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

October 23, 2002

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: (i) to select an accurate forecasting procedure for predicting volatilities based on high frequency data from various standard models and modern prediction tools; (ii) to evaluate the predictive potential of those volatility forecasts for both the realized and the true latent volatility; and (iii) to quantify the differences using volatility forecasts based on high frequency data and using a GARCH model for low frequency (e.g. daily) data, and study its implication in risk management for two widely used risk measures. The pay-off using high frequency data for the true latent volatility is empirically found to be still present, but magnitudes smaller than suggested by simple analysis.

JEL Classification: C22; C52; G10
Keywords: Forecasting, High-Frequency-Data, Predictive Potential, Risk Measures, Volatility

*Corresponding author. Tel.: +41-1-632-5319; fax: +41-1-632-1228; e-mail: dettling@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, accurate forecasts of the latent volatility are crucial for the successful implementation of these techniques.

We define the latent volatility σ_t for the time period of one day as $\sigma_t = \sqrt{Var(R_t \mid \mathcal{F}_{t-1})}$, where R_t is the daily return and $\mathcal{F}_{t-1}$ is the information available on the return process up to time $t-1$. Prediction of σ_t^2 can be done on the basis of daily data using parametric volatility models such as ARCH or GARCH, or by using implied volatility out of the Black-Scholes formula. A principal goal of the paper is to address the problem of quantifying possible gains by using high frequency rather than daily data, see also the discussion below.

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 intra-daily 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) apply 5-minute returns to estimate daily exchange rate volatilities.

Andersen et al. (1998) computed volatility estimates from aggregated high frequency returns and found them 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 an accurate forecasting procedure for volatilities based on high frequency data, and

b) to evaluate the quality of the volatility forecasts not only for future realized volatilities $\sigma_{t;RV}$, but also for the true latent volatilities σ_t, and

c) to quantify the improvement of using accurate forecasts based on high frequency data over a GARCH model for daily data and study its implication in the context of risk management.

Regarding issue a), we consider a variety of modern techniques and prediction tools for volatility forecasting based on 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 a simple linear autoregressive (AR) model was empirically found to have among the best overall 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 similarly accurate predictions.

Issue b) deals with judging volatility forecasts and addresses the problem of quantifying possible gains using models based on high frequency instead of daily data. 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 $\widehat{\sigma}_t$ by comparing it with $\sigma_{t;RV}$. This approach has been pursued by Andersen et al. (1998), but it does not fully reflect how accurately the target, namely the latent volatility σ_t, can be estimated; this, because high frequency sampling is only an approximation to

infinite sampling frequency and because the Brownian motion model may be (strongly) inadequate. We propose to evaluate volatility forecasts $\widehat{\sigma}_t$ by assuming stationarity and conditional mean zero of returns only: we then calibrate the squared prediction $\widehat{\sigma}_t^2$ against the squared returns R_t^2 which are very noisy but unbiased estimates of the true squared volatility σ_t^2. This noise may disturb much of the differences between forecasts for the true volatility, and we introduce here mean-type test statistics which reduce noise variance due to averaging.

Considering issue c), we could exploit advantages of our powerful forecasting procedure with high frequency data compared to a GARCH model with daily returns. Such gains can then be quantified in terms of p-values rather than (relative) differences between performance measures. An important application of volatility prediction 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 high frequency 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 a powerful method for volatility forecasting. The design and the results of our empirical study are presented in sections 4 and 5, whereas section 6 discusses the impact of the volatility predictions on the risk measures. Finally, we conclude with a brief summary in section 7.

2 High Frequency Volatility

Let $P_{(m),t}$, where $t = \frac{1}{m}, \frac{2}{m}, \ldots$, 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:

$$R_{(m),t} = log\Big(P_{(m),t}\Big) - log\Big(P_{(m),t-\frac{1}{m}}\Big), \qquad t = \frac{1}{m}, \frac{2}{m}, \ldots. \tag{1}$$

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 return[1] $R_{(1),t}$ 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 hope to use high frequency data is to improve the estimate of the latent volatility $\sigma_t = \sqrt{Var(R_t \,|\, \mathcal{F}_{t-1})}$. For a short time horizon Δt, the latent volatility can be estimated on the basis of the squared return, i.e. $\widehat{\sigma}_{(m),t} = \sqrt{R^2_{(m),t}}$. 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, although not a prediction which would be a function of $\mathcal{F}_{t-1}$ only, is obtained by summing up the squared high frequency returns:

$$\widehat{\sigma}_t = \sigma_{t;RV} = \sqrt{\sum_{i=0}^{m-1} R^2_{(m),t-\frac{i}{m}}} \tag{2}$$

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

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

that the sum of intra-day 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 $\widehat{\sigma}_t$ 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 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.

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$) from January 1, 1990 through December 31, 1999. To obtain greater variety of the data, the third exchange rate was chosen with respect to the US$, whereas the others are with respect to a non US currency. The construction for realized volatilities, using measurement intervals Δt of 5 minutes, is actually more complicated than described in formula (2): the preprocessing and bias corrections described in Corsi et al. (2001) have been used. Such preprocessing is important and there seems to be still room to improve upon that task. But this is beyond the scope of the present paper: we rather take the view that the preprocessed realized volatilities are given to us as "good" and useful auxiliary variables. All together, we have a total of 2600 such daily observations of realized volatilities and log-returns per dataset.

3 Forecasting Volatility

Plotting the realized volatilities $\sigma_{t;RV}$ against time, we recognize in figure 1 that its fluctuations show substantial persistence and the upward spikes are much larger than the downward spikes. The latter fact is an evidence of skewness in the data, which was also confirmed by an analysis of the empirical distribution.

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 decay. This decay could be interpreted as hyperbolic and thus as an evidence of long-range dependence. An alternative is to consider the presence of a deterministic component: we therefore apply the classical decomposition model to the transformed time series,

$$log(\sigma_{t;RV}) = X_t = m_t + \xi_t, \tag{3}$$

where $m_t = E[X_t]$ is deterministic and $(\xi_t)_t$ is a stochastic process with $E[\xi_t] = 0$, cf. Brockwell and Davis (1991). Note that on the basis of data, it is very difficult or even impossible to answer the question whether the (log-)realized volatilities are long-range dependent or whether they are nonstationary due to a deterministic component, see Künsch

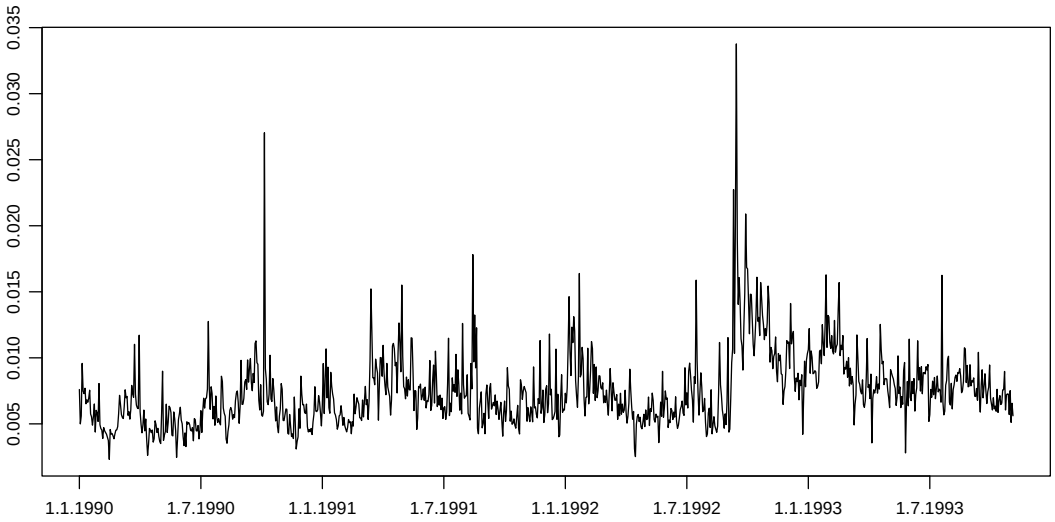

Figure 1: The realized volatility $(\sigma_{t;RV})_{t\in\{1,\ldots,1000\}}$ for the exchange rate GB£/US$.

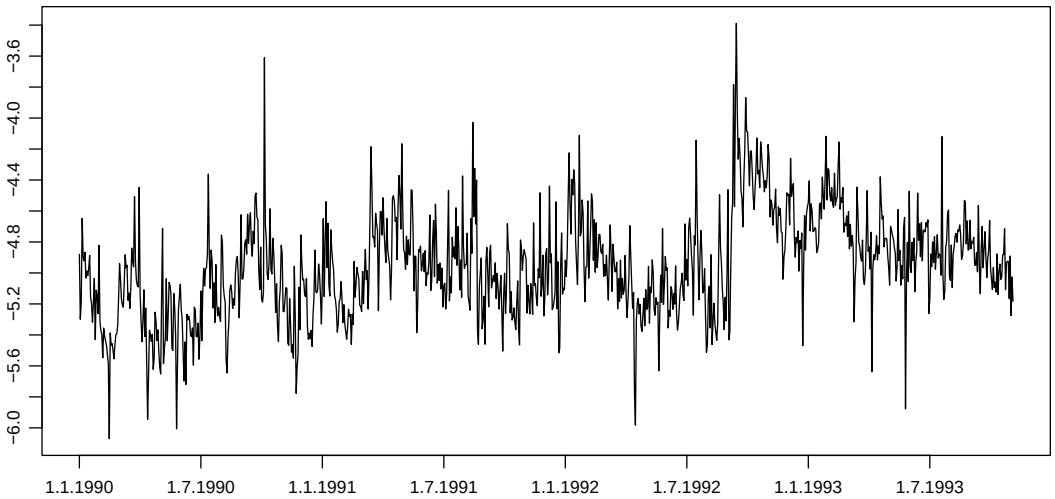

Figure 2: The log-transformed series $(X_t)_{t\in\{1,\ldots,1000\}}$ for the exchange rate GB£/US$.

(1986). In fact, we will see in section 4.2 that both long memory and decomposition models yield very similar results. This is not surprising, as pointed out in more detail in section 4.2.8.

We estimate the deterministic component with an asymmetric linear filter,

$$\widehat{m}_t = \sum_{j=0}^{t-1} a_j X_{t-j}, \tag{4}$$

which depends on a set of weights $\{a_j\}$. We choose them with an exponential decay and

standardize so that $\sum a_j = 1$,

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

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. See for example Gijbels et al. (1999), making a connection of exponential smoothing to kernel smoothing. Subtraction of $\widehat{m}_t$ then leaves us with residuals

$$Z_t = X_t - \widehat{m}_t, \tag{6}$$

the empirical analogues to the ξ_t from (3). An inspection by descriptive techniques shows that the residual process $(Z_t)_t$ has a symmetric and approximately normal distribution. Furthermore, it seems to be stationary. But most importantly, its sample autocorrelation function is decaying quickly, which indicates that the deterministic component has been successfully removed.

A simple parametric stochastic model for the residual time series $(Z_t)_t$ is a linear autoregressive model of order p, 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 one step ahead forecasts

$$\widehat{Z}_{t+1} = \hat{g}(Z_t, Z_{t-1}, \ldots), \tag{7}$$

where $\hat{g}(\cdot)$ is an estimated prediction function, e.g. from an AR or from a nonlinear model. We then use them in conjunction with a predictive estimate $\widehat{\widehat{m}}_{t+1}$ of the deterministic component for the prediction of the log-realized volatilities. We assume that the trend varies only slowly, so that $m_{t+1} \approx m_t$. The predictive estimate is then set as $\widehat{\widehat{m}}_{t+1} = \widehat{m}_t$. Therefore, the predicted log-realized volatility is given by

$$\widehat{X}_{t+1} = \widehat{m}_t + \widehat{Z}_{t+1}. \tag{8}$$

We will see in section 4.2 that, compared to more complex and nonlinear methods, we obtain very competitive forecasting results if we use the simple AR model to estimate the prediction function $\hat{g}(\cdot)$ for the residual process $(Z_t)_t$.

4 Predictive Potential for Realized Volatilities

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, without implying a better out-of-sample accuracy. 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

$$R_t = \sigma_t \varepsilon_t \tag{9}$$

with ε_t iid innovations having $E[\varepsilon_t] = 0$ and $Var(\varepsilon_t) = 1$, then

$$R_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2 = \sigma_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2 + \eta_t, \tag{10}$$

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

$$E[(R_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2)^2] = E[(\sigma_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2)^2] + E[\eta_t^2] \tag{11}$$

or its empirical analogue $\sum_{t\in\tau}(R_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2)^2$ is dominated by the large nuisance term $E[\eta_t^2]$. This implies that the values of the performance measure in (11) 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 (11) is to substitute the subsequent realizations σ_t with the realized volatilities $\sigma_{t;RV}$, which considerably 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. We discuss in section 5 how to compare accuracy of different prediction methods for the true latent volatility. However, to gain a first impression of the performance of a particular forecasting model $\mathcal{M}$, we will compute the residual sum of squares,

$$RSS_\sigma(\mathcal{M}) = \sum_{t\in\mathcal{T}} (\sigma_{t;RV} - \widehat{\sigma}_{t;\mathcal{M}})^2. \tag{12}$$

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:

$$RSAD_\sigma(\mathcal{M}) = \sum_{t\in\mathcal{T}} |\sigma_{t;RV} - \hat{\sigma}_{t;\mathcal{M}}|\,. \tag{13}$$

4.2 Prediction Methods

This section contains an empirical study, where we explore the performance of standard forecasting procedures and modern prediction tools for the volatility.

4.2.1 AR modelling

Our aim here is to find an accurate forecasting procedure for the empirical residual process $(Z_t)_t$ from the decomposition model in (6). Since a descriptive analysis of the residual process $(Z_t)_t$ shows that its sample autocorrelation function is decaying exponentially quick, it is sensible to model it with a simple linear autoregressive process. The order p of that model is chosen by Akaike's information criterion (AIC), and forecasts are computed in the usual manner.

In the next two subsections we present two state of the art prediction techniques, projection pursuit autoregression and dynamic combination of models, that could outperform the AR model as a forecasting procedure for the residual process $(Z_t)_t$. Both these methods are generally competitive nonlinear prediction procedures for conditional means and future observations in stationary time series.

4.2.2 Projection Pursuit

The projection pursuit autoregression model for the stationary mean zero residual process $(Z_t)_t$ from (6)is given by

$$Z_t = \mu_t + W_t,$$
$$\mu_t = \mathbb{E}[Z_t|Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p}] = \sum_{k=1}^{m} f_k(\alpha_k + \sum_{j=1}^{p} \phi_{k,j} Z_{t-j}), \tag{14}$$

where $f_k(\cdot)$ are unknown ridge functions, $\alpha_k, \phi_{k,j}$ are unknown parameters, and $(W_t)_t$ is 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 squared 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 even be achieved by much simpler AR predictions where the order p was chosen by AIC on the learning data $\mathcal{L}$ only.

4.2.3 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 (2002). The dynamic combination of AR(p) models for the stationary mean zero residual process $(Z_t)_t$ from (6) is given by

$$Z_t = \mu_t + W_t,$$
$$\mu_t = \mathbb{E}[Z_t|Z_{t-1}, Z_{t-2}, \ldots] = \sum_{k=1}^{m} P_{t;k}(\alpha_k + \sum_{j=1}^{p} \phi_{k,j} Z_{t-j}), \tag{15}$$

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 disjunct. The conditional probabilities are modelled as variable length Markov chains

$$P_{t;k} = \mathbb{P}[Z_t \in I_k|Z_{t-1}, Z_{t-2}, \ldots, Z_{t-\ell_t}], \tag{16}$$

where $\ell_t = \ell(Z_{t-1}, Z_{t-2}, \ldots)$ is the *variable* length of the memory, depending on the past values $Z_{t-1}, Z_{t-2}, \ldots$. 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 (2002). 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. But neither the dynamic combination of models with outsample-optimized number of variables p could outperform AR predictions where p was chosen by AIC on the learning set $\mathcal{L}$ only.

4.2.4 Long Memory Models

A totally different interpretation of the slow hyperbolic decay in the autocorrelation function of the log-realized volatilities $\log(\sigma_{t;RV})$ is given by a long-memory characteristic rather than a deterministic component as in (3). Such processes can be modelled by a fractionally integrated ARFIMA(p, d, q), where the parameter d is not an integer. As model order, we set $q = 0$ and choose p according to Akaike's information criterion as in the exponential smoothing setting. The model coefficients as well as the degree of fractional integration are then simultaneously estimated on the learning set $\mathcal{L}$.

4.2.5 Benchmark Methods

The popular GARCH(1,1) is a standard model for volatility forecasting without high frequency data. As in Andersen et al. (2002), we use it 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 (12) or (13).

In order to check whether the filtering process by exponential smoothing improves the accuracy of the predictions, we also forecast the unfiltered log-realized volatilities X_t with a linear AR model, where the order p is again chosen by AIC. The use of an AR model does not correspond well with the slowly decaying autocorrelation function of $(X_t)_t$, and therefore we expect a worse performance. As an extension of the autoregressive approach, we also consider projection pursuit directly applied on the unfiltered log-realized volatility X_t, see section 4.2.2.

4.2.6 Results

In the following tables 1-3 the deviance measures for each of the three exchange rates are given[2] . 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 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 sections 4.2.7 and 4.2.8.

The two benchmark methods, projection pursuit autoregression and AR(p), which are directly fitted on the log-realized volatilities X_t, cannot 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, using the GARCH(1,1)-model for daily data, we observe a very bad performance. We thus exploit the celebrated fact that high frequency data may drastically improve volatility forecasts. In section 5, we take a fresh look to see whether this is also true when

[2] Abbreviations in tables 1-3:

ES	Exponential Smoothing.
AR(p)	Linear autoregressive model of order p.
PPR(m, p)	Projection pursuit autoregression of order (m, p), see formula (14).
DCM(m, p)	Dynamic combination of m AR(p) models, see formula (15).
ARFIMA(p, d, q)	Fractionally integrated ARMA(p, q) with degree of fractional integration d.

Prediction Method	RSS_σ	*in%*	$RSAD_\sigma$	*in %*
ES & AR(4)	$35.65 \cdot 10^{-4}$	+ 0.65%	$157.62 \cdot 10^{-2}$	
ES & PPR(1,5)	$36.18 \cdot 10^{-4}$	+ 2.15%	$158.52 \cdot 10^{-2}$	+ 0.57%
ES & DCM(2,4)	$35.65 \cdot 10^{-4}$	+ 0.65%	$158.10 \cdot 10^{-2}$	+ 0.30%
ARFIMA(4,0.27,0)	$35.42 \cdot 10^{-4}$		$159.22 \cdot 10^{-2}$	+ 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. RSS_σ and $RSAD_\sigma$ are defined as in (12) and (13). The third and fifth column (*in %*) show the percentual deterioration against the best value. ES & AR(p), ARFIMA(p, d, q) and AR(p) are data-driven forecasting procedures using high frequency data. ES in combination with PPR and DCM, as well as PPR on its own are outsample-optimized benchmarks using high frequency data. Finally, GARCH(1,1) is a benchmark using daily data.

Prediction Method	RSS_σ	*in %*	$RSAD_\sigma$	*in %*
ES & AR(5)	$18.31 \cdot 10^{-4}$	+ 0.38%	$104.55 \cdot 10^{-2}$	
ES & PPR(1,3)	$18.60 \cdot 10^{-4}$	+ 1.97%	$105.70 \cdot 10^{-2}$	+ 1.10%
ES & 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$. Prediction methods and deviance measures are as explained in the caption of table 1.

Prediction Method	$RSAD$	*in %*	$RSAD_\sigma$	*in %*
ES & AR(5)	$87.25 \cdot 10^{-4}$		$202.53 \cdot 10^{-2}$	+ 0.79%
ES & PPR(1,7)	$88.56 \cdot 10^{-4}$	+ 1.50%	$202.17 \cdot 10^{-2}$	+ 0.61%
ES & 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. Prediction methods and deviance measures are as explained in the caption of table 1.

measuring the forecasting potential for the true volatility σ_t, rather than for the realized volatility $\sigma_{t;RV}$.

4.2.7 Linearity of the Detrended Process

The results given in the tables show that the three 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 and couldn't be beaten by the much more complex projection pursuit autoregression or the dynamically weighted combination of two AR models. Additionally, the simple AR(p) was handicapped by the fact that its order p was chosen by AIC on the learning set $\mathcal{L}$, whereas for the two nonlinear methods, p was selected to optimize the outsample performance.

We consider these observations as an evidence of linearity for the residual process $(Z_t)_t$. 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 $(Z_t)_t$ is that the ridge function estimated by the projection pursuit autoregression was found close to linear.

4.2.8 Trend or Long-Range dependence?

We observe that the decomposition model made up of exponential smoothing combined with AR(p) and the long memory ARFIMA model perform very similarly. Even though the models are fundamentally different, a closer look at their properties can explain this observation.

Both, fractional differentiation and exponential smoothing correspond to a linear filter of the log-realized volatilities. Moreover, both remaining time series after filtering are described with an AR model. Consequently, both these forecasting procedures are similar from a mathematical point of view, and it is no longer surprising that their results are almost the same.

5 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 in the basic model (9), the equation

$$R_t^2 = \sigma_t^2 \varepsilon_t^2 = \sigma_t^2 + \underbrace{\sigma_t^2(\varepsilon_t^2 - 1)}_{=\eta_t} = \sigma_t^2 + \eta_t \tag{17}$$

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

$$P(\mathcal{M}) = E[(R_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2)^2]. \tag{18}$$

Note that we assume here stationarity, implying independence of the right-hand side from t. Applying the notation of equation (17), we can convert the performance measure $P(\mathcal{M})$

into

$$P(\mathcal{M}) = E[(\sigma_t^2 - \widehat{\sigma}_{t;\mathcal{M}}^2)^2] + E[\eta_t^2]. \tag{19}$$

The first term on the right side of equation (19) apparently measures the predictive potential of a volatility forecast $\widehat{\sigma}_{t;\mathcal{M}}$, 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 also 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

$$U_t = (R_t^2 - \widehat{\sigma}_{t;\mathcal{M}_1}^2)^2 - (R_t^2 - \widehat{\sigma}_{t;\mathcal{M}_2}^2)^2 \quad \text{for all } t \in \mathcal{T} \tag{20}$$

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

$$\begin{aligned} E[U_t] &= E[(R_t^2 - \widehat{\sigma}_{t;\mathcal{M}_1}^2)^2] - E[(R_t^2 - \widehat{\sigma}_{t;\mathcal{M}_2}^2)^2] \\ &= E[(\sigma_t^2 - \widehat{\sigma}_{t;\mathcal{M}_1}^2)^2] - E[(\sigma_t^2 - \widehat{\sigma}_{t;\mathcal{M}_2}^2)^2] \\ &= P(\mathcal{M}_1) - P(\mathcal{M}_2). \end{aligned} \tag{21}$$

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 $E[U_t] \neq 0$ or $E[U_t] > 0$.

We estimate the expectation of U_t by the mean $\widehat{\mu}_m = \frac{1}{m}\sum_{t\in\mathcal{T}} U_t$, 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, including stationarity of U_t, the central limit theorem for dependent variables implies

$$\sqrt{m}(\widehat{\mu}_m - E[U_t]) \xrightarrow{d} \mathcal{N}(0, \sigma_\infty^2) \quad (m \to \infty), \tag{22}$$

where $\sigma_\infty^2 = \sum_{k=-\infty}^{\infty} Cov(U_0, U_k)$. Since the variance of $\hat{\mu}_m$ does not depend on the value of $E[U_t]$, the asymptotic variance σ_∞^2 remains the same under the null hypothesis H_0, and hence

$$\sqrt{m}\widehat{\mu}_m \xrightarrow{d} \mathcal{N}(0, \sigma_\infty^2) \quad (m \to \infty) \tag{23}$$

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 $\{U_t\}$

$$f_U(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \mathrm{Cov}(U_0, U_k)\exp(-i\lambda k) \ (0 \le \lambda \le \pi) \tag{24}$$

at frequency $\lambda = 0$. We do this using a classical smoothed periodogram $\widehat{f}_U(\lambda)$, but more formal approaches would be possible, see Bühlmann (1996). The estimate for the asymptotic variance is then

$$\widehat{\sigma}_\infty^2 = 2\pi \cdot \widehat{f}_U(0). \tag{25}$$

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

$$\left|\frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} U_t\right| > \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big) \cdot \frac{\widehat{\sigma}_\infty}{\sqrt{|\mathcal{T}|}}, \tag{26}$$

	US$/SFr	GB£/US$	US$/JPY
p-value (one-sided test): $\mathcal{M}_2$ better than $\mathcal{M}_1$	0.009	$1 \cdot 10^{-9}$	0.038
p-value (two-sided test): $\mathcal{M}_2$ different from $\mathcal{M}_1$	0.018	$2 \cdot 10^{-9}$	0.076

Table 4: GARCH(1,1) ($\mathcal{M}_1$) and accurate high frequency ($\mathcal{M}_2$) predictions, using exponential smoothing and AR modelling. Testing null hypothesis of equal performance against better performance of high frequency method (one-sided), and against non-equal performance (two-sided). Low p-values are always in favor of $\mathcal{M}_2$, the high frequency predictions with exponential smoothing and AR modelling.

where $|\mathcal{T}|$ is the number of observations in the test set, i.e. $|\mathcal{T}| = 1600$. Alternatively, a one-sided test would reject the null hypothesis of equal performance against favoring model $\mathcal{M}_2$ if

$$\frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} U_t > \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big)\cdot\frac{\widehat{\sigma}_\infty}{\sqrt{|\mathcal{T}|}}. \tag{27}$$

We use this version of the t-test, adapted to dependent observations, 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 one of the best high frequency data methods, 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 p-values as given in table 4.

In all three datasets the use of high frequency data significantly improves the predictive potential for the latent volatility σ_t. The high frequency method based on exponential smoothing and AR modelling is thus more reliable than the GARCH(1,1)-model with daily data. By pairwise comparison of the forecasting procedures which are based on high frequency data, we could not observe any significant differences in the predictive potential. This is why we omit the respective p-values.

5.1 A Good Volatility Forecasting Procedure

According to our empirical study, we conclude that a good forecasting procedure for the latent volatility σ_t is to filter the log-transformed realized volatility $\log(\sigma_{t;RV})$ by exponential smoothing, and to predict the residual process $(Z_t)_t$ with an AR(p) model. The ARFIMA model for long-range dependence performed similarly and can be used alternatively. Exponential smoothing and subsequent AR modelling have the advantage to be very popular in standard time series analysis, are easy to implement and computationally cheap. In figure 3, a comparison between the ex-post realized volatility, the good forecast based on exponential smoothing combined with AR(p), and the GARCH(1,1) predictions is shown.

6 Risk Estimation

6.1 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

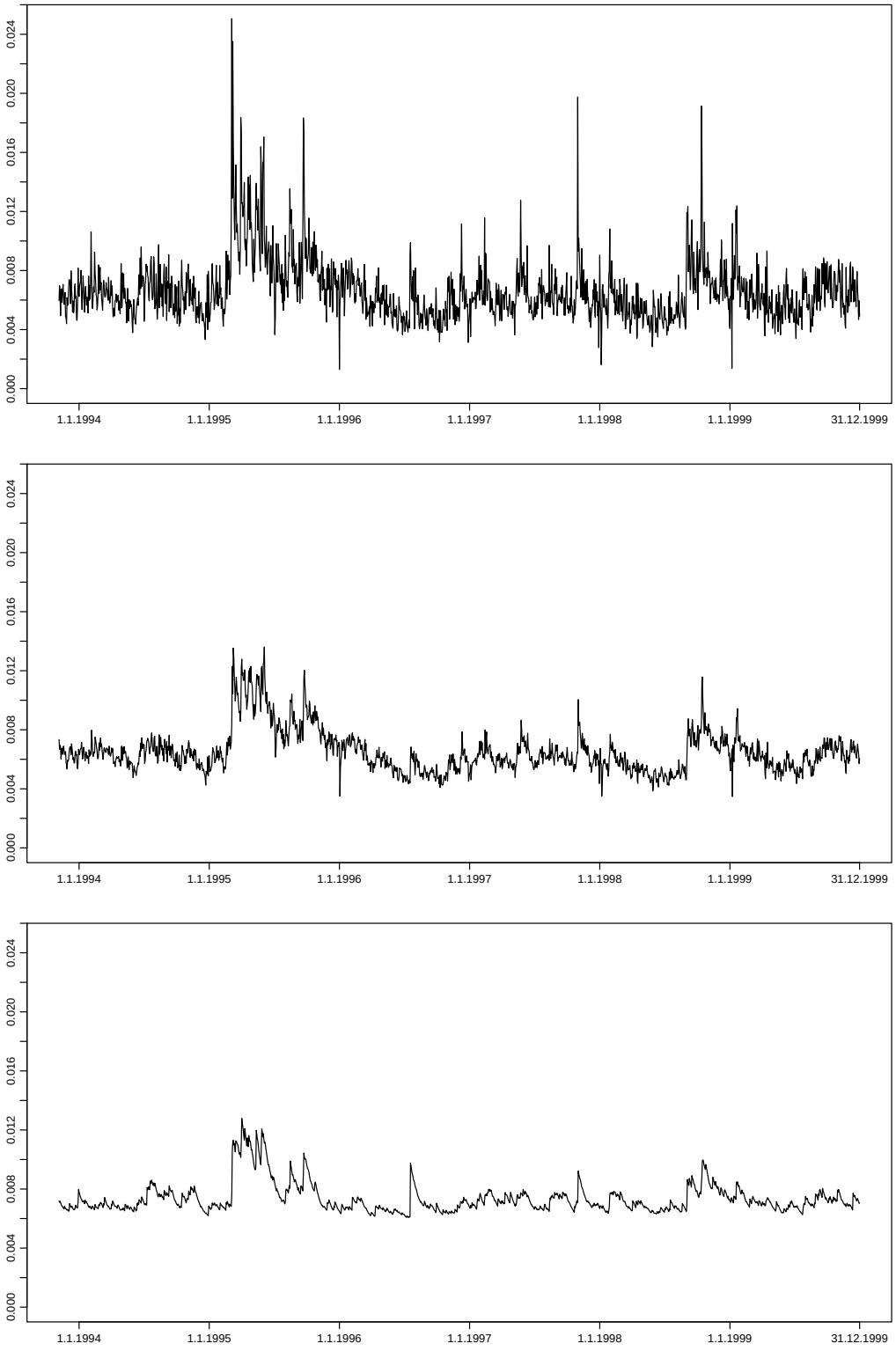

Figure 3: Comparison between ex-post realized volatility (top, not a forecast), the prediction based on exponential smoothing combined with an AR model using high frequency data (middle), and the daily GARCH(1,1) forecast (bottom) for the exchange rate US$/SFr.

prediction based on high frequency data improves the quality of the risk measures. In particular, we compare the impact of our "best" forecasting procedure, the combination of exponential smoothing with an AR-model, 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,

$$\mathcal{P}[R_t < q_t \,|\, \mathcal{F}_{t-1}] \leq \alpha. \tag{28}$$

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,

$$S_t = E[R_t \,|\, R_t < q_t, \mathcal{F}_{t-1}] \tag{29}$$

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

6.2 Estimating the Risk Measures

For the following computations we assume that the time series $(R_t)_t$, containing daily asset returns is stationary and follows the basic model

$$R_t = \sigma_t \varepsilon_t, \tag{30}$$

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

$$q_t = \sigma_t \cdot F_\varepsilon^{-1}(\alpha), \tag{31}$$

and

$$S_t = \sigma_t \cdot E[\varepsilon_t \,|\, \varepsilon_t < F_\varepsilon^{-1}(\alpha)] = \frac{\sigma_t}{\alpha} \cdot \int_{-\infty}^{F_\varepsilon^{-1}(\alpha)} x \, dF_\varepsilon(x), \tag{32}$$

respectively. To implement an estimation of the two risk measures, we must have a volatility forecast $\widehat{\sigma}_t$ 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 $\frac{R_t}{\widehat{\sigma}_t}$ 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.

Because of the empirically plausible symmetry, we use Student's leptokurtic t-distribution for the residuals $\frac{R_t}{\widehat{\sigma}_t}$, and scale it to have a variance equal to one. The degrees of freedom ν are then estimated with the maximum likelihood method.

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

Table 5: 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 $\frac{R_t}{\widehat{\sigma}_t}$, where $\widehat{\sigma}_t$ 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 $\widehat{\nu}$ in the range of 5 is much more appropriate than $\widehat{\nu}$ close to infinity which would correspond to approximately $\mathcal{N}(0,1)$ distributed innovations yielding potentially much to small quantiles for risk management.

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

$$\widehat{q}_t = \widehat{\sigma}_{t;\mathcal{M}} \cdot c(\widehat{\nu}) \cdot t_{\widehat{\nu}}^{-1}(\alpha) \quad \text{for all } t \in \mathcal{T}, \tag{33}$$

where $\widehat{\sigma}_{t;\mathcal{M}}$ denotes the volatility forecast from a particular forecasting procedure $\mathcal{M}$, $c(\widehat{\nu}) = \sqrt{\frac{\widehat{\nu}-2}{\widehat{\nu}}}$ is the constant scaling the innovation variance to unity, and $t_{\widehat{\nu}}^{-1}(\alpha)$ is the α-quantile of the $t_{\widehat{\nu}}$ 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.

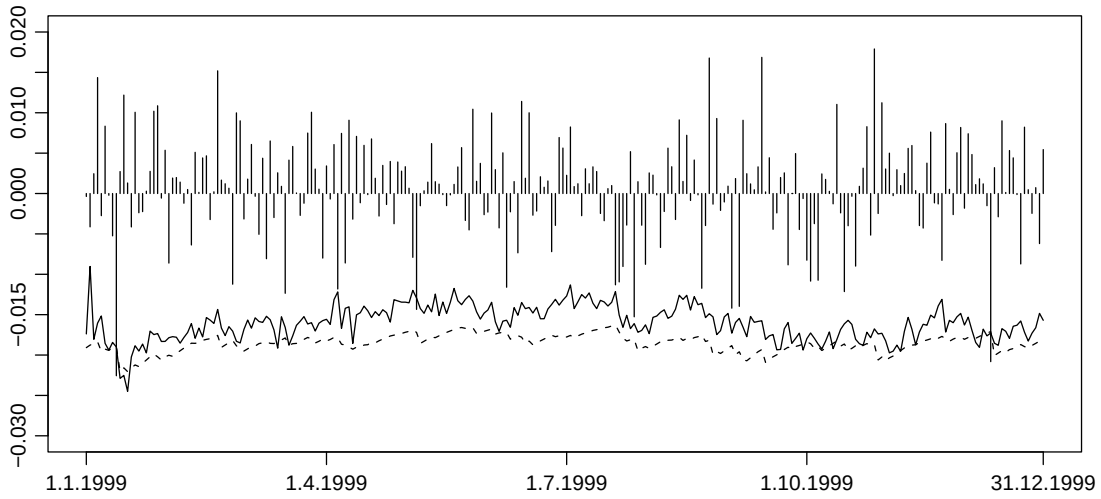

Figure 4: 1%-value at risk based on the "best" volatility forecast (solid line) and 1%-value at risk computed with GARCH(1,1) predictions (dashed line) superimposed on the returns of the exchange rate US$/SFr.

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 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 6.3.

An estimate for the conditional expected shortfall S_t is given by

$$\widehat{S}_t = \widehat{\sigma}_{t;\mathcal{M}} \cdot E_{t_{\hat{\nu}}}[\varepsilon_t \,|\, \varepsilon_t < t_{\widehat{\nu}}^{-1}(\alpha)] = \frac{\widehat{\sigma}_{t;\mathcal{M}}}{\alpha} \cdot \int_{-\infty}^{t_{\widehat{\nu}}^{-1}(\alpha)} x f_{\widehat{\nu}}(x)\, dx \quad \text{for all } t \in \mathcal{T}, \tag{34}$$

where $c(\widehat{\nu})$ and $t_{\widehat{\nu}}^{-1}(\alpha)$ are defined as in (33), 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.

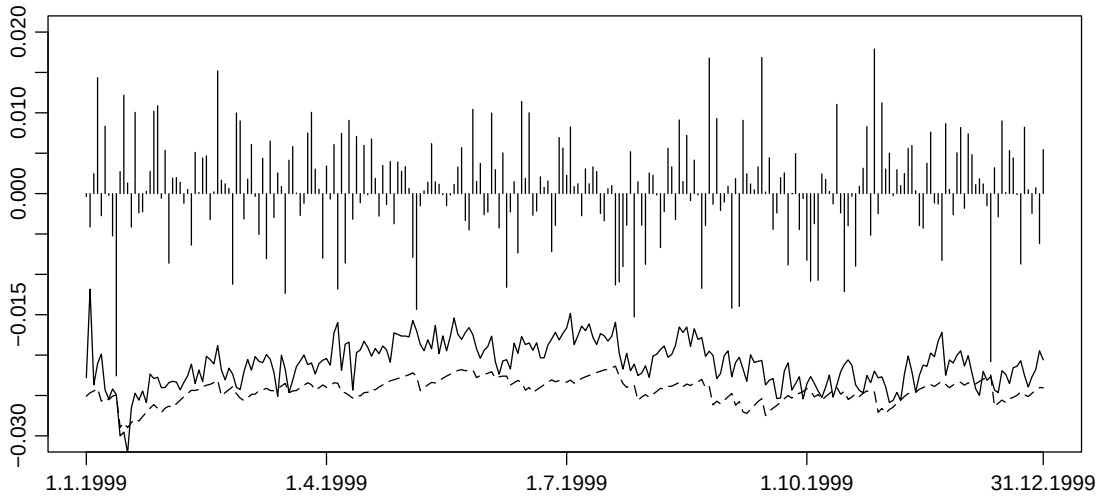

Figure 5: Conditional expected shortfall with $\alpha = 1\%$ based on the "best" volatility forecast (solid line) and conditional expected shortfall computed with GARCH(1,1) predictions (dashed line) superimposed on the returns of the exchange rate US$/SFr.

6.3 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:

$$N(\mathcal{M}) \sim Bin(m, \alpha), \tag{35}$$

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.

Table 6 indicates an advantage for the method with $\widehat{\sigma}_{BestPrediction}$: 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.

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

Table 6: Violations of the 1%-value at risk and p-values of the two-sided binomial test.

In a backtest for the conditional expected shortfall, we analyze

$$\widehat{r}_t = \frac{R_t - S_t}{\widehat{\sigma}_t} \quad \text{for } t \in \mathcal{T} \text{ with } R_t < \widehat{q}_t, \tag{36}$$

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

$$\widehat{r}_t \approx \varepsilon_t - E_{t_{\widehat{\nu}}}[\varepsilon_t \,|\, \varepsilon_t < t_{\widehat{\nu}}^{-1}(\alpha)], \tag{37}$$

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

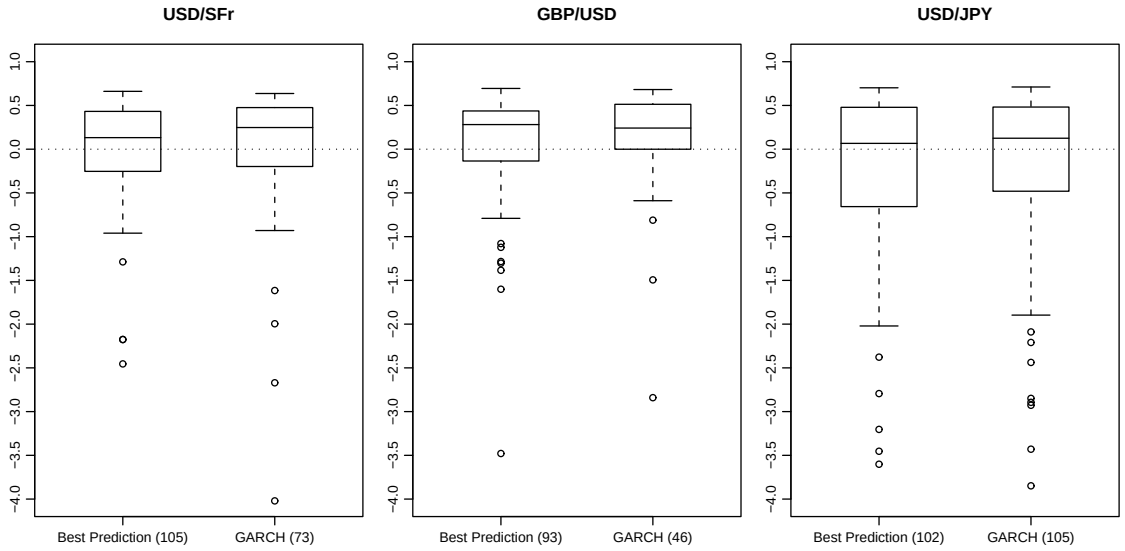

Figure 6: Boxplots for the empirical standardized residuals $\widehat{r}_t$ between the true return and the 5%-conditional expected shortfall in case of quantile violation. The numbers in parentheses are the number of quantile violations.

residuals $\widehat{r}_t$ as well as the ones of their squares $\widehat{r}_t^2$ 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.

7 Conclusions

We construct an accurate forecasting procedure for the volatility based on high frequency data and we quantify possible gains for predicting the true latent 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 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 much more accurate for predicting the auxiliary realized volatility than GARCH methods based on low frequency, daily data. More importantly, a novel statistical test also yields in all three datasets significantly better predictions with high frequency data for the principal target, namely the true underlying latent volatility. Our new and fresh look at evaluating performances supports that high frequency data methods perform better than low frequency techniques. But the gains are not so huge as reported in previous works (see e.g. Andersen et al., 2002) where only accuracy for the auxiliary realized volatilities has been considered.

- One of the best forecasting procedures for the volatility is derived from detrending the log-transformed realized volatilities $log(\sigma_{t;RV})$ by exponential smoothing, and forecasting the residual process $(Z_t)_t$ with an AR model. Surprisingly, nonlinear and sophisticated models for the filtered residual process $(Z_t)_t$ did not improve the accuracy of the forecast, even though their model specifications were chosen to optimize the outsample performance, whereas the model order p of the AR was determined via AIC on the learning set $\mathcal{L}$ only. We consider this as empirical evidence for linearity of the exponentially smoothed residual process $(Z_t)_t$. The alternative ARFIMA long-range dependence model for log-realized volatilities $\log(\sigma_{t;RV})$ yields a very similar performance and is thus an equally good tool for volatility prediction.

- 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.

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., Bollerslev, T., Diebold, F.X. and Labys, P. (Revised 2002). Modeling and Forecasting Realized Volatility. Econometrica , forthcoming.

[3] Artzner, P., Delbaen, F., Eber, J., Heath, D. (1999). Coherent Measures of Risk. Mathematical Finance 9, 203-228.

[4] Bickel, P.J. and Bühlmann, P. (1996). What is a Linear Process? Proceedings of the National Academy of Sciences USA 93, 12128-12131

[5] Bühlmann, P. (1996). Locally Adaptive Lag-Window Spectral Estimation. Journal of Time Series Analysis 17, 247-270.

[6] Bühlmann, P. and Ferrari, F. (2002). Dynamic Combination of Models. Preprint, ETH Zürich.

[7] Brockwell, P.J. and Davis, R.A. (1991). Time Series. Theory and Methods (Second Edition), Springer.

[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] Gijbels, I., Pope, A. and Wand, M.P. (1999). Understanding Exponential Smoothing via Kernel Regression. Journal of the Royal Statistical Society, Series B, 61, 39–50.

[10] Künsch H.R. (1986). Discrimination Between Monotonic Trends and Long-Range Dependence. Journal of Applied Probability 23, 1025-1030.

[11] 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.

[12] Merton, R.C. (1980). On Estimating the Expected Return on the Market: An Exploratory Investigation. Journal of Financial Economics, 8, 323-361

[13] Officer, R.R. (1973). The Variability of the Market Factor of the NYSE. Journal of Business, 46, 434-453.

[14] Ripley, B.D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge.

[15] Schwert, G.W. (1998). Stock Market Volatility: Ten Years After the Crash. Brookings-Wharton Papers on Financial Services, I, 65-114.

[16] Taylor, S.J. and Xu, X. (1997). The Incremental Volatility Information in One Million Foreign Exchange Quotations. Journal of Empirical Finance, 4, 317-340.