

Gürtler, Marc; Kreiss, Jens-Peter; Rauh, Ronald

Working Paper

A non-stationary approach for financial returns with nonparametric heteroscedasticity

Working Paper Series, No. IF31V2

Provided in Cooperation with:

Technische Universität Braunschweig, Institute of Finance

Suggested Citation: Gürtler, Marc; Kreiss, Jens-Peter; Rauh, Ronald (2009) : A non-stationary approach for financial returns with nonparametric heteroscedasticity, Working Paper Series, No. IF31V2, Technische Universität Braunschweig, Institut für Finanzwirtschaft, Braunschweig

This Version is available at:

<https://hdl.handle.net/10419/55240>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Working Paper Series

A Non-stationary Approach for Financial Returns with Nonparametric Heteroscedasticity

by Marc Gürtler, Jens-Peter Kreiss, and Ronald Rauh

No.: IF31V2/09
First Draft: 2009-04-28
This Version: 2009-09-17

Technische Universität Braunschweig
Institute of Finance
Abt-Jerusalem-Str. 7
D-38106 Braunschweig

A Non-stationary Approach for Financial Returns with Nonparametric Heteroscedasticity

by Marc Gürtler^{*}, Jens-Peter Kreiss^{**}, and Ronald Rauh^{***}

Abstract. A non-stationary regression model for financial returns is examined theoretically in this paper. Volatility dynamics are modelled both exogenously and deterministic, captured by a nonparametric curve estimation on equidistant centered returns. We prove consistency and asymptotic normality of a symmetric variance estimator and of a one-sided variance estimator analytically, and derive remarks on the bandwidth decision. Further attention is paid to asymmetry and heavy tails of the return distribution, implemented by an asymmetric version of the Pearson type VII distribution for random innovations. By providing a method of moments for its parameter estimation and a connection to the Student-t distribution we offer the framework for a factor-based VaR approach. The approximation quality of the non-stationary model is supported by simulation studies.

Keywords: heteroscedastic asset returns, non-stationarity, nonparametric regression, volatility, innovation modelling, asymmetric heavy-tails, distributional forecast, Value at Risk (VaR)

JEL classification: C14, C51, C52

^{*} **Professor Dr. Marc Gürtler**
Technische Universität Braunschweig
Institute of Finance
Abt-Jerusalem-Str. 7
D-38106 Braunschweig
Germany
Fon: +49 531 3912895
Fax: +49 531 3912899
E-mail: marc.guertler@tu-bs.de

^{**} **Professor Dr. Jens-Peter Kreiss**
Technische Universität Braunschweig
Institute of Mathematical Stochastics
Pockelsstr. 14
D-38106 Braunschweig
Germany
Fon: +49 531 3917566
Fax: +49 531 3917564
E-mail: j.kreiss@tu-bs.de

^{***} **Dipl.-Math. Oec. Ronald Rauh**
Öffentliche Versicherung Braunschweig
Department of Asset Risk Management
Theodor-Heuss-Str. 10
D-38122 Braunschweig
Germany
Fon: +49 531 2021568
Fax: +49 531 202331568
E-mail: ronald.rauh@oeffentliche.de

1 Introduction

Accurate modelling of financial time series is a fundamental task for investment decisions and the risk management of banks and insurance companies. In the course of regulatory developments by the Basel Committee on Banking Supervision and the Committee of European Insurance and Occupational Pensions Supervisors the *Value at Risk* (henceforth VaR) has become a benchmark risk measure for capital requirements. From the two main approaches to VaR measurement, *full valuation* and *local-valuation*, mentioned in Jorion (2006), the latter got famous in the form of the *delta-normal method*, that works with a first order Taylor approximation for the market value, e.g. the basis point value for fixed income. This approach, also named *variance-covariance model* or *parametric VaR*, assumes iid normal distributed benchmark returns that are mapped to the 'delta'-conception of exposures. The aim is a factor-based model where exposures are stressed by empirical standard deviations or risk factors to univariate risks, that are aggregated by correlation matrix multiplication to a portfolio VaR. As part of a companion paper, Gürtler and Rauh (2009) test that modelling assumptions for a multitude of financial time series carefully.¹ They have to be rejected in the majority of hypothesis tests, especially on a daily data base: The normality of the return distribution is denied as well as the serial identity and independence. The latter results query volatility measurement via empirical standard deviations and risk measurement in terms of (normal) quantiles entirely. Instead, modelling more general time-evolving return distributions with a focus on heteroscedasticity and querying the stationarity may be the ways to success.

These results are in line with the early analyses of Mandelbrot (1963) and Fama (1965) on several US stock series. They draw the following empirical conclusions for return series, often labeled as *stylized facts*:

- there is a serial data dependence,
- the volatility is changing over time (heteroscedasticity),
- the returns are asymmetrically distributed with heavy tails,
- a negative return amplitude entails a greater volatility than a positive return of same amount (Leverage Effects).

Let $\{P_t\}_{t=0,\dots,n}$ be the market value of an asset over n periods of time, and let $X_t = \ln P_t - \ln P_{t-1}$ ($t = 1, \dots, n$) be the corresponding series of *log-returns*. Evidence on the first stylized fact is often brought by correlograms, that are graphs of sample autocorrelations $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$ (SACF, where $\gamma(h)$ is the empirical estimate for $Cov(X_t, X_{t+h})$) versus discrete time lags $h \geq 0$ for a stationary return series $\{X_t\}_t$. Taylor (1986), Ding et al. (1993), Drees and Starica (2002), Straumann (2004), Mikosch and Starica (2004) and Herzel et al. (2005) came to the same typical picture for several financial instruments: Daily returns themselves contain little serial correlation, but their absolute values and their squares are significantly autocorrelated, with $\rho_{|X_t|}(h)$ or $\rho_{X_t^2}(h)$ being positive and decaying slowly over a large number of lags.² This effect is called *long range dependence* (LRD) and one might draw the conclusion of serial dependent time series.

On the ideas of serial dependence and conditional time-varying volatility Robert F. Engle and Tim Bollerslev developed in the 1980s nonlinear time series models as *autoregressive conditional heteroscedastic* (ARCH-processes). The general form is

$$X_t = \mu_t + \varsigma_t \varepsilon_t, \quad t \in \mathbb{Z}, \quad (1)$$

where $\{\varepsilon_t\}_t$ is an iid sequence with $E\varepsilon_1 = 0$ and $Var\varepsilon_1 = 1$, $\{\mu_t\}$ and $\{\varsigma_t\}$ are stochastic processes that depend only on past information, i.e. $\mu_{t+1}, \varsigma_{t+1}$ are measurable with respect to the σ -field $\mathcal{F}_t = \sigma(\{\varepsilon_j \mid j \leq t\})$. Hence X_t is \mathcal{F}_t -measurable for every t and $\mu_{t+1} = E(X_{t+1} \mid \mathcal{F}_t)$ is the conditional mean and $\varsigma_{t+1}^2 = Var(X_{t+1} \mid \mathcal{F}_t)$ is the conditional variance of X_{t+1} given the past returns. The starting point was set by the seminal paper of Engle (1982) who introduced the ARCH(p)-process, where conditional volatility dynamics $\{\varsigma_t\}_t$ are imposed via linear regression over past squared (centered) returns. An extension of Bollerslev (1986), named generalized ARCH (GARCH(p, q)-process), includes past variances next to historical returns into the parametric regressed volatility process, what enables a parsimonious parametrisation with a reasonable fit to empirical data. For purposes of parameter estimation, processes

¹The empirical studies are executed for 29 benchmark series as equity indices, interest rates, credit spreads and exchange rates.

²Further examples are provided in Gürtler and Rauh (2009). Ding et al. (1993) extended the result to power transformations of absolute returns, $|X_t|^d$ for $d > 0$, and showed them to be 'long-memory' for a S&P 500 return series, with quite high autocorrelations for large lags. Furthermore, they found that for a fixed lag τ the function $\rho_\tau(d) = Corr(|X_t|^d, |X_{t+\tau}|^d)$ has a unique maximum when $d \approx 1$.

of the ARCH-family are defined to be stationary, but finding conditions for the existence and uniqueness of a stationary solution was nontrivial.³ Besides the focus on heteroscedasticity and uncorrelated, serial dependence, the features of heavy-tailedness and asymmetry are mostly imposed on the distribution of innovations ε_t .⁴

Coming back to the LRD, Mikosch and Starica (2004) derive theoretically that the aforementioned SACF for absolute returns could alternatively arise from non-stationarities in the data. Correlograms are only a significant tool for detecting dependence under the assumption of stationarity, otherwise structural breaks in the data as shifts in the variance might cause identical results,⁵ and it cannot be discriminated between stationary, long memory and non-stationary time series. Secondly, Granger and Starica (2005) came by a long-term case study on the S&P 500 index to the conclusion, that the second interpretation is the more probable: The main reason of the example LRD is to be seen in the non-stationarities due to structural breaks of the unconditioned variance. From this follows a grave criticism on ARCH-type models, since they are parameterized as stationary processes (what implies a fortiori a fixed unconditional variance) and focus on modelling a long-range dependence structure of the second moments.

Another observation on ARCH modelling arises over longer periods of daily returns: The typical outcome of a GARCH(1, 1) implementation is that the sum of estimated parameters α_1, β_1 is approximately 1, leading to an IGARCH(1, 1) model, which is referred to as *IGARCH effect* in Starica (2003) or Mikosch and Starica (2004). But IGARCH-processes imply an infinite variance of the random variables, which contradicts to the results of a direct tail analysis, that indicate a finite second moment for daily financial returns (see De Haan et al. (1994)). Mikosch and Starica (2004) prove theoretically and empirically that the IGARCH effect may be generated by non-stationarities via shifts in the unconditional variance of the return series.⁶ Although some ARCH-extensions allow structural breaks in the volatility structure while holding up stationarity,⁷ one should question the stationarity assumption at all.

More and more sophisticated ARCH-type processes were developed to improve the characterisation of volatility and of single aspects in the return dynamics;⁸ Bollerslev et al. (1994) give a statistical overview of model extensions. Following Drees and Starica (2002), the need of an increasing complexity for volatility modelling can be possibly explained that a simple endogenous specification does not exist. In that case, the model fit can only be improved by a change of the working hypothesis: In their univariate approach the volatility is supposed to be exogenous to the return process. The evolution of market prices is interpreted as a manifestation of complex market conditions, driven by unknown exogenous factors, that evolve smoothly through time and that are expressed mainly in the level of unconditional variance of recent returns. Continuing with the multiplicative model (1) for log-return series $\{X_t\}_{t=1, \dots, n}$ we replace the volatility term ς_t by an unconditional variable $\sigma(t)$. The corresponding variance $\sigma(t)^2$ is modelled as a smooth, deterministic function of time via a nonparametric kernel regression⁹ over centered, squared returns R_t :

$$\hat{\sigma}^2(t) = \frac{\sum_{i=1}^n K_h(i-t) R_i^2}{\sum_{i=1}^n K_h(i-t)}, \quad t = 1, \dots, n. \quad (2)$$

The approach preserves the independence assumption for financial returns (unlike the LRD suggestion), but it abandons the hypothesis of stationarity, motivated in the expertise of Mikosch and Starica (2004) or Granger and Starica (2005) that the paradigm of time-varying unconditional variance is superior over some specifications of stationary long memory in explaining the return dynamics. The volatility estimates $\hat{\sigma}^2(t)$ may be employed to forecast close future returns since they are supposed to be within (nearly) the same market environment. Former attention is paid to asymmetry and heavy tails in the return distribution, involved by the features of random residuals ε_t . An accurate description is achieved by fitting the Pearson type VII distribution to positive and negative innovations separately, which enables a simple but flexible modelling. Drees and Starica (2002) show on a 12-year S&P 500 example that their non-stationary model fits the data adequately and gets better short-term forecasts on the return distribution than conventional GARCH models. Herzel et al. (2005) extend these ideas to a multivariate non-stationary framework. Vectors

³Bougerol and Picard (1992) gave the solution for stationary GARCH(p, q) processes via stochastic recurrence equations. Straumann (2004) summarizes this among other stochastic features on GARCH processes.

⁴The exponential GARCH (EGARCH) model of Nelson (1991) and the asymmetric power GARCH (APGARCH) model by Ding et al. (1993) include asymmetry and leverage effects directly in the volatility dynamics, that reacts different to financial gains and losses.

⁵The authors found that the stronger the non-stationarity, e.g. the difference of the variation of subsamples $X^{(1)}$ and $X^{(2)}$ measured as $(E|X^{(1)}| - E|X^{(2)}|)^2$, the more pronounced the LRD effect. This theoretical result is supplemented with a long-term empirical study of the S&P 500 index by ex-/ including the 1970s US-recession, generating the LRD effect.

⁶Concerning persistence in variance and long memory (IGARCH- and LRD effect) caused by structural changes, see also Lamoureux and Lastrapes (1990) or Diebold and Inoue (2001).

⁷E.g. regime-switching ARCH models by Hamilton and Susmel (1994) with transitions governed by an unobserved, fixed Markov chain.

⁸On the other hand stochastic volatility models specialized substantial to find endogenous (co-)variance descriptions for financial instruments.

⁹The rescaled kernel function K_h is defined as $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ with bandwidth h and an appropriate kernel K on a compact support $[-1, 1]$.

of financial returns are assumed to have a time-varying unconditional covariance matrix that evolves smoothly, captured by classical nonparametric regression. Vectors of standardized innovations are modelled again parametrically with an asymmetric version of Pearson VII. Another univariate extension with a time-varying expected return was introduced by Mikosch and Starica (2004a).

In this article, we will directly refer to the approaches of Herzel et al. (2005) and Drees and Starica (2002). After presenting the steps for fitting the non-stationary model to empirical return data in general, we specialize on the univariate description. The article is thought as the theoretical grounding for the companion paper Gürtler and Rauh (2009), where the non-stationary approach is applied to a multitude of financial time series. Here we focus on deriving the modelling components as the nonparametric volatility estimate (2) and its one-sided equivalent, which is based only on past/ recent data and should be used in forecasting. A lot of effort is spent to prove the consistency and asymptotic normality of both variance estimators analytically in a transformed version of the nonparametric regression model with an equidistant unit design. Remarks on the kernel selection and the optimal choice of bandwidths follow, a cross-validation setup is developed. The task of fitting innovations via the Pearson type VII distribution is simplified by providing a method of moments for parameter estimation. Moreover a connection to the Student- t distribution is derived, where the idea of a factor-based implementation of the non-stationary regression model for the purpose of VaR calculation emerged from: The univariate $VaR^{1-\alpha}(t)$ of an exposure $w(t)$ will be modelled as its product with the nonparametric estimated volatility $\hat{\sigma}_{(\cdot)}(t)$ and the $(1-\alpha)$ Pearson VII innovation quantile of the benchmark return distribution, adjusted by the mean return.

Our theoretical results are confirmed by simulation studies: We document for a predefined volatility function that the approximation of the two-sided and one-sided variance estimates improves with a more and more refined data base. Moreover we prove the non-stationary model being able to capture prices processes entirely and to produce good distributional forecasts. A Kupiec hypothesis testing evaluates forecasts in terms of predicted VaR limits $VaR^{1-\alpha,1d}(t)$ and their shortfall rates concerning the next day's returns X_{t+1} . We will show with that backtesting method that the non-stationary model significantly outperforms the parametric (Gaussian) VaR model. Furthermore, Gürtler and Rauh (2009) fit the regression model to daily return series of stock indices, exchange rates, interest rates and credit spreads and provide evidence for the empirical validity of that approach and its outperformance to the standard VaR model.

The rest of the paper is organized as follows: In section 2 we introduce the non-stationary model for asset returns formally. We delve into the nonparametric estimation of volatility theoretically, and focus on the procedure of fitting innovations. In section 3 we provide some simulation studies with the non-stationary approach, that underline the theoretical results and motivate a practical application in risk management. We conclude in section 4.

2 A non-stationary model for asset returns

As motivated in section 1, our following return model adopts the conceptual framework of Herzel et al. (2005) and Drees and Starica (2002) (univariate case) for analysing the return dynamics via classical nonparametric regression with fixed equidistant design points. The vectors of financial returns are assumed to have a time-varying unconditional covariance matrix that evolves smoothly through time and is modelled exogenously deterministic. The standardized residuals are modelled parametrically, allowing for asymmetry and heavy tails. This leads to a multiplicative approach, with a constant mean return μ added, for a non-stationary sequence of independent random vectors $\{\mathbf{X}_t\}_t$:

$$\begin{aligned} \mathbf{X}_t &= \mu + \mathbf{S}_t \varepsilon_t, \quad t = 1, \dots, n & (3) \\ \varepsilon_1, \dots, \varepsilon_n &\text{ iid random vectors with mutually independent coordinates,} \\ E\varepsilon_{k,1} &= 0, Var\varepsilon_{k,1} = 1, \quad \forall k = 1, \dots, d, \\ \mathbf{S}_t &: [0, n] \rightarrow \mathbb{R}^{d \times d} \text{ is an invertible matrix and a smooth function of time.} \end{aligned}$$

We emphasise that this regression-type model does not exclude random effects of the volatility dynamics. The basic idea is that recent past and the next future returns depend on the same unknown exogenous economic factors, that evolve gradually through time. Those factors are included in the recent asset returns and imply the level of the unconditional variance and the unconditional covariance structure. The aim is to estimate the multivariate return dynamics only by dint of recent returns and to build up short-term forecasts of future return distributions in a similar economic environment. Three steps have to be arranged to fit the regression-model to a financial time series:

1. Centering returns

The demeaned return series $\{\mathbf{R}_t\}_t$ is defined as

$$\mathbf{R}_t = \mathbf{X}_t - \bar{\mathbf{X}}_n, \quad t = 1, \dots, n, \quad (4)$$

with column vectors centered componentwise by the empirical mean of the whole series, $\bar{X}_{k,n} = \frac{1}{n} \sum_{t=1}^n X_{k,t}$ for all $k = 1, \dots, d$. Neglecting the estimation error of $\bar{\mathbf{X}}_t$ for μ (i.e. assumption: $\bar{\mathbf{X}}_t = \mu \implies \mathbf{R}_t = \mathbf{S}_t \varepsilon_t$) it follows that:

$$\begin{aligned} E(\mathbf{R}_t \mid \mathbf{R}_{t-1}, \mathbf{R}_{t-2}, \dots) &= E\mathbf{R}_t = \mathbf{S}_t E\varepsilon_t = 0 \\ E(\mathbf{R}_t \mathbf{R}_t' \mid \mathbf{R}_{t-1}, \mathbf{R}_{t-2}, \dots) &= E(\mathbf{R}_t \mathbf{R}_t') = \mathbf{S}_t \mathbf{S}_t' =: \Sigma^2(t) \end{aligned}$$

Hence, $\{\mathbf{R}_t \mathbf{R}_t'\}_t$ is an independent sequence of matrices with pointwise expectations $\Sigma^2(t)$, a smooth function of time. This is the framework for a nonparametric regression with equidistant design points $t = \{1, \dots, n\}$, where variances and covariances in matrices $\{\Sigma^2(t)\}_t$ may be estimated by kernel estimators applied to the sequence $\{\mathbf{R}_t \mathbf{R}_t'\}_t$.

2. Estimating volatilities

In any nonparametric regression the conditioned expectation of a variable Y relative to X can be written as

$$m(x) = E[Y \mid X = x],$$

where m is an unknown function with certain regularity conditions. We estimate m with a *local polynomial regression method* (Taylor approximation), i.e.

$$m(x) \approx \sum_{k=0}^p \beta_k (x - x_0)^k, \quad \beta_k := \frac{m^{(k)}(x_0)}{k!}$$

in a neighbourhood $U(x_0)$ of x_0 with m being p times differentiable, applying the *method of least squares* for the sample $(X_1, Y_1), \dots, (X_n, Y_n)$:

$$\sum_{i=1}^n \{Y_i - m(X_i)\}^2 \approx \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p \beta_k (X_i - x_0)^k \right\}^2 \xrightarrow{!} \min_{\beta_0, \dots, \beta_p}$$

Since this expression is only locally valid, the data is localized by weighting the summands with respect to the distance $|X_i - x_0|$ using symmetric kernel functions K and K_h on compact supports $[-1, 1]$ and $[-h, h]$, respectively:

$$K : \mathbb{R} \rightarrow [0, \infty) \text{ with } \int_{-\infty}^{\infty} K(u) du = \int_{-1}^1 K(u) du = 1, \quad K_h(u) := \frac{1}{h} K\left(\frac{u}{h}\right),$$

denoting K_h as rescaled kernel on bandwidth h . The minimization problem is modified to:

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p \beta_k (X_i - x_0)^k \right\}^2 K_h(X_i - x_0) \xrightarrow{!} \min_{\beta_0, \dots, \beta_p}$$

Moreover the degree of the regression polynomial is simplified to $p = 0$, so that m is approximated locally with a constant β_0 (*local constant regression*):¹⁰

$$\begin{aligned} &\sum_{i=1}^n \{Y_i - \beta_0\}^2 K_h(X_i - x_0) \xrightarrow{!} \min_{\beta_0} \\ \implies \hat{\beta}_0 &= \frac{\sum_{i=1}^n Y_i K_h(X_i - x_0)}{\sum_{i=1}^n K_h(X_i - x_0)} \text{ is the LS-estimator} \end{aligned}$$

The hereafter definition follows with $\hat{\beta}_0 = \hat{m}^{(0)}(x_0) = \hat{m}(x)$, $x \in U(x_0)$.

¹⁰For a symmetrical design around x_0 a local linear regression ($p = 1$) would achieve the same estimator, which is generally the case for the two-sided volatility estimation. Else an additive correcting term will improve the estimator.

Definition 2.1

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n K_h(X_i - x) Y_i}{\sum_{i=1}^n K_h(X_i - x)}, \quad (5)$$

where $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ with $h > 0$ and $K : \mathbb{R} \rightarrow [0, \infty)$, $x \mapsto K(x)$ symmetrically and $\int_{-\infty}^{\infty} K(u) du = 1$, is called *Nadaraya-Watson kernel estimator*.

Kernel estimators are linear smoothers, $m(x)$ is estimated via a weighted local average of observations Y_1, \dots, Y_n .

Applied to the regression model (3) the two-sided Nadaraya-Watson estimator (short: NWE) has the form:

$$\hat{\Sigma}^2(t) = \frac{\sum_{i=1}^n K_h(i-t) \mathbf{R}_i \mathbf{R}_i'}{\sum_{i=1}^n K_h(i-t)} \quad (6)$$

with $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$, where K is an appropriate kernel as defined in the sequel. Herzel et al. (2005) motivate the application of nonparametric regression by theoretical results of Müller and Stadtmüller (1987) in an asymptotic context (compare section 3.2 in Herzel et al. (2005)). That way, they additionally derive propositions on confidence intervals for $(\hat{\Sigma}_{i,j}(t))_{i,j}$. We restrict ourselves later to the univariate case and proof some common, useful statistical properties. Moreover, a one-sided volatility estimator for forecasting purposes will be introduced and analyzed.

3. Fitting innovations

In the last step we have to focus on the distribution of innovations $\{\varepsilon_t\}_t$. In the multivariate and the univariate context the approach for fitting innovations may be equal because of their mutual independent coordinates. Firstly, the innovations $\varepsilon_{k,t}$ are componentwise estimated by dint of demeaned returns $R_{k,t}$ and estimated volatilities $\hat{\sigma}_k(t) = \left(\hat{\Sigma}_{k,k}(t)\right)_{i,j}$ (the k -th diagonal element of the square root of the estimate $\hat{\Sigma}^2(t)$ for $\mathbf{S}_t \mathbf{S}_t'$) in each point of time t :

$$\hat{\varepsilon}_{k,t} = \frac{R_{k,t}}{\hat{\sigma}_k(t)}, \quad t = 1, \dots, n \quad (7)$$

Due to their independence it is sufficient to specify the distributions of $\hat{\varepsilon}_{k,t}$, $k = 1, \dots, d$ univariately. The easiest approach without any extra assumption could be the empirical distribution function $\hat{F}_n^{emp}(x)$ of the series $\{\hat{\varepsilon}_t\}_t$, but this is not able to capture heavy tails.¹¹ The normal distribution drops out due to neglecting skewness and heavy tails. Herzel et al. (2005) as well as Drees and Starica (2002) found the *Pearson type VII distribution* to be a flexible and parsimonious family of (heavy-tailed) distributions.¹² It has the following one-sided density with shape parameter m and scale parameter c :

$$f_{m,c}^{VII(1)}(x) = \frac{2\Gamma(m)}{c\Gamma\left(m - \frac{1}{2}\right)\pi^{1/2}} \left(1 + \left(\frac{x}{c}\right)^2\right)^{-m} I_{[0,\infty)}(x) \quad (8)$$

The one-sided Pearson VII presentation, concentrated on the positive axis, was chosen to allow for asymmetry: It is fitted separately to nonnegative innovations $\{\hat{\varepsilon}_t \mid \hat{\varepsilon}_t \geq 0\}_t$ and absolute values of negative innovations $\{-\hat{\varepsilon}_t \mid \hat{\varepsilon}_t < 0\}_t$. In the above mentioned literature the corresponding parameters (m_+, c_+) and (m_-, c_-) are estimated with maximum-likelihood methods. We solve that task with a method of moments, later. Because usually there are about as many positive as negative innovations in a financial time series, it may be assumed that the median of innovations is 0. Hence, the fitted one-sided Pearson type VII densities $f_{m_+,c_+}^{VII(1)}$ and $f_{m_-,c_-}^{VII(1)}$ are combined as

$$f_{m_+,c_+,m_-,c_-}^{VII}(x) = \frac{1}{2} \left(f_{m_-,c_-}^{VII(1)}(-x) I_{(-\infty,0)}(x) + f_{m_+,c_+}^{VII(1)}(x) I_{[0,\infty)}(x) \right), \quad (9)$$

its cdf is referred as $\hat{F}^{VII}(x)$ and called *asymmetric Pearson type VII distribution* (Drees and Starica (2002)) of random innovations ε_t .¹³

¹¹Being $x_{max} = \max_t \hat{\varepsilon}_t$ and $x_{min} = \min_t \hat{\varepsilon}_t$ of the innovation sample $\{\hat{\varepsilon}_t\}_{t=1,\dots,n}$, then $\hat{F}_n^{emp}(x_{max} + \delta) = 1$ for all $\delta \geq 0$ and $\hat{F}_n^{emp}(x_{min} - \epsilon) = 0$ for all $\epsilon > 0$. Consequently the probability for extreme future innovations $\varepsilon_N \notin [x_{min}, x_{max}]$, $N > n$ would equal 0. The empirical distribution function underestimates the extremes, which is an unacceptable shortcoming for risk management purposes.

¹²The Pearson VII family includes the t -distribution, the Cauchy distribution and (asymptotically) the Gaussian distribution. Other applications are e.g. to be seen in Kitagawa and Nagahara (1999) for standardized innovations in a stochastic volatility (state-space) model.

¹³An alternative approach for fitting innovations could be kernel density estimation, to stay completely within a nonparametric world.

Concluding, the estimated distributions $\hat{F}_{\varepsilon_k}^{VII}$ of the d independent random innovations together with (the square root of) the covariance matrix estimates $\hat{\Sigma}^2(t)$ and the mean vector $\bar{\mathbf{X}}_n$ completely specify the distribution of returns \mathbf{X}_t in the regression model (3).

From now on we specialize on the univariate non-stationary model, as introduced in Drees and Starica (2002):

$$\begin{aligned} X_t &= \mu + \sigma(t)\varepsilon_t, \quad t = 1, \dots, n, \\ \varepsilon_1, \dots, \varepsilon_n &\text{ iid with } E\varepsilon_1 = 0, \text{ Var } \varepsilon_1 = 1, \\ \sigma(t), \quad t &= 1, \dots, n, \text{ a smooth, deterministic function of time.} \end{aligned} \quad (10)$$

The series $\{X_t\}_{t=1, \dots, n}$ of log-returns or diff-returns¹⁴ preserves the independence assumption, but a time evolution of the return distribution is incorporated via the unconditional heteroscedasticity. Asymptotic and heavy tailed random innovations, modelled as Pearson-type VII distributed rvs, enable a flexible capturing of return series, leading together to an enhanced multiplicative approach.

2.1 Focus on univariate nonparametric volatility estimation

After demeaning the return series $\{X_t\}_{t=1, \dots, n}$ in the univariate regression model (10) we base the nonparametric regression on equidistant points t of squared returns $\{R_t\}_t$ with the methods developed above. The Nadaraya-Watson estimator for volatility estimation is obtained as:

(I) Two-sided NWE (smoother):

$$\hat{\sigma}^2(t) := \frac{\sum_{i=1}^n K_h(i-t)R_i^2}{\sum_{i=1}^n K_h(i-t)} \quad (11)$$

with $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ for an appropriate kernel K on a compact support $[-1, 1]$.

(II) One-sided NWE (filter):

$$\hat{\sigma}_{(1)}^2(t) := \frac{\sum_{i=1}^t K_h(i-t)\tilde{R}_i^2}{\sum_{i=1}^t K_h(i-t)} \quad (12)$$

with $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ and $\tilde{R}_i := X_i - \bar{X}_{i-1}$.

The first estimator immediately follows from equation (6). An asymmetric version is supplemented, that includes only past and current returns and that is consistently applicable to estimate the volatility at the last observed point n and to forecast close future returns. The distinction between both estimates is fundamentally: While the first estimation of $\sigma^2(t)$ depends on a symmetrical data base around t , using all returns R_i that are temporally close enough in a band $[t-h, t+h]$, the asymmetric estimator $\hat{\sigma}_{(1)}^2(t)$ inserts (past centered) returns \tilde{R}_i up to time t , restricted to values within the left-sided band $[t-h, t]$.¹⁵ While a symmetric kernel will be used to describe the dynamics of changes in the historical sample, the one-sided estimation is applied in forecasting volatility. Of course, the one-sided estimator delivers in a historical sample generally a bigger estimation error (with a bigger variance) than its two-sided counterpart due to the lack of information.

For deriving statistical results as consistency or asymptotic normality of our volatility estimates, we have to define an appropriate asymptotic framework, firstly. The following asymptotic does not only involve an increasing number of

¹⁴We define diff-returns as differences of prices, $X_t = P_t - P_{t-1}$ ($t = 1, \dots, n$), which are needed in Gürtler and Rauh (2009) as corresponding return type to interest rate- and credit-spread-exposures, measured as basis point values. The model and statistical features given below work for diff-returns as well as for log-returns. The asymptotical results use the asset return series as input but do not have any conditions on the return type.

¹⁵Values outside this radius are assumed not to influence $\sigma(t)$ and are zero-weighted. Nevertheless it requires a sufficient long sample: the calculation is only possible if $1 \leq t-h$ and $t+h \leq n$ (last condition only for symmetrical case). Else the required band would go beyond the data base and boundary effects occur. Several approaches for treating the boundary $t \in [0, h]$ and $t \in (n-h, n]$ of a regression region exist in the literature of nonparametric curve estimation. Fan and Yao (2003) list e.g. special *boundary kernels*, methods of *reflection* and *transformation* or local polynomial fitting of a higher degree. In general, the order of magnitude of the bias is different in the interior and near the boundaries. This is to be seen in the subsequent analysis as the optimal two-sided (interior) bandwidth is of order $n^{4/5}$ while the optimal bandwidth of the one-(left)-sided bandwidth has size $n^{2/3}$, that could be interpreted as a boundary corrected estimator for the right interval boundary.

observations but even more an increase of the frequency for observing data points on a fixed time-frame. Analytically, we rescale the observations to the interval $[0, 1]$ and refine the data base gradually. That way, an increase of sample size n means including more observations on closer design points, e.g. an increase of information via switching from monthly to daily returns within the predefined horizon. Henceforth we scan the (unknown) regression function more and more precisely, until it is infinitesimal granulated.

We assume that beyond the series of returns $\{X_t\}_{t=1,\dots,n}$ the discrete sequence $\{\sigma(t)\}_t$ is gathered from a continuous volatility function:

$$\sigma : [0, n] \longrightarrow \mathbb{R}_0^+, x \mapsto \sigma(x) \quad (13)$$

The aforementioned asymptotic is implemented by transforming the data series to the standardized window $[0, 1]$ with design points $\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$, that adopts the volatility values as $s\left(\frac{t}{n}\right) = \sigma(t)$ for all $t = 1, \dots, n$. Formally:

$$s : [0, 1] \longrightarrow \mathbb{R}_0^+, y \mapsto s(y) := \sigma(ny) \quad (14)$$

This produces the transformed multiplicative return model:

$$\begin{aligned} X_{t,n} &= \mu + s\left(\frac{t}{n}\right) \varepsilon_{t,n}, \quad t = 1, \dots, n, \\ \varepsilon_{1,n}, \dots, \varepsilon_{n,n} &\text{ iid with } E\varepsilon_{1,n} = 0, \text{ Var } \varepsilon_{1,n} = 1, \\ &s\left(\frac{t}{n}\right), \quad t = 1, \dots, n, \text{ a smooth, deterministic function of time.} \end{aligned} \quad (15)$$

The corresponding nonparametric estimators are:

(I) Two-sided transformed NWE:

$$\hat{s}_{h_n}^2(u) := \frac{\sum_{i=1}^n K_{h_n}\left(\frac{i}{n} - u\right) R_{i,n}^2}{\sum_{i=1}^n K_{h_n}\left(\frac{i}{n} - u\right)}, \quad (16)$$

with $u \in [0, 1]$, $h_n := h(n)$ and $K_{h_n}(\cdot) = \frac{1}{h_n} K\left(\frac{\cdot}{h_n}\right)$.

(II) One-sided transformed NWE:

$$\hat{s}_{(1)h_n}^2(u) := \frac{\sum_{i=1}^{\lfloor un \rfloor} K_{h_n}\left(\frac{i}{n} - u\right) \tilde{R}_{i,n}^2}{\sum_{i=1}^{\lfloor un \rfloor} K_{h_n}\left(\frac{i}{n} - u\right)}, \quad (17)$$

with $u \in [0, 1]$, $h_n := h(n)$, $K_{h_n}(\cdot) = \frac{1}{h_n} K\left(\frac{\cdot}{h_n}\right)$ and $\tilde{R}_{i,n} = X_{i,n} - \bar{X}_{i-1}$.

To reach the desired statistical properties of estimators $\hat{s}_{h_n}^2(u)$ and $\hat{s}_{(1)h_n}^2(u)$ the following conditions on the kernel K , the bandwidth h_n and the smoothness of the (transformed) volatility function $s(\cdot)$ are required:

(C1) Let $K : \mathbb{R} \rightarrow [0, \infty)$ be a symmetrical density with compact support $[-1, 1]$, i.e. the kernel has the following features: (i) $K(v) = 0 \forall v \notin [-1, 1]$, (ii) $\int_{-\infty}^{\infty} K(v)dv = 1$, (iii) $\int_{-\infty}^{\infty} vK(v)dv = 0$.

(C2) Let K be continuous with a limited first derivation K' .

(C3) $K_{h_n}(\cdot) = \frac{1}{h_n} K\left(\frac{\cdot}{h_n}\right)$ with restrictions to the bandwidth h_n : (i) $h_n \xrightarrow{n \rightarrow \infty} 0$, (ii) $nh_n, \dots, nh_n^4 \xrightarrow{n \rightarrow \infty} \infty$, $nh_n^6, nh_n^7, \dots \xrightarrow{n \rightarrow \infty} 0$, (iii) $nh_n^5 \xrightarrow{n \rightarrow \infty} C^2 \geq 0$.

(C4) Let s^2 be two times continuous differentiable.

(C5) Let rvs $\varepsilon_{1,n}, \dots, \varepsilon_{n,n}$ be iid with $E\varepsilon_{1,n} = 0$, $\text{Var } \varepsilon_{1,n} = 1$ and $E|\varepsilon_{1,n}|^{4+\delta} < \infty$ for a $\delta > 0$ and $n \in \mathbb{N}$.

Two-sided volatility estimation

We start the mathematical survey with the symmetrical estimate $\hat{s}_{h_n}(u)$ of the volatility function $s(u)$ on the standardized time frame $[0, 1]$. A minimum requirement to a developed, feasible estimator is consistency. Heuristically, an increase of the sample size should imply that the estimator converges to the parameter to be estimated. That way, a good estimate should be asymptotically unbiased and its variance should converge to zero as n goes to infinity. We prove both features and conclude the stochastic convergence.

Proposition 2.1 *Under the conditions (C1) - (C5) in setup (15) the sequence $(\hat{s}_{h_n}^2(u))_{n \in \mathbb{N}}$ of estimators for $s^2(u)$ is consistent for all $u \in (0, 1)$.*

Proof: Appendix A.

In the sequel we extend the consistency result by inspecting the rate of convergence. Searching for the maximum functional of n (h_n respectively), whose factorisation with $\hat{s}_{h_n}^2(u)$ still generates a limited sequence, we detected $\sqrt{nh_n}$ as the according convergence rate. $Z_{n,h_n}(u) := \sqrt{nh_n}(\hat{s}_{h_n}^2(u) - s^2(u))$ is limited (in probability) due to an asymptotical bias, a finite variance and an asymptotic normal distribution. Hence, it holds that:

Proposition 2.2 *Let $C \geq 0$ and $V := E\varepsilon_{1,n}^4 - 1 \in (0, \infty)$ and the conditions (C1) - (C5) be satisfied. Then the sequence of estimators $(\hat{s}_{h_n}^2(u))_{n \in \mathbb{N}}$ for $s^2(u)$ is asymptotic normally distributed for all $u \in (0, 1)$ in terms of*

$$\begin{aligned} \sqrt{nh_n}(\hat{s}_{h_n}^2(u) - s^2(u)) &\xrightarrow{D} \mathcal{N}(\beta(u), \tau^2(u)), \text{ where} \\ \beta(u) &= \frac{C}{2} (s^2(u))'' \int_{-1}^1 v^2 K(v) dv, \\ \tau^2(u) &= V s^4(u) \int_{-1}^1 K^2(v) dv. \end{aligned} \quad (18)$$

Proof: Appendix A.

The asymptotic normality result of the two-sided variance estimate entails some more insights: For the finite approximation of $s^2(u)$ by $\hat{s}_{h_n}^2(u)$ the slowest error terms have a $\sqrt{nh_n}$ rate of convergence. Concluding from proposition 2.2, for a sufficiently large n the pointwise approximation is nearly distributed as:

$$\hat{s}_{h_n}^2(u) - s^2(u) \approx \mathcal{N}\left(\frac{h_n^2}{2} (s^2(u))'' \int_{-1}^1 v^2 K(v) dv, \frac{V}{nh_n} s^4(u) \int_{-1}^1 K^2(v) dv\right) \quad (19)$$

The approximate bias has a negligible magnitude relative to the variance term. Hence, an approximative confidence interval for $s^2(u)$ can be simplistic implemented with Gaussian quantiles, built on a normal distribution centered at $\hat{s}_{h_n}^2(u)$ and a variance as above. Moreover, due to the approximate centered moments $\frac{\beta(u)}{\sqrt{nh_n}}$ and $\frac{\tau^2(u)}{nh_n}$ of the finite sample, discussions on optimal bandwidths could continue. Since the mean squared error is $MSE\hat{s}_{h_n}^2(u) = Bias^2\hat{s}_{h_n}^2(u) + Var\hat{s}_{h_n}^2(u)$, minimizing that function with respect to the bandwidth will provide the optimal trade-off between bias and variance.¹⁶ This typical trade-off from nonparametric curve estimation explains finally why the asymptotic bias of $\sqrt{nh_n}\hat{s}_{h_n}^2(u)$ does not equal zero, subjected to a maximum rate of convergence and the predefined bandwidth dimension from (C3). It is for the benefit of the asymptotic variance.

One-sided volatility estimation

We execute a similar statistical analysis for the historical, left-sided volatility estimator $\hat{s}_{(1)h_n}^2(u)$. The one-sided NWE for $s^2(u)$ in the transformed regression model has the exact definition:

$$\hat{s}_{(1)h_n}^2(u) = \begin{cases} \frac{\sum_{i=1}^{\lfloor un \rfloor} K_{h_n}(\frac{i}{n} - u) \tilde{R}_{i,n}^2}{\sum_{i=1}^{\lfloor un \rfloor} K_{h_n}(\frac{i}{n} - u)} & , \lfloor un \rfloor \geq 1 \\ 0 & , \lfloor un \rfloor = 0 \end{cases} , \quad u \in [0, 1] \quad (20)$$

The case differentiation is only for theoretical sake of completeness, because the later propositions are based on half-open intervals $u \in (0, 1]$ and for design points $u = \frac{t}{n}$, $t = 1, \dots, n$ there is no $\lfloor un \rfloor = 0$. The one-sided mean of past returns is defined as

$$\tilde{R}_{i,n} = X_{i,n} - \bar{X}_{i-1} \text{ where } \bar{X}_m = \begin{cases} \frac{1}{m} \sum_{j=1}^m X_{j,n} & , m \in \mathbb{N} \\ \mu & , m = 0 \end{cases} , \quad (21)$$

where the second case (for $i = 1$, with μ a technical, unspecified constant) is again asymptotically without influence.

For the one-sided NWE condition (C3) has to be replaced with (C3'), the other premises (C1) to (C5) persist:

¹⁶Our later considerations regarding optimal bandwidths are based on that pointwise result and extend it to the MISE-criterion.

(C3') $K_{h_n}(\cdot) = \frac{1}{h_n} K\left(\frac{\cdot}{h_n}\right)$ with restrictions to the bandwidth h_n : (i) $h_n \xrightarrow{n \rightarrow \infty} 0$, (ii) $nh_n, nh_n^2 \xrightarrow{n \rightarrow \infty} \infty$, $nh_n^4, nh_n^5, \dots \xrightarrow{n \rightarrow \infty} 0$, (iii) $nh_n^3 \xrightarrow{n \rightarrow \infty} D^2 \geq 0$.

Hence, a faster rate of convergence is claimed to the bandwidth h_n of one-sided estimators. This rate and the more complicated construction of the NWE require a multitude of effort in the convergence analysis, albeit the line of proofs is structurally similar to the both-sided case. We start again with the consistency of $\hat{s}_{(1)h_n}^2(u)$. The estimator is asymptotically unbiased, its variance tends to zero for large samples and the stochastic convergence follows.

Proposition 2.3 *Under the conditions (C1), (C2), (C3'), (C4), (C5) in setup (15) the sequence $\left(\hat{s}_{(1)h_n}^2(u)\right)_{n \in \mathbb{N}}$ of estimators for $s^2(u)$ is consistent for all $u \in (0, 1]$.*

Proof: Appendix A.

The consistency of $\hat{s}_{(1)h_n}^2(u)$ is not only valid at interior points of $[0, 1]$ (as restricted for the two-sided approach), but also at the right frontier. This makes the estimator consistently applicable for forecasting volatility and return distributions. Again $\sqrt{nh_n}$ is the (maximum) rate of convergence, where $Z_{n,h_n}^{(1)}(u) := \sqrt{nh_n} \left(\hat{s}_{(1)h_n}^2(u) - s^2(u)\right)$ is limited due to an asymptotical bias and a finite variance, and a convergence in distribution to a Gaussian follows.

Proposition 2.4 *Let $D \geq 0$ and $V := E\varepsilon_{1,n}^4 - 1 \in (0, \infty)$ and the conditions (C1), (C2), (C3'), (C4), (C5) be satisfied. Then the sequence of estimators $\left(\hat{s}_{(1)h_n}^2(u)\right)_{n \in \mathbb{N}}$ for $s^2(u)$ is asymptotic normally distributed for all $u \in (0, 1]$ in terms of*

$$\begin{aligned} \sqrt{nh_n} \left(\hat{s}_{(1)h_n}^2(u) - s^2(u)\right) &\xrightarrow{D} \mathcal{N}\left(\beta_{(1)}(u), \tau_{(1)}^2(u)\right), \quad \text{where} \\ \beta_{(1)}(u) &= 2D \left(s^2(u)\right)' \int_{-1}^0 vK(v)dv, \\ \tau_{(1)}^2(u) &= 4V s^4(u) \int_{-1}^0 K^2(v)dv. \end{aligned} \quad (22)$$

Proof: Appendix A.

We notice that the asymptotic variance $\tau_{(1)}^2(u)$ is twice the two-sided counterpart $\tau^2(u)$ from proposition 2.2, which is the consequence that only half the return information is used by $\hat{s}_{(1)h_n}^2(u)$ at the interior of $(0, 1]$. But at the right boundary the two-sided (uncorrected) estimator $\hat{s}_{h_n}^2(u)$ is implicitly limited to the identical data base and the same asymptotic error terms result. Moreover, we conclude from proposition 2.4 for the finite approximation:

$$\hat{s}_{(1)h_n}^2(u) - s^2(u) \approx \mathcal{N}\left(2h_n \left(s^2(u)\right)' \int_{-1}^0 vK(v)dv, \frac{4V}{nh_n} s^4(u) \int_{-1}^0 K^2(v)dv\right), \quad (23)$$

for a sufficiently large sample size n . An approximative confidence interval for $s^2(u)$ can be simplistic implemented via normal quantiles, centered at $\hat{s}_{(1)h_n}^2(u)$ and scaled by the above standard deviation. Last but not least, discussions on optimal bandwidths, that minimize the MSE or MISE of $\hat{s}_{(1)h_n}^2(u)$, employ the asymptotic normality result.

Choice of kernel and bandwidth

In the prevalent literature it is established, that the choice of a kernel function plays a relatively unimportant role compared with the chosen bandwidth for nonparametric regression. Different kernels perform very similar in large samples. But the simplest kernels, as a rectangle-, a triangular- or a normal kernel, are incompatible to the regression models (10) or (15), respectively, due to the conditions of continuous differentiability and a compact support, that are needed for the desired properties of the estimators. A polynomial kernel of fourth degree, with its maximum in zero and boundary roots in -1 and $+1$ is recommended:

$$K(u) := \begin{cases} \frac{15}{16}(1-u^2)^2 & , \quad |u| \leq 1 \\ 0 & , \quad \text{else} \end{cases}, \quad (24)$$

which is also called *biweight kernel*¹⁷ and depicted in figure 1.

¹⁷Following Fan and Yao (2003), this kernel is from the 'symmetric Beta family' $K_\gamma(u) = \left(\mathcal{B}\left(\frac{1}{2}, \gamma + 1\right)\right)^{-1} (1-u^2)^\gamma I_{[-1,1]}(u)$ with beta-integral $\mathcal{B}(\alpha_1, \alpha_2) = \int_0^1 (1-y)^{\alpha_1-1} y^{\alpha_2-1} dy$ as the special case $\gamma = 2$.

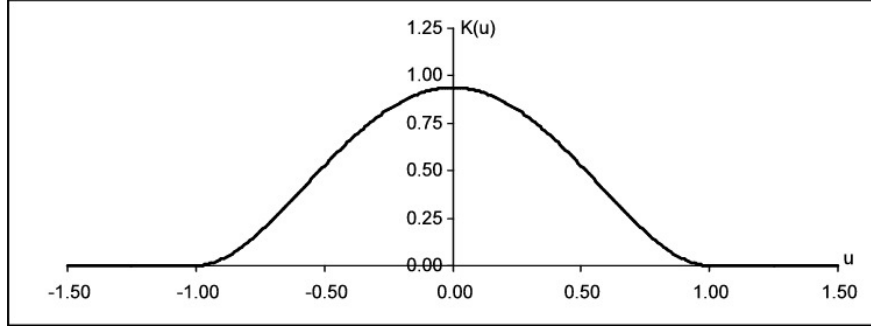


Figure 1: Biweight kernel.

Using the biweight kernel in the sequel, we turn to the task of bandwidth decision, where we have to find a trade-off between oversmoothing and undersmoothing. Oversmoothing means in terms of NWEs to build an average over a too large neighbourhood of return points (large bandwidth), where recent return information is dominated; a very smooth shape of the regression function (small variance, but biased) results. Undersmoothing averages over a very small neighbourhood (small bandwidth), where only a few recent data points are included; a rough shape (small bias, but large variance) of the estimated volatility function follows. Hence, the bandwidth is also called smoothing parameter.

Discussions on optimal bandwidths require an error measure. Local bandwidth optimization can be based on minimizing the MSE with respect to h_n for a point in time. The afore derived asymptotic results are constraining. An average of MSE optimal bandwidths over all design points corresponds to the MASE criterion. Superior to this is to integrate the squared deviation of the estimator $\hat{s}(\cdot)$ to the volatility function $s(\cdot)$ over the (standardized) horizon and to minimize this MISE. For sufficient large n we get the following global optimal bandwidths:

(I) For the two-sided (transformed) NWE (16):

$$h_n^{opt} = n^{-\frac{1}{5}} \left(\frac{V \left(\int_0^1 s^4(u) du \right) \left(\int_{-1}^1 K^2(v) dv \right)}{\left(\int_0^1 ((s^2(u))'')^2 du \right) \left(\int_{-1}^1 v^2 K(v) dv \right)^2} \right)^{\frac{1}{5}} \quad (25)$$

(II) For the one-sided (transformed) NWE (17):

$$h_n^{opt} = n^{-\frac{1}{3}} \left(\frac{V \left(\int_0^1 s^4(u) du \right) \left(\int_{-1}^0 K^2(v) dv \right)}{2 \left(\int_0^1 ((s^2(u))')^2 du \right) \left(\int_{-1}^0 v K(v) dv \right)^2} \right)^{\frac{1}{3}} \quad (26)$$

Those approaches can be used directly only for simulation studies, where the volatility function $s(\cdot)$ is a predefined input. For empirical samples, it is the task to estimate $s(\cdot)$, so bandwidth criteria that are based on that function or its derivatives are problematical.¹⁸ One usual way out is the *cross-validation* method, which we use for automatised bandwidth selection as in Gürtler and Rauh (2009). Basic idea is the 'leave-one-out prediction' over the discrete design. In our context, the variance $s^2(\cdot)$ has to be reestimated for each point $\frac{j}{n}$, $j = 1, \dots, n$, without using the actual observation $R_{j,n}^2$ (or $\tilde{R}_{j,n}^2$) itself. The resulting cross-validation estimators (CVE) for bandwidths $h_n > \frac{1}{n}$ are:

(I) Two-sided (transformed) CVE:

$$\hat{s}_{h_n}^{2(j)} \left(\frac{j}{n} \right) = \frac{\sum_{i=1, i \neq j}^n K_{h_n} \left(\frac{i-j}{n} \right) R_{i,n}^2}{\sum_{i=1, i \neq j}^n K_{h_n} \left(\frac{i-j}{n} \right)} \quad \forall j = 1, \dots, n \quad (27)$$

(II) One-sided (transformed) CVE:

$$\hat{s}_{h_n}^{2(j)} \left(\frac{j}{n} \right) = \frac{\sum_{i=1}^{j-1} K_{h_n} \left(\frac{i-j}{n} \right) \tilde{R}_{i,n}^2}{\sum_{i=1}^{j-1} K_{h_n} \left(\frac{i-j}{n} \right)} \quad \forall j = 2, \dots, n \quad (28)$$

¹⁸So called *plug-in* methods develop kernel estimators for the unknown volatility $s(\cdot)$ and its derivatives $s^2(\cdot)''$ or $s^2(\cdot)'$ and plug them into the above bandwidth formula, with an iterative procedure leading to optimal bandwidths, compare Gasser et al. (1991).

The bandwidth selection criterion resembles a MASE method. A sum of squared differences between returns $R_{j,n}^2$ and CVEs $\hat{s}_{h_n}^{2(j)}\left(\frac{j}{n}\right)$ (alternatively $\tilde{R}_{j,n}^2$ and $\hat{s}_{(1)h_n}^{2(j)}\left(\frac{j}{n}\right)$), called CV-function, is minimized:

$$(I) \quad CV(h_n) = \frac{1}{n} \sum_{j=1}^n \left(R_{j,n}^2 - \hat{s}_{h_n}^{2(j)}\left(\frac{j}{n}\right) \right)^2 = \frac{1}{n} \sum_{j=1}^n \left(\frac{\sum_{i=1}^n K_{h_n}\left(\frac{i-j}{n}\right) (R_{j,n}^2 - R_{i,n}^2)}{\sum_{i=1, i \neq j}^n K_{h_n}\left(\frac{i-j}{n}\right)} \right)^2 \xrightarrow{!} \min_{h_n > 1/n} \quad (29)$$

$$(II) \quad CV_{(1)}(h_n) = \frac{1}{n-1} \sum_{j=2}^n \left(\tilde{R}_{j,n}^2 - \hat{s}_{(1)h_n}^{2(j)}\left(\frac{j}{n}\right) \right)^2 = \frac{1}{n-1} \sum_{j=2}^n \left(\frac{\sum_{i=1}^{j-1} K_{h_n}\left(\frac{i-j}{n}\right) (\tilde{R}_{j,n}^2 - \tilde{R}_{i,n}^2)}{\sum_{i=1}^{j-1} K_{h_n}\left(\frac{i-j}{n}\right)} \right)^2 \xrightarrow{!} \min_{h_n > 1/n} \quad (30)$$

The CV-optimal bandwidths h_n^{CV} and $h_n^{CV(1)}$ for the two- and one-sided volatility estimation are numerically found via a value table for a sufficient fine grid of bandwidths h_n or via analysing the CV-plot. An example of functions $CV(h_n)$ and $CV_{(1)}(h_n)$ from the later executed simulation studies is presented in figure 2.

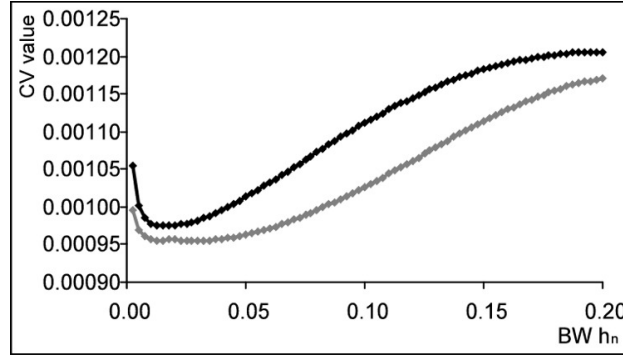


Figure 2: Cross validation graphs for the two-sided (grey) and the one-sided CV function (black) regarding the pre-defined volatility process and simulated log-returns in section 3, part B ($n = 5000$). The CV-plot is minimal for $h_{5000}^{CV} = 0.0300$ and $h_{5000}^{CV(1)} = 0.0150$ on a grid pattern of widths 0.005.

The CV-method determines the optimal smoothing parameter solely with the return series and without any knowledge about the regressed volatility function. For empirical samples $\{X_t\}_{t=1, \dots, n}$ it is not necessary to transform the setup first for estimating $s\left(\frac{t}{n}\right)$, rather the volatility $\sigma(t)$ should be estimated directly. We turn back the definition $h = nh_n$ and derive in the original regression model (10) likewise CV-functions to be minimized:

(I) Two-sided (untransformed):

$$\widetilde{CV}(h) = \frac{1}{n} \sum_{j=1}^n \left(R_j^2 - \hat{\sigma}_h^{2(j)}(j) \right)^2 = \frac{1}{n} \sum_{j=1}^n \left(\frac{\sum_{i=1}^n K_h(i-j) (R_j^2 - R_i^2)}{\sum_{i=1, i \neq j}^n K_h(i-j)} \right)^2 \quad (31)$$

$$\text{where } \hat{\sigma}_h^{2(j)}(j) = \frac{\sum_{i=1, i \neq j}^n K_h(i-j) R_i^2}{\sum_{i=1, i \neq j}^n K_h(i-j)}, \quad h > 1 \quad (\forall j = 1, \dots, n)$$

(II) One-sided (untransformed):

$$\widetilde{CV}_{(1)}(h) = \frac{1}{n-1} \sum_{j=2}^n \left(\tilde{R}_j^2 - \hat{\sigma}_{(1)h}^{2(j)}(j) \right)^2 = \frac{1}{n-1} \sum_{j=2}^n \left(\frac{\sum_{i=1}^{j-1} K_h(i-j) (\tilde{R}_j^2 - \tilde{R}_i^2)}{\sum_{i=1}^{j-1} K_h(i-j)} \right)^2 \quad (32)$$

$$\text{where } \hat{\sigma}_{(1)h}^{2(j)}(j) = \frac{\sum_{i=1}^{j-1} K_h(i-j) \tilde{R}_i^2}{\sum_{i=1}^{j-1} K_h(i-j)}, \quad h > 1 \quad (\forall j = 2, \dots, n)$$

It is easy to see that $\widetilde{CV}(h) = CV(h_n)$ and $h^{\widetilde{CV}} = h_n^{CV} n$ (two-sided example, one-sided approach works as well) and hence it does not matter on which setup the optimization is based on.

2.2 Focus on fitting innovations

In the sequel we develop the theory for fitting the Pearson type VII distribution to a series of estimated innovations $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ via a *method of moments*. Because the residual sample is evaluated in each point of time t by the ratio of a demeaned return and an estimated volatility from the regression-type model (10), we have to distinct between innovation estimates on a two-sided and one-sided data base:

- Two-sided innovation estimators:

$$\hat{\varepsilon}_t = \frac{R_t}{\hat{\sigma}(t)}, \quad t = 1, \dots, n \quad (33)$$

- One-sided innovation estimators:

$$\hat{\varepsilon}_t = \frac{\tilde{R}_t}{\hat{\sigma}_{(1)}(t)}, \quad t = 1, \dots, n \quad (34)$$

Agreeing with Drees and Starica (2002), Herzel et al. (2005), Mikosch and Starica (2004a) or Kitagawa and Nagahara (1999), and as shown in comprehensive empirical studies of Gurtler and Rauh (2009), the Pearson type VII distribution can capture some heavy tailed innovations quite nicely. Remember, that we fit the Pearson VII density (term (8)) separately to nonnegative $\{\hat{\varepsilon}_t \mid \hat{\varepsilon}_t \geq 0\}_t$ and absolutes of negative innovations $\{-\hat{\varepsilon}_t \mid \hat{\varepsilon}_t < 0\}_t$. Under the assumption that their median equals 0, the one-side fitted densities $f_{m_+, c_+}^{VII(1)}$ and $f_{m_-, c_-}^{VII(1)}$ are combined via $f_{m_+, c_+, m_-, c_-}^{VII}$ as described in equation (9), called asymmetric Pearson type VII density.

The Pearson VII density is part of the Pearson system, which can be studied in Johnson and Kotz (1970). Every member of the system has a probability density function $p(x)$ that solves a differential equation of form

$$\frac{1}{p} \frac{dp}{dx} = -\frac{a+x}{d_0 + d_1x + d_2x^2}$$

where a, d_0, d_1 and d_2 are real shape parameters. Type VII is a special case with $a = d_1 = 0$ and $d_0 > 0, d_2 > 0$. The explicit solution of the differential equation is

$$p(x) = k (d_0 + d_2x^2)^{-(2d_2)^{-1}},$$

with k chosen to satisfy $\int_{-\infty}^{+\infty} p(x)dx = 1$. This is called the symmetric Pearson VII density. Johnson and Kotz (1970) estimate the parameters of the Pearson system via a general method of moments, especially:

$$\begin{aligned} d_0 &= (4\beta_2 - 3\beta_1)(10\beta_2 - 12\beta_1 - 18)^{-1} \mu_2, \\ d_2 &= (2\beta_2 - 3\beta_1 - 6)(10\beta_2 - 12\beta_1 - 18)^{-1}, \end{aligned}$$

where $\mu_2(X) = EX^2$, $\beta_1(X) = \left(\frac{E(X-EX)^3}{(E(X-EX)^2)^{3/2}}\right)^2$ (squared skewness) and $\beta_2(X) = \frac{E(X-EX)^4}{(E(X-EX)^2)^2}$ (kurtosis). Since the Pearson VII density $p(x)$ is symmetrical, $\beta_1 = 0$, and $\beta_2 > 3$ is required due to a positive d_2 . This means, that the conception is only applicable for samples that are heavier than a normal distribution.

Substituting $m := (2d_2)^{-1}$ and interpreting d_0 as a rescaled constant $d_0 = \frac{c^2}{2m}$ (i.e. substitute $c := \sqrt{2m \cdot d_0}$) we derive an expression for k :

$$\begin{aligned} p(x) &= k \left(\frac{c^2}{2m} + \frac{1}{2m}x^2 \right)^{-m} = k \left(\frac{c^2}{2m} \left(1 + \left(\frac{x}{c} \right)^2 \right) \right)^{-m} \\ \Rightarrow 1 &\stackrel{!}{=} k \left(\frac{2m}{c^2} \right)^m 2 \int_0^\infty \frac{1}{\left(1 + \left(\frac{x}{c} \right)^2 \right)^m} dx \stackrel{y:=\left(\frac{x}{c}\right)^2}{=} k \left(\frac{2m}{c^2} \right)^m c \int_0^\infty \frac{y^{-1/2}}{(1+y)^m} dy = k \frac{(2m)^m}{c^{2m-1}} \mathcal{B} \left(\frac{1}{2}, m - \frac{1}{2} \right), \\ \Rightarrow k &= \frac{1}{\mathcal{B} \left(\frac{1}{2}, m - \frac{1}{2} \right)} \frac{c^{2m-1}}{(2m)^m}, \text{ where } \mathcal{B}(\alpha_1, \alpha_2) = \int_0^\infty \frac{u^{\alpha_1-1}}{(1+u)^{\alpha_1+\alpha_2}} du = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} \end{aligned}$$

Applying that term for k in the symmetric Pearson VII density $p(x)$ leads to:

$$\begin{aligned} p(x) &= \frac{1}{B\left(\frac{1}{2}, m - \frac{1}{2}\right)} \frac{c^{2m-1}}{(2m)^m} \cdot \left(\frac{2m}{c^2}\right)^m \left(1 + \left(\frac{x}{c}\right)^2\right)^{-m} \\ &= \frac{\Gamma(m)}{c\Gamma\left(m - \frac{1}{2}\right)\pi^{1/2}} \left(1 + \left(\frac{x}{c}\right)^2\right)^{-m} =: f_{m,c}^{VII(2)}(x) \end{aligned} \quad (35)$$

This expression corresponds to (8), but defined on the whole real axis, i.e. $f_{m,c}^{VII(1)}(x) = 2 \cdot f_{m,c}^{VII(2)}(x) I_{[0,\infty)}(x)$, $x \in \mathbb{R}$. That way we achieve moment estimators for parameters m, c of the one-sided Pearson VII density $f_{m,c}^{VII(1)}$:

$$m = \frac{1}{2d_2} = \frac{5\beta_2 - 9}{2\beta_2 - 6} \quad (36)$$

$$c = \sqrt{2m d_0} = \sqrt{2m \frac{2\beta_2\mu_2}{5\beta_2 - 9}} = \sqrt{\frac{2\beta_2\mu_2}{\beta_2 - 3}} \quad (37)$$

By inserting the empirical moment estimators for β_2 and μ_2 we can fit $f_{m,c}^{VII(1)}$ to the data set. More precisely, the asymmetric Pearson VII parameterisation m_+, c_+, m_-, c_- is to estimate successively for the innovation sets $\varepsilon_+ := \{\hat{\varepsilon}_t \mid \hat{\varepsilon}_t \geq 0\}_t$ and $\varepsilon_- := \{-\hat{\varepsilon}_t \mid \hat{\varepsilon}_t < 0\}_t$ concerning the right and left tail. Special attention should be paid to that method of moments in the one-sided implementation: For estimating $\hat{\beta}_2(\varepsilon_+)$ we recommend to include the hypothetical counterpart $-\varepsilon_+$ of the one-sided set and to work on the symmetrical base $\{\varepsilon_+, -\varepsilon_+\}$ with standard formula.¹⁹ It follows, that estimators can be reduced to:

$$\hat{\beta}_2(\varepsilon_+) = \frac{E\varepsilon_+^4}{E\varepsilon_+^2}, \quad \text{analogous: } \hat{\beta}_2(\varepsilon_-) = \frac{E\varepsilon_-^4}{E\varepsilon_-^2}$$

Next to the self created asymmetry, the Pearson VII setup detects heavy tailed distributions quite flexible²⁰ and its density can be parameterized to decay slower than any arbitrary decreasing polynomial. For $\beta_2 \searrow 3$ and $\mu_2 = 1$ ($\implies m \rightarrow \infty, c \rightarrow \infty$) the normal distribution follows from the symmetric Pearson VII asymptotically. Special cases are the Cauchy distribution²¹ and a scaled Student- t distribution with $2m - 1$ df.²² With respect to

$$g_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty,$$

as the Student- t_n density, it holds for a t -distributed rv with $2m - 1$ df which is rescaled by $\gamma := \frac{c}{\sqrt{2m-1}}$ that:

$$\begin{aligned} \bar{g}_{2m-1,\gamma}(x) &= \frac{1}{\gamma} g_{2m-1}\left(\frac{x}{\gamma}\right) \\ &= \frac{\sqrt{2m-1}}{c} \frac{\Gamma(m)}{\sqrt{2m-1}\Gamma\left(m - \frac{1}{2}\right)\sqrt{\pi}} \left(1 + \left(\frac{\sqrt{2m-1}}{c} \frac{x}{\sqrt{2m-1}}\right)^2\right)^{-m} = f_{m,c}^{VII(2)}(x) \end{aligned}$$

Because of the simple transformation from the Pearson type VII to a Student- t distribution, whose quantiles are looked up in tables, we developed the idea to implement the non-stationary model (10) for the task of VaR calculation as a factor-based model. The univariate Value at Risk $VaR^{1-\alpha}(t)$ of an exposure $w(t)$ at time t with yields X_t following the regression model (10) is then to be modelled as the product of $w(t)$ with a nonparametric estimated volatility $\hat{\sigma}_{(\cdot)}(t)$ and the α -quantile $u_{m,c;\alpha}$ of a Pearson VII innovation. Since the cdfs are connected as

$$F_{m,c}^{VII(2)}(x) = \int_{-\infty}^x \frac{1}{\gamma} g_{2m-1}\left(\frac{y}{\gamma}\right) dy \stackrel{z:=y/\gamma}{=} \int_{-\infty}^{x/\gamma} \frac{1}{\gamma} g_{2m-1}(z) \gamma dz = G_{2m-1}\left(\frac{x}{\gamma}\right),$$

¹⁹Else $E\varepsilon_+ \neq 0$ and $\hat{\beta}_2$ builds powers on a falsely centered rv.

²⁰But the method does not work, if innovations are not heavy tailed, i.e. if $\beta_2(\varepsilon_t) \leq 3$. That case one can conservatively (with a pure risk aversion view) apply the normal distribution.

²¹The Cauchy density follows from $p(x)$ with $k = \frac{1}{2\lambda\pi}$ ($\lambda > 0$) and $d_0 = d_2 = \frac{1}{2}$.

²²Student's df is a measure for the heavy-tailedness and is called *tail index point estimate* in the common literature, e.g. Drees and Starica (2002).

it follows immediately for the corresponding quantiles:

$$t_{2m-1;\alpha} = \frac{u_{m,c;\alpha}}{\gamma} \iff u_{m,c;\alpha} = \gamma t_{2m-1;\alpha} \quad (38)$$

Regarding the one-sided presentation, some attention has to be paid to:

$$F_{m,c}^{VII(1)}(x) = 2 \cdot F_{m,c}^{VII(2)}(x) \implies \alpha^{(1)} = 2\alpha \quad (\text{e.g. for confidence level } \alpha = 1\%)$$

Moreover, from the Student- t connection follows a restriction to the shape parameter m , that ensures the asymptotic normality of the NWEs $\hat{s}_{h_n}^2(u)$ and $\hat{s}_{h_n(1)}^2(u)$, respectively. Amongst others, (C5) $E|\varepsilon_{1,n}|^{4+\delta} < \infty$ ($\delta > 0$) was required theoretically. As all (central) moments μ_k with $n > k$ exist for a t_n -distributed rv, the Pearson VII fit of innovations has to be conform with

$$2m - 1 > 4 \iff m > \frac{5}{2}.$$

If the set of innovations was symmetric (special case, not observed in general), the assumption $Var\varepsilon_1 = 1$ for iid $\{\varepsilon_t\}_t$ implies additionally the relation $c = \sqrt{2m - 3}$ of shape parameters.²³

3 Simulation studies with the non-stationary approach

So far we have provided the necessary statistical tools for estimating the volatility structure $(\sigma(t))_t$ nonparametrically and for approximating estimated innovations $\{\hat{\varepsilon}_t\}_t$ by the asymmetric Pearson type VII distribution via a method of moments. These are the main steps for fitting the regression model to a financial time series. In this section we apply the theory to a simulation study, where firstly the focus is set on the goodness of volatility (re-)estimation, dependent on the sample size and bandwidth optimization. Afterwards we formulate a price process in terms of the regression model (10), perturbed by a Pearson VII noise, and finally evaluate the approximation by the non-stationary model in a backtesting framework.

A) For the beginning we predefine a heteroscedastic volatility function (standard deviation) as:

$$\sigma^a(t) = \frac{1}{10} \left(\sin\left(\frac{2\pi}{100}t\right) + 1 \right) \quad t \in [0, 500] \quad (39)$$

$$s^a(u) = \frac{1}{10} (\sin(10\pi u) + 1) \quad u \in [0, 1] \quad (40)$$

This corresponds to a multimodal oscillation with 5 periods in the interval $[0, 500]$. The values on the discrete design $1, \dots, 500$ could be thought as annualized volatilities $\sigma^a(t)$ at the end of days t , observed over two years (250 trading days p.a.). The transformed version is $s^a(u)$. The example focusses first on reestimating the volatility in a simplified case of the standardized regression model (15), with an expectation $\mu = 0$ for log-returns $\{X_{t,n}\}_{t=1,\dots,n}$ and innovations distributed as $\varepsilon_{1,n} \sim \mathcal{N}(0, 1)$. The main point is to estimate the heteroscedastic part of the simulated series²⁴ with the two-sided $\hat{s}_{h_n}^2(u)$ and the one-sided NWE $\hat{s}_{h_n(1)}^2(u)$. We observe the influence of sample size to the volatility estimator's fit.²⁵ Bandwidths are optimized via the MISE-criterion (formulas (25), (26)).

As discussed in section 2.1 the asymptotic works by increasing the data density on a fixed time-frame $[0, 1]$. A larger sample size n enables more and closer observations, the predefined function is scanned more precisely. Figure 3 displays the nonparametric curve estimation of s^a for different sample sizes in a median simulation of 65 repeats. By calculating the sum of squared errors $SSE \hat{s}_{h_n} := \sum_{i=1}^n \left(\hat{s}_{h_n}\left(\frac{i}{n}\right) - s\left(\frac{i}{n}\right) \right)^2$ of the estimator relative to the given volatility function for each simulated sample and ordering paths by their SSE we select the median simulation.

²³This is to be seen with $VarY = \frac{n}{n-2}$ for a (unscaled) t_n -distributed rv Y ($n > 2$). Then $Var\varepsilon_1 = \gamma^2 VarY = \frac{c^2}{2m-1} \cdot \frac{2m-1}{2m-3} \stackrel{!}{=} 1$.

²⁴We include later the proposed innovation modelling. The alternative choice of an expectation $\mu = const$ has a negligible influence on the simulation example due to centering the returns first of all. Assuming $\mu = 0$ is just for the sake of convenience.

²⁵The simulation was implemented in a C programme, using the Box-Muller method for transforming uniform to normal random numbers. Secondly we wrote VBA-code for MS Excel, to be used for smaller samples.

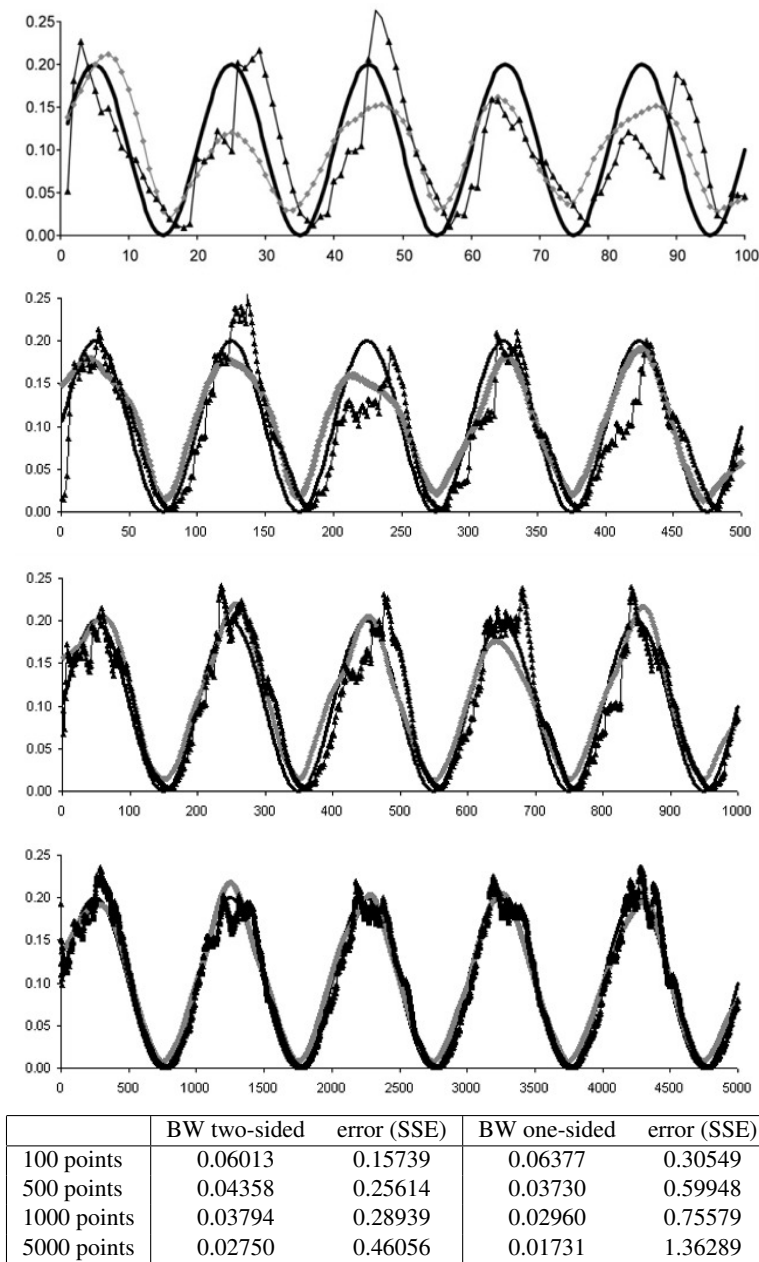


Figure 3: Median simulation (65 samples ordered by SSE) of volatility curve estimation on s^a (continuous line, black) for 100, 500, 1000 and 5000 equidistant design points (from top to bottom) by two-sided NWEs (rhombuses, grey) and one-sided NWEs (triangles, black) with MISE-optimal bandwidths (BW).

We discover visually that the fit improves with the sample size for both NWEs. The SSE increases considerably slower than the sample size, for very large samples it decreases absolutely. The approximation of $s^a(u)$ in terms of nonparametric volatility estimators $\hat{s}_{h_n}(u)$ and $\hat{s}_{(1)h_n}(u)$ is already noticeable for smaller samples of 100 or 500 points and quite satisfying using 1000 design points. Due to its additional future information, the two-sided estimator produces generally a better and smoother fit. The left-sided approach lags behind, since the volatility does not increase until the first extremal event happens, and after a series of shocks it decays typically slower. We did not correct boundary effects, so the first nh_n points are distorted, for the two-sided estimator also the last nh_n values. On the right boundary both estimators are nearly the same (neglecting different bandwidths). For the sample size of 5000 the estimates cover the predetermined volatility function visually excellent.

B) To appreciate how good the cross-validation method works, we estimate the optimal bandwidth in the simulation example directly without including the knowledge of s^a . The corresponding CV functions of a two- and one-sided 5000-point setup were plotted in figure 2. The resulting volatility estimations $\hat{s}_{h_n}(u)$ and $\hat{s}_{(1)h_n}(u)$, presented in graph 4, deliver again an excellent fit to the predetermined $s^a(u)$. Compared with the MISE-optimal smoothing parameters the bandwidths from CV are very close, with absolute differences lower/ equal 0.0025, which may also be caused by the grid pattern of widths 0.005. The SSEs are marginally larger. Deviations from figure 3 (bottom graph) are mainly caused due to other random innovations.

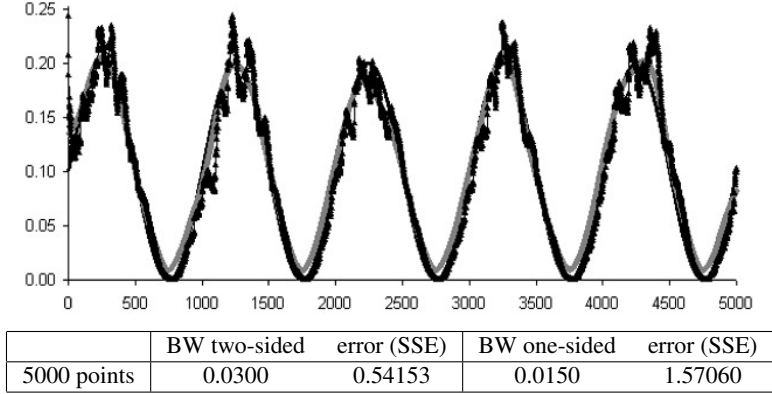


Figure 4: Two-sided (rhombuses, grey) and one-sided estimation (triangles, black) of the hypothetical volatility s^a (continuous line, black) for 5000 design points with optimal bandwidths from cross-validation.

C) In the next step we expand the volatility simulation to a discrete price process $\{P_t\}_{t=0,\dots,500}$, where $\sigma^a(t)$ on $[0, 500]$ is involved as time-variant part of an annualized volatility function $\tilde{\sigma}^a(t)$, that additionally includes a base volatility σ_0 . Moreover a constant trend μ is modelled and a heavy-tailed approach for the innovations is chosen:

$$\begin{aligned}
 P_t &= P_{t-1} e^{X_t}, \quad P_0 := 1000, \quad t = 1, \dots, 500, \quad \text{where} & (41) \\
 X_t &= \mu + \sigma(t)\varepsilon_t, \quad \text{with } \mu := \frac{15\%}{250} \text{ and} \\
 \sigma(t) &= \frac{\tilde{\sigma}^a(t)}{\sqrt{250}}, \quad \tilde{\sigma}^a(t) := \sigma^a(t) + \sigma_0, \quad \text{with } \sigma_0 := 10\%, \\
 \varepsilon_t &\sim \text{Pearson}_{m,c}^{VII(2)} \quad \text{where } m := 4 \text{ and } c := \sqrt{5}.
 \end{aligned}$$

The log-return of prices, $X_t = \ln(P_t/P_{t-1}) = \ln e^{X_t} = \mu + \sigma(t)\varepsilon_t$ corresponds to the (untransformed) regression approach (10). Innovations ε_t are modelled as symmetrical Pearson VII distributed rvs, i.e. following density $f_{m,c}^{VII(2)}$ from (35). Figure 5 displays a representative simulated path of prices²⁶ and the nonparametric estimated volatilities relative to the original volatility in the course of time.

For that example the fit is quite acceptable despite the small sample size of 500 and the strong influence of random heavy-tailed innovations ε_t . The comparison to the price graph shows, that the nonparametric estimators are able to capture fast phases of market shocks or increased volatility, respectively. Repeats of the simulation setup may produce very different paths, because a high volatility level (medium 20% p.a.) dominates the price process more than its trend ($\mu = 15\%$ p.a.) and the incident of extreme innovations leads to considerable shocks. Due to the random innovations, the NWEs may have different optimal bandwidths and, of course, some different developing of estimates adapted to the realized volatility of the simulated sample.

D) We finish our simulation studies with a forecasting experiment, continuing on the price process (41) and its realisation displayed in figure 5, i.e. realized innovations $\{\varepsilon_t\}_t$ were frozen. We imagine, that we observed only the first 251 prices (P_0, \dots, P_{250}) up to a forecast starting point $t_0 = 250$, and refer to that half-series as 'in-sample part'.

²⁶Innovations were simulated in MS Excel via uniform random numbers α , that where interpreted as probabilities corresponding to the α -quantile of a student- t_{2m-1} rv. The Pearson VII random number is then $u_{m,c;\alpha} = \frac{c}{\sqrt{2m-1}} t_{2m-1;\alpha}$.

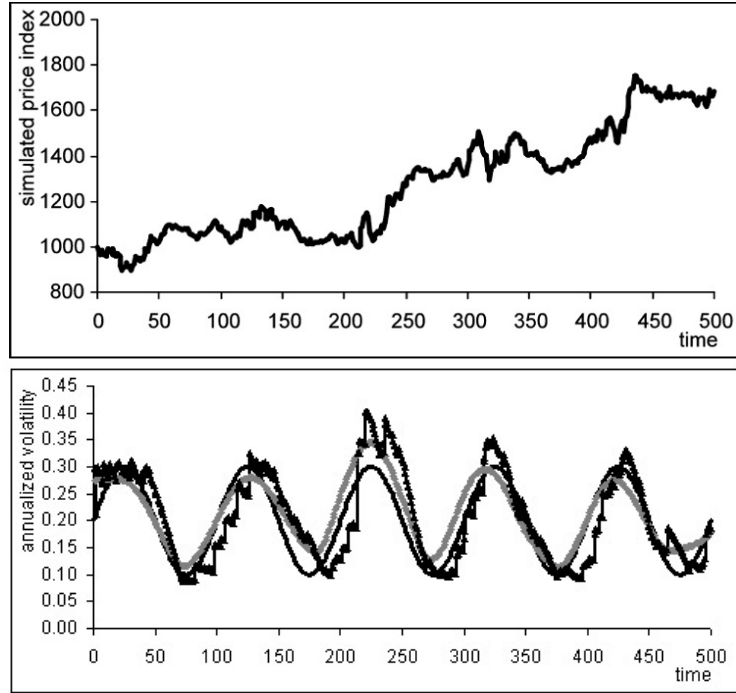


Figure 5: Simulation path for price process (41) including the regression model (10) (top) and annualized volatility estimates for log-returns (bottom): $\hat{\sigma}(t)$ (rhombuses, grey) and $\hat{\sigma}_1(t)$ (triangles, black) of the predefined volatility $\tilde{\sigma}^a(t)$ (continuous line, black) for 500 design points with CV-optimal bandwidths ($h^{\widehat{CV}} = 34$, $h^{\widehat{CV}^{(1)}} = 29$).

The other half, called 'out-of-sample part', will follow within the next year. We plan to forecast the distribution of the 1-day ahead return X_{t+1} with the information available in $t \geq t_0$, successively. For forecasting purposes we naturally have to work with the one-sided setup of the regression model (10) since only past information is available.

The model is calibrated in-sample and optimal parameters are fixed for the out-of-sample part:²⁷ After estimating a centered return series $\tilde{R}_1, \dots, \tilde{R}_{t_0}$, the optimal smoothing parameter $h_n^{\widehat{CV}^{(1)}}$ is identified with cross-validation for the purpose of nonparametric volatility estimation. Having calculated the series of standard deviation estimates $\hat{\sigma}_{(1)}(1), \dots, \hat{\sigma}_{(1)}(t_0)$ we estimate the innovation series $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_{t_0}$ and fit the asymmetric Pearson VII density $f_{m_+, c_+, m_-, c_-}^{VII}$ to nonnegative and absolute values of negative innovations.²⁸ That way, we use the distribution of X_{t_0} as a forecast of X_{t_0+1} . Afterwards, we develop with the new returns, their volatility estimates and the fixed innovation parametrization the distributions of X_{251}, \dots, X_{500} successively.

We obtain an optimal one-sided bandwidth $h_n^{\widehat{CV}^{(1)}} = 36$ days. This differs slightly from the optimum on the full sample due to the impact of random innovations in the quite small series. The one-sided annualized, estimated volatilities of the price process are shown in figure 6. The in-sample estimates are grey depicted, while the later pointwise estimators of the out-of-sample part are printed in black. The optimal Pearson VII parameters for estimated innovations $\{\hat{\varepsilon}_t\}_{t=1, \dots, 250}$ in terms of one-sided volatilities and returns are $m_+ = 5.6299$, $c_+ = 2.9038$ and $m_- = 7.1976$, $c_- = 3.2758$. They define the right and the left tail of the asymmetric Pearson VII density $f_{m_+, c_+, m_-, c_-}^{VII}$, that are presented in figure 7 relative to the histogram of innovations.

A pretty good approximation of the (scaled) innovation frequencies by the Pearson VII densities (black line) is observed on both sides. Moreover, this illustration presents a comparison to a standardized Gaussian density (grey line). Deviations from the normal densities are to be seen in the middle of the distribution and in the tails, where

²⁷ An alternative approach could be to re-calibrate the model every day or every fixed period out-of-sample, to incorporate the new parameter information.

²⁸ Even though we have assumed a symmetric Pearson VII distribution, the random innovations of the small sample cause different optimal parameters m_+, c_+ vs. m_-, c_- on both sides following the method of moments.

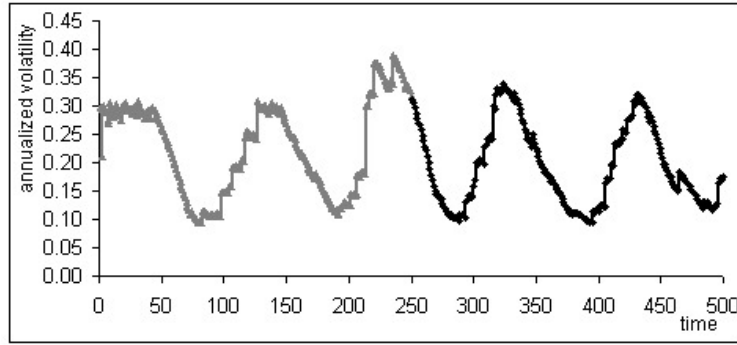


Figure 6: One-sided nonparametric estimator of volatilities $\hat{\sigma}_1(t)$ in the price process (41) with an CV-optimal bandwidth of 36 days. The in-sample estimates are printed in grey, out-of-sample volatilities are black depicted.

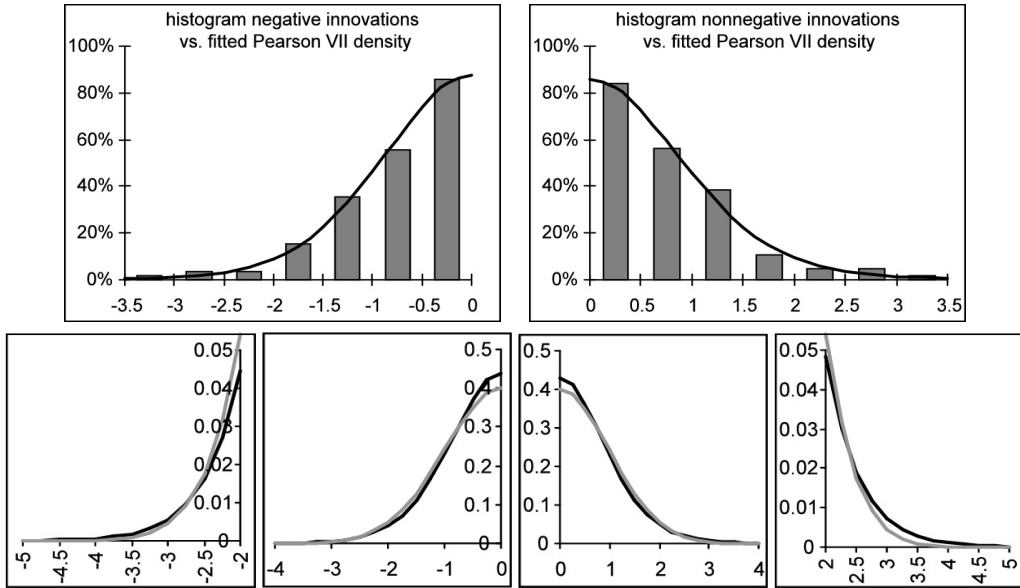


Figure 7: Pearson VII fit of negative and nonnegative innovations by the density $f_{5.63,2.90,7.20,3.28}^{VII}$. The bottom graphs compare the asymmetric Pearson VII (black) and the standard normal density (grey).

extreme Pearson quantiles differ significantly from normal quantiles.²⁹ There are only slight differences of the optimal left and right fitted distribution of innovations due to the simulation setup.

It should be attended, that in general a trade-off between the smoothness of volatility estimates and the innovation’s distribution was observed: The bigger the bandwidth in nonparametric volatility estimation the smoother is the volatility graph and the more heavy-tailed are the innovations.³⁰

To measure the modelling performance we apply the Kupiec test to shortfall rates of the out-of-sample-part. The two-sided hypothesis test is an extension of a binomial test for the likelihood of N shortfalls in a sample of size n , where the true shortfall probability is hypothetical $H_0 : p = \alpha$ for an $(1 - \alpha)$ VaR-level. Based on a normal

²⁹Concerning the input parameters (Pearson density $f_{4,\sqrt{5}}^{VII(2)}$), we observe that lower confidence levels (in terms of maximum losses), as 95%, may have Pearson quantiles that are absolutely smaller than the normal (-1.6012 Pearson vs. -1.6449 Gaussian). But for extreme shortfall levels, as 99% or 99.5%, Pearson overtakes the Gaussian quantiles (99%: -2.5337 vs. -2.3263 ; 99.5%: -2.9576 vs. -2.5758). In other words, the probability of falling short the 0.5%-normal quantile is for the Pearson VII distributed rv almost twice as much (0.93%).

³⁰On the other hand, this is the reason why the Pearson VII fit of innovations may fail in some cases for good volatility estimates with small bandwidths, entailing a kurtosis < 3 of estimated innovations. Then the normal distribution may be conservative for the task of extreme return quantile approximation.

approximation, Kupiec (1995) developed approximate 95% confidence regions of failure rates. The log-likelihood-ratio

$$LR_p = -2 \ln [(1-p)^{n-N} p^N] + 2 \ln \left[\left(1 - \frac{N}{n}\right)^{n-N} \left(\frac{N}{n}\right)^N \right] \quad (42)$$

is χ_1^2 -distributed under H_0 . Thus the risk measure and H_0 are rejected on a 5% level of significance if $LR_p > 3.84$. This test is widely used by financial risk managers to evaluate their risk models, even the penalty zones of the Basel II committee are based on this methodology (see e.g. Jorion (2006)). The Kupiec test is used by scientific as well to support return models, e.g. in Choi and Nam (2008).

Primarily, we use Kupiec backtesting for the non-stationary modelling of the price process (41) to evaluate the forecasted daily returns in terms of certain quantiles. We focus on the loss tail and deduce a relative 1-day Value at Risk $VaR^{1-\alpha, 1d}(t)$ for the next day's return X_{t+1} corresponding to the regression model (10):

$$VaR^{1-\alpha, 1d}(t) = \bar{X}_t + \hat{\sigma}_{(1)}(t) u_{m,c;\alpha^{(1)}} = \bar{X}_t + \hat{\sigma}_{(1)}(t) \gamma t_{2m-1;\alpha^{(1)}}, \quad \gamma = \frac{c}{\sqrt{2m-1}}, \quad (43)$$

where $u_{m,c;\alpha^{(1)}}$ and $t_{2m-1;\alpha^{(1)}}$, with $\alpha^{(1)} = 2\alpha$, are the quantiles of the left-side Pearson VII fit and the corresponding Student- t expression. This forecast for a maximum loss, that is not exceeded at the end of the horizon $(t, t+1]$ with probability $1-\alpha$, is compared to the realized next days return X_{t+1} . A shortfall is observed if $X_{t+1} \leq VaR^{1-\alpha, 1d}(t)$. The number of shortfalls in the out-of-sample part is accumulated and evaluated with the Kupiec test statistic LR_p . Illustration 8 summarizes for a 99% confidence level of maximum losses.

LR test (Kupiec-test), 5% level	
VaR confidence level (as $1-\alpha$)	99.0%
observations (sample size)	249
number of VaR exceedances	5
shortfall probability p	1.0%
expected exceedances of VaR	2.5
significance level of LR-Test (as $1-\beta$)	95%
c-value χ^2 -distribution (one-sided)	3.84
test statistic (LR)	1.98
	1.98 < 3.84
Non-stationary model acceptance	yes

Figure 8: Kupiec backtesting of forecasted $VaR^{99\%, 1d}(t)$ against realized returns X_{t+1} (out-of-sample) in the non-stationary model.

The non-stationary forecasts experienced 5 exceedances over the VaR threshold, which deviates slightly from the expected number of shortfalls, but is within the allowed range of $\{1, \dots, 6\}$ for the 5% level of significance. The model is accepted by the Kupiec test. Furthermore, acceptance was shown for all possible confidence levels greater-than-or-equal 80%. Most exceedances occurred generally, when the volatility was at a low and started again to increase. The example was chosen with quite extreme volatility changes and short periods of the oscillation. Nevertheless the non-stationary approach works.

It could be supposed that a parametric VaR based on a long-term standard deviation is advantaged, since it covers more periods of the recurring volatility structure. That's why we backtest a simple risk model, based on Gaussian quantiles and scaled by empirical standard deviations of historical returns (from a moving window of certain length), the same way and compare both model performances. Figure 9 gives the number of exceedances for the different models. Based on a 250-day standard deviation (fourth column), the basic risk model seems to benefit from the higher average volatility: Less exceedances occur for the most confidence levels and the number of shortfalls is closer to the expectation for levels from 95% to 99.5%. But this VaR model is too conservative, it is rejected on lower levels where certain exceedances are required. Based on a window adjusted to the bandwidth of the one-sided NWE (right column) the parametric VaR model passes the lower but fails at some higher confidence levels.³¹

³¹The basic risk models, based on standard deviations over 50 days and 36 days, respectively, show on the given confidence levels the same number of shortfalls.

VaR confidence level (as $1-\alpha$)	number of exceedances, model rejection			
	expectation	non-stationary model	normal model (250 days)	normal model (50d, 36d)
80.00%	49.8	45	32	38
90.00%	24.9	29	15	19
95.00%	12.5	16	10	16
98.00%	5.0	6	4	7
99.00%	2.5	5	2	5
99.50%	1.2	3	1	4
99.90%	0.2	1	1	1
99.95%	0.1	0	1	1
99.99%	0.0	0	0	1

Figure 9: Kupiec test results for the non-stationary model and two types of a Gaussian risk model, scaled by empirical standard deviations (over different horizons), on several levels of confidence. Model rejections on a 5% level of significance are grey highlighted.

Concluding, the simulation experiment proved that the non-stationary model is not only of theoretical quality: Moreover it was able to capture simulated return dynamics and to provide satisfactory distributional forecasts as well. Finally, the regression approach outperformed a standard VaR model, although a fast oscillating price process was chosen to challenge that implementation.

In Gürtler and Rauh (2009), the same non-stationary model is executed for real financial times series, as series of equity indices, interest rates, credit spreads and currencies. They observe satisfactory model approximations to the return series, where volatility clusters are detected by the nonparametric estimates and extremal or asymmetric behavior is incorporated by the Pearson VII fit of innovations. An excellent modelling of empirical data is confirmed again by Kupiec backtesting and an outperformance against the standard Gaussian VaR model is proved.

4 Summary and conclusion

In this paper we discuss a non-stationary, heteroscedastic model for financial returns theoretically. The multiplicative return model, introduced first by Drees and Starica (2002) and Herzel et al. (2005), works with a nonparametric regression approach for unconditional volatility and an asymmetric, heavy-tailed fit of random innovations. Based on the idea that recent past and future returns depend on the same unknown exogenous economic factors, that evolve smoothly through time and are manifested in pricing, the current level of (co-)variance is dominated exogenously. The vectors of returns are assumed to be independent while having an unconditional covariance structure that is a smooth deterministic function of time, modelled via nonparametric regression on equidistant centered returns (Nadaraya-Watson estimators, NWE). For the purpose of accurate modelling of the residuals, an asymmetric version of the Pearson type VII distribution is applied. We specialize on a univariate description.

Our article is thought as the theoretical background for the empirical research of Gürtler and Rauh (2009) in the non-stationary approach. So, we focus here on deriving the modelling components. A lot of effort is spent to statistical properties as consistency and asymptotic normality of the variance estimates in the nonparametric regression model (with constant trend) on an equidistant design. Both features are proved analytically for the symmetrical NWE as well as for the one-sided NWE, which is based only on past/ recent data and should be used in forecasting. The statistical analysis requires a special kernel, where the biweight kernel is established, and appropriate bandwidths. Cross-validation is adopted for an automatised bandwidth selection. The task of fitting innovations via Pearson type VII is simplified by providing a method of moments for parameter estimation and by deriving a connection to the Student- t distribution. By dint of the latter presentation of residuals a factor-based VaR calculation can be implemented in terms of the regression model: The univariate $VaR^{1-\alpha}(t)$ of an exposure $w(t)$ can be modelled as its product with the nonparametric estimated volatility $\hat{\sigma}_{(\cdot)}(t)$ and the $(1-\alpha)$ Pearson VII innovation quantile of the benchmark return distribution, adjusted by the mean return.

Besides the statistical results of volatility estimates, we document in simulation studies how their fit to a predefined function is improved by a more and more refined data base, and that the non-stationary model is able to capture price processes entirely. A Kupiec hypothesis testing confirms the goodness of distributional forecasts, where value at risk

limits are compared to the next day's returns and are evaluated by their shortfall rates. Regarding that backtesting method the non-stationary model clearly outperforms a parametric (normal) VaR model.

Gürtler and Rauh (2009) provide evidence that the regression model approximates daily return series of several financial instruments successfully. It outperforms parametric VaR implementations as they test model assumptions and evaluate shortfall rates. Extensive empirical studies focus on the practical usage of the nonparametric volatility estimations and the connection to the Pearson VII fit of innovations. Drees and Starica (2002) establish on a S&P 500 return series that their equivalent approach fits the data significantly better than conventional GARCH-type models, and provided a forecasting analysis with 1-, 20-, and 40-days ahead distributional forecasts. Herzel et al. (2005) show on a tri-variate example (exchange rate EUR/USD, FTSE 100 index, 10-year US T-bond rate) that the non-stationary paradigm describes the dynamics well and delivers good multivariate distributional forecasts. They proved an outperformance against the *RiskMetricsTM* (JP Morgan) approach. Last but not least, Mikosch and Starica (2004a) extend the non-stationary, nonparametric framework by including a time-varying expected return in the univariate case, giving statistical evidence that the return expectation and the market price of risk vary significantly through time.

Having outlined the practical advantages and having proved the theoretical model quality in this article, we see the following fields for future research: The main task will be to develop an adequate multivariate setup for a broad exposure conception. This could be based on the direct multivariate setup of Herzel et al. (2005). Otherwise, a risk aggregation in a simulation approach could be fruitful, e.g. approximatively with Cholesky decomposition of correlations, that continues on the empirically approved univariate implementation for VaR-purposes. A full nonparametric setup is conceivable, where the (still restrictive) parametric approach for the innovation's distribution could be substituted with a nonparametric kernel density. Following Mikosch and Starica (2004a), the inclusion of a time-dependent expected yield, modelled with kernel regression, may be a further step. Based on our basic belief that both recent past and future returns are manifestations of the same unspecified, exogenous economic factors, that evolve smoothly through time, we may use that frame for portfolio optimization in terms of a tactical asset allocation.

A Proofs for section 2.1

Proof of proposition 2.1:

1.) We first show that $\hat{s}_{h_n}^2(u)$ is asymptotically unbiased for all $u \in (0, 1)$.

$$E\hat{s}_{h_n}^2(u) = \frac{\frac{1}{n} \sum_{i=1}^n K_{h_n}\left(\frac{i}{n} - u\right) ER_{i,n}^2}{\frac{1}{n} \sum_{i=1}^n K_{h_n}\left(\frac{i}{n} - u\right)}$$

1.a.) The denominator is for sufficient large n approximatively

$$\frac{1}{h_n} \sum_{i=1}^n \frac{1}{n} K\left(\frac{\frac{i}{n} - u}{h_n}\right) \approx \frac{1}{h_n} \int_0^1 K\left(\frac{x - u}{h_n}\right) dx = \frac{1}{h_n} \int_{u-h_n}^{u+h_n} K\left(\frac{x - u}{h_n}\right) dx \stackrel{v:=\frac{x-u}{h_n}}{=} \frac{1}{h_n} \int_{-1}^1 K(v) h_n dv \stackrel{(C1)}{=} 1$$

The goodness of the approximation is to be seen with the mean value theorem for integrals and the mean value theorem for differentiation (where $\tilde{x}_i \in \left[\frac{i-1}{n}, \frac{i}{n}\right]$ and $\xi_i \in \left(\tilde{x}_i, \frac{i}{n}\right)$):

$$\int_0^1 K\left(\frac{x - u}{h_n}\right) dx = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} K\left(\frac{x - u}{h_n}\right) dx = \sum_{i=1}^n \frac{1}{n} K\left(\frac{\tilde{x}_i - u}{h_n}\right)$$

$$\begin{aligned} \frac{1}{nh_n} \left| \sum_{i=1}^n K\left(\frac{\frac{i}{n} - u}{h_n}\right) - K\left(\frac{\tilde{x}_i - u}{h_n}\right) \right| &= \frac{1}{nh_n} \left| \sum_{i=1}^n \underbrace{K'\left(\frac{\xi_i - u}{h_n}\right)}_{\leq M := \max_i |K'(\cdot)|} \cdot \underbrace{\left(\frac{\frac{i}{n} - \tilde{x}_i}{h_n}\right)}_{\leq \frac{1}{nh_n}} \right| \leq \frac{1}{nh_n} n \left(M \frac{1}{nh_n}\right) \\ &= O\left(\frac{1}{nh_n^2}\right) \stackrel{(C3)}{\rightarrow} 0 \end{aligned}$$

And the denominator has the deterministic limit value 1.

1.b.) Regarding the enumerator it holds for the expectation of $R_{i,n}^2$ that

$$ER_{i,n}^2 = s^2 \left(\frac{i}{n} \right) - \frac{2}{n} s^2 \left(\frac{i}{n} \right) + \frac{1}{n^2} \sum_{j=1}^n s^2 \left(\frac{j}{n} \right),$$

where the most effort is spent to the first terms scalar product with the kernel.

1.b.i.) Here we start again with an approximation, substitute $v := \frac{x-u}{h_n}$ and use a Taylor expansion (where $\hat{u} \in (u, u + vh_n)$ if $v > 0$, else $\hat{u} \in (u + vh_n, u)$) for assessing convergence.

$$\begin{aligned} & \frac{1}{h_n} \sum_{i=1}^n \frac{1}{n} K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) \approx \frac{1}{h_n} \int_0^1 K \left(\frac{x-u}{h_n} \right) s^2(x) dx = \frac{1}{h_n} \int_{u-h_n}^{u+h_n} K \left(\frac{x-u}{h_n} \right) s^2(x) dx \\ &= \frac{1}{h_n} \int_{-1}^1 K(v) s^2(u + vh_n) h_n dv = s^2(u) \underbrace{\int_{-1}^1 K(v) dv}_{=1, (C1)} + (s^2(u))' h_n \underbrace{\int_{-1}^1 v K(v) dv}_{=0, (C1)} + \frac{h_n^2}{2} \underbrace{\int_{-1}^1 (s^2(\hat{u}))'' v^2 K(v) dv}_{< \infty, (C4), (C2)} \\ &= s^2(u) + O(h_n^2) \rightarrow s^2(u) \end{aligned}$$

The goodness of the approximation is shown with mean value theorems (where $\tilde{x}_i \in [\frac{i-1}{n}, \frac{i}{n}]$ and $\xi_i, \zeta_i \in (\tilde{x}_i, \frac{i}{n})$):

$$\begin{aligned} & \int_0^1 K \left(\frac{x-u}{h_n} \right) s^2(x) dx = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} K \left(\frac{x-u}{h_n} \right) s^2(x) dx = \sum_{i=1}^n \frac{1}{n} K \left(\frac{\tilde{x}_i - u}{h_n} \right) s^2(\tilde{x}_i) \\ & \frac{1}{nh_n} \left| \sum_{i=1}^n K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) - K \left(\frac{\tilde{x}_i - u}{h_n} \right) s^2(\tilde{x}_i) \right| \\ &= \frac{1}{nh_n} \left| \sum_{i=1}^n s^2 \left(\frac{i}{n} \right) \left[K \left(\frac{\frac{i}{n} - u}{h_n} \right) - K \left(\frac{\tilde{x}_i - u}{h_n} \right) \right] + K \left(\frac{\tilde{x}_i - u}{h_n} \right) \left[s^2 \left(\frac{i}{n} \right) - s^2(\tilde{x}_i) \right] \right| \\ &= \frac{1}{nh_n} \left| \sum_{i=1}^n s^2 \left(\frac{i}{n} \right) \underbrace{K' \left(\frac{\xi_i - u}{h_n} \right)}_{\leq M := \max_i |K'(\cdot)|} \cdot \underbrace{\left(\frac{\frac{i}{n} - \tilde{x}_i}{h_n} \right)}_{\leq \frac{1}{nh_n}} + K \left(\frac{\tilde{x}_i - u}{h_n} \right) \underbrace{(s^2(\zeta_i))'}_{\leq N := \max_i |(s^2(\zeta_i))'|} \cdot \underbrace{\left(\frac{i}{n} - \tilde{x}_i \right)}_{\leq \frac{1}{n}} \right| \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n s^2 \left(\frac{i}{n} \right) \frac{M}{nh_n} + K \left(\frac{\tilde{x}_i - u}{h_n} \right) \frac{N}{n} = O \left(\frac{1}{nh_n^2} \right) + O \left(\frac{1}{nh_n} \right) \rightarrow 0 \end{aligned}$$

Putting both limit considerations together, we get with the triangle inequality:

$$\begin{aligned} & \left| \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) - s^2(u) \right| \leq \left| \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) - \frac{1}{h_n} \int_0^1 K \left(\frac{x-u}{h_n} \right) s^2(x) dx \right| \\ &+ \left| \frac{1}{h_n} \int_0^1 K \left(\frac{x-u}{h_n} \right) s^2(x) dx - s^2(u) \right| = O \left(\frac{1}{nh_n^2} \right) + O(h_n^2) \stackrel{(C3)}{\rightarrow} 0 \end{aligned}$$

And it follows that $\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) s^2 \left(\frac{i}{n} \right) \rightarrow s^2(u)$.

1.b.ii.) Paying attention to the residual components of the enumerator it is easy to see that:

$$-\frac{2}{n^2} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) s^2 \left(\frac{i}{n} \right) \rightarrow 0$$

with factorisation in $(-\frac{2}{n})$ and the expression of the previous part. Finally,

$$\frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n K_{h_n} \left(\frac{i}{n} - u \right) s^2 \left(\frac{j}{n} \right) \leq \frac{1}{n^3 h_n} n^2 \underbrace{\max_{i=1, \dots, n} K \left(\frac{\frac{i}{n} - u}{h_n} \right) \max_{j=1, \dots, n} s^2 \left(\frac{j}{n} \right)}_{< \infty, (C2), (C4)} \rightarrow 0$$

It immediately follows with the convergence in numerator and denominator: $E\hat{s}_{h_n}^2(u) \xrightarrow{n \rightarrow \infty} s^2(u) \forall u \in (0, 1)$.

2.) With a similar, more expensive analysis we prove that the variance of $\hat{s}_{h_n}^2(u)$ converges to zero as n goes to infinity:

$$\text{Var } \hat{s}_{h_n}^2(u) = \frac{\frac{1}{n^2} \text{Var} \left\{ \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) \left(s \left(\frac{i}{n} \right) \varepsilon_{i,n} - \frac{1}{n} \sum_{j=1}^n s \left(\frac{j}{n} \right) \varepsilon_{j,n} \right)^2 \right\}}{\frac{1}{n^2} \left(\sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) \right)^2}$$

2.a.) The denominator is the squared expression of 1.a.), i.e. its limit value is 1.

2.b.) We decompose the numerator in several sums of variances and covariances.

2.b.i.) The central component (with the lowest convergence rate) is:

$$\frac{1}{n^2} \text{Var} \left\{ \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) s^2 \left(\frac{i}{n} \right) \varepsilon_{i,n}^2 \right\} \stackrel{iid}{=} \frac{1}{n^2} \sum_{i=1}^n K_{h_n}^2 \left(\frac{i}{n} - u \right) s^4 \left(\frac{i}{n} \right) V,$$

where $V := \text{Var} \varepsilon_{1,n}^2 = E\varepsilon_{1,n}^4 - 1 \stackrel{(C5)}{<} \infty$. With the same technique as above, we approximate initially:

$$\begin{aligned} & \frac{V}{nh_n^2} \sum_{i=1}^n \frac{1}{n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \approx \frac{V}{nh_n^2} \int_0^1 K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx = \frac{V}{nh_n^2} \int_{-1}^1 K^2(v) s^4(u + vh_n) h_n dv \\ &= \frac{V s^4(u)}{nh_n} \int_{-1}^1 K^2(v) dv + \underbrace{\frac{V (s^4(u))'}{n} \int_{-1}^1 v K^2(v) dv}_{=0, \text{ symmetry}} + \frac{V h_n}{2n} \int_{-1}^1 (s^4(\hat{u}))'' v^2 K^2(v) dv \\ &= O\left(\frac{1}{nh_n}\right) + O\left(\frac{h_n}{n}\right) \rightarrow 0 \end{aligned}$$

The absolute approximation error is again a zero-sequence:

$$\begin{aligned} & \int_0^1 K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx = \sum_{i=1}^n \frac{1}{n} K^2 \left(\frac{\tilde{x}_i - u}{h_n} \right) s^4(\tilde{x}_i) \\ & \frac{V}{n^2 h_n^2} \left| \sum_{i=1}^n K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) - K^2 \left(\frac{\tilde{x}_i - u}{h_n} \right) s^4(\tilde{x}_i) \right| \\ &= \frac{V}{n^2 h_n^2} \left| \sum_{i=1}^n s^4 \left(\frac{i}{n} \right) \left(K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) - K^2 \left(\frac{\tilde{x}_i - u}{h_n} \right) \right) + K^2 \left(\frac{\tilde{x}_i - u}{h_n} \right) \left(s^4 \left(\frac{i}{n} \right) - s^4(\tilde{x}_i) \right) \right| \\ &= \frac{V}{n^2 h_n^2} \left| \sum_{i=1}^n s^4 \left(\frac{i}{n} \right) \underbrace{\left(K^2 \left(\frac{\xi_i - u}{h_n} \right) \right)'}_{=2K \left(\frac{\xi_i - u}{h_n} \right) K' \left(\frac{\xi_i - u}{h_n} \right) \leq 2\tilde{M}} \left(\frac{\frac{i}{n} - \tilde{x}_i}{h_n} \right) + K^2 \left(\frac{\tilde{x}_i - u}{h_n} \right) \underbrace{\left(s^4(\zeta_i) \right)'}_{\leq \tilde{N}} \left(\frac{i}{n} - \tilde{x}_i \right) \right| \\ &\leq \frac{V}{n^2 h_n^2} \sum_{i=1}^n s^4 \left(\frac{i}{n} \right) \frac{2\tilde{M}}{nh_n} + K^2 \left(\frac{\tilde{x}_i - u}{h_n} \right) \frac{\tilde{N}}{n} = O\left(\frac{1}{n^2 h_n^3}\right) + O\left(\frac{1}{n^2 h_n^2}\right) \rightarrow 0 \end{aligned}$$

By dint of the triangle inequality it follows altogether:

$$\begin{aligned} 0 &\leq \frac{V}{n^2 h_n^2} \sum_{i=1}^n K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \leq \left| \frac{V}{n^2 h_n^2} \sum_{i=1}^n K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) - \frac{V}{nh_n^2} \int_0^1 K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx \right| \\ &+ \frac{V}{nh_n^2} \int_0^1 K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx = O\left(\frac{1}{n^2 h_n^3}\right) + O\left(\frac{1}{nh_n}\right) \rightarrow 0 \end{aligned}$$

From this results that $\frac{1}{n^2} \sum_{i=1}^n K_{h_n}^2 \left(\frac{i}{n} - u \right) s^4 \left(\frac{i}{n} \right) V \rightarrow 0$.

2.b.ii.) All other components can be absolutely restricted to zero (with a faster rate of convergence). For lack of space we do not write out the full calculation.³²

Herewith it immediately follows that $\text{Var} \hat{s}_{h_n}^2(u) \xrightarrow{n \rightarrow \infty} 0 \forall u \in (0, 1)$.

3. The stochastic convergence is finally derived with the Markov inequality. Formally, let $\{P_{n,s^2} : s^2 \in \mathbb{R}_0^+\}_{n \in \mathbb{N}}$ be the corresponding sequence of statistical models for the sequence of estimators $(\hat{s}_{h_n}^2)_{n \in \mathbb{N}}$:

$$P_{n,s^2} \{ |\hat{s}_{h_n}^2(u) - s^2(u)| \geq \epsilon \} \leq \frac{E |\hat{s}_{h_n}^2(u) - s^2(u)|^2}{\epsilon^2} = \frac{\text{Var}(\hat{s}_{h_n}^2(u))}{\epsilon^2} + \frac{(E(\hat{s}_{h_n}^2(u) - s^2(u)))^2}{\epsilon^2}$$

$$\xrightarrow{n \rightarrow \infty} 0 \forall \epsilon > 0$$

I.e. $\hat{s}_{h_n}^2(u) \xrightarrow{P} s^2(u)$ for all $s^2 \in \mathbb{R}_0^+$, $u \in (0, 1)$. □

Proof of proposition 2.2:

A useful decomposition of $Z_{n,h_n}(u) = \sqrt{nh_n} (\hat{s}_{h_n}^2(u) - s^2(u))$ is:

$$Z_{n,h_n}(u) = \left. \sqrt{nh_n} \frac{\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) (R_{i,n}^2 - ER_{i,n}^2)}{\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right)} \right\} \text{variance-part}$$

$$+ \left. \sqrt{nh_n} \frac{\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) (ER_{i,n}^2 - s^2(u))}{\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right)} \right\} \text{bias-part}$$

Since the first summand is a centered rv we refer to as variance-part. The second term is deterministic and may contain a systematic bias of expected squared returns to the curvature $s^2(u)$.

1.) We concentrate first on both moments:

1.a.) With the expression of $ER_{i,n}^2$ as used in the proof of proposition 2.1 we rewrite the expectation of $Z_{n,h_n}(u)$:

$$EZ_{n,h_n}(u) = \frac{\frac{\sqrt{nh_n}}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) \left[s^2 \left(\frac{i}{n} \right) - \frac{2}{n} s^2 \left(\frac{i}{n} \right) + \frac{1}{n^2} \sum_{j=1}^n s^2 \left(\frac{j}{n} \right) - s^2(u) \right]}{\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right)}$$

The denominator converges to 1 as seen before. For the numerator we pick up the results of proof 2.1 as well. Concerning the first and fourth summand in brackets we repeat the steps of 1.b.i.) and arrive at:

$$\frac{\sqrt{nh_n}}{h_n} \sum_{i=1}^n \frac{1}{n} K \left(\frac{\frac{i}{n} - u}{h_n} \right) \left(s^2 \left(\frac{i}{n} \right) - s^2(u) \right) \rightarrow \frac{C}{2} (s^2(u))'' \int_{-1}^1 v^2 K(v) dv,$$

where $C \geq 0$ with $nh_n^5 \xrightarrow{n \rightarrow \infty} C^2$ as postulated in (C3). The sums of kernels multiplied with the second and third term in brackets are still zero-sequences (compare point 1.b.ii.) of proof 2.1), the factorisation with $\sqrt{nh_n}$ only slows up its convergences. This implies for the numerator as well as for the whole expectation of $Z_{n,h_n}(u) \forall u \in (0, 1)$ a pointwise limiting value of $\beta(u)$:

$$EZ_{n,h_n}(u) \xrightarrow{n \rightarrow \infty} \frac{C}{2} (s^2(u))'' \int_{-1}^1 v^2 K(v) dv$$

³²Interested readers are welcome to ask us for the full proof. The methods are to a certain extent similar to the previous. The Cauchy-Schwarz inequality was applied for covariance terms.

1.b.) We rephrase the variance of $Z_{n,h_n}(u)$ as:

$$\begin{aligned} \text{Var} Z_{n,h_n}(u) &= nh_n \text{Var} \left\{ \frac{\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) R_{i,n}^2}{\frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right)} \right\} \\ &= \frac{nh_n \text{Var} \left\{ \frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) \left(s \left(\frac{i}{n} \right) \varepsilon_{i,n} - \frac{1}{n} \sum_{j=1}^n s \left(\frac{j}{n} \right) \varepsilon_{j,n} \right)^2 \right\}}{\frac{1}{n^2} \left(\sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) \right)^2} \end{aligned}$$

Again the denominator converges to 1. By enhancing the results of proof 2.1, part 2.b.i.), the dominant term of the numerator and its limit value are:

$$nh_n \frac{1}{n^2} \sum_{i=1}^n K_{h_n}^2 \left(\frac{i}{n} - u \right) s^4 \left(\frac{i}{n} \right) V = \frac{V}{h_n} \sum_{i=1}^n \frac{1}{n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \rightarrow V s^4(u) \int_{-1}^1 K^2(v) dv,$$

where $V := E\varepsilon_{1,n}^4 - 1$. The other components hold the zero limit value (due to faster convergence in proof 2.1). This implies for the numerator of the variance-part as well as for the whole variance of $Z_{n,h_n}(u) \forall u \in (0, 1)$ a pointwise limiting value of $\tau^2(u)$:

$$\text{Var} Z_{n,h_n}(u) \xrightarrow{n \rightarrow \infty} V s^4(u) \int_{-1}^1 K^2(v) dv$$

2.) To prove the convergence in distribution, we have to focus mainly on the stochastic numerator of the variance-part, which can be decomposed in:

$$\begin{aligned} &\underbrace{\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n K \left(\frac{\frac{i}{n} - u}{h_n} \right) \left(s^2 \left(\frac{i}{n} \right) (\varepsilon_{i,n}^2 - 1) \right)}_{=: m_n} \\ &+ \underbrace{\sqrt{nh_n} \frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) \left(-\frac{2}{n} \sum_{j=1}^n s \left(\frac{i}{n} \right) s \left(\frac{j}{n} \right) \varepsilon_{i,n} \varepsilon_{j,n} + \frac{2}{n} s^2 \left(\frac{i}{n} \right) \right)}_{=: p_n} \\ &+ \underbrace{\sqrt{nh_n} \frac{1}{n} \sum_{i=1}^n K_{h_n} \left(\frac{i}{n} - u \right) \left(\frac{1}{n^2} \left[\sum_{j=1}^n s \left(\frac{j}{n} \right) \varepsilon_{j,n} \right]^2 - \frac{1}{n^2} \sum_{j=1}^n s^2 \left(\frac{j}{n} \right) \right)}_{=: q_n} \end{aligned}$$

2.a.) We apply the central limit theorem of Lindeberg-Feller³³ to show the distributional convergence of m_n . The triangle-schema is

$$Y_{n,i} := \frac{1}{\sqrt{nh_n}} K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) (\varepsilon_{i,n}^2 - 1),$$

with $\mu_{n,i} = \frac{1}{\sqrt{nh_n}} K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) E(\varepsilon_{i,n}^2 - 1) = 0$, $\sigma_{n,i}^2 = \frac{V}{nh_n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \in (0, \infty)$ and $s_n^2 = \frac{V}{h_n} \sum_{i=1}^n \frac{1}{n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \rightarrow V s^4(u) \int_{-1}^1 K^2(v) dv$.

³³CLT of Lindeberg-Feller: For each $n \in \mathbb{N}$ let $Y_{n,1}, \dots, Y_{n,n}$ be independent rvs with $EY_{n,i} = \mu_{n,i}$, $\text{Var} Y_{n,i} = \sigma_{n,i}^2 \in (0, \infty)$, $s_n^2 := \sum_{i=1}^n \sigma_{n,i}^2$ and $L_n(\epsilon) := \frac{1}{s_n^2} E \sum_{i=1}^n [(Y_{n,i} - \mu_{n,i})^2 I\{|Y_{n,i} - \mu_{n,i}| > \epsilon s_n\}]$. If $L_n(\epsilon) \rightarrow 0$ for all $\epsilon > 0$ ('Lindeberg-condition'), then $\frac{1}{s_n} \sum_{i=1}^n (Y_{n,i} - \mu_{n,i}) \xrightarrow{D} \mathcal{N}(0, 1)$.

For the verification of the Lindeberg condition we apply the Hölder inequality and the Markov inequality to get the following upper bound:³⁴

$$L_n(\epsilon) \leq \frac{\sum_{i=1}^n \left(\frac{1}{nh_n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \right)^{1 + \frac{\delta'}{1 + \delta'}} W}{\epsilon^{\frac{2\delta'}{1 + \delta'}} (s_n^2)^{1 + \frac{\delta'}{1 + \delta'}}},$$

with an upper bound $W \in (0, \infty)$ and $\delta' := \frac{\delta}{4}$ with δ from condition (C5). The finite, non-zero convergence of the denominator follows from the limit value of s_n^2 . For the convergence analysis of the numerator similar methods as seen before can be applied, leading to a limiting value of 0. Hence the Lindeberg condition holds. With a Slutsky argument it follows that

$$\begin{aligned} & \sum_{i=1}^n Y_{n,i} \xrightarrow{D} \mathcal{N} \left(0, \lim_{n \rightarrow \infty} s_n^2 \right) \\ \iff & \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) (\varepsilon_{i,n}^2 - 1) \xrightarrow{D} \mathcal{N} \left(0, V s^4(u) \int_{-1}^1 K^2(v) dv \right) \end{aligned}$$

2.b.) It can be proven that $E p_n = 0$ and $Var p_n \rightarrow 0$. With the Markov inequality we conclude (compare proof 2.1, part 3.) that $p_n \xrightarrow{P} 0$. It follows the same way that $q_n \xrightarrow{P} 0$. With the Slutsky theorem we combine the results:

$$m_n + p_n + q_n \xrightarrow{D} \mathcal{N} \left(0, V s^4(u) \int_{-1}^1 K^2(v) dv \right)$$

Due to the deterministic limit 1 of the denominator (and by applying Slutsky again) the limit distribution persists for the whole variance-part.

2.c.) Finally, we add the deterministic bias-part and the asymptotic distribution of $Z_{n,h_n}(u)$ results. □

Proof of proposition 2.3:

1.) We first show that $\hat{s}_{(1)h_n}^2(u)$ is asymptotically unbiased. For sufficiently large n (else case differentiation as in equation (20) necessary) its expectation is:

$$E \hat{s}_{(1)h_n}^2(u) = \frac{\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) E \tilde{R}_{i,n}^2}{\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right)}.$$

1.a.) The denominator is approximately

$$\frac{1}{h_n} \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K \left(\frac{\frac{i}{n} - u}{h_n} \right) \approx \frac{1}{h_n} \int_0^u K \left(\frac{x - u}{h_n} \right) dx = \frac{1}{2},$$

the last equality is to be seen by executing the steps of proof 2.1, part 1.a.). The approximation quality follows with two mean value theorems (where $\tilde{x}_i \in [\frac{i-1}{n}, \frac{i}{n}]$, $\tilde{x} \in [\frac{\lfloor un \rfloor}{n}, u]$ and $Res_n := (un - \lfloor un \rfloor) K \left(\frac{\tilde{x} - u}{h_n} \right) < \infty \forall n$; $\xi_i \in (\tilde{x}_i, \frac{i}{n})$):

$$\begin{aligned} & \int_0^u K \left(\frac{x - u}{h_n} \right) dx = \sum_{i=1}^{\lfloor un \rfloor} \int_{\frac{i-1}{n}}^{\frac{i}{n}} K \left(\frac{x - u}{h_n} \right) dx + \int_{\frac{\lfloor un \rfloor}{n}}^u K \left(\frac{x - u}{h_n} \right) dx \\ & = \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K \left(\frac{\tilde{x}_i - u}{h_n} \right) + \frac{un - \lfloor un \rfloor}{n} K \left(\frac{\tilde{x} - u}{h_n} \right) = \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K \left(\frac{\tilde{x}_i - u}{h_n} \right) + \frac{Res_n}{n} \end{aligned}$$

³⁴The complete calculation is available upon request.

$$\begin{aligned}
& \frac{1}{nh_n} \left| \sum_{i=1}^{\lfloor un \rfloor} K \left(\frac{i-u}{h_n} \right) - K \left(\frac{\tilde{x}_i-u}{h_n} \right) - Res_n \right| \leq \frac{1}{nh_n} \left| \sum_{i=1}^{\lfloor un \rfloor} \underbrace{K' \left(\frac{\xi_i-u}{h_n} \right)}_{\leq M := \max_i |K'(\cdot)|} \cdot \underbrace{\left(\frac{i-\xi_i}{h_n} \right)}_{\leq \frac{1}{nh_n}} \right| + \frac{Res_n}{nh_n} \\
& \leq \frac{1}{nh_n} un \left(M \frac{1}{nh_n} \right) + \frac{Res_n}{nh_n} = O \left(\frac{1}{nh_n^2} \right) + O \left(\frac{1}{nh_n} \right) \xrightarrow{(C3')} 0
\end{aligned}$$

And the denominator has the deterministic limit value $\frac{1}{2}$.

1.b.) Regarding the enumerator we use $E\tilde{R}_{1,n}^2 = s^2 \left(\frac{1}{n} \right)$ by definition and

$$E\tilde{R}_{i,n}^2 = s^2 \left(\frac{i}{n} \right) + \frac{1}{(i-1)^2} \sum_{j=1}^{i-1} s^2 \left(\frac{j}{n} \right) \text{ for } i \geq 2.$$

Summing up the numerator has the form

$$\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) s^2 \left(\frac{i}{n} \right) + \frac{1}{n} \sum_{i=2}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \frac{1}{(i-1)^2} \sum_{j=1}^{i-1} s^2 \left(\frac{j}{n} \right)$$

1.b.i.) We turn to the limit of the first sum. After approximating, we repeat the steps (integral transformation, Taylor expansion) of proof 2.1, part 1.b.):

$$\begin{aligned}
& \frac{1}{h_n} \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K \left(\frac{i}{n} - u \right) s^2 \left(\frac{i}{n} \right) \approx \frac{1}{h_n} \int_0^u K \left(\frac{x-u}{h_n} \right) s^2(x) dx \\
& = \underbrace{s^2(u) \int_{-1}^0 K(v) dv}_{=\frac{1}{2}} + \underbrace{(s^2(u))' h_n \int_{-1}^0 v K(v) dv}_{< \infty} + \underbrace{\frac{h_n^2}{2} \int_{-1}^0 (s^2(\hat{u}))'' v^2 K(v) dv}_{< \infty} \\
& = \frac{s^2(u)}{2} + O(h_n) + O(h_n^2) \rightarrow \frac{s^2(u)}{2}
\end{aligned}$$

The approximation goodness is shown using mean value theorems (with variables as above and $\zeta_i \in (\tilde{x}_i, \frac{i}{n})$):

$$\begin{aligned}
& \int_0^u K \left(\frac{x-u}{h_n} \right) s^2(x) dx = \sum_{i=1}^{\lfloor un \rfloor} \int_{\frac{i-1}{n}}^{\frac{i}{n}} K \left(\frac{x-u}{h_n} \right) s^2(x) dx + \int_{\frac{\lfloor un \rfloor}{n}}^u K \left(\frac{x-u}{h_n} \right) s^2(x) dx \\
& = \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K \left(\frac{\tilde{x}_i-u}{h_n} \right) s^2(\tilde{x}_i) + \frac{Res_n}{n} \\
& \frac{1}{nh_n} \left| \sum_{i=1}^{\lfloor un \rfloor} K \left(\frac{i}{n} - u \right) s^2 \left(\frac{i}{n} \right) - K \left(\frac{\tilde{x}_i-u}{h_n} \right) s^2(\tilde{x}_i) - Res_n \right| \\
& \leq \frac{1}{nh_n} \left| \sum_{i=1}^{\lfloor un \rfloor} s^2 \left(\frac{i}{n} \right) \left(K \left(\frac{i}{n} - u \right) - K \left(\frac{\tilde{x}_i-u}{h_n} \right) \right) + K \left(\frac{\tilde{x}_i-u}{h_n} \right) \left(s^2 \left(\frac{i}{n} \right) - s^2(\tilde{x}_i) \right) \right| + \frac{Res_n}{nh_n} \\
& = \frac{1}{nh_n} \left| \sum_{i=1}^{\lfloor un \rfloor} s^2 \left(\frac{i}{n} \right) \underbrace{K' \left(\frac{\xi_i-u}{h_n} \right)}_{abs \leq M} \cdot \underbrace{\left(\frac{i-\tilde{x}_i}{h_n} \right)}_{\leq \frac{1}{nh_n}} + K \left(\frac{\tilde{x}_i-u}{h_n} \right) \underbrace{(s^2(\zeta_i))'}_{abs \leq N} \cdot \underbrace{\left(\frac{i-\tilde{x}_i}{n} \right)}_{\leq \frac{1}{n}} \right| + \frac{Res_n}{nh_n} \\
& \leq \frac{1}{nh_n} \left(\sum_{i=1}^{\lfloor un \rfloor} s^2 \left(\frac{i}{n} \right) \frac{M}{nh_n} + K \left(\frac{\tilde{x}_i-u}{h_n} \right) \frac{N}{n} \right) + \frac{Res_n}{nh_n} = O \left(\frac{1}{nh_n^2} \right) + O \left(\frac{1}{nh_n} \right) \rightarrow 0
\end{aligned}$$

Putting both limit considerations together, we get by using the triangle inequality

$$\left| \frac{1}{nh_n} \sum_{i=1}^{\lfloor un \rfloor} K \left(\frac{i-u}{h_n} \right) s^2 \left(\frac{i}{n} \right) - \frac{s^2(u)}{2} \right| \leq \left| \frac{1}{nh_n} \sum_{i=1}^{\lfloor un \rfloor} K \left(\frac{i-u}{h_n} \right) s^2 \left(\frac{i}{n} \right) - \frac{1}{h_n} \int_0^u K \left(\frac{x-u}{h_n} \right) s^2(x) dx \right| + \left| \frac{1}{h_n} \int_0^u K \left(\frac{x-u}{h_n} \right) s^2(x) dx - \frac{s^2(u)}{2} \right| = O \left(\frac{1}{nh_n^2} \right) + O(h_n) \xrightarrow{(C3')} 0$$

And it immediately follows that $\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i-u}{h_n} \right) s^2 \left(\frac{i}{n} \right) \rightarrow \frac{s^2(u)}{2}$.

1.b.ii.) We now pay attention to the second component of the enumerator:

$$\begin{aligned} 0 \leq \frac{1}{nh_n} \sum_{i=2}^{\lfloor un \rfloor} K \left(\frac{i-u}{h_n} \right) \frac{1}{(i-1)^2} \underbrace{\sum_{j=1}^{i-1} s^2 \left(\frac{j}{n} \right)}_{\leq (i-1) \max_j s^2(\cdot)} &\leq \frac{1}{nh_n} \underbrace{\max_{j=1, \dots, \lfloor un \rfloor - 1} s^2 \left(\frac{j}{n} \right)}_{=: A_n < \infty} \max_{i=2, \dots, \lfloor un \rfloor} K \left(\frac{i-u}{h_n} \right) \sum_{i=2}^{\lfloor un \rfloor} \frac{1}{i-1} \\ &\leq \frac{A_n}{nh_n} \sum_{i=1}^n \frac{1}{i} \leq \frac{A_n}{nh_n} \left(1 + \int_1^n \frac{1}{x} dx \right) = \frac{A_n}{nh_n} + \frac{A_n \ln(n)}{nh_n} \rightarrow 0 \end{aligned}$$

It immediately follows with the convergence in numerator and denominator:³⁵

$$E \hat{s}_{(1)h_n}^2(u) \xrightarrow{n \rightarrow \infty} \frac{s^2(u)}{2} \cdot 2 = s^2(u) \quad \forall u \in (0, 1].$$

2.) We execute an analogous analysis to the variance of $\hat{s}_{(1)h_n}^2(u)$ proving that its limiting value is zero. For a sufficiently large n ,

$$Var \hat{s}_{(1)h_n}^2(u) = \frac{Var \left(\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i-u}{h_n} \right) \tilde{R}_{i,n}^2 \right)}{\frac{1}{n^2} \left(\sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i-u}{h_n} \right) \right)^2}.$$

2.a.) The denominator is the squared expression of 1.a.), i.e. its limit value is $\frac{1}{4}$.

2.b.) Regarding the numerator, we first fragment the sum in brackets similar to the examination of the expected return of $\tilde{R}_{i,n}^2$ in part 1.b.) of this proof, and then again decompose it in several sums of variances and covariances.

2.b.i.) We concentrate on the central component (with lowest convergence rate):

$$\frac{1}{n^2} Var \left\{ \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i-u}{h_n} \right) s^2 \left(\frac{i}{n} \right) \varepsilon_{i,n}^2 \right\} \stackrel{iid}{=} \frac{1}{n^2} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n}^2 \left(\frac{i-u}{h_n} \right) s^4 \left(\frac{i}{n} \right) V,$$

where $V := Var \varepsilon_{1,n}^2 = E \varepsilon_{1,n}^4 - 1 \stackrel{(C5)}{<} \infty$. With the same technique as above, we approximate initially:

$$\begin{aligned} &\frac{V}{nh_n^2} \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K^2 \left(\frac{i-u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \approx \frac{V}{nh_n^2} \int_0^u K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx = \frac{V}{nh_n^2} \int_{-1}^0 K^2(v) s^4(u + vh_n) h_n dv \\ &= \frac{V s^4(u)}{nh_n} \int_{-1}^0 K^2(v) dv + \frac{V (s^4(u))'}{n} \int_{-1}^0 v K^2(v) dv + \frac{V h_n}{2n} \int_{-1}^0 (s^4(\hat{u}))'' v^2 K^2(v) dv \\ &= O \left(\frac{1}{nh_n} \right) + O \left(\frac{1}{n} \right) + O \left(\frac{h_n}{n} \right) \rightarrow 0 \end{aligned}$$

³⁵ Altogether, much more effort was to spent due to by Gaussian brackets truncated sums. Moreover the power of zero-sequences was assessed in a way, that holds in the later analysis on asymptotic normality despite the stronger condition (C3') of the bandwidth dimension.

The absolute approximation error is again a zero sequence, following a similar argumentation to 1.b.i) (this proof) and 2.b.i.) in proof 2.1, respectively:

$$\frac{V}{nh_n^2} \left| \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) - \int_0^u K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx \right| \leq O \left(\frac{1}{n^2 h_n^3} \right) \rightarrow 0$$

We conclude with the triangle inequality at:

$$0 \leq \frac{V}{n^2 h_n^2} \sum_{i=1}^{\lfloor un \rfloor} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \leq \left| \frac{V}{n^2 h_n^2} \sum_{i=1}^{\lfloor un \rfloor} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) - \frac{V}{nh_n^2} \int_0^u K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx \right| + \frac{V}{nh_n^2} \int_0^u K^2 \left(\frac{x-u}{h_n} \right) s^4(x) dx = O \left(\frac{1}{n^2 h_n^3} \right) + O \left(\frac{1}{nh_n} \right) \rightarrow 0$$

And it results that $\frac{1}{n^2} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n}^2 \left(\frac{i}{n} - u \right) s^4 \left(\frac{i}{n} \right) V \rightarrow 0$.

2.b.ii.) All other components can be absolutely restricted to zero (with a faster rate of convergence, compare part 2.b.ii.) in proof 2.1).

Herewith it immediately follows that $Var \hat{s}_{(1)h_n}^2(u) \xrightarrow{n \rightarrow \infty} 0 \forall u \in (0, 1]$.

3. The stochastic convergence is finally derived with the Markov inequality. With $\{P_{n,s^2} : s^2 \in \mathbb{R}_0^+\}_{n \in \mathbb{N}}$ as the sequence of statistical models corresponding to the estimators sequence $(\hat{s}_{(1)h_n}^2)_{n \in \mathbb{N}}$ we conclude:

$$P_{n,s^2} \left\{ \left| \hat{s}_{(1)h_n}^2(u) - s^2(u) \right| \geq \epsilon \right\} \leq \frac{E \left| \hat{s}_{(1)h_n}^2(u) - s^2(u) \right|^2}{\epsilon^2} = \frac{Var \left(\hat{s}_{(1)h_n}^2(u) \right)}{\epsilon^2} + \frac{\left(E \left(\hat{s}_{(1)h_n}^2(u) - s^2(u) \right) \right)^2}{\epsilon^2} \xrightarrow{n \rightarrow \infty} 0 \quad \forall \epsilon > 0$$

I.e. $\hat{s}_{(1)h_n}^2(u) \xrightarrow{P} s^2(u)$ for all $s^2 \in \mathbb{R}_0^+, u \in (0, 1]$. □

Proof of proposition 2.4:

We decompose $Z_{n,h_n}^{(1)}(u) = \sqrt{nh_n} \left(\hat{s}_{(1)h_n}^2(u) - s^2(u) \right)$ into a centered rv (variance-part) and a deterministic bias-part. For sufficiently large n it can be expressed as:

$$Z_{n,h_n}^{(1)}(u) = \left. \sqrt{nh_n} \frac{\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \left(\tilde{R}_{i,n}^2 - E \tilde{R}_{i,n}^2 \right)}{\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right)} \right\} \text{variance-part} + \left. \sqrt{nh_n} \frac{\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \left(E \tilde{R}_{i,n}^2 - s^2(u) \right)}{\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right)} \right\} \text{bias-part}$$

1.) We concentrate first on both moments.

1.a.) The expectation of $Z_{n,h_n}^{(1)}(u)$ can be written as used in proof of proposition 2.3:

$$\frac{\frac{\sqrt{nh_n}}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \left(s^2 \left(\frac{i}{n} \right) - s^2(u) \right) + \sum_{i=2}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \frac{1}{(i-1)^2} \sum_{j=1}^{i-1} s^2 \left(\frac{j}{n} \right)}{\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right)}$$

The denominator converges to $\frac{1}{2}$ as seen before. For the numerator we exploit the results of proof 2.3, part 1.b.) as well. Concerning the first sum we arrive at:

$$\frac{\sqrt{nh_n}}{h_n} \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K \left(\frac{\frac{i}{n} - u}{h_n} \right) \left(s^2 \left(\frac{i}{n} \right) - s^2(u) \right) \rightarrow D \left(s^2(u) \right)' \int_{-1}^0 v K(v) dv,$$

where $D \geq 0$ and $nh_n^3 \rightarrow D^2$ as postulated in (C3'). The second sum can still be restricted absolutely to zero as n goes to infinity, the factorisation with $\sqrt{nh_n}$ slows up the convergence rate. This implies for the expectation of $Z_{n,h_n}^{(1)}(u) \forall u \in (0, 1]$ a pointwise limiting value of $\beta_{(1)}(u)$:

$$EZ_{n,h_n}^{(1)}(u) \xrightarrow{n \rightarrow \infty} 2D(s^2(u))' \int_{-1}^0 vK(v)dv$$

1.b.) We address ourselves now to the variance-part, enlarging the asymptotic results for the variance of $\hat{s}_{(1)h_n}^2(u)$ from proof 2.3:

$$\text{Var}Z_{n,h_n}^{(1)}(u) = \frac{nh_n \text{Var} \left(\frac{1}{n} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \tilde{R}_{i,n}^2 \right)}{\frac{1}{n^2} \left(\sum_{i=1}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \right)^2}$$

The denominator converges to $\frac{1}{4}$. From the aforementioned decomposition (proof 2.3, part 2.b.i)) it follows that the predominant part of the numerator and its limit value are (with $V := E\varepsilon_{1,n}^4 - 1$):

$$nh_n \frac{1}{n^2} \sum_{i=1}^{\lfloor un \rfloor} K_{h_n}^2 \left(\frac{i}{n} - u \right) s^4 \left(\frac{i}{n} \right) V = \frac{V}{h_n} \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \rightarrow Vs^4(u) \int_{-1}^0 K^2(v)dv$$

The other components hold the zero convergence (due to a faster rate in proof 2.3). This implies for the whole variance of $Z_{n,h_n}^{(1)}(u) \forall u \in (0, 1]$ a pointwise limiting value of $\tau_{(1)}^2(u)$:

$$\text{Var}Z_{n,h_n}^{(1)}(u) \xrightarrow{n \rightarrow \infty} 4Vs^4(u) \int_{-1}^0 K^2(v)dv$$

2.) To prove the convergence in distribution, we have to focus mainly on the stochastic numerator of the variance-part, which can be decomposed as:

$$\begin{aligned} & \underbrace{\frac{1}{\sqrt{nh_n}} \sum_{i=1}^{\lfloor un \rfloor} K \left(\frac{\frac{i}{n} - u}{h_n} \right) \left(s^2 \left(\frac{i}{n} \right) (\varepsilon_{i,n}^2 - 1) \right)}_{=:m_n} \\ & + \underbrace{\sqrt{nh_n} \frac{1}{n} \sum_{i=2}^{\lfloor un \rfloor} K_{h_n} \left(\frac{i}{n} - u \right) \left\{ \frac{-2}{i-1} \sum_{j=1}^{i-1} s \left(\frac{i}{n} \right) s \left(\frac{j}{n} \right) \varepsilon_{i,n} \varepsilon_{j,n} \right.}_{=:p_n} \\ & \quad \left. + \frac{1}{(i-1)^2} \left(\sum_{j=1}^{i-1} s \left(\frac{j}{n} \right) \varepsilon_{j,n} \right)^2 - \frac{1}{(i-1)^2} \sum_{j=1}^{i-1} s^2 \left(\frac{j}{n} \right) \right\}} \end{aligned}$$

2.a.) We apply again the central limit theorem of Lindeberg-Feller to show the distributional convergence of m_n . The triangle-schema is

$$Y_{n',i} = \frac{1}{\sqrt{nh_n}} K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) (\varepsilon_{i,n}^2 - 1),$$

where $Y_{n',1}, \dots, Y_{n',n'}$ are independent rvs with $n' := \lfloor un \rfloor$. Its moments are:

$$\mu_{n',i} = \frac{1}{\sqrt{nh_n}} K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) E(\varepsilon_{i,n}^2 - 1) = 0, \quad \sigma_{n',i}^2 = \frac{V}{nh_n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \in (0, \infty) \quad \text{and}$$

$$s_{n'}^2 := \sum_{i=1}^{n'} \sigma_{n',i}^2 = \frac{V}{h_n} \sum_{i=1}^{\lfloor un \rfloor} \frac{1}{n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \rightarrow Vs^4(u) \int_{-1}^0 K^2(v)dv.$$

For the verification of the Lindeberg condition we apply amongst others the Hölder inequality and the Markov inequality to get the following upper bound:

$$L_{n'}(\epsilon) \leq \frac{\sum_{i=1}^{\lfloor un \rfloor} \left(\frac{1}{nh_n} K^2 \left(\frac{\frac{i}{n} - u}{h_n} \right) s^4 \left(\frac{i}{n} \right) \right)^{1 + \frac{\delta'}{1 + \delta'}} W}{\epsilon^{\frac{2\delta'}{1 + \delta'}} (s_{n'}^2)^{1 + \frac{\delta'}{1 + \delta'}}$$

with an upper bound $W \in (0, \infty)$ and $\delta' := \frac{\delta}{4}$ with δ from condition (C5). The finite, non-zero convergence of the denominator follows from the limit value of $s_{n'}^2$. For the convergence analysis of the numerator similar methods as seen before can be applied, leading to a limiting value of 0. Hence the Lindeberg condition holds. With a Slutsky argument it follows that

$$\begin{aligned} \sum_{i=1}^{n'} Y_{n',i} &\xrightarrow{D} \mathcal{N} \left(0, \lim_{n \rightarrow \infty} s_{n'}^2 \right) \\ \Leftrightarrow \frac{1}{\sqrt{nh_n}} \sum_{i=1}^{\lfloor un \rfloor} K \left(\frac{\frac{i}{n} - u}{h_n} \right) s^2 \left(\frac{i}{n} \right) (\varepsilon_{i,n}^2 - 1) &\xrightarrow{D} \mathcal{N} \left(0, V s^4(u) \int_{-1}^0 K^2(v) dv \right) \end{aligned}$$

2.b.) It can be proved that $E p_n = 0$ and $Var p_n \rightarrow 0$, which implies (e.g. with Markov inequality) that $p_n \xrightarrow{P} 0$. We combine both convergence results and include the limit $\frac{1}{4}$ of the denominator with the Slutsky theorem. We gain the following convergence in distribution:

$$'variance-part' \xrightarrow{D} \mathcal{N} \left(0, 4 V s^4(u) \int_{-1}^0 K^2(v) dv \right)$$

2.c.) Finally, we add the limit value of the deterministic bias-part and the asymptotic distribution of $Z_{n,h_n}^{(1)}(u)$ results. \square

References

- BOLLERSLEV, TIM, 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31:307-327.
- BOLLERSLEV, TIM, ENGLE, ROBERT F. AND NELSON, DANIEL B., 1994. ARCH models. In: ENGLE, ROBERT F. AND MCFADDEN, DANIEL L. (Ed.) *Handbook of Econometrics*. Elsevier B.V., Amsterdam.
- BOUGEROL, PHILIPPE AND PICARD, NICO, 1992. Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* 52:115-127.
- CHOI, PILSUN AND NAM, KISEOK, 2008. Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the S_U -normal distribution. *Journal of Empirical Finance* 15:41-63.
- DE HAAN, LAURENS, JANSEN, DENNIS W., KOEDIJK, KEES G. AND DE VRIES, CASPER G., 1994. Safety first portfolio selection, extreme value theory and long run asset risks. In: GALAMBOS, JANOS, LECHNER, JAMES AND SIMIU, EMIL (Ed.) *Extreme value theory and its applications*. Kluwer Academic Publishers, Dordrecht.
- DIEBOLD, FRANCIS X. AND INOUE, ATSUSHI, 2001. Long memory and regime switching. *Journal of Econometrics* 105:131-159.
- DING, ZHUANXIN, ENGLE, ROBERT F. AND GRANGER, CLIVE W.J., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1:83-106.
- DREES, HOLGER AND STARICA, CATALIN, 2002. A simple non-stationary model for stock returns (Preprint). Available at (27.09.2004) <http://www.math.uni-sb.de/PREPRINTS/preprint69.pdf>.

- ENGLE, ROBERT F., 1982. Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987-1007.
- FAMA, EUGENE, 1965. The behavior of stock market prices. *Journal of Business* 38:34-105.
- FAN, JIANQING AND YAO, QIWEI, 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics. Springer-Verlag, New York.
- GASSER, THEO, KNEIP, ALOIS UND KÖHLER, WALTER, 1991. A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association* 86:643-652.
- GRANGER, CLIVE W. J. AND STARICA, CATALIN, 2005. Non-stationarities in stock returns. *Review of Economics and Statistics* 87:503-522.
- GÜRTLER, MARC AND RAUH, RONALD, 2009. Shortcomings of a Parametric VaR Approach and Nonparametric Improvements Based on a Non-stationary Return Series Model. IF Working Paper Series: IF32, TU Braunschweig.
- HAMILTON, JAMES D. AND SUSMEL, RAUL, 1994. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64:307-333.
- HERZEL, STEFANO, STARICA, CATALIN AND TÜTÜNCÜ, REHA, 2005. A non-stationary paradigm for the dynamics of multivariate financial returns. In: BERTAIL, PATRICE AND DOUKHAN, PAUL (Ed.). *Dependence in Probability and Statistics (Lecture Notes in Statistics, Vol. 187)*. Springer, New York 2006.
- JOHNSON, NORMAN L. AND KOTZ, SAMUEL, 1970. *Continuous univariate distributions - 1. Distributions in statistics (Vol. 2)*. John Wiley & Sons, New York.
- JORION, PHILIPPE, 2006. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill Professional, 3rd edition, New York.
- KITAGAWA, GENSHIRO AND NAGAHARA, YUICHI, 1999. A Non-Gaussian Stochastic Volatility Model. *The Journal of Computational Finance* 2:33-47.
- KUPIEC, PAUL, 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3:73-84.
- LAMOUREUX, CHRISTOPHER G. AND LASTRAPES, WILLIAM D., 1990. Persistence in variance, structural changes and the GARCH model. *Journal of Business and Economic Statistics* 8:225-234.
- MANDELBROT, BENOIT, 1963. The variation of certain speculative prices. *Journal of Business* 36:394-419.
- MIKOSCH, THOMAS AND STARICA, CATALIN, 2004. Non-stationarities in financial time series, the long range dependence and IGARCH effects. *Review of Economics and Statistics* 86:378-390.
- MIKOSCH, THOMAS AND STARICA, CATALIN, 2004a. Stock market risk-return inference. An unconditional non-parametric approach (Preprint). Available at (05.06.2007) <http://www.math.ku.dk/~mikosch/Preprint/Catalin/jf020904.pdf>.
- MÜLLER, HANS-GEORG AND STADTMÜLLER, ULRICH, 1987. Estimation of heteroscedasticity in regression analysis. *Annals of Statistics* 15:610-625.
- NELSON, DANIEL B., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59:347-370.
- STARICA, CATALIN, 2003. Is GARCH(1, 1) as good a model as the Nobel prize accolades would imply? (Preprint). Available at (07.03.2005) <http://www.math.chalmers.se/~starica/papersoft.pdf>.
- STRAUMANN, DANIEL, 2004. *Estimation in Conditionally Heteroscedastic Time Series Models*. Springer, Berlin.
- TAYLOR, STEPHEN J., 1986. *Modelling financial time series*. John Wiley & Sons, New York.