Gürtler, Marc; Rauh, Ronald

Working Paper

Challenging traditional risk models by a non-stationary approach with nonparametric heteroscedasticity

Working Paper Series, No. IF41V1

Provided in Cooperation with:
Technische Universität Braunschweig, Institute of Finance

Challenging Traditional Risk Models
by a Non-stationary Approach
with Nonparametric Heteroscedasticity

Marc Gürtler* and Ronald Rauh†

Braunschweig, September 2, 2012

Abstract

In this paper we analyze an econometric model for non-stationary asset returns. Volatility dynamics are modelled by nonparametric regression; consistency and asymptotic normality of a symmetric and of a one-sided kernel estimator are outlined with remarks on the bandwidth decision. Further attention is paid to asymmetry and heavy tails of the return distribution, involved by the framework for innovations. We survey the practicability and automatization of the implementation. For simulated price processes and a multitude of financial time series we observe a satisfying model approximation and good short-term forecasting abilities of the univariate approach. The non-stationary regression model outperforms parametric risk models and famous ARCH-type implementations.

JEL classification: C14, C5
Keywords: heteroscedastic asset returns, non-stationarity, nonparametric regression, volatility, innovation modelling, forecasting, Value at Risk (VaR), ARCH-models

*University of Braunschweig - Institute of Technology, Department of Finance, Abt-Jerusalem-Str. 7, 38106 Braunschweig, Germany. Tel.: +49-531-391-2895, fax: +49-531-391-2899, e-mail: marc.guertler@tu-bs.de.

†Öffentliche Versicherung Braunschweig, Department of Asset Risk Management, Theodor-Heuss-Str. 10, 38122 Braunschweig, Germany. Tel.: +49-531-202-1568, fax: +49-531-202-331568, e-mail: ronald.rauh@oeffentliche.de

1 Introduction

The dynamics of financial time series have been extensively studied by scientifics in the past years. State of the art became *autoregressive conditional heteroscedastic* models based on the seminal papers of Engle (1982) and Bollerslev (1986), although their stationary implementations are critically reviewed. On the other hand financial risk management has become an important task for banks and insurance companies, to some extent encouraged by regulatory capital requirements. In this context a certain quantile of the return distribution, the *Value at Risk* (VaR), became a benchmark risk measure, compare Jorion (2006).[1] A widespread approach is the *delta-normal-model* (also named *parametric VaR*), whose risk factors are theoretically still founded in the *random walk hypothesis*. Against this background it is obvious, that the distance between practically manageable risk models with meaningful key numbers and the academic possibilities for a return distribution forecast with nontrivial price dynamics is considerable.

In this paper we analyze an econometric model for non-stationary asset returns, that was first introduced by Drees and Starica (2002) and Herzel et al. (2005). Volatility dynamics are modelled both exogenously and deterministic, captured by nonparametric regression. We outline results of Gürtler et al. (2009) on consistency and asymptotic normality of a symmetric and of a one-sided kernel estimator of unconditional volatility and give remarks on the bandwidth decision. Further attention is paid to asymmetry and heavy tails of the return distribution, implemented by the framework for innovations. We fit estimated innovations with an asymmetric version of the Pearson type VII distribution and provide a method of moments for parameter estimation. By dint of a Student-t connection to the Pearson VII distribution we develop a factor-based VaR implementation of the non-stationary model, in order to make it practically amenable. Hence, next to providing a solid statistical background we survey the practicability and potentials of automatization of the implementation for real financial time series. Simulations on predefined price and volatility processes as well as an empirical study on the dynamics of equity indices, exchange rates, interest rates and credit spreads confirm in general a good approximation by the non-stationary paradigm. Moreover, an outperformance to the delta-normal-model and even to famous implementations of ARCH-type models (t-GARCH$(1,1)$, t-EGARCH$(1,1)$) is demonstrated.

The rest of the paper is organized as follows: We start with a brief review of financial time series analysis in section 2 to motivate our research. The multivariate non-stationary regression model is introduced in section 3, including a statistical discussion of nonparametric volatility estimates on the univariate frame and an examination of Pearson type VII distributed innovations. In section 4 we test whether the non-stationary approach offers an adequate adaption to hypothetical price processes, and in section 5 the model is calibrated for a multitude of individual financial time series including performance evaluations relative to standard parametric models and traditional ARCH-models. We conclude in section 6.

2 Shortfalls of common financial time series models

In this section we provide a brief, but incomplete overview of the history of financial time series analysis in order to reinvestigate opportunities and shortfalls of common models and to motivate our modelling approach at the end. The starting point of diffusion processes goes back to the random walk model for security and commodity markets of Bachelier (1900). The successive differences $P(t+\Delta) - P(t)$ of market prices $\{P_t\}_{t=0,\dots,n}$ are assumed to be independent, normally distributed random variables (rvs) with zero mean and variance proportional to the differencing interval Δ. Later on, Samuelson proposed modelling prices in continuous time by a geometric Brownian motion, whose discretization leads to a random walk for discrete log-prices $\ln P_t = \ln P_{t-1} + X_t$ with independent identically distributed (iid) normal increments X_t; Samuelson (1973) summarizes. Due to the great features of the normal distribution (e.g. invariance property) the hypothesis was preserved in a plenty of practical applications up to now: The Markowitz (1952) portfolio theory postulates multivariate Gaussian returns. Black and Scholes (1973) or Merton (1973) built up their option pricing theory on the geometric Brownian motion of prices. Last but not least the parametric VaR, one competitive model in section 5, exploits the mathematical tractability of iid normal returns.

In the 1960s the random walk hypothesis was first rejected statistically. Following Mandelbrot (1963) empirical distributions of price changes are usually too peaked and heavy-tailed relative to a normal distribution, which he replaced by the stable Paretian family. Fama (1965) proved this approach empirically for daily log-returns of the

[1] The VaR is an estimate for the maximum loss at a predefined confidence level $(1-\alpha)$ for holding a position over a target horizon. Generally, $\mathrm{VaR}^{1-\alpha} = u_\alpha = \inf\{u : F(u) \geq \alpha\}$, for a given positive value α close to zero corresponding to the shortfall probability.

30 DJ stocks from 1957 to 1962, where Gaussian features could be clearly rejected, but no strong support against the iid assumption was found. Lo and MacKinlay (1988) rejected the random walk hypothesis for several aggregate stock indices and size-sorted portfolios (weekly returns from 1962 to 1985 and subperiods) with a specification test for the linearity of variance estimates on different sampling frequencies. In a version of their variance-ratio-test that is consistent to non-normal features and time-varying volatility the rejections hold, implying the independence assumption to be neglected. The same procedure was executed by Liu and He (1991) for five pairs of weekly nominal exchange rates with evidence rejecting the random walk hypothesis.[2] Moreover, a Box-Pierce test rejected iid currency returns, but this was justified mainly by heteroscedasticity. Hsieh (1989) observed for log-returns of daily exchange rates little linear dependence, but the Brock-Dechert-Scheinkman test on independence and autocorrelations of the squared data detected strong nonlinear dependence.

A graphical access to linear dependence of a (stationary) return series $\{X_t\}_{t=1,\ldots,n}$ is often derived from correlograms, whether there is any sample autocorrelation $\rho(h) = \gamma(h)/\gamma(0)$ (SACF), with $\gamma(h)$ the empirical estimate for $\mathrm{Cov}\,(X_t, X_{t+h})$ and discrete time lags $h \geq 0$. At the typical picture of liquid financial instruments, daily returns contain themselves little serial correlation, but their absolutes $|X_t|$ are significantly positive correlated over a large number of lags with a slowly declining SACF. Taylor (1986) established this stylized fact for stock log-returns and brought it into agreement with the efficient market theory.[3] This effect is also named *long range dependence* (LRD), as one might draw the conclusion of a serial dependence of the time series.

At our preliminary work we observed this feature for the majority of the 30 benchmarks, that are analysed later on. Descriptions of those market indices are provided in table 1.[4] All data series are daily closing prices from January 4, 1999 to December 29, 2006, where log-returns $X_t = \ln P_t - \ln P_{t-1}$ are used for equity and currency series and differences of prices $X_t = P_t - P_{t-1}$ (short: diff-returns) for interest rates and credit spreads. Amongst others, the equity index MSCI North America (*EquNA*), the 5-year US swap rate (*RateUSD*), a global government to swap spread (*CreditSta*) and the exchange rate EUR/USD (*CurrUSD*) generate SACFs of original returns with only a few outliers from a 95% confidence band. But a systematic and significant autocorrelation appears for absolute returns, that is strictly positive and declines slowly with the number of lags. Furthermore, if a stationary return sample is serially uncorrelated but its absolutes (or power transformations) are not, then a normal distribution cannot hold. Gürtler and Rauh (2009) provide the corresponding correlograms and test additionally all aspects of the iid Gaussian return assumption in a comprehensive survey regarding the 30 benchmark series.[5] On the 8-year monthly data base the hypothesis can be barely sustained, albeit for 7 indices the normal assumption was denied and serial identity was rejected 19 times; testing for independence brought only 3 rejections. The result worsens for daily samples: The assumption of normality was rejected in all cases (especially Jarque-Bera statistics indicate heavy-tailed, asymmetric return distributions) and the serial identity was denied 28 times (particularly F-test statistics call for heteroscedasticity). Independence tests performed with 22 rejections still best. Altogether the random walk hypothesis for daily return dynamics of all exposure types has to be denied entirely.[6]

Nowadays the following general statistical features of empirical return series, often labelled as *stylized facts*, seem to be well established, compare e.g. Straumann (2004):

- 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).

[2] This result is contrary to the consensus in the 1980s that nominal exchange rates follow a random walk process.

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

[4] Our general data source is Bloomberg. A special thanks includes also the index data providers MSCI Barra, JP Morgan and Merrill Lynch.

[5] The normal hypothesis was checked using the Jarque-Bera test, the χ^2-goodness of fit test, the Kolmogorov-Smirnov- and the Anderson-Darling test, and visually supplemented by quantile-quantile-plots. Concerning serial identity of the distribution, subsamples were tested pairwise with the two-sample Kolmogorov-Smirnov- and the χ^2-test, supplemented by the t-test and F-test for identical expectations and variances. Next to SACF analyses of original and absolute returns, serial independence was tested via the Ljung-Box- and the Brock-Dechert-Scheinkman statistic. All hypotheses were tested on a 5% level of significance.

[6] Among the exposure classes credit spreads caused most approximation problems, but also equities, interest rates and currencies follow with a bad fit to the assumptions. A by-product of the survey was the introduced, best applicable return conception with log-returns for equity and currency series and diff-returns for interest rates and credit spreads.

exposure class	amount	benchmark (BM), description
EquEUR	w_1	MSCI Daily TR Gross EMU Local
EquexEUR	w_2	MSCI Daily TR Gross Europe ex EMU Local
EquNA	w_3	MSCI Daily TR Gross North America Local
EquAsP	w_4	MSCI Daily TR Gross (Asia) Pacific Local
EquEM	w_5	MSCI Emerging Markets (Free) Local
RateEUR	w_6	EUR Swap annual (30/360), 5 year
RateUSD	w_7	USD Swap annual (30/360), 5 year
RateJPY	w_8	JPY Swap annual (act/365), 5 year
CredSta	w_9	synthetic BM, models global spread Govt. to Swap
CredSwa	w_{10}	synthetic BM, models Pfandbrief/ Covereds Spread
CredAAA	w_{11}	JP Morgan Credit Index AAA Asset Swap Spread
CredAA	w_{12}	JP Morgan Credit Index AA Asset Swap Spread
CredA	w_{13}	JP Morgan Credit Index A Asset Swap Spread
CredBBB	w_{14}	JP Morgan Credit Index BBB Asset Swap Spread
CredEM	w_{15}	JP Morgan EMBI Global Divers. Sov. Spread
CredHY	w_{16}	Merrill Lynch HY US BB-B (Spread to US-Swap)
CurrEUR	w_{17}	synthetic BM, set constant 1 (home currency)
CurrGBP	w_{18}	ECB Euro Exchange Ref. Rate as EUR/GBP
CurrCHF	w_{19}	ECB Euro Exchange Ref. Rate as EUR/CHF
CurrSEK	w_{20}	ECB Euro Exchange Ref. Rate as EUR/SEK
CurrDKK	w_{21}	ECB Euro Exchange Ref. Rate as EUR/DKK
CurrNOK	w_{22}	ECB Euro Exchange Ref. Rate as EUR/NOK
CurrUSD	w_{23}	ECB Euro Exchange Ref. Rate as EUR/USD
CurrCAD	w_{24}	ECB Euro Exchange Ref. Rate as EUR/CAD
CurrJPY	w_{25}	ECB Euro Exchange Ref. Rate as EUR/JPY
CurrAUD	w_{26}	ECB Euro Exchange Ref. Rate as EUR/AUD
CurrNZD	w_{27}	ECB Euro Exchange Ref. Rate as EUR/NZD
CurrSGD	w_{28}	Bloomberg exchange rate as EUR/SGD
CurrHKD	w_{29}	Bloomberg exchange rate as EUR/HKD
CurrEM	w_{30}	synthetic BM for exchange rate of EM currencies

Table 1: Exposure classes and corresponding benchmarks.
The exposure conception is exemplified for a European (EUR) investor, the contribution of each asset to the several exposure classes has to be examined. Four general types of exposure classes are distinguished:
• Equity exposures, with subclasses for different economic areas. Exposure amounts $w_1, \ldots, w_5$ are measured as the effective (market valued) investments.
• Interest rate exposures, with subclasses for different currency areas. Exposure amounts $w_6, \ldots, w_8$ are measured as the basis point value (bpv) of attributed securities, with w_j being negative for long-positions.
• Credit spread exposures, with subclasses for different rating classes or types of coverage. Exposure amounts $w_9, \ldots, w_{16}$ are measured again as the bpv, with w_j being negative for long-positions.
• Currency exposures, with subclasses for different denominations. Exposure amounts $w_{17}, \ldots, w_{30}$ are measured as the effective investments in home currency.
Liquid market indices are matched to the exposure classes as risk factors. Equity benchmarks are MSCI equity indices (total return) in local currency. Interest rate benchmarks are 5 year swap rates (in bp). Credit spread benchmarks are asset swap spreads (in bp) from JP Morgan, Merrill Lynch or synthetic created (BM *CreditSta*: 75% Euro $+20\%$ US $+5\%$ Japan govt. to swap spread (5y.); BM *CredSwa*: 5y. PEX yield to swap rate). Currency benchmarks are mostly ECB exchange rates in price quotation (BM *CurrEUR*: index constant 1 (to cover non- risky EUR investments), BM *CurrEM*: derived from return differences of MSCI EM index (in EUR) and MSCI EM index (local)). Thanks to Bloomberg and the cited index data providers.

Cont (2001) presents an extended set of stylized facts and distributional properties, affirming inter alia (nonlinear) dependence, a volatility clustering and (conditional) heavy tails with a gain-loss asymmetry.[7]

Against the background 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

$$X_t = \mu_t + \varsigma_t \varepsilon_t, \;\; t \in \mathbb{Z}, \tag{1}$$

where $\{\varepsilon_t\}_t$ is an iid sequence with $\mathrm{E}\varepsilon_1 = 0$ and $\mathrm{Var}\varepsilon_1 = 1$, $\{\mu_t\}_t$ and $\{\varsigma_t\}_t$ are stochastic processes dependent on past information.[8] Engle (1982) introduced the ARCH(p)-process, where conditional volatility dynamics $\{\varsigma_t\}_t$ are constructed as linear regression over past squared (centred) returns. The extension of Bollerslev (1986), named generalized ARCH (GARCH(p,q)-process), includes past variances next to historical returns into the parametric regressed volatility. This enables often a parsimonious parametrization with a reasonable fit to empirical data. For the purpose of parameter estimation, processes of the ARCH-family are defined to be stationary, but finding conditions for the existence and uniqueness of a stationary solution is nontrivial.[9] 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. The exponential GARCH (EGARCH) model of Nelson (1991) and the asymmetric power GARCH (AGARCH) model by Ding et al. (1993) include asymmetry and leverage effects directly in the volatility dynamics. Furthermore, Nelson (1991) gives theoretical considerations that volatility tends to response asymmetrically to financial gains and losses. In his nobel price lecture Engle (2004) gives an alphabet of model extensions, but the 'GARCH$(1,1)$ specification is the workhorse of financial applications' (Engle (2004), p. 408), emphasising this to be a good starting point for the analysis of multifaceted financial returns. That is why we oppose t-GARCH$(1,1)$ and t-EGARCH$(1,1)$ implementations to our model evaluation in sections 4 and 5.

Interpreting once more the LRD, Mikosch and Starica (2004) derive theoretically that the aforementioned shape of the SACF for absolute returns could alternatively arise from non-stationarities in the data. Correlograms are only significant for detecting dependence under the assumption of stationarity, otherwise structural breaks in the data as shifts in the variance might cause identical results.[10] Granger and Starica (2005) came in a long-term case study on the S&P 500 index to the conclusion, that the main reason for the sample LRD is to be seen in non-stationarities due to structural breaks of the unconditioned variance. Severe criticism on ARCH-type models follows, since they are parametrized as stationary processes (i.e. with a fixed unconditional variance) and focus on modelling the dependence structure of second moments. More inconsistencies arise on ARCH modelling over longer periods of daily returns: The typical outcome of a GARCH$(1,1)$ implementation is that the sum of estimated parameters is approximately one, leading to an IGARCH$(1,1)$ model, which is referred to as *IGARCH effect* in Starica (2003) or Mikosch and Starica (2004). But an IGARCH-model implies an infinite variance of the observed random variables, which contradicts to the results of a direct tail analysis indicating that daily returns have a finite second moment (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.[11] Although some ARCH-models were developed that allow structural breaks in the volatility while maintaining stationarity,[12] the enhancement to more and more sophisticated volatility processes in that family (see Bollerslev et al. (1994) for a statistical overview) and its stationarity assumption should be questioned at all.

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.[13] In that case, the model fit can only be improved by a

[7]Engle and Patton (2001) present stylized facts about the asset price volatility, in which they additionally focus on the persistence and mean reversion of conditional volatility.

[8]Hence X_t is measurable with respect to the σ-field $\mathcal{F}_t = \sigma(\{\varepsilon_j \mid j \leq t\})$, and $\mu_{t+1} = \mathrm{E}(X_{t+1} \mid \mathcal{F}_t)$ is the conditional mean and $\varsigma^2_{t+1} = \mathrm{Var}(X_{t+1} \mid \mathcal{F}_t)$ the conditional variance of X_{t+1} given past returns.

[9]Bougerol and Picard (1992) gave the solution for stationary GARCH(p,q) processes via stochastic recurrence equations.

[10]The authors found that the stronger the non-stationarity, e.g. the variation difference of subseries $X^{(1)}$ and $X^{(2)}$ measured as $\left(\mathrm{E}|X^{(1)}| - \mathrm{E}|X^{(2)}|\right)^2$, the more pronounced is the seemingly long memory. This theoretical result is supported with a study on the S&P 500 index by ex-/ including the 1970s US-recession, generating the LRD effect.

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

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

[13]Engle and Patton (2001) provide examples of exogenous variables to the actual return series that influence its volatility (e.g. company announcements, macroeconomic data). Financial asset prices do not evolve independently from the markets around, other variables may contain relevant information. In a GARCH-X model the impact of US T-Bill rates to the Dow Jones 30 index volatility is exemplified.

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. In the multiplicative model (1) the variance term is replaced by an unconditional variable $\sigma(t)^2$. The corresponding volatility process $\{\sigma(t)\}_t$ is modelled as a discretization of a smooth, deterministic function of time via a nonparametric kernel regression over centred, squared returns. The approach preserves the independence assumption of log-returns, but it abandons the hypothesis of stationary, identically distributed (normal) returns. A special focus is set on an accurate description of the innovations $\{\varepsilon_t\}_t$ by fitting a Pearson type VII distribution to positive and negative innovations separately to allow asymmetry and heavy-tailedness. Drees and Starica (2002) provide one example of a 12-year S&P 500 return series where 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 of financial returns are assumed to have a time-varying unconditional covariance matrix that evolves smoothly, captured by classical nonparametric regression.[14] In the following we support the non-stationary paradigm statistically, provide a direct access for parameter estimation and model implementation, and present simulations and a broad empirical study on its forecasting abilities, compared to the delta-normal-model and ARCH-type models.

3 A non-stationary model for asset returns

As motivated in the previous section, we adopt the conceptual framework of Herzel et al. (2005) and Drees and Starica (2002) for describing the dynamics of financial returns. A non-stationary sequence of d-dimensional, independent random vectors $\{\boldsymbol{X}_t\}_{t=1,\ldots,n}$ is modelled in a multiplicative approach, with a constant mean return vector $\boldsymbol{\mu}$ added:

$$\begin{aligned} &\boldsymbol{X}_t = \boldsymbol{\mu} + \mathbf{S}_t\boldsymbol{\varepsilon}_t,\ t = 1,\ldots,n, \qquad (2)\\ &\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_n \text{ iid random vectors with mutually independent coordinates,}\\ &\qquad\qquad \mathrm{E}\varepsilon_{k,1} = 0, \mathrm{Var}\varepsilon_{k,1} = 1,\ \forall k = 1,\ldots,d,\\ &\mathbf{S}_t : [0,n] \to \mathbb{R}^{d\times d} \text{ is an invertible matrix and a smooth function of time.} \end{aligned}$$

The vectors of financial returns are assumed to have a time-varying unconditional covariance matrix that evolves smoothly through time.[15] The joint volatility of returns is modelled both exogenously and deterministic, captured by classical nonparametric regression with fixed equidistant design points. The standardized residuals (also called innovations) are modelled parametrically, allowing for asymmetry and heavy tails. Hence, the series $\{\boldsymbol{X}_t\}_t$ of log-returns or diff-returns preserves the independence assumption, but a time evolution of the return distribution is incorporated via unconditional heteroscedasticity.

We emphasize 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 recent asset returns and imply the level of unconditional (co)variance. The aim is to estimate the multivariate return dynamics only by dint of recent returns and to build up short-term forecasts of return distributions in a similar economic environment. The fit of the regression model to a financial time series is done in three steps.

3.1 Centring returns

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

$$\boldsymbol{R}_t = \boldsymbol{X}_t - \bar{\boldsymbol{X}}_n,\ \ t = 1,\ldots,n, \qquad (3)$$

with column vectors centred 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 $k = 1,\ldots,d$. If the estimation error of $\bar{\boldsymbol{X}}_t$ for $\boldsymbol{\mu}$ was neglected for once (i.e. if $\bar{\boldsymbol{X}}_t = \boldsymbol{\mu} \implies \boldsymbol{R}_t = \mathbf{S}_t\boldsymbol{\varepsilon}_t$) it follows:

$$\mathrm{E}\left(\boldsymbol{R}_t \mid \boldsymbol{R}_{t-1},\boldsymbol{R}_{t-2},\ldots\right) = \mathrm{E}\boldsymbol{R}_t = \mathbf{S}_t\mathrm{E}\boldsymbol{\varepsilon}_t = \mathbf{0}, \qquad (4)$$

$$\mathrm{E}\left(\boldsymbol{R}_t\boldsymbol{R}_t^{\mathrm{T}} \mid \boldsymbol{R}_{t-1},\boldsymbol{R}_{t-2},\ldots\right) = \mathrm{E}\left(\boldsymbol{R}_t\boldsymbol{R}_t^{\mathrm{T}}\right) = \mathbf{S}_t\mathbf{S}_t^{\mathrm{T}} =: \boldsymbol{\Sigma}^2(t) \qquad (5)$$

[14] Another univariate extension with a time-varying expected return was developed by Mikosch and Starica (2003).

[15] Since Granger and Starica (2005) have documented the superiority of the paradigm of time-varying unconditional variance over some specifications of stationary long memory (dependence in time), our approach is based on interpreting the slow decay of the SACF of absolute returns as a sign of non-stationarities in the second-moment structure.

Hence, $\left\{\boldsymbol{R}_t\boldsymbol{R}_t^{\mathrm{T}}\right\}_t$ is an independent sequence of matrices with pointwise expectations $\boldsymbol{\Sigma}^2(t)$, a smooth function of time. This offers the framework for nonparametric regression on an equidistant design $t = \{1, \ldots, n\}$ regarding the estimation of variances and covariances in matrices $\left\{\boldsymbol{\Sigma}^2(t)\right\}_t$.

3.2 Estimating volatilities

With the tools of classical nonparametric regression on equidistant points $t = \{1, \ldots, n\}$, we derive from a *local polynomial regression method* (local constant regression) on $\left\{\boldsymbol{R}_t\boldsymbol{R}_t^{\mathrm{T}}\right\}_t$, with data localized by kernel functions K_h, and the method of least squares a *Nadaraya-Watson estimator*:

$$\hat{\boldsymbol{\Sigma}}^2(t) = \frac{\sum_{i=1}^{n} K_h(i-t)\boldsymbol{R}_i\boldsymbol{R}_i^{\mathrm{T}}}{\sum_{i=1}^{n} K_h(i-t)}, \tag{6}$$

where $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ is a rescaled kernel on support $[-h, h]$ of a symmetric kernel K on the compact support $[-1, 1]$. This is the two-sided volatility estimate for the multivariate regression model (2). Herzel et al. (2005) motivate the application of nonparametric regression by asymptotic results of Müller and Stadtmüller (1987). Moreover, they derive propositions on confidence intervals for $\left(\Sigma_{i,j}(t)\right)_{i,j}$.

For the remainder of this subsection (and within the simulations and empirical studies in chapters 5 and 4) we restrict ourselves to the univariate case $d = 1$ and outline some statistical results. First of all, we distinct between the two-sided (symmetrical) and the one-sided (historical) estimation of variance. The corresponding univariate Nadaraya-Watson estimators (NWE) are:

(I) Two-sided NWE (smoother):

$$\hat{\sigma}^2(t) := \frac{\sum_{i=1}^{n} K_h(i-t)R_i^2}{\sum_{i=1}^{n} K_h(i-t)} \tag{7}$$

(II) One-sided NWE (filter):

$$\hat{\sigma}^2_{(1)}(t) := \frac{\sum_{i=1}^{t} K_h(i-t)\widetilde{R}_i^2}{\sum_{i=1}^{t} K_h(i-t)}, \tag{8}$$

with $\widetilde{R}_i := X_i - \bar{X}_{i-1}$.

The first volatility estimate includes past and future returns. The estimation of $\sigma^2(t)$ depends on a symmetrical data base around t, using all returns R_i that are temporally close enough, i.e. with design points i inside the bandwidth h around t. The calculation requires a sufficient long sample and is only possible if $1 + h \leq t \leq n - h$. Else the required band $[-h, +h]$ would go beyond the data base and boundary effects occur.[16] The asymmetric version (8), includes only past and current returns, and therefore is consistently applicable to estimate volatility at the last point n and to forecast close future returns. The asymmetric estimator $\hat{\sigma}^2_{(1)}(t)$ inserts returns $\widetilde{R}_i$ up to time t, restricted to values within the left-sided band $[t-h, t]$. Again a sufficient long past series is required, the (unbiased) estimation is possible for points t with $1 + h \leq t$. Else a boundary effect results on the left end. While the two-sided NWE is preferred to describe the dynamics of changes in a historical sample,[17] the one-sided NWE is applied in forecasting volatility.

In an asymptotic framework, that enables an increase of the frequency for observing data points on a fixed time-frame, Gürtler et al. (2009) studied the consistency and asymptotic normality of the nonparametric volatility estimates. Analytically, the observations were rescaled to a unit interval with design points $\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$ and the data base

[16]Several approaches for treating the boundary region $t \in [0, h)$ and $t \in (n-h, n]$ 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 left-sided estimator has size $n^{2/3}$, that could be interpreted as a boundary corrected estimator for the right interval boundary.

[17]Of course, inside a historical sample the one-sided NWE delivers generally a bigger estimation error than its two-sided counterpart due to the lack of information.

was refined gradually (by increasing n) to scan the regression function $s\left(\frac{t}{n}\right) = \sigma(t)$ more and more precisely. In the transformed multiplicative return model,

$$\begin{aligned} & X_{t,n} = \mu + s\left(\frac{t}{n}\right)\varepsilon_{t,n}, \quad t = 1, \ldots, n, \\ & \varepsilon_{1,n}, \ldots, \varepsilon_{n,n} \ \text{ iid rvs with } \ \mathrm{E}\varepsilon_{1,n} = 0, \ \mathrm{Var}\varepsilon_{1,\mathrm{n}} = 1, \\ & s\left(\frac{t}{n}\right), \quad t = 1, \ldots, n, \ \text{ a smooth, deterministic function of time,} \end{aligned} \tag{9}$$

the two-sided transformed NWE $\hat{s}^2_{h_n}(u)$ and the one-sided NWE $\hat{s}^2_{(1)h_n}(u)$ are structurally analogical to equations (7) and (8). Under certain regularity conditions[18] on the kernel K, the bandwidth h_n and the smoothness of the volatility function $s(\cdot)$, the following propositions are derived by Gürtler et al. (2009):

(P1) *Under the regularity conditions and setup (9) the sequence* $\left(\hat{s}^2_{h_n}(u)\right)_{n\in\mathbb{N}}$ *of two-sided estimators for* $s^2(u)$ *is consistent for all* $u \in (0,1)$.

(P2) *Let* $C \geq 0$ *and* $V := \mathrm{E}\varepsilon^4_{1,n} - 1 \in (0,\infty)$ *and the regularity conditions be satisfied. Then the sequence of estimators* $\left(\hat{s}^2_{h_n}(u)\right)_{n\in\mathbb{N}}$ *for* $s^2(u)$ *is asymptotically normally distributed for all* $u \in (0,1)$ *in terms of*

$$\begin{aligned} \sqrt{nh_n}\left(\hat{s}^2_{h_n}(u) - s^2(u)\right) &\xrightarrow{\mathrm{D}} \mathcal{N}\left(\beta(u), \tau^2(u)\right), \quad \textit{where} \\ \beta(u) &= \frac{C}{2}\left(s^2(u)\right)'' \int_{-1}^{1} v^2 K(v)\mathrm{d}v, \\ \tau^2(u) &= V s^4(u) \int_{-1}^{1} K^2(v)\mathrm{d}v. \end{aligned} \tag{10}$$

(P3) *Under the regularity conditions and setup (9) the sequence* $\left(\hat{s}^2_{(1)h_n}(u)\right)_{n\in\mathbb{N}}$ *of one-sided estimators for* $s^2(u)$ *is consistent for all* $u \in (0,1]$.

(P4) *Let* $D \geq 0$ *and* $V := \mathrm{E}\varepsilon^4_{1,n} - 1 \in (0,\infty)$ *and the regularity conditions be satisfied. Then the sequence of estimators* $\left(\hat{s}^2_{(1)h_n}(u)\right)_{n\in\mathbb{N}}$ *for* $s^2(u)$ *is asymptotically normally distributed for all* $u \in (0,1]$ *in terms of*

$$\begin{aligned} \sqrt{nh_n}\left(\hat{s}^2_{(1)h_n}(u) - s^2(u)\right) &\xrightarrow{\mathrm{D}} \mathcal{N}\left(\beta_{(1)}(u), \tau^2_{(1)}(u)\right), \quad \textit{where} \\ \beta_{(1)}(u) &= 2\,D\left(s^2(u)\right)' \int_{-1}^{0} vK(v)\mathrm{d}v, \\ \tau^2_{(1)}(u) &= 4\,V s^4(u) \int_{-1}^{0} K^2(v)\mathrm{d}v. \end{aligned} \tag{11}$$

From (P1) follows that the sequence of estimates is asymptotically unbiased at interior points of $[0,1]$ and its variance converges to zero as n goes to infinity. The stochastical convergence is to be concluded. The result is extended by inspecting the rate of convergence: $\sqrt{nh_n}\left(\hat{s}^2_{h_n}(u) - s^2(u)\right)$ is limited (in probability) due to an asymptotical bias, a finite variance and an asymptotic normal distribution. Concluding from (P2), the finite (n sufficiently large) pointwise approximation of $s^2(u)$ by $\hat{s}^2_{h_n}(u)$ is nearly distributed as:

$$\hat{s}^2_{h_n}(u) - s^2(u) \approx \mathcal{N}\left(\frac{h_n^2}{2}\left(s^2(u)\right)'' \int_{-1}^{1} v^2 K(v)\mathrm{d}v, \frac{V}{nh_n} s^4(u) \int_{-1}^{1} K^2(v)\mathrm{d}v\right). \tag{12}$$

[18](C1) Let $K : \mathbb{R} \to [0,\infty)$ be a symmetrical density with compact support $[-1,1]$, i.e. (i) $K(v) = 0\ \forall v \notin [-1,1]$, (ii) $\int_{-\infty}^{\infty} K(v)\mathrm{d}v = 1$, (iii) $\int_{-\infty}^{\infty} vK(v)\mathrm{d}v = 0$; (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) $h_n \overset{n\to\infty}{\longrightarrow} 0$, (ii) $nh_n, \ldots, nh_n^4 \overset{n\to\infty}{\longrightarrow} \infty$, $nh_n^6, nh_n^7, \ldots \overset{n\to\infty}{\longrightarrow} 0$, (iii) $nh_n^5 \overset{n\to\infty}{\longrightarrow} C^2 \geq 0$; (C4) Let s^2 be two times continuous differentiable; (C5) Let random variables $\varepsilon_{1,n}, \ldots, \varepsilon_{n,n}$ be iid with $\mathrm{E}\varepsilon_{1,n} = 0$, $\mathrm{Var}\varepsilon_{1,n} = 1$ and $\mathrm{E}\left|\varepsilon_{1,n}\right|^{4+\delta} < \infty$ for a $\delta > 0$ and $n \in \mathbb{N}$. Concerning the one-sided NWE, condition (C3) has to be replaced with: (C3') $K_{h_n}(\cdot) = \frac{1}{h_n}K\left(\frac{\cdot}{h_n}\right)$ with bandwidth restrictions: (i) $h_n \overset{n\to\infty}{\longrightarrow} 0$, (ii) $nh_n, nh_n^2 \overset{n\to\infty}{\longrightarrow} \infty$, $nh_n^4, nh_n^5, \ldots \overset{n\to\infty}{\longrightarrow} 0$, (iii) $nh_n^3 \overset{n\to\infty}{\longrightarrow} D^2 \geq 0$.

Here, the approximate bias has a negligible magnitude relative to the variance term. Thus, an approximative confidence interval for $s^2(u)$ can be simplistic implemented with Gaussian quantiles, centred at $\hat{s}^2_{h_n}(u)$ and with the above variance term. Discussions concerning optimal bandwidths continue. Since the mean squared error is $\mathrm{MSE}\hat{s}^2_{h_n}(u) = \mathrm{Bias}^2\hat{s}^2_{h_n}(u) + \mathrm{Var}\hat{s}^2_{h_n}(u)$, minimizing the function with respect to the bandwidth will provide the optimal trade-off between bias and variance.

Generally speaking, the statistical results of the two-sided estimate can be transmitted to the one-sided counterpart $\hat{s}^2_{(1)h_n}(u)$, conditioned on a faster convergence rate of the bandwidth h_n. The consistency result (P3) is not only valid at interior points of $[0,1]$, but also at the right frontier. This makes the estimator consistently applicable for forecasting volatility and return distributions. From the asymptotic normality result (P4), it follows again for sufficiently large n pointwise:

$$\hat{s}^2_{(1)h_n}(u) - s^2(u) \approx \mathcal{N}\left(2h_n\left(s^2(u)\right)'\int_{-1}^{0} vK(v)\mathrm{d}v, \frac{4\,V}{nh_n}s^4(u)\int_{-1}^{0}K^2(v)\mathrm{d}v\right). \tag{13}$$

Hence, an approximative confidence interval for $s^2(u)$ can be simplistic implemented in terms of normal quantiles, centred at $\hat{s}^2_{(1)h_n}(u)$ and scaled by the standard deviation resulting from above. Moreover, optimal one-sided bandwidths that minimize the MSE or MISE of $\hat{s}^2_{(1)h_n}(u)$ are concluded.

In the prevalent literature it is established, that the choice of the kernel function plays a relatively unimportant role compared with the optimal bandwidth for nonparametric regression. As we attend the smoothness conditions, we recommend a polynomial of fourth degree for the sequel, also called *biweight kernel*:[19]

$$K(u) := \begin{cases} \frac{15}{16}(1-u^2)^2 & , \quad |u| \le 1, \\ 0 & , \quad else. \end{cases} \tag{14}$$

The bandwidth is also called smoothing parameter because the aim is to find a trade-off between over- and under-smoothing.[20] Local bandwidth optimization can be based on minimizing the MSE with respect to h_n for a point in time, an appropriate global error criterion is the MISE. Using the above asymptotic results we get for sufficiently large n global optimal bandwidths (compare Gürtler et al. (2009)):

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

$$h_n^{\mathrm{opt}} = n^{-\frac{1}{5}}\left(\frac{V\left(\int_0^1 s^4(u)\mathrm{d}u\right)\left(\int_{-1}^1 K^2(v)\mathrm{d}v\right)}{\left(\int_0^1\left((s^2(u))''\right)^2\mathrm{d}u\right)\left(\int_{-1}^1 v^2K(v)\mathrm{d}v\right)^2}\right)^{\frac{1}{5}}. \tag{15}$$

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

$$h_{(1)n}^{\mathrm{opt}} = n^{-\frac{1}{3}}\left(\frac{V\left(\int_0^1 s^4(u)\mathrm{d}u\right)\left(\int_{-1}^0 K^2(v)\mathrm{d}v\right)}{2\left(\int_0^1\left((s^2(u))'\right)^2\mathrm{d}u\right)\left(\int_{-1}^0 vK(v)\mathrm{d}v\right)^2}\right)^{\frac{1}{3}}. \tag{16}$$

Those criteria can be used directly for simulations (as in section 4) where the actual volatility $s(\cdot)$ is a predefined input. But for empirical samples it is problematic,[21] since it is the prior task to estimate $s(\cdot)$. One solution is the *cross-validation method* (CV), that determines the optimal smoothing parameter solely with the return series. In terms of real-world samples $\{X_t\}_{t=1,\ldots,n}$ transforming the setup for estimating $s\left(\frac{t}{n}\right)$ is redundant, rather the volatility $\sigma(t)$

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

[20] Oversmoothing means 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 results (small variance, but biased). Undersmoothing averages over a very small neighbourhood (small bandwidth), where only a few recent data points are included; a rough shape results (small bias, but large variance).

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

should be estimated directly (as in section 5). Thus, we turn back the transformation, $h = nh_n$, and motivate the CV-criterion in the regression model (2) with $d = 1$.

Implementing the 'leave-one-out prediction' for CV, the variance $\sigma^2(j)$ has to be reestimated for each point $j = 1, \ldots, n$ without using the actual observation R_j^2 (or $\widetilde{R}_j^2$, respectively) itself. We modify the nonparametric volatility estimators (7) and (8) to accordant cross-validation estimators (CVEs) $\hat{\sigma}_h^{2(j)}(j)$ and $\hat{\sigma}_{(1)h}^{2(j)}(j)$. The bandwidth selection criterion is to minimize the CV-function, which is the sum of squared differences between returns R_j^2 and CVEs $\hat{\sigma}_h^{2(j)}(j)$ (alternatively $\widetilde{R}_j^2$ and $\hat{\sigma}_{(1)h}^{2(j)}(j)$).

(I) The two-sided CV-function and CVE are:

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{j=1}^{n}\left(R_j^2 - \hat{\sigma}_h^{2(j)}(j)\right)^2 = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\sum_{i=1}^{n} K_h(i-j)\left(R_j^2 - R_i^2\right)}{\sum_{i=1,i\neq j}^{n} K_h(i-j)}\right)^2 \xrightarrow{!} \min_{h>1},$$

$$\text{where } \hat{\sigma}_h^{2(j)}(j) = \frac{\sum_{i=1,i\neq j}^{n} K_h(i-j) R_i^2}{\sum_{i=1,i\neq j}^{n} K_h(i-j)}, \quad h > 1 \quad (\forall j = 1, \ldots, n). \tag{17}$$

(II) The one-sided CV-function and CVE are:

$$\mathrm{CV}_{(1)}(h) = \frac{1}{n-1}\sum_{j=2}^{n}\left(\widetilde{R}_j^2 - \hat{\sigma}_{(1)h}^{2(j)}(j)\right)^2 = \frac{1}{n-1}\sum_{j=2}^{n}\left(\frac{\sum_{i=1}^{j-1} K_h(i-j)\left(\widetilde{R}_j^2 - \widetilde{R}_i^2\right)}{\sum_{i=1}^{j-1} K_h(i-j)}\right)^2 \xrightarrow{!} \min_{h>1}$$

$$\text{where } \hat{\sigma}_{(1)h}^{2(j)}(j) = \frac{\sum_{i=1}^{j-1} K_h(i-j)\widetilde{R}_i^2}{\sum_{i=1}^{j-1} K_h(i-j)}, \quad h > 1 \quad (\forall j = 2, \ldots, n). \tag{18}$$

The CV-optimal bandwidths h^{CV} and $h^{\mathrm{CV}_{(1)}}$ for the two-sided and one-sided volatility estimation are numerically found via a value table for integers $h \geq 2$ or via analyzing the resulting CV-plot.

3.3 Fitting innovations

In the last step of calibrating the non-stationary regression approach we have to model the distribution of innovations $\{\varepsilon_t\}_t$. First, the residuals $\varepsilon_{k,t}$ are componentwise estimated by dint of demeaned returns $R_{k,t}$ and volatility estimates $\hat{\sigma}_k(t) = \left(\hat{\Sigma}_{k,k}(t)\right)_{i,j}$ (the k-th diagonal element of the square root of the estimate $\hat{\mathbf{\Sigma}}^2(t)$ for $\mathbf{S}_t\mathbf{S}_t^{\mathrm{T}}$) at each point t. Again we distinguish the two setups dependent on return information:

(I) Two-sided innovation estimator:

$$\hat{\varepsilon}_{k,t} = \frac{R_{k,t}}{\hat{\sigma}_k(t)} \tag{19}$$

(II) One-sided innovation estimator:

$$\hat{\varepsilon}_{(1)k,t} = \frac{\widetilde{R}_{k,t}}{\hat{\sigma}_{(1)k}(t)} \tag{20}$$

Due to their independence it is sufficient to specify the distributions of $\hat{\varepsilon}_{k,t}$, $k = 1, \ldots, d$ univariate.[22] The easiest approach without any assumptions could be the use of the empirical distribution function $\hat{F}_n^{\mathrm{emp}}(x)$ of the series $\{\hat{\varepsilon}_t\}_t$, but this is not able to capture heavy tails.[23,24] Against this background, 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.[25]

[22] For the sake of simplicity we omit the coordinate k, that is arbitrary constant, subsequently. Furthermore, we omit the identification $_{(1)}$, as we require to estimate a series within the same setup, and as the techniques are identical.

[23] Being $\varepsilon_{\max} = \max_t \hat{\varepsilon}_t$ and $\varepsilon_{\min} = \min_t \hat{\varepsilon}_t$ of the innovation sample $\{\hat{\varepsilon}_t\}_t$, then $\hat{F}_n^{\mathrm{emp}}(\varepsilon_{\max} + \delta) = 1$ for all $\delta \geq 0$ and $\hat{F}_n^{\mathrm{emp}}(\varepsilon_{\min} - \epsilon) = 0$ for all $\epsilon > 0$. Consequently the empirical distribution function underestimates extremes, since the probability for extreme future innovations $\varepsilon_N \notin [\varepsilon_{\min}, \varepsilon_{\max}]$, $N > n$ would equal 0.

[24] Moreover, the normal distribution function drops out due to neglecting skewness and heavy tails. Hu and Kercheval (2010) approximate stock return series (after filtering with GARCH(1, 1)) with distributions from the Generalized Hyperbolic family. They conclude that the Student-t and a Skewed-t distribution are amongst the best and can be efficiently fitted to (multivariate) data for the purpose of portfolio optimization.

[25] The Pearson VII family includes the Student 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.

It has the following one-sided density with shape parameter m and scale parameter c:

$$f_{m,c}^{\mathrm{VII}(1)}(x) = \frac{2\Gamma(m)}{c\Gamma\left(m-\frac{1}{2}\right)\pi^{1/2}}\left(1+\left(\frac{x}{c}\right)^2\right)^{-m}\mathrm{I}_{[0,\infty)}(x). \tag{21}$$

The 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 to absolute values of negative innovations $\{-\hat{\varepsilon}_t \mid \hat{\varepsilon}_t < 0\}_t$. Because there are usually about as many positive as negative innovations in a financial time series, the median of innovations may be assumed to be zero. Hence, the fitted one-sided Pearson VII densities $f_{m_+,c_+}^{\mathrm{VII}(1)}$ and $f_{m_-,c_-}^{\mathrm{VII}(1)}$ are combined as

$$f_{m_+,c_+,m_-,c_-}^{\mathrm{VII}}(x) = \frac{1}{2}\Big(f_{m_-,c_-}^{\mathrm{VII}(1)}(-x)\mathrm{I}_{(-\infty,0)}(x) + f_{m_+,c_+}^{\mathrm{VII}(1)}(x)\mathrm{I}_{[0,\infty)}(x)\Big), \tag{22}$$

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

Agreeing with Drees and Starica (2002), Herzel et al. (2005), Mikosch and Starica (2003) or Kitagawa and Nagahara (1999), the Pearson VII distribution can capture some heavy tailed innovations quite nicely. In their articles the parameters (m_+, c_+) and (m_-, c_-) are estimated with maximum-likelihood methods. In contrast, we solve that task with a method of moments, based on the Pearson systems in Johnson and Kotz (1970) and a transformation from Gürtler et al. (2009). It results in:

$$m = \frac{5\beta_2 - 9}{2\beta_2 - 6}\,,\ c = \sqrt{\frac{2\beta_2\mu_2}{\beta_2 - 3}}\,, \tag{23}$$

where $\mu_2(\varepsilon) = \mathrm{E}\varepsilon^2$ and $\beta_2(\varepsilon) = \frac{\mathrm{E}(\varepsilon-\mathrm{E}\varepsilon)^4}{(\mathrm{E}(\varepsilon-\mathrm{E}\varepsilon)^2)^2}$ (kurtosis), at which $\beta_2 > 3$ is required. Hence, the conception is only applicable for samples that are heavier than a normal distribution. By inserting empirical estimates for β_2 and μ_2 of subsamples $\varepsilon_+ := \{\hat{\varepsilon}_t \mid \hat{\varepsilon}_t \geq 0\}_t$ and $\varepsilon_- := \{-\hat{\varepsilon}_t \mid \hat{\varepsilon}_t < 0\}_t$ we fit $f_{m,c}^{\mathrm{VII}(1)}$ to gains and to losses, respectively. The estimation of β_2 can be reduced to:

$$\hat{\beta}_2(\varepsilon_+) = \frac{\mathrm{E}\varepsilon_+^4}{\mathrm{E}\varepsilon_+^2}\,,\ \hat{\beta}_2(\varepsilon_-) = \frac{\mathrm{E}\varepsilon_-^4}{\mathrm{E}\varepsilon_-^2}\,. \tag{24}$$

Moreover, the symmetric Pearson type VII distribution equates to a scaled Student-t distribution with $2m-1$ degrees of freedom (df).[26] With $g_\cdot$ and $G_\cdot$ being the density and the cdf of a Student-t rv we get:

$$f_{m,c,m,c}^{\mathrm{VII}}(x) = \frac{1}{\gamma}\,g_{2m-1}\left(\frac{x}{\gamma}\right),\ \text{with } \gamma := \frac{c}{\sqrt{2m-1}}\,, \tag{25}$$

$$F_{m,c,m,c}^{\mathrm{VII}}(x) = \int_{-\infty}^{x}\frac{1}{\gamma}\,g_{2m-1}\left(\frac{y}{\gamma}\right)\mathrm{d}y \overset{z:=y/\gamma}{=} \int_{-\infty}^{x/\gamma}\frac{1}{\gamma}\,g_{2m-1}(z)\,\gamma\,\mathrm{d}z = G_{2m-1}\left(\frac{x}{\gamma}\right). \tag{26}$$

Because of the simple transformation from the Pearson VII to the Student-t distribution, whose quantiles are to be looked up in tables, the non-stationary model (2) can be implemented as a factor-based approach for the task of VaR calculation. The univariate Value at Risk $\mathrm{VaR}^{1-\alpha}(t)$ of an exposure $w(t)$, with a yield X_t following the regression model, can be modelled as the product of $w(t)$, the nonparametric estimated volatility $\hat{\sigma}_{(\cdot)}(t)$ and the α-quantile of a Pearson VII innovation. Being $u_{m,c;\alpha}$ the latter and $t_{2m-1;\alpha}$ the Student α-quantile it immediately follows: $u_{m,c;\alpha} = \gamma\, t_{2m-1;\alpha}$. With respect solely to the loss density $f_{m_-,c_-}^{\mathrm{VII}(1)}(-x)\mathrm{I}_{(-\infty,0)}(x)$ of the asymmetric Pearson VII distribution, a transformation of confidence levels $\alpha \in (0, 0.5)$ has to be kept in mind: $\alpha^{(1)} = 2\alpha$.

Concluding, asymmetric and fat tailed random innovations, modelled as Pearson-type VII distributed rvs, enable a flexible capturing of the return series, and lead altogether to an enhanced multiplicative approach. The estimated distributions $\hat{F}_{\varepsilon_k}^{\mathrm{VII}}$ of the d independent random innovations together with (the square root of) the covariance matrix estimates $\hat{\boldsymbol{\Sigma}}^2(t)$ and the mean vector $\bar{\boldsymbol{X}}_n$ completely specify the distribution of returns $\boldsymbol{X}_t$ in the regression model (2).

[26] From the Student-t connection and the regularity conditions of NWEs follows a restriction to the shape parameter: $m > \frac{5}{2}$. Student's df is a measure for the heavy-tailedness and is also called *tail index point estimate*, e.g. in Drees and Starica (2002). The tail index of the distribution may be defined as the order of the highest finite absolute moment.

4 Simulation experiment

So far we have prepared the necessary statistical tools. From now on we specialize on the univariate non-stationary modelling. In this section we analyze the goodness of fit of the approach (2) for simulations of volatility functions and price processes.

4.1 Nonparametric volatility estimation

First, we predefine a heteroscedastic volatility function (standard deviation) as:

$$\sigma^{\mathrm{a}}(t) = \frac{1}{10}\left(\sin\left(\frac{2\pi}{100}t\right)+1\right) \quad t \in [0,500], \tag{27}$$

$$s^{\mathrm{a}}(u) = \frac{1}{10}\left(\sin(10\pi u)+1\right) \quad u \in [0,1]. \tag{28}$$

The values $\sigma^{\mathrm{a}}(t)$ on the discrete design $1,\ldots,500$ could be thought as annualized return volatilities at the end of days t, observed over two years (250 trading days p.a.), and $s^{\mathrm{a}}(u)$ is the transformed version. Both functions correspond to a multimodal oscillation with 5 periods. The subsequent study reestimates the volatility in a simplified case of the standardized regression model (9) with return expectation $\mu = 0$ and innovations distributed as $\varepsilon_{1,n} \sim \mathcal{N}(0,1)$.[27] The heteroscedasticity of the simulated series is estimated via the two-sided $\hat{s}^2_{h_n}(u)$ and the one-sided NWE $\hat{s}^2_{(1)h_n}(u)$. The influence of sample size (i.e. increasing the data density on the time-frame $[0,1]$) to the estimator's fit is focussed especially.[28] Bandwidths are optimized with the MISE-criterion.

Figure 1 displays the nonparametric curve estimation of the volatility function s^{a} for different sample sizes in a median simulation of 65 repeats.[29] We detect that the fit improves with the number of observations for both NWEs. The median SSE increases considerably slower than the sample size. The approximation of $s^{\mathrm{a}}(u)$ by nonparametric estimates $\hat{s}_{h_n}(u)$ and $\hat{s}_{(1)h_n}(u)$ is noticeable for smaller samples of 100 or 500 points and satisfying using 1000 design points. Due to additional future information, the two-sided estimator achieves generally a better and smoother fit. The historical approach lags behind, as the volatility does not increase before the first extremal event, and after a series of shocks it decays typically slower. We did not correct boundary effects, implying the first nh_n points to be distorted, and also the last nh_n values for the two-sided NWE. On the right boundary both estimators are nearly the same (apart from different bandwidths). For the sample size of 5000 both estimates reproduce the predefined function excellent. To appreciate the goodness of cross-validation, we estimate optimal bandwidths again with the CV method. Although we do not include any knowledge of s^{a} except realized returns, the outcomes are quite similar parameters and volatility graphs. Regarding the 5000-point setup the resulting bandwidths are 0.030 (two-sided) and 0.015 (one-sided), that are very close to the MISE-optimal parameters.[30] Consequently, the volatility estimates $\hat{s}_{h_n}(u)$ or $\hat{s}_{(1)h_n}(u)$ are almost indistinguishable from the bottom graph of figure 1.

We extend the simulation to a simple, discrete price process $\{P_t\}_{t=0,\ldots,500}$, where $\sigma^{\mathrm{a}}(t)$ is the time-variant part of an annualized volatility function $\widetilde{\sigma}^{\mathrm{a}}(t)$, added by a time-invariant component σ_0. Moreover, a constant trend μ and a heavy-tailed innovation approach of returns are modelled:

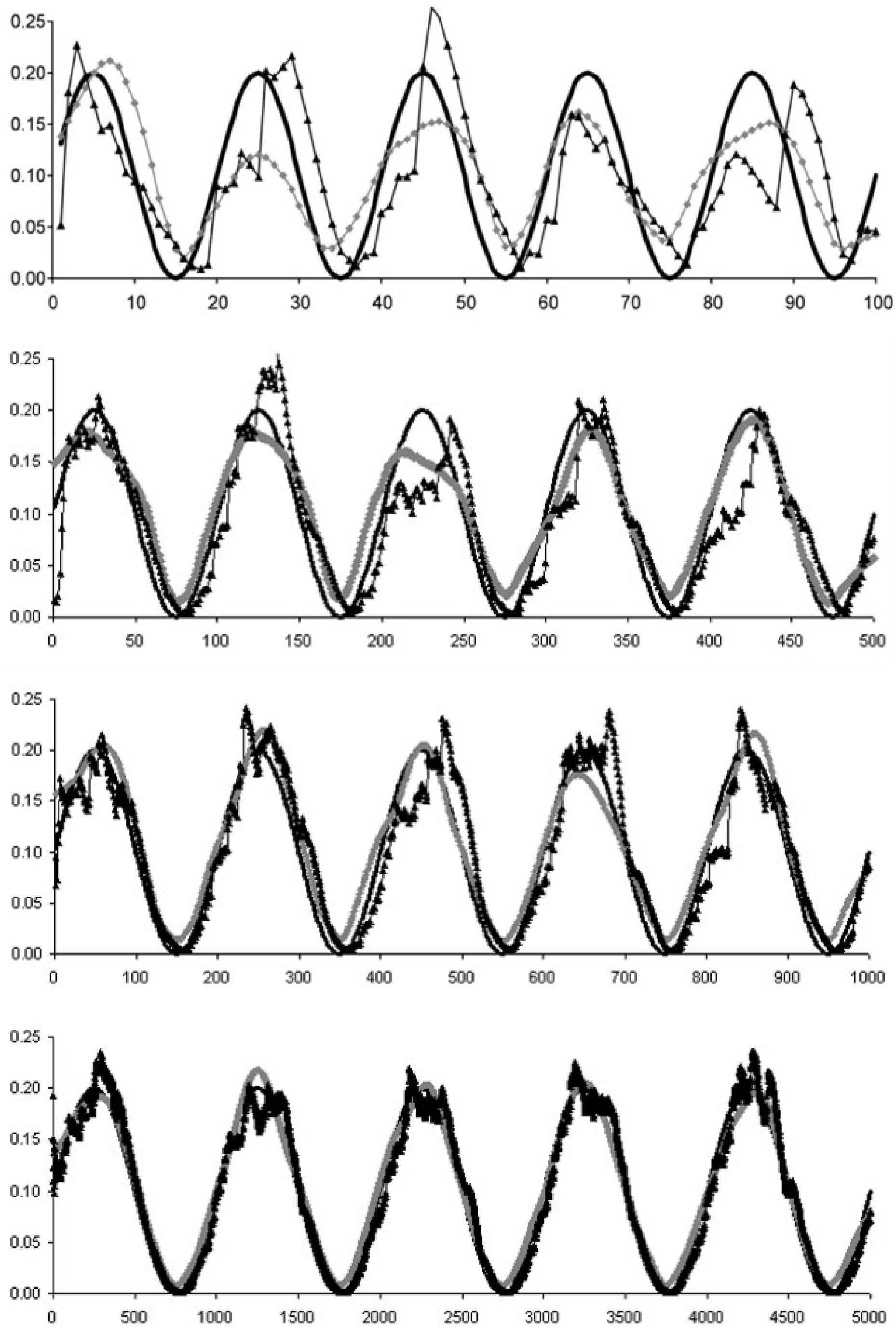
$$\begin{aligned}
P_t &= P_{t-1}\, e^{X_t}\ ,\ P_0 := 1000\ ,\ t = 1,\ldots,500, \quad \text{where} \\
X_t &= \mu + \sigma(t)\varepsilon_t, \ \text{with } \mu := \frac{15\%}{250}, \\
\sigma(t) &= \frac{\widetilde{\sigma}^{\mathrm{a}}(t)}{\sqrt{250}},\ \widetilde{\sigma}^{\mathrm{a}}(t) := \sigma^{\mathrm{a}}(t) + \sigma_0 \ \text{ and } \sigma_0 := 10\%, \\
\varepsilon_t &\sim F^{\mathrm{VII}}_{m,c,m,c} \ \text{with } m := 4 \text{ and } c := \sqrt{5}.
\end{aligned} \tag{29}$$

[27] We later include the proposed innovation modelling. The choice of an expectation $\mu = const$ had a negligible influence on the simulation example due to centring returns first. Assuming $\mu = 0$ is for the sake of convenience.

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

[29] The median simulation is determined as we calculate a sum of squared estimation errors $\mathrm{SSE}\hat{s}_{h_n} := \sum_{i=1}^{n}\left(\hat{s}_{h_n}\left(\frac{i}{n}\right) - s\left(\frac{i}{n}\right)\right)^2$ for each simulated sample and order paths by the extent of their SSE.

[30] The absolute bandwidth differences are lower/ equal 0.0025 and might even be caused by the chosen grid pattern of widths 0.005. The SSEs are with 0.54153 for the two-sided and 1.57060 for the one-sided setup marginally larger.



	BW two-sided	error (SSE)	BW one-sided	error (SSE)
100 points	0.06013	0.15739	0.06377	0.30549
500 points	0.04358	0.25614	0.03730	0.59948
1000 points	0.03794	0.28939	0.02960	0.75579
5000 points	0.02750	0.46056	0.01731	1.36289

Figure 1: 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 $\hat{s}_{h_n}$ (rhombuses, grey) and one-sided NWEs $\hat{s}_{(1)h_n}$ (triangles, black) with MISE-optimal bandwidths (BW). The abscissa intercepts are labeled according to design points.

The log-returns of prices, $X_t = \ln(P_t/P_{t-1}) = \mu + \sigma(t)\varepsilon_t$, correspond to the untransformed regression approach (2). Innovations ε_t are modelled as symmetrical Pearson VII distributed rvs with equal shape parameters m, c for gains and losses.[31]

A representative simulated path of prices and its nonparametric estimated return volatilities (relative to the predetermined volatility without noise) are depicted in figure 2. Here, the volatility fit is quite acceptable despite the small sample size of 500 and the strong influence of heavy-tailed random innovations ε_t. The comparison to the price graph shows, that the NWEs $\hat{\sigma}(t)$ and $\hat{\sigma}_{(1)}(t)$ are able to capture fast phases of market shocks and increased volatility, respectively.[32]

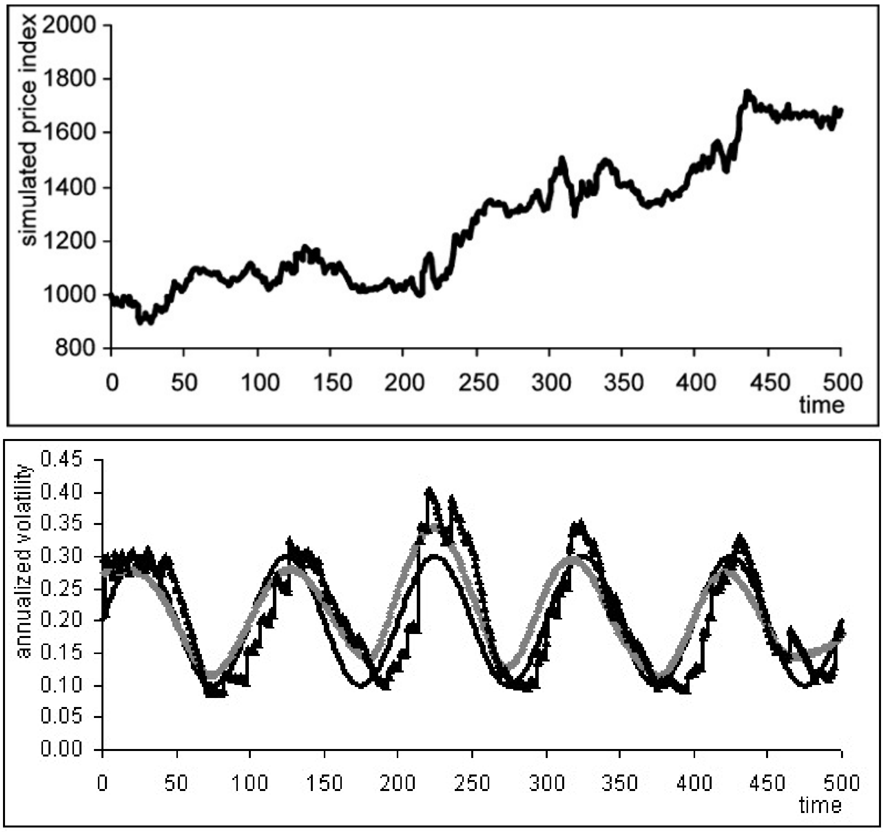


Figure 2: Simulation path for the price process (29) (top) and annualized volatility estimates (bottom) from the regression model (2) for log-returns: $\hat{\sigma}(t)$ (rhombuses, grey) and $\hat{\sigma}_1(t)$ (triangles, black) of the predefined volatility $\widetilde{\sigma}^{\mathrm{a}}(t)$ without noise (continuous line, black) for 500 design points and CV-optimal bandwidths ($h^{\mathrm{CV}} = 34$, $h^{\mathrm{CV}_{(1)}} = 29$).

4.2 Forecasting experiment

We continue with a forecasting experiment, holding up the price process (29) and its realization displayed in figure 2, i.e. realized innovations $\{\varepsilon_t\}_t$ were frozen. We imagine having observed only the first 251 prices $(P_0, \ldots, P_{250})$ up to a forecast starting point $t_0 = 250$, and call that half of the sample as 'in-sample' part. The second, 'out-of-sample' half will follow within the next year. We forecast the distribution of the 1-day ahead return X_{t+1} with the information available at $t \geq t_0$. Concretely, we use the distribution of X_{t_0} as a forecast of X_{t_0+1} (with $\hat{\sigma}_{(1)}(t_0)$ being the volatility

[31]For the simulation, uniform$[0,1]$ random numbers α from MS Excel were interpreted as probabilities determining α-quantiles of a Student-t_{2m-1} rv. Then the Pearson VII random number is $u_{m,c;\alpha} = \frac{c}{\sqrt{2m-1}} t_{2m-1;\alpha}$.

[32]The simulation may replicate very different paths, because the volatility level (average 20% p.a.) dominates the price process more than its trend ($\mu = 15\%$ p.a.) and the incident of extreme innovations may cause considerable shocks. This implies a slight variation of optimal bandwidths and NWEs, adapted to the realized volatility.

forecast for $\sigma(t_0+1)$; naturally we work with the one-sided setup). Afterwards, we develop the distributions of X_{251}, $\ldots$, X_{500} successively with the new returns, their volatility estimates and the innovation parametrization.

The model is calibrated in-sample and parameters are fixed in the out-of-sample analysis:[33] After assessing the centred returns $\widetilde{R}_1, \ldots, \widetilde{R}_{t_0}$, the optimal smoothing parameter $h^{\mathrm{CV}_{(1)}} = 36$ days is identified for nonparametric volatility estimation. This differs slightly from the full sample optimum, caused by random innovations, but the graph of one-sided volatility estimates $\hat{\sigma}_{(1)}(1), \ldots, \hat{\sigma}_{(1)}(t_0)$ and its out-of-sample continuation is still close to figure 2. The in-sample series of estimated innovations $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_{t_0}$ is fitted by the asymmetric Pearson VII density $f^{\mathrm{VII}}_{m_+,c_+,m_-,c_-}$, resulting in optimal parameters $m_+ = 5.6299, c_+ = 2.9038$ for the right tail and $m_- = 7.1976, c_- = 3.2758$ for the left tail.[34] We observe in figure 3 a pretty good approximation of the innovation frequencies by the Pearson VII density (black line), the asymmetry in the distribution comes along with deviations in the histograms of realized negative and positive innovations. A comparison to the standard normal density reveals deviations in the middle of the distribution and in the tails, in a way that extreme Pearson VII quantiles are significantly greater than Gaussian quantiles.[35]

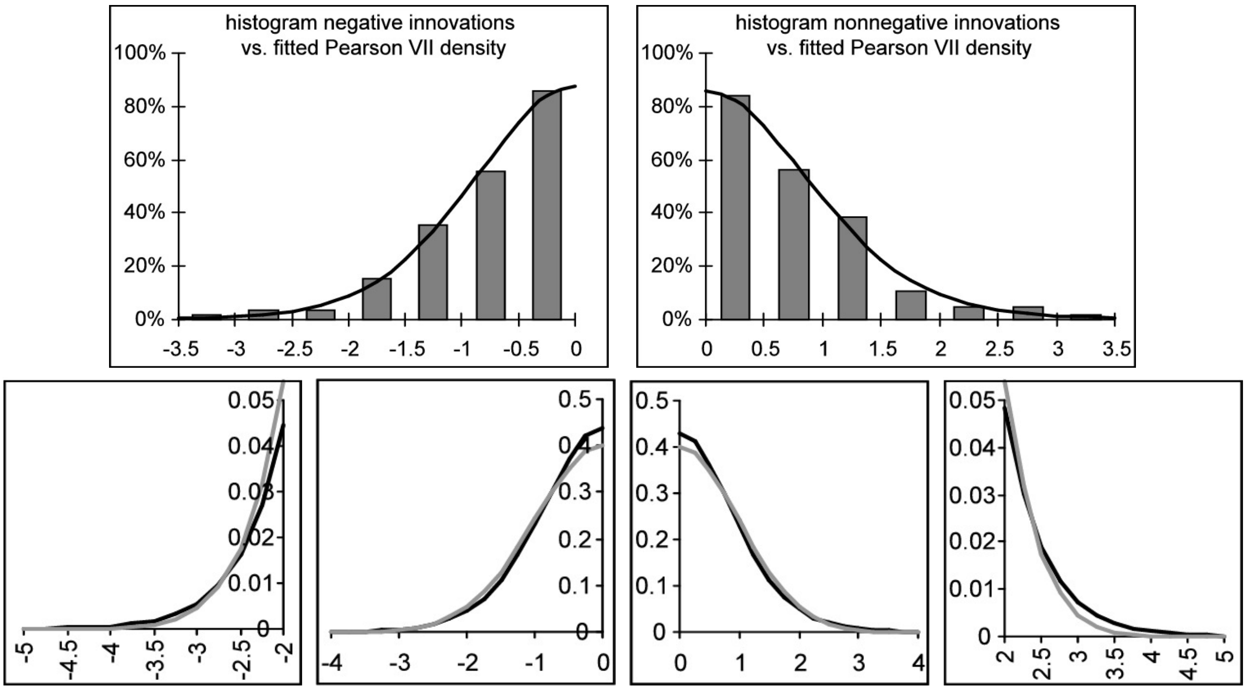


Figure 3: In-Sample Pearson VII fit of negative and nonnegative innovations by the density $f^{\mathrm{VII}}_{5.63,2.90,7.20,3.28}$. The bottom graphs focus on the tails of asymmetric Pearson VII (black) vs. standard normal density (grey).

It should be attended, that in general a trade-off between the smoothness of volatility estimates and the innovation's distribution is observed: The bigger the bandwidth in nonparametric volatility estimation the smoother is the volatility graph and the more heavy-tailed are the innovations. Vice versa, this could be the reason why the Pearson VII fit of innovations fails for good volatility estimates with small bandwidths. Then the requirement of a innovation kurtosis greater than 3 might be violated. In this case the normal distribution may be conservative for the task of extreme quantile approximation.

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 $\mathrm{H}_0 : p = \alpha$ for a $(1-\alpha)$ VaR-level. Based on a normal

[33] Alternatively one could recalibrate the model every day or with fixed period out-of-sample, to incorporate new information.

[34] Even though we had assumed a symmetric Pearson VII distribution, the method of moments on random residuals (small sample) may estimate different parameters m_+, c_+ vs. m_-, c_- for positive and negative innovations.

[35] Concerning the left tail, we see that lower confidence levels (in terms of maximum losses), as 95%, may have Pearson quantiles that are absolutely smaller than the normal (-1.5850 Pearson vs. -1.6449 Gaussian), but for extreme shortfall levels, as 99.5%, Pearson overtakes the Gaussian quantiles (-2.6961 vs. -2.5758).

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

$$\mathrm{LR}_{n,p}(N) = -2\ln\left[(1-p)^{n-N}p^N\right] + 2\ln\left[\left(1-\frac{N}{n}\right)^{n-N}\left(\frac{N}{n}\right)^N\right] \tag{30}$$

is χ^2_1 distributed under H_0. Therefore, a risk measure is rejected on a 5% level of significance if $\mathrm{LR}_{n,p}(N) > 3.84$. The Kupiec test is widely used to evaluate risk models, compare e.g. Choi and Nam (2008). In addition the penalty zones of the Basel II backtesting are based on this methodology (see Jorion (2006)).

Regarding the regression model implementation of the price process (29), we apply Kupiec backtesting to evaluate the forecasted return distribution. With focus on the loss tail, the relative 1-day Value at Risk $\mathrm{VaR}^{1-\alpha,1\mathrm{d}}_{\mathrm{RM}}(t)$ for the next day's return X_{t+1} has the factor-based form:

$$\mathrm{VaR}^{1-\alpha,1\mathrm{d}}_{\mathrm{RM}}(t) \quad = \quad \bar{X}_t + \hat{\sigma}_{(1)}(t)\, u_{m,c;\alpha^{(1)}} = \bar{X}_t + \hat{\sigma}_{(1)}(t)\,\frac{c}{\sqrt{2m-1}} t_{2m-1;\alpha^{(1)}}\,, \tag{31}$$

where $u_{m,c;\alpha^{(1)}}$ and $t_{2m-1;\alpha^{(1)}}$ are the quantiles of the left-side Pearson VII fit and the corresponding Student-t expression, $\alpha^{(1)} = 2\alpha$. This forecast for a maximum loss, that is not exceeded in $(t, t+1]$ with probability $1-\alpha$, is compared to the realized return X_{t+1}. A shortfall is observed if $X_{t+1} \leq \mathrm{VaR}^{1-\alpha,1\mathrm{d}}_{\mathrm{RM}}(t)$. The out-of-sample returns entailed 5 exceedances over the VaR threshold concerning a 99% confidence level. This deviates slightly from the expected number 2.5, but is within the allowed range $\{1,\ldots,6\}$ of shortfalls for a 5% level of significance ($\mathrm{LR}_{249,1\%}(5) = 1.98$). Consequently, the return model (2) is accepted.

Because the simulation example is chosen with quite extreme volatility changes and a short period of oscillation, a few exceedances occur when the volatility increases again after a low. Nevertheless, model acceptance is shown for all confidence levels greater-than-or-equal 80% with shortfall numbers being quite close to its expectations, as presented in the third column of figure 6. Hence, the non-stationary modelling of the price process (29) is proved to deliver good forecasts of 1-day return distributions.

4.3 Comparison to ARCH-models

We finish the simulation study with a performance comparison to traditional risk models. The following approaches are based on the initial return process (1) for $\{X_t\}_{t=1,\ldots,n}$, but differ in their (one-sided) volatility and innovation description. Estimates $\mathrm{VaR}^{1-\alpha,1\mathrm{d}}_{\cdot}(t)$ for the relative 1-day Value at Risk of the next day's return X_{t+1} are deduced in each case.

(1) Univariate delta-normal-model (parametric VaR model):

$$\begin{aligned}\bar{\sigma}_t &= \frac{1}{\min\{t,250\}-1}\sum_{i=\max\{t-249,1\}}^{t}\left(X_i - \bar{X}_t\right)\,, \\ \mathrm{VaR}^{1-\alpha,1\mathrm{d}}_{\mathrm{PM}}(t) &= \bar{X}_t + \bar{\sigma}_t z_\alpha\,,\end{aligned} \tag{32}$$

where $\bar{\sigma}_t$ is the empirical standard deviation (sd) on a 250 points moving window (cut on the left boundary) and the volatility forecast for $\sigma(t+1)$. $\bar{X}_t$ is the corresponding empirical mean and z_α the α-quantile of the standard normal distribution.

(2) GARCH$(1,1)$ model with Student-t distributed innovations (t-GARCH$(1,1)$):

$$\begin{aligned}\varsigma^2_{t+1} &= \alpha_0 + \alpha_1 \widetilde{R}^2_t + \beta_1\varsigma^2_t\,, \\ \mathrm{VaR}^{1-\alpha,1\mathrm{d}}_{\mathrm{GARCH}}(t) &= \bar{X}_t + \varsigma_{t+1} t_{n;\alpha}\,,\end{aligned} \tag{33}$$

where ς^2_{t+1} is the conditional volatility estimate of X_{t+1} given past centred returns $\widetilde{R}_t := X_t - \bar{X}_{t-1}$. Parameters α_0, α_1 and β_1 are required to be nonnegative and estimated with quasi-maximum likelihood methods from the software EViews. $t_{n;\alpha}$ is the α-quantile of a t-distribution with n df, that is assessed with a method of moments over realized innovations $\hat{\epsilon}_t = \widetilde{R}_t/\varsigma_t$.[36]

(3) EGARCH$(1,1)$ model with Student-t distributed innovations (t-EGARCH$(1,1)$):

$$\begin{aligned} \ln \breve{\varsigma}_{t+1}^2 &= \alpha_0 + \alpha_1 \frac{\left|\widetilde{R}_t\right|}{\breve{\varsigma}_t} + \gamma_1 \frac{\widetilde{R}_t}{\breve{\varsigma}_t} + \beta_1 \ln \breve{\varsigma}_t^2 \\ \mathrm{VaR}_{\mathrm{EGARCH}}^{1-\alpha,1\mathrm{d}}(t) &= \bar{X}_t + \breve{\varsigma}_{t+1} t_{n;\alpha}\,, \end{aligned} \tag{34}$$

where $\breve{\varsigma}_{t+1}^2$ is the conditional volatility, modelled and estimated in the same framework as item (2), with $\alpha_0, \alpha_1, \gamma_1 \in \mathbb{R}$ and $|\beta_1| < 1$.

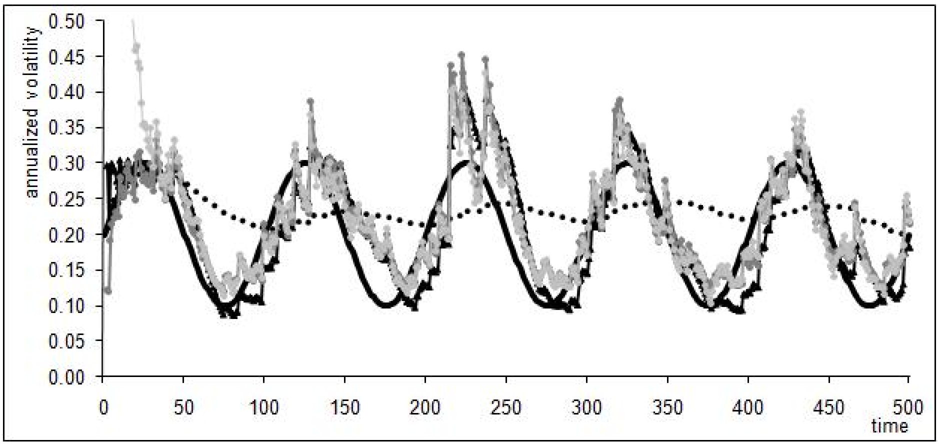


Figure 4: Annualized volatility estimates for log-returns in the price process (29) and comparison to the predefined volatility $\widetilde{\sigma}^{\mathrm{a}}(t)$ without noise (continuous line, black): Standard deviation $\bar{\sigma}_t$ (dotted line, black; 250-day), one-sided NWE $\hat{\sigma}_1(t)$ (triangles, black; $h^{\mathrm{CV}(1)} = 29$), GARCH$(1,1)$ ς_{t+1} (bullets, dark grey; $\alpha_0 = 6.46 \cdot 10^{-6}$, $\alpha_1 = 0.1346, \beta_1 = 0.8430$) and EGARCH$(1,1)$ volatility forecast $\breve{\varsigma}_{t+1}$ (bullets, light grey; $\alpha_0 = -0.6069$, $\alpha_1 = 0.2757, \gamma_1 = 0.0212, \beta_1 = 0.9535$).

First, we oppose the volatility structures concerning the price process (29) with its realization from the end of section 4.1. Figure 4 exhibits some interesting features: The 250-day sd oversmoothes strongly and lags behind the extrema. The unconditional one-sided NWE and the conditional volatility estimates from GARCH$(1,1)$ and EGARCH$(1,1)$ scarcely differ from each other.[37] Even their peaks are nearly congruent, with GARCH$(1,1)$ leading slightly in the maxima and NWE being somewhat lower in phases of decreased volatility. Deviations from the predefined volatility occur due to realized shocks from random innovations. As volatility estimates initialize themselves with respect to the calibration, about the first 25 values should be excluded from volatility comparisons and former model adaptions. The Student-t fit of GARCH (6.06 df) and EGARCH innovations (6.17 df) is assessed to be plausible, albeit there is too less probability mass around the mean. Focussing only the volatility fit, no dominances of the modelling types are concluded, but the non-stationary approach seems to be at least an equipollent alternative.

In terms of the Kupiec backtesting we compare the forecasting abilities of the different approaches. According to section 4.2, models are calibrated in-sample for prices $P_0, \ldots, P_{250}$ and employed out-of-sample for forecasting return distributions of X_t, $t > 250$. The in-sample fit of the non-stationary model was introduced in figure 3. The delta-normal-approach does not require additional calibration. As presented in figure 5, the GARCH$(1,1)$ volatility calibration on 250 return points deviates slightly from the full sample adaption. In contrast the EGARCH$(1,1)$

[36]Testing the whole residual series $\{\hat{\epsilon}_t\}_t$ for normality with Jarque-Bera fails (on a 5% level of significance), indicating a leptokurtic distribution. We choose Student-t, as it is broadly used (next to the Gaussian) in econometric literature and as the innovation assumption is heavy-tailed. From the empirical kurtosis K_{ϵ_t} we assess the df as $n = \frac{4K-6}{K-3}$.

[37]Moreover, GARCH(p,q)- and EGARCH(p,q)-adaptions of higher order have very similar volatility graphs for that example. Hence, the GARCH$(1,1)$ and EGARCH$(1,1)$ processes are adequate benchmarks for ARCH-type models.

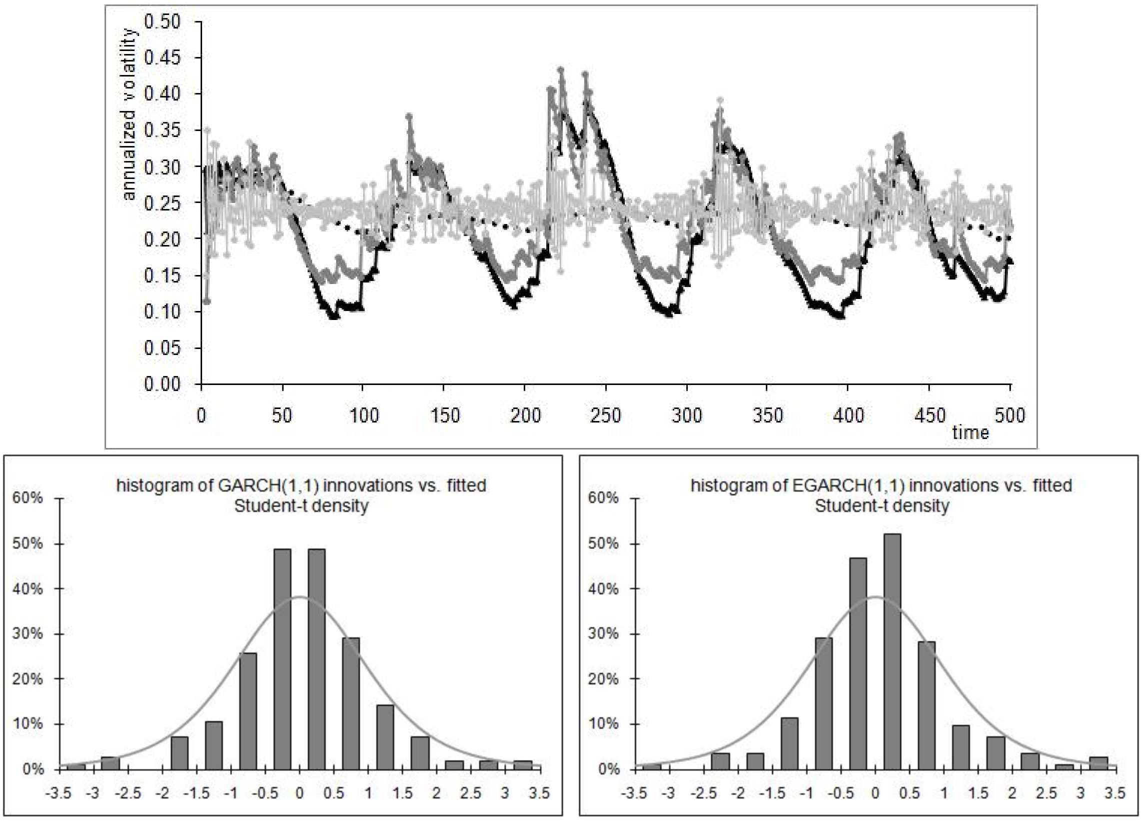


Figure 5: Top: Sd $\bar{\sigma}_t$ (dotted, black; 250-day), NWE $\hat{\sigma}_1(t)$ (triangles, black; $h^{\mathrm{CV}_{(1)}} = 36$), GARCH$(1,1)$ ς_{t+1} (bullets, dark grey; $\alpha_0 = 6.84 \cdot 10^{-6}, \alpha_1 = 0.1090, \beta_1 = 0.8725$) and EGARCH$(1,1)$ forecast $\breve{\varsigma}_{t+1}$ (bullets, light grey; $\alpha_0 = -14.7082, \alpha_1 = -0.0249, \gamma_1 = 0.1706, \beta_1 = -0.7570$), based on the in-sample calibration of process (29). Bottom: Histogram and Student-t fit of GARCH innovations (left; 5.43 df) and EGARCH innovations (right; 5.96 df). For the innovation fit of the non-stationary model see figure 3.

parametrization deviates strongly, presenting a peculiar noise behavior. The histograms of GARCH and EGARCH innovations relative to their Student-t approximation (5.43 df and 5.96 df) are displayed supplementary. Both t-densities have a lower, wider peak around the mean and thinner tails.

The four competitive implementations are used out-of-sample to assess quantiles $\mathrm{VaR}_{\cdot}^{1-\alpha,1\mathrm{d}}(t)$ of the next day's return and are evaluated by the number of shortfalls X_{t+1} lower or equal that limit. Figure 6 outlines the exceedances and Kupiec test results at a 5% level of significance. As mentioned before, the non-stationary model is approved for all confidence levels $1 - \alpha \geq 80\%$. The parametric VaR, based on a 250-day sd, seems to benefit from a higher average volatility as less exceedances occur for most levels and their number is closer to the expectation for 95% to 99.5% levels. But this VaR-model is too conservative and rejected on weaker levels, where certain exceedances are required.[38] Even more conservative are the t-GARCH$(1,1)$ and t-EGARCH$(1,1)$ forecasts for risk levels till 99%. The GARCH approximation is rejected five times since significantly less exceedances were observed. Despite good extreme quantile forecasts, the complete return distribution is misspecified. The EGARCH implementation is rejected for risk levels up to 95%. We expected even more misbehavior from the frantic volatility structure, but overall a higher average volatility and the innovation fit enable still applicable forecasts.

Concluding, the simulation experiment proved that the non-stationary regression model is able to capture simulated return dynamics and to provide satisfactory distributional forecasts as well. For the executed example and in terms of the Kupiec backtesting, this approach works best in the comparison against traditional risk models. It significantly outperforms the parametric VaR as well as famous ARCH-specifications.

[38] Based on a window adjusted to the bandwidth of the one-sided NWE (36 days), the parametric VaR-model passes the weaker confidence levels but fails some higher ones, e.g. with four exceedances at the 99.5% confidence level.

VaR confidence level (as 1-α)	number of exceedances, model rejection				
	expectation	non-stationary model	parametric VaR	t-GARCH(1,1) model	t-EGARCH(1,1) model
80.00%	49.8	45	32	33	24
90.00%	24.9	29	15	12	12
95.00%	12.5	16	10	4	2
98.00%	5.0	6	4	1	2
99.00%	2.5	5	2	0	1
99.50%	1.2	3	1	0	0
99.90%	0.2	1	1	0	0
99.95%	0.1	0	1	0	0

Figure 6: Kupiec test results of the non-stationary model, the parametric VaR with a 250-day sd, the t-GARCH$(1,1)$-model and the t-EGARCH$(1,1)$-model on several confidence levels regarding the forecasting of price process (29). Rejections on a 5% level of significance are grey highlighted.

5 Empirical study

In the previous paragraphs we got the theoretical confirmation to apply the non-stationary model (2) to financial time series. Now we adopt the benchmark conception for equity-, interest rate-, credit spread- and currency exposures, which was introduced in section 2 and completely described in table 1. Our study implements the approach univariate to the daily series of log-returns (equity indices and exchange rates) or diff-returns (interest rates and credit spreads; in bp) and evaluates the model fit.

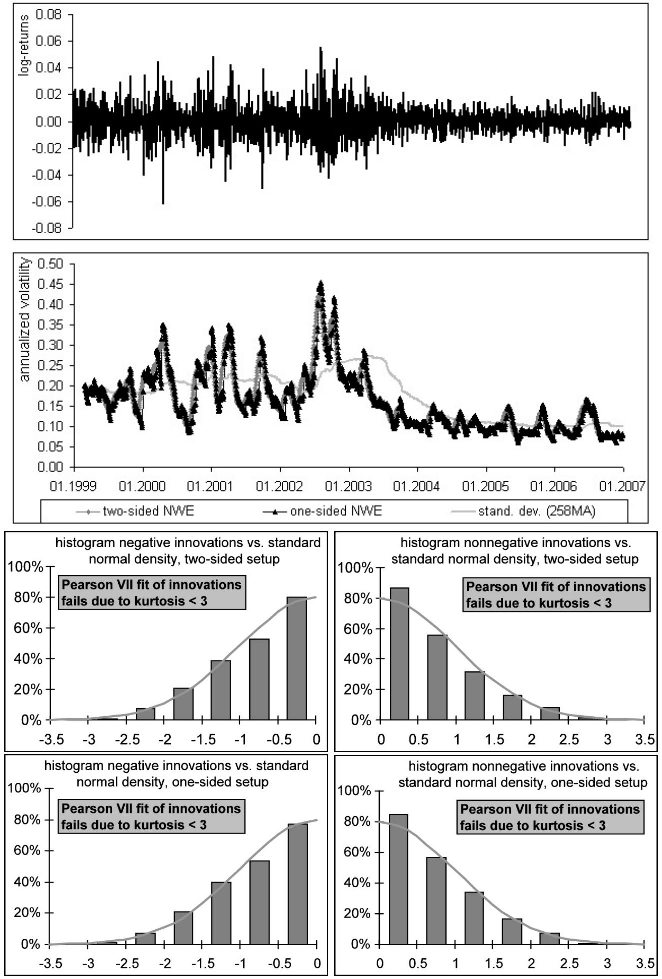
5.1 Nonparametric volatility estimation and residual fitting

Our special focus is set on the representative examples *EquNA* (MSCI North America index), *RateUSD* (5-year USD swap rate), *CredSta* (a global government to swap spread) and *CurrUSD* (EUR/USD exchange rate) of the four exposure types. The non-stationary modelling is illustrated in figures 7 till 10. After demeaning the financial returns (top graph), the volatility of the series is estimated nonparametrically. The symmetric, two-sided NWE (7) and the historical, one-sided NWE (8) are presented over time (middle graph).[39] The bottom graphs present the fit of the asymmetric Pearson VII distribution to the residuals, estimated as ratio of demeaned returns and volatility estimators. Regarding the data input we distinguish between the two-sided and the one-sided approach. The tables in each figure report on the corresponding Pearson VII moment estimators and the optimal bandwidths from cross validation.

The *EquNA* example reflects the development of one of the major stock markets in the period from 1999 to 2006, with a negative highlight in the burst of the 'I.T. bubble' in the early new millennium. First significant peaks in stock returns appear in March 2000, markets were deflating with full speed in the years 2001 and 2002. The nonparametric volatility estimates immediately react on sequences of extreme log-returns, reaching volatility levels that are more than the double of the long-term average. The one-sided NWE has a certain delay to the both-sided equivalent, but it similarly detects phases of high and low volatility. Contrary is the empirical sd, where it takes long until extreme changes get an impact on the average of 258 centred squared returns, and after the crisis it declines slowly while the market volatility was on a low level from summer 2003. The peaked nonparametric volatility graphs are consequence of quite small optimal bandwidths ($h^{\mathrm{CV}} = 24$, $h^{\mathrm{CV}(1)} = 30$). The volatility estimators alone catch the market dynamics excellent and the residuals are not as heavy tailed as assumed. Indeed the Pearson VII fit of innovations fails, since the kurtosis of realized innovations is lower than 3. In figure 7 alternatively a standard Gaussian density is compared to the histogram of innovations, that sufficiently approximates here and might be adequate in combination with 'heavy-tailed' volatility estimates.

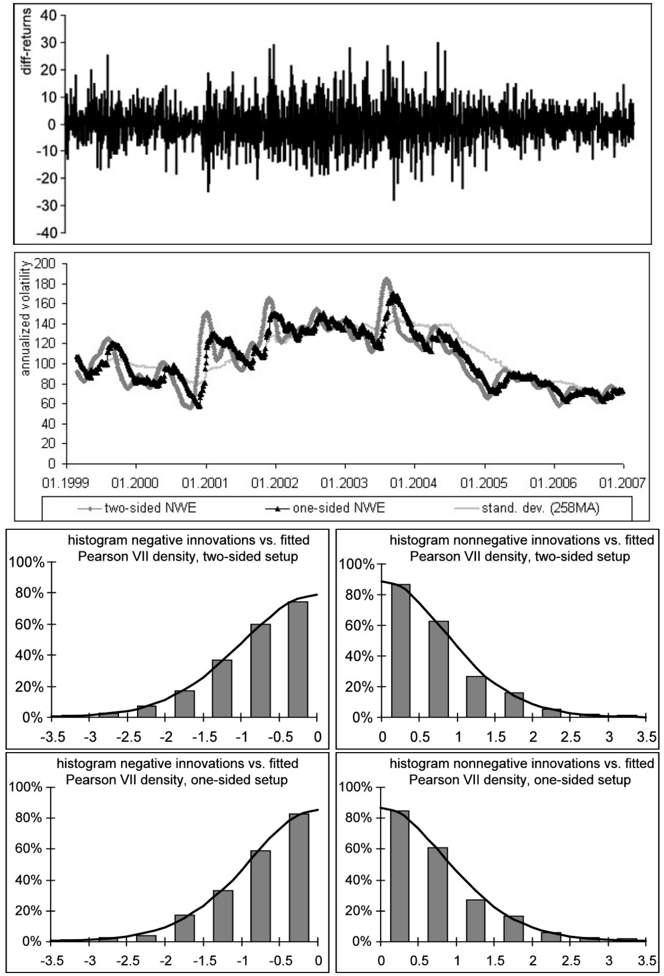
For the following examples the Pearson VII innovation fit works and we can present the full non-stationary implementation (2). *RateUSD* exhibits three phases (figure 8), with low volatilities in the first and last two years. Shocks in interest rates accompany the financial and economic crisis from 2001 to 2004, where the 5-year swaps declined from

[39] The volatility series starts at March 1, 1999 to initialize the estimators with two months of past returns. That reduces boundary effects on the left interval end, but does not eliminate them completely if the bandwidth is greater than 40 days. Moreover, boundary corruptions occur at the end of the series for the two-sided estimates. The average trading days p.a. in the 8-year financial series are 258, used for annualization.



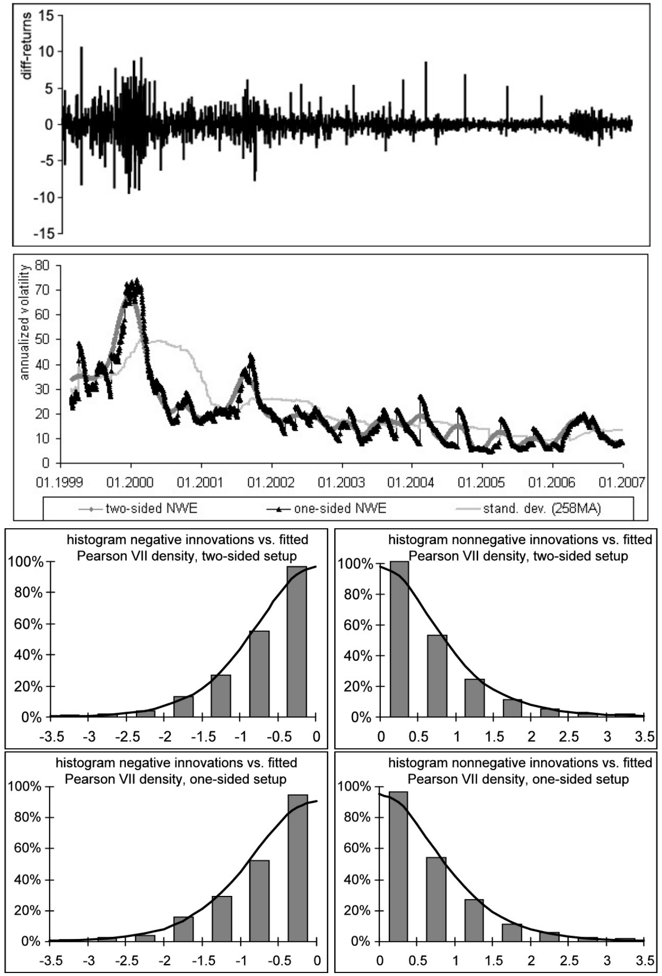
setup	BW h^{CV}	m_-	c_-	m_+	c_+
two-sided	24	$n.a.$	$n.a.$	$n.a.$	$n.a.$
one-sided	30	$n.a.$	$n.a.$	$n.a.$	$n.a.$

Figure 7: Top: Daily log-returns of *EquNA* from 1999 to 2006. Middle: Corresponding nonparametric volatility estimators in model (2) compared to standard deviation. Bottom: As left and right Pearson VII fits of innovations fail for the two-sided and one-sided approach, the Gaussian density is overlaid to the histograms alternatively.



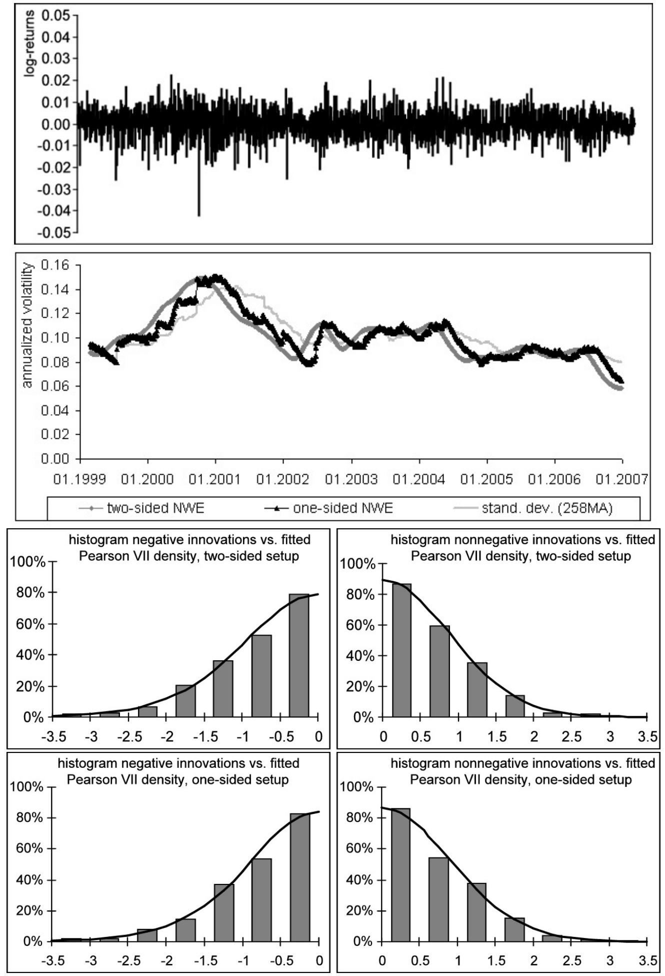
setup	BW h^{CV}	m_-	c_-	m_+	c_+
two-sided	41	15.2753	5.4369	7.8298	3.3805
one-sided	120	7.1818	3.3445	6.0330	2.9925

Figure 8: Top: Daily diff-returns of *RateUSD* from 1999 to 2006. Middle: Corresponding nonparametric volatility estimators in model (2) compared to standard deviation. Bottom: Left and right Pearson VII fit $f^{\mathrm{VII}(1)}_{m.,c.}$ of innovations for the two-sided and one-sided approach; the loss tail for a fixed income portfolio is the right side.



setup	BW h^{CV}	m_-	c_-	m_+	c_+
two-sided	63	4.7266	2.3271	2.8040	1.6532
one-sided	51	5.8272	2.8017	3.4236	1.9443

Figure 9: Top: Daily diff-returns of *CredSta* from 1999 to 2006. Middle: Corresponding nonparametric volatility estimators in model (2) compared to standard deviation. Bottom: Left and right Pearson VII fit $f_{m.,c.}^{\mathrm{VII}(1)}$ of innovations for the two-sided and one-sided approach; again the loss tail for a credit portfolio is the right side.



setup	BW h^{CV}	m_-	c_-	m_+	c_+
two-sided	100	6.2818	3.2537	27.3716	6.5059
one-sided	178	5.5069	2.9202	20.1207	5.7119

Figure 10: Top: Daily log-returns of *CurrUSD* from 1999 to 2006. Middle: Corresponding nonparametric volatility estimators in model (2) compared to standard deviation. Bottom: Left and right Pearson VII fit $f_{m.,c.}^{\mathrm{VII}(1)}$ of innovations for the two-sided and one-sided approach; the loss tail of a currency portfolio is the left.

more than 7% to a low of 2.5% p.a., as a consequence of the FED policies to reanimate US economy with low funding rates. Those shocks increased the annualized volatility from a level of 60bp to about 180bp. Here, the optimal one-sided bandwidth was adjusted manually to $h^{\text{CV}_{(1)}} = 120$, since the CV-function was almost constant for bandwidths greater than 100 with minima in the ranges $[115, 145]$ and $[185, 200]$. The Pearson VII fit of innovations reflects the asymmetry of the return distribution, where the right tail is heavier, as to be seen in lower Student's df $2m - 1$.

The *CredSta* series (figure 9) fluctuates strongly from summer 1999 to spring 2000, decreasing from a swap spread of -35bp to -65bp. The NWEs jump to a level of 74bp annualized volatility, which is more than three times the sample average. A second maximum is arrived in autumn 2001, where a run into the save haven of treasuries after the 9/11 terrorist attacks temporarily dropped the rates. Single peaks in the years 2004 and 2005 cause a sawtooth structure in the one-sided estimator, i.e. a rapid increase and a quite smooth decline. The Pearson VII fit of innovations works well, the spread changes turn out to be very heavy tailed, expressed by low optimal values of $m_.$ and $c_.$. The return distribution is leptokurtic and right skewed, i.e. extreme credit losses due to sudden spread expansions are discovered with the non-stationary approach.

The last example is the exchange rate EUR to USD in figure 10. For the two-sided volatility estimation the optimal bandwidth has to be manually adjusted ($h^{\text{CV}} = 100$) with respect to other horizons, other currencies and the one-sided equivalent. The annualized volatilities evolve most time in a range from 8% to 15%, clusters of higher returns are detected by the NWEs during the years 2000 and 2001. The Pearson VII approximation of the residuals leads to an asymmetric distribution that is heavier on the left side (loss tail of an European investor in USD).

We executed this analysis to all 30 benchmark return series from 1999 to 2006. Figure 11 reports on the optimal bandwidths for nonparametric volatility estimation and the Pearson VII parametrization (if existent) of the estimated residuals. The first rows of the table show the two-sided implementation, the parametrization based on historical data is shown below. As highlighted grey in the tableau, some bandwidth optimizations fail due to a plane or slowly declining CV-function, that inhibits detecting a global minimum. Generally, bandwidths were restricted to 200 days for reasons of heteroscedasticity and boundary effects, especially for the subsequent smaller samples. Manual bandwidth adjustments have to be conducted to the minor of series. We derive these optima by studying the CV-function, considering similar benchmarks, other time horizons or the opposite (one-/ two-sided) appropriate optima.

The Pearson VII fit of innovations fails several times, since their kurtosis is lower than 3. In some cases the approximation is successful at least in one tail. We observe slightly more successes in the historical approach. It is noticeable that the Pearson VII method falls through many times when the volatility estimates are based on small (CV-optimal) bandwidths. We developed the following explanation: The bigger the bandwidth in nonparametric curve estimation the smoother is the volatility estimator and the more heavy-tailed are the innovations. Vice versa, a perfectly calibrated NWE with a small bandwidth may imply an innovation distribution that is weaker than Gaussian and the Pearson VII fit fails. On the other hand, oversmoothed volatility estimates on the same series could produce heavier tailed innovations and the complete non-stationary framework (2) holds. Hence, there might be a trade-off between the quality in volatility estimation and a successful fit of innovations.[40]

We continued the empirical study as we analyzed the non-stationary modelling (2) with respect to three time setups: 1. long-term horizon 1999 to 2006, 2. time horizon 1999 to 2000, and 3. time horizon 2005 to 2006. The complete results are presented in Gürtler and Rauh (2009) and summarized as follows: Partially, significant differences are observed in nonparametric volatility estimation between the periods of time. For instance, the respective bandwidths of equity indices and credit spreads are still close to each other between 1999-2006 and 2005-2006, while the horizon 1999-2000 has very different outcomes. Probably, extremal returns during the financial crises after 2000 had a strong influence on the automatized parameter choice. Different overall information settings can cause deviating volatility estimates (or total model calibrations) in overlapping periods. Alternatively, time-varying bandwidths, that are reestimated with a certain frequency, could be reconsidered.

The smoothness of the volatility estimates affects the magnitude of estimated innovations and their distributional fitting. Hence, the same benchmarks on different horizons entail different Pearson VII approximations of innovations. Several times the asymmetric Pearson VII fit of innovations fails.[41] It holds in general, that the smoother the volatility estimate the more heavy-tailed are the innovations and the smaller are its parameters m_+, c_+, m_-, c_-.

[40] For instance, Drees and Starica (2002) choose in their S&P 500 example a bandwidth manually, that is significantly larger than the CV-optimum we derived for that series. In the next step they observed heavy-tailed innovations and fit them asymmetrically by Pearson VII. But as a consequence of our improved NWEs, a Pearson VII fit is impossible for positive innovations since those are no more heavy tailed.

[41] The Pearson VII method works for more than 60% of the 30 benchmark examples regarding all considered cases of horizons, two-sided and one-sided implementations and separate tail approximations. Manual bandwidth adjustments after CV were necessary in about 15% of all cases.

	Parameters of complete implementation non-stationary model									
Two-sided setup	**EquEUR**	**EquexEUR**	**EquNA**	**EquAsP**	**EquEM**	**RateEUR**	**RateUSD**	**RateJPY**	**CredSta**	**CredSwa**
=> NWE volatility:										
CV-optimal BW	26	19	24	35	22	34	41	23	63	75
=> innovation fitting:										
m-	*n.a.*	*n.a.*	*n.a.*	9.5314	*n.a.*	26.4848	15.2753	20.1651	4.7266	7.3859
c-	*n.a.*	*n.a.*	*n.a.*	4.2113	*n.a.*	6.6996	5.4369	5.2953	2.3271	3.3564
m+	*n.a.*	*n.a.*	*n.a.*	*n.a.*	15.7506	6.6866	7.8298	*n.a.*	2.8040	5.7025
c+	*n.a.*	*n.a.*	*n.a.*	*n.a.*	4.4192	3.2924	3.3805	*n.a.*	1.6532	2.8941
One-sided setup										
=> NWE volatility:										
CV-optimal BW	24	18	30	38	32	106	120	25	51	117
=> innovation fitting:										
m-	*n.a.*	*n.a.*	*n.a.*	11.0480	65.8141	12.3610	7.1818	*n.a.*	5.8272	7.3385
c-	*n.a.*	*n.a.*	*n.a.*	4.4531	11.7928	4.4445	3.3445	*n.a.*	2.8017	3.3763
m+	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	5.2567	6.0330	53.4319	3.4236	5.4114
c+	*n.a.*	*n.a.*	*n.a.*	*n.a.*	*n.a.*	2.8549	2.9925	10.2646	1.9443	2.8190

Two-sided setup	**CredAAA**	**CredAA**	**CredA**	**CredBBB**	**CredEM**	**CredHY**	**CurrEUR**	**CurrGBP**	**CurrCHF**	**CurrSEK**
=> NWE volatility:										
CV-optimal BW	80	25	28	83	5	7		24	16	19
=> innovation fitting:										
m-	6.3178	*n.a.*	8.4697	3.3385	*n.a.*	*n.a.*		*n.a.*	*n.a.*	*n.a.*
c-	3.1625	*n.a.*	3.6754	1.9523	*n.a.*	*n.a.*		*n.a.*	*n.a.*	*n.a.*
m+	4.0434	8.4898	25.7181	3.7359	*n.a.*	*n.a.*		*n.a.*	256.4915	*n.a.*
c+	2.1762	3.6155	6.8492	2.0024	*n.a.*	*n.a.*		*n.a.*	22.2233	*n.a.*
One-sided setup										
=> NWE volatility:										
CV-optimal BW	120	33	40	141	14	103		101	21	45
=> innovation fitting:										
m-	5.9376	*n.a.*	14.3047	3.3855	*n.a.*	7.1959		5.4396	*n.a.*	*n.a.*
c-	3.0460	*n.a.*	5.1000	1.8276	*n.a.*	3.2015		2.9060	*n.a.*	*n.a.*
m+	3.9626	9.6422	10.6654	3.5661	*n.a.*	4.4542		62.3075	*n.a.*	*n.a.*
c+	2.1004	3.8218	4.1354	2.0555	*n.a.*	2.5230		10.3192	*n.a.*	*n.a.*

Two-sided setup	**CurrDKK**	**CurrNOK**	**CurrUSD**	**CurrCAD**	**CurrJPY**	**CurrAUD**	**CurrNZD**	**CurrSGD**	**CurrHKD**	**CurrEM**
=> NWE volatility:										
CV-optimal BW	30	41	100	97	55	45	55	126	106	17
=> innovation fitting:										
m-	8.5323	7.6567	6.2818	6.4734	8.2892	12.1496	7.8549	21.0223	29.9300	*n.a.*
c-	3.7541	3.7998	3.3527	3.4973	3.6349	5.2782	3.9188	6.6021	8.0164	*n.a.*
m+	4.5259	17.8446	27.3716	22.7077	6.6031	7.1995	7.7123	10.6278	17.7757	*n.a.*
c+	2.3732	5.0011	6.5059	5.7618	3.1892	2.8562	3.0982	4.0050	5.3440	*n.a.*
One-sided setup										
=> NWE volatility:										
CV-optimal BW	52	47	178	88	59	163	167	135	110	17
=> innovation fitting:										
m-	4.9225	7.1637	5.5069	6.2964	11.3313	5.0402	5.2303	16.3051	25.2519	*n.a.*
c-	2.5754	3.4522	2.9202	3.1971	4.3151	2.7502	2.8721	5.4866	7.0335	*n.a.*
m+	5.1287	62.8801	20.1207	*n.a.*	8.6155	8.1105	8.5277	12.6410	15.4323	*n.a.*
c+	2.7431	10.5449	5.7119	*n.a.*	3.7398	3.2905	3.4501	4.5491	5.0632	*n.a.*

Figure 11: Parameters for the univariate non-stationary modelling (2) of daily benchmark returns from 1999 to 2006, with reference to the two-sided and the one-sided implementation. Grey highlighted elements designate manually adjusted bandwidths after CV. Cells with entry '*n.a.*' instead of Pearson VII parameters identify series where the innovation fitting failed.

5.2 Forecasting experiment

We proceed with a forecasting analysis, similar to the setup in section 4.2. Again, the sample series is divided into 'in-sample' for model calibration and 'out-of-sample' for forecasting and model evaluation. We examine the daily price series of the MSCI North America from 1999 to 2002. The first two years are treated as in-sample: On its demeaned log-returns $\widetilde{R}_1, \ldots, \widetilde{R}_{516}$ the bandwidth $h^{\mathrm{CV}_{(1)}} = 63$ days is found to be optimal for nonparametric volatility estimation. After determination of one-sided NWEs $\hat{\sigma}_{(1)}(1), \ldots, \hat{\sigma}_{(1)}(516)$ and return residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_{516}$, the asymmetric Pearson VII density $f^{\mathrm{VII}}_{m_+,c_+,m_-,c_-}$ with $m_+ = 9.8325$, $c_+ = 3.9547$, $m_- = 7.1698$, $c_- = 3.4640$ fits the random innovations best. Accordingly we implement the regression approach (2) and model the return distribution of X_{256}, that is our prediction for the distribution of X_{257}. Figure 12 represents the in-sample calibration (grey), and extends the return and volatility series out-of-sample (black). Although the optimal *EquNA* bandwidth differs in the short-term period from the longer optimum, the one-sided NWE is able to capture the stock market shocks during 2001 and 2002 visually well. A sequence of extreme log-returns increases the volatility estimator to a maximum of 39% p.a. in August 2002. The Pearson VII distribution is leptokurtic and skewed to the left, i.e. extreme losses are more probable than extreme gains.

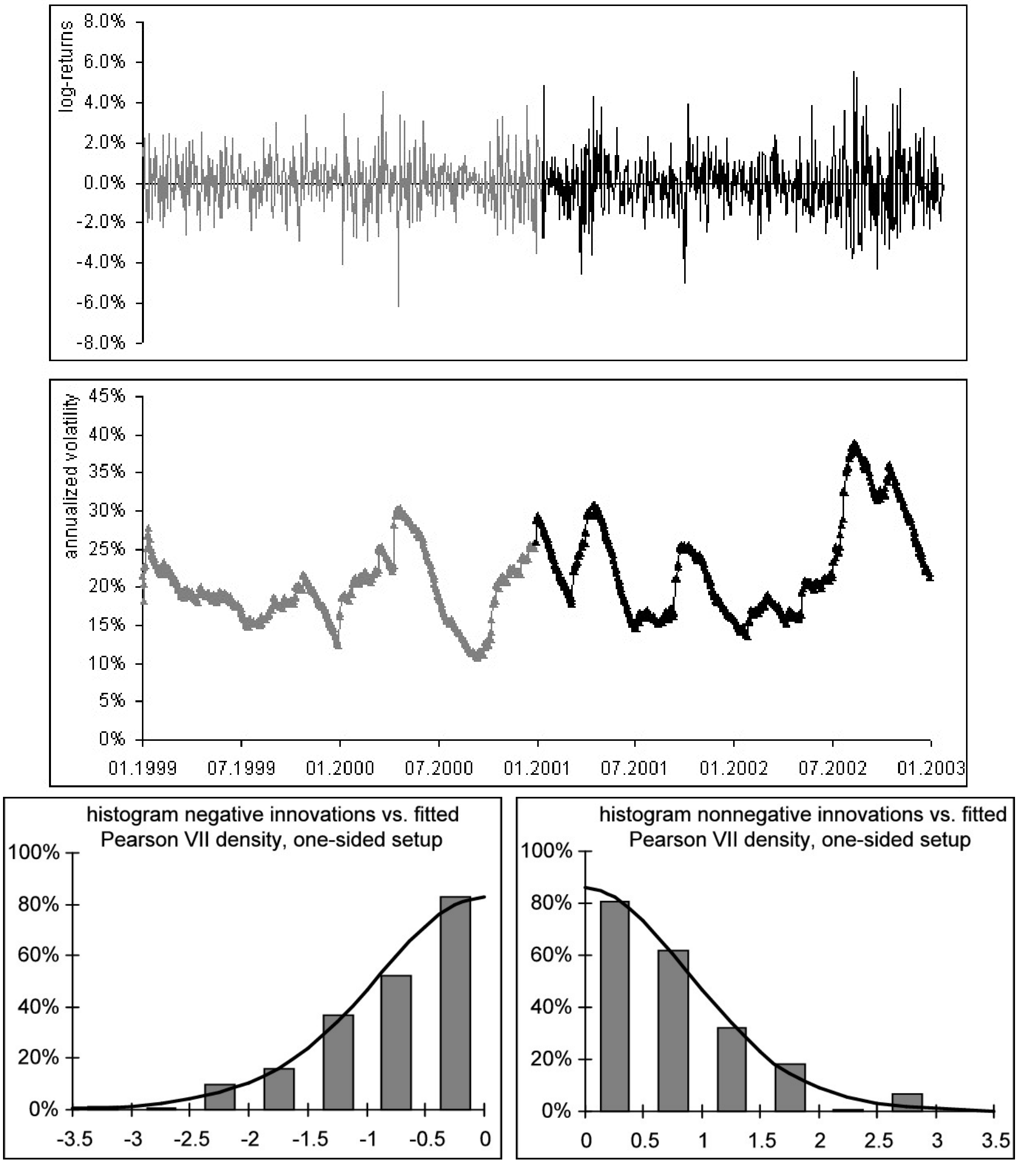


Figure 12: Non-stationary model calibration for the *EquNA* series from 1999 to 2000 (grey, 'in-sample') and model application to the years 2001 and 2002 (black). Top: daily log-returns; middle: nonparametric one-sided volatility estimation with $h^{\mathrm{CV}_{(1)}} = 63$; bottom: left and right Pearson VII fit of innovations with density $f^{\mathrm{VII}}_{9.83,3.95,7.17,3.46}$.

In terms of 1-day forecasting, we develop the return series and their volatility estimates (with fixed bandwidth) in the out-of-sample part gradually. We assess the quality of the regression model (2) with Kupiec backtesting, as we compare forecasted return distributions or certain quantiles, respectively, to the realized returns X_{t+1}. We focus on the loss tail and deduce at each point of time $t \geq 516$ a relative Value at Risk $\mathrm{VaR}^{1-\alpha,1\mathrm{d}}(t)$, applying formula (31). The number of shortfalls, where $X_{t+1} \leq \mathrm{VaR}^{1-\alpha,1\mathrm{d}}(t)$, is evaluated with the test statistic $\mathrm{LR}_{n,p}(N)$ from (30). Regarding the 513 observations and a 99% confidence level of maximum losses the VaR forecast was exceeded 5 times, that is very close to the expected number of 5.1 shortfalls. The test statistic is approximately zero ($\mathrm{LR}_{513,1\%}(5) = 0.003$) and the risk measure is accepted by the Kupiec test based on a 5% level of significance. Furthermore, acceptance was derived for all confidence levels greater-than-or-equal 80%, which is documented at the end of next section (figure 15). The non-stationary implementation of the empirical series of *EquNA* returns works excellent.

5.3 Comparison to ARCH-models

We finish again with a performance comparison of the non-stationary regression model to the delta-normal-model, the t-GARCH$(1,1)$- and t-EGARCH$(1,1)$ model, as introduced in section 4.3. The last two approaches are fitted to the empirical return series with the software EViews. For the beginning we oppose the variance structures of the approaches with reference to the 8-year daily returns of *EquNA*, *RateUSD*, *CreditSta* and *CurrUSD*. Figure 13 displays their volatility estimates in time and the underlying parameters. It is noticeable that the nonparametric estimation is quite close to the GARCH$(1,1)$ and EGARCH$(1,1)$ volatility in most cases (except the *CreditSta* example).[42]

Especially for *EquNA*, the three graphs almost cover each other even in strong peaks, where the one-sided NWE leads in maxima and minima due to a small bandwidth. The EGARCH$(1,1)$ volatility seems to run ahead the others in some phases, but this is because of its asymmetric reaction to gains and losses.[43] The empirical sd, based on a moving window of 258 returns, oversmoothes strongly. It lags behind and exhibits some 'ghosting features' of declines without any present incidence as one-year-old extremes are omitted. The dynamics of *RateUSD* and *CurrUSD* are less versatile, but approximated similarly by the approaches. Some peaks of the *CreditSta* volatility are dominated by EGARCH$(1,1)$ and its asymmetry becomes visible at the sawtooth structure of estimates. Here, GARCH$(1,1)$ is comparatively smooth and the empirical sd delivers a moving average that fails resembling recent dynamics. For most examples the Student-t fit of (E)GARCH residuals is visually satisfying with dfs that range from 9.2 to 10.8.[44] Furthermore, we found out that (E)GARCH parameters and the approximation of innovations vary similarly in time, as observed for the non-stationary model at the end of section 5.1.[45]

For the final judgement of dominances we require full model implementations and the backtesting of return forecasts. The forecasting experiment from the previous section is continued, the models are calibrated with in-sample prices of the *EquNA* series (1999-2000) and are evaluated via return distributions out-of-sample (2001-2002). Figure 14 presents the implementation of the alternatives. Similarly to the regression model, the GARCH$(1,1)$ volatility is smoother parametrized on the 2-year subsample. The EGARCH$(1,1)$ volatility is less persistent (since $\beta_1 < 0.9$) and seemingly undersmoothed, combined with its asymmetric behavior. In general the differences of the conditional heteroscedastic approaches are increased. The fit of realized GARCH and EGARCH innovations by the t-distribution (7.72 df and 9.59 df) is quite satisfying, apart from the median.

The four modelling approaches compete out-of-sample in assessing the $\mathrm{VaR}^{1-\alpha,1\mathrm{d}}_{\cdot}(t)$ of the next day's return X_{t+1} and are evaluated by the shortfall number with the Kupiec statistic. Their results are compared in figure 15. As mentioned before, the non-stationary approach is convincing with exceedances close to the expectations in most cases and no rejection of the shortfall probability at all considered confidence levels $1-\alpha$. The parametric VaR performs worst with four rejections (of nine tests), since too many exceedances were observed. The t-GARCH$(1,1)$ model fails one time because its loss forecast is too conservative and less shortfalls than expected were observed for risk levels greater or equal 95%. The t-EGARCH$(1,1)$ predictability is surprisingly good and no model rejections are discovered. Dependent on the confidence level its exceedances are partially more distant from the expectation than in the non-stationary case.

Concluding on the *EquNA* example, the performance of the non-stationary regression model is at least as good as the ARCH-type specifications. Even a weak dominance to t-GARCH$(1,1)$ was observed. The parametric VaR approach is clearly outperformed. Analogically, we executed forecasting studies for a plenty of benchmark indices and time horizons with qualitatively similar results. For instance, a survey on daily exchanges rates EUR/USD from January 1999 to March 2008, with models calibrated within the first 6 years and applied to the next 15 months, gains a successful fit by the non-stationary model, that outperforms even the (E)GARCH implementations significantly.[46] Consequently, our regression model is an excellent alternative to traditional risk models for a broad field of financial time series. The approach is at least equipollent or case by case even dominant. With the methods provided in section 3 our model implementation is more straightforward and its adaption is better interpretable than ARCH-specifications.

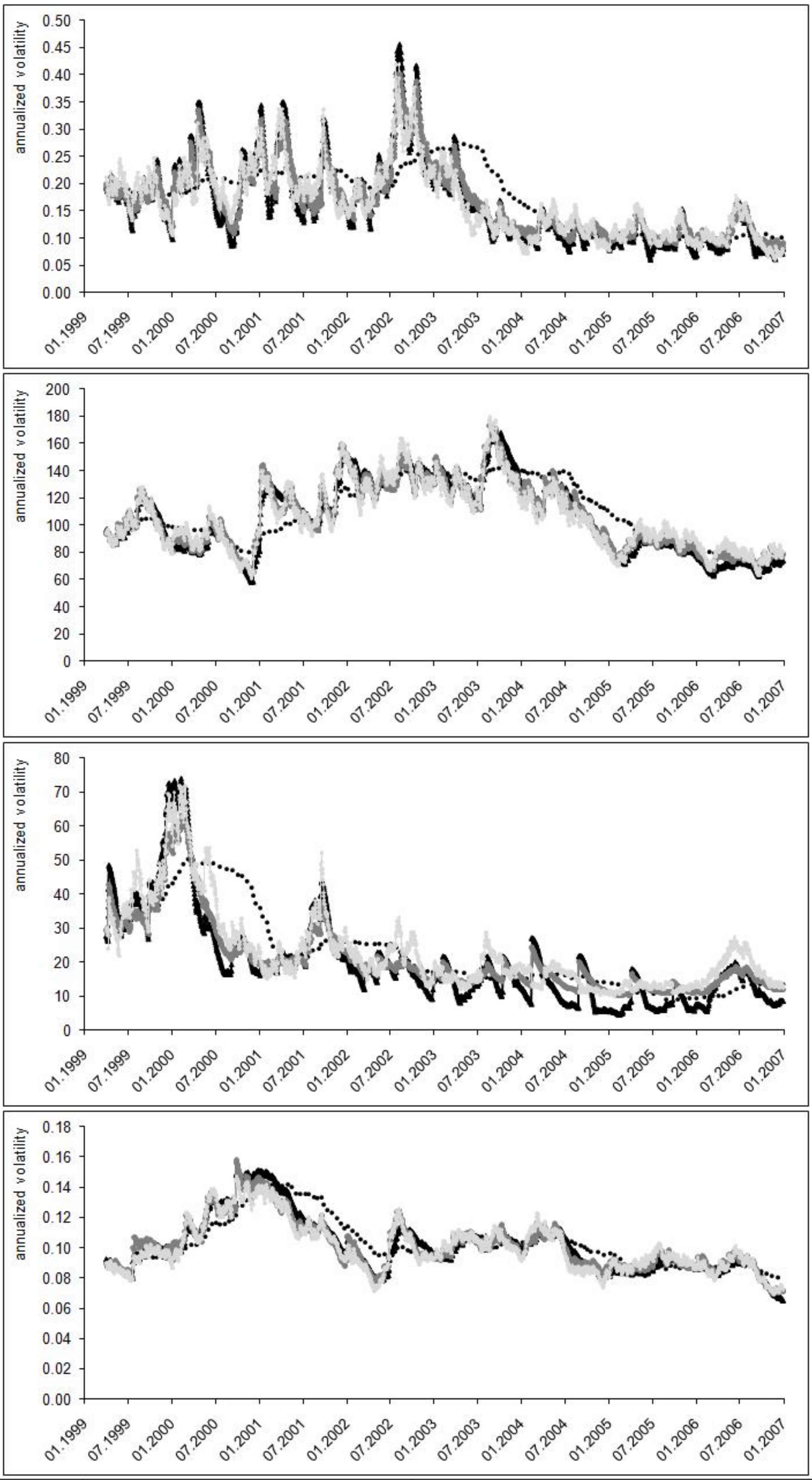
[42] Additionally their GARCH(p,q) and EGARCH(p,q) implementations of higher order have almost congruent volatility graphs to the base approaches. Hence, (E)GARCH models with $p=q=1$ are appropriate references.

[43] If after negative shocks a counter movement of large gains follows, the EGARCH$(1,1)$ estimate declines immediately, while GARCH$(1,1)$ and NWE are still increased by describing volatility classical based on squared returns.

[44] Again *CreditSta* is the outlier with a very heavy tailed innovation fit (about 4.5 df at t-fit in both cases).

[45] We repeated the comparative survey for the 2-year subsamples. Next to significant deviations in volatility estimates and innovation distributions, some faults of ARCH parametrizations were observed for the small samples.

[46] The empirical study of *CurrUSD* returns is available upon request. In terms of Kupiec tests the non-stationary approach was rejected once at the 80% level. Parametric VaR failed two times. t-GARCH$(1,1)$- and t-EGARCH$(1,1)$ modelling experienced too less exceedances for most risk levels, resulting in four respectively three rejections.



	GARCH(1,1)			EGARCH(1,1)			
Parameters	α_0	α_1	β_1	α_0	α_1	γ_1	β_1
EquNA	$4.58 \cdot 10^{-7}$	0.0542	0.9422	-0.1553	0.0647	-0.1006	0.9888
RateUSD	0.1789	0.0283	0.9677	-0.0312	0.0844	-0.0016	0.9910
CredSta	0.0086	0.0231	0.9705	-0.0363	0.0622	-0.1181	0.9933
CurrUSD	$1.57 \cdot 10^{-7}$	0.0199	0.9760	-0.1016	0.0538	0.0062	0.9941

Figure 13: Annualized volatility estimates for return series in the order of the table: Sd $\bar{\sigma}_t$ (dotted, black; 258-day), one-sided NWE $\hat{\sigma}_1(t)$ (triangles, black; BWs from figure 11), GARCH$(1,1)$ ς_{t+1} (bullets, dark grey) and EGARCH$(1,1)$ volatility forecast $\check{\varsigma}_{t+1}$ (bullets, light grey) with parametrization given below.

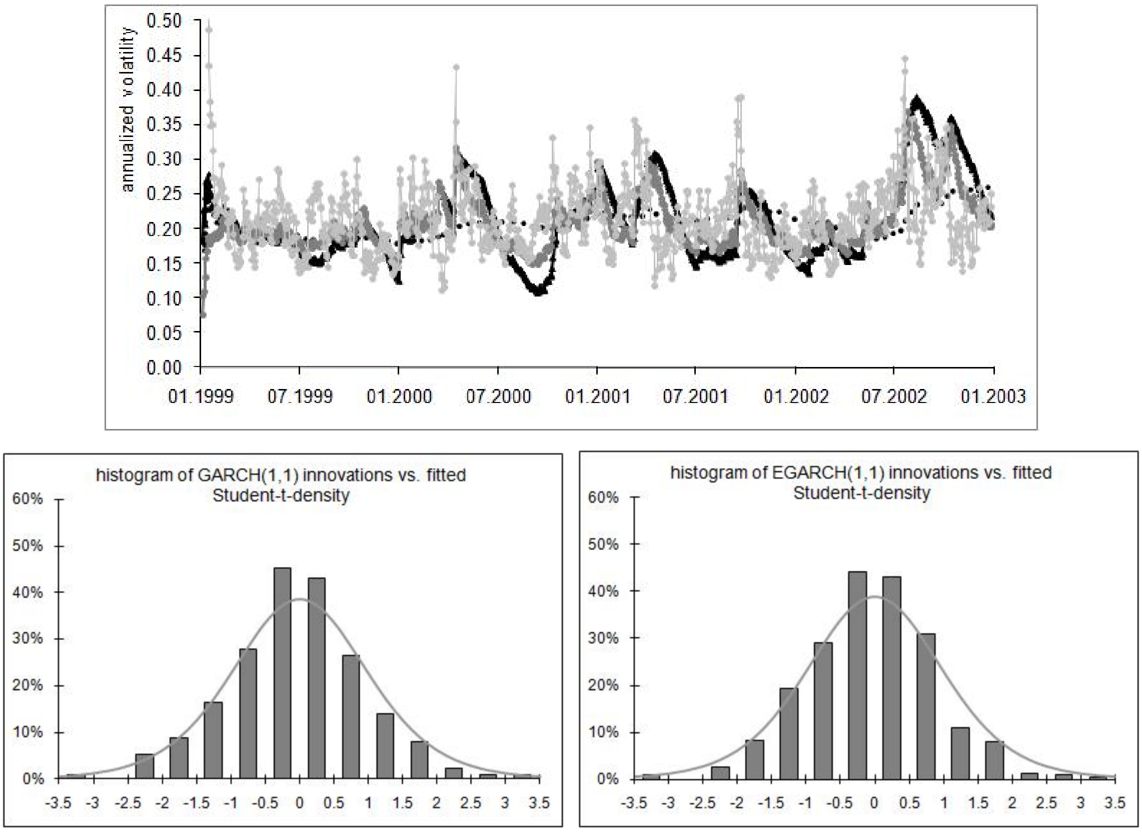


Figure 14: Top: Sd $\bar{\sigma}_t$ (dotted, black; 258-day), NWE $\hat{\sigma}_1(t)$ (triangles, black; $h^{\mathrm{CV}_{(1)}} = 63$), GARCH$(1,1)$ ς_{t+1} (bullets, dark grey; $\alpha_0 = 5.35 \cdot 10^{-6}, \alpha_1 = 0.0472, \beta_1 = 0.9192$) and EGARCH$(1,1)$ forecast $\check{\varsigma}_{t+1}$ (bullets, light grey; $\alpha_0 = -0.9567, \alpha_1 = 0.0687, \gamma_1 = -0.2153$, $\beta_1 = 0.8978$), based on the in-sample calibration of *EquNA* returns. Bottom: Histogram and Student-t fit of GARCH innovations (left; 7.72 df) and EGARCH innovations (right; 9.59 df). For the innovation fit of the non-stationary model see figure 12.

VaR confidence level (as 1-α)	number of exceedances, model rejection				
	expectation	non-stationary model	parametric VaR	t-GARCH(1,1) model	t-EGARCH(1,1) model
80.00%	102.6	111	114	111	114
90.00%	51.3	61	66	51	50
95.00%	25.7	26	33	18	25
98.00%	10.3	12	17	4	9
98.50%	7.7	8	15	3	5
99.00%	5.1	5	8	2	3
99.50%	2.6	3	7	2	1
99.90%	0.5	2	2	0	0
99.95%	0.3	1	2	0	0

Figure 15: Kupiec test results for the non-stationary model, the parametric VaR with a 258-day sd, the t-GARCH$(1,1)$-model and t-EGARCH$(1,1)$-model on several confidence levels regarding the forecasting of *EquNA* returns in 2001-2002. Rejections on a 5% level of significance are grey highlighted.

6 Summary and conclusion

In this article we examine a non-stationary, heteroscedastic model for the dynamics of financial returns. The motivation is founded in the shortfalls of the random walk hypothesis, that is still preserved for many practical applications (e.g. parametric VaR), and criticism to nonlinear time series models of the ARCH-type. The latter are parametrized as stationary processes and focus on the time-dependence of second moments. But recent research suggests that central features as the LRD might be caused by structural breaks of the unconditioned variance. Our non-stationary approach is based on the ideas of Herzel et al. (2005) and Drees and Starica (2002).

In a regression framework the volatility is assumed to be exogenous to the return process, since no explanatory variables are at hand. Instead, the evolution of prices contains the complex market conditions. The vectors of returns are assumed to be independent while having an unconditional covariance structure that changes smoothly in time. By dint of nonparametric regression on equidistant design points we estimate the unconditional (co-)variance directly on centred returns (Nadaraya-Watson estimators, NWE). Further effort is spent to an accurate modelling of the random residuals. An asymmetric version of the Pearson type VII distribution, that enables heavy tails, is fitted to the estimated innovations. We specialize on the univariate description. Regarding the included information we distinguish between a two-sided approach and a one-sided implementation, that is based only on past data and applied for forecasting return dynamics.

After outlining the statistical properties of consistency and asymptotic normality for both the symmetric NWE and the one-sided NWE, based on the results of Gürtler et al. (2009), we deduce requirements for their kernels and bandwidths. The biweight kernel is established. Cross-validation is adopted for an automatized bandwidth selection. The task of fitting innovations is simplified by providing a method of moments for Pearson VII parameter estimation and via a connection to the Student-t distribution. This presentation of residuals facilitates a factor-based VaR calculation in terms of the regression model: The univariate $\mathrm{VaR}^{1-\alpha}(t)$ of an exposure $w(t)$ can be modelled as the product of $w(t)$ with a nonparametric estimated volatility $\hat{\sigma}_{(\cdot)}(t)$ and the Pearson VII innovation α-quantile $u_{m,c;\alpha^{(1)}} = \gamma\, t_{2m-1;\alpha^{(1)}}$ of its benchmark, adjusted by the return expectation $w(t)\bar{X}_t$. This idea is picked up in simulations and a broad empirical study for forecasting and model evaluation.

Our simulation study documents how the fit of NWEs 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 at all. The quality of 1-day return distribution forecasts or VaR estimates, respectively, are confirmed with Kupiec tests on shortfall rates of simulated returns. Moreover, our approach outperforms the delta-normal-model and even a t-GARCH$(1,1)$ and t-EGARCH$(1,1)$ implementation. An extensive empirical study concerning daily return series of equity indices, interest rates, credit spreads and exchange rates (30 benchmarks) completes our analysis. The nonparametric estimates detect clusters of market volatility and the asymmetric Pearson VII distribution fits heavy-tailed innovations well. The goodness of fit is exemplified on North American equity returns in terms of the Kupiec backtesting of return forecasts. The non-stationary approach is compared to the aforementioned alternatives with the outcome, that it is at least equipollent or occasionally significantly better. Nonparametric volatility estimates (one-sided NWEs) are often close to GARCH$(1,1)$ and EGARCH$(1,1)$ volatilities, the Pearson VII innovation distribution is superior to Student-t residuals. Consequently, the regression model is an intuitive and strong alternative to traditional risk models.

On the other hand we notice that the bandwidth selection via cross validation cannot be automatized at all: Rarely the method fails due to finding no accurate minimum in the CV-function. But we provide other quantitative criteria for that choice, so that we are able to avoid smoothing by eyes. Moreover, we observe a trade-off between volatility estimation and Pearson VII innovation fitting: The bigger the bandwidth in nonparametric curve estimation the smoother is the volatility estimate and the more heavy-tailed will be the innovations. Thus, a perfectly calibrated (small) bandwidth might cause that the Pearson VII fit of innovations fails, since the condition of a kurtosis greater than 3 is not satisfied. A failed Pearson VII fit could be compensated conservatively by a Gaussian distribution for innovations. Another observation is that the optimal parameters (bandwidths and Pearson VII coefficients) change through time. Hence, a time-varying setup, that is reestimated with a certain frequency, could be a task for further research.

In a nutshell, we think to have developed the following novelties: The non-stationary regression model is applied univariate to a variety of financial time series. It is probably used for the first time to approximate credit spreads and to model diff-returns. Based on the statistical theory of Gürtler et al. (2009) we derive consistent criteria and quantitative methods for nonparametric volatility estimation. Regarding the Pearson VII modelling of random innovations we employ a method of moments. By dint of a Student-t description of the innovation's distribution we derive a factor-based VaR presentation of the non-stationary approach, that is easy to implement in real-world practice. The modelling performance is successfully evaluated with Kupiec tests, and an outperformance to the delta-normal-model and some ARCH-type models is proved.

Beside the successful aggregate test of the model approximation, we do not explicitly verify all singular assumptions of the non-stationary return model. For that task we refer back to Drees and Starica (2002), who elaborately proved on a 12-year S&P 500 return series that estimated innovations are iid random variables. Moreover, in a comparison with Student's t-GARCH$(1,1)$ and GED-EGARCH$(1,1)$ the non-stationary model fitted this one data set significantly better and forecasted 1-, 20- and 40-days ahead (conditional) return distributions best. Herzel et al. (2005) showed on a tri-variate example (exchange rate EUR/USD, FTSE 100 index, 10-year US T-bond rate) that their congruent paradigm describes the joint dynamics of the risk factors well and delivers good distributional forecasts. They proved an outperformance against the $RiskMetrics^{TM}$ (JP Morgan) approach.[47] The careful modelling of the extremal behavior of innovations was identified as one factor of success, making their approach 'amenable for precise VaR calculations' (Herzel et al. (2005), chapter 8). This is one of the targets, we had on our own agenda. Last but not least, Mikosch and Starica (2003) extend the univariate, non-stationary framework by a time-varying expected return. With reference to a 50-year S&P 500 daily return series they give statistical evidence that the expected return and market price of risk vary significantly in time.

Having seen the advantages of our non-stationary regression model, we can imagine the following fields for future research: The main task will be to develop an adequate multivariate setup for a broad exposure conception. Instead of a direct multivariate approach, a risk aggregation via simulation could be fruitful, e.g. with Cholesky decomposition of correlations, that continues on our univariate factor-based implementation for VaR purposes. A full nonparametric setup is conceivable, where the (still restrictive) parametric approach for the distribution of innovations could be substituted with a nonparametric kernel density. Going along with the ideas of Mikosch and Starica (2003), the inclusion of a time-dependent expected yield may be a further step. Based on the 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 framework for portfolio optimization in terms of a tactical asset allocation.

References

Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure* 3(17):21-86.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy* 81:637-654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31:307-327.

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994). ARCH models. In: Engle, R. F., McFadden, D. L., eds. *Handbook of Econometrics*. Amsterdam: Elsevier B.V..

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* 52:115-127.

Chan, N. H. (2002). *Time Series: applications to finance*. New York: John Wiley & Sons.

Choi, P. and Nam, K. (2008). Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the S_U-normal distribution. *Journal of Empirical Finance* 15:41-63.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1:223-236.

De Haan, L., Jansen, D. W., Koedijk, K. G. and de Vries, C. G. (1994). Safety first portfolio selection, extreme value theory and long run asset risks. In: Galambos, J., Lechner, J., Simiu, E., eds. *Extreme value theory and its applications*. Dordrecht: Kluwer Academic Publishers.

[47] As investigated by Drees and Starica (2002) and Herzel et al. (2005), $RiskMetrics^{TM}$ is built on the same multiplicative setup as our non-stationary model, but with return expectation 0 and normal innovations. Exponential weighted moving average volatilities can be interpreted as NWE with exponential kernel: $K(x) = a^x \mathrm{I}_{[-1,1]}(x)$, $a = \lambda^m$ and $K_h(t) = h^{-1} K\left(\frac{t}{h}\right)$ with $h = m$; by default $\lambda = 0.94$ and $m = 74$ for daily returns. Hence it is a special case, violating some regularity conditions in volatility estimation and is less flexible in bandwidth selection.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics* 105:131-159.

Ding, Z., Engle, R. F. and Granger, C. W. J. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1:83-106.

Drees, H. and Starica, C. (2002). A simple non-stationary model for stock returns. Preprint, Saarland University and Chalmers University. Available online at: http://www.math.uni-sb.de/PREPRINTS/preprint69.pdf.

Engle, R. F. (1982). Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987-1007.

Engle, R. F. (2004). Risk and Volatility: Econometric Models and Financial Practice. *The American Economic Review* 94:405-420.

Engle, R. F. and Patton, A. J. (2004). What good is a volatility model?. *Quantitative Finance* 1:237-245.

Fama, E. (1965). The behavior of stock market prices. *Journal of Business* 38:34-105.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer-Verlag.

Gasser, T., Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association* 86:643-652.

Granger, C. W. J. and Starica, C. (2005). Non-stationarities in stock returns. *Review of Economics and Statistics* 87:503-522.

Gürtler, M., Kreiss, J.-P. and Rauh, R. (2009). A Non-stationary Approach for Financial Returns with Nonparametric Heteroscedasticity. IF Working Paper Series: IF31, University of Braunschweig.

Gürtler, M. and Rauh, R. (2009). Shortcomings of a Parametric VaR Approach and Nonparametric Improvements Based on a Non-stationary Return Series Model. IF Working Paper Series: IF32, University of Braunschweig.

Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64:307-333.

Herzel, S., Starica, C. and Tütüncü, R. (2005). A non-stationary paradigm for the dynamics of multivariate financial returns. In: Bertail, P., Doukhan, P., eds. *Dependence in Probability and Statistics*. New York: Springer-Verlag.

Hsieh, D. A. (1989). Testing for Nonlinear Dependence in Daily Foreign Exchange Rates. *The Journal of Business* 62:339-368.

Hu, W. and Kercheval, A. N. (2010). Portfolio optimization for student t and skewed t returns. *Quantitative Finance* 10:91-105.

Johnson, N. L. and Kotz, S. (1970). *Continuous univariate distributions - 1. Distributions in statistics*. New York: John Wiley & Sons.

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*. New York: McGraw-Hill Professional.

Kitagawa, G. and Nagahara, Y. (1999). A Non-Gaussian Stochastic Volatility Model. *The Journal of Computational Finance* 2:33-47.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3:73-84.

Lamoureux, C. G. and Lastrapes, W. D. (1990). Persistence in variance, structural changes and the GARCH model. *Journal of Business and Economic Statistics* 8:225-234.

Liu, C. Y. and He, J. (1991). A Variance-Ratio Test of Random Walks in Foreign Exchange Rates. *Journal of Finance* 46:773-785.

Lo, A. W. and MacKinlay, A. C. (1988). Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification test. *The Review of Financial Studies* 1:41-66.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* 36:394-419.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance* 7:77-91.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science* 4:141-183.

Mikosch, T. and Starica, C. (2003). Stock market risk-return inference. An unconditional non-parametric approach. Working Paper, University of Copenhagen and Chalmers University. Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=882820.

Mikosch, T. and Starica, C. (2004). Non-stationarities in financial time series, the long range dependence and IGARCH effects. *Review of Economics and Statistics* 86:378-390.

Müller, H.-G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics* 15:610-625.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59:347-370.

Samuelson, P. A. (1973). Mathematics of speculative prices. *SIAM Review* 15:1-42.

Starica, C. (2003). Is GARCH$(1, 1)$ as good a model as the Nobel prize accolades would imply?. Working Paper, Chalmers University. Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=637322.

Straumann, D. (2004). *Estimation in Conditionally Heteroscedastic Time Series Models*. Berlin: Springer-Verlag.

Taylor, S. J. (1986). *Modelling financial time series*. New York: John Wiley & Sons.