Projektbereich A
Discussion Paper No. A–565

Estimating the Functional Components of Asset Price Volatilities

by

Frank Heid

November 1997

Acknowledgements: Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn is gratefully acknowledged.

SONDERFORSCHUNGSBEREICH 303
Rheinische Friedrich-Wilhelms-Universität Bonn
Lennéstraße 37
D–53113 Bonn
Germany
Telefon (0228) 73 92 42
Telefax (0228) 73 79 40
E–Mail: with2@econ2.uni-bonn.de

Abstract

The volatility of asset prices as a function of past prices is estimated by nonparametric regression. The estimates show that these functions are concave and obtain a minimum at a value close to zero. We use ideas of Principal Component Analysis to evaluate common functional components of different sets of volatility functions.

Keywords: volatility estimation, noparametric regression, model selection, principal components, dimension estimation, selfmodelling regression
JEL-classification: C14, C50, C12

1 Introduction

Assessing and explaining the volatility of financial assets was always at the centre of financial research. Three points may suffice to illustrate this claim:

1. Investors base their investment decisions on price forecasts. The accuracy of these estimates crucially depends on the volatility of prices.

2. In the capital asset pricing model (CAPM) the risk premium of an asset arises from the equilibrium of financial markets and is equal to the covariance between the return on an asset and the return on the market portfolio.

3. Arbitrage-free prices of derivatives are functions of the volatility of the underlying asset. The price of a put or a call option in the Black and Scholes model, for instance, depends *only* on the price and the volatility of the underlying asset as well as the exercise price, the time of maturity and the interest rate.

It has long been recognized in previous research that the volatility of asset prices does not stay constant: times of high volatility alternate with those of low volatility. Parametric models such as the ARCH model and the like which capture heteroscedasticity are now well established in applied financial research. More and more models have been added to the abounding literature to incorparate other relevant effects found in the data. A different stance has recently been taken by Härdle and Tsybakov (1995) and Heid (1996) who proposed to estimate the

volatility function in a more flexible way by nonparametric regression.
In chapter 2 we will reproduce the main results of Heid (1996). In addition, we will propose an alternative way to estimate the volatility for autoregressive processes. Some further insight into the structure of volatility curves is presented at the end of the chapter. A closer look at the estimates will prove that the volatility curves are of fairly similar and simple shape which leads us to test the hypothesis of them being polynomials of low degree (chapter 3). In chapter 4 we continue to model similarity in a more thorough way. This approach uses results of Selfmodelling Regression introduced by Gasser and Kneip (1992) and Kneip (1993) and goes back to ideas of Principal Component Analysis.

2 Nonparametric Estimation of the Volatility of Financial Prices

In this paper we use two data sets of financial time series. The first one consists of daily noon exchange rates of 19 currencies vis à vis the Canadian Dollar in a period from January 1993 to June 1996, equivalent to 858 trading days. The data was made available by the Pacific Exchange Rate Service of the University of British Columbia, Vancouver, Canada. The second set of data reports daily closing prices of 175 stocks listed at the New York Stock Exchange provided by the MIT Experimental Stock Market Data Server. The recording time is from September 1993 to May 1996, i.e. 623 trading days. The third seris consists of daily dax values, a German stock market index, within a period from January 1988 to November 1996 (2214 trading days). For a detailed discussion of the data sets we refer to Heid (1996). The return process of an asset (exchange rate or stock) is defined by

$$r_t = \log x_t - \log x_{t-1},$$

where x_t is the price of the considered asset. In the sequel we will assume that the return processes are strictly stationary. Let $\mu(r)$ be the mean of tomorrow's return conditioned on that of today's, then we obtain for the conditional variance

$$\sigma^2(r) := \mathrm{var}[r_{t+1}|r_t = r] := E[r_{t+1}^2|r_t = r] - E^2[r_{t+1}|r_t = r] \tag{1}$$

$$= E[r_{t+1}^2|r_t = r] - \mu^2(r). \tag{2}$$

We call $\mu(\cdot)$ the conditional mean and $\sigma(\cdot)$ the (conditional) volatility of the process $\{r_r\}$. Consequently, it follows that

$$r_{t+1} = \mu(r_t) + \sigma(r_t)\xi_{t+1} \tag{3}$$

with some random variables ξ_t. Where necessary we will assume that the ξ_t are i.i.d. in which case equation (3) would constitute an autoregression model. In the next two chapters we will briefly summarize the results of Heid (1996).

2.1 Nonparametric Regression of Time Series

Assume a stationary process $\{r_t\}_{t=1,\dots T+1}$ for which we want to estimate

$$G(r) = \mathrm{E}[g(r_{t+1})|r_t = r].$$

Then a nonparametric estimator of G is given by

$$\widehat{G}_T(r) = \sum_{t=1}^{T} g(r_{t+1})\ k\left(\frac{r_t - r}{h_T}\right) \Bigg/ \sum_{t=1}^{T}\ k\left(\frac{r_t - r}{h_T}\right),$$

with a function k, called kernel. Let

$$H(r) = \mathrm{E}[g^2(r_{t+1})|r_t = r]$$

and

$$\begin{aligned} G &= (G(\psi_1), ..., G(\psi_n)) \\ \widehat{G} &= (\widehat{G}(\psi_1), ..., \widehat{G}(\psi_n)) \end{aligned}$$

then the following theorem holds

Theorem 1 *Let $\{r_t\}$ be strongly mixing and let assumption* (N1) *to* (N9) *of* Heid (1996) *hold. Further define*

$$S_T = \sqrt{h_T T}(\widehat{G} - G).$$

Assume that as T converges to infinity

$$h_T^{2(r+1)} T \to 0 \quad , \quad h_T T \to \infty.$$

Then S_T converges to a vector of independent normal random variables with zero mean the i'th element of the vector having variance

$$\int k^2(u)du\,(H(\psi_i) - G^2(\psi_i))\ /\ f(\psi_i), \tag{4}$$

where f is the density function of r_t. (4) is consistently estimated by

$$\int k^2(u)du\,(\widehat{H}(\psi_i) - \widehat{G}^2(\psi_i))\ /\ \hat{f}(\psi_i), \tag{5}$$

$\widehat{H}$ being the kernel estimator of H and $\hat{f}(\psi)$ equal to $\sum_{t=1}^{T+1} k(\frac{\psi_i - r_t}{h_T})$.

2.2 Direct Estimation of the Volatility Function

Applying theorem 1 of chapter 2.1, equation (2) will lead to straightforward estimates of the volatility functions of the financial time series. However, to gain comparability of the different series, we chose first to standardize the returns by their means and standard deviations, i.e.

$$\tilde{r}_{it} = \frac{r_{it} - m_i}{s_i},$$

where m_i and s_i denote the (unconditional) mean and the (unconditional) standard deviation of the financial series i $(i = 1, ..., I)$.[1] Some easy calculations show that

$$\mu_i(r) := s_i\ \tilde{\mu}\left(\frac{r - m^i}{s^i}\right) + m_i$$
$$\sigma_i(r) := s_i\ \tilde{\sigma}\left(\frac{r - m^i}{s^i}\right)$$

with $\tilde{\mu}_i(r) := \mathrm{E}[\tilde{r}_{i,t+1}|\tilde{r}_{it} = r]$ and $\tilde{\sigma}_i^2(r) := \mathrm{var}[\tilde{r}_{i,t+1}|\tilde{r}_{it} = r]$. Thus, we will make all our calculations on the standardized returns, and, to simplify notation, we will ommit the tilde on all variables. With $m_{i2}(r) := \mathrm{E}[r_{i,t+1}^2|r_t = r]$ an estimate of $\sigma_i^2(r)$ is given by

$$\hat{\sigma}_i^2(r) = \hat{m}_{i2} - \hat{\mu}_i^2(r), \tag{6}$$

where each term on the right hand side of the equation is calculated by the means of the kernel estimator presented in the previous chapter. The bandwidth was

[1] We analyzed the different sets of data seperately so that i refers to a series either in the set of stocks or in the set of exchange rates.

chosen according to the method of cross validation (cf Györfi Sarda and Vieu 1989). The plots of the volatility functions for the different data sets are depicted in figure 1. It it eye-catching that the curves have such similar patterns as nearly all of them are U-shaped with a minimum at a value close to zero. This especially applies to the volatility curves of the exchange rates which seem to be running on roughly parallel lines. We will investigate this similarity in more detail in chapter 4.

2.3 Volatility Estimation for Autoregressive Processes

In this section we will present another estimator of $\sigma^2(\cdot)$ for the case that the ξ_t in (3) are independently distributed. Rearranging this equation yields

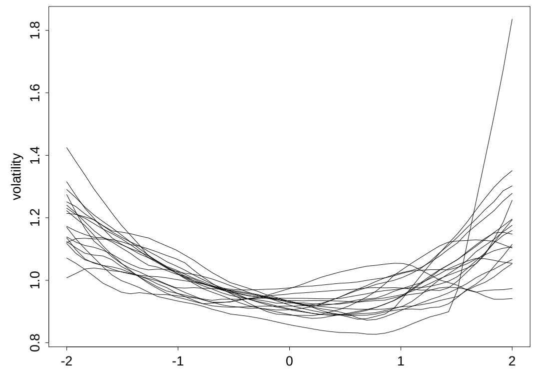
$$\log((r_{t+1} - \mu(r_t))^2) - \mathrm{E}[\log \xi_{t+1}^2] = \log \sigma^2(r_t) + (\log \xi_{t+1}^2 - \mathrm{E}[\log \xi_{t+1}^2]). \tag{7}$$

If we additionally assume that ξ_{t+1} is N(0,1), then $\mathrm{E}[\log \xi_{t+1}^2] \approx -1.272$ and an estimate of $\sigma^2(r)$ is obtained by carrying out the following steps:

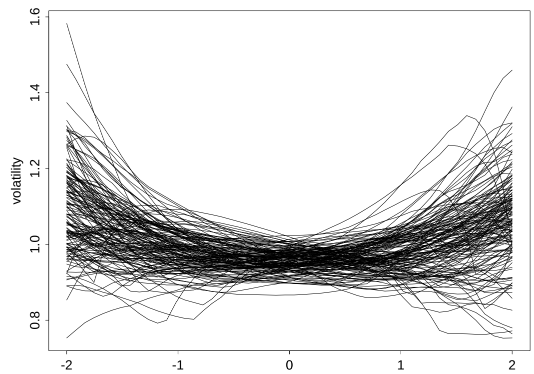
1. Estimate $\mu(r)$ as in the previous section.
2. Estimate $\log \sigma^2(r)$ by regression of $\log(r_{t+1} - \hat{\mu}(r_t))^2 + 1.272$ on r_t.
3. Determine an estimate of σ^2 by $\exp \widehat{\log \sigma^2}$.

Figure 2 provides the plots of these estimates for the different sets of data and it can be seen that their shape roughly corresponds to those of figure 1. Nevertheless, they also display some significant deviations. Generally speaking, they do not possess the same degree of uniformity as the latter. Although most of the curves are still U-shaped, there are some curves which have different pattern. Additionally, the stock plots are more scattered than the plots of the exchange rates, though it is not quite clear whether this is due to vertical shifts only.

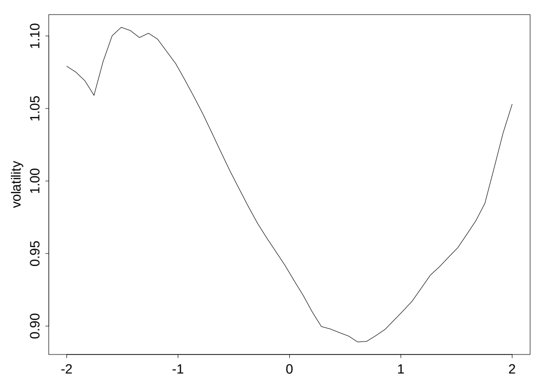
Theorem 1 can be used to test the hypothesis that steps 1) to 3) indeed estimate the volatility function σ^2. More generally, suppose that we want to test the hypothesis that $\sigma^2(r)$ is equal to some function $\tilde{\sigma}^2(r)$ (hypothesis H_0). Recall that $\hat{\sigma}^2$ is the estimator of σ^2 obtained by (6). From theorem 1 we know



volatility functions of exchange rates

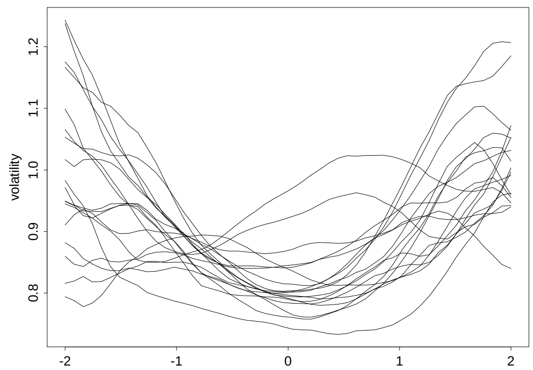


volatility functions of stock returns

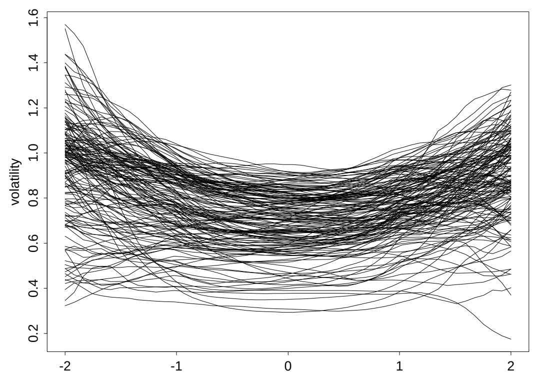


volatility function of dax returns

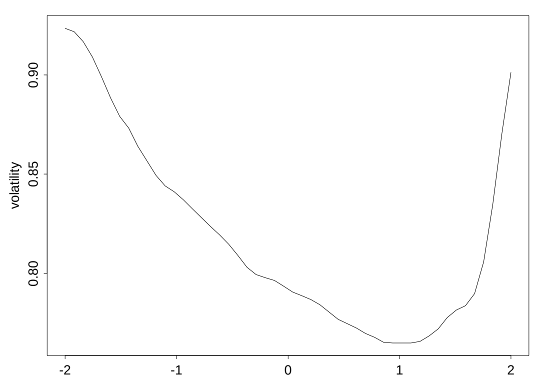
Figure 1: volatility functions



volatility functions of exchange rates



volatility functions of stock returns



volatility functions of dax returns

Figure 2: volatility functions

that $\hat{\sigma}(\psi_1), \ldots, \hat{\sigma}(\psi_n)$ are assymptotically independently normal with, under H_0, $\hat{\sigma}(\psi_i)$ having zero mean and variance v_i^2 given by (4). Hence,

$$d = \sum_{i=1}^{n} \left(\frac{\hat{\sigma}^2(\psi_i) - \tilde{\sigma}^2(\psi_i)}{v_i} \right)^2$$

is $\chi^2(n)$ distributed and, therefore, H_0 will be rejected if d is larger than a critical value.

The test applied to our setting rejects H_0 for all exchange rates as well for the dax and 95% of the stock returns. Two reasons may be responsible for this: a) the random variables ξ_t in (3) are not independent or b) the ξ_t are not normal. We now want to test the autoregression model without assuming the innovations to be normal. For this reason we choose a value for $c := \mathrm{E}[\log \xi_{t+1}^2]$ that minimizes the distance between $\hat{\mathrm{E}}[\log(r_{t+1} - \hat{\mu}(r_t))^2 - c|r_t]$ and $\log \hat{\sigma}^2$, where $\hat{\mathrm{E}}$ denotes the nonparametric estimator for the corresponding conditional mean.[2] Then we proceed as in steps 1. to 3. but replace 1.272 with -c. Afterwards this estimator, which we denote by $\tilde{\sigma}^2$, is submitted to the test described above. As regards the exchange rates, Figure 3 validates that the graphs of $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ are quite close. Indeed, the test whether $\tilde{\sigma}^2$ is equal to $\hat{\sigma}^2$ is rejected in only 7 out of 19 exchange rates. Looked upon the exchange rates as being independent, in which case rejection of 5% would be expected, this would still be a high fraction. Whereas the hypothesis is also confirmed for the dax series the number of rejections is very high for the stocks (113 out of 175). All in all, model (7) seems to be more appropriate for exchange rates than for stock returns.

[2] c is given by $\frac{1}{n}\sum_{i=1}^{n}(\log \hat{\sigma}^2(\psi_i) - \mathrm{E}[\log(r_{t+1} - \hat{\mu}(\psi_i))^2 | r_t = \psi_i])$.

Figure 3: Estimated volatility functions of exchange rates with 95% confidence bounds (solid lines) and volatility functions in the autoregression case (dotted line)

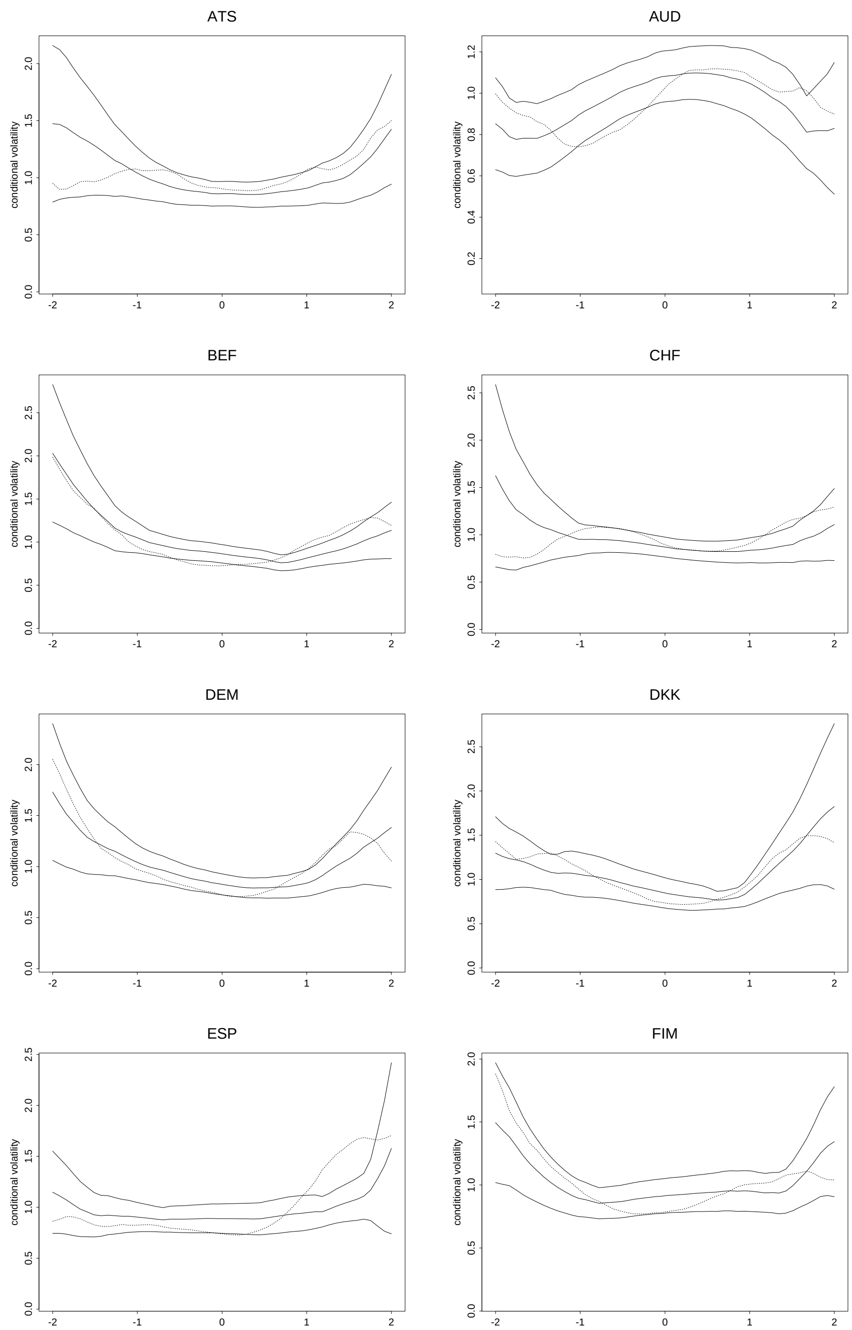


Figure 3 *continued*

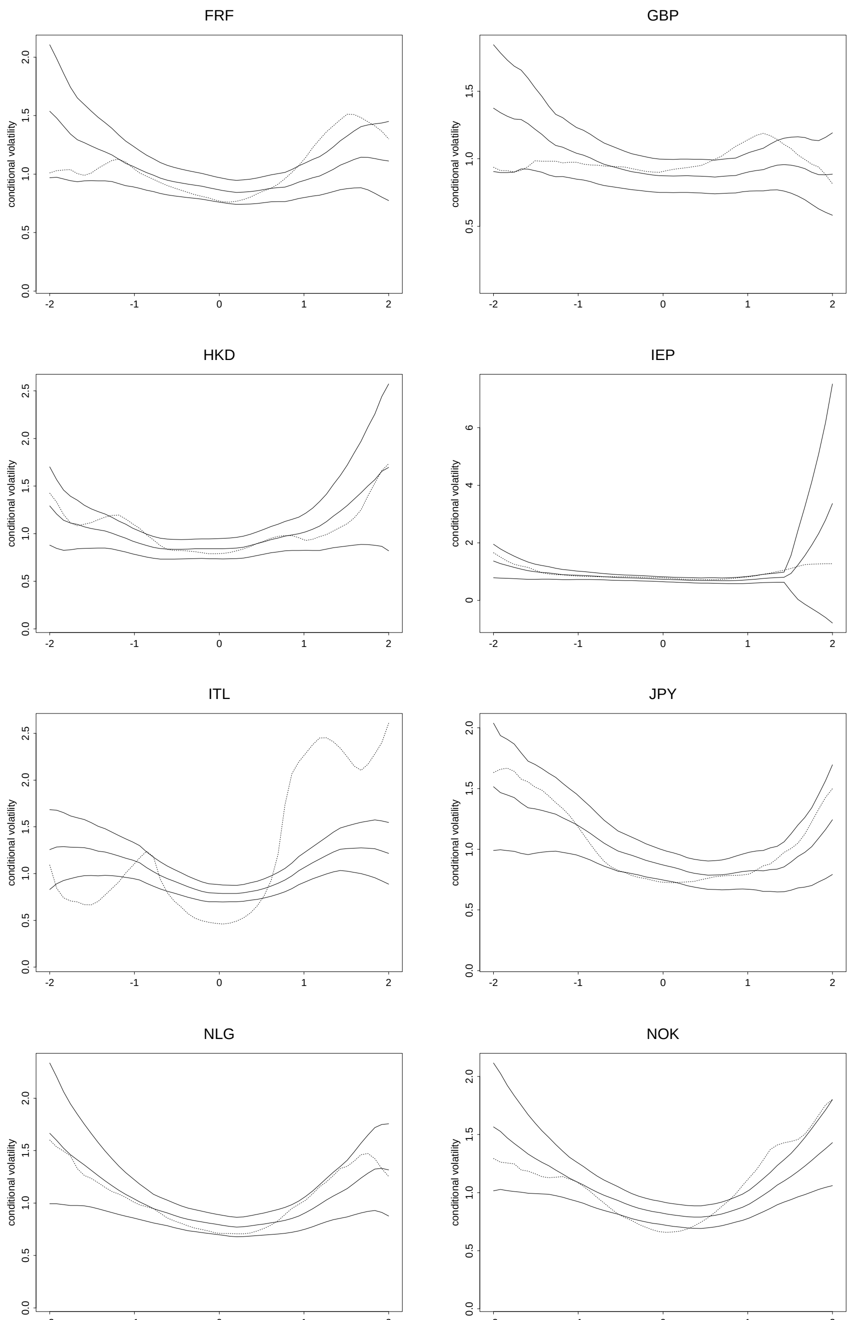
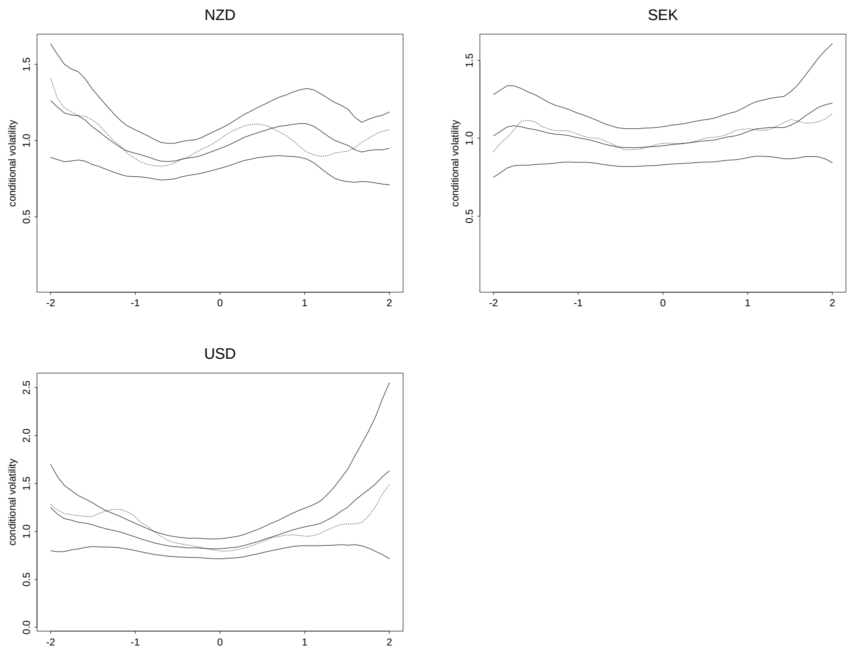


Figure 3 *continued*



2.4 Test on Symmetry

There are two things to be learned from the plots of figure 1: 1. Large absolute returns bring about large future volatility. 2. As a function of returns, the volatility is not always symmetrical. On the other hand there is no clear trend indicating that the left wing of the volatility curves are predominantly flatter than their right wings or vice versa. The hypothesis of the symmetry of the curves can be put to a formal test with the help of theorem 1. Fix $2n$ points $\psi_{-n}, ..., \psi_{-1}, \psi_1, ..., \psi_n$ such that $\psi_{-i} = \psi_i$ for $i = 1, ..., n$. We know that the $\hat{\sigma}^2(\psi)$ are assymptotically independent $\mathrm{N}(0, v_i)$ with variance v_i^2 calculated from (4). Hence, the test-statistic

$$d = \sum_{i=1}^{n} \left(\frac{\hat{\sigma}^2(\psi_i) - \hat{\sigma}^2(\psi_{-i})}{\sqrt{v_i^2 + v_{-i}^2}} \right)^2$$

is assymptotically $\chi^2(n)$ distributed and, consequently, the test rejects the hypothesis of symmetry of the volatility function if d is larger than a critical value.

The test applied to our data shows that the volatility function of the dax returns is symmetrical whereas the returns of the Belgian Francs (BEF), the Japanese Yen (JPY), the Norwegian Kroner (NOK) and the New Zealand Dollar (NZD) have asymmetrical volatility. For the stock data the test rejects symmetry for 48 out of 175 series. Overall it can be said that significant assymmetry does exist, albeit slight.

3 Polynomial Fit of the Volatility Function

In the remainder of this paper we will focus on the similarity of the volatility functions already mentioned in the previous section. For this reason we will begin by modelling the volatility as a specific function which depend on returns and some further parameters. The number of included parameters can be viewed as a yardstick measuring the diversity of the respective curves. A simple as well as flexible way to model volatilities is given by polynomial regression, so let us assume that

$$\sigma_i^2(r) = \sum_{l=0}^{L_i} \theta_l^i r^l, \tag{8}$$

and we impose the important restriction that the degree of the polynomial (as well as θ_l) is unknown.(The index i indicates the name of the series in the considered data set.) A brief glance at the plots in figure 1 suggests that the L_i are small, perhaps being 2 or 3.

More generally, assume that the volatilities are the weighted sums of some known functions $g_1, g_2, ...$, i.e.,

$$\sigma_i^2(r) = \sum_{l=1}^{L_i} \theta_{il} g_l(r)$$

where L_i and θ_l, $l = 1, ..., L_i$ are unknown.[3] Let $\psi_1, ..., \psi_n$ be some predetermined points at which we are to evaluate the $\sigma_i^2(\cdot)$. It is tempting to estimate the unknown dimension L_i by solving the following problem:

$$\min_L \min_{\theta_1,\dots,\theta_L} \sum_{j=1}^{n} \left(\hat{\sigma}_i^2(\psi_j) - \sum_{l=1}^{l} \theta_l g_l(\psi_j) \right)^2 \tag{9}$$

[3]In the case of polynomial regression $g_l(r) = r^{l-1}$.

where $\hat{\sigma}_i^2$ is the nonparametric estimate of σ^2 given in equation (6). This problem, however, has in general no solution since it leads one to add as many functions g_l as possible. This can be avoided by adding to (9) a second term which is increasing in L, thereby introducing a penalty for adding additional functions g_l. Along this lines an estimator was proposed by Mallows (1973) (Appendix A). From theorem 1 we know that

$$\begin{aligned} \hat{\sigma}_i^2(\psi_j) &= \sigma_i^2(\psi_j) + \epsilon_{ij} \\ &= \sum_{l=1}^{L} \theta_{il} g_l(\psi_j) + \epsilon_{ij} \end{aligned} \tag{10}$$

with $\epsilon_{ij}, j = 1, ..., n$ being assymptotically independent $N(0, v_{ij})$-distributed and variances v_{ij}^2 given by (4). To be able to apply the method described in Appendix A we have to divide both sides of (10) by v_{ij}, i.e.

$$\frac{\hat{\sigma}_i^2(\psi_j)}{v_{ij}} = \sum_{l=1}^{L} \theta_{il} \frac{g_l(\psi_{ij})}{v_{ij}} + \tilde{\epsilon_{ij}}$$

with $\tilde{\epsilon_{ij}}$ being i.i.d. N(0,1). The dimension L is then estimated by solving the problem

$$\min \mathrm{C_L}^i(L) = \min_L \left(\min_{\theta_1,\ldots,\theta_l} \sum_{j=1}^{n} \left(\frac{\hat{\sigma}_i^2(\psi_j)}{v_{ij}} - \sum_{l=1}^{L} \theta_l \frac{g_l(\psi_j)}{v_{ij}} \right)^2 + 2L \right). \tag{11}$$

Setting $g_l = x^{l-1}$ we have a case of polynomial regression.

In determining the number of points ψ_j two opposing goals need to be balanced. On the one hand n should be large enough in order to give sufficient structure to the points $\hat{\sigma}_i^2(\psi_j)$ to be approximated. Noting that it is always possible to interpolate L points perfectly by a polynomial of degree $L-1$ the number of points should at least exceed $L^* + 1$ if L^* is a preliminary guess of the polynomial degree. On the other hand we cannot expect the points $\hat{\sigma}_i^2(\psi_j)$ to be approximately independent by means of theorem 1 if n is too large. For our purpose we chose 6 points equally spaced in the interval $[-2, 2]$.
Table 1 states the estimates of the polynomial degree for the different exchange rates, most of which clearly equal 2. If, in fact, we think that the degree of the polynomial is *always* 2 (no contradiction to table 1 since the entrants are realisations of random numbers) we can improve on the estimation above. Instead

ATS	**AUD**	**BEF**	**CHF**	**DEM**
2	2	2	0	2
DKK	**ESP**	**FIM**	**FRF**	**GBP**
2	0	2	2	0
HKD	**IEP**	**ITL**	**JPY**	**NLG**
2	2	2	2	2
NOK	**NZD**	**SEK**	**USD**	
2	0	0	2	

Table 1: Estimated degree of the polynomial fit of σ^2.

of minimizing all $\mathrm{C_L}^i$-functions seperatly, it would be better to choose L as to minimize

$$\overline{\mathrm{C_L}}(L) := \mathrm{C_L}^1(L) + ... + \mathrm{C_L}^N(L).$$

As expected, the estimated degree for the exchange rates is 2. We also applied this estimator to the stocks and the dax series obtaining the same result.

To evaluate the fit of σ_i^2 by quadratic polynomials we propose a simple χ^2-test based on least squares regression. In a general setting this works as follows: assume that

$$\hat{\sigma}^2(\psi_j) = \sigma^2(\psi_j) + \epsilon_j = \sum_{l=1}^{L} \theta_l g_l(\psi_j) + \epsilon_j, \quad j = 1, ...n$$

where ϵ_j are i.i.d. $N(0, v_j)$. Denote by

$$\begin{aligned} \underline{\tilde{\sigma}}^2 &= (\hat{\sigma}^2(\psi_1)/v_1, ..., \hat{\sigma}^2(\psi_n)/v_n)^{\mathrm{T}} \\ g_l &= (g_l(\psi_1)/v_1, ..., g_l(\psi_n)/v_n)^{\mathrm{T}} \\ \tilde{\epsilon} &= (\epsilon_1/v_1, ..., \epsilon_n/v_n)^{\mathrm{T}} \end{aligned}$$

and

$$G = (\tilde{g}_1, ..., \tilde{g}_L).$$

Let $\hat{m}$ be the least squares projection of $\tilde{\sigma}^2$ on $\tilde{g}_1, ...\tilde{g}_L$, then the residual sum of squares is equal to

$$\begin{aligned} \mathrm{RSS} &= \|\underline{\tilde{\sigma}}^2 - \hat{m}\|_2^2 = \|y - G(G^{\mathrm{T}}G)^{-1}G^{\mathrm{T}}\underline{\tilde{\sigma}}^2\|_2^2 \\ &= \tilde{\epsilon}^{\mathrm{T}}(I - G(G^{\mathrm{T}}G)^{-1}G^{\mathrm{T}})\tilde{\epsilon} = \tilde{\epsilon}^{\mathrm{T}}M\tilde{\epsilon} \end{aligned}$$

ATS	**AUD**	**BEF**	**CHF**	**DEM**
0.26	1.68	0.87	0.70	0.26
DKK	**ESP**	**FIM**	**FRF**	**GBP**
1.68	0.86	1.87	0.31	0.58
HKD	**IEP**	**ITL**	**JPY**	**NLG**
0.12	1.33	4.04	0.92	0.11
NOK	**NZD**	**SEK**	**USD**	
0.42	3.39	0.29	0.14	

Table 2: Residual sum of squares of polynomial fit of σ^2. Critical value: 6.0

M is idempotent of rank $T - L$, hence, under the correct choice of L, RSS is $\chi^2(n - L)$-distributed. Thus, the hypothesis of L being equal to some L^* is rejected if RSS is larger than a critical value. The residual sums of squares for the exchange rates are listed in table 2. It clearly confirms the hypothesis that σ_i^2 is always a quadratic polynomial.

So far we have calculated the polynomial fit for each series seperately. If we assume that the series are independent, we can test the hypothesis that *all* volatility functions are of degree 2.[4] Then, under the correct choice of the polynomial degree, the overall sum of the individual RSS_i, $i = 1, ..., N$, is $\chi^2(N(n-L))$ distributed. As it turnes out this test also accepts the hypothesis of σ_i^2 being quadratic.

Polynomial fit of transformed volatility

Since returns are random and σ^2 cannot be negative, equation (8) is only sensible if the θ_l are chosen in such a way that the term on the right side is always positive. Thus, it is sometimes more appropriate to model suitable *transformations* of σ^2 instead. In the following paragraph we will therefore assume that $\log \sigma^2$ is a

[4]Independence seems to be no realistic assumption for exchange rates, but note that we are investigating returns on exchange rates and not exchange rates themselves. While exchange rates may be dependent, independence can nevertheless hold true for the returns.

polynomial of some degree L, i.e.,

$$\log \sigma^2(r) = \sum_{l=0}^{L} \theta_l r^l. \tag{12}$$

Unfortunately, devising a test similar to that of the previous section is rather intricate since it necessitates calculating the joint distribution of the $\log \hat{\sigma}^(\psi_i)$, $i = 1, \ldots n$. If we however assume that the innovations ξ_t in equation (3) are identically independently distributed, which seems to be a reasonable assumption for many exchange rates, a suitable test is much easier constructed since

$$\mathrm{E}[\log(r_{t+1} - \mu(r_t))^2 | r_t] = \log \sigma^2(r_t) + \mathrm{E}[\log \xi_t^2]. \tag{13}$$

Hence, $\log \sigma^2$ and $\mathrm{E}[\log(r_{t+1} - \mu(r_t))^2 | r_t]$ are polynomials of the same degree and we are free to choose either of the functions for estimating the dimension. Denoting by $\widehat{\mathrm{E}}[y|x]$ the nonparametric estimate of $\mathrm{E}[y|x]$, we know from theorem 1 that

$$\widehat{\mathrm{E}}[\log(r_{t+1} - \hat{\mu}(\psi_j))^2 | r_t = \psi_j] = \mathrm{E}[\log(r_{t+1} - \mu(\psi_j))^2 | r_t = \psi_j] + \epsilon_j \tag{14}$$

where the ϵ_j are asymptotically independently normal. The results of Mallows' estimator are listed in table 3. Most of the values are equal to 2, which is also the result of the overall estimator. The results of the χ^2-test can be seen in table 4. They show that the hypothesis of $\log \sigma^2(r)$ being a polynomial of degree is accepted.

4 Linear Selfmodelling Regression

In the previous chapter we saw that the volatility functions are of rather similar shape. In fact, one might get the impression that simple linear transformations suffice to make any two different volatility functions congruent. If this holds true, all volatility functions could be written in the form

$$\sigma_i^2(r) = a_i g(r) + c$$

with some function $g(\cdot)$. In this particular case the volatility functions would span a linear space of dimension 1. However, since we do not want to fix the

ATS	**AUD**	**BEF**	**CHF**	**DEM**
2	0	2	2	2
DKK	**ESP**	**FIM**	**FRF**	**GBP**
2	2	0	0	2
HKD	**IEP**	**ITL**	**JPY**	**NLG**
3	2	0	2	2
NOK	**NZD**	**SEK**	**USD**	
4	3	0	2	

Table 3: Estimated degree of the polynomial fit of $\log \sigma^2$.

ATS	**AUD**	**BEF**	**CHF**	**DEM**
1.2	2.6	2.7	4.6	3.8
DKK	**ESP**	**FIM**	**FRF**	**GBP**
3.1	1.9	4.7	4.0	4.2
HKD	**IEP**	**ITL**	**JPY**	**NLG**
3.4	0.9	4.4	1.6	2.9
NOK	**NZD**	**SEK**	**USD**	
7.2	4.7	0.1	0.5	

Table 4: Residual sum of squares of polynomial fit of $\log \sigma^2$. Critical value: 7.8

dimension beforehand we generally ask whether the functional space given by

$$S = \{s : \mathbb{R} \to \mathbb{R} | s(x) = \sum_{i=1}^{N} a_i \sigma_i^2(x)\}.$$

is of low dimension. It is clear that the dimension of S is at most N, but what we would like to know is, whether it could also be less, say L_0. In order to determine L_0 we need to find a suitable set of basis functions $g_1, ..., g_{L_0}$. These will then completely characterize S since

$$\sigma_i^2(x) = \sum_{l=1}^{L_0} a_{il} g_l(x) \tag{15}$$

for some parameters a_{il}. It should be emphasized that we do not want to prespecify the basis functions to be of a specific parametric form as in chapter 3 (eg polynomials), but rather intend to model them as flexible as possible. Following the more general approach of Kneip and Gasser (1992), we call our method *Linear Selfmodelling Regression*.

4.1 Selecting Basis Functions

Let us now fix some points $\psi_1, ..., \psi_n$ in an interval $[a, b]$ at which the volatility functions as well as the basis functions $g_1, ..., g_{L_0}$ are to be calculated. Define

$$\begin{aligned} \underline{\sigma}_i^2 &= (\sigma_i^2(\psi_1), ..., \sigma_i^2(\psi_n))^{\mathrm{T}} \\ \underline{g}_l &= (g_l(\psi_1), ..., g_l(\psi_n))^{\mathrm{T}}. \end{aligned}$$

Then

$$\underline{\sigma}_i^2 = \sum_{l=1}^{L_0} a_{il} \underline{g}_l.$$

Clearly, the basis vectors $\underline{g}_l$ are not uniquely determined, therefore we impose some normalization conditions:

a) $\underline{g}_r^{\mathrm{T}} \underline{g}_s = \delta_{rs}$,

b) $\sum_{i=1}^{N} a_{ir} a_{is} = 0 \quad r \neq s$,

c) $\sum_{i=1}^{N} a_{i1}^2 \geq \sum_{i=1}^{N} a_{i2}^2 \geq ... \geq \sum_{i=1}^{N} a_{iL_0}^2 > 0$.

Of course, the normalization conditions depend on the particular choice of $\psi_1, ..., \psi_n$. If these points are equidistantly chosen in the interval $[a, b]$, then $\frac{1}{n}\underline{g}_r^{\mathrm{T}}\underline{g}_s$ converges to $\frac{1}{b-a}\int g_r(x)g_s(x)\,dx$, the standard product form of continuous functions on $[a, b]$. Given a) to c) the following equations hold:

$$
\begin{aligned}
W := \sum_{i=1}^{N} \underline{\sigma}_i^2 \underline{\sigma}_i^{2\mathrm{T}} &= \sum_{i=1}^{N} \left(\sum_{l=1}^{L_0} a_{il}\underline{g}_l\right)\left(\sum_{l=1}^{L_0} a_{il}g_l\right)^{\mathrm{T}} \\
&= \sum_{l=1}^{L_0}\sum_{i=1}^{N} a_{il}^2 \underline{g}_l \underline{g}_l^{\mathrm{T}}.
\end{aligned}
$$

Hence, $\sum_{i=1}^{N} a_{i1}^2, ..., \sum_{i=1}^{N} a_{iL_0}^2$ are the eigenvalues of W in descending order. We shall now prove that it is always possible to identify a basis satisfying a) to c). Let $g_1, ...g_N$ be the eigenvectors of W corresponding to eigenvalues in descending order. Define the matrix G as

$$G = (\underline{g}_1, ..., \underline{g}_N)$$

and the vector a_i by

$$\underline{\sigma}_i^2 = Ga_i.$$

Since the image space of W is $\mathrm{span}(\underline{g}_1, ..., \underline{g}_{L_0})$ the parameters $a_{i,L_0+1}, ..., a_{i,N}$ are equal to zero. Because W is positive semidefinite, there exists a diagonal matrix Λ such that

$$W = G\Lambda G^{\mathrm{T}}.$$

Let $A := (a_1, ..., a_N)$, then

$$G\Lambda G^{\mathrm{T}} = W = GA$$

yielding

$$
\begin{aligned}
A &= \Lambda G^{\mathrm{T}} \\
AA^{\mathrm{T}} &= \Lambda G^{\mathrm{T}} G \Lambda^{\mathrm{T}} = \Lambda\Lambda^{\mathrm{T}}.
\end{aligned}
$$

Condition c) is thus established by noting that $\Lambda\Lambda^{\mathrm{T}}$ is diagonal.

Well-known results of Linear Algebra teach us with respect to c) that if the strict instead of the weak inequality holds parameters and basis functions are uniquely determined up to sign changes. Clearly, one may replace a_{ir} by $-a_{ir}$ and $\underline{g}_r$ by $-\underline{g}_r$ without invalidating conditions a) to c). This indeterminacy might be removed by an appropriate additional normalizing condition such as $a_{i1} > 0$.

4.2 Estimation of Basis Function

Having shown how we can identify a suitable basis, we move on to estimate the basis function. From theorem 1 we know that nonparametric estimates of the volatility functions satisfy

$$\hat{\sigma}_i^2(\psi_j) = \sigma_i^2(\psi_j) + \epsilon_{ij} \tag{16}$$

with ϵ_{ij} being assymptotically $N(0, v_{ij})$ distributed. Hence, it is straightforward to estimate W by

$$\widehat{W} = \sum_{i=1}^{N} \underline{\hat{\sigma}}_i^2 \underline{\hat{\sigma}}_i^{2\mathrm{T}}$$

where $\underline{\hat{\sigma}}_i^2 = (\hat{\sigma}_i^2(\psi_1), ..., \hat{\sigma}_i^2(\psi_n))^{\mathrm{T}}$ is the vector of nonparametric estimates of the volatilities. For reasons which will become clear later in this chapter it is more appropriate to consider another matrix $\widetilde{W}$ which is defined by

$$\widetilde{W} = \sum_{i=1}^{N} \underline{\tilde{\sigma}}_i^2 \underline{\tilde{\sigma}}_i^{2\mathrm{T}}$$

where $\underline{\tilde{\sigma}}_i^2 = (\hat{\sigma}_i^2(\psi_1)/v_{i1}, ..., \hat{\sigma}_i^2(\psi_n)/v_{in})^{\mathrm{T}}$. The difficulty of this approach, however, is that a model for $(\sigma_i^2(\psi_1)/v_{i1}, ..., \sigma_i^2(\psi_n)/v_{in})$ does not necessarily lead to a model of $\underline{\sigma}_i^2$. This problem can be solved if v_{ij}^2 can be decomposed: there exists a function w such that

$$v_{ij}^2 = \bar{v}_i^2 \cdot w^2(\psi_j).$$

Without restriction we recquire that

$$\frac{1}{N} \sum_{j=1}^{n} w^2(\psi_j) = 1.$$

Consequently

$$\bar{v}_i^2 = \frac{1}{n} \sum_{j=1}^{n} v_{ij}^2$$

and

$$w^2(\psi_j) = \frac{1}{N} \sum_{i=1}^{N} v_{ij}^2 / \bar{v}_i^2.$$

The original model

$$\sigma_i^2(\psi_j) = \sum_{l-1}^{L_0} a_{il} g_l(\psi_j)$$

implies for the transformed model

$$\frac{\hat{\sigma}_i(\psi_j)}{v_{ij}} = \sum_{l=1}^{L_0} \frac{a_{il}}{\bar{v}_i} \cdot \frac{g_l(\psi_j)}{w(\psi_j)}.$$

Hence, the transformed model has the same dimension as the original one and basis function of the latter are easily calculated from the other. However, since $\widetilde{W}$ has rank N, it is difficult to determine which of the eigenvectors may function as estimates of the basis functions of the transformed model.

Apart from this problem the eigenvectors $\gamma(\widetilde{W})$ still have an important geometrical meaning since they are optimal in a least squares sense (Appendix B): For all L

$$\min_{p_1,\dots p_L \in \mathbb{R}^n} \sum_{j=1}^{L} \min_{a_{i1},\dots,a_{iL} \in \mathbb{R}} \left\| \underline{\hat{\sigma}}_i^2 - \sum_{l=1}^{L} a_{il} p_l \right\|_2^2 = \sum_{i=1}^{N} \min_{a_{i1},\dots,a_{iL} \in \mathbb{R}} \left\| \underline{\hat{\sigma}}_i^2 - \sum_{l=1}^{L} a_{il} \gamma_l(\widetilde{W}) \right\|_2^2$$

$$= \sum_{l=L+1}^{N} \lambda_l(\widetilde{W}) \quad (17)$$

where $\lambda_1(\widetilde{W}), \dots, \lambda_n(\widetilde{W})$ are the eigenvalues of $\widetilde{W}$ in descending order. Equation (17) justifies considering the eigenvectors of $\widetilde{W}$ independently of the modelling motivation outlined in the previous section. This approach incorporates idea of the *Principal Component Analysis* which simplifies the analysis of a large number of observed variables by replacing these with a smaller set of linear combinations. References can be found in Rao (1958).

We calculated nonparametric estimates of the volatility functions at 50 points equidistantly spaced in the interval [-2,2] which, on average, covered 90% of the returns of a series. From table 5, which lists the 10 largest eigenvalues of the matrix $\widetilde{W}$ for the two data sets it is evident that the first eigenvector plays a predominant role in explaining the variability of the volatility curves. The eigenvectors shown in figure 4 correspond to the first four eigenvalues. Eigenvectors corresponding to eigenvalue 5 and more are nearly insignificant. Our initial impression regarding the nonparametric estimates of the volatility functions as positive functions with unique minimums at values close to zero is confirmed.

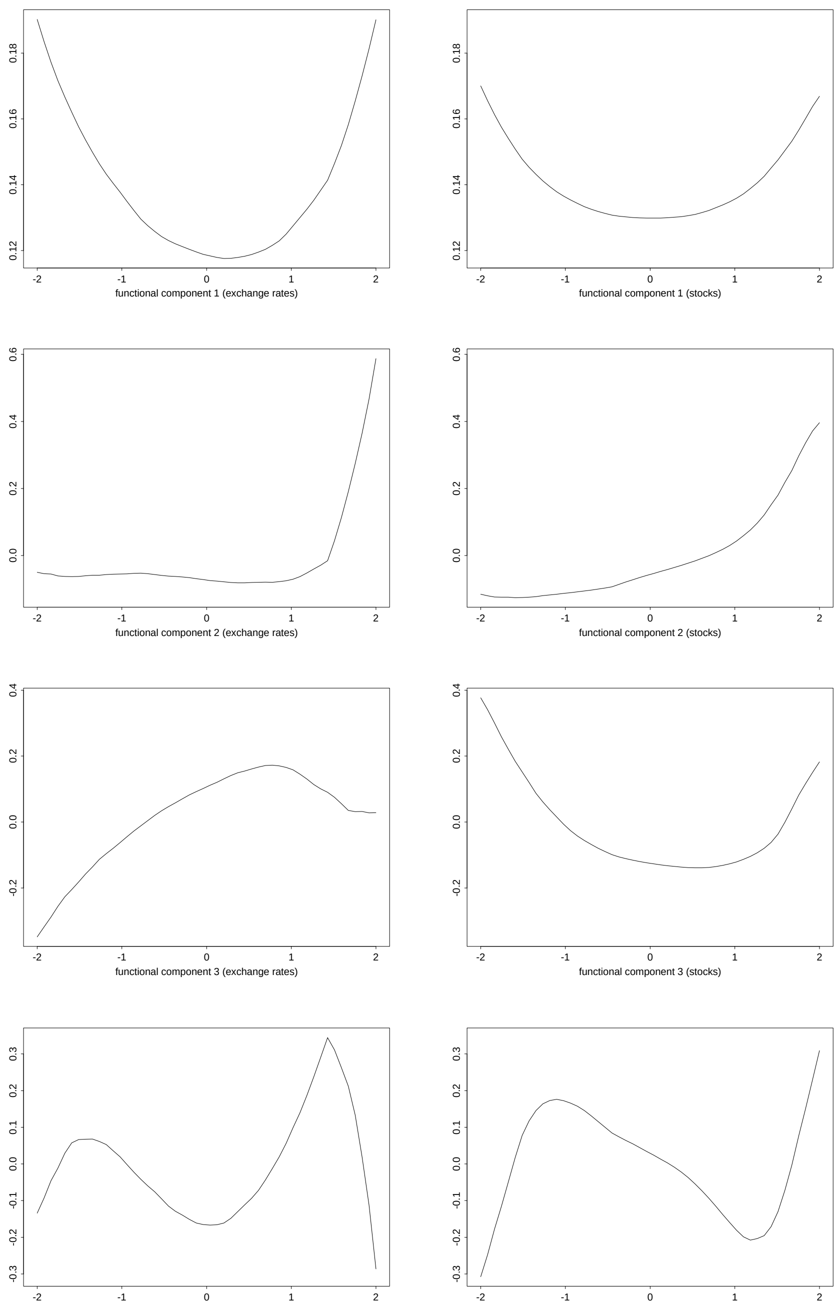


Figure 4: First four functional components: exchange rates (left column), stocks (right column)

number	exchange rates	stocks
1	1013	8757
2	12.5	236
3	6.3	77
4	1.5	17.6
5	0.68	4.4
6	0.21	0.8
7	0.032	0.43
8	0.019	0.31
9	0.0074	0.27
10	0.006	0.078

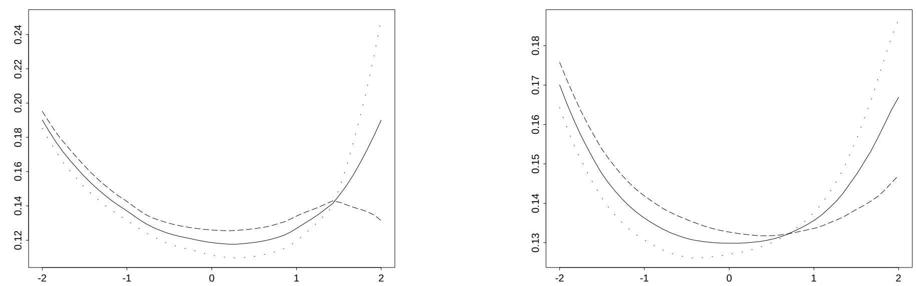
Table 5: 10 largest eigenvalues of $\widetilde{W}$

(Note, that $\gamma_1(\widetilde{W})$ is just a multiple of the average of the vectors $\underline{\tilde{\sigma}}_1^2, ..., \underline{\tilde{\sigma}}_N^2$.) However, the eigenvector $\gamma_1(\widetilde{W})$ does not explain all asymmetries found in the volatility functions. This is where other eigenvectors come into play. Figure (5) elucidates the effect of adding and subtracting additional components to $\gamma_1(\widetilde{W})$. Let us only consider the components of the exchange rates. Clearly, the second component is mainly responsible for upward or downward shifts of the first component where some asymmetry effect is due to the upward slope of $\gamma_2(\widetilde{W})$ at the far right of the domain. The decisive factor in producing asymmetry is the third component, which is also able to shift the value at which the minimum is obtained to the left or the right of the x-axis.

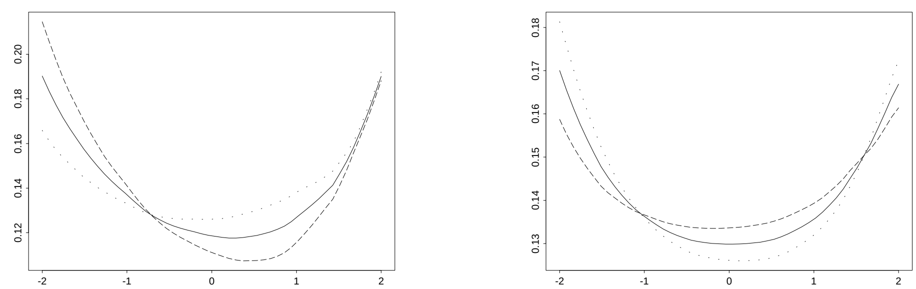
4.3 Estimation of Dimension

In chapter 3 we discussed two methods determining the degree of the polynomial fit of the volatility functions. These methods can also be applied to estimate the dimension of the space S. Similar to the stance taken there, Mallows' estimator can be used to determine the optimal number of included components in least

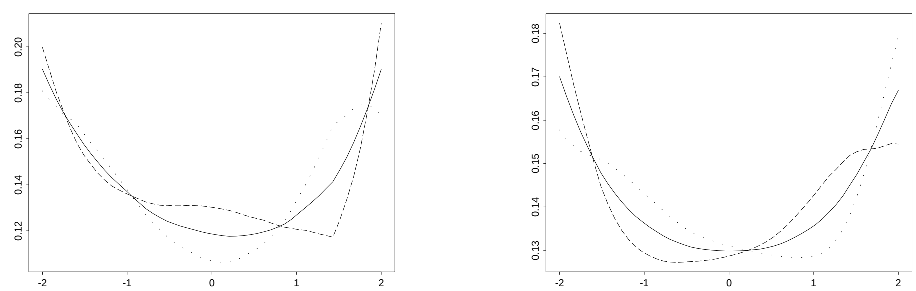
exchange rates **stocks**



(a)



(b)



(c)

Figure 5: First functional component and effect of adding and subtracting multiples of additional components: (a) component 2, (b) component 3, (c) component 4

ATS	**AUD**	**BEF**	**CHF**	**DEM**
1	4	3	1	1
DKK	**ESP**	**FIM**	**FRF**	**GBP**
1	1	1	2	2
HKD	**IEP**	**ITL**	**JPY**	**NLG**
1	1	4	1	1
NOK	**NZD**	**SEK**	**USD**	
1	3	3	1	

Table 6: Estimated degree of the functional component fit of σ^2.

square regression by solving for each series i:

$$\min_L C_L^i(L) = \min_L \left(\min_{\theta_1,\ldots,\theta_L} \left\| \underline{\tilde{\sigma}}_i^2 - \sum_{l=1}^{L} \theta_l \gamma_l(\widetilde{W}) \right\|_2^2 + 2L \right).$$

As already mentioned in chapter 3, the number of design points has to be small enough to guarantee some degree of independence of the entries in $\underline{\hat{\sigma}}_i^2$. For this purpose we have chosen 6 equidistantly spaced points in the interval [-2,2]. We want to emphasize, however, that the results below are quite insensitive to the number actually chosen. Table 6 summarizes the results of this analysis for the exchange rates. As can be seen, for the majority of the series it is optimal to include only one component. A similar outcome was obtained for the stocks where Mallows' estimator selected only one single component in 57% of the series. We cannot, however, infer from these estimates the dimension of S since each C_L^i is itself apt to noise. A suitable estimator which we applied here is determined by minimizing *the sum of all* C_L^i. The estimated dimension was surprisingly low: it was 3 for the exchange rates and 2 for the stocks. The χ^2-test described in chapter 3 confirmed this result: The hypothesis of the dimension being 3 for the exchange rates and 2 for the stocks was accepted whereas the hypothesis of it being 2 and 1 resp. was rejected. Nonetheless, one could argue that the estimated dimension of S depends on the chosen reference interval $[a, b]$. We would expect the dimension to rise if we increase the interval, yet further analysis showed that the dimension is quite insensitive to the length of the reference interval. Figure 6 demonstrates how well the three components approximate the nonparamtric estimates of the volatility functions. Obviously, the two curves are nearly always extremely close, the least squares approximation being nearly always completely in the 95% confidence bounds of the nonparametric estimate.[5]

[5] It is important to note that this constitues no formal test of the curves being equal. In that case it would be necessary to consider joint confidence bounds.

Figure 6: Volatility functions of exchange rates (solid line) with 95% confidence bounds and least squares approximation by 3 functional components

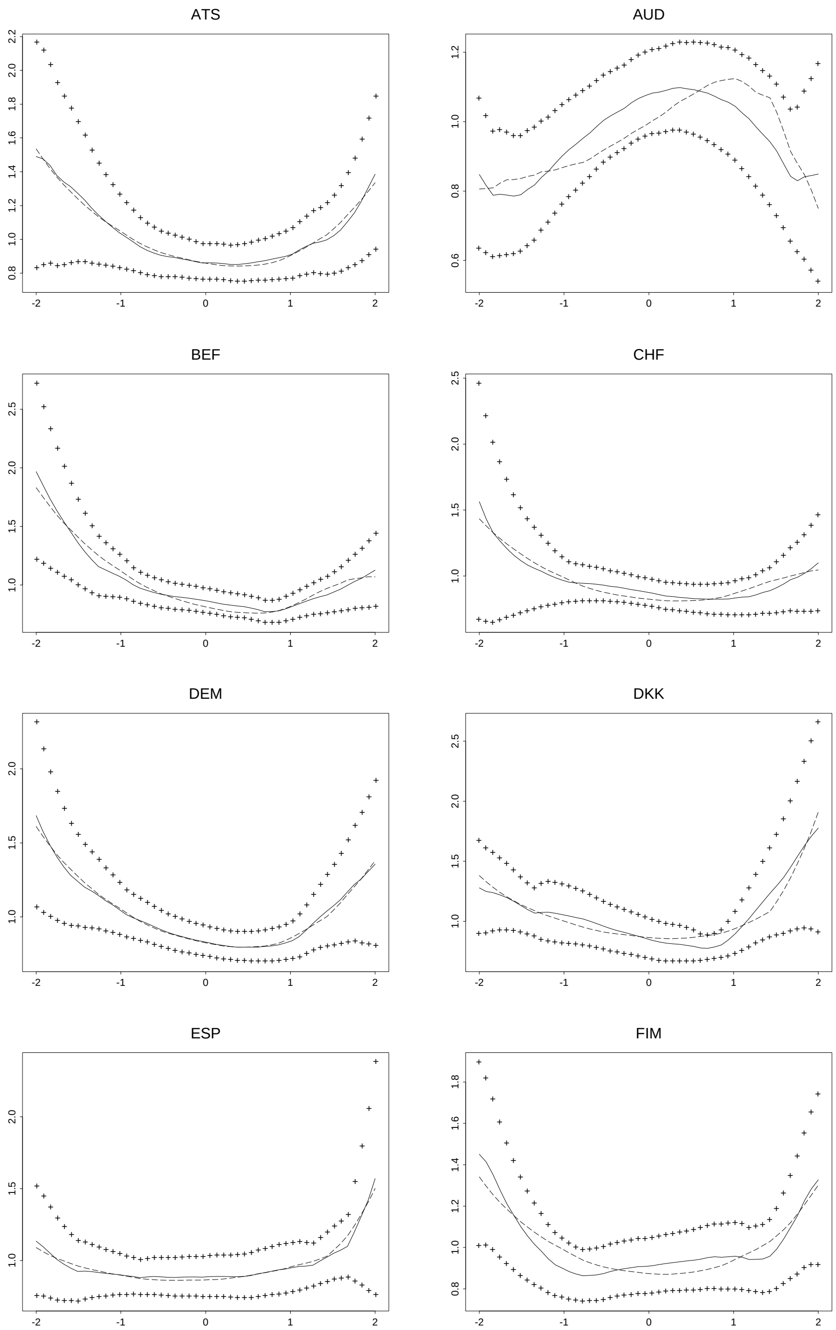


Figure 6 *continued*

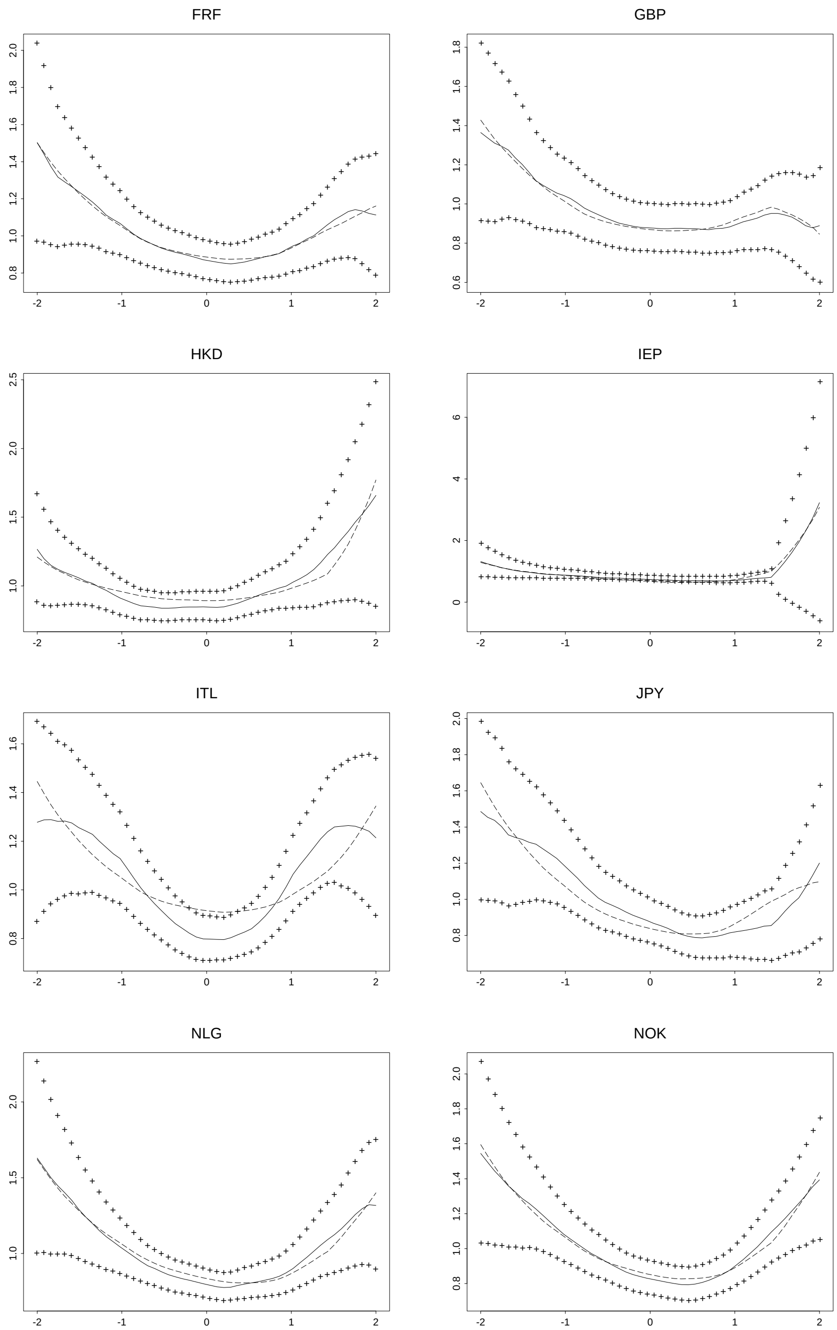
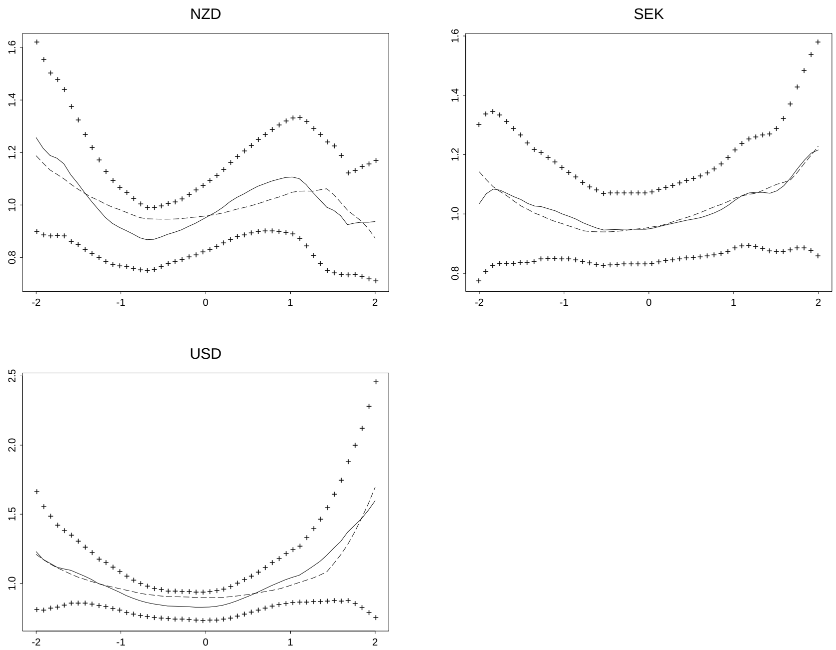


Figure 6 *continued*



4.4 Functional Components of Transformed Volatility

By modelling volatility functions as in (15) one has to take into account that the linear combination of the basis functions has to be positive. To avoid this drawback one sometimes prefers to model $\log \sigma^2$ instead which gives rise to the hypothesis that

$$\log \sigma_i^2(x) = \sum_{l=1}^{L} a_{il} h_l(x) \tag{18}$$

with some basis function $h_1(x), ..., h_L(x)$.
Of course, the logarithm is not a linear transformation so that if (15) holds true for $L_0 < N$, then $\mathrm{span}\{\log \sigma_1^2, ..., \log \sigma_N^2\}$ generally has dimension N rather than dimension less than N. Nevertheless, from a statistical point of view it makes sense to investigate model (15) as well as model (18) as statistical tests may not be able to decide between the two alternatives. If we assume that in (3) the random variables ξ_t are independent normally distributed then

$$\log \sigma_i^2(x) = \mathrm{E}[\log(r_{i,t+1} - \mu(r_{it}))^2 + 1.272 | r_{it} = x] := \phi_i(x). \tag{19}$$

ATS	**AUD**	**BEF**	**CHF**	**DEM**
1	2	2	1	2
DKK	**ESP**	**FIM**	**FRF**	**GBP**
2	3	2	1	1
HKD	**IEP**	**ITL**	**JPY**	**NLG**
2	1	1	2	2
NOK	**NZD**	**SEK**	**USD**	
3	1	1	1	

Table 7: Estimated degree of the functional component fit of $\log \sigma^2(x) + 1.272$.

The functions $\phi_i(x)$ can be estimated by nonparamtric regression as described in chapter 2.1, and by replacing σ_i^2 with ϕ_i in chapter 4.1 and 4.2 we are equally able to select functional components of $\phi_i(\cdot)$. Of course, as we have shown before, equation (19) holds not true for many series. Nevertheless, it still makes sense to model ϕ_i as a proxy for the volatility of an process since however we chose $\mathrm{E}[\log \xi_{it}^2]$ (here: -1.272) a model for the functions ϕ_i has the same dimension as the model for the functions $\log \sigma_i^2$. In figure 7 we plotted the first 4 components for the exchange rate data and for the stock data respectively. In table 7 we are presenting the results of Mallows' estimator applied to each exchange rate seperately. In 9 out of 19 series one component suffices for a satisfactory approximation of the volatility function, for the rest one or more components have to be added. With respect to the stocks, about 43% were optimally fitted by 1 component, 36% by 2, and 21% by 3 and more. The overall estimate of the dimension was 3 for both data sets.

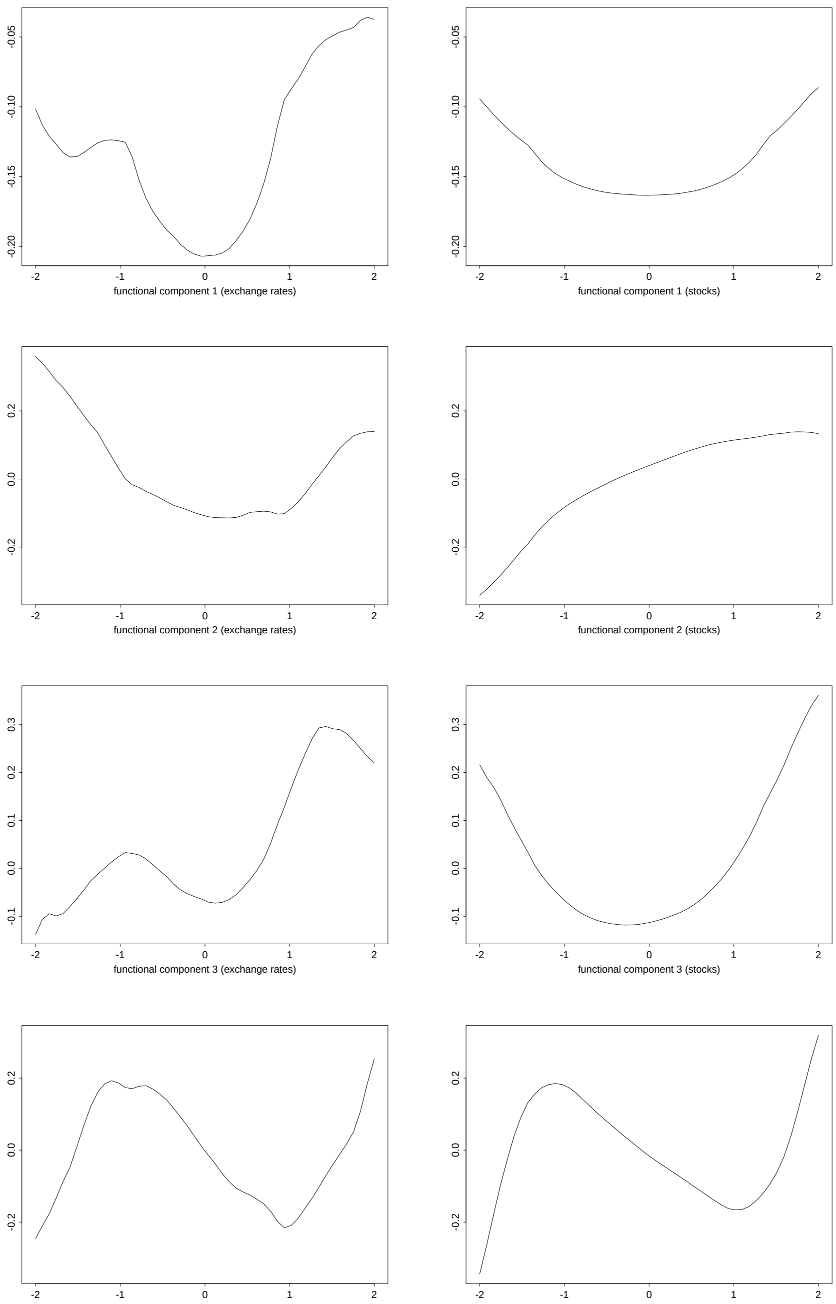


Figure 7: First four functional components of $\log \sigma^2(x) + 1.272$: exchange rates (left column), stocks (right column)

Appendix A

Assume that the relationship between response values y_t and predictors x_t is given by

$$y_t = f(x_t) + \epsilon_t, \quad t = 1, ..., T, \tag{20}$$

where the ϵ_t are assumed to be independent random variables with zero mean and variance equal to 1. Putting equation (20) into vector form we have

$$y = w + \epsilon. \tag{21}$$

The goal is then to find the best approximation of w in a pre-specified set of admissable estimates W, i.e. to solve the problem

$$\min_{\hat{w} \in W} \mathrm{MSE}(\hat{w})$$
$$\mathrm{MSE} := \mathrm{E}\frac{1}{T}\|w - \hat{w}\|_2^2.$$

Frequently, estimators of $\hat{w}$ are linear functions of y:

$$\hat{w} = S_h y, \tag{22}$$

where the "smoother matrix" depends on a parameter h. For example estimation may be based on least squares approximation on functions $g_1, ..., g_h$ of x_t. In this case an estimate of w is obtained by

$$\hat{w} = S_h y := V_h (V_h^{\mathrm{T}} V_h)^{-1} V_h^{\mathrm{T}} y,$$

where

$$V_h = [g_r(x_t)]_{t,r}.$$

Here the parameter h denotes the number of included functions in V_h.

If, in fact, the estimator is linear as in (22) then the optimal parameter h may be chosen such as to minimize Mallows' C_L:

$$C_L(h) = \frac{1}{T}\|y - S_h y\|_2^2 + \frac{2}{T}\mathrm{tr}(S_h).$$

This is motivated by the fact that

$$\mathrm{E}[C_L(h) - 1] = \mathrm{MSE}(h). \tag{23}$$

In other words for fixed h C_L provides an unbiased estimator of the risk regarding quadratic loss.

Proof of (23):

$$\begin{aligned}
\mathrm{E}[C_L(h)-1] &= \mathrm{E}\frac{1}{T}\|w+\epsilon-S_h y\|_2^2+\frac{2}{T}\mathrm{tr}(S_h)-1\\
&= \mathrm{MSE}(h)+\frac{2}{T}\mathrm{E}[\epsilon^{\mathrm{T}}(w-S_h y)]+\frac{2}{T}\mathrm{tr}(S_h)\\
&= \mathrm{MSE}(h)-\frac{2}{T}\mathrm{E}[\epsilon^{\mathrm{T}}S_h\epsilon]+2\mathrm{tr}(S_h)\\
&= \mathrm{MSE}(h)
\end{aligned}$$

Note: If S_h is a projection matrix of rank h then $\mathrm{tr}(S_h)=h$

Appendix B

Let $X_1,...,X_N$ be vectors in $\mathbb{R}^m$. Suppose we want to solve the problem

$$\min_{p_1,\dots,p_L\in\mathbb{R}^m}\sum_{i=1}^{N}\min_{a_{i1},\dots,a_{iL}\in\mathbb{R}}\left\|X_i-\sum_{l=1}^{L}a_{il}p_l\right\|_2^2. \tag{24}$$

for $L\le N$. We shall write (24) more conveniently in matrix notation which is

$$\min_{p_1,\dots,p_L\in\mathbb{R}^m}\min_{a_1,\dots,a_L\in\mathbb{R}^N}\left\|X-\sum_{l=1}^{L}p_l a_l^{\mathrm{T}}\right\|_2^2, \tag{25}$$

where X is the matrix $(X_1,...,X_N)$ and $\|M\|_2^2$ means the sum of the squared entries of the matrix M. As a solution of the problem is obviously not uniquely given, we may assume without loss of generality that the set of vectors p_l is orthonormal. Before giving a solution of (25), we prove the following lemma.

Lemma 1 *Let $\{p_1,...,p_L\}$ be a set of orthonormal vectors in $\mathbb{R}^m$ and X be a symmetrical $m\times m$ matrix with eigenvalues $\lambda_1\ge...\ge\lambda_m$ and corresponding orthonormal eigenvectors $v_1,...,v_m$. Then*

$$\sup_{p_1,\dots,p_L}\sum_{l=1}^{L}p_l^{\mathrm{T}}Xp_l=\sum_{l=1}^{L}v_l^{\mathrm{T}}Xv_l=\sum_{l=1}^{L}\lambda_l.$$

Proof: We can set $p_l = \sum_{j=1}^m c_{lj} v_j$. Then

$$\sum_{l=1}^{L} p_l^{\mathrm{T}} X p_l = \lambda_1 (c_{11}^2 + ... + c_{L1}^2) + ... + \lambda_m (c_{1m}^2 + ... + c_{Lm}^2). \tag{26}$$

The coefficients of the λ_i are all smaller than 1: their sum is L. To maximize the right hand side of (26) we choose the coefficients of $\lambda_1, ..., \lambda_L$ to be 1 the others to be 0, i.e. we choose $p_l = v_l$, $l = 1, ..., L$.

$\square$

We return again to problem (25). Note that the sum of squares of all elements a_{ij} of a $m \times m$ matrix A can be written as the trace of AA^{T} so that solving problem (25) is equivalent to minimizing

$$\begin{aligned} &\mathrm{tr}(X - \sum_{l=1}^{L} p_l a_l^{\mathrm{T}})^{\mathrm{T}} (X - \sum_{l=1}^{L} p_l a_l{}^{\mathrm{T}}) = \\ &\quad \mathrm{tr}(X^{\mathrm{T}} X) - \sum_{l=1}^{L} \mathrm{tr}(a_l p_l^{\mathrm{T}} X) - \sum_{l=1}^{L} \mathrm{tr}(X^{\mathrm{T}} p_l a_l^{\mathrm{T}}) + \sum_{k=1}^{L} \sum_{l=1}^{L} \mathrm{tr}(a_l p_l^{\mathrm{T}} p_k a_k^{\mathrm{T}}). \end{aligned} \tag{27}$$

Now

$$\mathrm{tr}(X^{\mathrm{T}} p_l a_l^{\mathrm{T}}) = \mathrm{tr}(a_l p_l^{\mathrm{T}} X) = \mathrm{tr}(p_l^{\mathrm{T}} X a_l) = p_l^{\mathrm{T}} X a_l,$$

where the second equality follows from

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$$

whenever the products of the matrices A, B, C are defined. By the same argument and the orthonormality conditions on the p_l one can show

$$\mathrm{tr}(a_k p_k^{\mathrm{T}} p_l a_l^{\mathrm{T}}) = \mathrm{tr}(p_l^{\mathrm{T}} p_k a_l^{\mathrm{T}} a_k) = p_l^{\mathrm{T}} p_k a_l^{\mathrm{T}} a_k = \delta_{lk} a_l^{\mathrm{T}} a_k.$$

Hence, the right hand side of (27) becomes

$$\mathrm{tr}(X^{\mathrm{T}} X) - 2 \sum_{l=1}^{L} p_l^{\mathrm{T}} X a_l + \sum_{l=1}^{L} a_l^{\mathrm{T}} a_l. \tag{28}$$

We equate the derivative of (28) with respect to a_l with 0, which yields

$$a_l = X^{\mathrm{T}} p_l.$$

When this is inserted into (28), one gets

$$\mathrm{tr}(X^{\mathrm{T}}X) - \sum_{l=1}^{L} p_l^{\mathrm{T}} X X^{\mathrm{T}} p_l = \mathrm{tr}(XX^{\mathrm{T}}) - \sum_{l=1}^{L} p_l^{\mathrm{T}} X X^{\mathrm{T}} p_l.$$

By means of lemma (1) the second term is minimized by setting p_l equal to the l'th eigenvector of XX^{T}. Thus we get

$$\begin{aligned}\min_{p_1,\dots,p_L \in \mathbb{R}^m} \sum_{i=1}^{N} \min_{a_{i1},\dots,a_{iL}} \left\| X_i - \sum_{l=1}^{L} a_{il} p_l \right\|_2^2 &= \sum_{i=1}^{N} \min_{a_{i1},\dots,a_{iL}} \left\| X_i - \sum_{l=1}^{L} a_{il} \gamma_l(XX^{\mathrm{T}}) \right\|_2^2 \\ &= \sum_{l=L+1}^{N} \lambda_l(XX^{\mathrm{T}}),\end{aligned}$$

where $\lambda_1(XX^{\mathrm{T}}) \geq \dots \geq \lambda_N(XX^{\mathrm{T}})$ are the N largest eigenvalues of XX^{T} (the others being 0) and $\gamma_l(XX^{\mathrm{T}})$ the corresponding eigenvectors.

Appendix C

List of exchange rates:

ATS	Austrian Schillings
AUD	Australian Dollars
BEF	Belgian Francs
CHF	Swiss Francs
DEM	German Marks
DKK	Danish Kroner
ESP	Spanish Pesetas
FIM	Finnish Markka
FRF	French Francs
GBP	British Pounds
HKD	Hong Kong Dollars
IEP	Irish Punt
ITL	Italian Lira
JPY	Japanese Yen
NLG	Dutch Guilders
NOK	Norwegian Kroner
NZD	New Zealand Dollars
SEK	Swedish Krona
USD	American Dollars

References

GASSER, T., AND A. KNEIP (1992): "Searching for structure in curve samples," Discussion Paper A-368, Sonderforschungsbereich, University of Bonn.

GYÖRFI, L.AND HÄRDLE, W., P. SARDA, AND P. VIEU (1989): *Nonparametric Curve Estimation from Time Series*, no. 60 in Lecture Notes in Statistics. Springer Verlag.

HÄRDLE, W., AND A. TSYBAKOV (1995): *Local Polynomial Estimators of the Volatility Function in Nonparametric Autoregression.*

HEID, F. (1996): "Non-Parametric Volatility Estimation of Exchange Rates and Stock Prices," Discussion Paper A-533, Sonderforschungsbereich 303, University of Bonn.

KNEIP, A. (1993): "Heterogeneity of Demand Behaviour and the Space of Engel Curves," .

KNEIP, A., AND T. GASSER (1992): "Statistical tools to analyze data representing a sample of curves," *Annals of Statistics*, 20, 1266–1305.

MALLOWS, C. (1973): "Some comments on C_p," *Technometrics*, 15, 661–675.

RAO, C. (1958): "Some statistical methods for the comparison of growth curves," *Biometrics*, 14, 1–17.