

Modelling Different Volatility Components in High-Frequency Financial Returns

Yuanhua Feng

Department of Mathematics and Statistics

University of Konstanz

D-78457 Konstanz, Germany

Email: Yuanhua.Feng@uni-konstanz.de

Abstract

This paper considers simultaneous modelling of seasonality, slowly changing unconditional variance and conditional heteroskedasticity in high-frequency financial returns. A new approach, called a seasonal SEMIGARCH model, is proposed to perform this by introducing multiplicative seasonal and trend components into the GARCH model. A data-driven semiparametric algorithm is developed for estimating the model. Asymptotic properties of the proposed estimators are investigated briefly. An approximate significance test of seasonality and the use of Monte Carlo confidence bounds for the trend are proposed. Practical performance of the proposal is investigated in detail using some German stock price returns. The approach proposed here provides a useful semiparametric extension of the GARCH model.

Keywords: High-frequency financial data, nonparametric regression, seasonality in volatility, semiparametric GARCH model, trend in volatility.

1 Introduction

Financial returns exhibit conditional heteroskedasticity (CH). Well known approaches for modelling the CH are the ARCH (autoregressive conditional heteroskedastic, Engle, 1982), GARCH (generalized ARCH, Bollerslev, 1986) models and their extensions. In spite of their conditional heteroskedastic property the ARCH and GARCH models are stationary with constant unconditional variance and are hence *time homoskedastic*. In recent years it is however realized that financial returns also exhibit *time heteroskedasticity* (TH) or

unconditional heteroskedasticity, that is the unconditional variance varies over time and hence the process is no longer stationary but at most local stationary.

Evidence of TH in financial time series was reported in the literature together with different approaches for modelling it. Mikosch and Stărică (1999) showed that the phenomenon $\hat{\alpha}_1 + \hat{\beta}_1 \approx 1$ by a fitted GARCH(1, 1) model indicates nonstationarity and proposed the use of a piecewise GARCH model. Beran and Ocker (2001) fitted SEMIFAR (semiparametric fractional autoregressive) models proposed by Beran (1999) to some volatility series and found significant trend in volatility. Härdle et al. (2001) introduced a time-inhomogeneous stochastic volatility model with time varying coefficients. A local time-homogeneous model with change points is proposed by Mercurio and Spokoiny (2002), where the volatility is assumed to be constant in an unknown local time interval. A semiparametric GARCH model with a slowly changing scale function (called SEMIGARCH) is proposed by Feng (2002) for simultaneously modelling the CH and TH.

The current paper considers modelling of volatility in high-frequency financial returns. Now the volatility also exhibits daily periodicity (see e.g. Dacorogna et al., 2001). For modelling different components in the volatility of high-frequency financial returns a new approach, called a seasonal SEMIGARCH model, is proposed by introducing an additional multiplicative seasonal component into the SEMIGARCH model, which extends the traditional component model of economic time series to the current context. A data-driven semiparametric algorithm is developed for estimating the model. Although the focus of this paper is on applications, necessary asymptotic properties of the proposed estimators are investigated briefly. An approximate significance test of seasonality and the use of Monte Carlo confidence bounds for the trend are proposed. Practical performance of the proposal is investigated in detail using some German stock price returns. It is shown that the proposal works well in practice. This new approach provides a useful semiparametric extension of the well known GARCH model.

The paper is organized as follows. Section 2 introduces the model and proposes the data-driven semiparametric algorithm. Asymptotic properties of the proposed estimators are investigated in Section 3. Section 4 describes the significance test of seasonality and the Monte Carlo confidence bounds for the trend. Applications and discussion on the practical performance of the proposal are given in Section 5. Final remarks in Section 6 conclude the paper. Proofs of results are put in the appendix.

2 The model and the estimation procedure

2.1 The model

Assume that the log-returns of a financial time series follow the model:

$$Y_i = \mu + V_0^{1/2} v^{1/2}(t_i) S_i^{1/2} h_i^{1/2} \eta_i, \quad (1)$$

where $t_i = i/n$ is the re-scaled time which guarantees the availability of consistent estimation, $V_0 > 0$ is a constant, $v(\cdot) > 0$ is a smooth, bounded function, $S_i > 0$ is a periodic function with period T and $\epsilon_i := h_i^{1/2} \eta_i$ is assumed to follow a GARCH(p, q) model with

$$h_i = \alpha_0 + \sum_{j=1}^p \alpha_j \epsilon_{i-j}^2 + \sum_{k=1}^q \beta_k h_{i-k} \quad (2)$$

(Bollerslev, 1986). $\sigma(t) = v^{1/2}(t)$ is called the scale (or volatility trend) function. Let $\theta = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)'$. It is assumed that $\alpha_0 > 0$, $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \geq 0$ and $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$, which guarantees the existence of a unique strictly stationary solution of (2). Without loss of generality let $T^{-1} \sum_{i=1}^T S_i = 1$, $\int_0^1 v(t) dt = 1$ and $\text{var}(\epsilon_i) = 1$. The last condition implies $\alpha_0 = 1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j$. An equivalent representation of model (1) and (2) is

$$Y_i = \mu + V_0^{1/2} v^{1/2}(t_i) S_i^{1/2} \epsilon_i. \quad (3)$$

The moment conditions $E(\epsilon_i^4) < \infty$ and $E(\epsilon_i^8) < \infty$ are required for the derivation of the asymptotic results and the practical implementation of a nonparametric estimator of $v(\cdot)$ respectively. Necessary and sufficient conditions which guarantee the existence of high order moments of a GARCH process may be found e.g. in Ling and Li (1997), Ling (1999) and Ling and McAleer (2002). See also Bollerslev (1986) for results in the special case of a GARCH(1, 1) model.

Model (1) and (2) is a semiparametric extension of the GARCH model, which provides a tool for simultaneous modelling of conditional heteroskedasticity (h_i), slow scale change ($v(\cdot)$) and seasonality (S_i) in high-frequency financial returns. V_0 quantifies the averaged unconditional variance of Y_i . And the total volatility of Y_i is hence $V_0^{1/2} v^{1/2}(t_i) S_i^{1/2} h_i^{1/2}$. Note that $V_0^{1/2}$ is the absolute level of the standard deviation. All other three components are defined relatively and move around the unit level. The introduction of the trend

function $v(\cdot)$ and the seasonal component S_i extends the applicability of the well known GARCH model to a wide class of nonstationary process. If the local variance of the return series changes over time and/or if there is a nonconstant periodic term in the unconditional variance, then the use of model (1) and (2) will lead to significantly theoretical and practical improvements.

Furthermore, define $Z_i = (Y_i - \mu)$, $X_i = Z_i/\sqrt{V_0}$, $R_i = Z_i/\sqrt{V_0 S_i}$ and $\xi_i = (\epsilon_i^2 - 1)$. Then we have

$$Z_i^2 = V_0 v(t_i) S_i + V_0 v(t_i) S_i \xi_i, \quad (4)$$

$$X_i^2 = v(t_i) S_i + v(t_i) S_i \xi_i. \quad (5)$$

and

$$R_i^2 = v(t_i) + v(t_i) \xi_i. \quad (6)$$

Model (6) is a nonparametric regression with dependent and heteroskedastic errors, which indicates that $v(\cdot)$ can be estimated using well known techniques in nonparametric regression (see e.g. Efromovich, 1999 and Feng, 2002).

The assumption that ϵ_i follow a GARCH model is made here for simplicity, which allows us to use well known theoretical results for investigating the properties of the proposed model and to estimate the conditional heteroskedasticity using an existing GARCH packet. This assumption is however not necessary and can be replaced by other parametric or nonparametric assumptions.

2.2 Estimation of the model

If $S_i \equiv 1$, equations (1) and (2) reduce to the SEMIGARCH model introduced by Feng (2002). Estimation of the SEMIGARCH model was investigated there in detail. The transformation from (4) to (6) shows that, if V_0 and S_i can be estimated properly without pre-estimation of $v(\cdot)$ and θ , then $v(\cdot)$ and θ can be estimated from the seasonal adjusted data following the proposal of Feng (2002). In this paper a semiparametric estimation procedure following this idea will be proposed.

Let y_1, y_2, \dots, y_n denote the observations. At first μ can be estimated by the sample mean \bar{y} . And one obtains the centralized observations $\hat{z}_i = y_i - \bar{y}$. An estimator of the

averaged variance V_0 is given by

$$\hat{V}_0 = n^{-1} \sum_{i=1}^n \hat{z}_i^2. \quad (7)$$

Now define $\hat{x}_i = \hat{z}_i / \sqrt{\hat{V}_0}$. Let $m = [n/T]$, where $[\cdot]$ denotes the integer part. The seasonal component can be estimated from \hat{x}_i^2 as follows

$$\hat{S}_i = m^{-1} \sum_{j=0}^{m-1} \hat{x}_{jT+i}^2, \quad i = 1, 2, \dots, T, \quad (8)$$

and $\hat{S}_i = \hat{S}_{i-kT}$ for $i > T$, where $k = [(i-1)/T]$. In the next section it will be shown that the effect of the unknown volatility trend on \hat{S}_i is negligible. That is, the pre-eliminating of $v(\cdot)$ is not necessary for estimating S_i . Alternatively, $v(\cdot)$ can also be estimated from \hat{x}_i without seasonal adjustment. However, there are two disadvantages, if this is done: 1. One is faced with a bandwidth selection problem in a model with periodic errors. 2. The error in the nonparametric estimate $\hat{v}(\cdot)$ will cause a non-negligible bias in \hat{S}_i . Hence, this alternative estimation procedure will not be considered here.

After obtaining \hat{V}_0 and \hat{S}_i , define $\hat{r}_i = \hat{z}_i / \sqrt{\hat{V}_0 \hat{S}_i}$ to be the standardized, seasonal adjusted data. Let $K(u)$ denote a second order kernel function. Following Feng and Heiler (1998) and Feng (2002), a Nadaraya-Watson kernel estimator of $v(t)$ is given by

$$\hat{v}(t) = \frac{\sum_{i=1}^n K\left(\frac{t_i-t}{b}\right) \hat{r}_i^2}{\sum_{i=1}^n K\left(\frac{t_i-t}{b}\right)} =: \sum_{i=1}^n w_i \hat{r}_i^2, \quad (9)$$

where $w_i = K\left(\frac{t_i-t}{b}\right) [\sum_{i=1}^n K\left(\frac{t_i-t}{b}\right)]^{-1}$ and b is the bandwidth. And we define $\hat{\sigma}(t) = \hat{v}^{1/2}(t)$. It is assumed that $b \rightarrow 0$, $nb \rightarrow \infty$ as $n \rightarrow \infty$, which together with other regular conditions ensures the consistency of $\hat{v}(\cdot)$ or $\hat{\sigma}(\cdot)$.

Finally, define the standardized residuals by

$$\hat{\epsilon}_i = \hat{z}_i / \sqrt{\hat{V}_0 \hat{v}(t_i) \hat{S}_i}.$$

Following the idea for estimating the parameters in the SEMIFAR model (Beran, 1999 and Beran and Feng, 2002), it is proposed to estimate θ using the maximum likelihood estimator (MLE) of Bollerslev (1986) with ϵ_i there being replaced by $\hat{\epsilon}_i$. That is, $\hat{\theta}$ is defined as the maximizer of the (approximate) conditional log-likelihood (apart from a constant)

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n l_i(\theta) \quad (10)$$

with

$$l_i(\theta) = -\frac{1}{2} \log h_i - \frac{1}{2h_i} \hat{\epsilon}_i^2. \quad (11)$$

It is assumed that $\hat{\epsilon}_i = 0$ and $h_i = \hat{\epsilon}_i^2 = \frac{1}{n} \sum_{j=1}^n \hat{\sigma}_j^2$ for $i \leq 0$. This will not affect the asymptotic properties of $\hat{\theta}$. For more details about the MLE of θ see Bollerslev (1986) and Ling and Li (1997). For computing $\hat{\theta}$ we propose to use the S+GARCH packet.

Asymptotic properties of \bar{y} , \hat{V}_0 and \hat{S}_i given in the next section ensure that the data-driven SEMIGARCH algorithm proposed by Feng (2002) can be directly adapted to estimate the seasonal SEMIGARCH model. Such a data-driven algorithm processes as follows:

1. Estimate μ by \bar{y} .
2. Estimate V_0 and S_i by \hat{V}_0 and \hat{S}_i as defined above.
3. Run the SEMIGARCH program (in S-Plus) using the standardized, seasonal adjusted data \hat{r}_i . Then we obtain data-driven estimation of $v(\cdot)$ and θ .

For more details see Feng (2002).

The return between the last observation on one day and the first observation on the next day is called the overnight return. When analyzing high-frequency data one is also faced with the problem of how to deal with the overnight returns, because they are quite different from the intraday returns, i.e. those between observations on the same day. Here, we propose to carry out the proposed algorithm twice with and without the overnight returns. By the former approach the effect of the overnight returns may be estimated, since now the overnight returns are treated as returns in a special phase of the daily period. However, including the overnight returns will cause underestimation of the seasonality in intraday returns. The latter approach provides us more detailed information about the daily periodic change in the volatility of the intraday returns.

3 Asymptotic properties

The practical implementation proposed in the last section based on some asymptotic properties of the proposed estimators, which will be discussed in this section briefly. For

this purpose following assumptions are required.

- A1. Model (1) and (2) holds with i.i.d. $N(0, 1)$ η_i and strictly stationary ϵ_i such that $E(\epsilon_i^4) < \infty$.
- A2. The trend function $v(t)$ is strictly positive, bounded and at least twice continuously differentiable on $[0, 1]$.
- A3. S_i are strictly positive and exactly periodic with period T .

The condition of $E(\epsilon_i^4) < \infty$ is sufficient for the derivation of the asymptotic properties of the proposed estimators. However, for the practical implementation, $E(\epsilon_i^4)$ has to be estimated (see Feng, 2002). Now, the existence of finite eighth moment of ϵ_i is required.

The asymptotic properties of \bar{y} are given in the following proposition.

Proposition 1. *Under assumptions A1 to A3 we have $E(\bar{y}) = \mu$ and $\sqrt{n}(\bar{y} - \mu) \xrightarrow{\mathcal{D}} N(0, V_0)$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.*

The proof of Proposition 1 is given in the appendix. Note in particular that the asymptotic variance of \bar{y} does not depend on $v(\cdot)$ and S_i . We see \bar{y} is unbiased and has the same asymptotic variance as that of the sample mean of an i.i.d. series with variance V_0 , because ϵ_i are uncorrelated and $v(\cdot)$, S_i and ϵ_i are all standardized.

The asymptotic properties of \hat{V}_0 are given by

Theorem 1. *Under assumptions A1 to A3 we have*

- i) $E(\hat{V}_0 - V_0) = O(n^{-1})$, $\text{var}(\hat{V}_0) \doteq n^{-1}\sigma_{V_0}^2$.
- ii) $\sqrt{n}(\hat{V}_0 - V_0) \xrightarrow{\mathcal{D}} N(0, \sigma_{V_0}^2)$,

where

$$\sigma_{V_0}^2 = (nT)^{-1}V_0^2 \int_0^1 v^2(t)dt \left\{ \sum_{j=0}^{T-1} a_j b_j \right\} \quad (12)$$

with

$$a_j = \sum_{i=1}^T S_i \frac{S_{i+j} + S_{i-j}}{2}, j = 0, 1, \dots, T-1, \quad (13)$$

$$b_0 = \sum_{k=-\infty}^{\infty} \gamma_{\epsilon^2}(kT) \quad (14)$$

and

$$b_j = 2 \sum_{k=0}^{\infty} \gamma_{\epsilon^2}(kT + j), j = 1, \dots, T-1, \quad (15)$$

where $\gamma_{\epsilon^2}(k)$ are the autocovariances of the squared GARCH process ϵ_i^2 .

The proof of Theorem 1 is given in the appendix. If $v(t) \equiv 1$ and $S_i \equiv 1$, model (1) and (2) reduces to the GARCH model. Now, results in Theorem 1 reduce to known limit theorem on the sample variance of a GARCH process (see e.g. Davis et al. 1999 and Mikosch and Stărică, 2000).

Asymptotic properties of \hat{S}_i are given by the following theorem.

Theorem 2. *Under assumptions A1 to A3 we have*

i) $E(\hat{S}_i - S_i) = O(m^{-1}) = O(n^{-1})$.

ii) *The asymptotic variance of \hat{S}_i is given by*

$$\text{var}(\hat{S}_i) \doteq m^{-1} c_f^* S_i^2 \int_0^1 v^2(t) dt = m^{-1} \sigma_{S_i}^2, \quad (16)$$

where $\sigma_{S_i}^2 = c_f^* S_i^2 \int_0^1 v^2(t) dt$, $c_f^* = \sum_{i=-\infty}^{\infty} \gamma_{\epsilon^2}(iT)$ and $\gamma_{\epsilon^2}(k)$ are as defined in Theorem 1.

iii) $\sqrt{n}(\hat{S}_i - S_i) \xrightarrow{\mathcal{D}} N(0, T\sigma_{S_i}^2)$.

The proof of Theorem 2 is straightforward and is omitted. Theorem 2 shows in particular that S_i can be estimated \sqrt{n} -consistently without pre-eliminating the volatility trend and the GARCH effect.

Note that \bar{y} , \hat{V}_0 and \hat{S}_i are all \sqrt{n} -consistent. Hence, under conditions A1 to A3 and additional regular conditions in nonparametric regression, the asymptotic properties

of $\hat{v}(\cdot)$ and $\hat{\theta}$ as given in Theorems 1 to 3 in Feng (2002) hold for the corresponding estimators proposed in the last section. This is the reason, why the data-driven algorithm for estimating the SEMIGARCH model (Feng, 2002) can be directly used for estimating model (1) and (2) after seasonal adjustment.

4 Significance test and confidence bounds

4.1 An approximate significance test of seasonality

An important question is, whether the seasonal component in a return series is significant. To answer this question a test should be carried out. In the following we will propose an approximate significance test of the null hypotheses

$$H_0^i : S_i = 1, i = 1, 2, \dots, T.$$

It is proposed to reject H_0^i at the level α , if

$$\sqrt{m} \left| \hat{S}_i - 1 \right| > z_{\alpha_1/2} \hat{S}_i \sqrt{\hat{c}_f^* \int \hat{v}^2(t) dt},$$

where $z_{\alpha_1/2}$ is the $N(0, 1)$ - $\alpha_1/2$ -quantile and

$$\hat{c}_f^* = \sum_{i=-K}^K \hat{\gamma}_{\epsilon^2}(iT)$$

with an integer K such that $K \rightarrow \infty$ and $K/m \rightarrow 0$ as $n \rightarrow \infty$. This condition ensures that $\sum_{i=-K}^K \hat{\gamma}_{\epsilon^2}(iT)$ is consistent. α_1 is chosen so that the joint significance level of the test is about α . Hence α_1 should be much smaller than α . If the correlation between \hat{S}_i , $i = 1, 2, \dots, T$, is omitted, then an approximate value of α_1 may be obtained from the relationship $(1 - \alpha_1)^T = (1 - \alpha)$. One side tests can be carried out similarly.

4.2 Monte Carlo confidence bounds

Another question is, if there is a volatility trend in the data. This means that we should test the null hypothesis $H_0: v(t) \equiv 1$ or give confidence bounds of $\hat{v}(\cdot)$ under H_0 . In

the following we propose to calculate the confidence bounds based on the Monte Carlo method (see Feng, 2002 for a similar idea). The use of Monte Carlo confidence bounds in nonparametric regression is also proposed e.g. by Efromovich (1999). Assume that we have obtained a fitted SEMIGARCH model from the standardized, seasonal adjusted data. The $100(1 - \alpha)\%$ Monte Carlo confidence bounds are obtained as follows.

1. Generate a time series of length n following the estimated GARCH model.
2. Fit a SEMIGARCH model to the simulated data using the bandwidth \hat{b} .
3. Repeatedly carry out steps 1 and 2 until a given number of replications.
4. Find out proper lower and upper bounds so that the number of estimated trends, which exceed these bounds at some places, is not larger than $100\alpha\%$.

The null hypothesis $H_0: v(t) \equiv 1$ will be rejected at level α , if $\hat{v}(t)$ obtained from the real data exceeds these simulated confidence bounds at some places. The confidence level is asymptotically $(1 - \alpha)$, since $\hat{\theta}$ is consistent. Here, the bandwidth \hat{b} is used to keep the estimated trends from different replications to be comparable with each other and with $\hat{v}(\cdot)$ obtained from the real data. The confidence bounds in this paper are determined such that the numbers of estimated trends which exceed the lower and the upper bounds are the same. Note also that for calculating the total number of exceeding estimates those, which exceed the lower and the upper bounds at the same time, should not be calculated twice.

5 Applications

In the following we will apply the proposal to the 20 minute stock price returns (log-returns) of four German firms: Allianz AG, BASF AG, Henkel KGaA and Linde AG. The data are the observations of the German Xetra electronic trading system. The observation period is from November 28, 1997 to December 30, 1999 including 524 observation days. Here $T = 24$ for all returns and $T = 23$ for intraday returns only. For the parametric part a GARCH(1, 1) model is used. The estimated parameters $\sqrt{\hat{V}_0}$, \hat{b} , $\hat{\alpha}_1$ and $\hat{\beta}_1$ for all examples are given in Table 1. We see that $\sqrt{\hat{V}_0}$ with overnight returns is larger than that

obtained for the same return series without overnight returns, since the volatility of the overnight returns is much larger than the volatility in any other phase (see Figures 1 to 5 below). The selected bandwidths for the same data set with or without overnight returns are quite similar, because $v(\cdot)$ is estimated from the seasonal adjusted data. The fitted GARCH models are all highly significant. The difference between the fitted GARCH models in cases with or without overnight returns is not clear. Following the existence condition for finite high order moments of a GARCH(1, 1) model given by Bollerslev (1986), it can be checked that all of the fitted GARCH models have at least finite eighth moment. This together with the results given below shows that financial return series often have finite high order moments but are in general nonstationary (see Mikosch and Stărică, 1999 and Feng, 2002 for related findings).

Results of one side significance tests of $H_0^i: S_i = 1$ against $H_1^i: S_i > 1$ (for $\hat{S}_i > 1$) or $H_1^i: S_i < 1$ (for $\hat{S}_i < 1$), $i = 1, 2, \dots, T$, are listed in Table 2 where the codes “1”, “0” and “-1” stand for $S_i > 1$, $S_i = 1$ and $S_i < 1$ respectively. In these tests $\alpha_1 = 0.0022$ was used so that $\alpha \approx 0.05$. Here only results in cases without overnight returns are given. The observation time intervals are: 9:20–9:40, 9:40–10:00, ..., 16:20–16:40 and 16:40–17:00. For calculating \hat{c}_f^* , $K = \lceil \sqrt{n}/T + 0.5 \rceil = \lceil \sqrt{12052}/23 + 0.5 \rceil = 5$ is used. These results show that the seasonality is for all examples significant. Observing the Monte Carlo confidence bounds for $\hat{v}(\cdot)$ shown in Figures 1 to 5, we can see that h_i , $v(t)$ as well as S_i are significantly non-constant for all examples.

Detailed results obtained following the seasonal SEMIGARCH model for the BASF return series are shown in Figures 1 and 2. The data for this series with overnight returns are shown in Figure 1a. Estimation results for this example are shown in Figures 1b through h. The estimated seasonal component in Figure 1b shows that the overnight returns are clearly different from those in other phases. The estimated trend $\hat{v}(\cdot)$ is shown in Figure 1c together with 95% (long dashes) and 99% (short dashes) Monte Carlo confidence bounds calculated from 400 replications. Figure 1d shows the standardized residuals from which the GARCH(1, 1) model was fitted. The conditional standard deviations calculated following this GARCH model are shown in Figure 1e. Figure 1f displays the total volatility, i.e. the product of the three components shown in Figures 1b, c and e and the averaged standard deviation $\sqrt{\hat{V}_0}$. The zoomed total standard deviations for the last ten days are shown in Figure 1g. Figure 1h shows the prediction of the volatility for five

days in the future, where the estimated trend for the last observation, i.e. $\sqrt{\hat{v}(1)}$, is used as the scale function in the recent future, the averaged standard deviation is assumed to be unchanged and the prediction of the conditional standard deviations is obtained using the S-Plus GARCH function *predict* (see Martin et al., 1996). We see, the GARCH effect in the prediction decay very quickly and hence the seasonal component plays a more important role.

Figure 2 shows the same results as given in Figure 1 but for the BASF return series without overnight returns. We see, the seasonality in this case is more regular. Note that the difference between the two values of \hat{S}_i obtained with and without overnight returns is only due to the difference of \hat{V}_0 in these two cases. The estimated trends in these two cases are almost the same. Estimation results for the Allianz, Henkel and Linde return series are shown in Figures 3 to 5 respectively for cases with (Figures a to d) and without (Figures e to h) overnight returns, where some details are omitted to save space. From Figures 2 to 5 we see that the seasonal component in the case without overnight returns has a “U” form over one day. That is the volatility near the open and close time is generally larger than that near the noon. But the change from one phase to another is not smooth, especially by the Henkel and Linde returns. The seasonality is most regular by the Allianz returns.

The fitted trend in the considered time period has a \cap form. That is the volatility is larger in the middle of this observation period and small at both ends. The volatility trend is smallest at the current end of these series except for the BASF returns. This property is important for predicting the future volatility, because it shows that the non-seasonal unconditional variance at the current end is much smaller than the averaged level. Hence one can obtain more reasonable prediction for future volatility by introducing the trend function into the parametric GARCH model.

Figures 1 to 5 also show that all the confidence bounds for $\hat{v}(\cdot)$ are not symmetric. The distance between the upper bound and the unit level is always larger than that between the lower bound and the unit level. This means that the estimated trend from data generated by a GARCH model without trend often has some larger peaks. Furthermore, following the asymptotic normality of $\hat{v}(t)$ (see Feng, 2002) we can easily calculate confidence intervals for the trend at a given point t_0 . However, this does not provide correct confidence bounds for the trend function on the whole support $[0, 1]$. The length

between the lower and upper Monte Carlo confidence bounds is much larger than that of a corresponding confidence interval of the trend at a given point.

The idea behind the seasonal SEMIGARCH model can be well understood, if we compare the ACF's of different time series transformed from the same return series. Let $\hat{x}_i = \hat{z}_i/\sqrt{\hat{V}_0}$, $\hat{r}_i = \hat{z}_i/\sqrt{\hat{V}_0\hat{S}_i} = \hat{x}_i/\sqrt{\hat{S}_i}$ and $\hat{\epsilon}_i = \hat{z}_i/\sqrt{\hat{V}_0\hat{v}(t_i)\hat{S}_i} = \hat{x}_i/\sqrt{\hat{v}(t_i)\hat{S}_i}$ as defined in Section 2. Note that the time series \hat{x}_i should have both trend and seasonality in volatility as for the return series itself. \hat{r}_i are seasonal adjusted data and should only have trend in volatility. $\hat{\epsilon}_i$ are consistent estimates of ϵ_i and should have neither trend nor seasonality in volatility. Also define $\hat{\zeta}_i = \hat{x}_i/\sqrt{\hat{v}(t_i)}$ to be the trend adjusted data for comparison, which should have seasonality in volatility.

The ACF's of $|\hat{x}_i|$, $|\hat{r}_i|$, $|\hat{\zeta}_i|$ as well as $|\hat{\epsilon}_i|$ in all cases are displayed in Figures 6 and 7. Figures 6a to d show these results for the Allianz return series with overnight returns. The same results for the Allianz return series without overnight returns are displayed in Figures 6e to h. Figures 6i to p, Figures 7a to h and Figures 7i to p show the same results as those in Figures 6a to h, but for the BASF, Henkel and Linde return series respectively. Both, the trend and seasonal effects can be seen clearly from the ACF's of $|\hat{x}_i|$. The ACF's of $|\hat{r}_i|$ exhibit only trend effect. This means that the seasonality is well modelled and eliminated following the proposed algorithm, in both cases with or without overnight returns. Note also that the trend effect becomes more clear after eliminating the seasonality. The ACF's of $|\hat{\zeta}_i|$ exhibit only seasonal effect as expected. That is the trend is properly estimated and eliminated from these series. Note that all of the ACF's of $|\hat{x}_i|$, $|\hat{r}_i|$ and $|\hat{\zeta}_i|$ indicate non-stationarity in these series. Again, we see that the seasonality in the volatility of the Henkel and Linde returns is not so regular as that in the other two return series. But the volatility trend in these two series is more clear.

The ACF's of $|\hat{\epsilon}_i|$ displayed in Figures d, h, l and p of Figures 6 and 7 show that these series seem to be stationary and that there is clear GARCH effect in the data after eliminating both the seasonality and the trend. Stationary CH models, for instance the GARCH model as considered in this paper, can then be fitted to these seasonal and trend adjusted series. Figures 1d and h show that the persistence level in $\hat{\epsilon}_i$ is sometimes very high. This indicates that a CH model with long-memory property is sometimes more preferable. However, this is not considered here.

6 Final remarks

In this paper the estimation of different volatility components in high-frequency financial returns is investigated. A new approach to perform this is introduced and the data-driven algorithm proposed by Feng (2002) is adapted to estimate the model in this paper. Asymptotic results, significance test of seasonality and Monte Carlo confidence bounds of the trend are investigated. Data examples show that the proposal works well in practice. However, there are still some open questions including the development of a joint significance test of seasonality and the development of a theoretical significance test of the whole trend function. The latter is also an important open question in standard nonparametric regression. Also the problem of the model selection is not discussed. This problem may perhaps be solved by using the AIC or BIC information criteria. Finally, it is worthwhile to extend the idea in Feng (2002) and in this paper to other GARCH variants, e.g. FARIMA-GARCH model.

Acknowledgements

The paper was financially supported by the *Center of Finance and Econometrics (CoFE)*, University of Konstanz, Germany. We are very grateful to Prof. Jan Beran, Department of Mathematics and Statistics, University of Konstanz, for the advice. We would very like to thank Prof. Winfried Pohlmeier and Mr. Nikolaus Hautsch, Department of Economics, University of Konstanz, for providing us the data. Without their help this paper would not be finished. Our special thanks go to Mr. Erik Lüders, CoFE/ZEW, for helpful discussions and suggestions, which lead to improve the quality of this paper. Finally, we are grateful to Prof. Wolfgang Härdle, Prof. Olaf Bunke and Dr. Woocheol Kim, Humboldt University, Berlin, Prof. Vladimir Spokoiny, Weierstraß-Institut, Berlin, and Prof. Rohit Deo, New York University and Humboldt University, for useful comments.

Appendix. Proofs of results

Proof of Proposition 1. It is obvious that \bar{y} is unbiased. Hence we need to check $\text{var}(\bar{y})$ and the asymptotic normality of \bar{y} . Note that the autocovariances of ϵ_i are $\gamma_\epsilon(0) = 1$ and $\gamma_\epsilon(k) = 0$ for $|k| > 0$, since ϵ_i is a standardized GARCH process. We have

$$\text{var}(\bar{y}) = n^{-2}V_0 \sum_{i=1}^n v(t_i)S_i. \quad (\text{A.1})$$

Let $m = \lfloor n/T \rfloor$ as defined in (8). Note that $v(t_i) \doteq v(t_j)$ for $|i - j| < T$ and observe the standardizing assumptions on S_i and $v(t)$. We have

$$\begin{aligned} n^{-1} \sum_{i=1}^n v(t_i)S_i &\doteq n^{-1} \sum_{j=0}^{m-1} \sum_{k=1}^T v(t_{jT+k})S_{jT+k} \\ &\doteq n^{-1} \sum_{j=0}^{m-1} v(t_{jT+1}) \sum_{k=1}^T S_{jT+k} \\ &= n^{-1}T \sum_{j=0}^{m-1} v(t_{jT+1}) \doteq \int_0^1 v(t)dt. \end{aligned} \quad (\text{A.2})$$

Under the assumptions we have $\int_0^1 v(t)dt = 1$, that is $\text{var}(\bar{y}) \doteq n^{-1}V_0$.

Furthermore, under the assumptions of model (1) and (2) it can be shown that \bar{y} is asymptotically normal, if and only if the sample mean of the GARCH process ϵ_i is. It is well known that the sample mean of a GARCH process with finite fourth moment is asymptotically normal (see e.g. Beran and Feng, 2001). Proposition 1 is proved. \diamond

Proof of Theorem 1. Following the results of Proposition 1 it can be shown that

$$E(\hat{V}_0) = n^{-1} \sum_{i=1}^n E[Z_i^2] + O(n^{-1}), \quad (\text{A.3})$$

where $Z_i = Y_i - \mu$ are as defined in Section 2.

$$\begin{aligned} n^{-1} \sum_{i=1}^n E[Z_i^2] &= n^{-1}V_0 \sum_{i=1}^n v(t_i)S_i \text{var}(\epsilon_i) \\ &= n^{-1}V_0 \sum_{i=1}^n v(t_i)S_i \\ &= V_0 + O(n^{-1}). \end{aligned} \quad (\text{A.4})$$

The last equation is due to the same argument used in (A.2). One obtains $E(\hat{V}_0 - V_0) = O(n^{-1})$.

Since \bar{y} is consistent, we have for the variance of \hat{V}_0

$$\begin{aligned}\text{var}(\hat{V}_0) &\doteq n^{-2} \text{var} \left(\sum_{i=1}^n Z_i^2 \right) \\ &\doteq n^{-2} V_0^2 \sum_{i=1}^n \sum_{j=1}^n v(t_i) v(t_j) S_i S_j \gamma_{\epsilon^2}(i-j).\end{aligned}\quad (\text{A.5})$$

The autocovariances $\gamma_{\epsilon^2}(k)$ of the squared GARCH process ϵ_i^2 decay exponentially (see e.g. He and Teräsvirta, 1999). Hence $\sum_{j=1}^n v(t_i) v(t_j) S_i S_j \gamma_{\epsilon^2}(i-j)$ converges absolutely. Let $h > 0$ such that $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$. And let $N_n = [nh]$. We have $\sum_{j=1}^n v(t_i) v(t_j) S_i S_j \gamma_{\epsilon^2}(i-j) \doteq \sum_{|i-j| \leq N_n} v(t_i) v(t_j) S_i S_j \gamma_{\epsilon^2}(i-j)$ and $v(t_i) \doteq v(t_j)$ for $|i-j| \leq N_n$. This analysis leads to

$$\begin{aligned}\text{var}(\hat{V}_0) &\doteq n^{-2} V_0^2 \sum_{i=N_n}^{n-N_n} \sum_{j=1}^n v(t_i) v(t_j) S_i S_j \gamma_{\epsilon^2}(i-j) \\ &\doteq n^{-2} V_0^2 \sum_{i=N_n}^{n-N_n} \sum_{|i-j| \leq N_n} v(t_i) v(t_j) S_i S_j \gamma_{\epsilon^2}(i-j) \\ &\doteq n^{-2} V_0^2 \sum_{i=N_n}^{n-N_n} v^2(t_i) \sum_{|i-j| \leq N_n} S_i S_j \gamma_{\epsilon^2}(i-j) \\ &\doteq n^{-2} V_0^2 \sum_{i=N_n}^{n-N_n} v^2(t_i) \sum_{k=-\infty}^{\infty} S_i S_{i-k} \gamma_{\epsilon^2}(k).\end{aligned}\quad (\text{A.6})$$

Furthermore, note that $\sum_{k=-\infty}^{\infty} S_i S_{i-k} \gamma_{\epsilon^2}(k)$ is periodic in i with the same period T and $v^2(t_i) \doteq v^2(t_j)$ for $|i-j| < T$. Let $M_1 = [N_n/T]$ and $M_2 = [(n - N_n)/T]$. We have

$$\begin{aligned}\text{var}(\hat{V}_0) &\doteq n^{-2} V_0^2 \sum_{j=M_1}^{M_2} v^2(t_{jT}) \left\{ \sum_{i=1}^T \sum_{k=-\infty}^{\infty} S_i S_{i-k} \gamma_{\epsilon^2}(k) \right\} \\ &\doteq (nT)^{-1} V_0^2 \int_0^1 v^2(t) dt \left\{ \sum_{i=1}^T \sum_{k=-\infty}^{\infty} S_i S_{i-k} \gamma_{\epsilon^2}(k) \right\}.\end{aligned}\quad (\text{A.7})$$

Straightforward calculation leads to

$$\left\{ \sum_{i=1}^T \sum_{k=-\infty}^{\infty} S_i S_{i-k} \gamma_{\epsilon^2}(k) \right\} = \sum_{j=0}^{T-1} a_j b_j, \quad (\text{A.8})$$

where a_j and b_j , $j = 0, 1, \dots, T-1$, are as defined in Theorem 1.

Note again that \hat{V}_0 is asymptotically normally distributed, if and only if the sample variance of the GARCH process ϵ_i is. The latter result is shown by Davis et al. (1999) (see also Mikosch and Stărică, 1999 and Feng, 2002). Theorem 1 is proved. \diamond

References

- Beran, J. (1999), “SEMIFAR models – A semiparametric framework for modelling trends, long range dependence and nonstationarity,” Discussion paper, CoFE, No. 99/16, University of Konstanz.
- Beran, J. and Y. Feng (2001), “Local polynomial estimation with a FARIMA-GARCH error process,” *Bernoulli*, 7, 733 – 750.
- — (2002), “SEMIFAR models - A semiparametric framework for modelling trends, long range dependence and nonstationarity,” *Computational Statistics and Data Analysis*, 40, 393–419.
- Beran, J. and Ocker, D. (2001), “Volatility of Stock Market Indices - An Analysis based on SEMIFAR models,” *Journal of Business and Economic Statistics*, 19, 103-116.
- Bollerslev, T. (1986), “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.
- Dacorogna, M.M., Gencay, R., Müller, U.A., Olsen, R.B. and Pictet, O. V. (2001), *An introduction to high-frequency finance*, San Diego: Academic Press.
- Davis, R.A., Mikosch, T. and Basrak, B. (1999), “Sample ACF of Multivariate stochastic recurrence equations with application to GARCH,” Preprint, University of Groningen.
- Efromovich, S. (1999), *Nonparametric Curve estimation: Methods, Theory, and Applications*, New York: Springer.
- Engle, R.F. (1982), “Autoregressive conditional heteroskedasticity with estimation of U.K. inflation,” *Econometrica*, 50, 987–1008.
- Feng, Y. (2002), “Simultaneously Modelling Conditional Heteroskedasticity and Scale Change,” Discussion Paper, CoFE, No. 02/12, University of Konstanz (revised).
- Feng, Y. and Heiler, S. (1998), “Locally weighted autoregression,” in *Econometrics in Theory and Practice – Festschrift für Hans Schneeweiß*, eds. R. Galata and H. Küchenhoff, Heidelberg: Physica-Verlag.

- Härdle, W., Spokoiny, V. and Teyssière, G. (2001), “Adaptive estimation for a time inhomogeneous stochastic-volatility model,” Discussion Paper, SFB 373, Humboldt University.
- He, C. and Teräsvirta, T. (1999), “Forth moment structure of the GARCH(p, q) process,” *Econometric Theory*, 15, 824–846.
- Ling, S. (1999), “On probability properties of a double threshold ARMA conditional heteroskedasticity model,” *Journal of Applied Probability*, 36, 688–705.
- Ling, S. and Li, W.K. (1997), “On fractional integrated autoregressive moving-average time series models with conditional heteroskedasticity,” *Journal of the American Statistical Association*, 92, 1184–1194.
- Ling, S. and McAleer, M. (2002), “Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models,” *Econometric Theory*, 18, 722–729.
- Mercurio, D. and Spokoiny, V. (2002), “Statistical inference for time-inhomogeneous volatility models,” Discussion Paper, SFB 373, Humboldt University.
- Mikosch, T. and Stărică, C. (1999), “Change of structure in financial time series, long range dependence and the GARCH models,” Preprint, University of Groningen.
- — (2000), “Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process,” *Annals of Statistics*, 28, 1427–1451.

Table 1. Estimated parameters for all examples

	For all returns				Intraday returns only			
	Allianz	BASF	Henkel	Linde	Allianz	BASF	Henkel	Linde
\hat{b}_{opt}	0.1588	0.1150	0.1196	0.1262	0.1567	0.1113	0.1213	0.1268
$\sqrt{\hat{V}_0}$	0.0050	0.0047	0.0063	0.0063	0.0045	0.0042	0.0058	0.0057
$\hat{\alpha}_1$	0.0592	0.0773	0.1232	0.0911	0.0631	0.0769	0.1473	0.0991
$\hat{\beta}_1$	0.8988	0.8380	0.7468	0.8542	0.8908	0.8305	0.6935	0.8379

Table 2. Relative volatility strength during one day (intraday returns only)

Allianz	9:00 - 13:00	—	1	1	1	1	0	0	-1	-1	-1	-1	-1
	13:00 - 17:00	-1	-1	-1	-1	0	-1	-1	-1	1	1	1	1
BASF	9:00 - 13:00	—	1	1	1	1	0	-1	-1	-1	0	-1	0
	13:00 - 17:00	-1	-1	-1	-1	0	0	-1	-1	0	1	1	1
Henkel	9:00 - 13:00	—	1	1	0	1	0	0	0	-1	-1	0	-1
	13:00 - 17:00	-1	-1	-1	0	-1	-1	-1	-1	1	1	0	1
Linde	9:00 - 13:00	—	1	1	1	1	0	0	-1	-1	0	0	0
	13:00 - 17:00	-1	-1	-1	-1	-1	-1	-1	0	1	0	0	1

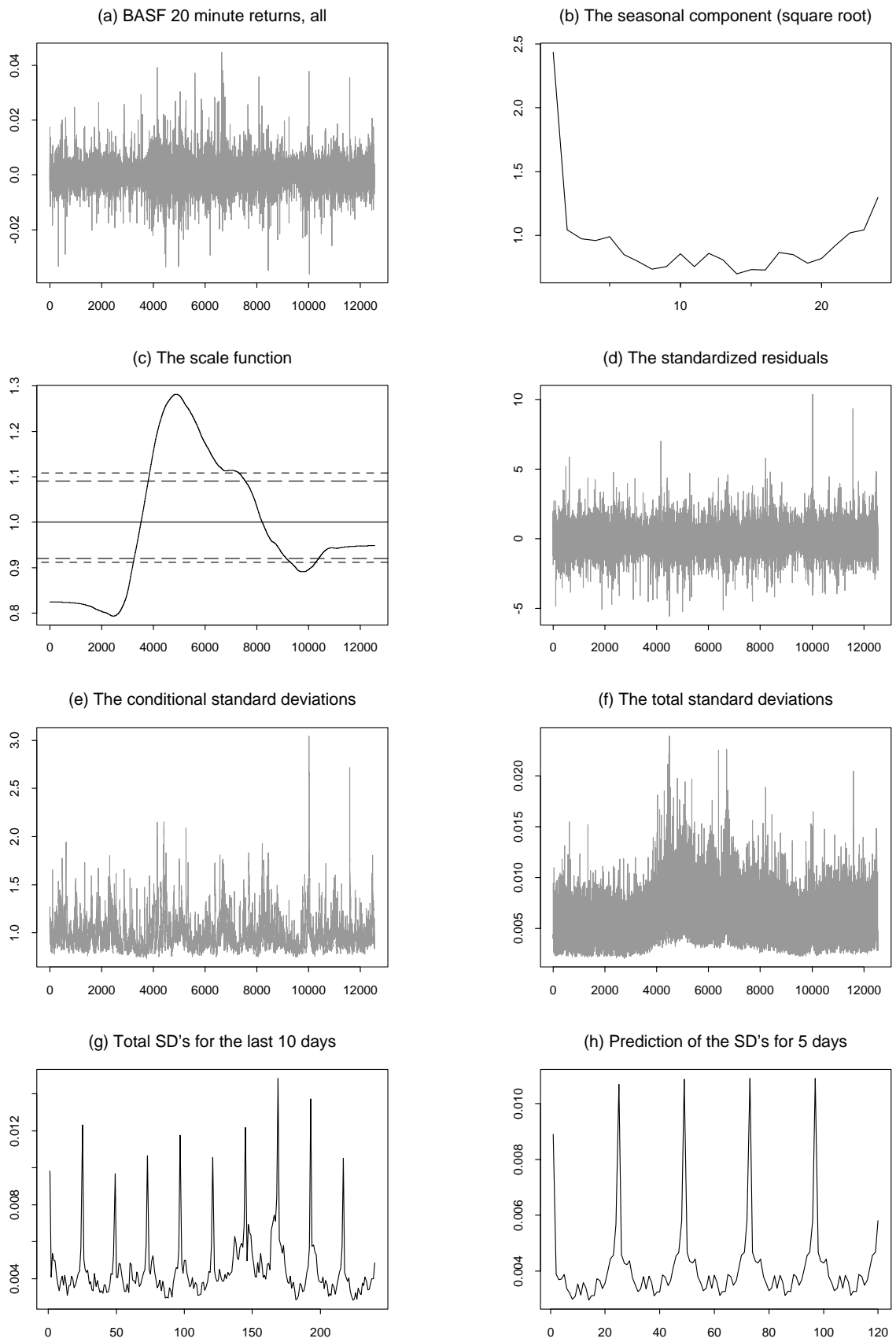


Figure 1: Estimation results for the BASF returns (with overnight returns).

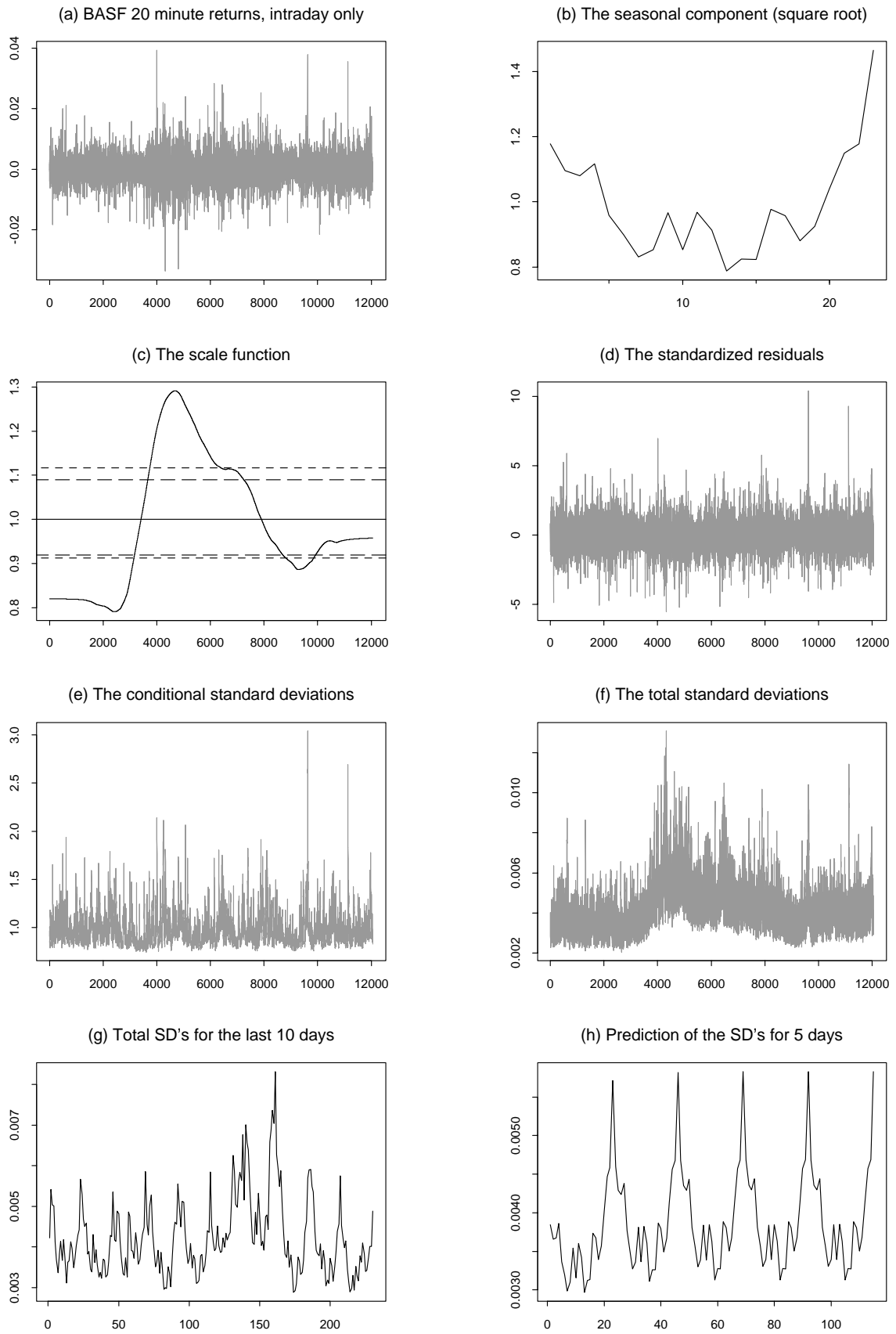


Figure 2: Estimation results for the BASF returns (intraday returns only) .

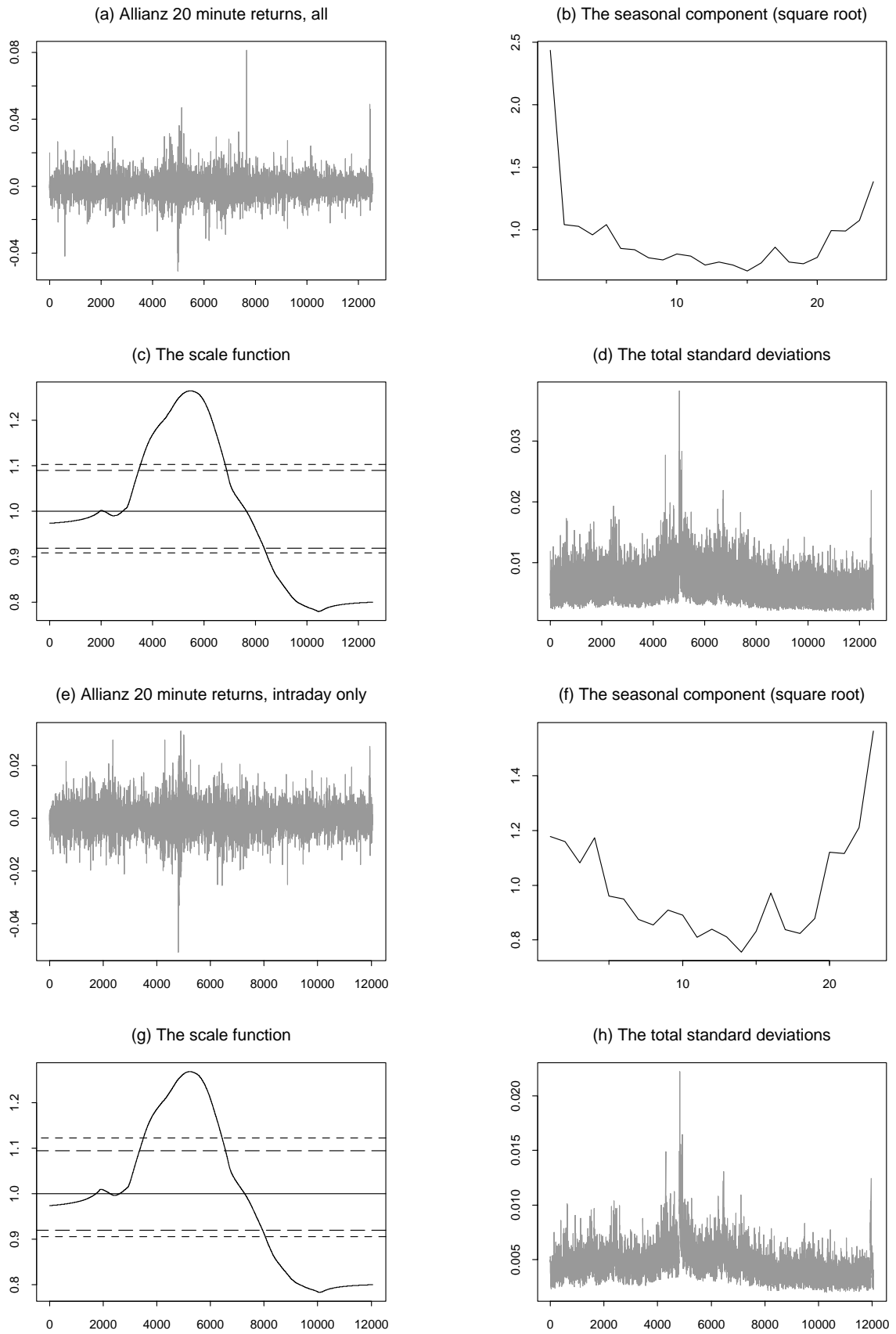


Figure 3: Estimation results for the Allianz returns with (a to d) and without (e to h) overnight returns.

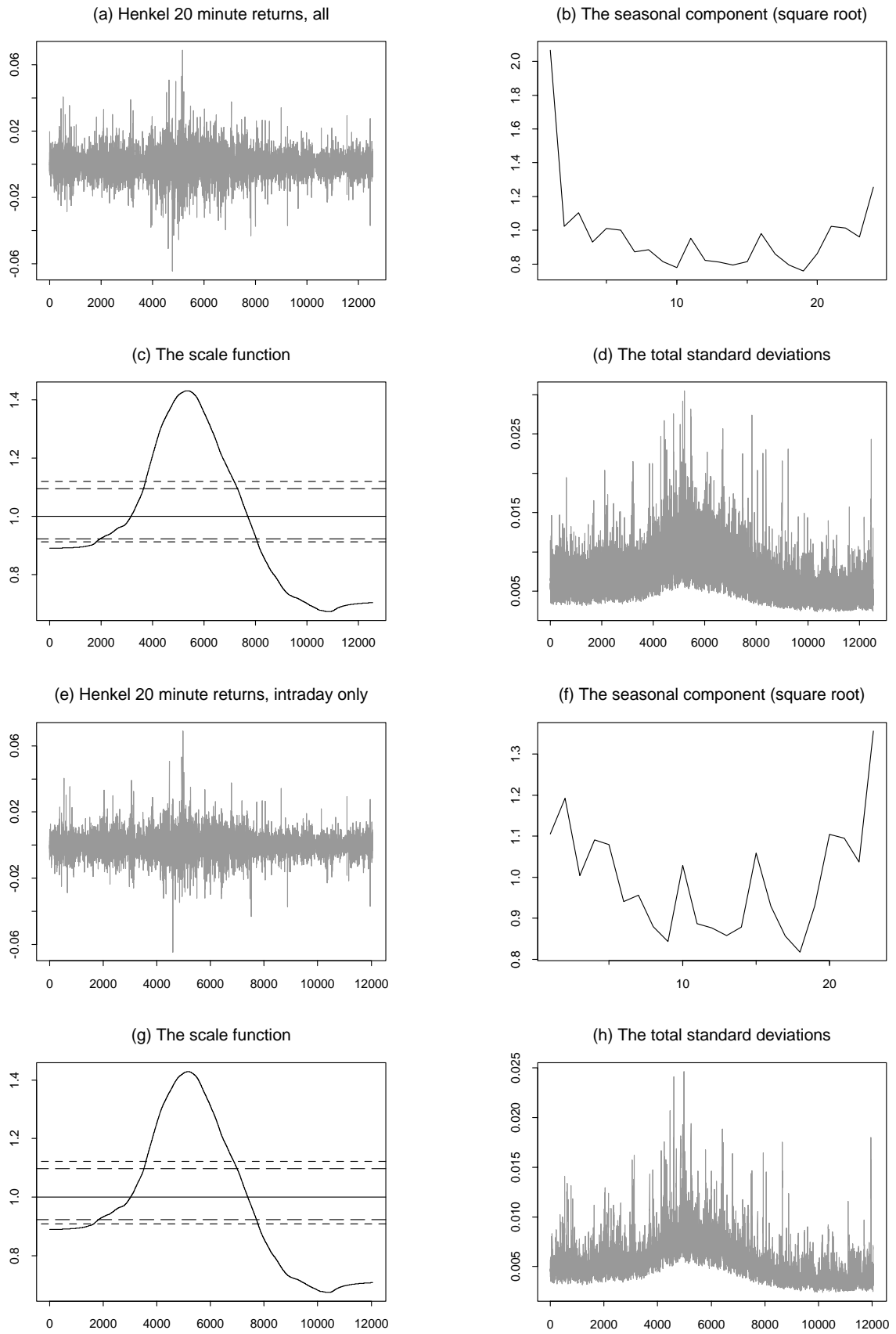


Figure 4: Estimation results for the Henkel returns with (a to d) and without (e to h) overnight returns.

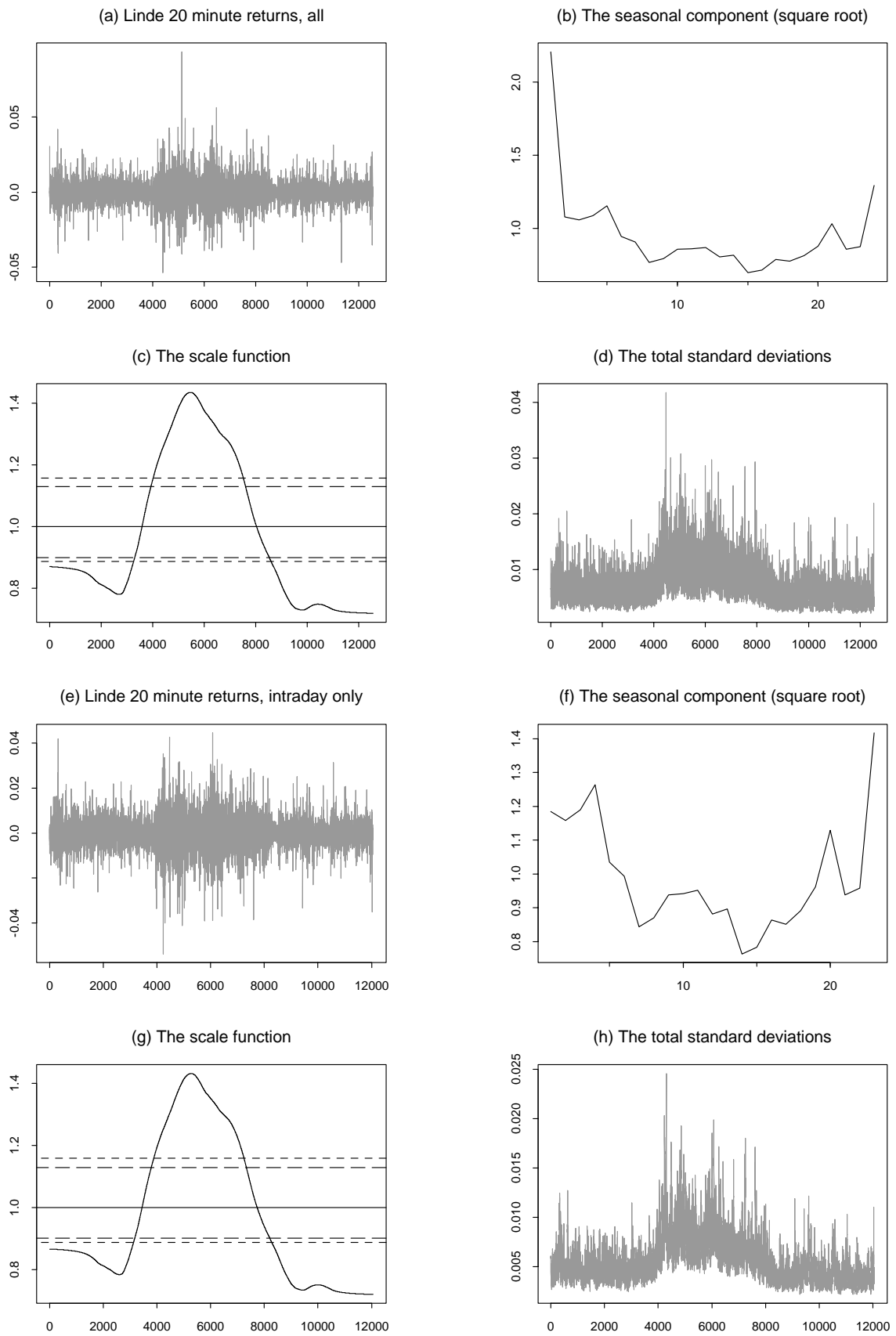


Figure 5: Estimation results for the Linde returns with (a to d) and without (e to h) overnight returns.

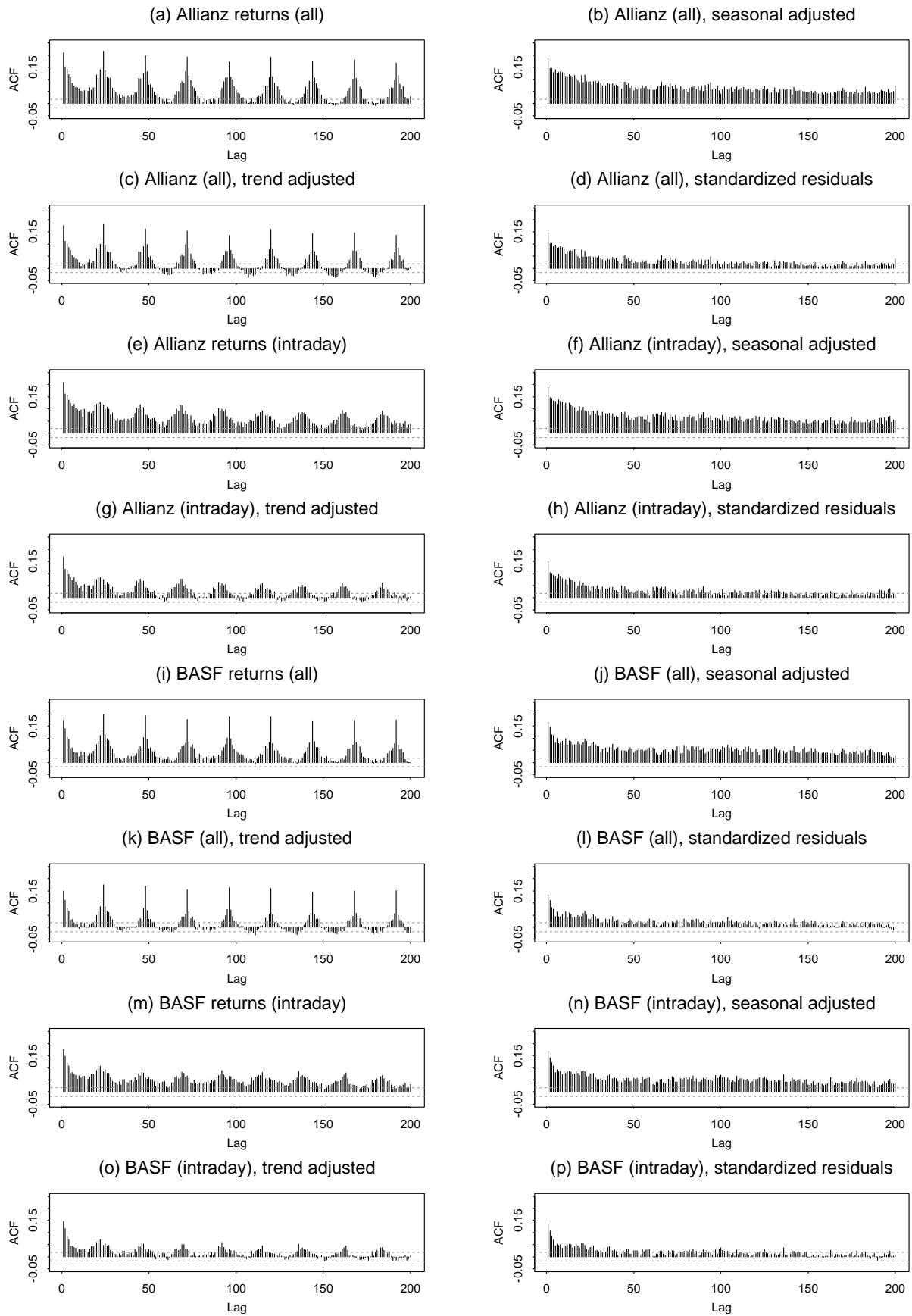


Figure 6: ACF's for different transformed time series obtained from the Allianz returns (Figures a to h) and the BASF returns (Figures i to p).

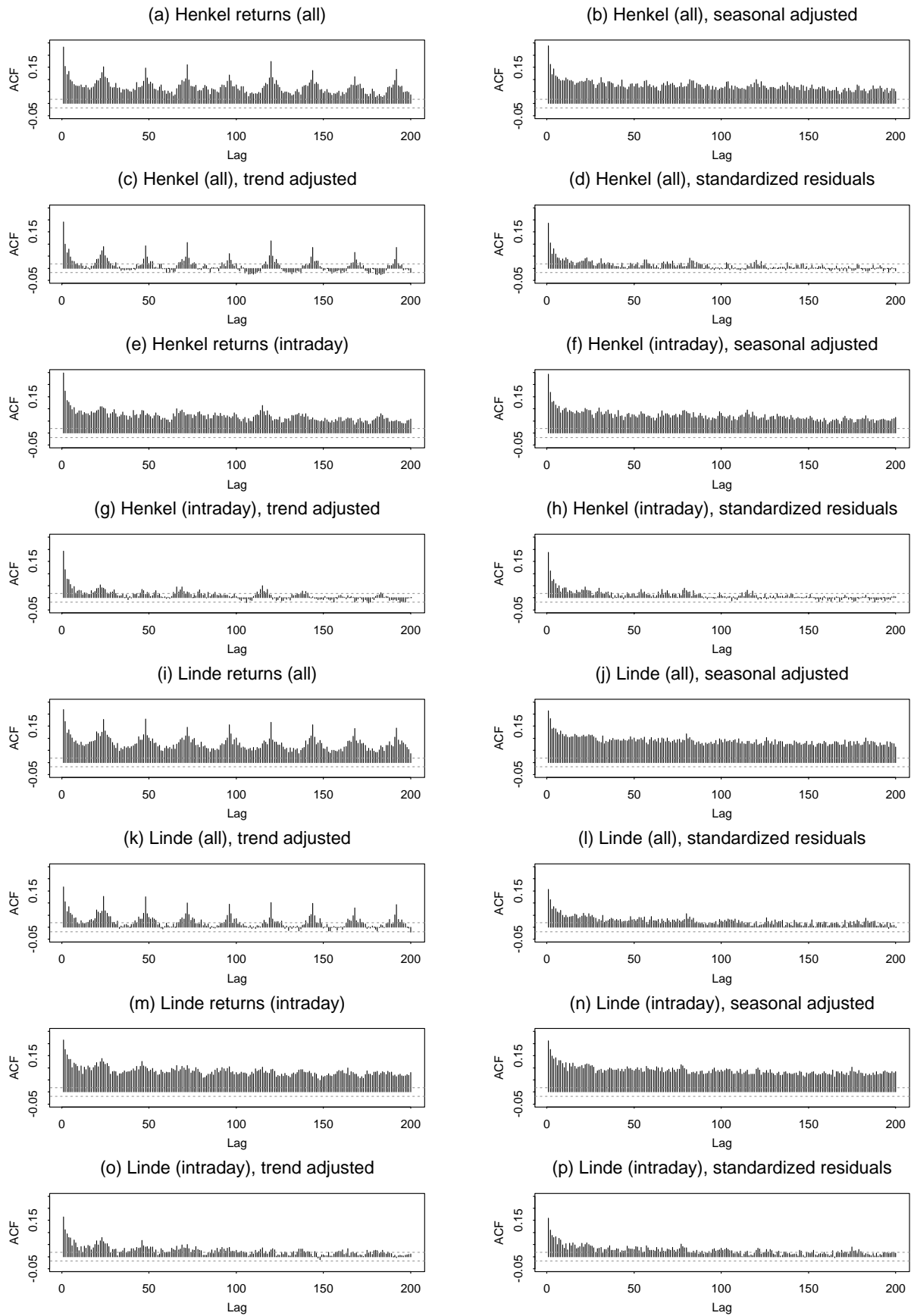


Figure 7: ACF's for different transformed time series obtained from the Henkel returns (Figures a to h) and the Linde returns (Figures i to p).