

Volatility and risk estimation with linear and nonlinear methods based on high frequency data

Working Paper**Author(s):**

Dettling, Marcel; Bühlmann, Peter

Publication date:

2001

Permanent link:

<https://doi.org/10.3929/ethz-a-004218133>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

Research Report / Seminar für Statistik, Eidgenössische Technische Hochschule (ETH) 96

VOLATILITY AND RISK ESTIMATION WITH LINEAR AND
NONLINEAR METHODS BASED ON HIGH FREQUENCY
DATA

by

Marcel Dettling

and

Peter Bühlmann¹

Research Report No. 96
February 5, 2001

Seminar für Statistik

Eidgenössische Technische Hochschule (ETH)

CH-8092 Zürich

Switzerland

¹Corresponding author. Tel.: +41-1-632-7338; fax: +41-1-632-1228; e-mail: buhlmann@stat.math.ethz.ch

VOLATILITY AND RISK ESTIMATION WITH LINEAR AND NONLINEAR METHODS BASED ON HIGH FREQUENCY DATA

Marcel Dettling

and

Peter Bühlmann[†]

Seminar für Statistik
ETH Zentrum
CH-8092 Zürich, Switzerland

February 5, 2001

Abstract

Accurate volatility predictions are crucial for the successful implementation of risk management. The use of high frequency data approximately renders volatility from a latent to an observable quantity, and opens new directions to forecast future volatilities. Our goals in this paper are to find a general powerful forecasting procedure for volatilities based on high frequency data, to evaluate the predictive potential of volatility forecasts for the true latent volatility, and to analyze the impact of more reliable volatility predictions on the quality of two widely used risk measures. For that purpose, we explore the performance of various models and modern prediction tools.

JEL Classification: C22; C52; G10

Keywords: Forecasting, High-Frequency-Data, Predictive Potential, Risk Measures, Volatility

[†]Corresponding author. Tel.: +41-1-632-7338; fax: +41-1-632-1228; e-mail: buhlmann@stat.math.ethz.ch

1 Introduction

Financial market volatility is the key ingredient in the theory of risk management, asset pricing and asset allocation. Thus, precisely defined measures and accurate forecasts of the latent volatility factor are crucial for the successful implementation of these techniques.

The volatility σ_t for the time period of one day is defined as $\sigma_t = \sqrt{\text{Var}(R_t | \mathcal{F}_{t-1})}$, where R_t is the daily return, and \mathcal{F}_{t-1} is the information available on the return process up to time $t - 1$. Since σ_t is not observable, it has to be estimated. This can be done by the squared daily return R_t^2 , which constitutes an unbiased but very noisy estimator of the squared volatility σ_t^2 . An alternative approach is to use parametric volatility models such as ARCH or GARCH, or to extract the implied volatility out of the Black-Scholes formula, but it has been found that all these approaches lead to volatility estimates which are highly biased.

The idea to use high frequency data for more reliable volatility estimation turned up more than twenty years ago. In order not to lose all the information about the price process in between, Officer (1973) computed annual volatilities from monthly returns, whereas Merton (1980) used daily returns for the measurement of monthly volatilities. Yet only recently the idea of using high frequency intradaily data for estimating daily volatility came up. Schwert (1998) was working with 15-minute returns, while Taylor and Xu (1997) as well as Andersen et al. (1998, 1999a) apply 5-minute returns to compute daily exchange rate volatilities.

Andersen et al. (1998, 1999a) found daily volatility estimates computed from aggregated high frequency returns to be so accurate that they introduced the term *realized volatility*. Those realized volatilities $\sigma_{t,RV}$, being observable approximations for the latent volatility σ_t , render new possibilities to forecast future volatilities. Our goals in this paper are

- a) to find a general powerful forecasting procedure for volatilities based on high frequency data, and
- b) to evaluate the quality of the volatility forecasts for the true latent volatilities σ_t , and not primarily for future realized volatilities $\sigma_{t,RV}$.

Regarding issue a), we consider a variety of modern techniques for predicting realized volatilities. But surprisingly, detrending the log-transformed realized volatility $\log(\sigma_{t,RV})$ by *exponential smoothing*, and modelling the remaining stationary time series with the simple *AR(p)-method* has empirically proven to have the best predictive potential. We could not exploit any advantages by using nonlinear modelling for the detrended log-realized volatilities. An alternative approach using fractional differencing for long memory processes and subsequent AR-modelling yields similar predictions.

Issue b) deals with the problem of judging volatility forecasts. Assuming a Brownian motion model for the continuous time prices, the realized volatilities $\sigma_{t,RV}$ converge to the corresponding true latent volatilities σ_t as sampling frequency grows, see for example Andersen et al. (1998). Hence it is plausible to evaluate a forecast $\hat{\sigma}_t$ by comparing it with $\sigma_{t,RV}$. This approach has been pursued by Andersen et al. (1998), but it should not be the only way. We also propose evaluating volatility forecasts $\hat{\sigma}_t$ by assuming stationarity and conditional mean zero of returns only: we then calibrate the squared prediction $\hat{\sigma}_t^2$ against the squared returns R_t^2 which are very noisy estimates of the true squared volatility σ_t^2 . But mean-type test statistics are able to reduce noise variance due to averaging, and we could exploit significant advantages of our most powerful forecasting procedure with high frequency data compared to a GARCH model with daily returns.

An important application of volatility predictions is in risk management, i.e. the computation of risk measures like *value at risk* or *conditional expected shortfall*. We explore the advantage of more reliable volatility forecasts for improving the quality of those two risk measures in comparison to the GARCH benchmark model.

The plan for the rest of the paper is as follows: In section 2 we explain how volatilities on a high frequency basis are computed, and why this is sensible. Section 3 presents the most powerful method for volatility forecasting. The design and the results of our empirical study are presented in section 4, whereas section 5 discusses the impact of the volatility predictions on the risk measures. Finally, we conclude with a brief summary in section 6.

2 High Frequency Volatility

Let $P_{(m),t}$, where $t = \frac{1}{m}, \frac{2}{m}, \dots$, denote the price of a financial asset, which is recorded equidistantly m times per day. The time series of high frequency returns with m observations per day, corresponding to a return horizon of $\Delta t = \frac{1}{m}$, is then defined by the logarithmic difference of the price process:

$$R_{(m),t} = \log\left(P_{(m),t}\right) - \log\left(P_{(m),t-\frac{1}{m}}\right), \quad t = \frac{1}{m}, \frac{2}{m}, \dots \quad (1)$$

Throughout the whole paper we assume that the conditional expected returns are equal to zero, i.e. $E[R_{(m),t} | \mathcal{F}_{t-\frac{1}{m}}] = 0$, where \mathcal{F}_s stands for the information about the return process up to time s . In risk management, the main focus is on the daily return¹ $R_{(1),t}$ which equals $R_t = \sum_{i=0}^{m-1} R_{(m),t-\frac{i}{m}}$. Hence, high frequency returns do not provide any new information about the daily return.

The reason to use high frequency data is to improve the estimate of the latent volatility $\sigma_t = \sqrt{\text{Var}(R_t | \mathcal{F}_{t-1})}$. For a short time horizon Δt , the latent volatility can be estimated on the basis of the squared return, i.e. $\hat{\sigma}_{(m),t} = \sqrt{R_{(m),t}^2}$. Again, in risk management the variable of interest is the daily volatility σ_t , and not the one for a short time interval. An estimate of the daily volatility σ_t is obtained by summing up the squared high frequency returns:

$$\hat{\sigma}_t = \sigma_{t;RV} = \sqrt{\sum_{i=0}^{m-1} R_{(m),t-\frac{i}{m}}^2} \quad (2)$$

Under the assumption of a continuous-time diffusion for the logarithmic price process and sufficient regularity conditions, it can be shown by the theory of quadratic variation that the sum of intraday squared returns in (2) converges almost surely to the true squared volatility σ_t^2 as the sampling frequency m goes to infinity. Hence, we can get an asymptotically error-free estimate of the latent volatility factor by simply increasing the sampling frequency and, therefore designate $\hat{\sigma}_t$ as *realized volatility*, denoted by $\sigma_{t;RV}$.

In practice, due to market microstructure effects, logarithmic asset prices do not evolve according to a diffusion process. As time intervals become shorter and shorter, especially the assumption of independent increments becomes less and less realistic. Thus, the theoretical result mentioned above will no longer hold, and the volatility estimator made up of high frequency returns is no longer unbiased and consistent, see Corsi et al. (2001).

Thus, we are in a trade-off situation between stochastic error and bias. It has been found that the return interval for which the bias is not significant is at the level of 2-3

¹In order to simplify the notation, we will write R_t for $R_{(1),t}$ and σ_t for $\sigma_{(1),t}$.

hours even for the most liquid assets. Furthermore, the interval for which the bias has no influence changes considerably from asset to asset. It is nowadays common to use realized volatilities based on a return interval of 5 minutes, which is equivalent to $m = 288$ measurements per day. Thereafter, a bias correction is carried out: how to do this is beyond the scope of this paper, and we refer to Corsi et al. (2001) for further details.

Olsen&Associates kindly provided three datasets containing daily observations of high frequency realized volatilities for the exchange rates between US Dollar and Swiss Franc (US\$/SFr), between US Dollar and Japanese Yen (US\$/JPY), and between GB Pound and US Dollar (GB£/US\$) from January 1, 1990 through December 31, 1999. The realized volatilities were computed as in (2) with measurement intervals Δt of 5 minutes and then treated with a bias correction. This results in a total of 2600 observations per dataset.

3 Volatility Forecasting with a Powerful Method

Plotting the realized volatilities $\sigma_{t,RV}$ against time, we recognize in figure 1 that its fluctuations show substantial persistence and all the spikes point in the same direction. The latter fact is an evidence of skewness in the data, which is confirmed furtherly by an analysis of the empirical distribution.

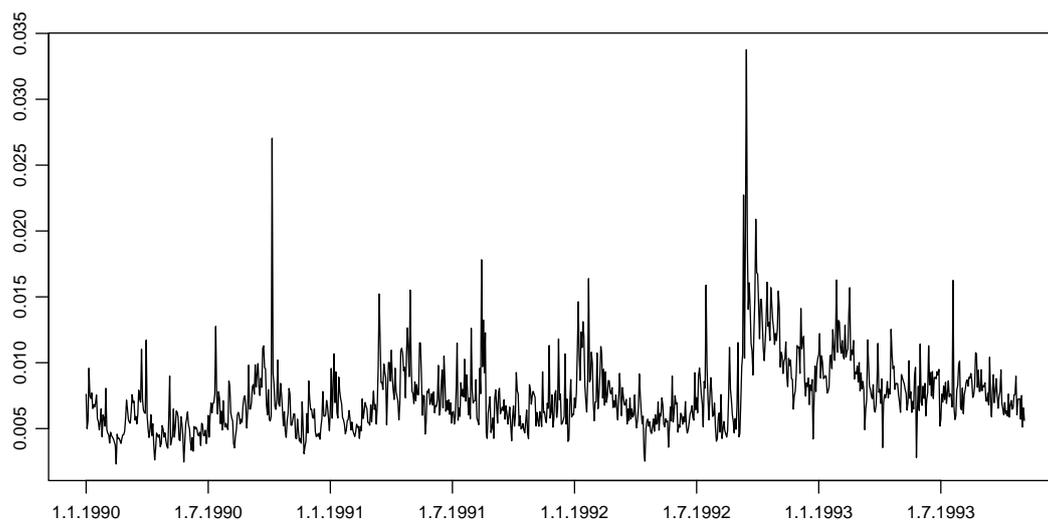


Figure 1: The realized volatility $(\sigma_{t,RV})_{t \in \{1, \dots, 1000\}}$ for the exchange rate GB£/US\$.

Since modelling of time series is easier under the assumption of normally distributed data, we log-transform the realized volatilities. This results in a symmetric and approximately normal distribution of the transformed data. But still, the transformed time series in figure 2 exhibits clearly defined periods of high and low observations. Thus, not surprisingly, the empirical autocorrelation function shows a slow hyperbolic decay. This is often interpreted as an evidence of long-range dependence, for example in Andersen et al. (1999b). But alternatively, we consider the possibility of a trend, and therefore apply the classical decomposition model to the transformed time series,

$$\log(\sigma_{t,RV}) = X_t = m_t + Z_t, \quad (3)$$

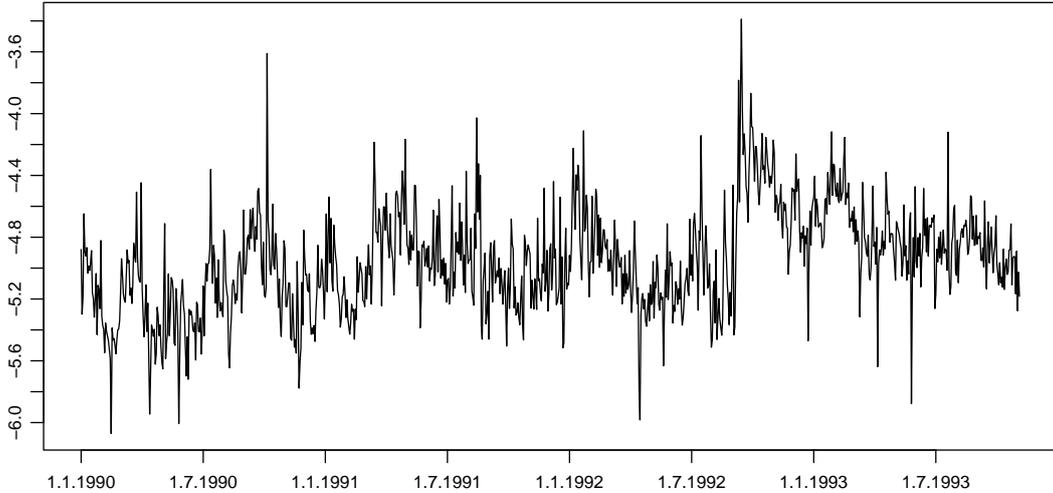


Figure 2: The log-transformed series $(X_t)_{t \in \{1, \dots, 1000\}}$ for the exchange rate GB£/US\$.

where $m_t = E[X_t]$ and $E[Z_t] = 0$. As pointed out more detailed in section 4.2.7, it is mainly a question of taste whether a long memory or a deterministic trend model is preferred. We will see in section 4.2 that long memory models yield similar results.

Our aim is to estimate and extract the deterministic component m_t in the hope that the residuals Z_t will turn out to be a stationary random process. We will determine the trend component with an asymmetric linear filter,

$$\hat{m}_t = \sum_{j=0}^{t-1} a_j X_{t-j}. \quad (4)$$

The filter is depending on a set of weights $\{a_j\}$. We choose them with an exponential decay and standardize so that $\sum a_j = 1$,

$$a_j = \frac{\alpha^j}{\sum_{k=0}^{t-1} \alpha^k}, \text{ where } \alpha \in (0, 1) \text{ and } j = 0, 1, 2, \dots, t-1. \quad (5)$$

The technique specified above is known as exponential smoothing. The parameter α is a tuning element. Empirically, we found a good performance with $\alpha = 0.05$, but more formal data-driven choices using outsample-optimization or cross-validation for dependent data would be possible. Subtraction of \hat{m}_t then leaves us with a residual process $\{Z_t = X_t - \hat{m}_t\}$. An inspection by descriptive techniques shows that these residuals have a symmetric and approximately normal distribution. Furthermore, they seem to be stationary. But most importantly, the sample autocorrelation function is decaying quickly, which indicates that the trend has been successfully removed.

A simple parametric stochastic model for the detrended time series $\{Z_t\}$ is a linear autoregressive model of order p , abbreviated by $AR(p)$. On the other hand nonlinear state of the art forecasting procedures such as *projection pursuit autoregression* and *dynamic combination of models* may represent the residual process more accurately, and provide

better predictions. We will compute model based forecasts $\{\widehat{Z}_t\}$, and use them in conjunction with the predictive estimate \widehat{m}_t of the deterministic trend component for the prediction of the log-realized volatilities. Since we assume the trend to vary only slowly, the predicted value is given by

$$\widehat{X}_{t+1} = \widehat{m}_{t+1} + \widehat{Z}_{t+1} = \widehat{m}_t + \widehat{Z}_{t+1}. \quad (6)$$

We will see in section 4.2 that an AR model for $\{Z_t\}$ yields very competitive forecasting results compared to more complex and nonlinear methods.

4 Predictive Potential

4.1 Measuring Performance

The time series methods we will apply have unequal complexity. For a fair comparison, it is therefore necessary to evaluate the predictive potential out-of-sample. It is important to keep in mind that more complex forecasting procedures usually show a better performance in-sample, but they must not minimize the out-of-sample forecast errors. Thus, we divide each of the three datasets into a learning set \mathcal{L} containing the first $n = 1000$ consecutive observations, and a test set \mathcal{T} containing the remaining $m = 1600$ observations. Once the models have been fitted on the learning set \mathcal{L} , we explore their performance on the test set \mathcal{T} .

The standard approach for judging the predictive potential of any forecasting procedure is to compare predictions and subsequent realizations. For the volatility σ_t which is a latent variable, the subsequent realizations are unknown. If we utilized the noisy estimate R_t^2 for the squared volatility σ_t^2 , we could look at the magnitude of the difference $R_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2$, where $\widehat{\sigma}_{t;\mathcal{M}}^2$ is the volatility prediction obtained with a particular forecasting model \mathcal{M} . If $R_t = \sigma_t \varepsilon_t$ with ε_t iid mean zero innovations, then

$$R_t^2 = \sigma_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2 + \eta_t, \quad (7)$$

where $\eta_t = \sigma_t^2(\varepsilon_t^2 - 1)$, satisfying $E[\eta_t] = 0$. This regression type representation has typically extremely low signal to noise ratio, and a measure like

$$E[(R_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2)^2] = E[(\sigma_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2)^2] + E[\eta_t^2] \quad (8)$$

is dominated by the large nuisance term $E[\eta_t^2]$. This implies that the values of the performance measure in (8) are very similar for a large collection of forecasting models \mathcal{M} , and hence not suitable for discrimination. One way to improve the evaluation method in (8) is to substitute the subsequent realizations σ_t with the realized volatilities $\sigma_{t;RV}$, which lowers the magnitude of the nuisance term $E[\eta_t^2]$. This approach was pursued by Andersen et al. (1998), but it may be biased since it only measures the predictive potential for the realized volatility $\sigma_{t;RV}$, and not for the true latent volatility σ_t which is our objective. However, to gain a first impression of the performance of a particular forecasting model \mathcal{M} , we will compute the residual sum of squares,

$$RSS_\sigma(\mathcal{M}) = \sum_{t \in \mathcal{T}} (\sigma_{t;RV} - \widehat{\sigma}_{t;\mathcal{M}})^2. \quad (9)$$

This criterion is not very robust and easily influenced by outliers. We will therefore also compute the residual sum of absolute differences which is less sensitive to outliers:

$$RSAD_\sigma(\mathcal{M}) = \sum_{t \in \mathcal{T}} |\sigma_{t;RV} - \widehat{\sigma}_{t;\mathcal{M}}|. \quad (10)$$

4.2 Comparison with Other Methods

This section contains an empirical study, where we explore the performance of different forecasting procedures for the volatility. Our aim is to find the forecasting method, which yields the best results in connection with exponential smoothing, and to compare it against other prediction models for the volatility.

A descriptive analysis of the residual process $\{Z_t\}$ obtained after exponential smoothing shows that it is sensible to model it with the simple linear autoregressive process. The order p of the model is chosen by Akaike's information criterion, and forecasts are computed in the usual manner. We then tried to outperform the combination of exponential smoothing and AR modelling by two state of the art prediction techniques which are presented in the next two sections.

4.2.1 Projection Pursuit

The projection pursuit autoregression model for the stationary mean zero process $\{Z_t\}$ is given by

$$\begin{aligned} Z_t &= \mu_t + W_t, \\ \mu_t &= \mathbb{E}[Z_t | Z_{t-1}, Z_{t-2}, \dots] = \sum_{k=1}^m f_k(\alpha_k + \sum_{j=1}^p \phi_{k,j} Z_{t-j}), \end{aligned} \quad (11)$$

where $f_k(\cdot)$ are unknown ridge functions, $\alpha_k, \phi_{k,j}$ are unknown parameters, and $\{W_t\}_t$ a sequence of iid innovations such that W_t is independent of $\{Z_s; s < t\}$. This is a surprisingly general class of models, as it can approximate arbitrary continuous conditional mean functions over compacta and it also encompasses feed-forward neural networks, see Ripley (1996). We consider here mainly $m = 1$ with one ridge function, which is sometimes referred as a single-index model: already this and even more complex structures did not prove to be useful for our application. Note that for $m = 1$ and $f_1(\cdot)$ linear, we obtain the linear AR(p) model. The projection pursuit estimate of μ_t is also the best (with respect to mean square error) predictor for the random variable Z_t . The number of lagged variables p has been optimized with respect to outsample performance. Although this is not a rule on the learning set \mathcal{L} only, it yields the benchmark with this method; we will see later that this benchmark can be achieved by the much simpler AR predictions.

4.2.2 Dynamic Combination of Models

Another powerful method for predicting conditional means and future observations is given by dynamic combination of models, see Bühlmann and Ferrari (2001). The dynamic combination of AR(p) models for the stationary mean zero process $\{Z_t\}_t$ is given by

$$\begin{aligned} Z_t &= \mu_t + W_t, \\ \mu_t &= \mathbb{E}[Z_t | Z_{t-1}, Z_{t-2}, \dots] = \sum_{k=1}^m P_{t;k}(\alpha_k + \sum_{j=1}^p \phi_{k,j} Z_{t-j}), \end{aligned} \quad (12)$$

where $P_{t;k}$ is the conditional probability that Z_t falls into a quantization interval I_k ; here $\cup_{k=1}^m I_k = \mathbb{R}$, and the intervals I_k are pairwise disjoint. The conditional probabilities are modelled as variable length Markov chains

$$P_{t;k} = \mathbb{P}[Z_t \in I_k | Z_{t-1}, Z_{t-2}, \dots, Z_{t-\ell_t}], \quad (13)$$

where $\ell_t = \ell(Z_{t-1}, Z_{t-2}, \dots)$ is the *variable* length of the memory, depending on the past values Z_{t-1}, Z_{t-2}, \dots . Thus, the $P_{t;k}$ can be viewed as mixture weights, dynamically changing over time, for different AR models. For prediction, such dynamic combination of AR models seems to be as good or even slightly better than projection pursuit autoregression, see Bühlmann and Ferrari (2001). Here, we choose the simplest version with $k = 2$ and $I_1 = (\infty, 0], I_2 = (0, \infty)$, since more complex structures did not show any better performance in our application. As for projection pursuit autoregression, the number of lagged variables p has been optimized with respect to outsample performance, yielding the benchmark for this procedure.

Both projection pursuit autoregression and dynamic combination of models are generally among the best nonlinear prediction procedures for conditional means and future observations in stationary time series.

4.2.3 Long Memory Models

A totally different interpretation of the slow hyperbolic decay in the autocorrelation function of the log-realized volatilities is given by a long-memory characteristic rather than a deterministic trend. Such processes can be modelled by a fractionally integrated ARFIMA(p, d, q), where the parameter d is not an integer. We choose the model order $p = 5$ and $q = 0$ according to the p given by Akaike's information criterion in the exponential smoothing setting. The model coefficients as well as the degree of fractional integration are then simultaneously estimated on the learning set \mathcal{L} .

4.2.4 Benchmark Methods

The popular GARCH(1,1) is the standard model for volatility forecasting without high frequency data. It therefore serves as a benchmark for a comparison with the forecasting procedures mentioned above. Since this volatility model is not based on realized volatilities, we expect an inferior performance with respect to the measures in (9) or (10).

In order to check whether detrending the time series by exponential smoothing improves the accuracy of the predictions, we also forecast the log-realized volatilities X_t with a linear AR model, where the order p is again chosen by Akaike's information criterion. The use of an AR model does not correspond well with the slowly decaying autocorrelation function, and therefore we expect a worse performance. As an extension of the autoregressive approach, we also consider projection pursuit directly applied on the log-realized volatility X_t , see section 4.2.1.

4.2.5 Deviance Measures

In the following tables 1-3 the deviance measures for each of the three exchange rates are given². The second column contains the percentual deterioration against the smallest value.

We observe that the magnitude of the deviance measures is very different across the three datasets. This is due to the different signal to noise ratio of the three exchange rates. But qualitatively, the results given in the tables are consistent both across the two

²Abbreviations in tables 1-3:

| | |
|---------------------|--|
| AR(p) | Linear autoregressive model of order p . |
| PPR(m, p) | Projection pursuit autoregression of order (m, p) , see formula (11). |
| DCM(m, p) | Dynamic combination of m AR(p) models, see formula (12). |
| ARFIMA(p, d, q) | Fractionally integrated ARMA(p, q) with degree of fractional integration d . |

| <i>Prediction Method</i> | RSS_σ | <i>in %</i> | $RSAD_\sigma$ | <i>in %</i> |
|----------------------------------|-----------------------|-------------|------------------------|-------------|
| Exponential smoothing & AR(4) | $35.65 \cdot 10^{-4}$ | + 0.65% | $157.62 \cdot 10^{-2}$ | |
| Exponential smoothing & PPR(1,5) | $36.18 \cdot 10^{-4}$ | + 2.15% | $158.52 \cdot 10^{-2}$ | + 0.57% |
| Exponential smoothing & DCM(2,4) | $35.65 \cdot 10^{-4}$ | + 0.65% | $158.10 \cdot 10^{-2}$ | + 0.30% |
| ARFIMA(5,0.27,0) | $35.42 \cdot 10^{-4}$ | | $159.22 \cdot 10^{-2}$ | + 1.02% |
| PPR(1,18) | $35.74 \cdot 10^{-4}$ | + 0.90% | $167.57 \cdot 10^{-2}$ | + 6.31% |
| AR(9) | $36.34 \cdot 10^{-4}$ | + 2.60% | $163.28 \cdot 10^{-2}$ | + 3.59% |
| GARCH(1,1) | $50.90 \cdot 10^{-4}$ | + 43.70% | $219.09 \cdot 10^{-2}$ | + 39.00% |

Table 1: Deviance measures for the exchange rate US\$/SFr.

| <i>Prediction Method</i> | RSS_σ | <i>in %</i> | $RSAD_\sigma$ | <i>in %</i> |
|----------------------------------|-----------------------|-------------|------------------------|-------------|
| Exponential smoothing & AR(5) | $18.31 \cdot 10^{-4}$ | + 0.38% | $104.55 \cdot 10^{-2}$ | |
| Exponential smoothing & PPR(1,3) | $18.60 \cdot 10^{-4}$ | + 1.97% | $105.70 \cdot 10^{-2}$ | + 1.10% |
| Exponential smoothing & DCM(2,5) | $18.80 \cdot 10^{-4}$ | + 3.07% | $106.63 \cdot 10^{-2}$ | + 1.99% |
| ARFIMA(5,0.47,0) | $18.24 \cdot 10^{-4}$ | | $106.20 \cdot 10^{-2}$ | + 1.58% |
| PPR(1,11) | $18.72 \cdot 10^{-4}$ | + 2.63% | $111.54 \cdot 10^{-2}$ | + 6.69% |
| AR(9) | $18.80 \cdot 10^{-4}$ | + 3.07% | $112.85 \cdot 10^{-2}$ | + 7.94% |
| GARCH(1,1) | $45.87 \cdot 10^{-4}$ | + 151.48% | $229.12 \cdot 10^{-2}$ | + 119.50% |

Table 2: Deviance measures for the exchange rate GB£/US\$.

| <i>Prediction Method</i> | $RSAD$ | <i>in %</i> | $RSAD_\sigma$ | <i>in %</i> |
|----------------------------------|------------------------|-------------|------------------------|-------------|
| Exponential smoothing & AR(5) | $87.25 \cdot 10^{-4}$ | | $202.53 \cdot 10^{-2}$ | + 0.79% |
| Exponential smoothing & PPR(1,7) | $88.56 \cdot 10^{-4}$ | + 1.50% | $202.17 \cdot 10^{-2}$ | + 0.61% |
| Exponential smoothing & DCM(2,5) | $87.31 \cdot 10^{-4}$ | + 0.07% | $202.37 \cdot 10^{-2}$ | + 0.71% |
| ARFIMA(5,0.33,0) | $88.37 \cdot 10^{-4}$ | + 1.28% | $200.94 \cdot 10^{-2}$ | |
| PPR(1,15) | $100.46 \cdot 10^{-4}$ | + 15.14% | $207.47 \cdot 10^{-2}$ | + 3.25% |
| AR(5) | $90.17 \cdot 10^{-4}$ | + 3.35% | $202.90 \cdot 10^{-2}$ | + 0.98% |
| GARCH(1,1) | $109.69 \cdot 10^{-4}$ | + 25.72% | $244.02 \cdot 10^{-2}$ | + 21.44% |

Table 3: Deviance measures for the exchange rate US\$/JPY.

deviance measures and the three datasets. The three forecasting procedures applied to the exponentially smoothed log-realized volatilities X_t and the fractional ARIMA show approximately the same predictive potential. We will explore the practical significance of this observation in the sections 4.2.6 and 4.2.7.

The two benchmark methods, projection pursuit autoregression and AR(p) which are directly fitted on the log-realized volatilities X_t , can not keep up with the performance of the best four methods. This matches our expectations, since these two forecasting procedures are not suitable for time series which are nonstationary or long-range dependent. Finally, the GARCH(1,1) model exhibits a very bad performance. We thus exploit the celebrated fact that high frequency data may drastically improve volatility forecasts. In section 4.3, we take a fresh look to see whether this is also true when measuring the forecasting potential for the true volatility σ_t , rather than for the realized volatility $\sigma_{t,RV}$.

4.2.6 Linearity of the Detrended Process

The results given in the tables show that the forecasting procedures applied after exponential smoothing perform equally well. In five of six cases, the most accurate predictions were even produced by the simple linear AR(p) model. It is a surprising fact that a linear model yields so competitive results compared to the much more complex projection pursuit autoregression or the dynamically weighted combination of two AR models.

We consider this observation as an evidence of linearity for the detrended process $\{Z_t\}$. It would be tempting to test this hypothesis statistically, but Bickel and Bühlmann (1996) argue that such testing problems are in a sense ill-posed, saying that it is impossible (or extremely difficult) to detect nonlinearity from observed data. But for the task of prediction, we have collected empirical evidence that the linear AR(p) model performs among the best. In the light of the ill-posed testing problem, this is “all” we can do. A second argument for linearity of the residual process $\{Z_t\}$ is that the ridge function estimated by the projection pursuit autoregression is close to linear.

4.2.7 Trend or Long-Range dependence?

We observe that exponential smoothing combined with AR(p) and the ARFIMA model perform very similarly. Even though the purpose of these two forecasting procedures is fundamentally different, a closer look at the models’ properties can explain the observation.

Both, fractional differentiation and exponential smoothing correspond to a linear filter of the log-realized volatilities. Even its coefficients, although differently defined, are decaying in a similar manner. Moreover, both remaining time series are explained with an AR model. Consequently, both these forecasting procedures are almost equivalent from a mathematical point of view, and it is no longer surprising that their results are similar.

The question whether realized volatilities are long-range dependent, or whether they are nonstationary due to a deterministic trend, can not be answered on the basis of data, see Künsch (1986).

4.3 Testing for Better Prediction of the Latent Volatility σ_t

As we now have insights how forecasting realized volatility $\sigma_{t,RV}$ performs, our goal is to judge the predictive potential for the latent volatility σ_t . For the squared return R_t^2 , the equation

$$R_t^2 = \sigma_t^2 \varepsilon_t^2 = \sigma_t^2 + \underbrace{\sigma_t^2 (\varepsilon_t^2 - 1)}_{=\eta_t} = \sigma_t^2 + \eta_t \quad (14)$$

holds, where η_t are uncorrelated but not independent martingale differences with $E[\eta_t] = 0$. To evaluate the predictive potential of a particular forecasting procedure \mathcal{M} , consider the expected squared loss

$$P(\mathcal{M}) = E[(R_t^2 - \hat{\sigma}_{t;\mathcal{M}}^2)^2]. \quad (15)$$

Applying the notation of equation (14), we can convert the performance measure $P(\mathcal{M})$ into

$$P(\mathcal{M}) = E[(\sigma_t^2 - \hat{\sigma}_{t;\mathcal{M}}^2)^2] + E[\eta_t^2]. \quad (16)$$

The first term on the right side of equation (16) apparently measures the predictive potential of a volatility forecast $\hat{\sigma}_{t;\mathcal{M}}$, whereas the second term on the right side is a nuisance term. Unfortunately, its magnitude is so big that it renders $P(\mathcal{M})$ almost constant to $E[\eta_t^2]$ for a whole collection of different models, see Andersen et al. (1998). Hence, $P(\mathcal{M})$ is not suitable to quantify the performance of different models. But a very simple, yet fundamental idea helps: for the *difference* between two models \mathcal{M}_1 and \mathcal{M}_2 , the noise term $E[\eta_t^2]$ disappears. The random variables

$$U_t = (R_t^2 - \hat{\sigma}_{t;\mathcal{M}_1}^2)^2 - (R_t^2 - \hat{\sigma}_{t;\mathcal{M}_2}^2)^2 \quad \text{for all } t \in \mathcal{T} \quad (17)$$

can be easily recorded. The expected difference between two forecasting procedures \mathcal{M}_1 and \mathcal{M}_2 is then given by

$$\begin{aligned} E[U_t] &= E[(R_t^2 - \hat{\sigma}_{t;\mathcal{M}_1}^2)^2] - E[(R_t^2 - \hat{\sigma}_{t;\mathcal{M}_2}^2)^2] \\ &= E[(\sigma_t^2 - \hat{\sigma}_{t;\mathcal{M}_1}^2)^2] - E[(\sigma_t^2 - \hat{\sigma}_{t;\mathcal{M}_2}^2)^2] \\ &= P(\mathcal{M}_1) - P(\mathcal{M}_2). \end{aligned} \quad (18)$$

The price we have to pay when building differences is that a relative comparison between forecasting procedures is no longer possible. Yet it is possible to statistically test whether the difference between two forecasting procedures \mathcal{M}_1 and \mathcal{M}_2 is significantly different from zero. As a null hypothesis H_0 , we will use $E[U_t] = 0$, and test it against the alternative $E[U_t] \neq 0$.

We estimate the expectation of U_t by the mean $\hat{\mu}_m = \frac{1}{m} \sum_{t \in \mathcal{T}} U_t$, where $m = |\mathcal{T}| = 1600$, but we have to pay attention to the fact that U_t are not uncorrelated random variables. Under sufficient regularity conditions, the central limit theorem for dependent variables implies

$$\sqrt{m}(\hat{\mu}_m - E[U_t]) \xrightarrow{d} \mathcal{N}(0, \sigma_\infty^2) \quad (m \rightarrow \infty), \quad (19)$$

where $\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} \text{Cov}(U_0, U_k)$. Under H_0 , the asymptotic variance σ_∞^2 remains the same and hence

$$\sqrt{m}\hat{\mu}_m \xrightarrow{d} \mathcal{N}(0, \sigma_\infty^2) \quad (m \rightarrow \infty) \quad (20)$$

under H_0 . For practical use, the asymptotic variance σ_∞^2 has to be estimated. This problem is equivalent to estimating the spectral density of the variables $\{U_t\}$ at zero. We do this using a classical smoothed periodogram $\hat{f}_U(\lambda)$, but more formal approaches would be possible, see Bühlmann (1996). The estimate for the asymptotic variance is then

$$\hat{\sigma}_\infty^2 = 2\pi \cdot \hat{f}_U(0). \quad (21)$$

Assuming that asymptotic properties hold in our particular setting, we reject the null hypothesis on significance level α if

$$\left| \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} U_t \right| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\hat{\sigma}_\infty}{\sqrt{|\mathcal{T}|}}, \quad (22)$$

where $|\mathcal{T}|$ is the number of observations in the test set, i.e. $|\mathcal{T}| = 1600$. We use this test to check whether a volatility forecast based on high frequency data results in more reliable predictions for the latent volatility σ_t . Hence, we compare the GARCH(1,1) as the benchmark model \mathcal{M}_1 with the most powerful high frequency data method, exponential smoothing combined with an AR model as \mathcal{M}_2 . The test described above always points towards better accuracy of the high frequency data method with the following p-values.

| | US\$/SFr | GB£/US\$ | US\$/JPY |
|---------|----------|-------------------|----------|
| p-value | 0.018 | $2 \cdot 10^{-9}$ | 0.076 |

In all three datasets the use of high frequency data improves the predictive potential for the latent volatility σ_t . For the exchanges rates US\$/SFr and GB£/US\$ the most powerful high frequency data forecasting method even performs significantly better than the GARCH model. By comparing the forecasting procedures which are based on high frequency data, we could not observe any significant differences in the predictive potential.

4.4 The Best Forecasting Procedure

We conclude that the best forecasting procedure for the latent volatility σ_t is to detrend the log-transformed realized volatility $\log(\sigma_{t,RV})$ by exponential smoothing, and to predict the residual process with an AR(p) model. Our empirical findings show that this yields very good results. Furthermore, it is built with popular standard time series methods which are easy to implement and computationally cheap. In figure 3, a comparison between the ex-post realized volatility, the best forecast based on exponential smoothing combined with AR(p), and the GARCH(1,1) predictions is shown.

5 Risk Estimation

5.1 Introduction

The large increase in the number of traded assets in portfolios of most financial institutions has made the measurement of risk exposure a primary concern for regulators and for internal risk control. A key ingredient for the practically relevant risk measures is a volatility forecast. The aim of this section is an evaluation whether a more accurate volatility prediction based on high frequency data improves the quality of the risk measures. In particular, we compare the impact of the best forecasting procedure against the one of GARCH(1,1) predictions.

In practice, an important time horizon for the risk measures is one day. The most popular measure for the risk exposure within the next 24 hours is the conditional quantile q_t of the return R_t , also known as value at risk,

$$\mathcal{P}[R_t < q_t | \mathcal{F}_{t-1}] \leq \alpha. \quad (23)$$

An alternative risk measure, which is coherent, is the conditional expected shortfall S_t , see Artzner et al. (1999). It measures the magnitude of the expected loss given that the loss exceeds the value at risk q_t ,

$$S_t = E[R_t | R_t < q_t, \mathcal{F}_{t-1}] \quad (24)$$

In the next section we explore the estimation of these risk measures.

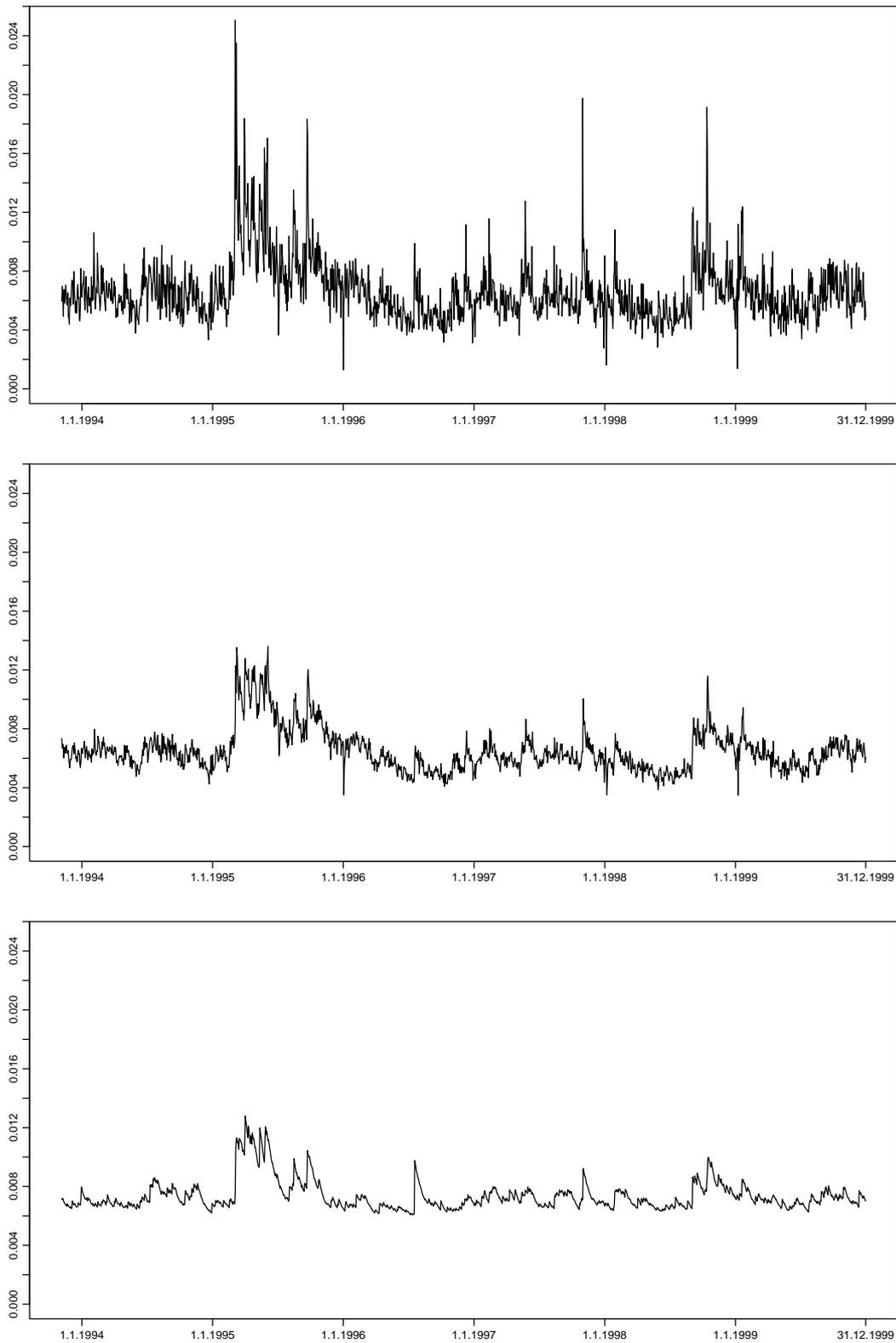


Figure 3: Comparison between ex-post realized volatility (top, not a forecast), the best prediction method, exponential smoothing combined with an AR model (middle), and the GARCH(1,1) forecast (bottom) for the exchange rate US\$/SFr.

5.2 Estimating the Risk Measures

For the following computations we assume the time series $\{R_t\}$, containing daily asset returns, to be stationary and to have its dynamics given by

$$R_t = \sigma_t \varepsilon_t, \quad (25)$$

where the innovations ε_t are iid according to the cumulative distribution function F_ε with zero mean and unit variance. Under these model assumptions, value at risk q_t and conditional expected shortfall S_t simplify to

$$q_t = \sigma_t \cdot F_\varepsilon^{-1}(\alpha), \quad (26)$$

and

$$S_t = \sigma_t \cdot E[\varepsilon_t | \varepsilon_t < F_\varepsilon^{-1}(\alpha)] = \frac{\sigma_t}{\alpha} \cdot \int_0^{F_\varepsilon^{-1}(\alpha)} x dF_\varepsilon(x), \quad (27)$$

respectively. To implement an estimation of the two risk measures, we must have a volatility forecast $\hat{\sigma}_t$ and the distribution F_ε of the innovations ε_t . In the previous sections, we already investigated the issue of volatility forecasting. We will now focus on the distribution of the innovations ε_t . For both volatility forecasts, the residuals $\frac{R_t}{\hat{\sigma}_t}$ and their squares did not show any significant autocorrelation structure. Hence, the iid assumption for the residuals seems plausible. Their empirical distribution looks symmetric but clearly leptokurtic.

In order not to underestimate the conditional quantiles we use Student's leptokurtic t-distribution for the residuals $\frac{R_t}{\hat{\sigma}_t}$, and scale it to have a variance equal to one. The degrees of freedom ν are estimated with the maximum likelihood method.

| standardized by: | US\$/SFr | GB£/US\$ | US\$/JPY |
|----------------------------------|----------|----------|----------|
| $\hat{\sigma}_{Best Prediction}$ | 5.247 | 4.615 | 4.601 |
| $\hat{\sigma}_{GARCH}$ | 5.184 | 4.832 | 4.457 |

Table 4: Degrees of freedom $\hat{\nu}$ estimated with maximum likelihood.

Since we want to construct the risk measures in a *predictive* fashion, we have to estimate the distribution F_ε with the residuals $\frac{R_t}{\hat{\sigma}_t}$, where $\hat{\sigma}_t$ is a forecast. If we were using $\frac{R_t}{\sigma_{t,RV}}$ instead, we would *underestimate* the tail of the innovations ε_t : it is unrealistic to mimic an unpredictable innovation by $\frac{R_t}{\sigma_{t,RV}}$, where $\sigma_{t,RV} \in \mathcal{F}_t$ rather than \mathcal{F}_{t-1} . That is why we believe that $\hat{\nu}$ in the range of 5 is much more appropriate than $\hat{\nu}$ close to infinity, what would correspond to approximately $\mathcal{N}(0, 1)$ distributed innovations.

We now estimate the risk measures as follows. The value at risk is given by

$$\hat{q}_t = \hat{\sigma}_{t,\mathcal{M}} \cdot c(\hat{\nu}) \cdot t_{\hat{\nu}}^{-1}(\alpha) \quad \text{for all } t \in \mathcal{T}, \quad (28)$$

where $\hat{\sigma}_{t,\mathcal{M}}$ denotes the volatility forecast from a particular forecasting procedure \mathcal{M} , $c(\hat{\nu}) = \sqrt{\frac{\nu-2}{\nu}}$ is the constant scaling the innovation variance to unity, and $t_{\hat{\nu}}^{-1}(\alpha)$ is the α -quantile of the $t_{\hat{\nu}}$ distribution. We choose $\alpha = 1\%$ and plot in figure 4 the value at risk based on the best volatility forecast and the one based on the GARCH(1,1) predictions.

We observe that the value at risk estimated with the best volatility prediction lies most of the times above its analogue from the GARCH forecast. Nevertheless, the value at risk estimated with the best volatility forecast seems to be able to avoid violations

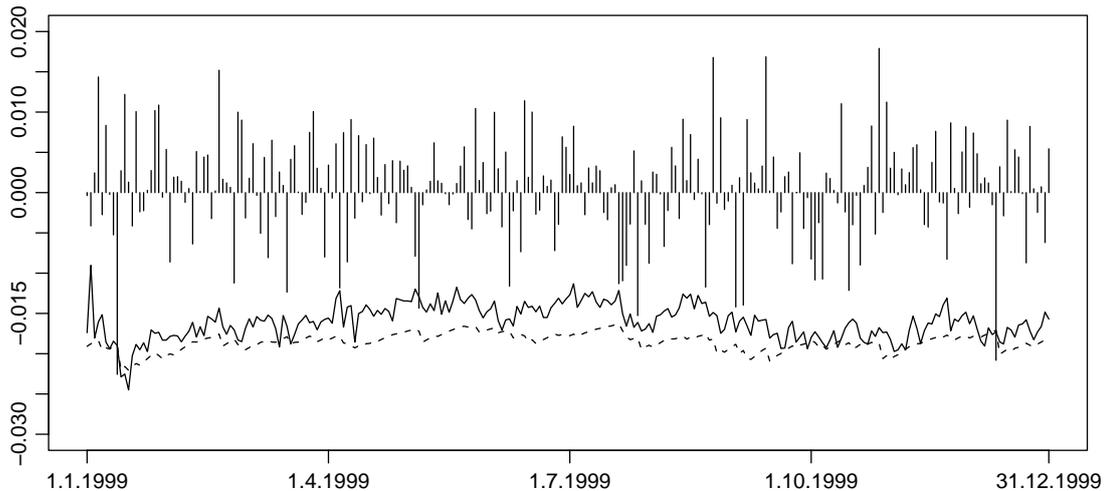


Figure 4: 1%-value at risk based on the best volatility forecast (solid line) and 1%-value at risk computed with GARCH(1,1) predictions (dashed line) superimposed on the returns of the exchange rate US\$/SFr.

despite generally being on a higher level. This is easy to explain, because the best forecast describes the conditional variance of the returns much more accurately than the GARCH model. The best forecasting procedure allows therefore more progressive risk management, which is of great importance in practice. The question whether it is not too progressive and underestimates the conditional quantile will be addressed in section 5.3.

An estimate for the conditional expected shortfall S_t is given by

$$\widehat{S}_t = \widehat{\sigma}_{t;\mathcal{M}} \cdot E_{t_{\widehat{\nu}}}[\varepsilon_t | \varepsilon_t < t_{\widehat{\nu}}^{-1}(\alpha)] = \frac{\widehat{\sigma}_{t;\mathcal{M}}}{\alpha} \cdot \int_0^{t_{\widehat{\nu}}^{-1}(\alpha)} x f_{\widehat{\nu}}(x) dx \quad \text{for all } t \in \mathcal{T}, \quad (29)$$

where $c(\widehat{\nu})$ and $t_{\widehat{\nu}}^{-1}(\alpha)$ are defined as in (28), and $f_{\widehat{\nu}}(\cdot)$ is the density function of the $t_{\widehat{\nu}}$ -distribution. As for the value at risk, we will compare the expected shortfall based on the two volatility forecasting procedures graphically. The plot in figure 5 is very similar to the one in figure 4. Again, the expected shortfall estimated with the best forecast lies above its analogue from the GARCH prediction. The similarity is not surprising since both risk measures are primarily depending on the volatility forecast. The best volatility prediction yields a more progressive and optimistic estimate of the conditional expected shortfall. From a practical point of view, we prefer this more progressive risk measure based on the best forecast, provided it is not unaware of the true existing risk. This is indeed the case as argued in the next section.

5.3 Backtesting the Risk Measures

We compare the predicted conditional quantile \widehat{q}_t with the true return R_t , and simply count how often the quantile was violated. A violation is said to occur whenever $R_t < \widehat{q}_t$. The total number of violations $N(\mathcal{M})$ is approximately binomially distributed if the model is correct:

$$N(\mathcal{M}) \sim \text{Bin}(m, \alpha), \quad (30)$$

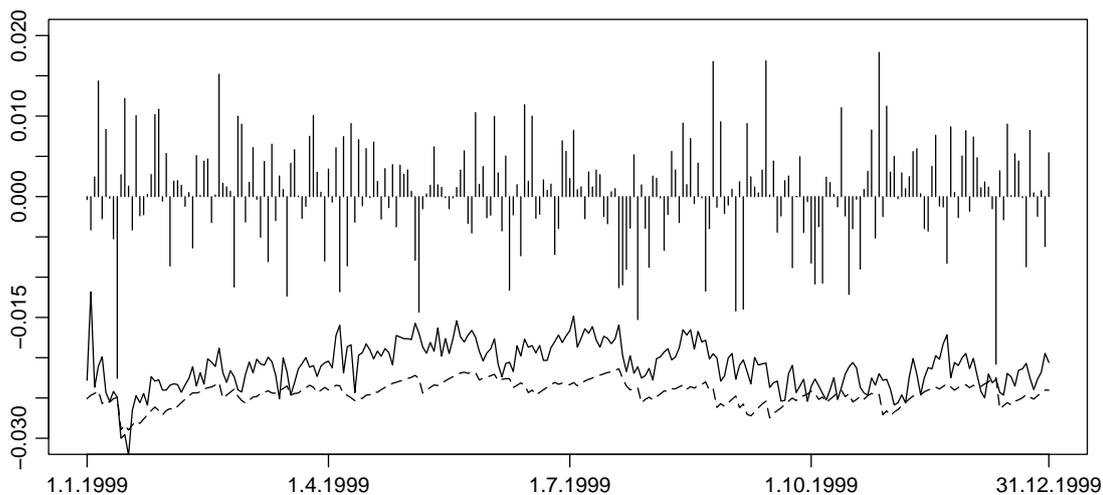


Figure 5: Conditional expected shortfall with $\alpha = 1\%$ based on the best volatility forecast (solid line) and conditional expected shortfall computed with GARCH(1,1) predictions (dashed line) superimposed on the returns of the exchange rate US\$/SFr.

see McNeil and Frey (2000). We perform a two-sided binomial test for the null hypothesis that a forecasting method \mathcal{M} correctly estimates the conditional quantiles q_t against the alternative that the method has a systematic error. We assessed the quantile violations for the value at risk based on the best volatility forecast and the one computed with GARCH predictions.

| | US\$/SFr | | GB£/US\$ | | US\$/JPY | |
|--------------------------------------|------------|---------|------------|---------|------------|---------|
| | violations | p-value | violations | p-value | violations | p-value |
| expected | 16 | - | 16 | - | 16 | - |
| with $\hat{\sigma}_{BestPrediction}$ | 21 | 0.256 | 20 | 0.312 | 36 | 0.000 |
| with $\hat{\sigma}_{GARCH}$ | 15 | 0.900 | 6 | 0.006 | 35 | 0.000 |

Table 5: Violations of the 1%-value at risk and p-values of the two-sided binomial test.

Table 5 indicates an advantage for the method with $\hat{\sigma}_{BestPrediction}$: it rejects the model only once, whereas the GARCH prediction yields two rejections. Moreover, according to our backtest, the value at risk obtained with the best forecasting model is not too progressive. However, one should keep in mind that the power of the backtest is fairly low, due to relatively few expected violations with $\alpha = 1\%$.

In a backtest for the conditional expected shortfall, we analyze

$$\hat{r}_t = \frac{R_t - S_t}{\hat{\sigma}_t} \quad \text{for } t \in \mathcal{T} \text{ with } R_t < \hat{q}_t, \quad (31)$$

some standardized residuals between the true return and the conditional expected shortfall in case of quantile violation. If the model is correct, the residuals \hat{r}_t are approximately equal to

$$\hat{r}_t \approx \varepsilon_t - E_{t_{\hat{v}}}[\varepsilon_t | \varepsilon_t < t_{\hat{v}}^{-1}(\alpha)], \quad (32)$$

which indicates that they should be approximately iid with mean zero. Such residuals are given in figure 6: the procedures with $\hat{\sigma}_{BestPrediction}$ and with the GARCH model yield residuals which vary similarly around zero. Moreover, the autocorrelation of the empirical

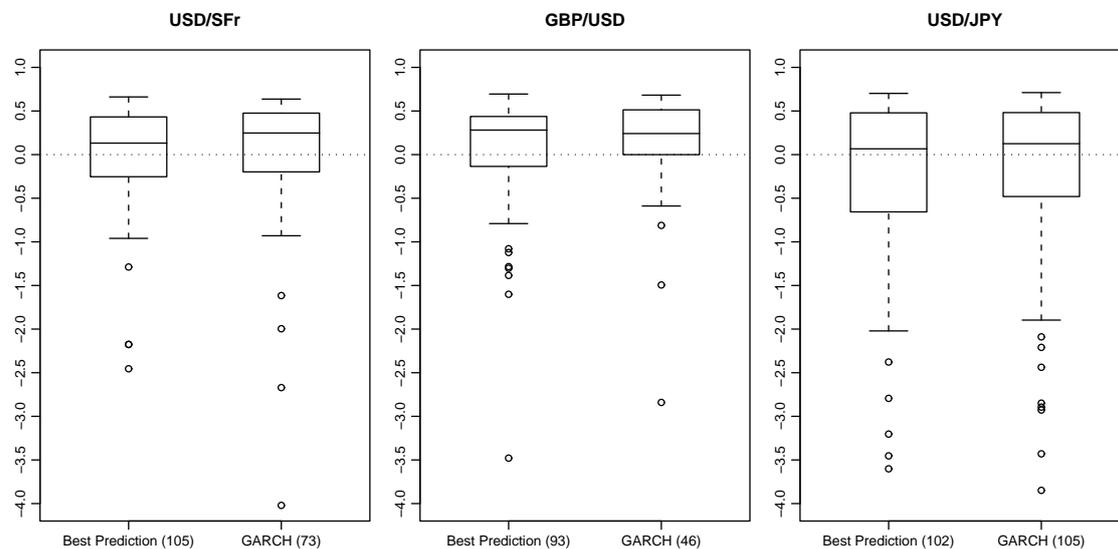


Figure 6: Boxplots for the empirical standardized residuals \hat{r}_t between the true return and the 5%-conditional expected shortfall in case of quantile violation. The numbers in parantheses are the number of quantile violations.

residuals \hat{r}_t as well as the ones of their squares \hat{r}_t^2 do not indicate a specific structure. We conclude from these facts that the more progressive conditional expected shortfall based on the best volatility prediction is not significantly biased.

6 Conclusions

The present paper deals with constructing a powerful forecasting procedure for the volatility. The recognition that the use of high frequency data approximately renders volatility from a latent to an observable quantity opens new directions for its prediction. Powerful state of the art forecasting procedures can be directly fitted to the realized volatility without having to rely on the much more complicated GARCH-type volatility models with latent variables. In an empirical study we explore the performance of various models and modern prediction tools, and discuss the problem of performance measurement. Finally, we investigate the practical impact of the improved volatility forecast by estimating the two most popular risk measures, value at risk and conditional expected shortfall. Our main findings can be summarized as follows:

- The volatility predictions based on high frequency realized volatility are of much better quality than the ones computed with a GARCH(1,1) model for daily data. Deviance measures show an improvement between 20% and 150% over the GARCH(1,1) benchmark when the target is the realized volatility $\sigma_{t,RV}$. Moreover, a fresh look at the problem of evaluating forecasts for the latent volatility σ_t motivates a test for

ranking performances. We find in two of the three datasets a significant advantage of our best forecast with high frequency data over the GARCH(1,1) model.

- The best forecasting procedure for the volatility is derived from detrending the log-transformed realized volatilities $\log(\sigma_{t,RV})$ by exponential smoothing, and forecasting the residual process $\{Z_t\}$ with an AR model. Surprisingly, nonlinear and sophisticated models for the filtered residual process $\{Z_t\}$ did not improve the accuracy of the forecast. We consider this as an empirical proof for linearity of the exponentially smoothed residual process $\{Z_t\}$.
- Volatility predictions are frequently used in risk management. We explore how our more reliable volatility forecast improves the two most popular risk measures, value at risk and conditional expected shortfall. We observe that the improved volatility forecasts lead to more progressive risk measures. According to backtests, we gained evidence that they are not underestimating the true underlying risks. Our findings about the risk measures should be interpreted with caution, because the focus is on tail events which are very delicate to quantify. Further consideration of the risk measures' quality as well as quantifying the profit from an economic point of view are left for future research.

Acknowledgements. We thank Olsen&Associates for providing the high frequency data.

References

- [1] Andersen, T.G. and Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 39, 885-905.
- [2] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (1999a). The Distribution of Exchange Rate Volatility. NBER Working Paper 6961.
- [3] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (1999b). Realized Volatility and Correlation. Preprint, 1-21.
- [4] Artzner, P., Delbaen, F., Eber, J., Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance* 9, 203-228.
- [5] Bickel, P.J. and Bühlmann, P. (1996). What is a Linear Process? *Proceedings National Academy of Sciences USA* 93, 12128-12131
- [6] Bühlmann, P. (1996). Locally Adaptive Lag-Window Spectral Estimation. *Journal of Time Series Analysis* 17, 247-270.
- [7] Bühlmann, P. and Ferrari, F. (2001). Dynamic Combination of Models. Preprint, ETH Zürich.
- [8] Corsi, F., Dacorogna, M., Müller, U. and Zumbach, G. (2001). Consistent High-Precision Volatility from High-Frequency-Data. Internal Paper, Olsen&Associates, Zürich, Switzerland.
- [9] Künsch H.R. (1986). Discrimination Between Monotonic Trends and Long-Range Dependence. *Journal of Applied Probability* 23, 1025-1030.

- [10] McNeil, A.J. and Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach. *Journal of Empirical Finance*, 7, 271-300.
- [11] Merton, R.C. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics*, 8, 323-361
- [12] Officer, R.R. (1973). The Variability of the Market Factor of the NYSE. *Journal of Business*, 46, 434-453.
- [13] Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge.
- [14] Schwert, G.W. (1998). *Stock Market Volatility: Ten Years After the Crash*. Brookings-Wharton Papers on Financial Services, Washington DC.
- [15] Taylor, S.J. and Xu, X. (1997). The Incremental Volatility Information in One Million Foreign Exchange Quotations. *Journal of Empirical Finance*, 4, 317-340.