , thus $Y_i^* = \sum_l \zeta_{il} \phi_l = \Phi \zeta_i$. Therefore, to reduce data we may approximate Y_i^* by a linear combination of the first $K \leq \text{rk}(G)$ principle components weighted by PC scores ζ_{ik} , $k = 1, \dots, K$, i.e.

$$Y_i^*(K) = \sum_{k=1}^K \zeta_{ik} \phi_k.$$

The number K of PCs used for approximation is chosen such that the variance explained by PC 1 to PC K is at least $\alpha \cdot 100\%$ of the total variance, i.e.

$$\frac{\sum_{i=1}^n \|Y_i^*(K)\|^2}{\sum_{i=1}^n \|Y_i^*\|^2} \geq \alpha, \quad \alpha \in (0, 1].$$

Depending on purposes in practical application α may be for example 0.95 or 0.99.

In terms of geometry the PC score vector $\zeta_k = Y^* \phi_k$ is the projection of the centered data onto the space defined by ϕ_k , thus $E[\zeta_{ik}] = 0$. From (2.17) we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \zeta_{ik}^2 &= \phi_k^T G \phi_k \\ &= \phi_k^T \rho_k \phi_k = \rho_k, \end{aligned}$$

hence $\text{Var}[\zeta_{ik}] = \rho_k$. The first K principal components define the directions of main variation of observations Y_{ij} . We can define a contour of constant probability by the m -dimensional ellipsoid $g(y) = y^T G^{-1} y = C_1$, $y \in \mathbb{R}^m$, where C_1 is a constant number. Note that

$$\begin{aligned} g(Y_i) &= \zeta_i^T \Phi^T G^{-1} \Phi \zeta_i \\ &= \zeta_i^T \Phi^T \Phi P^{-1} \Phi^T \Phi \zeta_i \\ &= \zeta_i^T P^{-1} \zeta_i \\ &= \sum_{k=1}^n \frac{\zeta_{ik}^2}{\rho_k}, \end{aligned}$$

where the second line holds by (2.18) and $\Phi^{-1} = \Phi^T$. Thus the ellipsoid is defined by

$$g(Y_i) = \sum_{k=1}^n \frac{\zeta_{ik}^2}{\rho_k} = C_1,$$

which is the construction of an ellipsoid referred to its principal axes.

For PCA in a continuous setting, we define the integral transformation

$$\mathbf{G}\phi(x') = \int \mathcal{G}(x', x)\phi(x)dx = \langle \mathcal{G}(x', \cdot), \phi \rangle$$

with kernel $\mathcal{G}(x', x) = \text{Cov}(Y(x'), Y(x))$, where Y is a stochastic process with observed trajectories Y_i . Smooth eigenfunctions $\phi_k : \mathbb{R} \rightarrow \mathbb{R}$ of the covariance operator $\mathbf{G} : \phi_k \rightarrow \int \mathcal{G}(\cdot, x)\phi_k(x)dx$ solve

$$\mathbf{G}\phi_k = \rho_k\phi_k,$$

with respect to $\langle \phi_{k_1}, \phi_{k_2} \rangle = \delta_{k_1 k_2}$, where the series of ordered positive eigenvalues ρ_k converge to zero. Here we have no bound due to the number m of observations per subject as in the discrete setting, such that

$$\text{rk}(\mathbf{G}) \leq n - 1$$

and equality holds if observed curves Y_i are linearly independent.

In figures 2 and 3 examples for a PCA are given. We plotted the estimated covariance matrix of the residuals of implied volatility log returns of DAX30X options, corresponding PCs and the effects of PCs on the mean data curve. The first three PCs explain about 99% of the total variation and thus determine the main modes of variation. PC 1 is overall positive, i.e. almost 80% of the variation among data curves $Y_i(x)$ is due to overall up-and-down shifts. Corresponding effects are shown in the upper right panel of figure 3, where the mean curve $m(x) = E[Y_i(x)]$ is plotted and $m(x) \pm C_2\phi_1(x)$. To better visualise the effects we introduced the constant factor $C_2 > 0$. The shifts are lightly stronger pronounced at the left and right border. PC 2 is negative over $I = [0.7, 1.2)$ and approximately zero at the right border. Thus second most variation is explained by further up-and-down shifts which are restricted to the interval I , see corresponding plot of the effects of PC 2. The third PC forms two contrasts, one over $[0.7, 1.1)$ and one over $[1.1, 1.4]$, meaning that data curves vary by positive shifts at the left side and negative shifts at the right side and vice versa. Thus a little part of variation is due to curvature changings of the left halves and the right halves of the curves in opposite directions.

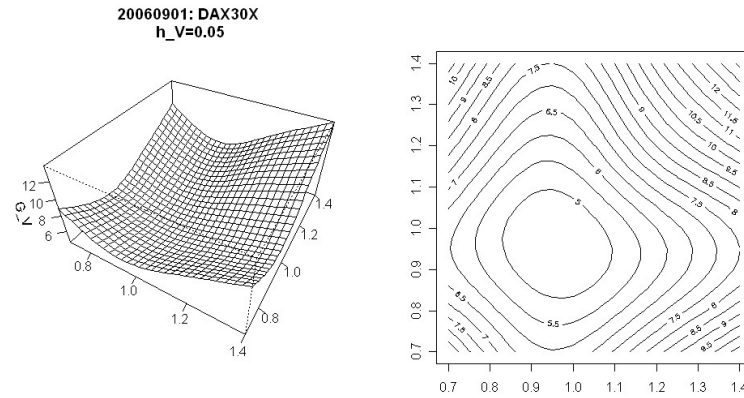


Figure 2: Covariance of residual curves of DAX30X IV log returns at 1 September 2006 according to option maturities $\tau_i \in \{2W, 1M, 2M, 3M\}$. Left panel shows the surface plot, right panel the corresponding contour plot.

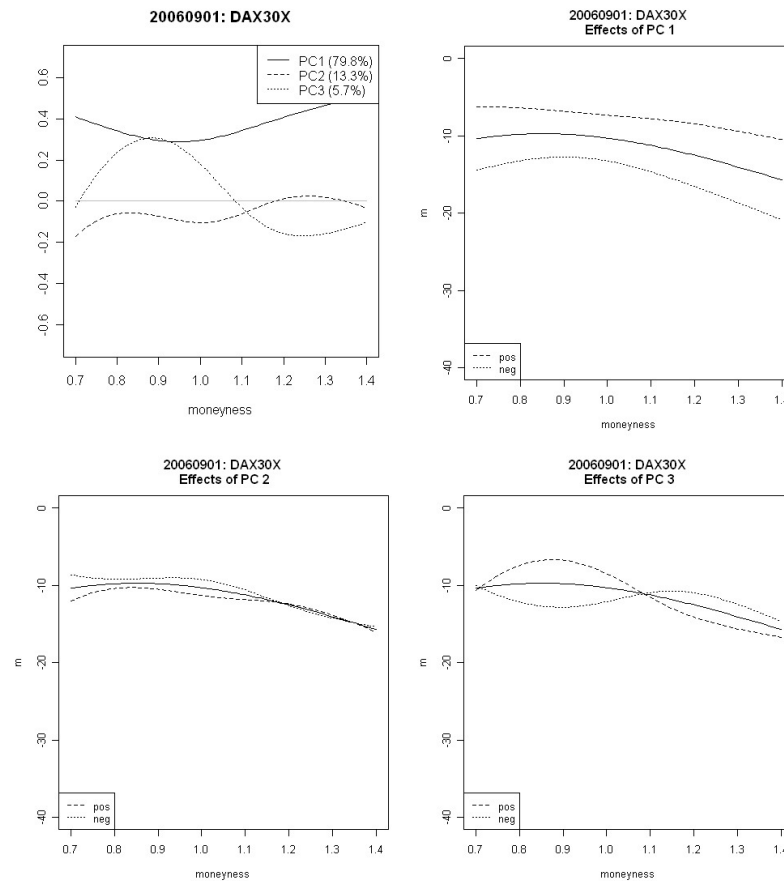


Figure 3: First three PCs of DAX30X IV data (upper left panel) and modes of variation explained by first three principal components. For each PC plots show the mean function (solid line), positive effects (dashed line) and negative effects (dotted line).

3 Functional Variance Processes

Originally introduced by Müller et al. [34] in 2006, we now present a new concept to analyse variation in functional data.

When analysing complex functional data it is a common tool to include a variance function into the non- or semiparametric model that quantifies heteroscedasticity, i.e. locally changing variances. When observed data scatter randomly around a fixed regression function, there is the need to add a variance of errors to the smooth mean regression function. This has traditionally been done in models of the form $Y_i = m(X_j) + \sigma(X_j)\epsilon_{ij}$ where m is the smooth mean function, σ is a smooth variance function and ϵ_{ij} are iid error terms with zero mean and variance 1. By $Var(Y_i) = \sigma^2(X_j)$ heteroscedasticity of the model is defined.

Now the concept of a variance function should be extended to a random variance process. When the data stem from a stochastic process, i.e. each observed trajectory is a random phenomenon itself, we should include random components for variation. We present a method combined of concepts of nonparametric smoothing and principal component analysis, to find a functional variance process, after an initial pre-smoothing step. This process generates smooth trajectories that jointly with pure white noise components determine the additive errors in the discretely observed data.

3.1 The Model

Our data is assumed to be realisations of smooth trajectories of a stochastic process observed under additive noise. The functional data is sampled on a discrete grid of m support points that may differ from trajectory to trajectory. The data comprises n curves Z_i which derive from a square integrable stochastic process S and some random noise R . The noise is generated by a smooth variance process V and an independent white noise component W . The model is

$$Z_{ij} = S_i(X_j) + R_{ij} \quad (3.1)$$

$$Y_{ij} = \log(R_{ij}^2) = m_V(X_j) + W_{ij} \quad (3.2)$$

$i = 1, \dots, n$, $j = 1, \dots, m$. S_i are identically distributed trajectories of the process S . $Z_{ij} = Z_i(X_j)$ are pointwise observed curves on the regular dense grid $\{X_j | X_j \in T = [a_1, a_2], j = 1, \dots, m\}$. The squared residuals R_{ij}^2 are assumed to be the product of the exponentiated mean function m_V of the variance process V at points X_j and an exponentiated white noise W_{ij} , i.e.

$$R_{ij}^2 = e^{m_V(X_j)} e^{W_{ij}}.$$

By this way of modeling we ensure nonnegativity of the estimated squared residuals. The white noise W_{ij} has constant variance and zero expectation, thus we can write

$$W_{ij} = \sigma_W \epsilon_{ij}$$

for iid random variables ϵ_{ij} , $E[\epsilon_{ij}] = 0$, $Var[\epsilon_{ij}] = 1$.

Hence, after filtering trajectories S_i we model the log-transformed residuals $Y_{ij} = \log(R_{ij}^2)$ according to the homoscedastic model (3.2), i.e. Y_{ij} are assumed to be additively decomposed of a variance process and a white noise. We assume the following properties

- S_i, m_V are smooth functions on T , i.e. twice continuously differentiable
- the domain $T = [a_1, a_2]$ is a compact interval
- $ER_{ij} = 0$, $Var(R_{ij}) < \infty$ and $R_{i_1j} \perp R_{i_2k}$ for $i_1 \neq i_2$
- $EW_{ij} = 0$, $Var(W_{ij}) = \sigma_W^2 < \infty$ and $W_{ij} \perp W_{ik}$ for $j \neq k$
- $W \perp S$, $W \perp V$

where \perp means independence. Let $m_V(x) = EV(x)$ and $\mathcal{G}_V(x, x') = Cov(V(x), V(x'))$ the smooth mean function and covariance structure of the functional variance process V , respectively. The model implies

$$EY_{ij} = EV(X_j) = m_V(X_j) \quad (3.3)$$

$$Var(Y_{ij}) = Var(V(X_j)) + \sigma_W^2 \quad (3.4)$$

$$Cov(Y_{ij_1}, Y_{ij_2}) = Cov(V(X_{j_1}), V(X_{j_2})) = \mathcal{G}_V(X_{j_1}, X_{j_2}), \quad j_1 \neq j_2. \quad (3.5)$$

Note that the diagonal values of the covariance matrix \mathcal{G}_V are influenced by the white noise variance. We focus on characterizing the variance function by the eigenfunctions and eigenvalues of the covariance operator which is

$$\mathbf{G}_V(f)(x') = \int_T \mathcal{G}_V(x', x) f(x) dx. \quad (3.6)$$

$\mathbf{G}_V : L^2(T) \rightarrow L^2(T)$, $f \mapsto \mathbf{G}_V(f)$. ϕ_k are smooth eigenfunctions with corresponding ordered eigenvalues $\rho_k \geq 0$, converging to zero. We approximate $V(x)$ by

$$V(K)(x) = m_V(x) + \sum_{k=1}^K \zeta_k \phi_k(x) \quad (3.7)$$

for some $K > 0$ and PC scores are $\zeta_k = \int_T (V(x) - m_V(x)) \phi_k(x) dx$.

3.2 Estimation of model components

We specify the different model components according to a PART algorithm (Principal Analysis of Random Trajectories), proposed by Müller et al. [34].

Each trajectory S_i is independently estimated by local linear kernel estimation. For all $i = 1, \dots, n$ and $x \in T$ minimize

$$RSS(S_i(x)) = \sum_{j=1}^m K_{b_{S_i}}^{[1]}(X_j - x) (Z_{ij} - \beta_0 - \beta_1(x - X_j))^2$$

with respect to $\beta = (\beta_0, \beta_1)$. Then $S_i(x)$ is estimated by $\hat{S}_i(x) = \hat{\beta}_0$, where $\hat{\beta}$ is the solution of the minimizing problem. The kernel $K_b^{[1]}(x) = K^{[1]}(x/b)/b$ is some symmetric, positive kernel that is supported on T . By $K^{[1]}$ a one-dimensional kernel function is denoted.

Optimal bandwidths b_{S_i} are chosen via cross validation, thus minimize

$$CV(b) = \sum_{j=1}^m \left(\hat{S}_i^{(-j)}(X_j) - Z_{ij} \right)^2$$

with respect to b . Now determine transformed residuals $Y_{ij} = \log(R_{ij}^2) = \log((Z_{ij} - \hat{S}_i(X_j))^2)$, and estimate the mean function m_V by local linear kernel estimation taking the whole data set into account, i.e. minimize

$$RSS(m_V(x)) = \sum_{i=1}^n \sum_{j=1}^m K_{b_V}^{[1]}(X_j - x) (Y_{ij} - \beta_0 - \beta_1(x - X_j))^2 \quad (3.8)$$

with respect to β and estimate $\hat{m}_V(x) = \hat{\beta}_0$. The bandwidth b_V is chosen by one-curve-leave-out cross validation, i.e. by minimizing

$$CV(b) = \sum_{i=1}^n \sum_{j=1}^m \left(\hat{m}_V^{(-i)}(X_j) - Y_{ij} \right)^2 \quad (3.9)$$

with respect to b . The next aim is to determine the covariance surface $\mathcal{G}_V(x', x)$. We have to guarantee that the estimate is

- symmetric
- positive definite
- appropriate at the diagonal, i.e. dispose of the white noise variance when estimating $Var(V(t))$.

Let $G_i = [G_i(X_{j_1}, X_{j_2})]_{j_1, j_2=1, \dots, m}$ the empirical covariance matrix of trajectory Y_i , i.e.

$$G_i(X_{j_1}, X_{j_2}) = (Y_{ij_1} - \hat{m}_V(X_{j_1}))(Y_{ij_2} - \hat{m}_V(X_{j_2})).$$

The covariance surface $\mathcal{G}_V(x', x)$ is then estimated by 2-dimensional local linear kernel estimation according to

$$RSS(\mathcal{G}_V(x', x)) = \sum_{i=1}^n \sum_{1 \leq j_1 \neq j_2 \leq m} K_{h_V}^{[2]}(X_{j_1} - x', X_{j_2} - x) \times \quad (3.10)$$

$$(G_i(X_{j_1}, X_{j_2}) - \beta_0 - \beta_1(x' - X_{j_1}) - \beta_2(x - X_{j_2}))^2$$

which is to be minimized with respect to $\beta = (\beta_0, \beta_1, \beta_2)$. If $\hat{\beta}$ is the solution, set $\hat{\mathcal{G}}_V(x', x) = \hat{\beta}_0$. We obtain a symmetric covariance surface by using a symmetric two dimensional kernel $K^{[2]}$, for example just choose the product kernel of two one-dimensional Gaussian kernels with same bandwidths. The bandwidth h_V is again chosen by one-curve-leave-out cross validation. Do not forget to leave out the diagonal values of the empirical covariance in sum (3.11). They are influenced by the white noise variance σ_W^2 and would bias the estimate.

We further improve diagonal estimates $\hat{\mathcal{G}}_V(x, x)$ by fitting a local quadratic component orthogonal to the diagonal of the empirical covariance, and a local linear component in the direction of the diagonal as proposed in [55], see remark at the end of this section.

Let now $\hat{\mathcal{G}}_V(x, x')$ the estimated covariance function corrected at $x = x'$. The last point we have to guarantee is the positive definiteness of the covariance function. Therefore determine eigenvalues ρ_k and eigenfunctions ϕ_k of the covariance operator with kernel $\hat{\mathcal{G}}_V$.

Then simply recalculate $\hat{\mathcal{G}}_V$ ignoring negative estimates of eigenvalues and corresponding eigenfunctions in the expansion of the covariance surface as proposed in [54], i.e.

$$\hat{\mathcal{G}}_V(x, x') = \sum_{\{k, \rho_k > 0\}} \rho_k \phi_k(x) \phi_k(x').$$

Now determine the white noise variance σ_W^2 . To estimate $Var(Y_{ij}) = Var(V(X_j)) + \sigma_W^2$ we just take the diagonal values of the empirical covariances, that is $G_i(X_j, X_j)$, $i = 1, \dots, n$, $j = 1, \dots, m$, and determine a local linear kernel estimator $\hat{Q}_V(x)$ of the diagonal by minimizing

$$RSS(Q_V(x)) = \sum_i \sum_j K_{b_{Q_V}}^{[1]}(X_j - x) (G_i(X_j, X_j) - \beta_0 - \beta_1(x - X_j))^2,$$

with respect to $\beta = (\beta_0, \beta_1)$. Approximate σ_W^2 from the difference between $\hat{Q}_V(x)$ and $\hat{\mathcal{G}}_V(x, x)$. To avoid boundary effects we restrain the support interval $T = [a_1, a_2]$ to $T_1 = [a_1 + |T|/4, a_2 - |T|/4]$ and obtain

$$\hat{\sigma}_W^2 = \frac{1}{|T_1|} \int_{T_1} \left(\hat{Q}_V(x) - \hat{\mathcal{G}}_V(x, x) \right)_+ dt \quad (3.11)$$

where $|T|$ is the length of the interval and $(\cdot)_+$ denotes the positive part of the term in brackets.

Finally calculate approximate trajectories V_i of the variance process V

$$V_i(K)(x) = \hat{m}_V(x) + \sum_{k=1}^K \hat{\zeta}_{ik} \phi_k(x) \quad (3.12)$$

where the principal component scores ζ_{ik} are discretely approximated by

$$\hat{\zeta}_{ik} = \sum_{j=2}^m (Y_{ij} - \hat{m}_V(X_j)) \phi_k(X_j) (X_j - X_{j-1}), \quad k = 1, \dots, K.$$

The number of included components K has to be chosen adequately. We determine it such that the total variance of chosen eigenvalues with corresponding eigenfunctions is more than 95%, i.e. $(\sum_{k=1}^K \rho_k) / (\sum_{k=1}^{\infty} \rho_k) \geq 0.95$. Other methods are possible, for example choose the number of model components by a one-curve-leave-out cross validation minimizing $CV(K) = \sum_{i=1}^n \sum_{j=1}^m (Y_{ij} - V_i^{(-i)}(K)(X_j))^2$ with respect to K where

$$V_i^{(-i)}(K)(x) = \hat{m}_V^{(-i)}(x) + \sum_{k=1}^K \hat{\zeta}_{ik}^{(-i)} \phi_k^{(-i)}(x),$$

see [34].

Remark: Yao et al. [55] found that the standard fitting of local planes around the diagonal of the covariance matrix G_V leads to overestimation of σ_W^2 . To adjust estimates of the variance of $V(x)$ the covariance surface is fitted by local quadratic kernel estimation in the

direction orthogonal to the diagonal.

In practice to obtain adjusted estimates rotate both axes by 45 degrees clockwise, i.e.

$$(X_{j_1}^*, X_{j_2}^*) = \frac{\sqrt{2}}{2}(X_{j_1}, X_{j_2}) \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

and get a surface estimate $G_V^*(x', x)$ by minimizing

$$RSS(G_V^*) = \sum_{i=1}^n \sum_{1 \leq j_1 \neq j_2 \leq m} K_{h_V}^{[2]}(X_{j_1}^* - x', X_{j_2}^* - x) \times \\ (G_i(X_{j_1}^*, X_{j_2}^*) - \beta_0 - \beta_1(x' - X_{j_2}^*) - \beta_2(x - X_{j_2}^*))^2$$

with respect to $\beta = (\beta_0, \beta_1, \beta_2)$. Then estimate the covariance diagonal according to $\hat{G}_V(x, x) = G_V^*(0, \sqrt{2}x)$, taking the rotation into account.

3.3 Asymptotics

In this section we want to present asymptotic results for the model introduced above. For each of the estimated components needed to approximate the functional variance process consistency is proven and consistency rates are given. Here consistency of an estimated random function $\hat{f}(x)$ means that $\|f - \hat{f}\|_\infty = \sup_{x \in T} |f(x) - \hat{f}(x)| \xrightarrow{P} 0$, or in case of estimated random numbers $\hat{u} \in \mathbb{R}$ consistency means $|u - \hat{u}| \xrightarrow{P} 0$. Assumptions and convergence are with respect to $n \rightarrow \infty$ and $m \rightarrow \infty$, where n is the number of observed curves (indexed by i) and m is the number of observations per subject i .

To insure the consistency results in theorems beneath, several conditions must be fulfilled. The stochastic process S and the variance process V are assumed to be smooth, that is to be twice continuously differentiable. V , S and its first and second derivative are assumed to be bounded, i.e.

(A1) $\exists C > 0 : \sup_x |V(x)| < C$ and $\sup_x |S^{(\nu)}(x)| < C$ for derivatives $\nu \in \{0, 1, 2\}$.

For the derivation of model components we used several bandwidths, which are b_{S_i} , $i = 1, \dots, n$, in the initial smoothing step to estimate S_i from observations Z_{ij} , further b_V to estimate m_V from the log transformed squared residuals Y_{ij} and the bandwidth h_V used twice in the two dimensional kernel when smoothing \hat{G}_V from empirical covariances G_i , $i = 1, \dots, n$ and to re-estimate the diagonal values. At last b_{Q_V} is used to estimate the white noise variance from the diagonal covariance values. Note, that all bandwidths were obtained by cross validation procedures and thus depend on the number of observations n and m .

We assume that bandwidths b_{S_i} to smooth trajectories S_i are connected by a common sequence of bandwidths $b_S = b_S(n)$,

(A2.1) $\exists b_S \exists c_1, c_2 \in (0, \infty) : c_1 < \inf_i \frac{b_{S_i}}{b_S} \leq \sup_i \frac{b_{S_i}}{b_S} < c_2$.

We impose further conditions that are essential for kernel smoothing asymptotics. For $n \rightarrow \infty$

$$(A2.2) \quad m \rightarrow \infty, \quad b_S \rightarrow 0, \quad mb_S^2 \rightarrow \infty,$$

$$(A2.3) \quad \text{for } b \in \{b_V, b_{Q_V}\} : \quad b \rightarrow 0, \quad nb^4 \rightarrow \infty, \quad \limsup_n nb^6 < \infty,$$

$$(A2.4) \quad h_V \rightarrow 0, \quad nh_V^6 \rightarrow \infty, \quad \limsup_n nh_V^8 < \infty,$$

$$(A2.5) \quad \text{for } b \in \{b_V, b_{Q_V}, h_V\} : \quad \limsup_n \frac{\sqrt{nb}}{m} < \infty.$$

We assumed in our model that the curves are observed on the same prediction points $X_1, \dots, X_m \in T$. For the consistency results given below this assumption can be relaxed to varying prediction points. I.e. let $X_{ij} \in T$, $j = 1, \dots, m$, the prediction points of the i -th curve, where we allow $X_{i_1j} \neq X_{i_2j}$ for $i_1 \neq i_2$. Asymptotically we assume regularly spaced prediction points X_{ij} , i.e. let $\Delta_n = \max_{j=2, \dots, m} (X_{ij} - X_{i,j-1})$, then

$$(A3) \quad \Delta_n = O\left(\frac{1}{m}\right) \text{ as } n, m \rightarrow \infty.$$

Further the fourth moments of observations Z_{ij} and transformed residuals Y_{ij} have to be uniformly bounded, i.e.

$$(A4) \quad \sup_j E[Z_{ij}^4] < \infty \text{ and } \sup_j E[Y_{ij}^4] < \infty.$$

We need specific assumptions to get consistency results for eigenfunctions and eigenvalues. Therefore we first introduce some more notation.

Let F denote a separable Hilbert space with norm $\|\cdot\|_F = \langle \cdot, \cdot \rangle_F^{1/2}$, where $\langle T_1, T_2 \rangle_F = \text{tr}(T_1 T_2^*)$ for Hilbert-Schmidt operators T_1, T_2 , and T_2^* is the adjoint of T_2 . The covariance operator \mathbf{G}_V defined in (3.6) is a Hilbert-Schmidt operator and we call $\mathbf{R}_V(y) = (\mathbf{G}_V - yI)^{-1}$ the resolvent of \mathbf{G}_V , where I is the identity operator. ρ_l and ϕ_l are the eigenvalues and corresponding eigenfunctions of \mathbf{G}_V , $l = 1, 2, \dots$. Let now

$$A_{\delta_j^V} = \sup_{y \in \Lambda_{\delta_j^V}} \|\mathbf{R}_V(y)\|_F,$$

where

$$\delta_j^V = \frac{1}{2} \min\{|\rho_l - \rho_j|, \rho_l \neq \rho_j\} \quad \text{and} \quad \Lambda_{\delta_j^V} = \{y \in \mathbf{C} : |y - \rho_j| = \delta_j^V\}$$

and \mathbf{C} is the space of complex numbers.

Let $K = K(n)$ the number of chosen PCs to model the variance process V according to (3.12), then we assume that K depends on n and m in such a way that

$$(A5) \quad \tau_n = \sum_{j=1}^K \frac{\delta_j^V A_{\delta_j^V} \|\phi_j\|_\infty}{\left(\sqrt{nh_V^2} - A_{\delta_j^V}\right)} \rightarrow 0, \text{ as } K \rightarrow \infty,$$

$$(A6) \quad \sum_{j=1}^K \|\phi_j\|_\infty = o(\min\{\sqrt{nb_V}, \sqrt{m}\}) \quad \text{and} \quad \sum_{j=1}^K \|\phi_j\|_\infty \|\phi_j^{(1)}\|_\infty = o(m),$$

when $n \rightarrow \infty$. $\phi^{(1)}$ is the first derivative of ϕ and $\|f\|_\infty = \sup_x |f(x)|$.

Further let $V(K)(x) = m_V(x) + \sum_{k=1}^K \zeta_k \phi_k(x)$. For $n \rightarrow \infty$ we assume

$$(A7) \ E \left[\left\{ \sup_{x \in T} |V(x) - V(K)(x)| \right\}^2 \right] = o(n),$$

$$(A8) \ \left\{ b_S^2 + \frac{1}{\sqrt{mb_S}} \right\} \sum_{j=1}^K \|\phi_j\|_\infty^2 = o_P(1) \quad \text{and} \quad \gamma_n = \sum_{j=1}^K \frac{\delta_j^V A_{\delta_j^V} \|\phi_j\|_\infty}{\left(\left\{ b_S^2 + \frac{1}{\sqrt{mb_S}} \right\}^{-1} - A_{\delta_j^V} \right)} \rightarrow 0.$$

At last, we need regularity conditions on the density and kernel functions. Let g and g_2 densities of observations Z_{ij} and pairs (Z_{ij_1}, Z_{ij_2}) and let f and f_2 densities of residuals Y_{ij} and pairs (Y_{ij_1}, Y_{ij_2}) , i.e.

$$\begin{aligned} Z_{ij} &\sim g(z, x) & \text{and} & & (Z_{ij_1}, Z_{ij_2}) &\sim g_2(z_1, z_2, x_1, x_2), \\ Y_{ij} &\sim f(y, x) & \text{and} & & (Y_{ij_1}, Y_{ij_2}) &\sim f_2(y_1, y_2, x_1, x_2). \end{aligned}$$

We assume these densities can be extended to smooth families of densities $g(\cdot, x)$, $g_2(\cdot, x_1, x_2)$, $f(\cdot, x)$ and $f_2(\cdot, x_1, x_2)$ that have second derivatives which are uniformly continuous on their domains $\mathbb{R} \times T$ and $\mathbb{R}^2 \times T^2$, respectively:

$$(B1.1) \ \frac{d^2}{dx^2} g(z, x) < \infty \quad \text{and} \quad \frac{d^2}{dx^2} f(z, x) < \infty$$

$$(B1.2) \ \frac{d^2}{dx_1^{l_1} dx_2^{l_2}} g_2(z_1, z_2, x_1, x_2) < \infty \quad \text{and} \quad \frac{d^2}{dx_1^{l_1} dx_2^{l_2}} f_2(z_1, z_2, x_1, x_2) < \infty,$$

where $l_1, l_2 \in \{0, 1, 2\}$, $l_1 + l_2 = 2$. Moreover, we have conditions on the kernel functions $K^{[1]} : \mathbb{R} \rightarrow \mathbb{R}$ and $K^{[2]} : \mathbb{R}^2 \rightarrow \mathbb{R}$. The kernels are compactly supported symmetric density functions with zero mean and finite variance (for $K^{[2]}$ in both arguments) and absolute integrable Fourier transforms, that is

$$(B2.1) \ \|K^{[1]}\|^2 = \int (K^{[1]})^2(u) du < \infty \quad \text{and} \quad \int |\chi_1(x)| dx < \infty,$$

$$(B2.2) \ \|K^{[2]}\|^2 = \int \int (K^{[2]})^2(u, v) dudv < \infty \quad \text{and} \quad \int \int |\chi_2(x, x')| dx dx' < \infty,$$

where $\chi_1(x) = \int e^{-iux} K^{[1]}(u) du$ and $\chi_2(x, x') = \int \int e^{-(iux + ivx')} K^{[2]}(u, v) dudv$ are the Fourier transforms.

There are several relaxing assumptions possible. For example we can think of random design points, that come from the same probability density function f or even from several f_i , that are bounded in the same interval but specific for each subject i . Further the number of observations m_i per subject may vary as long as there exists a common sequence m such that m_i/m , $i = 1, \dots, n$ is bounded. Asymptotic results in this context can be found in Yao et al. [55].

The following assertions are based on assumptions (A1)-(B2.2). They claim convergency of the estimates of each component included in the model presented in section 3.1 and give

convergency rates for most of them.

Theorem 1. *Under conditions (A1), (A2), (B1.1) and (B2.1), it holds for smoothed trajectories $\hat{S}_i(x)$ that*

$$E \left[\sup_{x \in T} |\hat{S}_i(x) - S_i(x)| \right] = O \left(b_S^2 + \frac{1}{\sqrt{mb_S}} \right).$$

The first theorem states that if trajectories S_i that drive the observations are smooth and individual bandwidths b_{S_i} are tied to b_S then local linear estimators of these trajectories are consistent. Hence, further decomposition of the transformed residuals Y_{ij} of the model is justified. A second theorem claims consistency of further model components, namely the mean function m_V , the covariance function $\mathcal{G}_V(x, x')$ of the process V , the white noise variance and also eigenfunctions and eigenvalues.

Theorem 2. *Under conditions (A1)-(A8), (B1.1)-(B2.2), it holds for the estimated components of the functional variance process that*

$$\begin{aligned} \sup_{x \in T} |\hat{m}_V(x) - m_V(x)| &= O_P \left(b_S^2 + \frac{1}{\sqrt{mb_S}} + \frac{1}{\sqrt{nb_V}} \right), \\ \sup_{x, x' \in T} |\hat{\mathcal{G}}_V(x, x') - \mathcal{G}_V(x, x')| &= O_P \left(b_S^2 + \frac{1}{\sqrt{mb_S}} + \frac{1}{\sqrt{nh_V^2}} \right), \\ |\hat{\sigma}_W^2 - \sigma_W^2| &= O_P \left(b_S^2 + \frac{1}{\sqrt{mb_S}} + \frac{1}{\sqrt{nh_V^2}} + \frac{1}{\sqrt{nb_{Q_V}}} \right). \end{aligned}$$

Considering eigenvalues ρ_k of multiplicity one, ϕ_k can be chosen such that

$$\sup_{x \in T} |\hat{\phi}_k(x) - \phi_k(x)| \xrightarrow{P} 0 \quad \text{and} \quad \hat{\rho}_k \xrightarrow{P} \rho_k.$$

Finally, consistency results for principle component scores ξ_{ik} justify consistency of the functional variance process.

Theorem 3. *Under conditions (A1)-(A8), (B1.1)-(B2.2), it holds for the estimated PC scores of the functional variance processes V that*

$$\sup_{k=1, \dots, K} |\hat{\zeta}_{ik} - \zeta_{ik}| \xrightarrow{P} 0,$$

where for the number of components in the model expansion $K = K(n) \rightarrow \infty$ as $n \rightarrow \infty$. Furthermore, for estimated trajectories $\hat{V}_i(K)(x)$ of the functional variance process V it holds that for $i = 1, \dots, n$

$$\sup_{x \in T} |\hat{V}_i(K)(x) - V_i(x)| \xrightarrow{P} 0.$$

Proofs of all three theorems can be found in [34]. They are mainly based on convergency results given in [44], [54] and [55].

4 Financial Options and Implied Volatilities

The aim of our work is to apply traditional and new developed statistical methods presented in section 3 in a financial context. To give a better understanding of applications presented in section 5 we now give an overview of financial options and terminology. For further information see [25].

Lets first introduce the notions stock index and swap, and then explain what options on financial assets are and what kind of options exist.

A *stock index* is an index monitoring the value of a portfolio of stocks. The percentage increase in the stock index over a small interval of time is set equal to the percentage increase in the value of the hypothetical portfolio. The weight of a stock in the portfolio equals the proportion of the portfolio invested in the stock. When the price of one particular stock in the portfolio rises more sharply than others, more weight is automatically given to that stock. The weights assigned to the stocks are proportional to their market prices or proportional to market capitalization, i.e. stock price \times number of outstanding shares. Examples are the German Stock Index (*DAX30*) or Standard & Poor's 500 (*S&P500*) index. A *swap* is an agreement between two companies to exchange cash flows in the future. The agreement defines the dates when the cash flows are to be paid and the way in which they are to be calculated. Popular types of swaps are interest rate swaps and currency swaps. In interest rate swaps a company agrees to pay cash flows equal to interest at a predetermined fixed rate on a notional principal for a number of years. In return, it receives interest at a floating rate on the same notional principal for the same period of time. A currency swap involves exchanging principal and interest payments in one currency for principal and interest payments in another currency.

An *option* is the right to buy or sell an asset at a certain date for a certain price. We distinguish between call options, which give the right to buy, and put options which give the right to sell. The price in the contract is called strike price, the date in the contract is called the expiration date or maturity. European options can only be exercised on the expiration date, American options can be exercised at any time up to the expiration date. An option on a stock gives the owner the right to buy or sell a stock. A stock index option is an option contract on an index, where one contract is to buy or sell 100 times the index at the specified strike price. A swap option or swaption is the option to enter into an interest rate swap where a specified fixed rate is exchanged for floating.

4.1 Basic Properties of Stock Options

Now concentrate on European call options on stocks. We can characterise options in terms of the terminal value, i.e. the payoff to the investor at maturity. If K is the strike price and S_T is the price of the underlying asset at expiration day T the payoff for the owner (long position) of an European call option is

$$\max(S_T - K, 0).$$

The owner will exercise his option when the predetermined strike is less than the actual price of the asset, $K < S_T$. He won't exercise it when $K \geq S_T$ not to make loss, thus the payoff is zero in this case. On the other side the payoff for the seller (short position) of the option is

$$\max(S_T - K, 0) = \min(K - S_T, 0).$$

To value a call option at time $t < T$ we discount the payoff at a continuously rate r for a time period $\tau = T - t$ which is the remaining time to maturity at t

$$e^{-r\tau} \max(S_T - K, 0),$$

when assuming a risk-free rate r on a non dividend paying stock. We call an option to be in-the-money if $K < S_t$ and out-of-the-money if $K > S_t$. An option is at-the-money if $K = S_t$.

Influencing factors of stock options are the current stock price S_0 , the strike price K , time to expiration T , the volatility of the stock price denoted by σ , the risk-free interest rate r and the dividends expected during the lifetime of the option. The option price is mainly influenced by the current stock price. For fixed strike K the option will become more valuable when the stock price increases. On the opposite side, for fixed strike price the option price will decrease for increasing strike price. The volatility of a stock price is a measure of how uncertain we are about stock price movements. When volatility is high the chance is high that the stock price will strongly increase or decrease. The owner of a call option will have great benefit from increasing stock price, but only a limited risk from decreasing stock price. He will maximal loose the price of the option. Thus increasing volatility has a positive influence on the call option price, i.e. it will increase the value of the option. The risk-free interest rate affects call options in a positive way thus call option prices increase for increasing rate. When dividends occur the stock price will decrease the day after the dividend is paid. As a consequence call options will loose value for increasing dividend values. How about the influence of time to maturity? Options become more valuable when the expiry date is far in future. Thus increasing maturity should increase the option price in general. But there exist situations, when the behavior is opposite. When a large dividend on a stock is expected, an option with expiry date before the dividend paying will be worth more than an option on the same stock with expiry date after the dividend paying. In this case the option price will decrease for increasing time to maturity. Therefore we cannot clearly determine the influence of maturity onto option prices. Similar considerations can be made in case of European put options and in case of American options.

4.2 Black-Scholes Option Pricing

Theory to value options is made under the risk neutral assumption. In a risk-neutral world investors are assumed to require no extra return on average for bearing risks and the expected return on all securities is the risk-free interest rate. The risk-free rate is the rate at which money is borrowed or lent when there is no credit risk, so that the money is certain to be repaid. When moving from a risk-neutral world to a risk-averse world two things

happen. On the one hand the expected growth rate in the stock prices change, and on the other hand the discount rate that must be used for any payoffs from the derivatives change. It turns out that these two changes always offset each other exactly. Therefore the option price we obtain in risk-neutral valuation is correct not just in a risk-neutral world but in the real world as well.

Stock prices are random variables changing their value in time, where any point in time and any value in the specified range is feasible. We assume stock prices to be driven by an Ito processes according to

$$\delta S = \mu S \delta t + \sigma S \delta z,$$

where z is a Wiener process. μ is the expected rate of return, we expect an increase of S to $\mu S \delta t$ in time δt . The variability of the percentage return of the stock price is the same regardless of the stock price, therefore the standard deviation of the change in δt should be proportional to the stock price as well. We may write equivalently

$$\frac{\delta S}{S} = \mu \delta t + \sigma \delta z$$

and $\frac{\delta S}{S} \sim \mathcal{N}(\mu \delta t, \sigma^2 \delta t)$.

By Ito's Lemma

$$\ln S_T - \ln S_0 \sim \mathcal{N}\left(\left(\mu - \frac{\sigma^2}{2}\right) T, \sigma^2 T\right)$$

and further

$$\ln S_T \sim \mathcal{N}\left(\ln S_0 + \left(\mu - \frac{\sigma^2}{2}\right) T, \sigma^2 T\right).$$

Stock prices are log-normal distributed with expectation $E(S_T) = S_0 e^{\mu T}$ and variance $Var(S_T) = S_0^2 e^{2\mu T} (e^{\sigma^2 T} - 1)$. When η is the continuously compounded rate of return, i.e. $S_T = S_0 e^{\eta T}$ then $\eta = \frac{1}{T}(\ln S_T - \ln S_0)$ thus

$$\eta \sim \mathcal{N}\left(\mu - \frac{\sigma^2}{2}, \frac{\sigma^2}{T}\right).$$

For further details see [25] or [27].

The expected value of a call on a European stock option at maturity T in a risk-neutral world is $E(\max(S_T - K, 0))$ where S_T is the price of the stock and K is the strike price. Thus at time t the option is valued by

$$C_t = e^{-r\tau} E(\max(S_T - K, 0)) \quad (4.1)$$

which is the discounted payoff at a continuously rate r when $\tau = T - t$ is the remaining time to maturity. We obtain the well known Black-Scholes formula for option pricing by the following theorem.

Theorem 4. Let V log-normal distributed with expectation $E[V]$ and variance $\text{Var}[\ln V] = s^2$. Then

$$E(\max(V - K, 0)) = E[V]\mathcal{N}(d_1) - K\mathcal{N}(d_2)$$

where

$$d_1 = \frac{\ln(E[V]/K) + s^2/2}{s}, \quad d_2 = d_1 - s.$$

A proof can be found in [25].

Now, S_T is log-normal with expectation $E[S_T] = S_0 e^{-rT}$ and $\text{Var}[\ln S_T] = \sigma^2 T$. Using theorem 4 we get $E(\max(S_T - K, 0)) = S_0 e^{rT} \mathcal{N}(d_1) - K \mathcal{N}(d_2)$ where

$$d_1 = \frac{\ln\left(\frac{S_0 e^{rT}}{K}\right) + \frac{\sigma^2}{2}T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T}.$$

By (4.1) we result in Black-Scholes formula for pricing European call options on stocks, first introduced by F. Black, M. Scholes and R. Merton in the early 1970s, i.e.

$$C_0 = S_0 \mathcal{N}(d_1) - K e^{-rT} \mathcal{N}(d_2). \quad (4.2)$$

When using Black-Scholes formula in practice the interest rate r is set equal to the zero coupon risk-free interest rate for a maturity T . (A zero coupon is a bond that provides no interest payments.) The volatility σ^2 is not known and has to be estimated from historical data or as implied volatility. We will explain these concepts later on in this section. The option price at time t according to Black-Scholes is

$$C_t = S_t \mathcal{N}(d_1) - K e^{-r\tau} \mathcal{N}(d_2),$$

where $d_1 = \frac{\ln(S_t/K) + (r + \sigma^2/2)\tau}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$.

There are several extensions of Black-Scholes formula to price options on further financial assets, like dividend paying stocks, indexes or swaps. We now present corresponding concepts.

For dividend paying stocks to value an European option lasting for time T , we reduce the current stock price from S_0 to $S_0 e^{-qT}$ where q is the annual rate of the dividend yield. Indeed a stock that starts at S_0 and pays a dividend has the same probability distribution as a stock starting at $S_0 e^{-qT}$ and paying no dividends. Therefore use Black-Scholes to value European options on dividend paying stocks substituting S_0 by $S_0 e^{-qT}$. Then by $\ln\left(\frac{S_0 e^{-qT} e^{rT}}{K}\right) = \ln(S_0/K) + (r - q)T$ we get

$$C_0 = S_0 e^{-qT} \mathcal{N}(d_1) - K e^{-rT} \mathcal{N}(d_2), \quad (4.3)$$

where $d_1 = \frac{\ln(S_0/K) + (r - q + \sigma^2/2)T}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$.

Options on stock indexes can be treated as options on dividend paying stocks. Recall

that a stock index is a benchmark of a hypothetical portfolio of stocks. Thus to value an index option we average dividend yields of the portfolio's stocks. Let now q the average annualized dividend yield on the index during the lifetime of the option, where only those dividends are included whose ex-dividend data occur in the lifetime. Then an European call option on a stock index is valued according to the extended Black-Scholes formula (4.3).

Valuation of options on interest rate swaps is somewhat more complicated. Let the agreement be to swap N payments of fixed and continuous rate m times a year. The duration of the swap will be N/m years. The exercise price K of an option on the swap is here the fixed rate s_K committed at the beginning of the option, at time $t = 0$. Let s_T the fixed rate of a newly issued swap offered at time $t = T$ with the same duration and number of cash flows. We assume that s_T is log-normal distributed. Let L the notional principal on which interest rates are paid. Then single payoffs of the swap are

$$\frac{L}{m} \max(s_T - s_K, 0)$$

paid m times a year. According to Black-Scholes we can value an European call option on the swap by

$$C_0 = \sum_{i=1}^N \frac{L}{m} p(0, T + i/m) (s_0 \mathcal{N}(d_1) - s_K \mathcal{N}(d_2)) \quad (4.4)$$

where $d_1 = \frac{\ln(s_0/s_K) + T\sigma^2/2}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$. $p(0, T)$ is the price at time $t = 0$ for a zero-coupon bond paying 1 at time T , and s_0 is the forward swap rate thus the expectation of rate s_T .

4.3 Implied Volatility

The Black-Scholes formula for option pricing is a function of various parameters that determine the price of an option. In case of stock options we have $C^{BS}(r, S_0, T, K, \sigma)$ where the only one parameter that cannot directly be observed is the volatility σ^2 of the stock price. When estimating this parameter there are two main approaches. First is to approximate σ from historical data. Under the log-normal assumption the volatility σ^2 is the variance of observations $x_t = \ln(S_t/S_{t-1})$. Then the next idea on hand is to estimate the variance at day t empirically by

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=t-n}^t (x_i - \bar{x})^2,$$

where \bar{x} is the observation's mean and n defines the time horizon taking into account. An interesting question is then how to choose the time horizon. If n is chosen too large, we include data that is to 'old' to have influence on the current values and thus leads to estimation bias. But the smaller we choose n the worser will be the empirical variance as a fit of σ^2 . Another approach to estimate σ is to use implied volatilities.

The implied volatility of a call option is the volatility resulting from Black-Scholes formula when the option price is observed on the market. In case of stock options we thus set

$$C^{obs} = C^{BS}(r, S_0, T, K, \sigma), \quad (4.5)$$

were C^{BS} denotes the option price under Black-Scholes formula and C^{obs} is the observed price. The option price regarded as a function of volatility σ is a monotonic mapping from \mathbb{R}^+ to $(0, S_0 - Ke^{-rT}]$, therefore the solution to equation (4.5) exists and is unique. We call the solution the implied volatility of the stock option at strike price K and maturity T . Implied volatilities of put options can be obtained by the put-call parity which is

$$C_t + Ke^{-r\tau} = P_t + S_t,$$

and allows us to determine call prices according to observed put prices. Implied volatilities are used to monitor the market's opinion about the volatility of a particular stock. Prices of deep-in-the-money options ($K \ll S$) and deep-out-of-the-money options ($K \gg S$) are relatively insensitive to volatility. IVs calculated from these options tend to be unreliable. Traders like to calculate IV from actively traded options on a certain asset and interpolate between them to calculate the appropriate volatility for pricing a less actively traded option on the same stock.

Implied volatilities are frequently analysed as functions of the strike price and time to maturity, for example for options on stocks or stock indexes. It is well known that implied volatilities are not equal for options on the same stock when maturity or strike price vary and thus contradict the Black-Scholes assumption of constant volatility σ . In fact IVs tend to be larger when the strike price is very low or very high. We are then talking of the IV smile, i.e. the shape of the IV curve with respect to strike resembles a smile. In direction of maturity we call the IV curve the term structure. The shape of term structure curves is not clearcut. When historical volatility has been high, volatility tends to be decreasing for decreasing option maturity. When historical volatility has been low, volatility tends to increase. This is a consequence of the traders expectation of market behavior, as the volatility seems to be largely caused by trading itself.

In case of swaptions implied volatilities are analysed as functions of swap maturity and option maturity. Resulting curves are not flat, as well and tend to have maxima for short (swap and option) maturities.

The influence of σ onto the option price is clearcut: According to (4.2) if $\sigma \rightarrow 0$ (i.e. the stock is virtually riskless) and $S_0 > Ke^{-rt}$ both, $\mathcal{N}(d_1)$ and $\mathcal{N}(d_2)$ tend to 1 and the option price becomes $S_0 - Ke^{-rt}$ and the call option is exercised. If $\sigma \rightarrow 0$ and $Ke^{-rt} > S_0$ both, $\mathcal{N}(d_1)$ and $\mathcal{N}(d_2)$ tend to 0 and the option price becomes 0 and the call option is not exercised. If $\sigma \rightarrow \infty$, $\mathcal{N}(d_1)$ tends to 1 and $\mathcal{N}(d_2)$ to 0 thus $C_0 \rightarrow S_0$, i.e. for increasing asset price volatility the option becomes as worth as the asset itself.

In general, the dependence of the option price and the volatility is analysed in terms of the option vega which is the proportion of price changes and volatility changes. We determine the vega by the derivative

$$vega = \frac{\partial C}{\partial \sigma}$$

which is $S_0\sqrt{\tau}\mathcal{N}'(d_1) = S_0\sqrt{\tau} \frac{\exp(-d_1^2/2)}{\sqrt{2\pi}}$ in case of European call options on non-dividend paying stocks. *vega* is maximal for at-the-money options as the normal density $\mathcal{N}'(d_1)$ is

largest for $K = S$. When *vega* is large in absolute terms little changes of σ lead to large changes in the option price. When *vega* is small in absolute terms changes of σ have to be larger to lead to the same amount of changes in the option price. Thus the option vega determines the sensitivity of option prices to volatility changes.

In this context it is clearcut that estimation of (implied) volatilities and their variation plays an important role in terms of risk controlling when trading options.

5 Analysis of Implied Volatilities

5.1 Introduction of the Data Set

We have got three big data sets of implied volatilities, first one resulting from options on stocks (Stock IV), second one resulting from options on indexes (Index IV) and third one on swaps (Swap IV). The data sets were kindly offered by the COMMERZBANK AG in Frankfurt a.M., Germany, section Risk Controlling. The external data source for the prices of options and underlyings to determine Stock and Index IV is REUTERS. BLOOMBERG is the data source for associated dividends. Swap IV are originally from REUTERS.

The implied volatilities of stock and index options are given for different strike prices and times to maturity of the option. We have index IVs of options on the indexes

DAX30X	Xetra DAX 30, Frankfurt,
DJ	Dow Jones Industrial, New York,
FTSE100	Financial Times Stock Exchange, London,
Nikkei225	Nikkei 225, Tokyo,
SANDP500	Standard & Poor's 500 (S&P500), New York,
SMI	Swiss Market Index, Zuerich.

Daily data is given from 03.01.2005 to 31.10.2006 (477 trading days).

Stock IV data is given for **Peugeot** and **Volkswagen** stocks from 03.01.2005 to 31.10.2006 (477 trading days) for **Air France** stocks it is given from 03.02.2005 to 31.10.2006 (450 trading days).

Implied volatilities are determined over a grid of moneyness values

$$\kappa \in \{0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.4\}$$

and time to maturity

$$\tau \in \left\{ \frac{1}{24}, \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 2, 3, 5, 10 \right\} \text{ years,}$$

where $\kappa = K / \{S * \exp(r * \tau)\}$ is the strike price divided by the future asset price. This makes a total of 36.729 data points per index and per stock, except Air France with 34.650 data points.

In contrast to previous IVs, swaption implied volatilities do not depend on strike price and option maturity, but depend on the maturity of the swap and on option maturity. Therefore Swap IV data is given on a grid of 14 different swap maturities τ^S and 14 different option maturities τ^O ,

$$\begin{aligned} \tau^S &\in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30\} \text{ years,} \\ \tau^O &\in \left\{ \frac{1}{12}, \frac{1}{4}, \frac{1}{2}, 1, 2, 3, 4, 5, 7, 10, 15, 20, 25, 30 \right\} \text{ years.} \end{aligned}$$

Data is observed at 477 trading days, namely from 03.01.2005 to 31.10.2006, such that a single SWAP IV data set consists of 93.492 data points. Such sets are given for IV of options on **EUR** (European Euro), **GBP** (Great Britain Pound), **JPY** (Japanese Yen) and

USD (United States Dollar) swaps.

Index and Stock IVs were calculated in-house by the COMMERZBANK AG from the market data given by external sources. In general such data appears as ‘floating point’ data. To get IVs on a fixed grid of moneyness and maturity values a 3-dimensional interpolation algorithm was applied. Furthermore, for remaining option maturity of more than two years IV data was extrapolated and should be handled suspiciously in the analysis.

Methods used to interpolate and extrapolate the data are documented in [50] and [30].

5.2 Functional Data Analysis of the IVS

As already mentioned we are interested in the variation of implied volatilities. There are three approaches of how to analyse the volatility of IVs. First of all we may regard changes of IVs within the implied volatility surface for a fixed day, i.e. fitting the IV smile and term structure. A second point is to describe the IV time series for fixed moneyness and maturity. How do IVs vary in time? How do time series of IV log returns look like? Third, the changes of IV log returns are of interest when moneyness or maturity values vary, i.e. variation within the ‘IV log return surface’.

In what follows, we will have a short look at the first and second point, we will follow the analysis in Fengler et al. [17], but our main focus is on the analysis of variation of IV log return curves in the daily surface.

5.2.1 First steps of data analysis

To get a first look at the properties of the data we distinguish between index and stock IVs on the one hand and swap IVs on the other hand. We limit the index and stock data to a maximum of 2 years to option maturity, as we do not rely on the extrapolated data afterwards, i.e. we have prediction points $\tau \in \{\tau_1, \dots, \tau_8\} = \{\frac{1}{24}, \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 2\}$ years and $\kappa \in \{\kappa_1, \dots, \kappa_7\} = \{0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.4\}$. In case of swap data, to guarantee a regular grid of prediction points we limit option maturity to maximal 5 years and swap maturity to maximal 10 years, i.e. prediction points $\tau^O \in \{\tau_1^O, \dots, \tau_{10}^O\} = \{\frac{1}{12}, \frac{1}{4}, \frac{1}{2}, 1, 2, 3, 4, 5, 7, 10\}$ years and $\tau^S \in \{\tau_1^S, \dots, \tau_{10}^S\} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ years.

First we generate daily plots of the IVS of options on indexes and stocks. We plot the IVs as two dimensional surfaces dependent on maturity of the option and on moneyness. Having data only over a grid of discrete values, we simply interpolate linearly between the grid points. Plots of the IVS at 1 September 2006 for two indexes and two stocks are shown in the upper four panels in figure 4.

For both, index and stock data in-the-money IVs are predominantly higher than out-of-the-money IVs. Curves in moneyness direction appear as the well-known IV smiles or smirks, i.e. IV is high for moneyness $\kappa = 0.7$, decreases until $\kappa = 1$ and increases again or remains nearly constant afterwards for increasing moneyness $\kappa > 1$. This behavior is well pronounced for short time to option maturity and weakens when option maturity increases. The reason for IV smile appearance is that the volatility of a traded option is higher, when

the strike price of an option is lower or higher than the asset price, i.e. moneyness $\kappa < 1$ or $\kappa > 1$. In those cases exercising the option makes sense for buyers or sellers of the option, respectively.

In direction of option maturity IV curves (i.e. term structure curves) have extreme values for short maturities and flatten for long maturities. In general the curves are decreasing. This makes sense as short maturity options are traded more frequently. But there exist cases when the behavior is reverse and the term structure is increasing. Term structure curves are best pronounced when options are in-the-money, but nearly flat for at-the-money.

IV time series for fixed κ and τ resemble trajectories of stochastic processes jumping up and down, with different drifts and variances. Examples are given in figures 5 and 6, where we plotted DAX30X and Volkswagen IV time series for fixed moneyness values over time. For both, index and stock data, when maturity is small (here 1 month) we notice that time series are extremely volatile for low moneyness. Time series become much less volatile when moneyness increases. The volatility of IV time series for at-the-money and out-of-the-money options seems to be the same. For increasing time to maturity the volatility of IVs decrease, compare therefore the time series plotted from upper to lower panels.

Note that in general IV time series for in-the-money options lie above time series for moneyness $\kappa \geq 1$, reflecting the typical smile or smirk behavior of IVs in moneyness direction.

Moreover, regard the IV log returns shown in figure 8, upper four panels, where we plot surfaces of log return data at 1 September 2006. As expected from previous data inspection, data are most volatile and reach extreme values in regions where moneyness is small and option maturity is short.

Now we will make similar considerations for the swaption data. The swap implied volatility surface depends on option maturity and the maturity of the traded swap. Thus features of these data differ from index and stock IVs. In general swap IVs are flatter than index or stock IVs, note therefore the different scales of the IV axes in figure 4.

In direction of option maturity, for EUR, GBP and USD options the IV curves are increasing up to 1 or 2 years option maturity, where the curves reach their maximum and then decrease for increasing option maturity. Curves for JPY options are decreasing. In swap maturity direction, in general IV curves are decreasing and the slope of the curves is almost constant. Only for option maturities less than one year and short swap maturities extreme IV values appear.

Time series of JPY implied volatilities for different fixed maturities are shown in figure 7 as an example of swap option IVs. In each panel time series for swap maturities $\tau^S \in \{1, 2, 4, 6, 8, 10\}$ years are plotted together. The different panels contain plots for different option maturities $\tau^O \in \{1/12, 1, 5\}$ years. Variation of the data is similar for swap maturities more than four years, for swap maturities underneath variation increases. IV time series according to short swap maturities lie above those for high maturities, i.e. IVs decrease for increasing τ^S . For increasing option maturity and increasing swap maturity

variation of the IV time series decreases to a very low level.

We have to add that data is more volatile for EUR, GBP and USD swaps, but the characteristics described so far are the same for all swaps.

IV log return surfaces are displayed in figure 8, lower two panels. Data is almost constant over the whole range of swap and option maturities and the variation of the data is much less than it is for index and stock log returns. Additionally, variation seems not to differ a lot between short and long (option and swap) maturities.

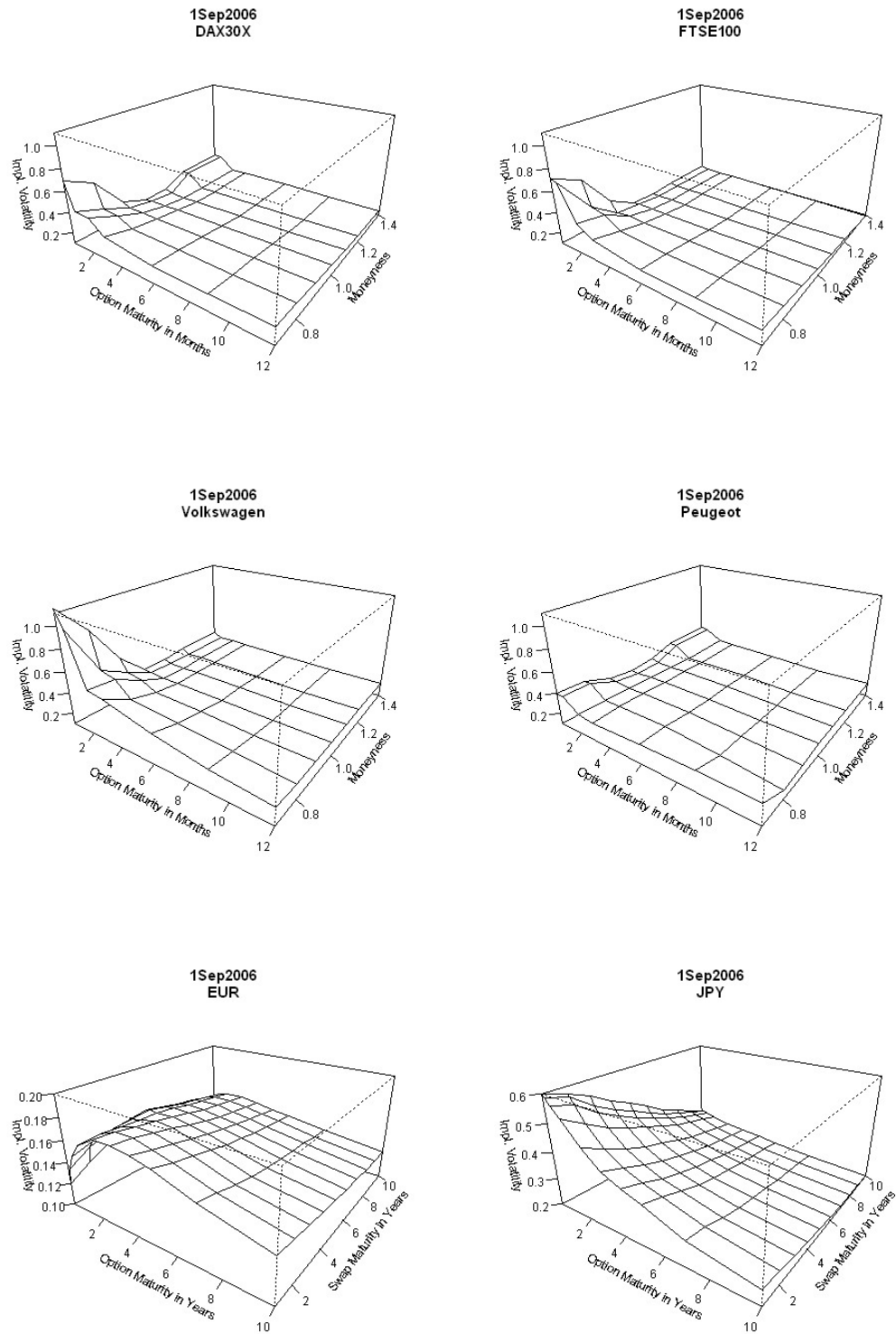


Figure 4: Implied volatility surfaces at September 1, 2006 for two index options (DAX30X, FTSE100), two stock options (Volkswagen, Peugeot) and two swap options (EUR, JPY). Typical smile curves occur for DAX and Peugeot options with short maturities. Smirk curves are well pronounced for FTSE and Volkswagen options with short maturities.

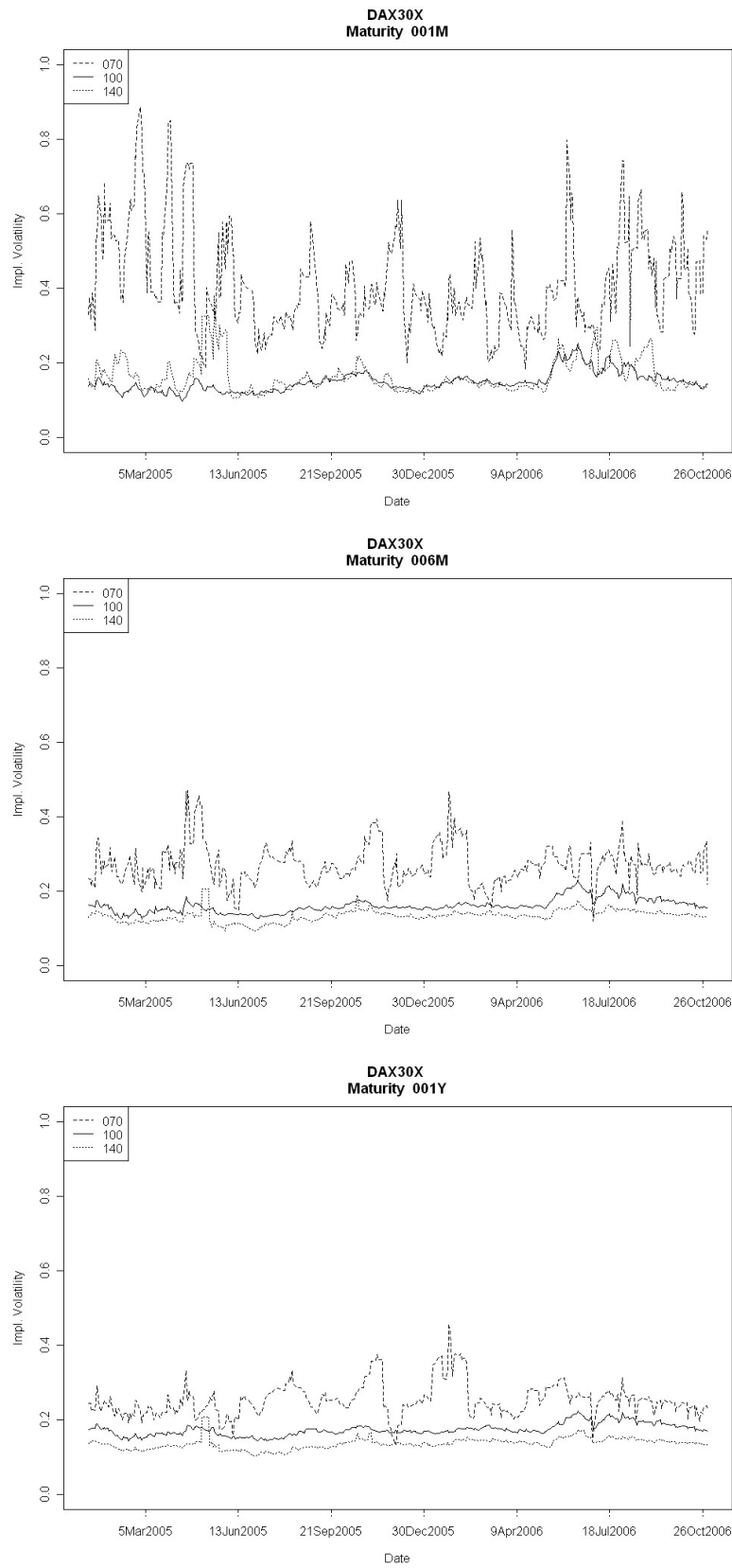


Figure 5: Dax IV time series for fixed maturities 1 month (top panel), 6 months (middle panel), 1 year (bottom panel). Different curves correspond to moneyness in $\{0.7, 1.0, 1.4\}$.

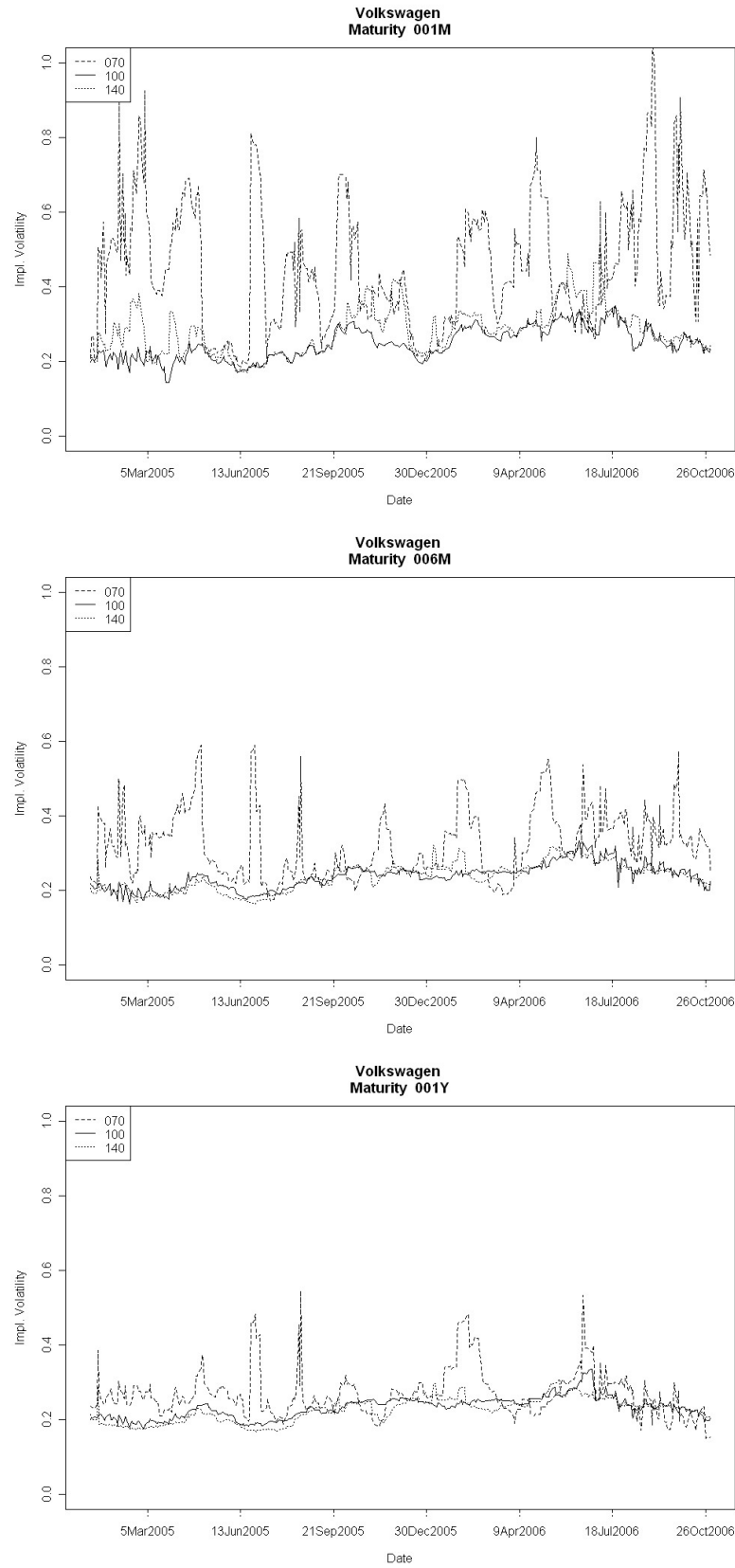


Figure 6: Volkswagen IV time series for fixed maturities 1 month (top panel), 6 months (middle panel), 1 year (bottom panel). Different curves correspond to moneyiness in $\{0.7, 1.0, 1.4\}$.

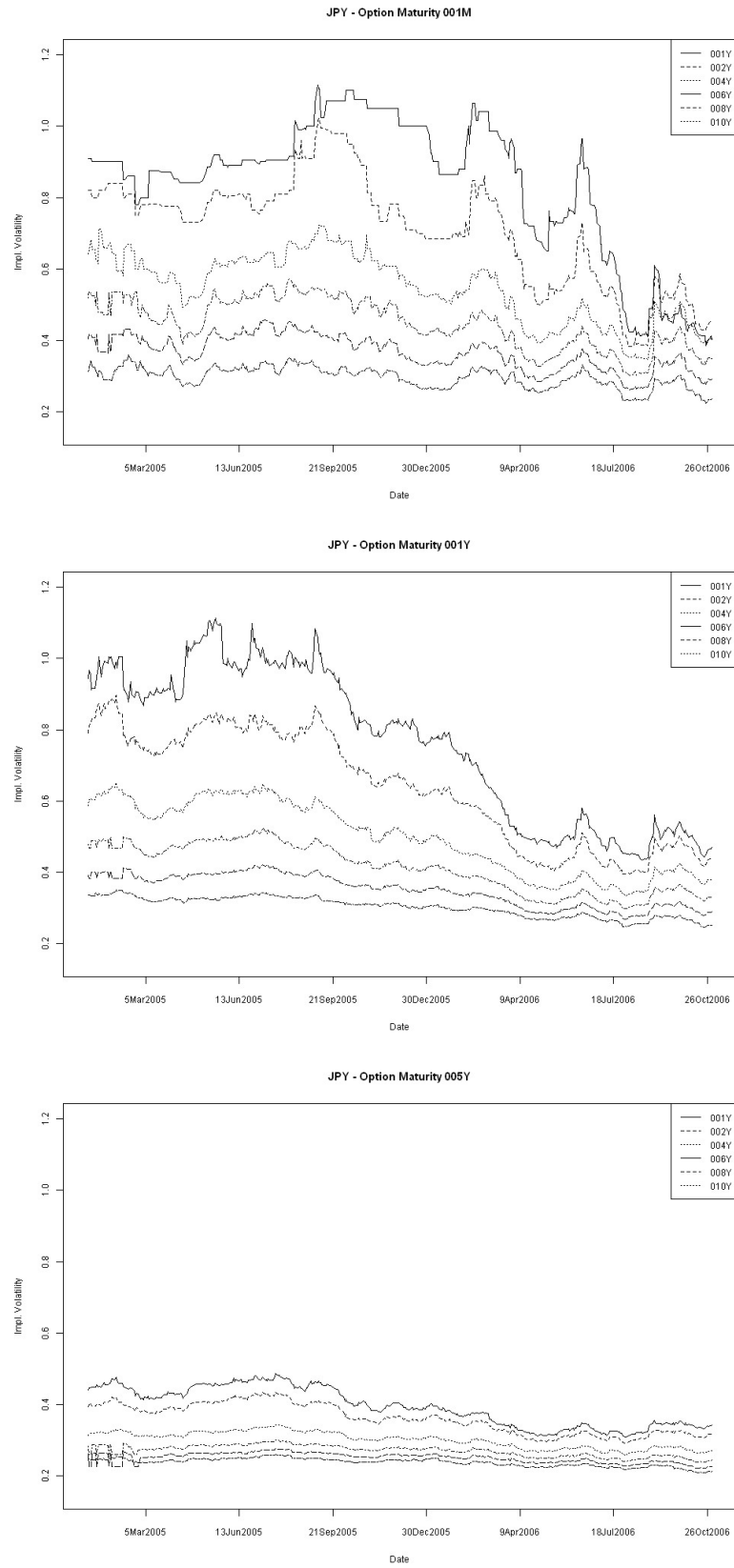


Figure 7: JPY IV time series for fixed option maturities 1 month (top panel), 6 months (middle panel), 1 year (bottom panel). Different curves correspond to swap maturities from 1 to 10 years.

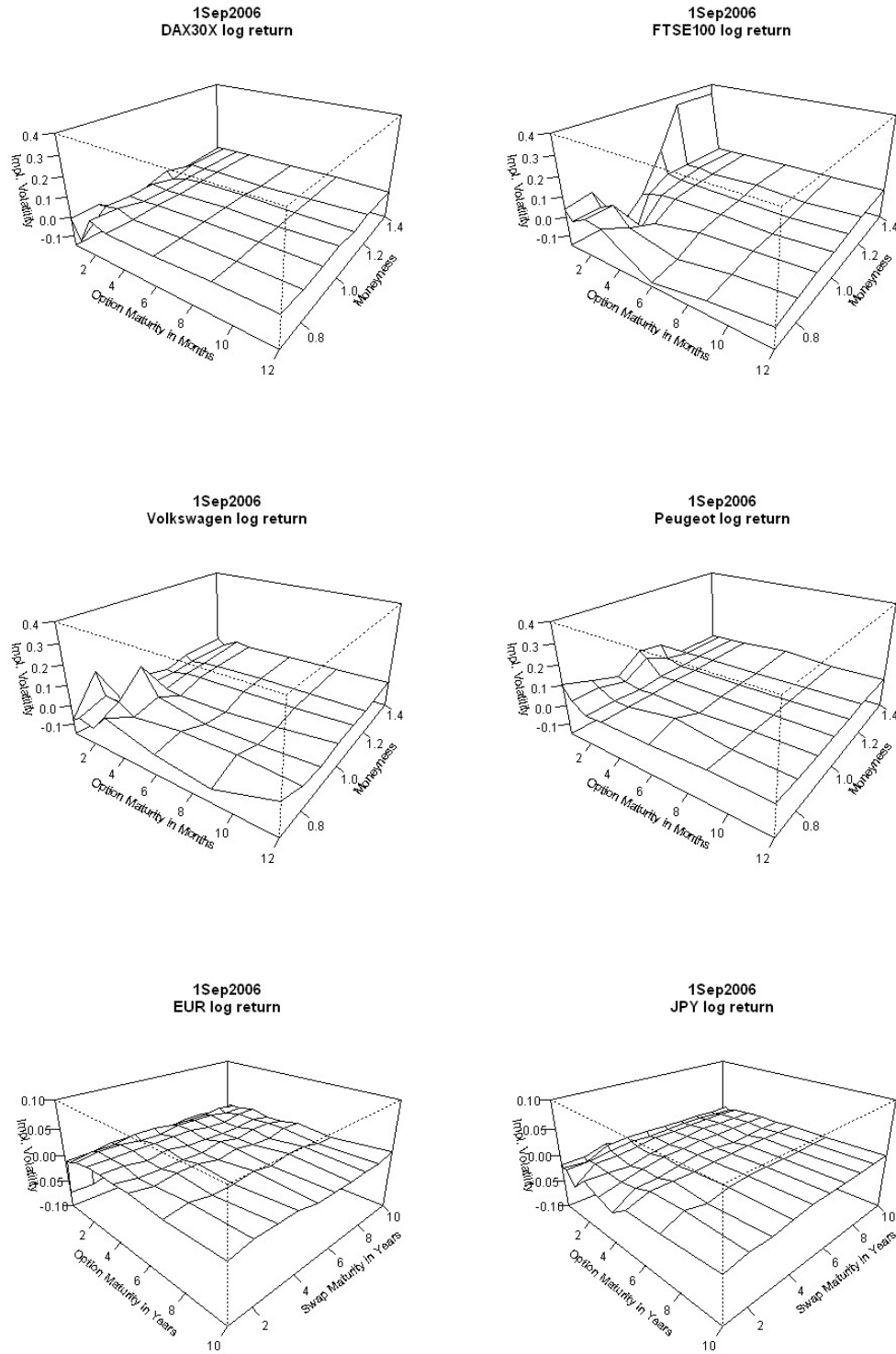


Figure 8: Log returns of implied volatility surfaces at September 1, 2006 for two index options (DAX30X, FTSE100), two stock options (Volkswagen, Peugeot) and two swap options (EUR, JPY). Index and stock data is more volatile when option maturity is short and moneyness is small.

5.2.2 Principal component analysis

Next we analyse principal components of the data to specify main directions of variation of the IVS in time. We proceed according to methods of Fengler et al. [17] and analyse the shocks, that drive IV smiles in time.

For further examination of the variation we analyse the log returns of implied volatilities sampled over time $t \in \{1, \dots, T\}$, which is

$$\sigma(t, \kappa, \tau) = \log \left\{ \frac{IV(t, \kappa, \tau)}{IV(t-1, \kappa, \tau)} \right\} = \log IV(t, \kappa, \tau) - \log IV(t-1, \kappa, \tau),$$

in case of index and stock data. For each single $\tau \in \{\tau_1, \dots, \tau_8\}$ we perform a PCA of the sampled data $Y_j(\tau) = (\sigma(1, \kappa_j, \tau), \dots, \sigma(T, \kappa_j, \tau))^T$, $j = 1, \dots, 7$. Thus for each τ we get a set of seven eigenvalues and eigenvectors. In case of swap IV, replace κ by τ^S and τ by τ^O . Then for each τ^O we get a set of ten eigenvalues and eigenvectors.

First, to detect dependencies among data for different option maturities, we scatterplot the log return data for moneyness $\kappa_1 = 1$ vs. $\kappa_2 = 0.7, 0.8, 0.9, 1.1, 1.2, 1.4$ respectively, when time to maturity is fixed. An example is given in figure 9 for DAX30X indexes. Further we scatterplot each pair of adjoined moneyness values, i.e. 0.7 vs. 0.8, 0.8 vs. 0.9, etc., see figure 10. Additionally we draw the two main directions of variation into the plots and an approximate 0.95 confidence region, which is an ellipse of constant standard distance. Here, testing for multivariate normal distributed data, we set the square root of the Mahalanobis distance equal to 2, i.e.

$$\{(y - \bar{y})^T S^{-1} (y - \bar{y})\}^{1/2} = 2,$$

where $y = (y_{j1}, y_{j2})^T$, mean vector $\bar{y} = (\bar{y}_{j1}, \bar{y}_{j2})^T$, $\bar{y}_j = \frac{1}{T} \sum_{t=1}^T y_{j,t}$, $i = 1, 2$. S is the sample covariance matrix of observations $Y_{j1}(\tau)$ and $Y_{j2}(\tau)$. The size of the ellipses reflects the variability of the data. The main directions are given by the eigenvectors of the covariance matrix of observation vectors $Y_{j1}(\tau)$ and $Y_{j2}(\tau)$, compare section 2.3.

The plots show again that the volatility of the data is highest when moneyness and maturity values are small. Deviation is biggest in the first row and the first column plots, where we have 1 month option maturity and moneyness 0.7. The main directions are similar in rowwise plots, where moneyness values are identic but maturity varies. Thus it seam reasonable to analyse the data on the basis of common principal components for groups of different option maturities, but to allow the variation, i.e. the eigenvalues to vary among groups. When the ellipsoid is very flat, i.e. one principal axis is short relative to the other and axes are not parallel to the x-axis or y-axis in the plot, we assume a linear dependence among data for different moneyness values. As expected this feature intensifies when κ_1 is close to κ_2 , which are middle row plots in figure 9 and all plots in figure 10. Thus for our data we can assume dependencies between neighboring moneyness groups.

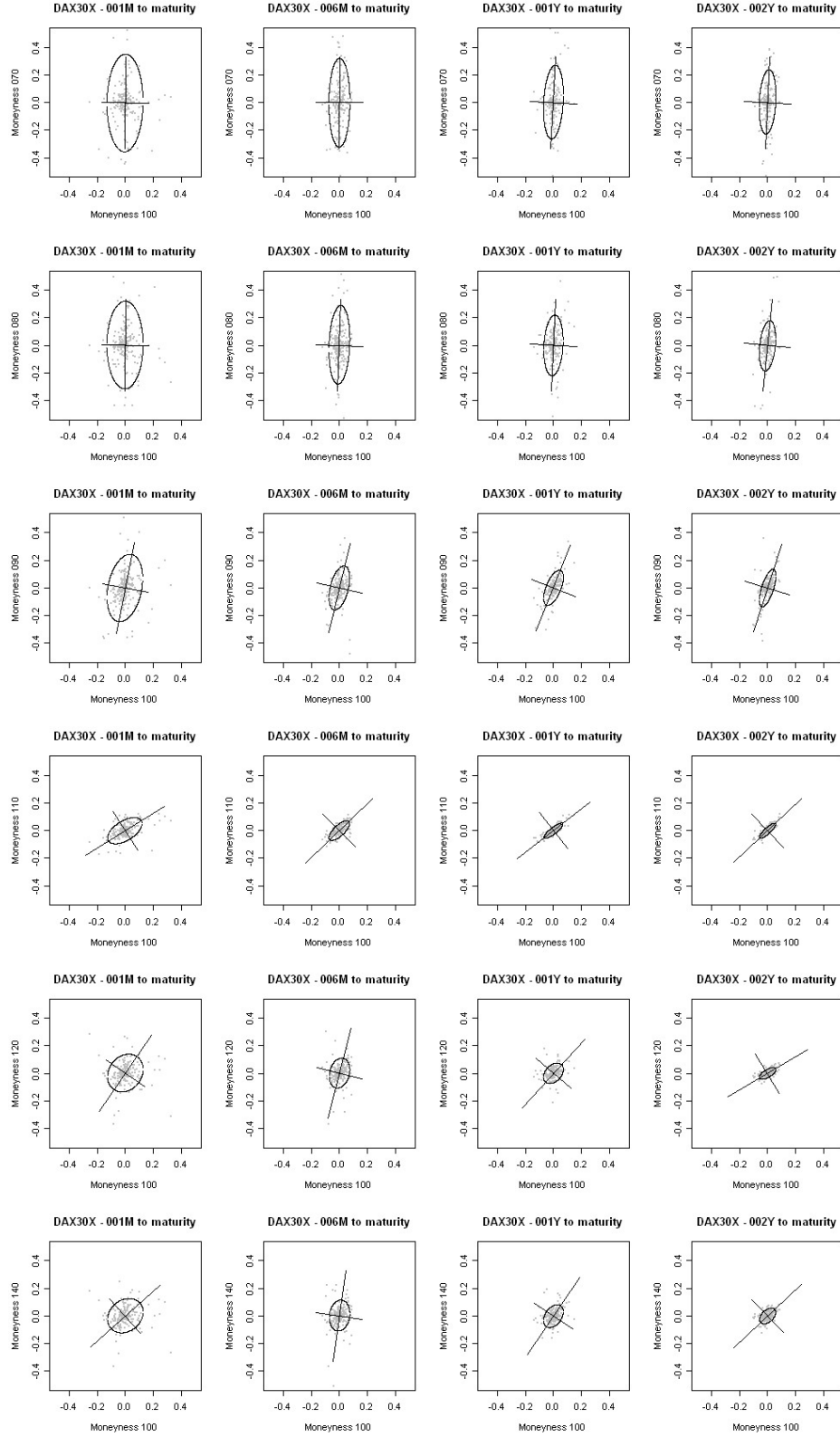


Figure 9: Scatterplot of Dax IV log returns, main directions of variation and 95% confidence region. Panels in different rows show plots of moneyiness $m_1 = 1$ vs. different $m_2 \in \{0.7, 0.8, 0.9, 1.1, 1.2, 1.4\}$. Panels in different columns show plots for different maturity groups, $\tau \in \{1M, 6M, 1Y, 5Y\}$.

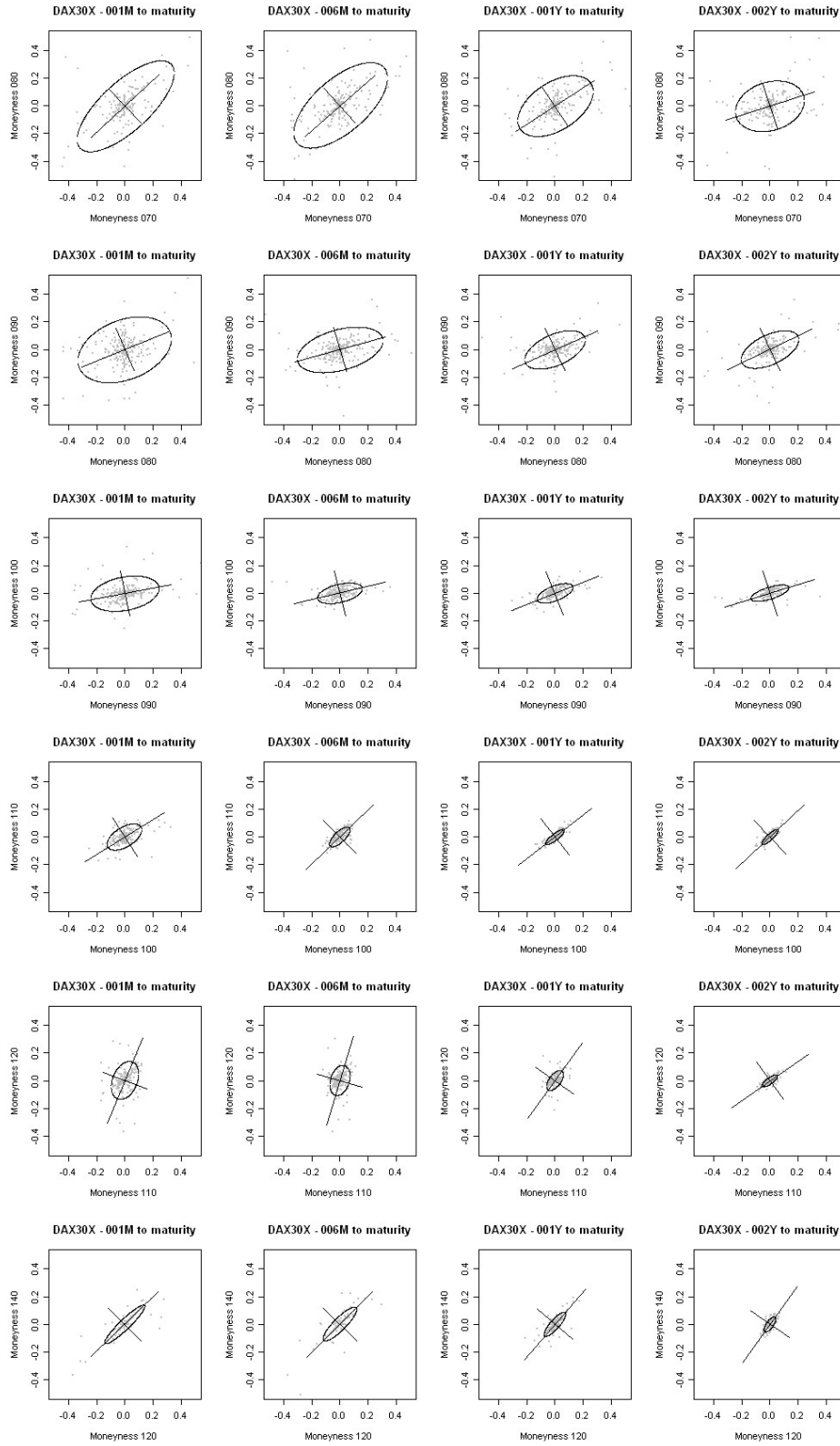


Figure 10: Scatterplot of Dax IV log returns, main directions of variation and 95% confidence region. Panels in different rows show scatterplots of for different neighbored moneyness values, i.e. 0.7 vs. 0.8, 0.8 vs. 0.9, etc. Panels in different columns show plots for different maturity groups, $\tau \in \{1M, 6M, 1Y, 5Y\}$.

Similar considerations are made for swap data. We made scatterplots for different fixed option maturities and plot observations for pairs of swap maturities against each other. This time the volatility of the data is much less than for index and swap data, but we can detect the same features concerning main directions and dependencies among groups. Eigenvectors are similar for identic swap maturities and different option maturities. And even for different swap maturities eigenvectors are almost identic. Linear dependencies are strongest among neighboring groups of swap maturities, they seem to intensify for long swap maturities. Obviously a (partial) common PCA as introduced in [17] could be appropriate to analyse the modes of variation of the data.

Now we take a closer look at the PCs to describe these different modes of variation that mainly influence the data, i.e. the shocks, that drive an IV smile for fixed option maturity in time. The number of chosen PCs to describe the data is determined by the explained variance, i.e.

$$Var_j = \frac{\rho_j}{\sum_{k=1}^n \rho_k}, \quad j \in \{1, \dots, n\}$$

for single PCs ϕ_j and

$$Var_{1:K} = \frac{\sum_{k=1}^K \rho_k}{\sum_{k=1}^n \rho_k}, \quad 1 \leq K \leq n$$

for the first K PCs ϕ_1, \dots, ϕ_K , where ρ_k is the k -th eigenvalue and $\sum_{k=1}^n \rho_k$ is the total variation. We choose the first K PCs to describe the data, if $Var_{1:K} \geq 0.95$.

In figures 13 to 15 eigenvalues of the data and the according explained variances are plotted. The first PC of DAX30X data already explains more than 60% of the total variation. For maturities shorter than one year the first four PCs explain about 95% of the variation, for maturities more than one year the first five PCs are sufficient to explain the variation. In case of Volkswagen data, for all maturity groups the first four PCs explain about 95% of the total variation. In case of swap data, here JPY swaps, the first PC already describes about 80% of the total variation, 95% are reached by PCs 1 to 3.

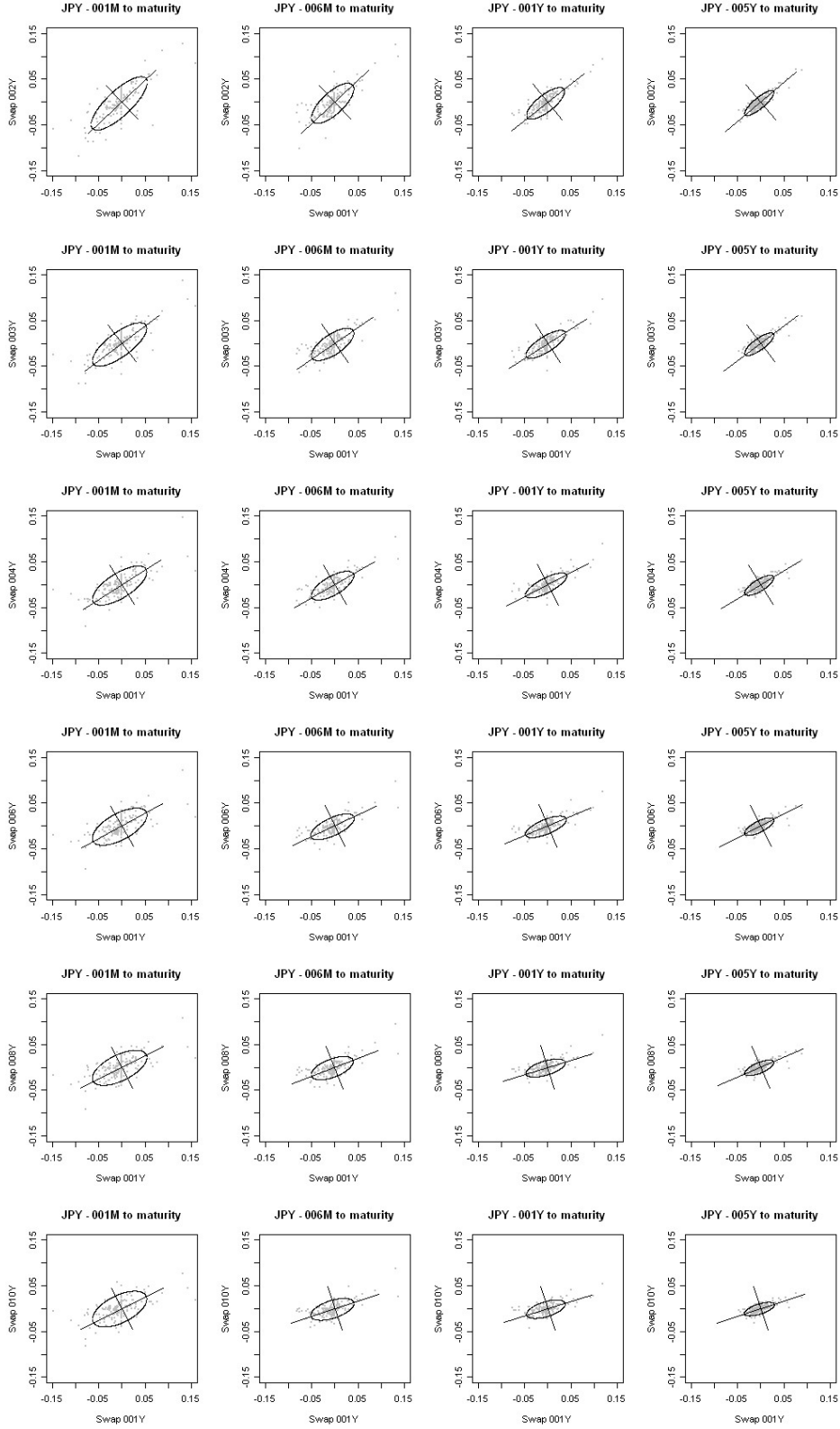


Figure 11: Scatterplot of JPY IV log returns, main directions of variation and 95% confidence region. Panels in different rows show scatterplots for $\tau_1^S = 1$ year vs. different $\tau_2^S \in \{2, 3, 5, 10, 20, 30\}$ years. Panels in different columns show plots for different option maturity groups, $\tau^O \in \{1M, 6M, 1Y, 5Y\}$.

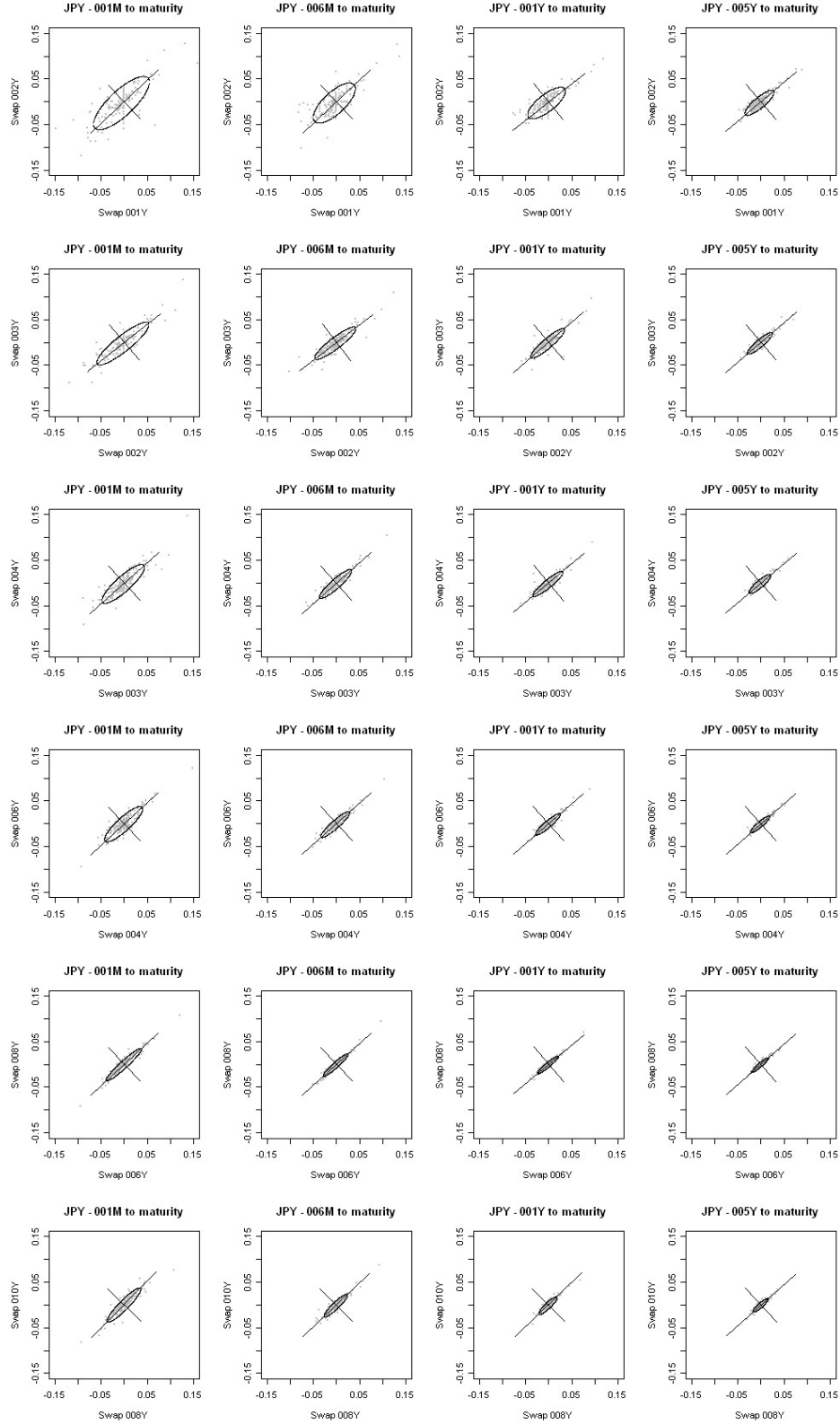


Figure 12: Scatterplot of JPY IV log returns, main directions of variation and 95% confidence region. Panels in different rows show scatterplots of for different neighbored swap maturities values, i.e. 1Y vs. 2Y, 2Y vs. 3Y, etc. Panels in different columns show plots for different option maturity groups, $\tau^O \in \{1M, 6M, 1Y, 5Y\}$.

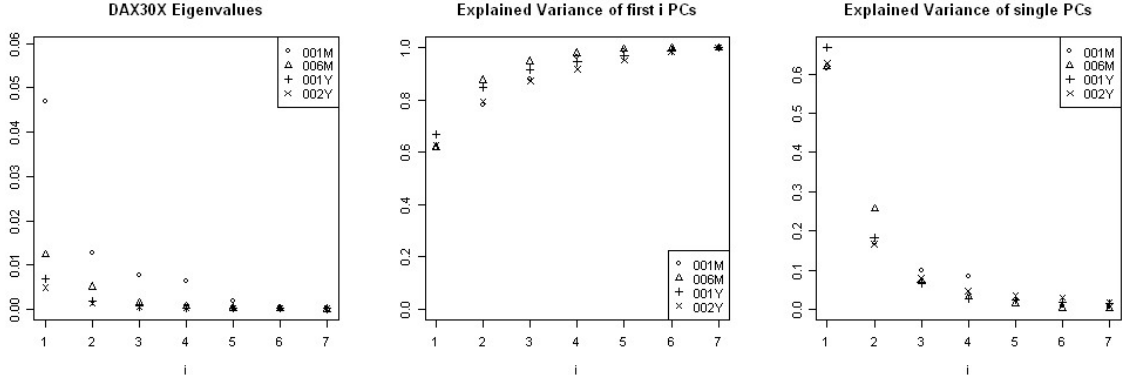


Figure 13: Eigenvalues of Dax IV log returns for different option maturities (left), explained variance explained by the first few eigenvalues (middle), i.e. $Var_{1:i}$, $i = 1, \dots, 7$, and variance explained by single eigenvalues (right), i.e. V_i , $i = 1, \dots, 7$.

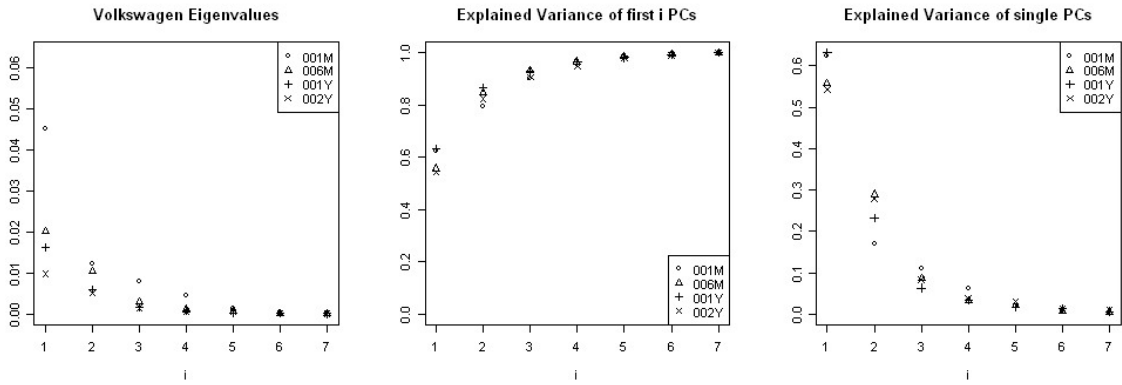


Figure 14: Eigenvalues of Volkswagen IV log returns for different option maturities (left), explained variance explained by the first few eigenvalues (middle), i.e. $Var_{1:i}$, $i = 1, \dots, 7$, and variance explained by single eigenvalues (right), i.e. V_i , $i = 1, \dots, 7$.

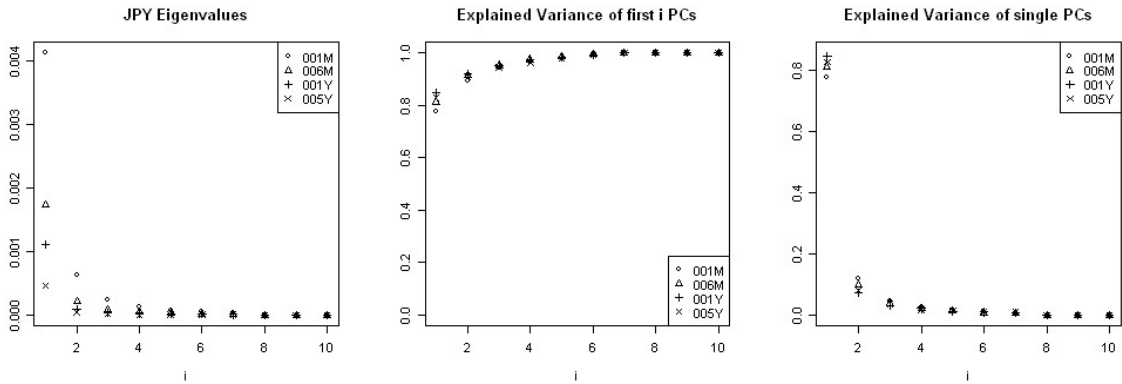


Figure 15: Eigenvalues of JPY IV log returns for different option maturities (left), explained variance explained by the first few eigenvalues (middle), i.e. $Var_{1:i}$, $i = 1, \dots, 10$, and variance explained by single eigenvalues (right), i.e. V_i , $i = 1, \dots, 10$.

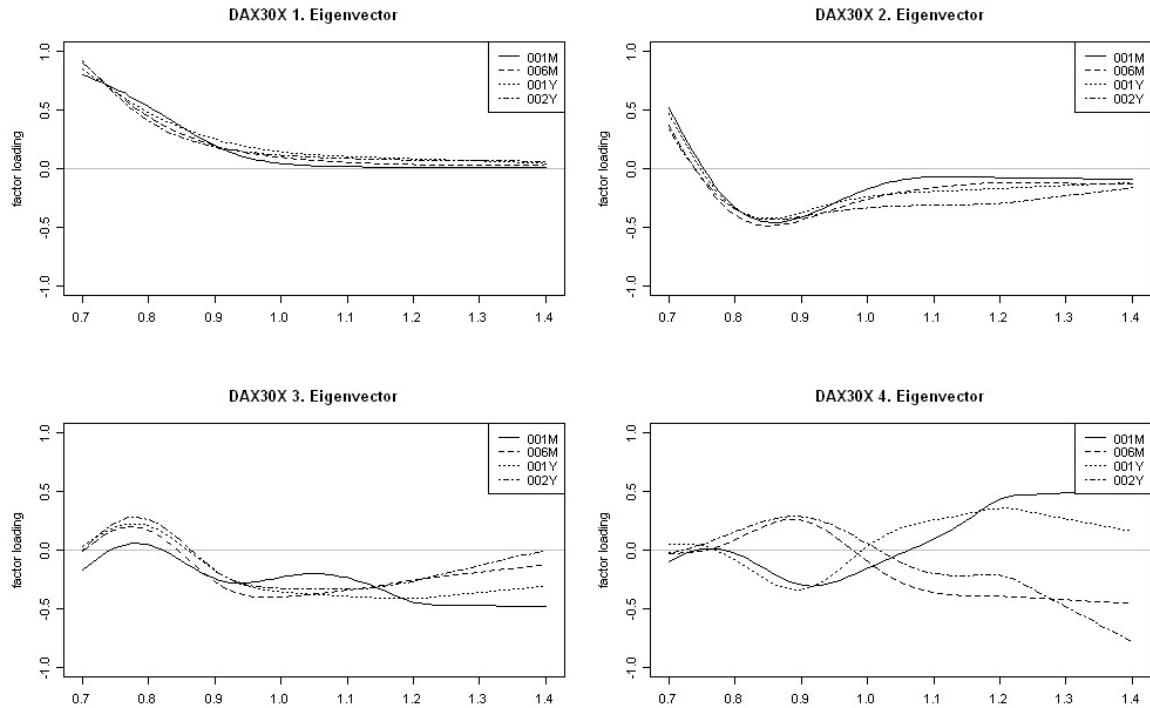


Figure 16: First four eigenvectors of Dax IV log returns for different option maturities $\tau \in \{1M, 6M, 1Y, 2Y\}$, smoothed by local linear kernel estimation.

We obtain smooth eigenfunctions from the calculated eigenvectors by local linear kernel estimation, and plot the smoothed functions according to different maturity groups together in one plot. Eigenfunctions of index and stock data are shown in figures 16 and 17, swap eigenfunctions are shown in figure 18. We only show the first four eigenfunctions as they explain most of the variation of the IV smile. Note, that eigenvectors can be mirrored at the x-axis by just changing their sign without changing the result.

First of all regard the DAX plots in figure 16. In general, the eigenfunctions for different option maturities show the same behavior, although eigenfunctions according to maturity $\tau = 1M$ always slightly differ from the other eigenfunctions. The first eigenfunction is highest for low moneyness values then decreases and reaches 0 at about $\kappa = 1$. The interpretation is that the curvature of the smile varies for in-the-money values and remains constant out-of-the-money. The second eigenfunction explains again a changing curvature in the first part of the smile and slightly up-and-down shifts in the second part, where $\kappa \geq 0.9$. The third eigenfunction again corresponds to up-and-down shifts of IV smiles at $\kappa \geq 0.9$ and thus strengthens the effect of the second eigenfunction. The fourth eigenfunction forms contrasts in the left and the right part. It therefore describes variation of the left and right part of the smile in opposite directions.

We resume that DAX30X IV smiles are mainly driven by shocks in the curvature which are strongest in the left part, i.e. in-the-money, and further by little up-and-down shifts of middle and right parts of the curve.

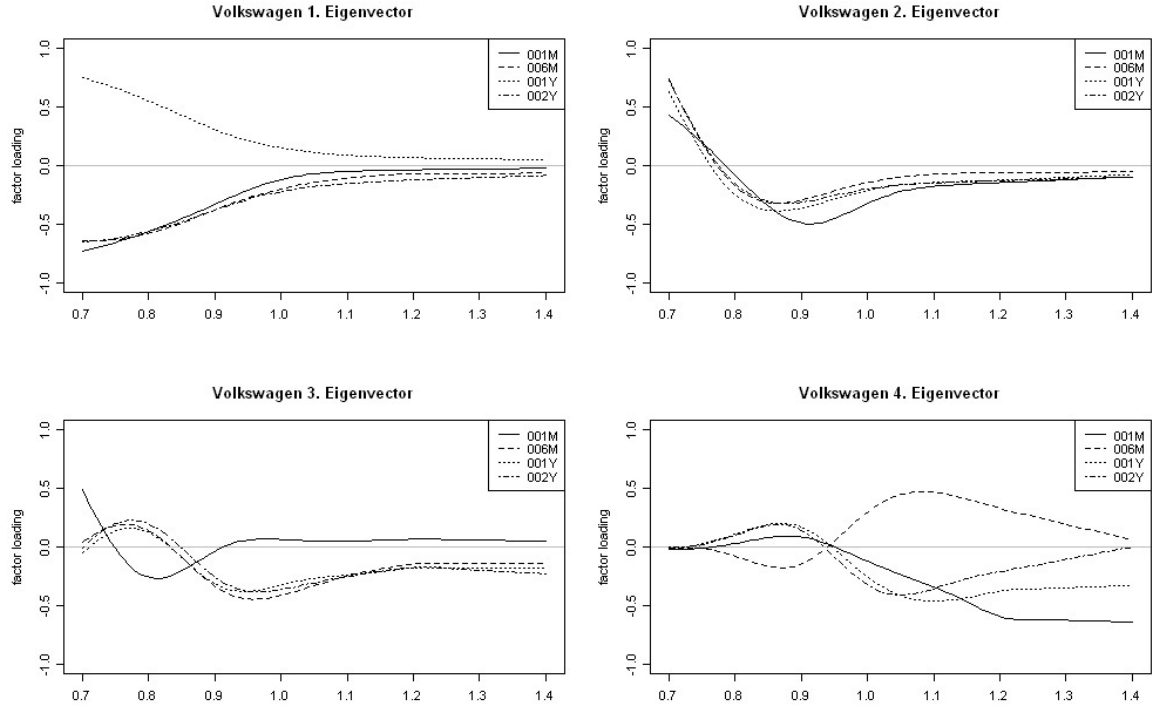


Figure 17: First four eigenvectors of Volkswagen IV log returns for different option maturities $\tau \in \{1M, 6M, 1Y, 2Y\}$, smoothed by local linear kernel estimation.

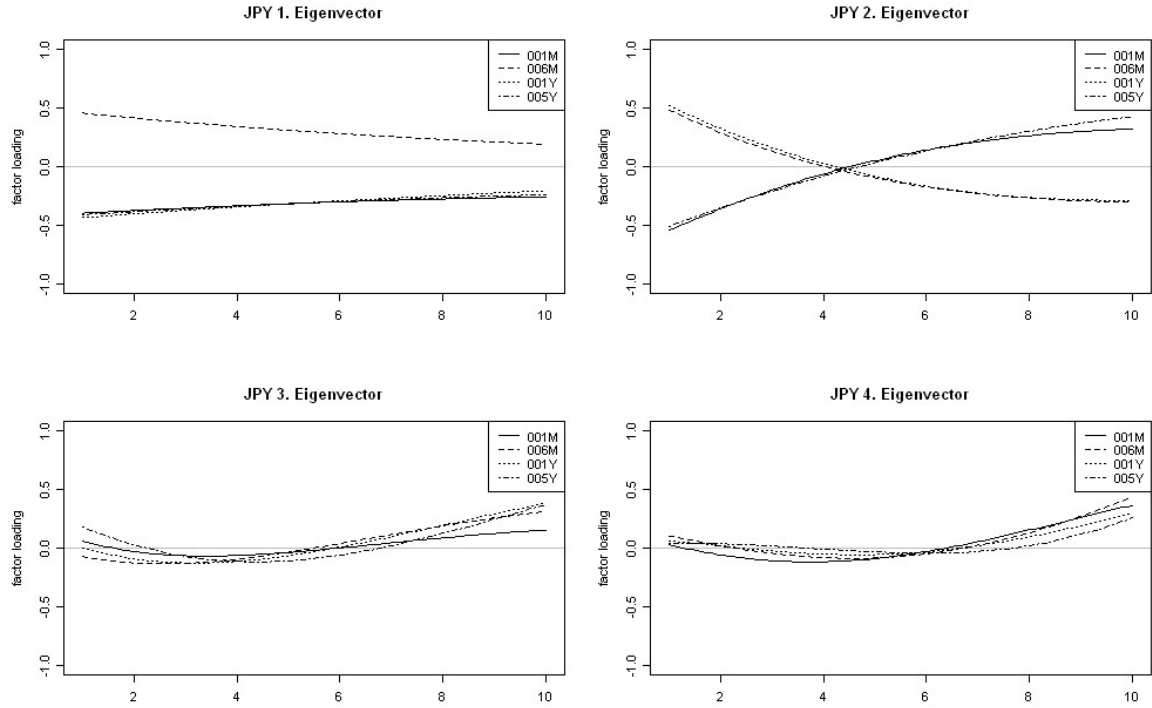


Figure 18: First four eigenvectors of JPY IV log returns for different option maturities in $\tau^O \in \{1M, 6M, 1Y, 5Y\}$, smoothed by local linear kernel estimation.

An example for stock IVs is given in figure 17 where we plot eigenfunctions of Volkswagen stock data. Again eigenfunctions are similar for different option maturities except $\tau = 1M$. The shocks explained by single eigenfunctions match those which we detected for DAX30X index data. The first eigenfunction corresponds to a changing curvature of the in-the-money part of the smile curve. The second and third eigenfunctions describe uniform shifts of the smiles for $\kappa \geq 0.9$, and changing curvature underneath. In case of short option maturities, here $\tau = 1M$, the up-and-down shifts are much less pronounced, but still existent.

Now regard JPY swap eigenfunctions in figure 18. We already mentioned that less PCs are necessary to explain a sufficient amount of variation. In general, the eigenfunctions are flatter than those of index and stock data. For different option maturities all eigenfunctions nearly coincide. Eigenfunctions are given over swap maturity values. Over the whole range the first eigenfunction is of the same sign. It corresponds to up-and-down shifts of the whole IV curve with respect to swap maturity. The second eigenfunction changes sign at τ^S shortly after 4Y. The effect on the IV curve is a changing slope. The third and fourth eigenfunctions are largely the same, and close to the zero line, except for very large swap maturity. These eigenfunctions can only explain very little changes in curvature of IV curves in time.

Resuming, the main shocks for swap IV curves with respect to swap maturity are up-and-down shifts and a changing slope of the whole curve.

In the analysis so far we detected large similarities among the different option maturity groups. The question arises how to pool the groups in PCA, with respect to same modes of variation. To clarify the situation we regard the PC scores that, if clusters appear, give us a hint of how to group the data. Figures 19 to 21 show scatterplots of first and second scores according to first and second PC. Scores that come from short option maturity data are plotted by small letters a, b, c, d , corresponding to $\tau \in \{2W, 1M, 2M, 3M\}$ for indexes and stocks and $\tau^O \in \{1M, 3M, 6M, 1Y\}$ for swaps. Scores according to long option maturities are plotted by big letters A, B, C, D , i.e. $\tau \in \{6M, 9M, 1Y, 2Y\}$ for indexes and stocks and $\tau^O \in \{2Y, 3Y, 4Y, 5Y\}$ for swaps. We plotted the scores at different days, namely at the first trading days of each month from January 2006 to October 2006. Even if clusters do not appear regularly, there are some days, when short and long option maturity groups appear separately in the plots. We detected the same feature in scatterplots of PC 3 scores vs. PC 4 scores. It thus lies on hand that grouping makes sense for data with respect to short and long option maturities.

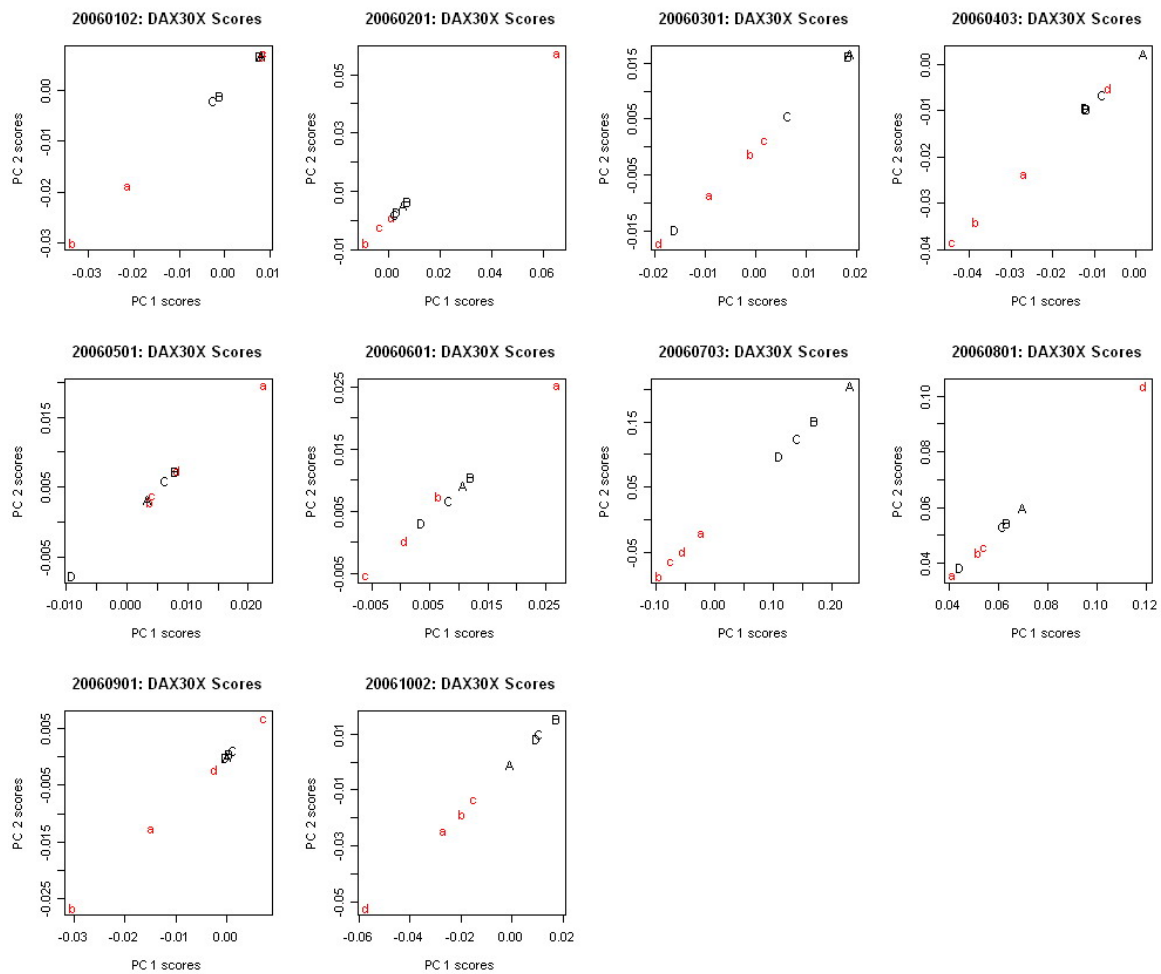


Figure 19: Scatterplot of PC scores of DAX30X IV from 2 January 2006 (20060102) to 2 October 2006 (20061002) at first trading days of each month, respectively.

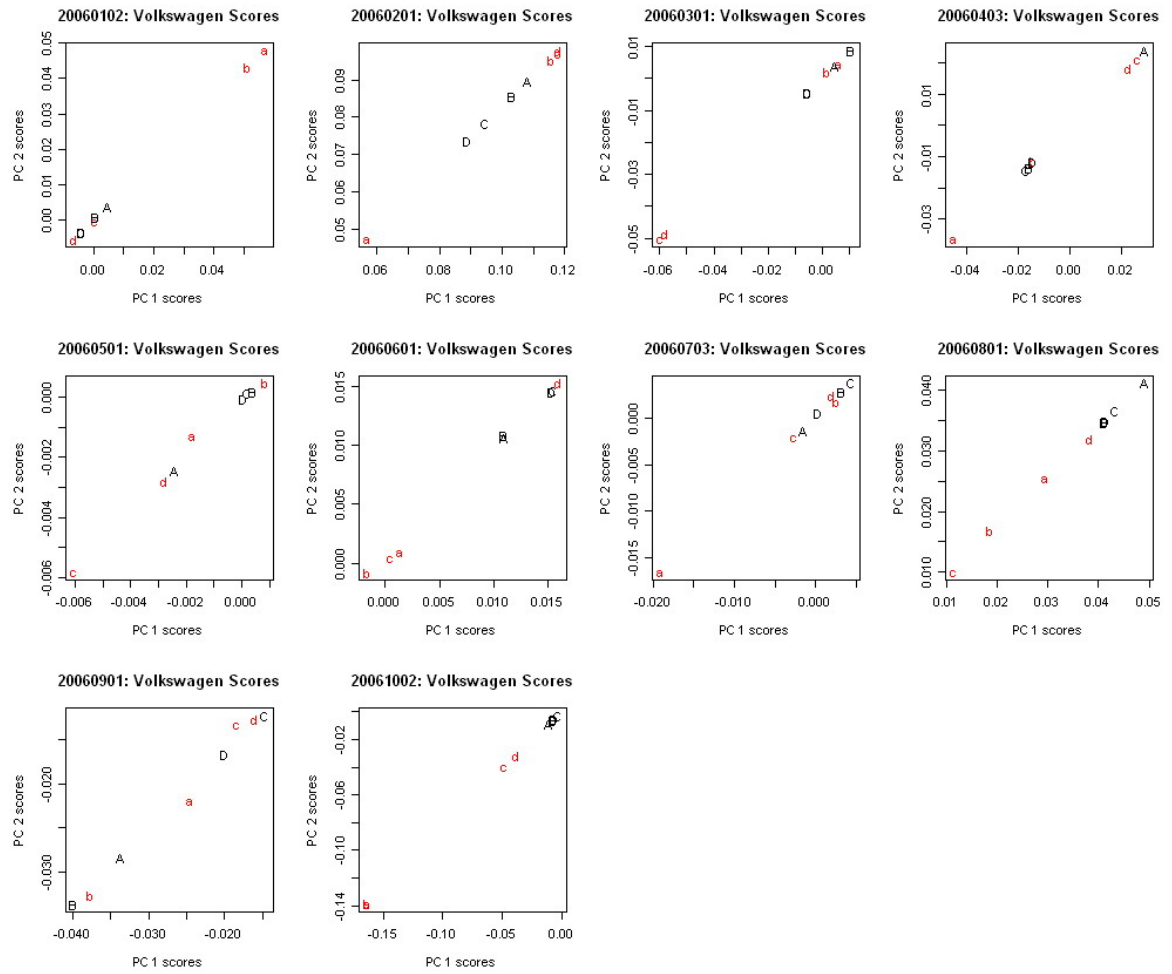


Figure 20: Scatterplot of PC scores of VOLKSWAGEN IV from 2 January 2006 (20060102) to 2 October 2006 (20061002) at first trading days of each month, respectively.

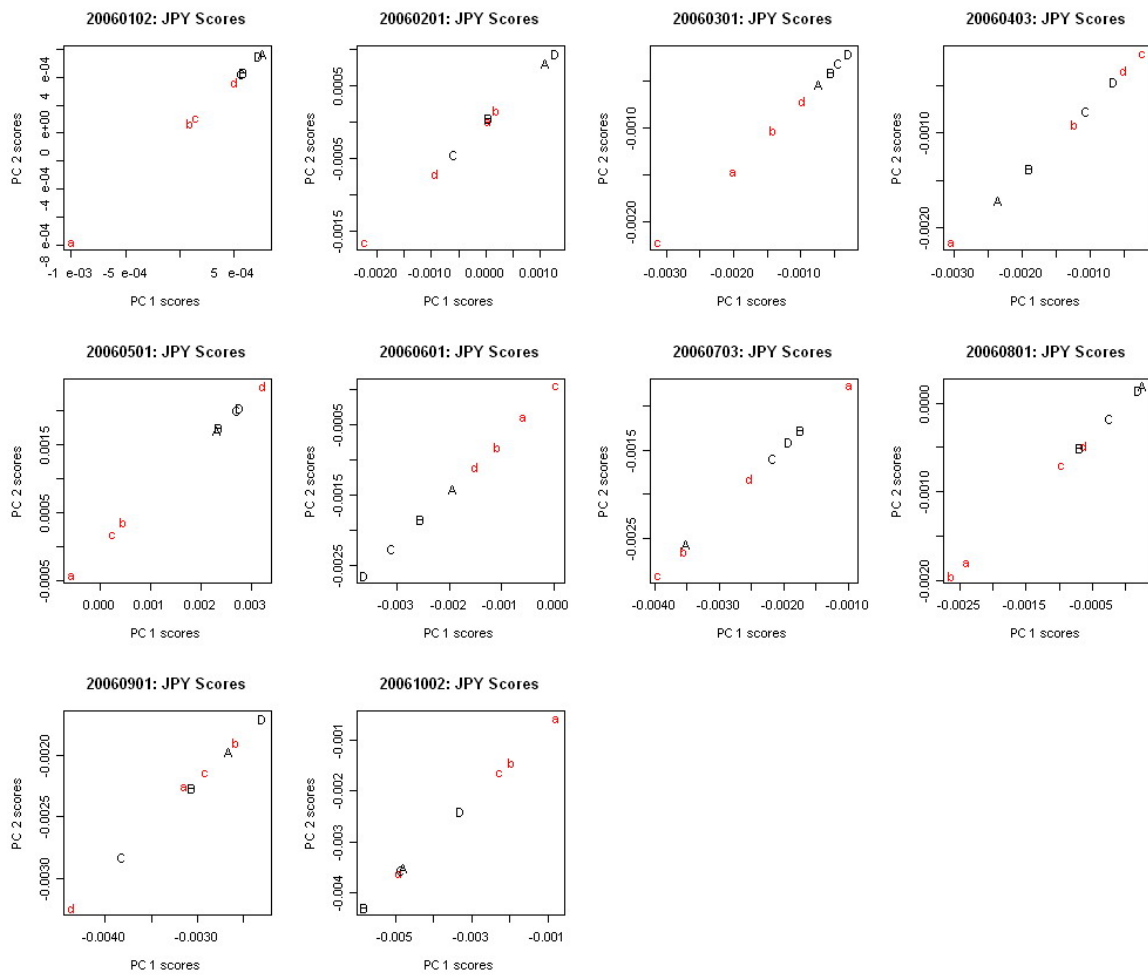


Figure 21: Scatterplot of PC scores of JPY IV from 2 January 2006 (20060102) to 2 October 2006 (20061002) at first trading days of each month, respectively.

5.3 Application of Functional Variance Processes

Until now we analysed the variation of single IV curves in time. We detected, that the curves for different option maturities are driven by similar modes of variation while the strength of variation varies. Now we want to analyse how changes of IV curves for different option maturities interact, to be more precise we do not analyse variation in time, but how IV log return curves vary from one option maturity to another within a single surface at day t .

To this end we will to analyse the additional noise in the data, i.e. we will first of all smooth single curves to obtain residuals, which we then analyse according to a common variance process that determines variation of curves within the surface. We proceed according to methods of Müller et al. [34], presented in section 3. As before we restrict the data to maximal two years to option maturity and maximal ten years to swap maturity.

Let $Z_{ij} = \sigma(\kappa_j, \tau_i)$ the log return IV for an option with moneyness κ_j and option maturity τ_i . We suppress the time index t , as we analyse the data at one single, fixed day. Further let $S_i(\kappa_j) = S(\kappa_j, \tau_i)$ smooth trajectories corresponding to option maturity groups $\tau_i, i = 1, \dots, n$. According to (3.1) and (3.2) we model the data

$$\begin{aligned} Z_{ij} &= S_i(\kappa_j) + R_{ij} \\ Y_{ij} &= \log(R_{ij}^2) = V(\kappa_j) + W_{ij}. \end{aligned}$$

In case of swap IVs replace moneyness κ_j by swap maturity τ_j^S and denote option maturity by τ_i^O . The additional noise R_{ij} is assumed to be the product of the exponentiated random function V and an exponentiated white noise W . V is common among all curves that build the IV log return surface. The regressors $\kappa_j, j = 1, \dots, m$, are equal for each path i . Note that asymptotics of functional variance processes hold for regressors that differ from path to path, see section 3.3. Therefore application would be possible for the non-smoothed, directly observed IV data, as well. One only would have to group the data into naturally given option maturity groups.

For the various local polynomial kernel estimators needed to approximate model components we use the Gaussian kernel function although in theory the Epanechnikov kernel performs better, compare section 2.1. The reason for this choice is that we want to avoid estimation problems in sparse regions of our data. Bandwidths are chosen by cross validation as described in section 3. For stock and index IV we have prediction points $\kappa_j \in \{0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.4\}$, thus we choose optimal bandwidths in the interval $[0.05, 1]$. For the swap IV data we have prediction points $\tau_j^S \in \{1, 2, \dots, 10 \text{ years}\}$ and choose optimal bandwidths in $[.5, 10]$.

We perform the analysis using the statistical software R. For efficiency reasons extensive subroutines for local polynomial estimation are programmed in C and embedded into our R routines. Eigenfunctions and eigenvalues are discretely calculated using the R function *eigen()* and functional forms are produced by local linear kernel estimation.

In practice when handling discrete data we have to omit zero residuals in (3.1) because we

want to log transform them. Thus after taking the log we filter $Y_{ij} = -\infty$ and replace by $Y_{ij}^* = \min\{Y_{ik} | Y_{ik} \neq -\infty, k = 1, \dots, m\}$.

The white noise variance is originally estimated by the integral

$$\hat{\sigma}_W^2 = \frac{1}{|T_1|} \int_{T_1} \left(\hat{Q}_V(\kappa) - \hat{G}_V(\kappa, \kappa) \right)_+ d\kappa,$$

where T_1 the shortened domain of regressors κ_j . We approximate the integral by a Riemann sum and estimate

$$\tilde{\sigma}_W^2 = \frac{1}{|T_1|} \sum_{k=1}^{m'} \left(\hat{Q}_V(\kappa_{j_k}) - \hat{G}_V(\kappa_{j_k}, \kappa_{j_k}) \right)_+ w_k, \quad (5.1)$$

where j_k , $k = 1, \dots, m'$, the indices of ordered regressors κ_j that fall into T_1 , $m' \leq m$. Weights are given according to $w_1 = (\kappa_{j_2} - \kappa_{j_1})/2$, $w_{m'} = (\kappa_{j_{m'}} - \kappa_{j_{m'-1}})/2$ and $w_k = (\kappa_{j_{k+1}} - \kappa_{j_{k-1}})/2$, $k = 2, \dots, m' - 1$.

Additionally, we want to analyse the change of the functional variance process over time. Therefore, in a second calculation of FVPs, we decide to take fixed bandwidths to estimate m_V and G_V to get comparable estimates of V . By visual inspection of the data we decide to take $b_V = 0.1$ and $h_V = 0.5$ for index and swap data, for swap data we choose $b_V = 2$ and $h_V = 0.5$.

How to interpret the results? What does changes of V mean to the IV data? To make the influence of V onto log return data directly visible we plot a function according to model (3.1) over the range $[-40, 0]$ wherein realisations V_i in our data example fall, see figure 22. We approximate the additive noise $R_{ij} \approx \sqrt{\exp V_i(\kappa_j)} \approx (V_i(\kappa_j))^{-2}$ over the given range, thus, for $V < 0$ increasing values correspond to increasing noise in the observed IV log returns $Z_{ij} = \sigma(\kappa_j, \tau_i)$.

5.3.1 Index IV example: DAX30X

As an example for index IV data we take the DAX30X data set introduced in section 5.1. In figure 23 we plotted observed log returns of implied volatility smiles at 1 September 2006 and the corresponding smoothed trajectories S_i for option maturity from two weeks to two years. The data is situated around zero with strong outliers in case of two weeks and one month option maturity. Data adaptively chosen bandwidths differ from 0.05 to 1. In general small bandwidths are chosen for short option maturities and estimated trajectories have stronger curvature. For long maturities bandwidth $b_{S_i} = 1$, $i = 5, \dots, 8$, is chosen, this means that for estimation all observations were similarly weighted and estimated trajectories resemble a straight line.

As already mentioned in previous analysis it seems reasonable to divide the data into a short maturity group, i.e. $\tau \in \{2W, 1M, 2M, 3M\}$, and a long maturity group, i.e. $\tau \in \{6M, 9M, 1Y, 2Y\}$. We thus continue the functional analysis separately for the two groups.

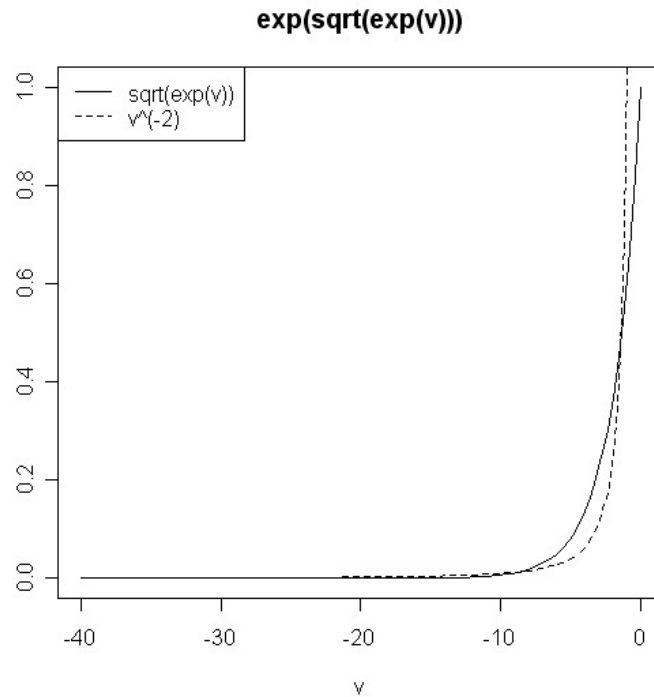


Figure 22: How does the changes in the functional variance process V determine the IV data. The functional relationship of the noise R of IV log returns and V at $[-40, 0)$ is approximately $\sqrt{e^V} \approx V^{-2}$.

In figure 24 estimates of the mean variance function m_V and variance trajectories V_i are drawn. Results for short and long option maturities in top and bottom panels, respectively, are calculated independently. Trajectories V_i lightly move up and down, while their the shape stays almost similar. We note that resulting variance processes for short and long maturity groups differ. The short maturity variance function has a light curvature and is decreasing, the data adaptively chosen bandwidth to estimate m_V is a small one, namely $b_V = 0.15$. For long maturities the variance function is almost a straight line and lightly increasing for increasing moneyness. Data-adaptively chosen bandwidth is of middle size, $b_V = 0.4$. The bandwidth to estimate the covariance structure \mathcal{G}_V is $h_V = 0.05$ in both cases.

Plots of $\hat{\mathcal{G}}_V$ and the corresponding correlation surface are shown in figure 25 and 26. For short option maturities there is a large positive correlation throughout all values. Thus if a trajectory of the variance process is larger than the average in one point we can expect it to be larger than average everywhere. Correlation falls off only slightly for separating values κ_i . The variance process shows a slight departure from stationarity. In contrast the variance process for long option maturities seems to be stationary. Further the correlation falls off rapidly for separating moneyness values and reaches zero at opposite boundaries, i.e. $\kappa_1 = 0.7$ and $\kappa_2 = 1.4$.

Eigenfunctions are presented in figure 27. Only those eigenfunctions are plotted that are chosen to approximate V according to (3.12). In case of short maturities the first principal

component explains about 80% of the total variance and first three PCs together are sufficient to describe the variation in the data. The first eigenfunction is flat and positive. Thus main shocks that drive the variance process are up-and-down shifts. The second eigenfunction is close to the zero line, slightly negative for $\kappa \leq 1.2$. It thus strengthens the shifts of the variance curve at the left side. The third eigenfunction forms two contrasts, first one for $\kappa \leq 1.1$, second one for $\kappa \geq 1.1$. When trajectories V_i shift up in the left part, then they shift down in the right part and vice versa. The effects of the PCs are shown in figure 28, where we plot the mean curve m_V and curves according to $m_V(\kappa) \pm C\phi_k(\kappa)$ for a constant number $C > 0$, corresponding to the modes of variation explained by single PCs ϕ_k .

For long option maturities the first PC also determines up-and-down shifts, that explain 55% of the variation of the data. Further variation is explained by the second eigenfunction which is decreasing, crossing zero between $\kappa = 0.9$ and 1., i.e. it corresponds to a changing slope of original curves, see figures 27 bottom panels and figure 29.

Another quantity of interest is the constant white noise variance that is estimated from the covariance diagonal, see (5.1). Unfortunately we have very few data to estimate the variance and thus we have to handle the estimates with some mistrust. Here estimates are $\hat{\sigma}_W^2 = 3.104$ in case of short maturities and $\hat{\sigma}_W^2 = 1.919$ for long maturities.

Resuming we note that the variance process that determines variation among IV log return curves slightly differs from short to long option maturities. It shows approximately stationarity, which is more clearly for long maturities. Thus the mean curve m_V has rather little slope in both cases, for short option maturities m_V is slightly larger in short maturity regions. Variation between trajectories V_i are mainly driven by up-and-down shifts of the whole range. In case of long option maturities a changing slope is a second intense mode of variation. For the original IV data we have the following the variation from one IV log return curve to another is larger in-the-money than out-of-the-money. Thus time dependent variation of the IV smiles in the implied volatility surface differs stronger from one smile to another in regions where option maturity and moneyness is small.

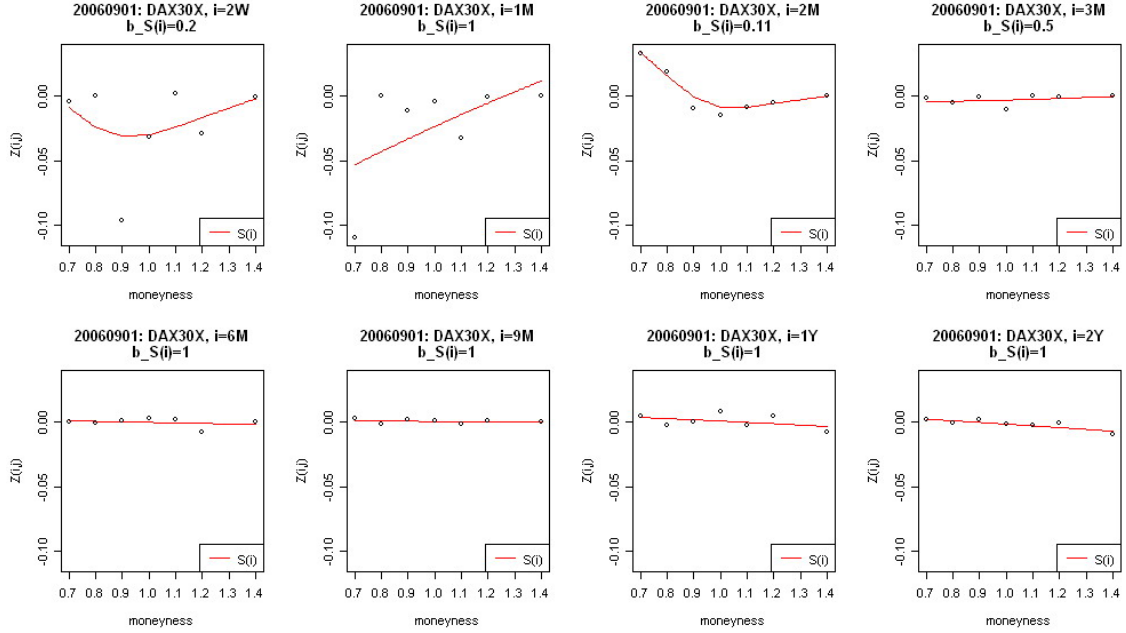


Figure 23: Smoothed log returns of smiles for DAX30X IV data at 1 September 2006 for option maturity from two weeks (2W) to two years (2Y). Data are discretely plotted (dots) and smoothed trajectories are continuously estimated by local linear kernel estimation (solid line). Bandwidths for each trajectory are chosen by cross validation.

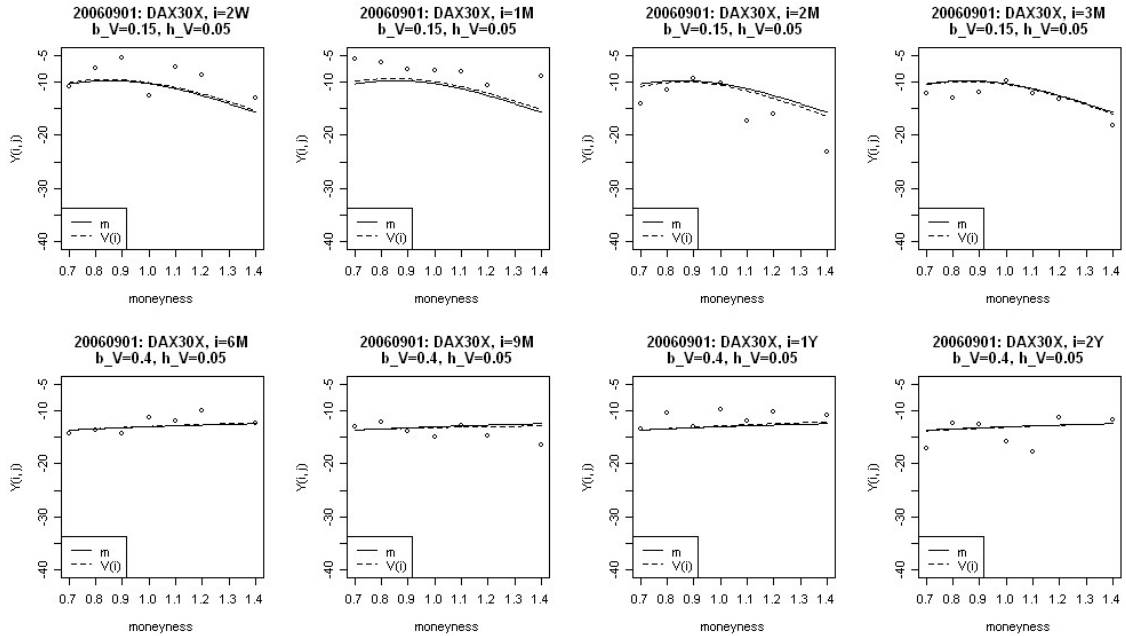


Figure 24: DAX30X residuals at 1 September 2006 for different small option maturities (top panels) and different long option maturities (bottom panels). Estimated mean variance function m_V (solid line) and variance trajectories V_i (dashed line) are drawn into the plots. Estimates for small maturities and long maturities were made independently. All bandwidths are chosen by cross validation.

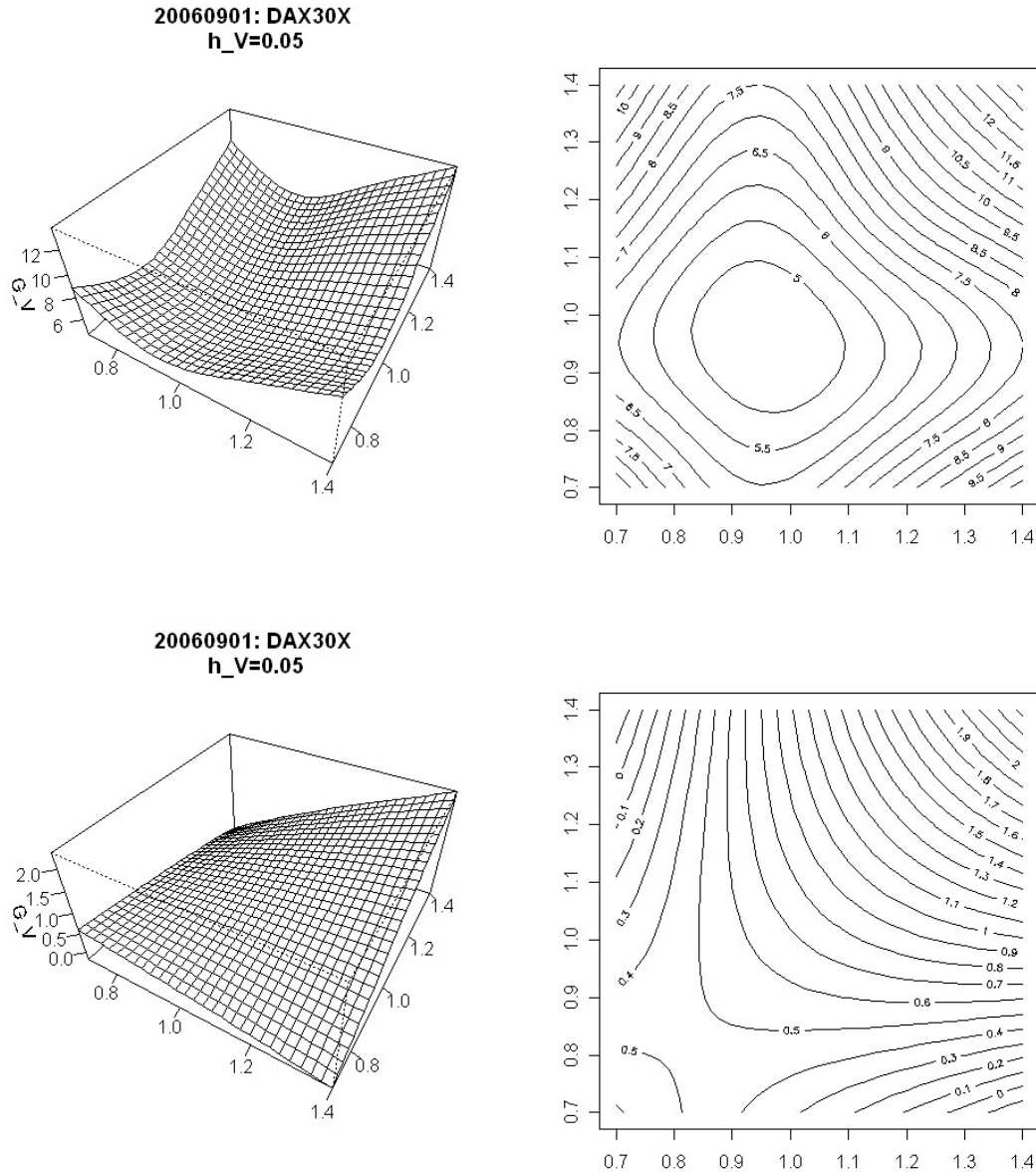


Figure 25: Estimated covariance of functional variance process for short maturities (top panels) and long maturities (bottom panels). The bandwidth h_V to estimate the surface is chosen by cross validation.

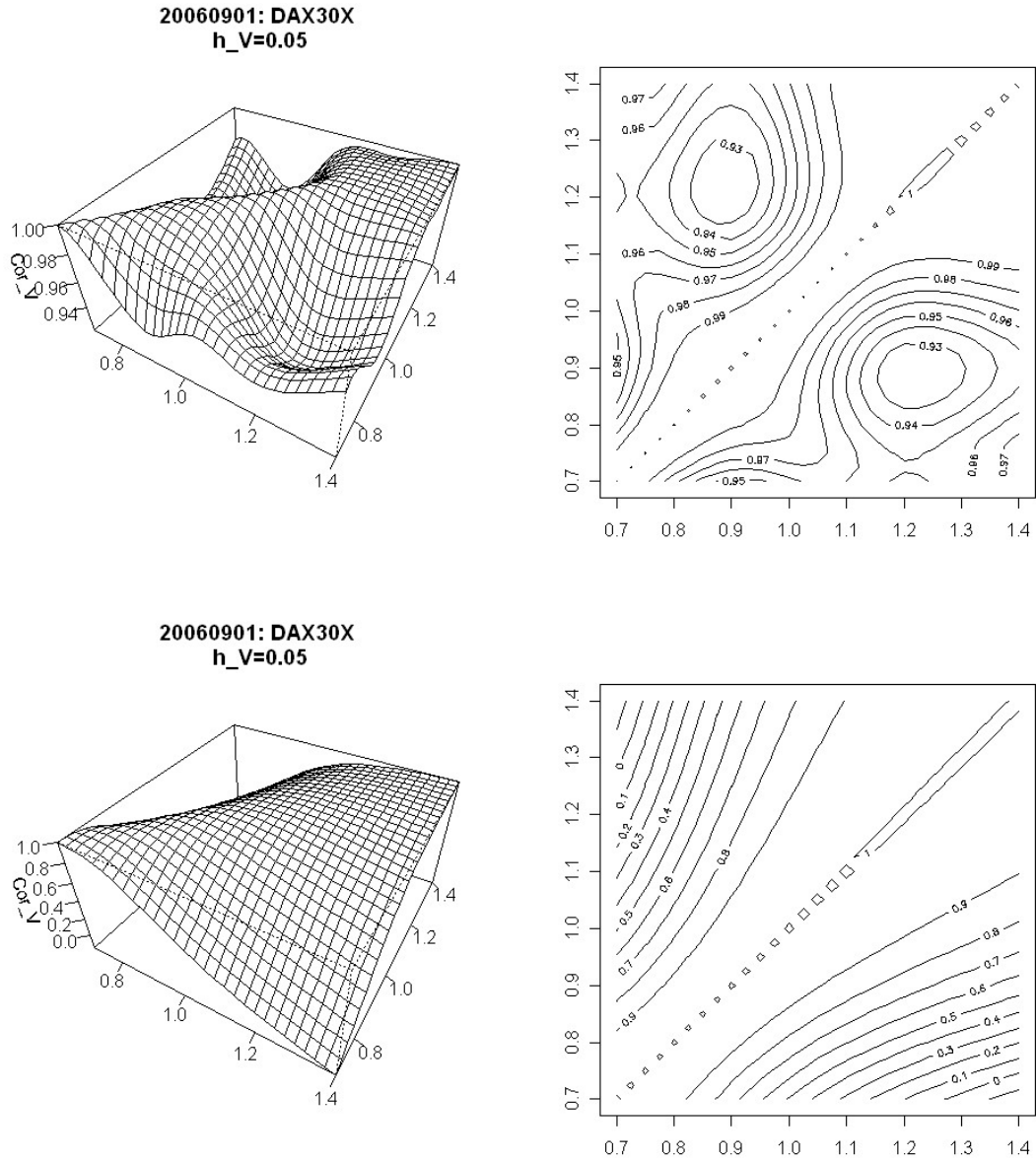


Figure 26: Estimated correlation of functional variance process for short maturities (top panels) and long maturities (bottom panels).

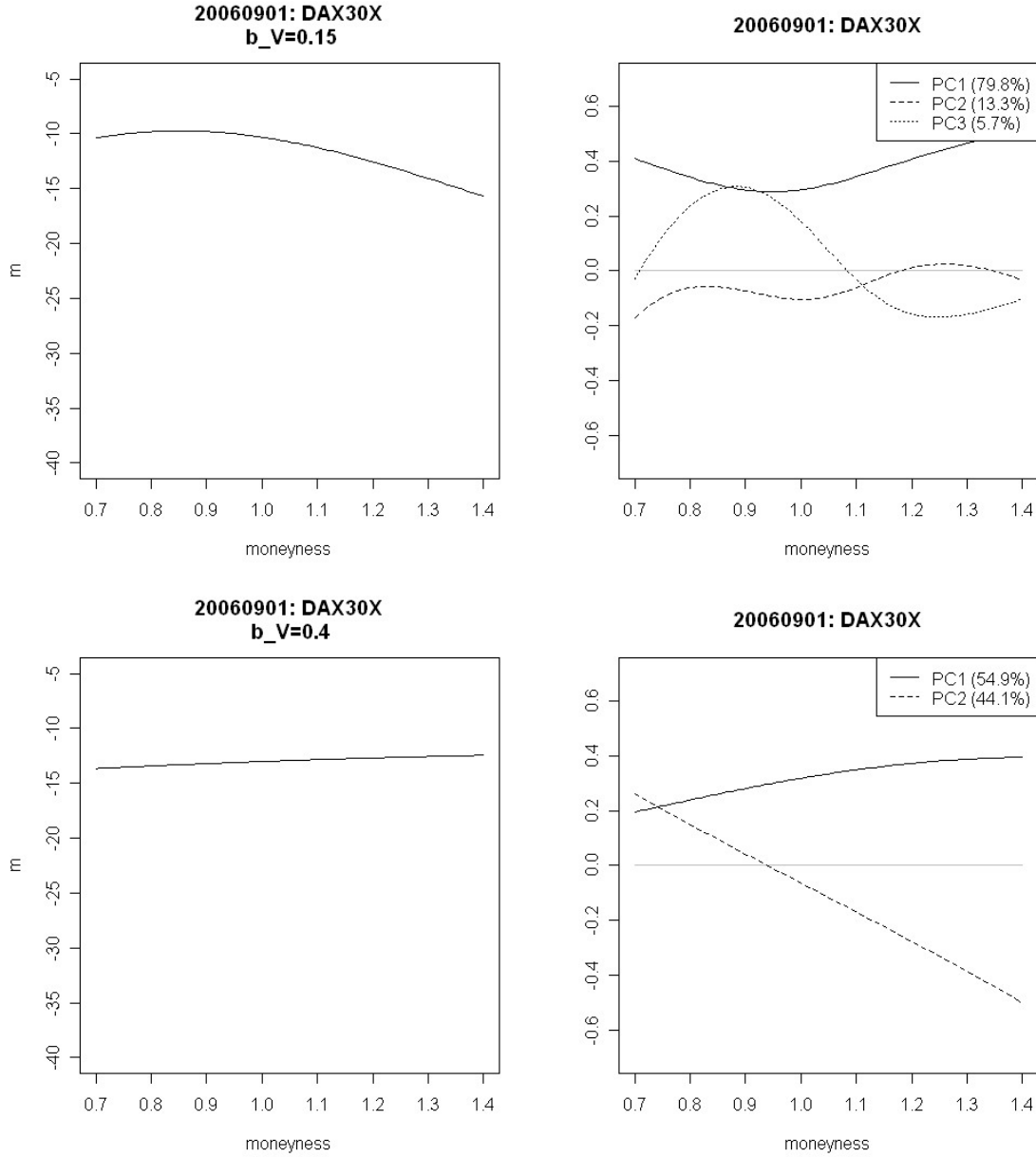


Figure 27: Functional variance process (left) and eigenfunctions (right) for small option maturities (top panels) and for large option maturities (bottom panels). Functional variance process V is approximated by local linear kernel estimation from single trajectories V_i . Bandwidth h is chosen by cross validation. First eigenfunction (solid line) and second eigenfunction (dashed line) determine the dominant modes of variation.

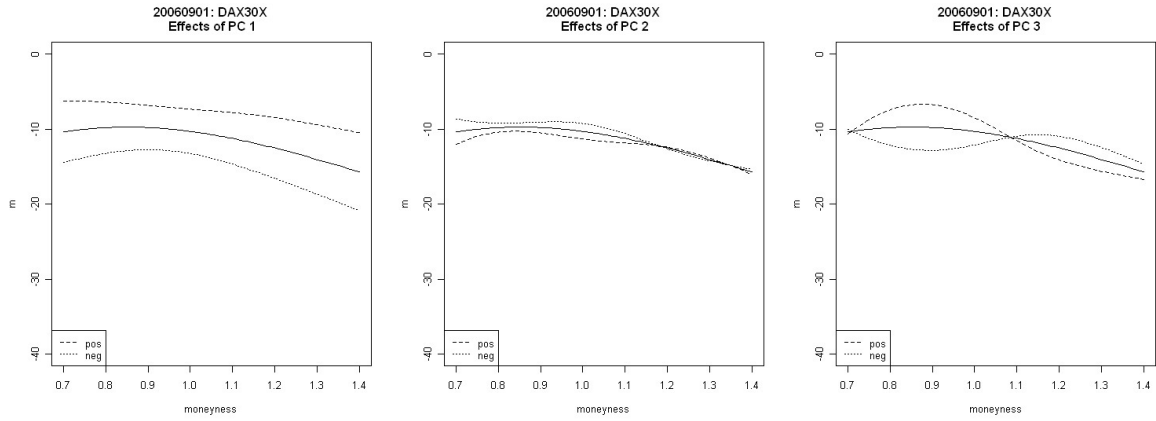


Figure 28: Positive and negative effects of principal components onto the mean variance process for short option maturities.

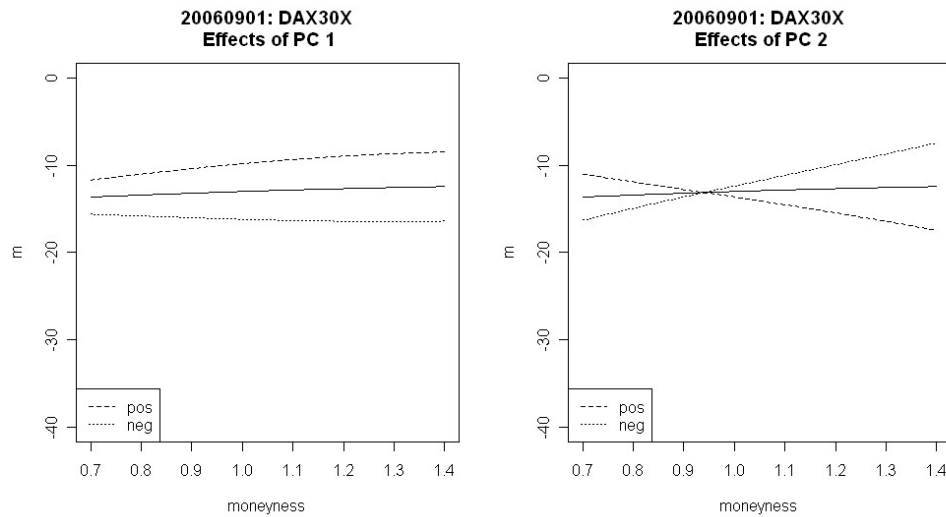


Figure 29: Positive and negative effects of principal components onto the mean variance process for long option maturities.

5.3.2 Stock IV example: VOLKSWAGEN

The next example is for stock options. We choose the Volkswagen IV data at 1 September 2006. Similar to the index data we decided to separate the data into implied volatilities for short time to option maturity $\tau \in \{2W, 1M, 2M, 3M\}$ and for long time to maturity $\tau \in \{6M, 9M, 1Y, 2Y\}$ according to structural differences between the IV curves.

Log return data is more volatile when time to maturity is short. Data adaptively chosen bandwidths to estimate smooth paths S_i vary from 0.05 to 1. In following smoothing steps, due to a large deviation of the residuals Z_{ij} , all automatically chosen optimal bandwidths are very small. That is $b_V = 0.1$ and $b_V = 0.05$ in case of short or long maturities, respectively, and $h_V = 0.05$ for both maturity groups. Plots of the estimated functions are shown in figures 30 and 31. The mean variance function is clearly decreasing for $\kappa > 0.8$. For long maturities m_V is much stronger decreasing, which is a result of large negative residuals for $\kappa = 1.4$ and $\tau \geq 1$ year.

Covariance and correlation matrices are shown in figure 32 and 33. We detect only a slight tendency for stationarity when option maturity is short and options are in and at-the-money. Correlation is positive over the whole range but much smaller in parts for long maturities. There seems to be a strong positive correlation between in-the-the-money and out-of-the-money data, especially for long maturities. But in case of long maturities at-the-money data has near zero correlation with data for nearly each κ outside a neighborhood of 1.

As a consequence of large data variability we need four PCs (short maturities) and three PCs (large maturities) to explain the variation. The first PC according to short maturities explains about 58% of the total variation, the second PC explains 26% and the third one still 10%. Main shocks are thus up-and-down shifts, that are more pronounced at the borders, but also heavily shocks over the whole range at different parts of the curves, resulting from various contrasts formed by second and third PCs, compare figures 34 and 35. In case of long option maturities the first and most important PC (89%) is negative over the whole range except at about $\kappa = 1.1$ when it reaches zero. Thus it explains up-and-down shifts at the left and right side of 1.1. The second PC forms a large contrast around $\kappa = 1.1$ and corresponds to heavy variation around this point, see figure 36.

Approximated variances of white noise processes are $\hat{\sigma}_W^2 = 5.6$ in case of short option maturities and $\sigma_W = 1.031$ in case of long ones. Similar to index data $\hat{\sigma}_W^2$ is larger for short maturities than for long maturities.

Summarizing, for short and long option maturities the variance process V for Volkswagen IV log returns is after a short increase until $\kappa = 0.8$ decreasing for increasing moneyness. Stationarity cannot be assumed. Shocks that mainly explain the variation are up-and-down shocks, which are more pronounced for long option maturities at left and right sides of $\kappa \approx 1.1$. The decreasing variance process indicates decreasing variability among the time variation of different IV smiles in the surface at regions where moneyness is large.

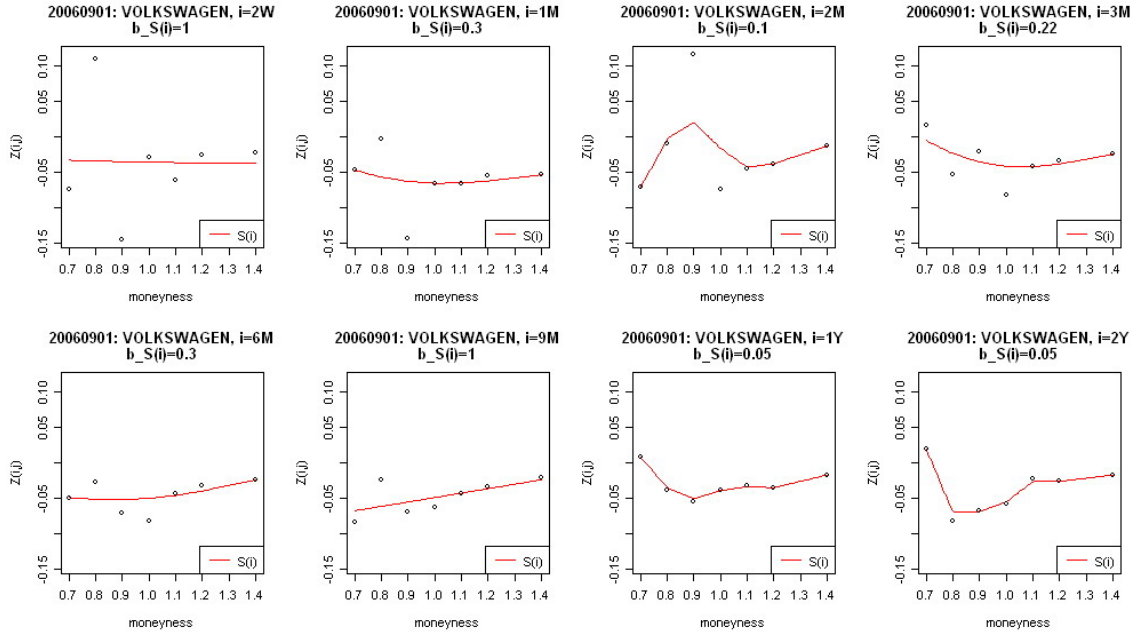


Figure 30: Smoothed log returns of smiles for Volkswagen IV data at 1 September 2006 for option maturity from two weeks (2W) to two years (2Y). Data are discretely plotted (dots) and smoothed trajectories are continuously estimated by local linear kernel estimation (solid line). Bandwidths for each trajectory are chosen by cross validation.

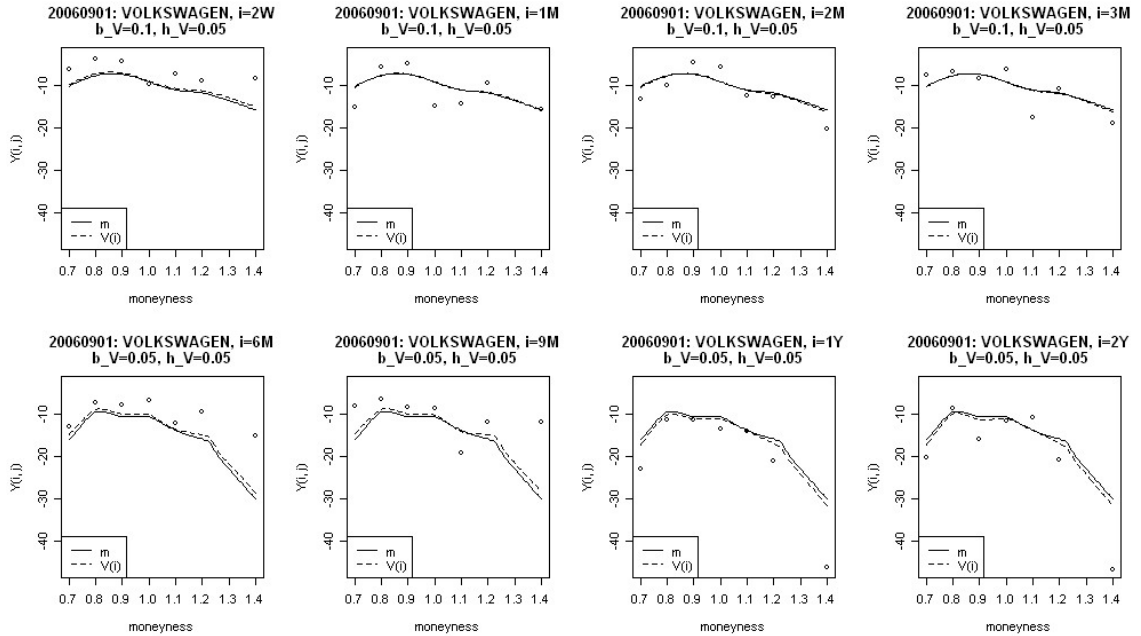


Figure 31: Volkswagen residuals at 1 September 2006 for different small option maturities (top panels) and different long option maturities (bottom panels). Estimated mean variance function m_V (solid line) and variance trajectories V_i (dashed line) are drawn into the plots. Estimates for small maturities and long maturities were made independently. All bandwidths are chosen by cross validation.

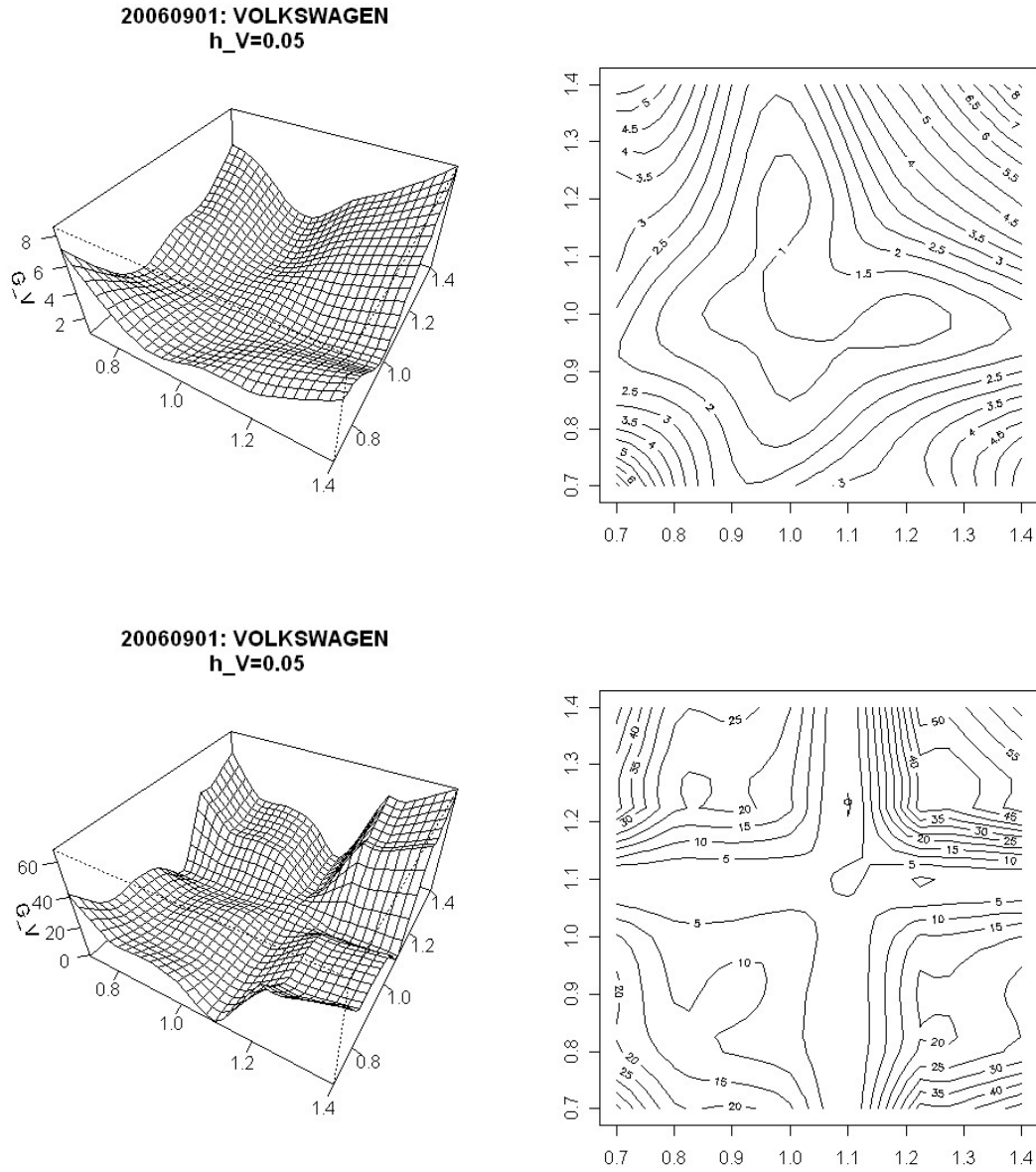


Figure 32: Estimated covariance of functional variance process for short maturities (top panels) and long maturities (bottom panels). The bandwidth h_V to estimate the surface is chosen by cross validation.

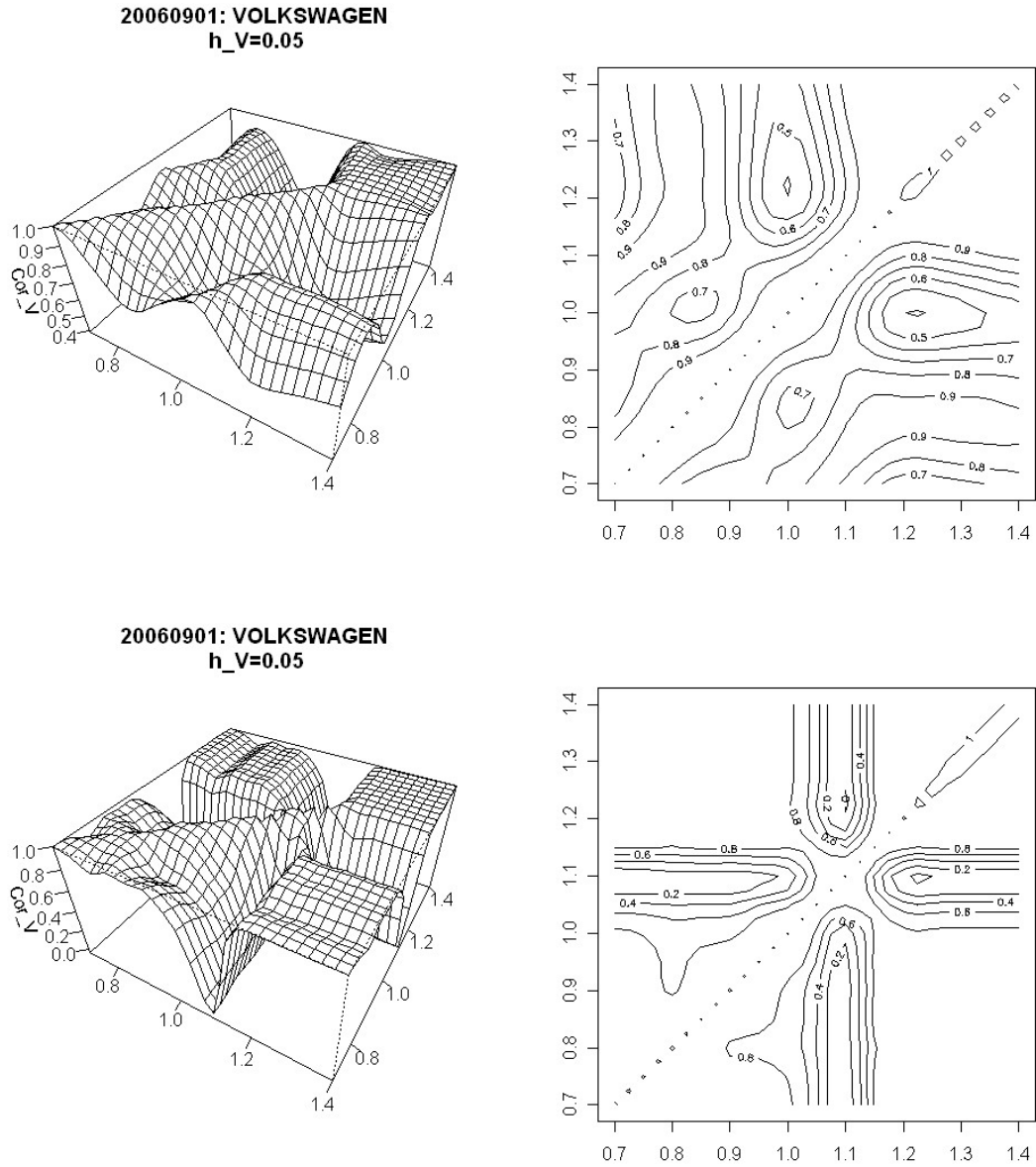


Figure 33: Estimated correlation of functional variance process for short maturities (top panels) and long maturities (bottom panels).

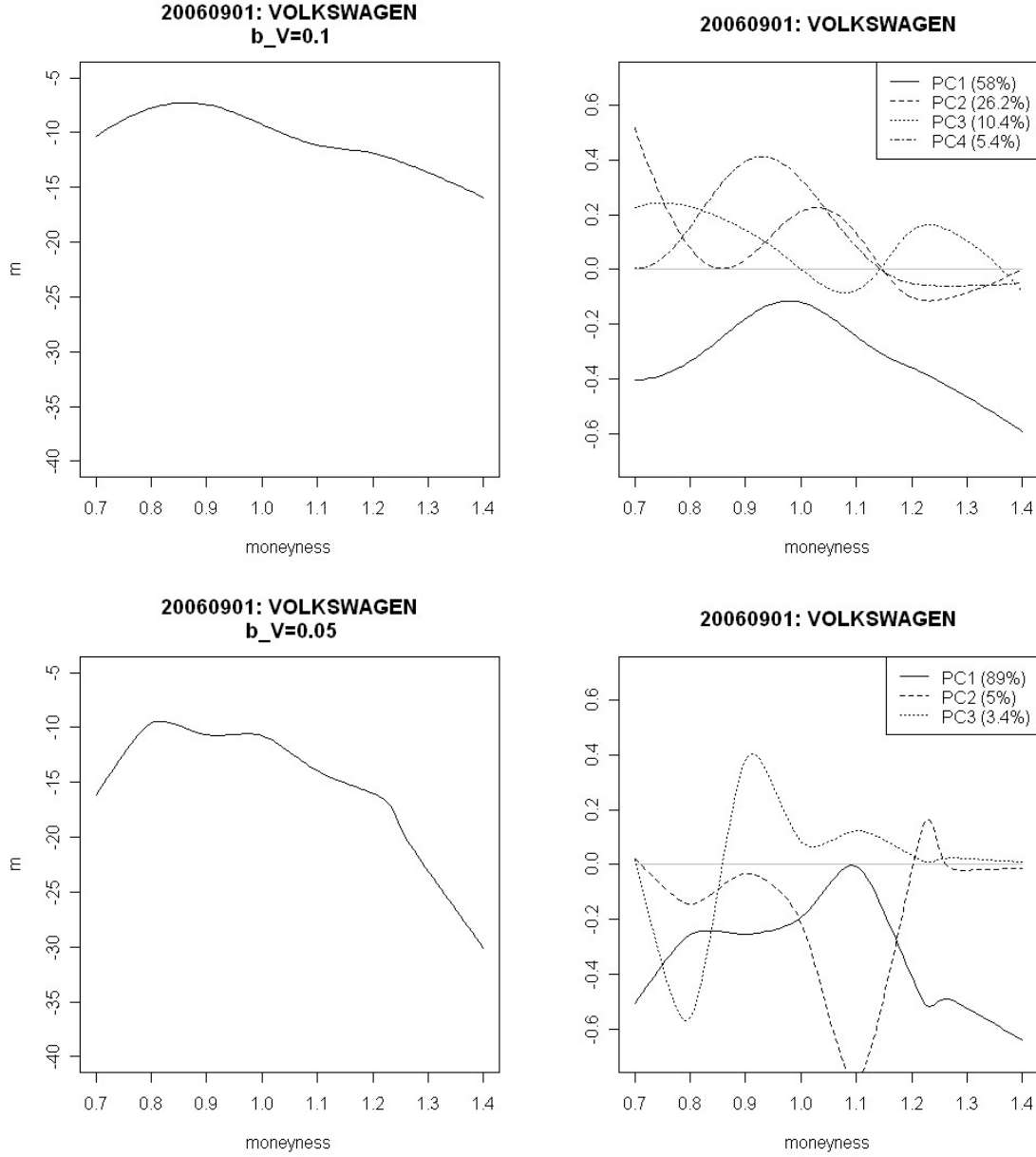


Figure 34: Volkswagen residuals at 1 September 2006 for different small option maturities (top panels) and for large option maturities (bottom panels). Estimated mean variance function m_V (dashed), variance trajectories V_i (dash-dotted) and estimated variance process V (solid) are drawn into the plots. All bandwidths are chosen by cross validation.

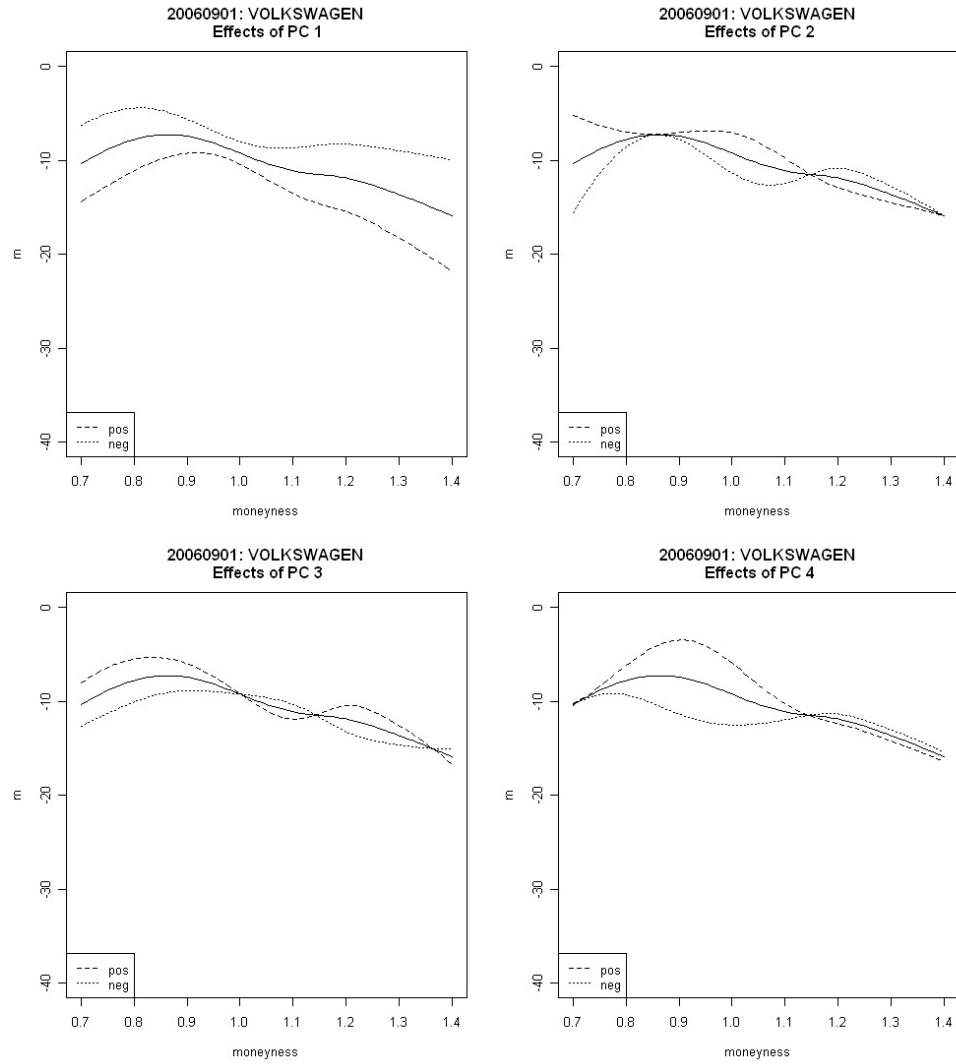


Figure 35: Positive and negative effects of principal components onto the mean variance process for short option maturities.

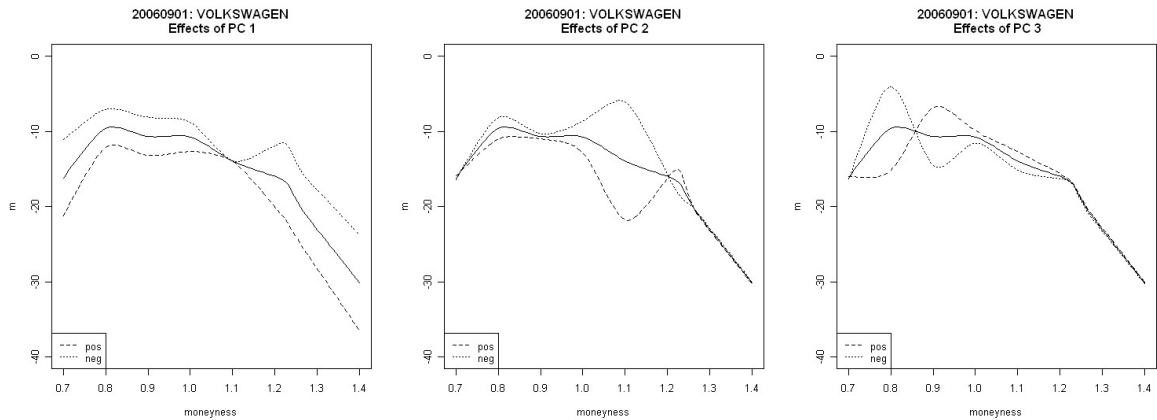


Figure 36: Positive and negative effects of principal components onto the mean variance process for long option maturities.

5.3.3 Swap IV example: JPY

At last the analysis is done for IV log returns of JPY swaptions at 1 September 2006. Note that implied volatilities now depend on option maturity τ^O and on swap maturity τ^S . The data is splitted into option maturities in $\{1M, 3M, 6M, 1Y\}$ and in $\{2Y, 3Y, 4Y, 5Y\}$.

The mean slopes of IV log return curves seem to be larger for long option maturities than for short ones due to some small data values when swap maturity is short, see figure 37. In general data seems to be less volatile than index and stock data. Bandwidths b_{S_i} were automatically chosen and range from 0.5 to 2.5. In figure 38 estimated mean variance processes and trajectories V_i are plotted. The trajectories clearly lie below or above the mean process and deviation from the mean is larger than it has been for index and swap data. The variance process decreases for decreasing $\tau^S > 3$ years. For long option maturities there is an exception at about $\tau^S = 7$ years where we find a little bump. Optimal bandwidth were chosen to be $b_V = 2.2$, $h_V = 0.5$ and $b_V = 0.5$, $h_V = 2.1$ for short and long option maturities, respectively.

The covariance and correlation plots are given in figures 39 and 40. The estimated covariance functions are very smooth, meaning that corresponding correlation functions are close to one almost everywhere. In case of short option maturities correlation differs from 1 maximal by 0.03, in case of long maturities the maximal difference is even less. Consequently, we assume strong dependencies among trajectories of the functional variance process for both option maturity groups.

The variance of V is driven by three PCs for short option maturities and by only one PCs for long option maturities. Even in the first case PC 1 explains already 87% of the total variation. The first PC corresponds to up-and-down shifts of the whole curves in both cases. For short option maturities second and third PCs explain little further variations of single parts of the curves. Corresponding plots are drawn in figures 41, 42 and 43.

The approximate white noise variance is larger for short option maturities than for long ones. We estimated $\hat{\sigma}_W^2 = 4.928$ and $\hat{\sigma}_W^2 = 1.082$, respectively.

For the original data we get the following. As the variance process is decreasing for swap maturities more than two years, the variation of IV log returns at 1 September 2006 among option maturity decreases for increasing swap maturity. Time variation of the IV Smiles thus varies strongest from one option maturity to another when swap maturity is about two or three years. For long option maturities there is a little increase of variation among curves within the surface at swap maturity $\tau^S = 7$ years. A huge amount of the variation of the data is explained by regular up-and-down shifts of the whole curves.

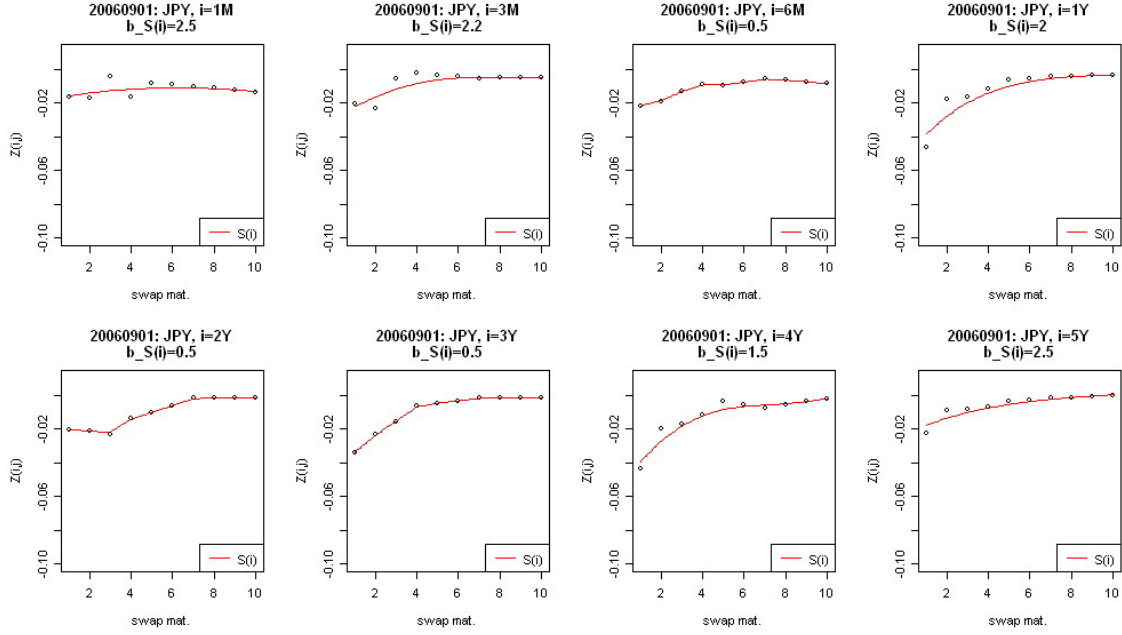


Figure 37: Smoothed log returns of IV curves for JPY IV data at 1 September 2006 for option maturity from one month (1M) to two years (2Y). Data are discretely plotted (dots) and smoothed trajectories are continuously estimated by local linear kernel estimation (solid line). Bandwidths for each trajectory are chosen by cross validation.

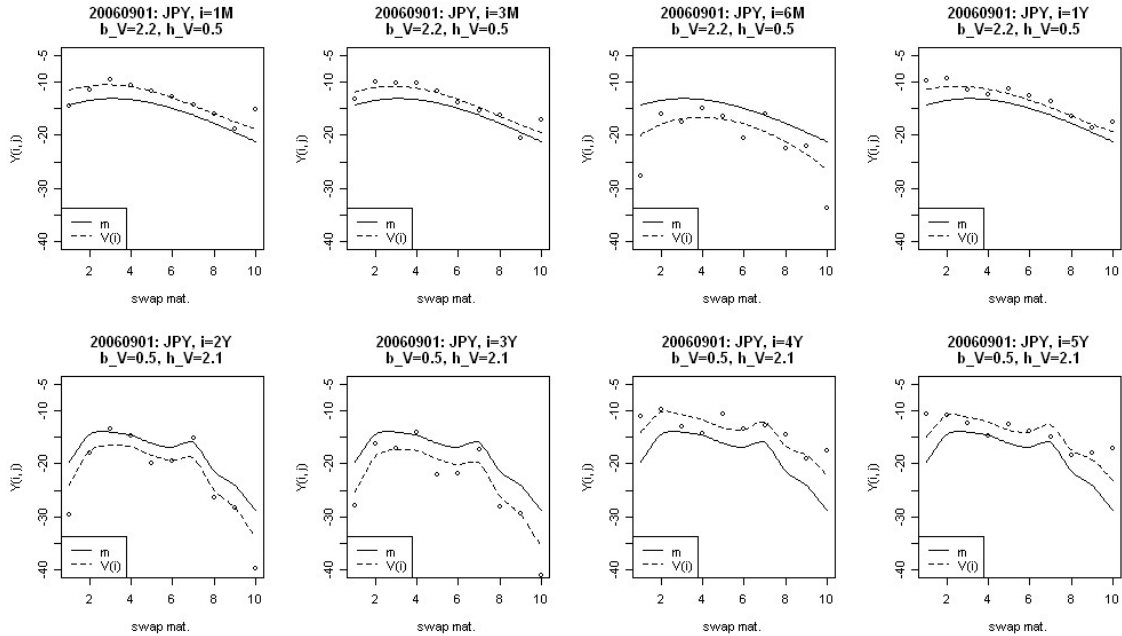


Figure 38: JPY residuals at 1 September 2006 for different small option maturities (top panels) and different long option maturities (bottom panels). Estimated mean variance function m_V (solid line) and variance trajectories V_i (dashed line) are drawn into the plots. Estimates for small maturities and long maturities were made independently. All bandwidths are chosen by cross validation.

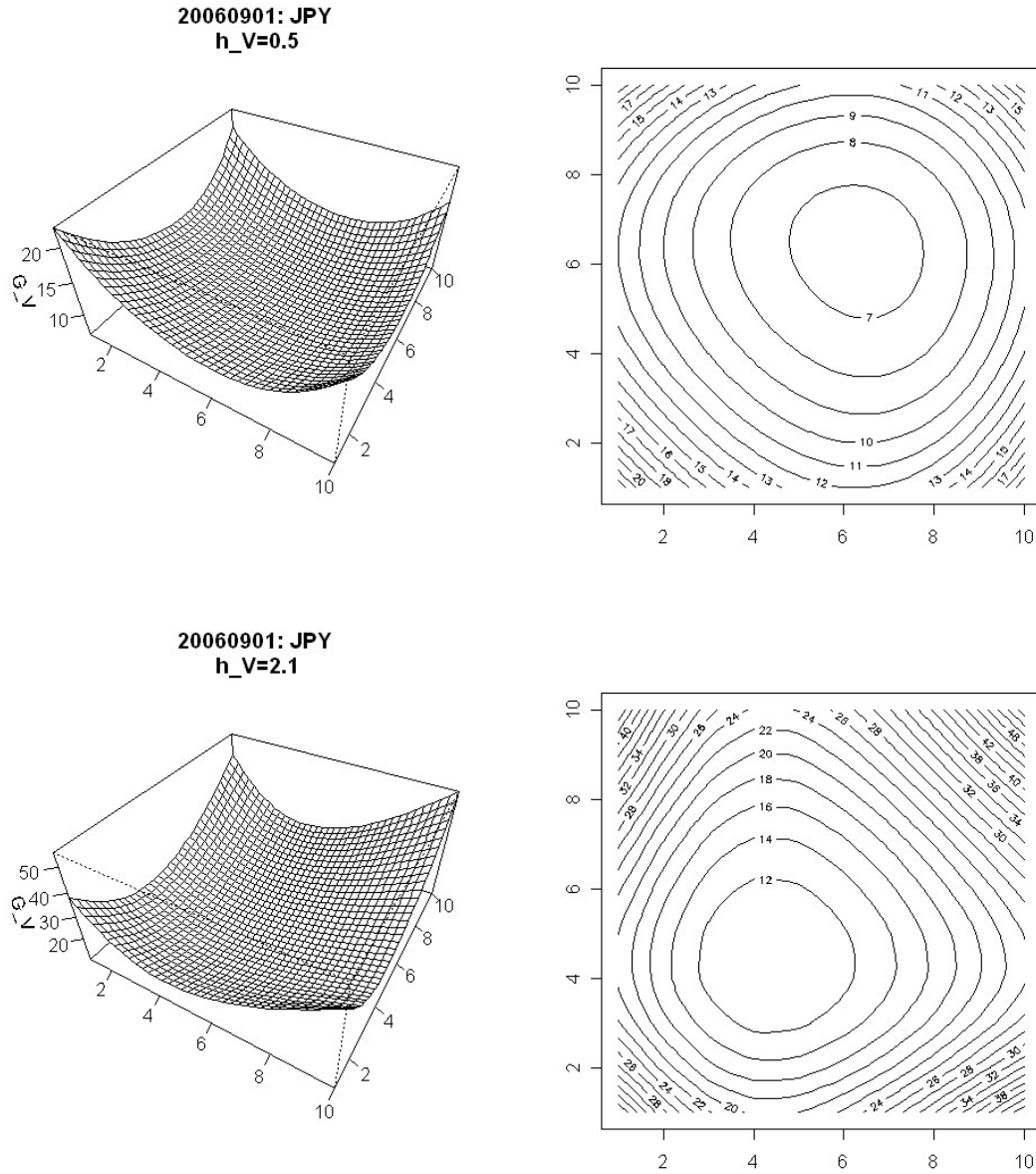


Figure 39: Estimated covariance of functional variance process for short maturities (top panels) and long maturities (bottom panels). The bandwidth h_V to estimate the surface is chosen by cross validation.

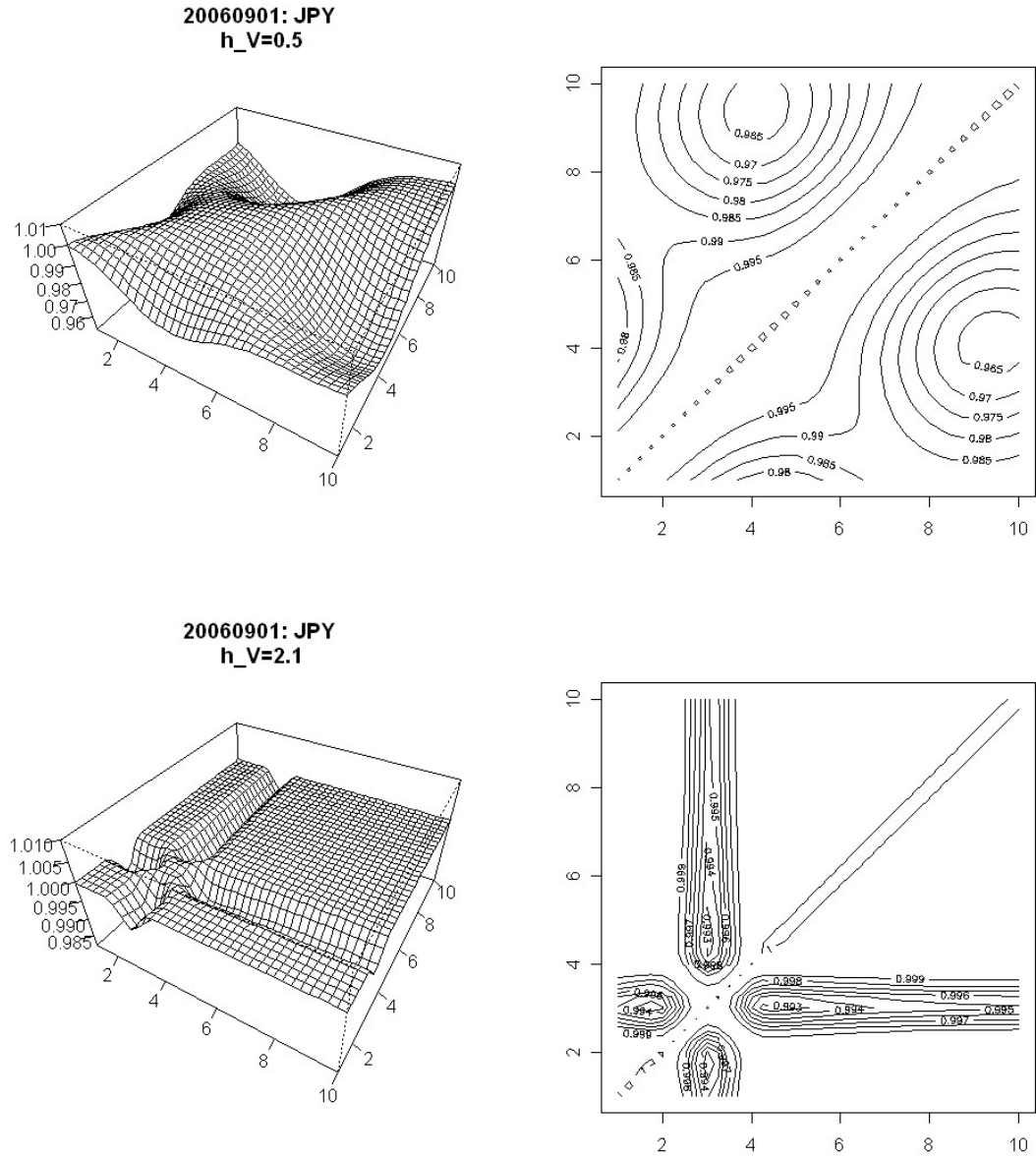


Figure 40: Estimated correlation of functional variance process for short maturities (top panels) and long maturities (bottom panels).

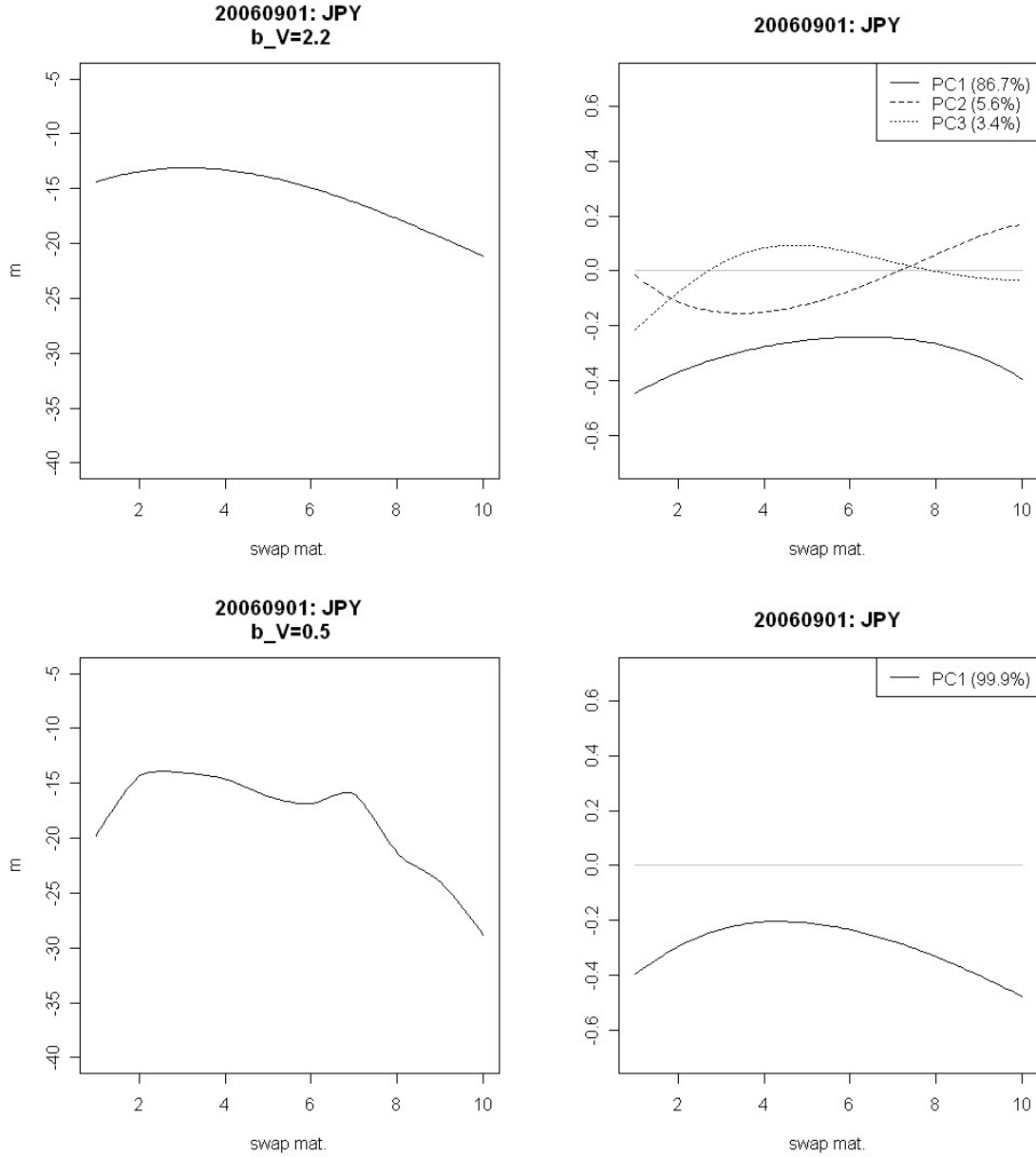


Figure 41: Functional variance process (left) and eigenfunctions (right) for small option maturities (top panels) and for large option maturities (bottom panels). Functional variance process V is approximated by local linear kernel estimation from single trajectories V_i . Bandwidth h is chosen by cross validation. First eigenfunction (solid), second (long-dash), third (short-dash) and fourth eigenfunction (dash-dot) determine the dominant modes of variation.

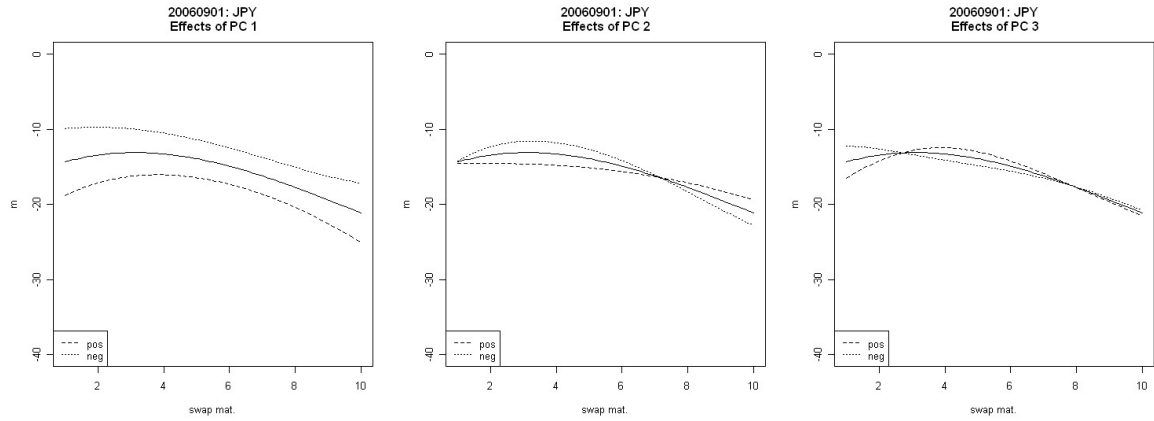


Figure 42: Positive and negative effects of principal components onto the mean variance process for short option maturities.

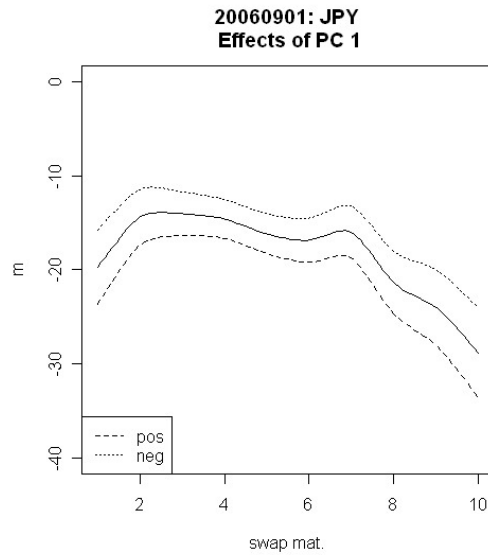


Figure 43: Positive and negative effects of principal components onto the mean variance process for long option maturities.

5.4 Time series of functional variance processes

Until now we analysed variation among IV log return curves at specific points in time in terms of the functional variance process V . An interesting question is how this process changes in time. We aim at finding a general feature of the variability within implied volatility surfaces that remains over time. Therefore we estimate functional variance processes for all daily data sets of log returns of stock, index and swap options, respectively. To get comparable results we decided to choose fixed bandwidths b_V and h_V to estimate the mean function m_V and the covariance structure \mathcal{G}_V of the variance process. Former analysis arises that for index and stock data, data-adaptive choice of bandwidths frequently leads to small bandwidths. Therefore we choose $b_V = 0.1$ and $h_V = 0.05$. For the swap data cross validation leads to small to middle size bandwidths b_V and to small bandwidths h_V . Therefore we decided to fix $b_V = 2$ and $h_V = 0.5$ to approximate the mean variance function m_V and the covariance \mathcal{G}_V , respectively.

As examples for index, stock and swap IVs we took again the data sets of section 5.3, i.e. DAX30X, Volkswagen and JPY. To give an overview, we only plotted the variance processes of each fifth trading day from 23 August to 25 October 2006. Data at selected days are the log differences of the IVS at Tuesdays and Wednesdays of each week, respectively.

In figure 44 plots are shown for DAX30X IVs when option maturity is in $\{2W, 1M, 2M, 3M\}$. In figure 45 data is plotted for long maturities in $\{6M, 9M, 1Y, 2Y\}$. From left to right and from top to bottom panels we draw the variance processes and corresponding PCs from end of August to end of October. The mean variance processes remain similar over the weeks and also the eigendecomposition is quite similar in most cases. m_V is concave, lightly increasing for $\kappa < 0.9$ and stronger decreasing for $\kappa > 1.1$. The shapes of m_V corresponding to short and long option maturities resemble, but in general variance processes are flatter in the long maturity group. Note that there are some exceptions, when the variance process becomes almost constant or even lightly increasing over the whole range, see for example the plot at 6 September (20060906) in figure 45. Comparing the first PCs, we detect that PC 1 often seems to align m_V , meaning that both functions resemble in their shape. PC 1 is of the same sign throughout at most days. Therefore we conclude that the up-and-down shifts among IV log return curves are a mode of variation that remains in time. These shifts are most pronounced outside the range $0.8 < \kappa < 1.1$. Of course we also detected outliers, here for example in figure 44 at 20 September (20060920) and in figure 45 at 6 September (20060906), when the first PC does not correspond thus clear to overall up-and-down shifts or even corresponds to completely different shocks. For the second PC our results vary from one day to another and similarities are only hard to find by visual inspection.

Functional variance processes are flatter for IV data according to long option maturities than to data according short option maturities. In general variance processes reach maximal values for $\kappa \in [0.8, 1.1]$ and decrease when running off this range. Therefore we resume that for options on DAX30X indexes daily changes of implied volatilities vary the most between options with different maturities when options are around $\kappa = 1$ or slightly below, i.e. at-the-money options and lightly in-the-money options. Variation decreases for deep in-the-money options and decreases even stronger for out-of-the-money options.

Results for options on Volkswagen stocks are similar to the previous, so we will not plot the results here. We noticed that the shape of m_V varies little more, meaning that that variance processes are sometimes flatter and sometimes more curved. Nevertheless m_V is convex, with maximal values for $\kappa \in [0.8, 1.1]$, and the first PC essentially corresponds to overall up-and-down shifts.

At last we check the time dependent development of functional variance processes for swap implied volatilities. Resulting functional estimates are shown in figures 46 and 47 for short maturities $\tau^O \in \{1M, 3M, 6M, 1Y\}$ and long maturities $\tau^O \in \{2Y, 3Y, 4Y, 5Y\}$, respectively. Here, m_V is given over swap maturities τ^S . Differences between the two maturity groups are not thus clear. In general, m_V is less curved than it is for the index and swap data and that shape of m_V does not vary as much. The process is decreasing, on some days there is a first little increase for $\tau^S < 3$ years. Thus variation among IV log return curves corresponding to different option maturities is strongest in parts where $3Y < \tau^S < 5Y$, and this result holds during the months.

The first PC nearly always corresponds to regular shifts of the functional variance. Even the second PC is similar at various days, it explains up-and-down-shifts in the middle part, i.e. $\tau^S \in [2, 8]$ years, see for example 23 August (20060823) in figure 47. These features are more clearly pronounced for data according to long option maturities.

In general, we found a common functional variance process that explains the variation among IV daily changes according to different option maturities. The mean process is maximal around 4 years swap maturity and lightly decreases afterwards. Thus variation among daily IV log return curves is largest in parts where $\tau^S \approx 4$ years. Trajectories of the functional variance process mainly differ by overall up-and-down shifts from the mean process.

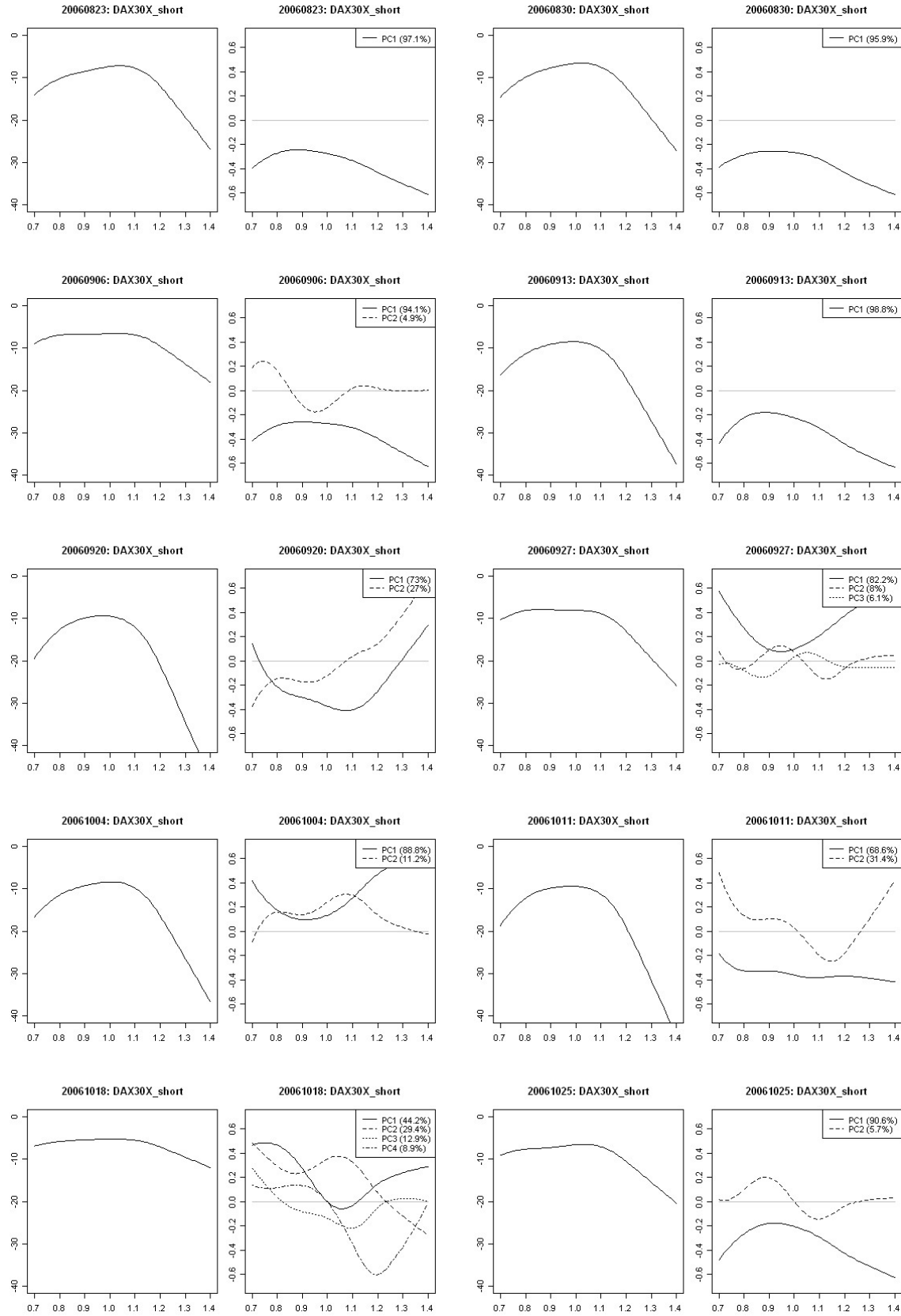


Figure 44: FVPs and according PCs for DAX30X options of the short maturity group. From left to right and top to bottom panels functions are given from 23 August (20060823) to 25 October 2006 (20061025) at each fifth trading day.

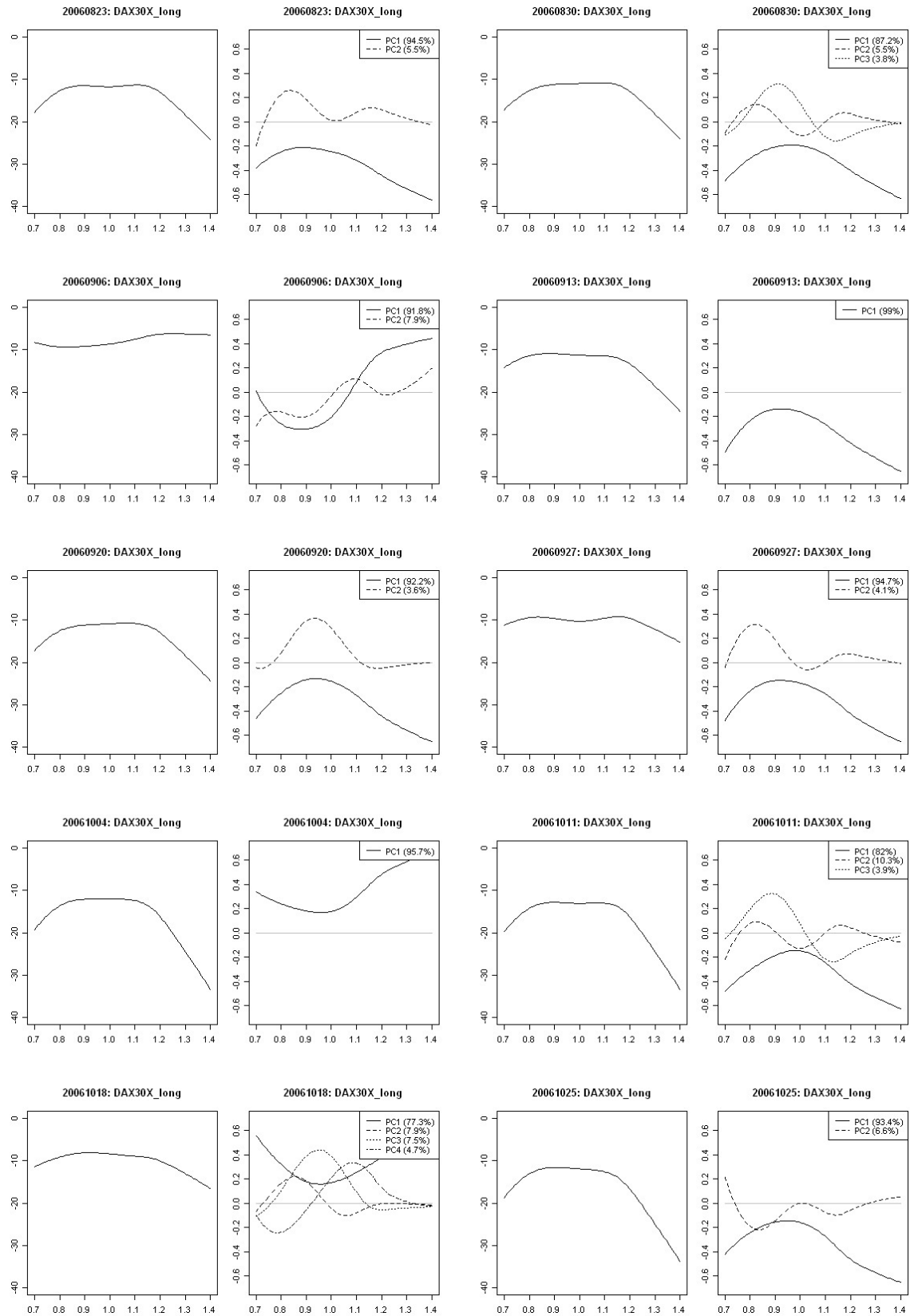


Figure 45: FVPs and according PCs for DAX30X options of the long maturity group. From left to right and top to bottom panels functions are given from 23 August (20060823) to 25 October 2006 (20061025) at each fifth trading day.

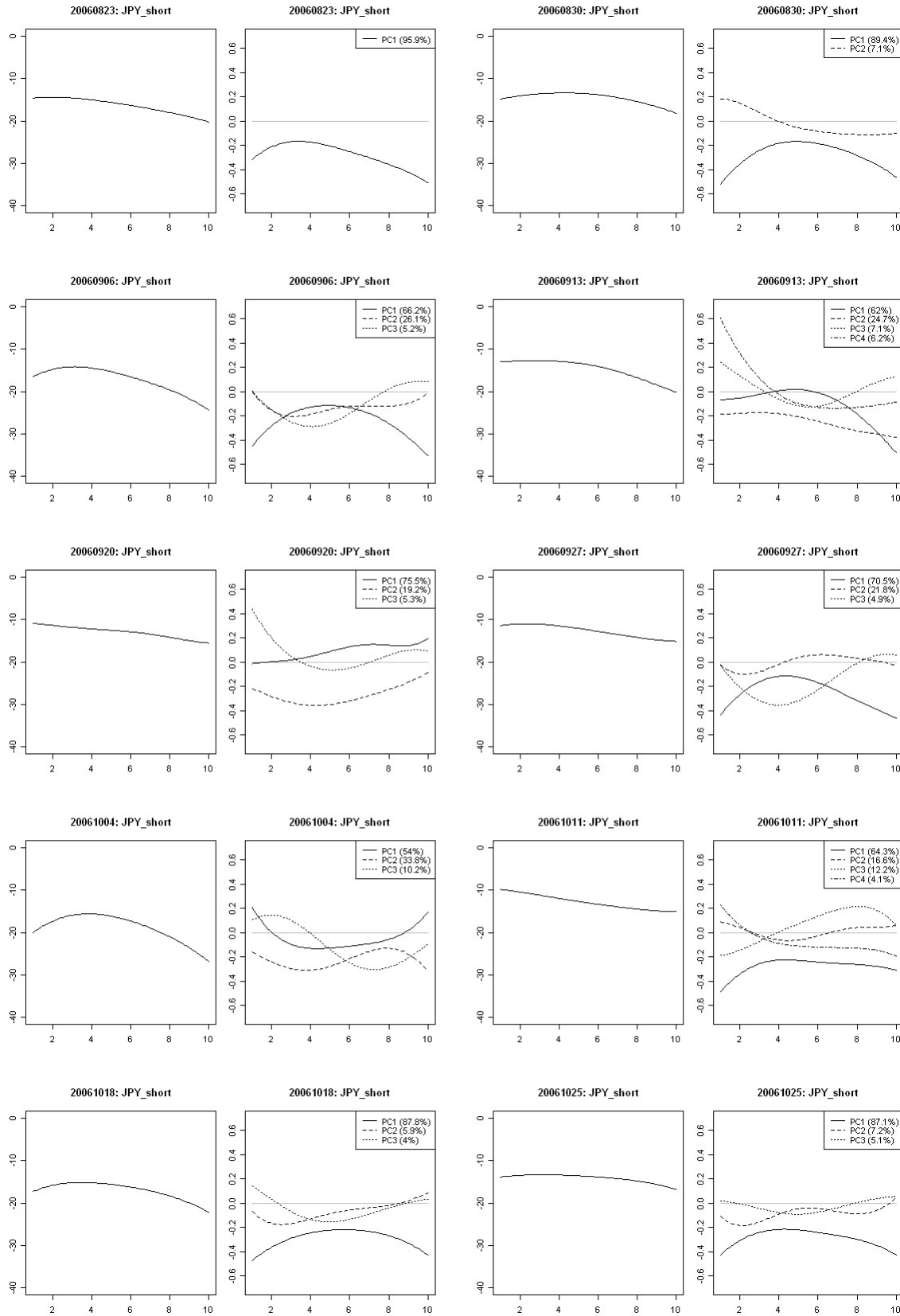


Figure 46: FVPs and according PCs for JPY options of the short maturity group. From left to right and top to bottom panels functions are given from 23 August (20060823) to 25 October 2006 (20061025) at each fifth trading day.

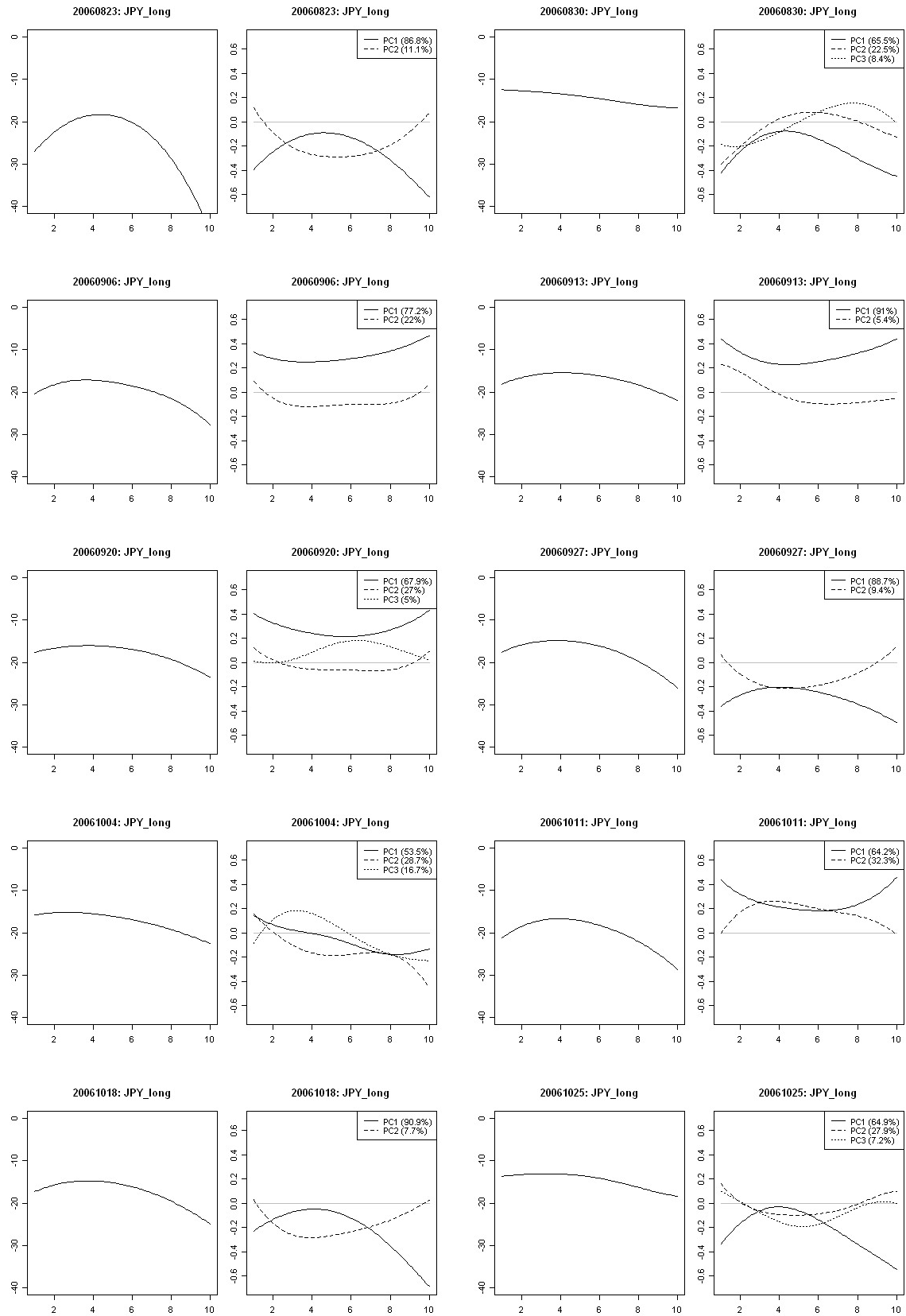


Figure 47: FVPs and according PCs for JPY options of the short maturity group. From left to right and top to bottom panels functions are given from 23 August (20060823) to 25 October 2006 (20061025) at each fifth trading day.

6 Summary and Open Questions

Implied volatilities of options on financial assets appear as two dimensional surfaces on the remaining time to option maturity and on a second variable. For options on indexes or stocks this second variable is the moneyness (or the strike price), for swap options the second variable is the swap maturity. In the analysis we describe implied volatility surfaces as a set of IV curves on the second variable, i.e. into curves on moneyness (IV smiles) or IV curves on swap maturity, indexed by discrete values of option maturity.

In literature there are different approaches to analyse time dependent variation of IV smiles. We followed some of these methods, not only to analyse daily variation of IV smiles but also IV curves on swap maturity. Our results align with former results. The main modes of variation that drive IV curves in time are shift, slope and curvature changes. These modes are similar for different curves in the IVS. Only strength of variation varies. Therefore it lies on hand to analyse IV curves for several option maturities simultaneously.

Moreover, in the data we analysed we detected structural differences between IV log returns according to options with short time to maturity and to options with long time to maturity. Although these differences do not appear regularly, we decided to separate data into two groups according to option maturities in continuative analysis of variation within the data.

We further applied a new method in our data analysis to determine how daily variation of the IV curves varies among option maturities that belong to one group, i.e. how daily variation varies within the IVS. We described this kind of variation at day t in terms of a functional variance process that is determined from the log transformed and squared, functional residuals of the original IV log return curves. We detected that this variance process remains in time and that trajectories of the process mainly vary by all over up-and-down shifts. The result holds for different maturity groups and for IV data sets coming from options on different financial assets, i.e. indexes, stocks and swaps. Processes are maximal for moneyness at about $\kappa = 1$ and lightly below, meaning that daily changes of the IV curves vary the most among option maturities when options are at-the-money or lightly in-the-money. For swaptions functional variance processes are maximal around swap maturity in $[3Y, 5Y]$. This is the range, where daily variation of IVs differs the most among option maturities.

This work arises some questions for future research, concerning the analysis of functional variance processes and the data used in the analysis.

- We analysed variation of implied volatility curves according to their daily changes. In practice it is also of interest to analyse variation for larger lags, i.e. determine IV log differences of days t and $t - l$ for $l > 1$ and find modes of variation in time and within the surface.
- As we already mentioned, in practice implied volatility surfaces appear as a set of strings in moneyness direction with respect to some few option maturities. As asymptotics for the estimation of FVPs hold even when prediction points vary from one

trajectory to another, it is reasonable to apply the analysis directly onto observed implied volatility data without previously smoothing data onto a fixed grid of option maturity and moneyness.

- Until now, the time dependent development of functional variance processes is only analysed by visual inspection. It remains to perform a functional data analysis of the processes to determine trends and variation of these processes in time. Therefore one may regard the estimated mean processes μ_V at different days as functional data sampled in time. The analysis of these data may lead to deeper understanding of IV variation.
- We found that daily changes of stock and index implied volatilities are determined by similar modes of variation concerning both, time dependent variation and variation within the surface. Even for IVs of swap options there are similarities to stock and index data. Now, it would be interesting to determine the main differences with respect to variation among IVs corresponding to different financial assets. How do the corresponding functional variance processes differ?

References

- [1] Barndorff-Nielsen, O.E. (1997) Normal Inverse Gaussian Distributions and Stochastic Volatility Modeling. *Scandinavian Journal of Statistics*, **24**, 1–13. *Scandinavian Journal of Statistics*, **24**, 1–13.
- [2] Bates, D.S. (1997) Jumps and Stochastic Volatility: Exchange Rate Processes implicit in Deutsche Mark Options. *Review of financial Studies*, **9**, 69–107.
- [3] Bauer, H. (1992) Maß- und Integrationstheorie. 2. Auflage. *De Gruyter & Co., Berlin*.
- [4] Black, F., Scholes, M. (1973) The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, **81**, 637–654.
- [5] Capra, W.B., Müller, H.-G. (1997) An Accelerated-Time Model for Response Curves. *Journal of the American Statistical Association*, **92**, No. 437, 72–83.
- [6] Castro, P.E., Lawton, W.H., Sylvestre, E.A. (1986) Principal Modes of Variation for Processes with Continuous Sample Curves. *Technometrics*, **28**, No. 4, 329–337.
- [7] Cont, R., da Fonseca, J. (2002) Dynamics of Implied Volatility Surfaces. *Quantitative Finance*, **2**, 45–60.
- [8] Cont, R., da Fonseca, J., Durrleman, V. (2002) Stochastic Models of Implied Volatility Surfaces. *Economic Notes by Banca Monte dei Paschi di Siena SpA*, **31**, No. 2-2002, 361–377.
- [9] Cox, J., Ross, S. (1976) The valuation of Options for Alternative Stochastic Processes. *Journal of Financial Economics*, **3**, 145–166.
- [10] Cox, J., Ross, S., Rubinstein, M. (1979) Option Pricing: A Simplified Approach. *Journal of Financial Economics*, **7**, 87–106.
- [11] Derman, E., Kani, I., Chriss, N. (1996) Implied Trinomial Trees of the Volatility Smile. *The Journal of Derivatives*, **3**, 4, 7–22.
- [12] Dumas, B., Fleming, J., Whaley, R.E. (1998) Implied Volatility Functions: Empirical Tests. *Journal of Finance*, **53**, 6, 2059–2106.
- [13] Dupire, B. (1994) Pricing with a Smile. *Risk*, **7**, 18–20.
- [14] Fan, J. (1992) Design-adaptive Nonparametric Regression. *Journal of the American Statistical Association*, **87**, 998–1004.
- [15] Fan, J., Gijbels, I. (1997) Local Polynomial Modeling and Its Applications. *Chapman & Hall, London*.
- [16] Fengler, M., Härdle, W., Schmidt, P. (2001) The Analysis of Implied Volatilities. *SFB 373 Papers 2001-73*, Humboldt-University, Berlin.

-
- [17] Fengler, M., Härdle, W., Villa, C. (2003) The Dynamics of Implied Volatility: A Common Principal Components Approach. *Review of Derivative Research*, **6**, 179-202.
- [18] Fengler, M., Härdle, W., Mammen, E. (2005) A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics. *SFB 649 Discussion Paper 2005-020*, Humboldt-University, Berlin.
- [19] Härdle, W. (1993) Applied Nonparametric Regression. *Cambridge University Press*.
- [20] Hafner, R., Wallmeier, M. (2001) The Dynamics of DAX Implied Volatilities. *International Quarterly Journal of Finance*, **1**, 1–27.
- [21] Hall, P., Poskitt, D.S., Presnell, B. (2001) A Functional Data-Analytic Approach to Signal Discrimination. *Technometrics*, **43**, No. 1, 1–9.
- [22] Hall, P., Hosseini-Nasab, M. (2006) On Properties of Functional Principal Components Analysis. *Journal of the Royal Statistical Society. B*, **68**, Part 1, 109–126.
- [23] Hart, J.D. (1997) Nonparametric Smoothing and Lack-of-Fit Tests. *Springer-Verlag, New York*.
- [24] Heston, S. (1993) A closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, **6**, 327–343.
- [25] Hull, J.C. (2003) Options, Futures, and Other Derivatives. *Pearson Education, New Jersey*, Prentice Hall.
- [26] Hull, J.C., White, A. (1987) The Pricing of Options on assets with Stochastic Volatilities. *Journal of Finance*, **42**, 281–300.
- [27] Irle, A. (1998) Finanzmathematik, Die Bewertung von Derivaten. *Teubner, Stuttgart*.
- [28] James, G.M., Hastie, T.J., Sugar, C.A. (2000) Principal Component Models for Sparse Functional Data. *Biometrika*, **87**, No. 3, 587–602.
- [29] Jolliffe, I.T. (2002) Principal Component Analysis. *Springer-Verlag, New York*.
- [30] Knapp, C. (2000) Estimation of Long Term Implied Volatilities. Version 1.1, COMMERZBANK AG, Frankfurt.
- [31] Kneip, A. (1994) Nonparametric Estimation of Common Regressors for Similar Curve Data. *Annals of Statistics*, **22**, No. 3, 1386–1427.

- [32] Kneip, A., Utikal, J. (2001) Inference for Density Families Using Functional Principal Component Analysis. *Journal of the American Statistical Association*, **96**, No. 454, 519–542.
- [33] Lee, R. (2004) Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation. *Recent Advances in Applied Probability*, Springer.
- [34] Müller, H.G., Stadtmüller, U., Yao, F. (2006) Functional Variance Processes. *Journal of the American Statistical Association*, **101**, 1007–1018.
- [35] Merton, R.C. (1973) Theory of Rational Option Pricing. *Bell Journal of Econometrics and Management Science*, **4**, 141–183.
- [36] Mixon, S. (2002) Factors Explaining the Movements on the Implied Volatility Surface. *Journal of Futures Markets*, **22**, 915–937.
- [37] Nadaraya, E.A. (1964) On Estimating Regression. *Theory of Probability and its Applications*, **10**, 186–190.
- [38] Parzen, E. (1962) On the Estimation of a Probability Density Function and the Mode. *The Annals of Mathematical Statistics*, **33**, 1065–1076.
- [39] Ramsay, J.O., Silverman, B.W. (1997) Functional Data Analysis. *Springer-Verlag, New York*.
- [40] Ramsay, J.O., Silverman, B.W. (2002) Applied Functional Data Analysis, Methods and Case Studies. *Springer-Verlag, New York*.
- [41] Rice, J.A., Silverman, B.W. (1991) Estimating the Mean and Covariance Structure when the Data are Curves. *Journal of the Royal Statistical Society, Ser. B*, **53**, 233–243.
- [42] Rosenblatt, M. (1956) Remarks on some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, **27**, 832–837.
- [43] Rubinstein, M. (1994) Implied Binomial Trees. *Journal of Finance*, **49**, 3, 771–818.
- [44] Schuster, E., Yankowitz, S. (1979) Contributions to the Theory of Nonparametric Regression, with Applications to System Identification. *Ann. Statist.*, **7**, No. 1, 139–149.
- [45] Skiadopoulos, G., Hodges, S., Clewlow, L. (1999) The Dynamics of the S&P 500 Implied Volatility Surface. *Review of Derivatives Research*, **3**, 263–282.
- [46] Silverman, B.W. (1996) Smoothed Functional Principal Components Analysis by Choice of Norm. *Annals of Statistics*, **24**, 1–24.
- [47] Staniswalis, J.G., Lee, J.J. (1998) Nonparametric Regression Analysis of Longitudinal Data. *Journal of the American Statistical Association*, **93**, 1403–1418.

-
- [48] Stein, E.M., Stein, J.C. (1991) Stock Price Distributions with Stochastic Volatility: An Analytic Approach. *Review of Financial Studies*, **4**, 727–752.
- [49] Tompkins, R. G. (2001) Implied Volatility Surfaces: Uncovering Regularities for Options on Financial Futures. *The European Journal of Finance*, **7**, 3, 198–230.
- [50] Überfuhr, U. (2001) CB IV Interpolation Algorithm Amendment. COMMERZBANK AG, Frankfurt.
- [51] Wand, M.P., Jones, M.C. (1995) Kernel Smoothing. *Chapman & Hall, London*.
- [52] Watson, G.S. (1964) Smooth Regression Analysis. *Sankhyā*, Ser. A, **26**, 101–116.
- [53] Xu, X., Taylor, S.J. (1994) The Term Structure of Volatility Implied by Foreign Exchange Options. *Journal of Financial and Quantitative Analysis*, **29**, 1, 57–74.
- [54] Yao, F., Müller, H.G., Clifford, A.J., Duenker, S.R., Follet, J., Lin, Y., Buchholz, B.A., Vogel, J.S. (2003) Shrinkage Estimation for Functional Principal Component Scores with Application to the Population Kinetics of Plasma Folate. *Biometrics*, **59**, 676–685.
- [55] Yao, F., Müller, H.G., Wang, J.L. (2005) Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, **100**, No. 470, 577–590.
- [56] Yao, F., Lee, T.C.M. (2006) Penalized Spline Models for Functional Principal Component Analysis. *Journal of the Royal Statistical Society. B*, **68**, Part 1, 3–25.
- [57] Yao, F. (2006) Functional Principal Component Analysis for Longitudinal and Survival Data. *Statistica Sinica*, to appear.
- [58] Yao, F. (2007) Asymptotic Distributions of Nonparametric Regression Estimation for Longitudinal or Functional Data. *Journal of Multivariate Analysis*, **98**, 40–56.
- [59] Yao, Q., Tong, H. (2000) Nonparametric Estimation of Ratios of Noise to Signal in Stochastic Regression. *Statistica Sinica*, **10**, 751–770.
- [60] Yu, K., Jones, M.C. (2004) Likelihood-Based Local Linear Estimation of the Conditional Variance Function. *Journal of the American Statistical Association*, **99**, No. 465, 139–144.

Lebenslauf

Geburtstag 29.06.1979

Geburtsort Korbach

Schulbildung & Studium

1985–1989 Grundschule in Medebach

1989–1995 Gymnasium in Medebach

1995–1998 Geschwister-Scholl-Gymnasium in Winterberg
Leistungskurse: Mathematik und Deutsch
Abschluss: Abitur

09/1998–09/2003 Mathematikstudium an der Universität in Paderborn
Studienschwerpunkte: Mathematische Statistik,
Theoretische Informatik
Thema der Diplomarbeit: Neyman smooth Tests und
Anwendungen
Abschluss: Diplom

10/2003–heute Promotionsstudium an der Georg-August-Universität
Göttingen, Institut für Mathematische Stochastik
Teilnahme am interdisziplinären Promotionsstudiengang
“Angewandte Statistik und Empirische Methoden”
Betreuer der Dissertation: Herr Prof. Dr. A. Munk
Thema der Dissertation: Analysis of Implied Volatility
Surfaces

Stipendien

10/2003–09/2005 Lichtenberg Stipendium des Landes Niedersachsen

10/2006–12/2006 Stipendium des Graduiertenkollegs “Identifikation in
mathematischen Modellen: Synergie stochastischer und
numerischer Methoden” der Universität Göttingen

Praktika

02/2004–04/2004 Commerzbank AG in Frankfurt a.M.
Abteilung Risk Controlling

07/2005–09/2005 BHW Lebensversicherung AG in Hameln
Abteilung Versicherungsmathematik