STOCK MARKET RISK-RETURN INFERENCE. AN UNCONDITIONAL NON-PARAMETRIC APPROACH.

Abstract. By means of a detailed analysis of the returns of the Standard & Poors 500 (S&P 500) composite stock index over the last fifty years we show how theoretical results and methodological recommendations from the statistical theory of non-parametric curve inference allow one to consistently estimate expected return and volatility. In this approach we do not postulate an *a priori* relationship risk-return nor do we specify the evolution of the first two moments through covariates. Our analysis gives statistical evidence that the expected return of the S&P 500 index as well as the market price of risk (the ratio expected return minus risk free interest rate over volatility) vary significantly through time both in size and sign. In particular, the periods of negative (positive) estimated expected return and market price of risk coincide with the bear (bull) markets of the index as defined in the literature. A complex relationship between risk and expected return emerges which is far from the common assumption of a positive linear time-invariant relation.

Key words and phrases. Non-stationarity, expected excess return, risk free interest rate, volatility, kernel curve estimator.

Adequate models for the time-evolution of the distribution of returns are important for the investor and the risk manager. While early studies of stock market returns (King [28], Blume [3], Officer [40], Merton [33]) identified the variance as the main time-changing characteristic of the return distribution, more recent studies, including Fama and French [13], Harvey [24], Kandel and Stambaugh [27], Whitelaw [49], have suggested significant time variation in expected return, but also in the risk-return relationship apparently related to the business cycle. While changes in the volatility are relatively easy to document, changes of the expected return are more difficult to detect. Indeed, various analyzes (e.g. Merton [33], French et al. [15]) show that expected returns are of the same (or lower) order of magnitude as volatility. This observation turns inference on expected return into a particularly difficult statistical problem - a finding which is supported by the results of the present study.

In this paper we argue that the theory of statistical curve estimation offers a suitable and convenient set-up for simultaneous consistent non-parametric inference on time-changing expected return *and* volatility. To make things precise, denote by $R_t = (P_t - P_{t-1})/P_{t-1}$ the net returns of a price or stock index P_t observed at equidistant instants of time. Our analysis is conducted under the simple modeling assumption

$$R_t = \mu(t) + \sigma(t)\,\varepsilon_t, \qquad t = 1, 2, \ldots, n\,, \tag{1}$$

where the functions μ and σ are supposed to be smooth, the noise (ε_t) is iid with mean zero and unit variance, without further specification of the distribution. In particular, we do not assume the noise Gaussian. In words, returns are modeled as independent observations with unconditional mean and variance changing slowly through time. Our assumptions on the volatility function σ and the expected return function μ are close to those in Merton [33]. Our approach is motivated by the findings of the non-stationary analysis performed in Stărică and Granger [46]. The authors argue that most of the dynamics of the S&P 500 index is explained by shifts of the unconditional variance and show that a model close to 1, where the returns are independent and display a piece-wise constant unconditional variance, describes the dynamics of the data better than Garch-type or long memory-type models.

The modeling setting positions our paper in that recent vein of the literature on risk-return inference which takes distance from tight parametric estimation of the excess return-volatility relationship insofar that it assumes the dynamics of the expected return and volatility to be driven by exogeneous factors (Kandel and Stambaugh [27], Scruggs [44], Whitelaw [49], [50]).

We aim at extending the literature on risk-return estimation in three important directions. *First,* we refine the non-parametric approach to volatility estimation, initiated by Officer [40] and Merton [33], in the light of recent theoretical developments in non-parametric regression. Although early research on time-varying volatility obtained variance estimates from asset returns without specifying a parametric model (Officer [40], Merton [33], French et al. [15]), until recently[1] the dominant approach to volatility estimation has been tightly parametric[2].

---

[1]The non-parametric approach to volatility estimation introduced in Merton [33] and French et al. [15] which uses non-overlapping samples of higher frequency (daily) data to estimate standard deviations of lower frequency (monthly) returns, largely ignored for more then a decade of parametric modeling, has recently been revived in the burgeoning literature on realized volatility in the context of high-frequency data (see Andersen et al. [1] and the references therein).

[2]This means that a particular parametric model for the volatility is specified *a priori* and then used to extract volatility estimates from the returns. These models often belong to the *stationary* conditionally heteroscedastic class (see Bollerslev et al. [5] for an overview).

Lacking strong prior information about the functional form of the volatility, we argue that a non-parametric approach to its inference is more appropriate. The hypothesis of stationary volatility, which is implicitly assumed in most conditionally heteroscedastic models, is not plausible for financial asset returns over periods of time longer than a few years (see Stărică [45] and Herzel et al. [25] for an evaluation of the impact of the stationary assumption on financial returns modeling). It is worth emphasizing that our methodology yields consistent estimates of *non-stationary, time-changing volatility* even when the expected return varies with time. Since our approach to volatility estimation is free of parametric assumptions on the mean term it avoids misspecification which is likely to affect a related methodology proposed by Rodriguez-Poo and Linton [42][3].

Second, we employ non-parametric estimation of the expected return in close relation to volatility estimation. This technique, which differs from previous studies on expected return and volatility, is desirable for a number of reasons. Our estimation procedure does not impose an *a priori* functional form neither on the first moment nor on its relationship with volatility. This is important when consistent estimates of the expected return require correct specification of the underlying data generating process. In fact, the conflicting findings of the literature which focuses on parametric estimation of the risk-return relationship might

[3]Their paper does not discuss modeling of financial data. Their covariance estimation procedure is developed in the framework of non-parametric factor analysis and applied to a macro model of the US economy. The authors suppose the first moment is a linear functional of stationary exogeneous variables and the covariance structure changes slowly through time. As the authors point out, in their set-up, correct specification of the conditional mean is crucial for consistent variance estimation.

also reflect various misspecification[4]. For example, Campbell and Hentschel [8] and French et al. [15] find the expected excess return positively related to its variance, whereas Campbell [7], Glosten et al. [19], Nelson [39] report a negative relationship between the expected excess return and conditional volatility. Misspecification is also likely to affect the alternative approach to parametric risk-return inference that models the first two moments as linear specifications of predetermined financial variables (Scruggs [44], Whitelaw [49, 50])[5]. Although our modeling approach essentially assumes exogenously driven dynamics for the first two moments, we need not worry about specifying an appropriate group of endogenous explanatory variables (an issue faced when modeling the expected return and the volatility as covariates). It is well known that selecting the endogenous explanatory variables from the

[4]Most of the papers in this literature assume a static proportional relationship between the first two moments implied by capital asset pricing theory (Merton [33], Harvey [24], Glosten et al. [19], Campbell [7], Campbell and Hentschel [8], Chan et al. [10]). The volatility is also estimated parametrically by using, most often, an ARCH-type specification.

[5]This approach replaces the tight parametric assumptions on the direct relationship risk-return with other parametric assumptions (most likely less rigid) on the type of relationship (linear, time-invariant) between the moments and a group of explanatory variables. Whitelaw [50] reports instability of the coefficient estimates in a rolling regression estimating the linear relationship between the moments and the explanatory variables as evidence of possible misspecification. Besides, there is no *a priori* reason to believe that the relation between the first two moments and the pre-determined financial variables should be linear. In fact, Brandt [6] for example, detects significant non-linearities within the context of a portfolio choice problem.

large universe of potential regressors based on their demonstrated predicting power raises the concern about potential data snooping biases[6].

Third, our approach shows the feasibility of *unconditional modeling* of financial returns[7]. The unconditional approach described in this paper is intuitively appealing and has technical advantages[8]. Model (1) is an extension of the simple and elegant log-normal model of Samuelson [43] that describes the returns R_t as iid normal. Our approach preserves the independence assumption of the log-normal model but abandons the hypothesis of identically

[6]See Foster et al. [14] and Lo and MacKinley [30] for a discussion of data snooping and its implications.

[7]The current econometric literature on financial returns overwhelmingly uses the stationary conditional modeling paradigm of which the ARCH-type processes are an outstanding example.

[8]In a conceptual sense, the two modeling paradigms are two alternative approaches to the modeling of change. In the conditional approach, one usually assumes that the unconditional distribution does not change (i.e., the returns constitute a strictly stationary sequence) and the arrival of information is reflected in the time evolution of the conditional distribution given the information about the past. In the unconditional approach, the time evolution of the return distribution reflects the arrival and incorporation of information, while the assumption of stationarity is completely dropped. Based on our statistical experience on modeling return data and given the length of the time series under consideration, non-stationarity is an intuitively appealing modeling feature. The two frameworks can also be thought of as alternative approaches to modeling non-linearities present in the dynamics of returns.

distributed normal returns. On the technical side, the study of the probabilistic and statistical properties of conditional models is notoriously difficult. For example, the probabilistic structure (existence of a stationary solution, dependence structure, tails, extremes, etc.) and statistical properties of parameter estimators of the popular GARCH process (Bollerslev [4], Taylor [47]) are by no means easy to derive and not in all cases well understood (see Berkes et al. [2], Mikosch [34], and the references therein). By contrast, the regression-type model (1) has been studied for several decades, and therefore a solid body of theoretical results and methodological recommendations on the estimation of the functions μ and σ is available in the statistical literature. Moreover, these results yield rigorous measures of the estimation error providing the frame for testing hypotheses on expected return and volatility.

Of course, besides its benefits any method has its own costs. The non-parametric estimation approach (to which our methodology belongs) requires few assumptions on the nature of the dynamics in the data. However, it can be highly data-intensive, generally not efficient for smaller sample sizes and prone to over-fitting. The first two possible drawbacks are easily avoided in the present set-up since long time series of returns are often available. To rule out the third one, careful out-of-sample performance evaluations have been conducted in two related studies. With a closely related non-parametric methodology Drees and Stărică [12] produce out-of-sample forecasts of the conditional distribution of tomorrow's return on the S&P 500 composite index as well as of returns over longer time horizons. These forecasts clearly outperform those obtained from conventional parametric GARCH-type models. In Herzel et al. [26] a multivariate extension of model (1) is introduced. It produces significantly better out-of-sample distributional forecasts of the trivariate distribution of the returns on the foreign exchange rate Euro/Dollar, the FTSE 100 index, and the 10 year US T-bond than the industry standard RiskMetrics.

One of the main findings of our analysis is that the point estimate of the *market price of risk*, i.e., the ratio of expected excess return to volatility, is subject to changes varying between -20% and 30% (annually) with a tendency towards positive values (see Figure IV.2). Its sign indicates the type of market (bear/bull). Moreover, we find that expected return is of the same (or lower) order of magnitude as volatility. This fact implies wide confidence bands around the point estimates of expected return or market price of risk[9] (see Section IV). As a consequence, the assumption of constant expected return or constant market price of risk cannot be rejected for periods as long as fifteen years (1985-2000 is the longest example). On the other hand, our analysis individuates *statistically significant* variations of the market price of risk. In particular, we document the existence of periods of significantly positive/negative price of risk. Shortly, we find that the estimated expected return (market price of risk) changes level and sign significantly over time. Our analysis uncovers an almost perfect coincidence between the periods of negative (positive) estimated expected return and the bear (bull) markets of the S&P 500 index as defined in Klein and Niermira [29] and Pagan and Sossounov [41]. Moreover, our analysis also reveals a complex dynamic relationship risk-expected return. While high volatility is typical for many bear markets, it does not seem to characterize them (see Section III for details).

The paper is organized as follows. In Section I we discuss some results from the statistical theory of non-parametric heteroscedastic regression which motivates our methodology. Then

[9]Our approach is no remedy for the low ratio signal-to-noise in the estimation procedure for the expected return.

we consider its implications for the simultaneous estimation of volatility and expected return, and we also discuss models for the distribution of the noise sequence (ε_t). Sections II, III and IV contain detailed statistical analyzes and their interpretation of the returns on the S&P 500 composite stock index between January 3, 1950 and June 15, 2003. In Section V the goodness of fit of the model (1) is checked through a careful analysis of the marginal distribution and dependence structure of the estimated noise sequence. We conclude with Section VI where we summarize our findings.

I. The statistical estimation procedure

In this section relevant results from the statistical theory of kernel curve estimation are discussed. Our main reference in this context is Müller and Stadtmüller [37] on estimation in the heteroscedastic regression model

$$R_{k,n} = \mu(t_{k,n}) + \sigma(t_{k,n})\, \varepsilon_{k,n}, \qquad k = 1, 2, \dots, n. \tag{2}$$

We omit indices n whenever feasible and assume that the design is fixed at $t_k = k/n$, $t_k \in [0,1]$. The returns R_k are observations of the unknown regression function, the possibly time-varying expected return, $\mu(t) : [0,1] \to \mathbb{R}$ contaminated with heteroscedastic errors $\sigma(t_k)\varepsilon_k$. The sd of the errors is the volatility of the market, that is also possibly time-varying. The sequence (ε_k) is iid with mean zero and unit variance, but not necessarily Gaussian. The functions $\mu : [0,1] \to \mathbb{R}$, the expected return, and $\sigma : [0,1] \to \mathbb{R}_+$, the volatility, are assumed smooth. The smoothness hypothesis incorporates the intuitive assumption that the time evolution of the expected return and volatility of the data is slow due to the aggregational nature of the index. This model is a reformulation of (1) in the standard set-up of statistical curve estimation: the observational period (in the case of the S&P 500, fifty years of data) is simply rescaled to the unit interval. Notice that modeling returns over different periods (daily, weekly, monthly, etc.) yields expected return and volatility functions depending on the sampling frequency. The empirical investigation of Sections III and IV indicates that, while the level of the two functions varies according to the frequency of the observations, the overall shape does not change significantly; see Figure IV.5.

The smoothing technique employed in this paper is kernel regression. For an introduction on smoothing estimators and, in particular, on kernel estimators, see Section 12.3 of Campbell et al. [9] or Wand and Jones [48]. The following kernel estimator will be used in the various steps on mean and variance estimation in the heteroscedastic regression model (2):

$$\widehat{f}(t;h) = \sum_{k=1}^{n} W_k(t)\, U_k \,, \tag{3}$$

where the random variables U_k will be specified in the corresponding sections on estimation of μ, σ and μ/σ. The deterministic weights $W_k(t)$ are given by

$$W_k(t) = W_{k,n}(t) = \frac{1}{h}\int_{s_{k-1}}^{s_k} K\left(\frac{t-u}{h}\right) du\,, \quad s_k = \frac{t_{k-1}+t_k}{2}\,. \tag{4}$$

They depend on the bandwidth $h > 0$ of the kernel function K on $[-1, 1]$. The latter satisfies the basic condition $\int K(u)du = 1$ and some further assumptions which are satisfied, for example, by the celebrated Epanechnikov kernel which is used in our analysis.

I.A. **Estimation of the volatility.** We summarize some of the necessary theory for the estimation of $\widehat{\sigma}$ in the heteroscedastic model (2). The kernel estimator (3) of $\sigma(t)$ in the heteroscedastic regression model (2) is defined in two steps. *First*, a preliminary smoothing is conducted in order to remove the expected value function μ in (2) in some neighborhood of $t_k \in (0,1)$. The preliminary estimator of the squared volatility t_k is given by

(5) $$\widetilde{\sigma}^2(t_k) = \left(\sum_{j=-m_1}^{m_2} w_j \, R_{j+k} \right)^2 ,$$

where the weights w_j satisfy the conditions $\sum_{j=-m_1}^{m_2} w_j = 0$ and $\sum_{j=-m_1}^{m_2} w_j^2 = 1$ for some fixed $m_1, m_2 \geq 0$.

The initial estimates $\sigma^2(t_k)$ of the squared volatility are viewed as measurements from the following regression model:

(6) $$\widetilde{\sigma}^2(t_k) = \sigma^2(t_k) + \widetilde{\varepsilon}_k, \quad 1 \leq k \leq n,$$

where the errors $\widetilde{\varepsilon}_k$ form an $m_1 + m_2$-dependent sequence and $E\widetilde{\varepsilon}_k = 0$.

In the *second* step, the estimator of the squared volatility is obtained by choosing $U_k = \widetilde{\sigma}^2(t_k)$ in (3)

(7) $$\widehat{\sigma}^2(t) = \widehat{\sigma}^2(t; h_{\sigma^2}) = \sum_{k=1}^{n} W_k(t) \, \widetilde{\sigma}^2(t_k) \, ,$$

where the weights $W_k(t)$ are defined in (4).

The smoothness assumptions together with the statistical properties of the estimator $\widehat{\sigma}^2$ are specified in the Appendix.

I.B. **Estimation of the expected return.** The mean estimation in (2) can be approached in two different ways. The first is direct estimation in the heteroscedastic set-up (2) which yields $\widehat{\mu}_{\mathrm{He}}(t; h_\mu)$, the estimator given by (3) with $f := \mu$. The smoothness assumptions on μ and the asymptotics of the estimator $\widehat{\mu}_{\mathrm{He}}(t; h_\mu)$ are given in Appendix.

The second is a two-step procedure where we first standardize the returns with the estimated volatility function $\widehat{\sigma}$ in (7) and then estimates the ratio μ/σ in the approximative homoscedastic regression set-up

$$\frac{R_k}{\widehat{\sigma}(t_k)} = \frac{\mu(t_k)}{\widehat{\sigma}(t_k)} + \frac{\sigma(t_k)}{\widehat{\sigma}(t_k)}\varepsilon_k \,. \tag{8}$$

The estimator $\widehat{\mu}_{\mathrm{Ho}}$ of the expected returns is obtained by multiplying the estimated ratio μ/σ by $\widehat{\sigma}$. Note that in this step we do not distinguish between $\sigma(t_k)$ and its estimate $\widehat{\sigma}(t_k)$[10]. In particular, we identify (8) with the truly homoscedastic set-up

$$\frac{R_k}{\widehat{\sigma}(t_k)} = \frac{\mu(t_k)}{\widehat{\sigma}(t_k)} + \varepsilon_k \,. \tag{9}$$

The ratio μ/σ is estimated in the homoscedastic regression set-up (9) using the kernel estimator introduced in (3). Note that, the results for heteroscedastic regression set-up presented in the Appendix apply also to the homoscedastic case taking $\sigma^2(t) \equiv 1$.

I.C. **Modeling the distribution of the noise.** The *residuals* from the model (2) are given by

$$\hat{\varepsilon}_t = \frac{R_t - \widehat{\mu}(t)}{\widehat{\sigma}(t)}, \tag{10}$$

with $\widehat{\mu}(t)$ and $\widehat{\sigma}(t)$ denoting any of the estimators for the expected return and the volatility. In order to avoid further model assumptions, one might be tempted to use the empirical

[10]The main reason for this identification is that developing a statistical estimation theory in the regression model (8), with dependent noise $(\sigma(t_k)/\widehat{\sigma}(t_k)\varepsilon_t)$ would be quite a daunting task. Besides, a comparison between the heteroscedastic estimate $\widehat{\mu}_{\mathrm{He}}$ and the results of the homoscedastic inference based on (8) shows a rather close match; see Figure IV.5, Section IV. An identification of the volatility with its estimate does not substantially affect the homoscedastic mean inference.

distribution function of the residuals as an estimate of the distribution function of the innovations. However, statistical evidence shows that the distribution of the innovations can be rather heavy-tailed. Thus, using the empirical distribution would underestimate the risk of extreme innovations and, hence, the probability of extreme returns.

A flexible and parsimonious family of distributions for the noise in model (2) was introduced in Herzel et al. [26]. It allows for asymmetry between the positive and negative noise and for heavy tails of Pareto type. Start from the Pearson type VII distribution with shape parameter m and scale parameter c whose density is defined on the positive real line:

$$f(x; m, c) = \frac{2\Gamma(m)}{c\Gamma(m-1/2)\pi^{1/2}}\left(1 + \left(\frac{x}{c}\right)^2\right)^{-m}. \tag{11}$$

Note that f is the t-density with $\nu = 2m - 1$ degrees of freedom multiplied by the scale parameter $c\nu^{-1/2}$.

Judging from our experience, the density f fits the positive noise and the absolute values of the negative noise quite nicely. Assuming that the distribution of the noise has median 0 and denoting the densities of the negative and positive standardized innovations by $f_-(x) = f(x; m_-, c_-)$ and $f_+(x) = f(x; m_+, c_+)$, respectively, we propose the following density for the noise ε_t:

$$f^{\mathrm{VII}}(x; m_-, c_-, m_+, c_+) = \frac{1}{2}\Big(f_-(x)1_{(-\infty,0)}(x) + f_+(x)1_{[0,\infty)}(x)\Big). \tag{12}$$

We refer to the corresponding distribution as the *asymmetric Pearson type VII distribution* and denote its distribution function by F^{VII}. We mention that Markowitz and Usmen [31, 32] in their attempt to find realistic distributions fitting stock returns[11] suggested the Pearson

[11]A part of their analyzes is also based on evidence from the S&P 500 index.

type IV distribution as most appropriate. The Pearson type IV and VII densities are close variations on the same theme.

II. Statistical analysis of the S&P 500 index: the set-up

We perform a detailed analysis of the returns $R_t = (P_t - P_{t-1})/P_{t-1}$ on the S&P 500 composite stock index between January 1950 and June 2003; P_t denotes the index at day t. The goal is to estimate the functions μ and σ assuming the heteroscedastic regression model (2). Moreover, we want to evaluate the goodness of fit of this model, judging from the distribution and dependence structure of the residuals.

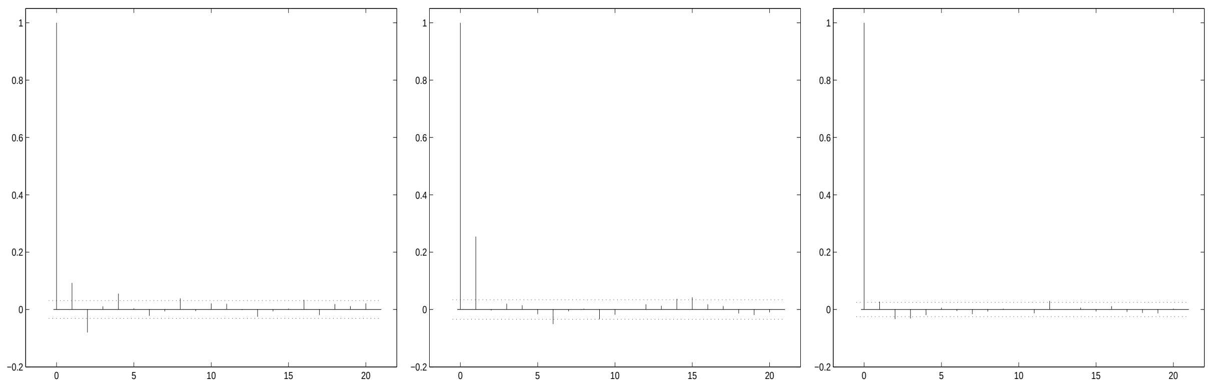

Figure II.1. *Sample ACFs for the three periods:* 1950-1965 (left), 1965-1979 (center) *and* 1979-2003 (right). *The linear dependence of the data indicated by a significant non-zero value at lag* 1 *disappears in the last period.*

II.A. **Choice of sampling frequency.** A glance at the sample autocorrelation function (ACF) of the daily returns on the S&P 500 index for different time periods unveils an evolving linear dependence. Figure II.1 displays the sample ACFs in the periods 1950-1965, 1966-1978 and 1979-2003. The first two periods are characterized by a certain degree of linear dependence as indicated by the non-zero values of the sample autocorrelations at the first few lags. This dependence disappears in the period 1979-2003[12]. Since we want to keep simple the modeling of the *whole* period 1950-2003 we have chosen to conduct our investigation on two-day returns. Figure II.2 shows that the sample of two-day returns exhibits negligible linear dependence.[13]

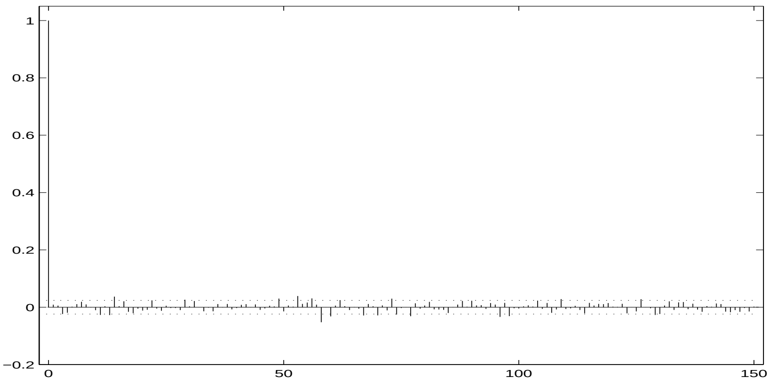

Figure II.2. *Sample ACF of the two-day returns. A comparison with Figure II.1 shows that the* 1*-lag dependence present in the daily returns disappears.*

II.B. **Asymptotically optimal bandwidths.** The issue of selecting the bandwidth h in (3), (4) is central in the non-parametric kernel smoothing methodology. Too small (large) a bandwidth h produces undersmoothed (over-smoothed) estimates of the function.

Equations (19) and (20) yield the asymptotic mean square error (MSE) and the asymptotic integrated square error (MISE) of $\widehat{\mu}_{\mathrm{He}}(t)$, the estimator (2) of μ as defined in Section I.B. The asymptotic bias and variance of the kernel estimator of μ/σ in the homoscedastic regression

[12]A possible explanation is that the CBOT started trading future contracts on the S&P 500 index in 1979. The introduction of this financial instrument might have improved the efficiency of the market.

[13]For the period 1979-2003, due to the absence of linear dependence, one could run the analysis based on daily returns.

(9) are similar to those in (19) and (20) (see Gasser et al. [16]). In a unified notation, the two errors are given by

$$MSE^{\widehat{f}}(t) = h_f^4\, B^2 (f''(t))^2 + \frac{\sigma^2(t)}{nh_f} V, \tag{13}$$

$$MISE^{\widehat{f}} = h_f^4\, B^2 \int (f''(u))^2 du + \frac{\int \sigma^2(u)du}{nh_f} V,$$

where the functions to be estimated are $f = \mu$ in the case of the heteroscedastic regression (2) and $f = \mu/\sigma$ for the homoscedastic regression (9). For the homoscedastic regression (9), $\sigma \equiv 1$. Minimization of MSE (MISE) with respect to the bandwidth h_f yields the locally (globally) optimal bandwidth

$$h_f^{(l)}(t) = \left(\frac{\sigma^2(t)V}{4nB^2(f''(t))^2} \right)^{1/5}, \quad h_f^{(g)} = \left(\frac{\int \sigma^2(u)duV}{4nB^2 \int (f''(u))^2 du} \right)^{1/5}. \tag{14}$$

Due to the importance of the bandwidth choice, we applied a set of different methods of bandwidth selection both for mean and variance estimation. We obtained bandwidths by cross-validation and a plug-in method, and we experimented with locally and globally optimal bandwidths; see the discussion in the next sections.

II.C. **Edge effects.** The importance of adequately treating the boundary $t \in [0, h)$ and $t \in (1 - h, 1]$ of a regression design defined on [0,1] has been repeatedly stressed in the literature of statistical curve estimation[14]. A wide body of work exists on how to overcome the boundary bias problem; see Wand and Jones [48]. In our analysis we chose to use the simple, practical method proposed in Hall and Wehrly [20]. The method is attractive in

[14]There exists a discrepancy between the order of magnitude of the bias in the interior and near the boundary. This phenomenon is usually referred to as a *boundary bias* problem. This leads to an optimal bandwidth of order $n^{1/5}$ in the interior of the interval [0,1], while near the boundary the optimal bandwidth is of the order $n^{1/3}$.

that it provides a simple way of extending traditional techniques of bandwidth selection to an entire design interval. In a nutshell, the method can be described as follows. Using a one-sided kernel, L, estimate the values of the regression mean at the extreme left and right ends of the design interval. Then reflect the entire data set in each of these points to obtain a new data set three times the extent of the old one. Finally, estimate the mean by a regular kernel estimator *over the original design interval* but using the new data set, which combines the original data with the new set of pseudo-data. The method allows the bandwidth for both the preliminary edge kernel estimators and the final kernel estimator to be estimated automatically by the cross-validation algorithm. It is shown in Hall and Wehrly [20] that the difference between the MISE of the estimator based on pseudo-data and that of a hypothetical (but unobtainable) estimator based on data from a larger interval is of the order $O(h^5)$. This is negligible relative to the whole MISE, which, if h is chosen optimally[15], is of the size $O(h^4)$.

[15]The steps in the choice of the bandwidth are as follows. For a given h the choice of bandwidth for the one-sided kernel estimation in the preliminary inference of the regression mean at the extreme left and right ends of the design interval is ch, where c is a constant that depends on the type of the kernel used (for the Epanechnikov kernel, $c = 1.86$). Using the kernel L and the bandwidth ch, estimate the regression mean at the two ends of the interval, $\widehat{f}(0)$, $\widehat{f}(1)$, respectively, and produce the pseudo-data by reflecting the data points interior to the design interval in each of the two points $(0, \widehat{f}(0))$ and $(1, \widehat{f}(1))$. Calculate a version $\widetilde{CV}(\mathrm{h})$ of the classical leave-one-out cross-validation $CV(\mathrm{h})$ (16), over the new pseudo-data set, producing $\widehat{f_i}(t)$ in (16) by leaving out not only the observation U_i but also those that are obtained from it through reflection. Choose h to minimize $\widetilde{CV}$. Under a set of usual conditions on the smoothness of the function to be estimated and the tails of the error distribution, the authors show that, with probability 1, their version of leave-one-out cross-validation

$$\begin{aligned} \widetilde{CV}(h) &= \sum_{i=1}^{n}(\widehat{f}(t_i;h) - f(t_i))^2 + \sum_{i=1}^{n} \varepsilon_i^2 w(t_i) + o(n^{1/5}) \\ &= \int_0^1 E(\widehat{f}(u;h) - f(u))^2 du + \sum_{i=1}^{n} \varepsilon_i^2 + o(n^{1/5}) \end{aligned} \tag{15}$$

uniformly in $h \in H := \{An^{1/5} \leq h \leq Bn^{1/5}\}$, for any $0 < A < B < \infty$. Hence, minimizing $\widetilde{CV}(h)$ is asymptotically equivalent to minimizing $\int_0^1 E(\widehat{f}(u) - f(u))^2 du$ and it produces a bandwidth $\widehat{h}$ that satisfies $\widehat{h}/h_f^{(g)} \to 1$ with probability 1, where $h_f^{(g)}$ is the MISE asymptotically optimal bandwidth defined in (14).

III. Statistical analysis of the S&P 500 index: volatility estimation

We start with the estimation of the *squared volatility function* $\sigma^2(t)$ as described in Section I.A. For the preliminary estimate $\widetilde{\sigma}^2$ in (5) we have chosen $m_1 = 1, m_2 = 0$ with corresponding optimal (in the sense of Müller and Stadtmüller [37]) weights $w_{-1} = 1/\sqrt{2}$ and $w_0 = -1/\sqrt{2}$. Other choices for m_1, m_2 are possible. However, the larger the window $[j - m_1, j + m_2]$ the more dependency one introduces in the data. For this reason, we prefer a small window which leads to moving averages of two two-day returns in the definition of $\widetilde{\sigma}^2$.

III.A. **Bandwidth selection.** Cross-validation[16] was used for choosing h_{σ^2} in estimation of $\sigma^2(t)$ in (6). However, caution is required since the errors $\widetilde{\varepsilon}_k$ in (6) form a 1-dependent sequence[17]. One could follow the general methodology for incorporating covariance estimates into the choice of the bandwidth proposed by Hart [22]. An easier alternative is available due to the special dependency structure at hand. Since the sequence $(\widetilde{\varepsilon}_k)$ is assumed 1-dependent, each of the sequences $(\widetilde{\sigma}^2(t_{2k}))$ and $(\widetilde{\sigma}^2(t_{2k+1}))$, $k = 1, \ldots, [n/2]$ consists of independent random variables and therefore standard cross-validation is applicable. To obtain the optimal

[16]Cross-validation is a method which is based on the minimization of the residual mean squared error and it is frequently used to infer the optimal smoothing parameter. With the notation used in (3), define

$$\widehat{f}_i(t) = \sum_{k \neq i} W_k(t) U_k, \quad CV(h) = \sum_{i=1}^{n} (\widehat{f}_i(t_i) - U_i)^2 w(t_i), \tag{16}$$

where w is a weight function. The cross-validation approach chooses the bandwidth that minimizes the function $h \to CV(h)$.

[17]It is well-known (see Diggle and Hutchinson [11], Hart and Wehry [23], Hart [22]) that the traditional form of cross-validation procedure fails when data are correlated.

global bandwidth for the original sample of size n, a correction with a factor of $2^{1/5}$, motivated by the asymptotic theory, is needed.

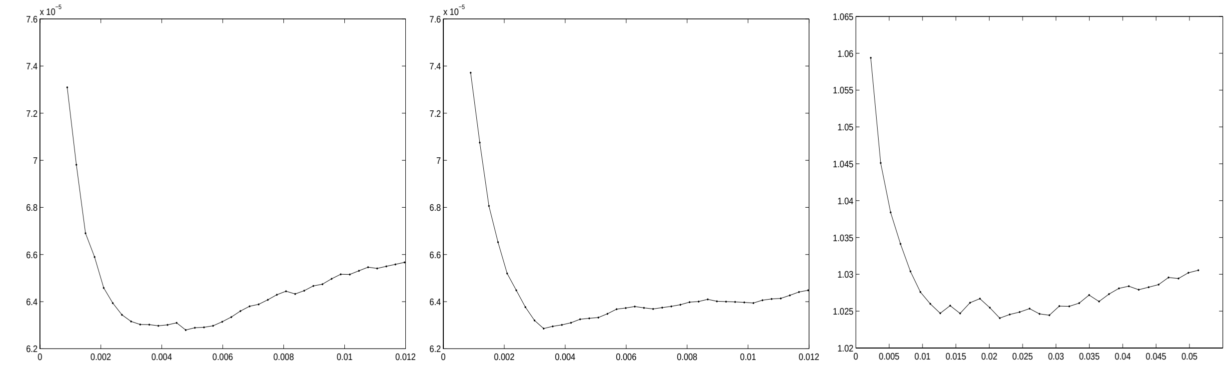

Figure III.1. Left, center: *The cross-validation graphs for the choice of the bandwidth* $h^{(c)}_{\sigma^2}$ *for* $\widehat{\sigma}^2$ *in* (7). *The bandwidths* h_{σ^2} *minimizing the cross validation functions fall in the interval* $[0.003, 0.005]$. Right: *The cross-validation graph for the choice of the bandwidth* $h_{\mu/\sigma}$ *in the homoscedastic regression model* (9). *The minimizing bandwidth* $h^{(c)}_{\mu/\sigma} = 0.021$ *is chosen.*

On the left and in the center of Figure III.1 the cross validation graphs for the two subsamples are displayed. Other methods of bandwidth selection (plug-in, the bandwidth choice suggested in Drees and Stărică [12][18]) produced practically identical results.

[18]There the choice of the bandwidth is based on the sample ACF of the absolute standardized noise. The sample ACF of the absolute values of the returns centered by the sample mean displays almost constant autocorrelations even at large lags. This phenomenon can be explained by changes in the volatility; see Mikosch and Stărică [35, 36] for a theoretical explanation. Given the heteroscedastic regression model, the centered (by the sample mean) and scaled (by the volatility) residuals which correspond to the optimal choice of bandwidth should be "almost" independent. Hence the bandwidth is chosen such that the sample ACF of the absolute values of the returns is negligible at all lags.

III.B. **Discussion of the estimation results.** Figure III.2 displays the estimate $\widehat{\sigma}^2$ in (7) where the bandwidth $h^{(c)}_{\sigma^2} = 0.0045$ was used. The graph indicates that during the '60s and early '70s the bear markets of the S&P 500 composite index (1961, 1966, 1968, 1971 and 1973) were periods of increasing volatility compared to the preceding and succeeding bull markets. For the '50s and the period after the 1973-1974 oil crisis the connection between the type of market and volatility level is not so clear-cut. The 1983 bear market had roughly the same level of volatility as the following bull market. The extremely low volatility in the beginning of the long bull market that covered the second half of the '90s was followed by high volatility at the end of the decade. The high level attained towards the end of the bull market continued, seemingly without further augmentation, also during the bear market of 2000. We conclude that, *while high volatility is typical for many bear markets, it does not seem to characterize them.* We will see soon that the relation of the type of market (bear or bull) to expected return is more clear-cut than that to volatility .

IV. Statistical analysis of the S&P 500 index: estimation of expected return

We continue with the estimation of the expected return function $\mu(t)$ as described in Section I.B.

IV.A. **Bandwidth selection.** Since the choice of bandwidth is crucial in our approach we applied different methods of bandwidth selection both in the homoscedastic and the heteroscedastic frameworks. Cross validation has already been mentioned. Another method builds on inferring the asymptotically optimal (local or global) bandwidth (14) from the data by replacing the residual variance and the asymptotic expression of the bias (19) by sample estimates. Such selection rules are called 'plug-in' estimators. For $\sigma^2(t)$, in the case

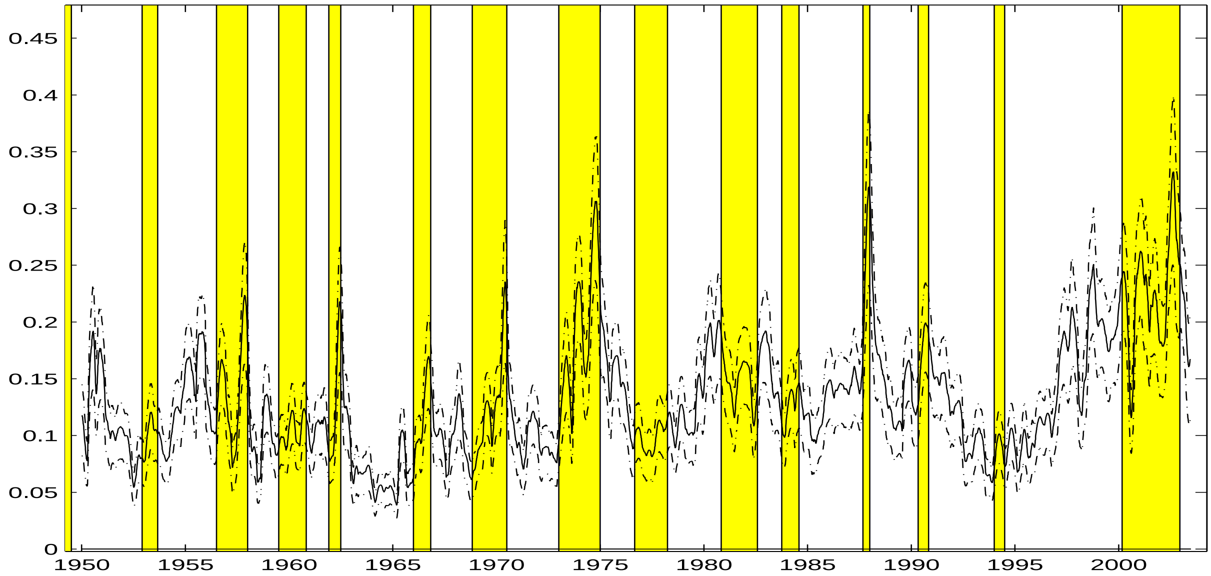

Figure III.2. *The estimated volatility function* $\widehat{\sigma}(t; h^{(c)}_{\sigma^2})$ (solid line) *with* 95% *asymptotic confidence bands of the two-day returns of the* S&P 500 *with bandwidth* $h^{(c)}_{\sigma^2} = 0.0045$. *Bear markets are shown in a shade of grey. High volatility is typical for many bear markets, but does not seem to characterize them.*

of the heteroscedastic regression, the estimator (7) is used. The functional that quantifies bias is approximated by the integrated squared second derivative of the regression function. It is determined by an iterative procedure introduced in Gasser et al. [17] based on a kernel estimator $\widehat{f''}(t; h_{f''})$ for the second derivative (the integrals are easily obtained from the point estimates). Such an estimator has the form (3) with the kernel K tailored to estimate second derivatives; see Gasser et al. [18] (we used the optimal (2,4)-kernel)[19].

[19]The iteration procedure goes as follows. For a given k, based on asymptotic theory, the bandwidth $\widehat{h}^{\text{opt}}_{f,k}$ yields a value for $\widehat{h}^{\text{opt}}_{f'',k} = \widehat{h}^{\text{opt}}_{f,k}\, n^{1/10}$. This is the bandwidth to be used in the estimation of $f''(t)$ with sample size n. The estimated function $\widehat{f''}(t; \widehat{h}^{\text{opt}}_{f'',k})$ is then used in (14) to produce the next bandwidth $\widehat{h}^{\text{opt}}_{f,k+1}$ for estimation of $f(t)$. The iterative procedure quickly converges to the asymptotically optimal bandwidth both in theory and practice. A

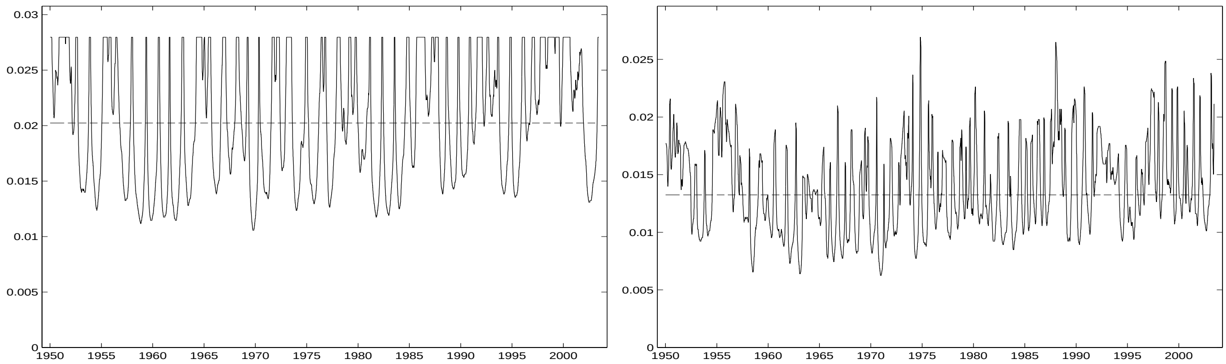

Figure IV.1. *The local bandwidth* $h^{(l)}_{\mu/\sigma}(t)$ (left) *and* $h^{(l)}_{\mu}(t)$ (right) *obtained by using the iterative method for automatic smoothing of Gasser et al.* [17]. *The dotted lines represent the median local bandwidth of* 0.020 *and* 0.013 *respectively.*

IV.B. **Homoscedastic regression.** . Once we estimated the volatility, the homoscedastic regression model yields the estimate $\widehat{\mu}_{\mathrm{Ho}}$ of the expected return as outlined in Section I.B. For the ratio μ/σ in (9), the cross-validation function, displayed in the right-hand graph of Figure III.1 attains a minimum of $h^{(c)}_{\mu/\sigma} = 0.021$ with a plateau covering the interval [0.012, 0.025] [20]. The iterative method of Gasser et al. [17] yields an estimate of the globally optimal bandwidth of $h^{(g)}_{\mu/\sigma} = 0.017$ in (14) and a locally optimal bandwidth, $h^{(l)}_{\mu/\sigma}$ in (14) with median 0.020; see Figure IV.1 (left graph).

IV.C. **Heteroscedastic regression.** In the heteroscedastic framework the expected return is estimated by $\widehat{\mu}_{\mathrm{He}}(t; h_\mu)$ as explained in Section I.B. The plug-in approach to global and local bandwidth selection was implemented (no results about the cross validation are available

theoretical large sample analysis shows that the plug-in estimator is attractive in terms of variability, with a relative rate $O_p(n^{-1/2})$ for smooth functions. In contrast, cross-validation leads to a relative rate $O_p(n^{-1/10})$.

[20]The use of penalizing methods leads to a similar choice.

in this context). The procedure yields $h_\mu^{(g)} = 0.012$ and a function $h_\mu^{(l)}(t)$ that is displayed in Figure IV.1 (right graph), median value, 0.013.

As a conclusion, we see that all methods employed for bandwidth selection in the homoscedastic/heteroscedastic set-ups produce comparable results. In particular, the global optimal bandwidth for μ/σ seems to be in the range $[0.017, 0.021]$, while that for μ seems to belong to the interval $[0.012, 0.013]$.

IV.D. **Discussion of the estimation results.** In Figure IV.2 we exhibit the estimated market price of risk $(\widehat{\mu}_{\mathrm{Ho}}(t; h_{\mu/\sigma}^{(l)}) - r_t^f)/\widehat{\sigma}(t)$, inferred in the set-up of the homoscedastic

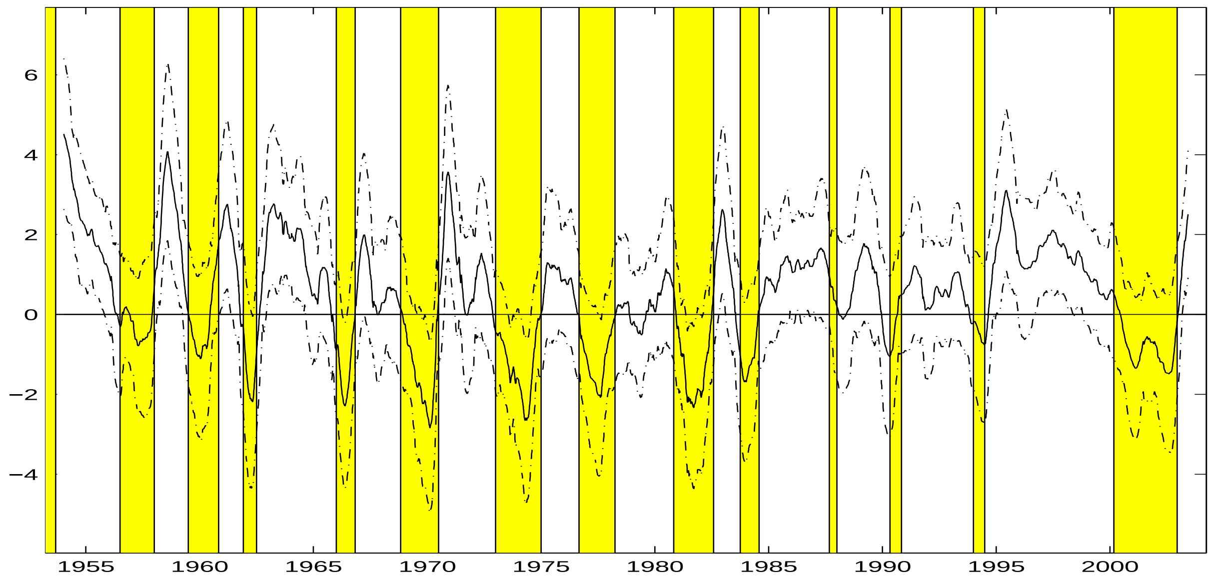

Figure IV.2. *Kernel curve estimation of the market price of risk* (solid line) *of the two-day returns of the* S&P 500 *index together with* 95% *asymptotic confidence bands. The estimates are based on homoscedastic non-parametric regression; see Section I.B for details on estimation. The periods of a negative market price of risk mostly coincide with those of a bear market shown in a shade of grey. See also Section IV.*

regression (9) (r_t^f denotes the risk-free interest rate which, in the case of our analysis is taken to be the return on the three month US Treasury bill).

In Figure IV.3 the homoscedastic expected return $\widehat{\mu}_{\text{Ho}}(t; h^{(l)}_{\mu/\sigma})$ is displayed. The bandwidth $h^{(l)}_{\mu/\sigma}$ is the locally optimal bandwidth displayed in Figure IV.1 (left graph).

The asymptotic confidence bands in Figures IV.2 and IV.3 were calculated by using the asymptotic formula of the variance of the kernel estimator similar to (20); see Gasser et al. [16]. The 95% asymptotic confidence bands for the market price of risk are $(\widehat{\mu}_{\text{Ho}} - r_t^f)/\widehat{\sigma}(t) \pm 1.96\sqrt{V/(nh^{(l)}_{\mu/\sigma})}$ and were obtained by assuming $\sigma^2(t) \equiv 1$ in (20). The 95% asymptotic confidence bands for the expected return are given by $\widehat{\mu}_{\text{Ho}}(t) \pm 1.96\,\widehat{\sigma}(t)\sqrt{V/(nh_{\mu/\sigma})}$. The low signal-to-noise ratio in the estimation procedure of the expected return translates into rather wide confidence bands both for $\mu_{\text{Ho}}(t)$ and $(\mu_{\text{Ho}}(t) - r_t^f)/\sigma(t)$.

Figure IV.4 displays the kernel estimate $\widehat{\mu}_{\text{He}}(t; h^{(l)}_{\mu})$ of the expected return $\mu(t)$ in the heteroscedastic regression model (2) obtained by using the estimated locally optimal bandwidth $h^{(l)}_{\mu}(t)$; see Figure IV.1 (right graph)[21]. The confidence bands for $\widehat{\mu}_{\text{He}}(t)$ in Figure IV.4 were calculated using the asymptotic formula for the variance of the kernel estimator given in (20). The 95% asymptotic confidence bands for the expected return are given by $\widehat{\mu}_{\text{He}}(t) \pm 1.96\widehat{\sigma}(t)\sqrt{V/(nh^{(l)}_{\mu}(t))}$.

For comparison, Figure IV.5 presents both estimators $\widehat{\mu}_{\text{He}}(t; h^{(l)}_{\mu})$ and $\widehat{\mu}_{\text{Ho}}(t; h^{(l)}_{\mu/\sigma})$ of μ[22]. Figure IV.5 shows that the two different estimation procedures for the expected returns $\mu(t)$ lead to very similar results with the graph for $\widehat{\mu}_{\text{He}}$ having slightly deeper troughs and higher

[21]The kernel estimate of μ that uses the globally optimal bandwidth $h^{(g)}_{\mu} = 0.014$ is practically identical.

[22]The specific locally optimal bandwidth was used to produce the estimates.

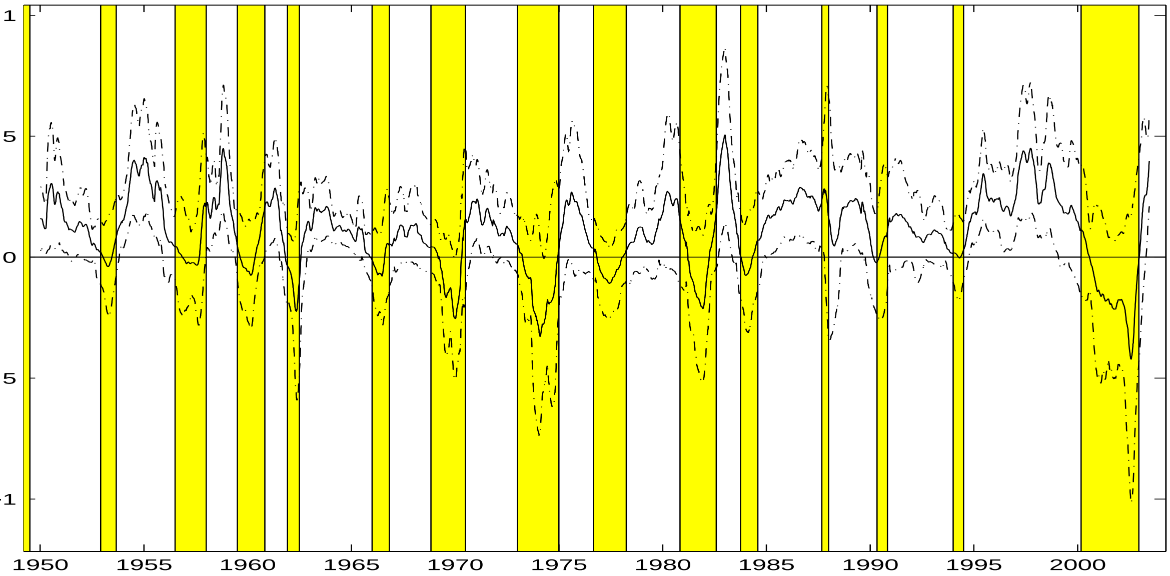

Figure IV.3. *Kernel estimate* $\widehat{\mu}_{\mathrm{Ho}}(t; h^{(l)}_{\mu/\sigma})$ (solid line) *with* 95% *asymptotic confidence bands of the expected excess two-day returns of the* S&P 500 *index. The bandwidth* $h^{(l)}_{\mu/\sigma}$*, the locally optimal one from Figure IV.1* (left graph). *Bear market periods are shown in a shade of grey; they coincide with the periods of negative expected excess return.*

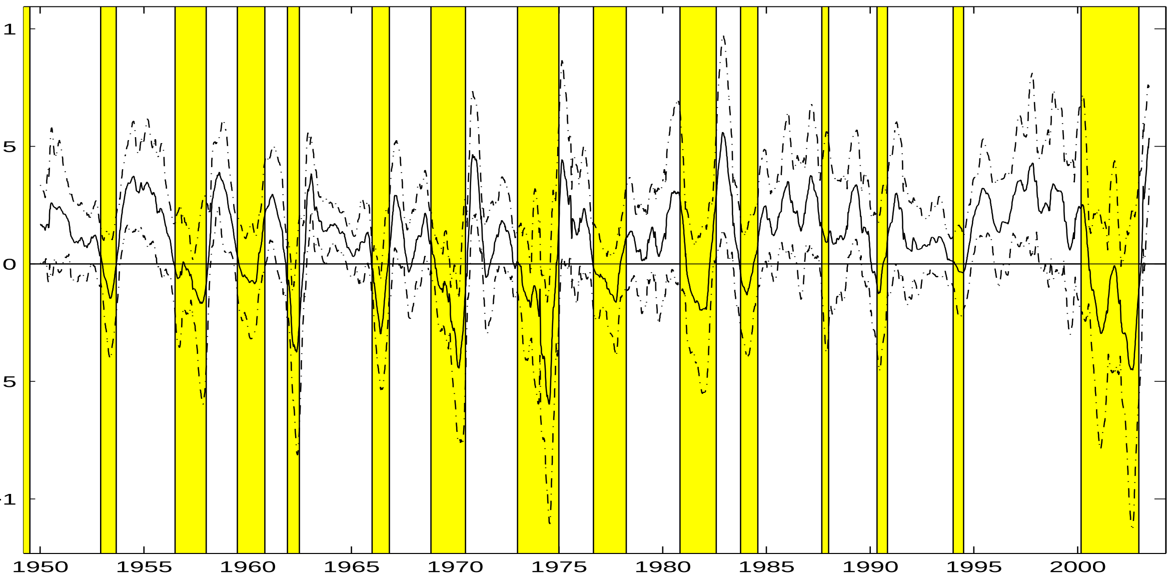

Figure IV.4. *The kernel estimate* $\widehat{\mu}_{\mathrm{He}}(t; h^{(l)}_{\mu})$ (solid line) *with* 95% *asymptotic confidence bands of the expected excess two-day returns of the* S&P 500 *index. The bandwidth* $h^{(l)}_{\mu}(t)$ *is the estimated locally optimal bandwidth; see Figure IV.1* (right graph). *Bear market periods are shown in a shade of grey; they coincide with the periods of negative expected returns.*

peaks than that for $\widehat{\mu}_{\mathrm{Ho}}$. The impact of the 1987 market crash on the volatility estimation is noticeable in the homoscedastic estimate $\widehat{\mu}_{\mathrm{Ho}}$.

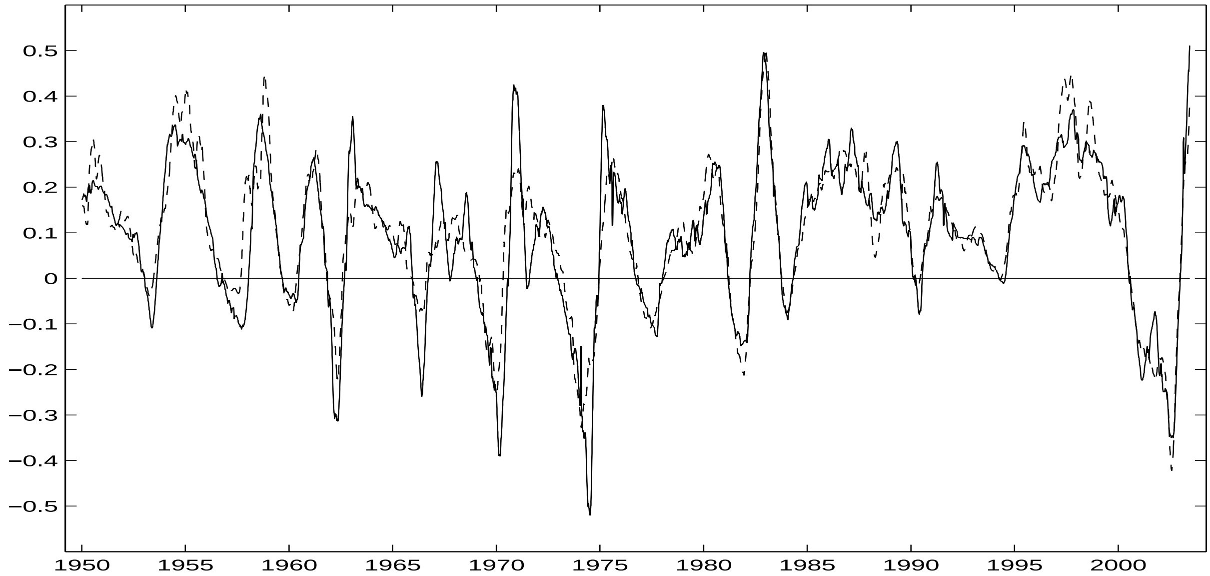

Figure IV.5. *The estimators* $\widehat{\mu}_{\mathrm{He}}(t; h_{\mu}^{(l)})$ (solid line) *and* $\widehat{\mu}_{\mathrm{Ho}}(t; h_{\mu/\sigma}^{(l)})$ (dashed line) *of the expected returns in the* S&P 500 *two-day return series are rather close to each other. The estimator* $\widehat{\mu}_{\mathrm{He}}$ *has slightly deeper troughs and higher peaks than* $\widehat{\mu}_{\mathrm{Ho}}$.

Although displaying slight differences, Figures IV.2, IV.3 and IV.4 give an overall similar picture. First, they show that the point estimates of the market price of risk or the expected return are subject to strong changes varying between -20% and 30% (annually), -40% and 40% respectively, with a tendency towards positive values. Second, they suggest that bear market periods for the S&P 500 composite index as defined by Klein and Niermira [29] and Pagan and Sossounov [41] are periods of negative point estimates of expected returns and often coincide with the periods of negative point estimate for the market price of risk. Third, we note that expected return is of the same (or lower) order of magnitude as volatility. This implies wide confidence bands around the point estimates of expected return or market price

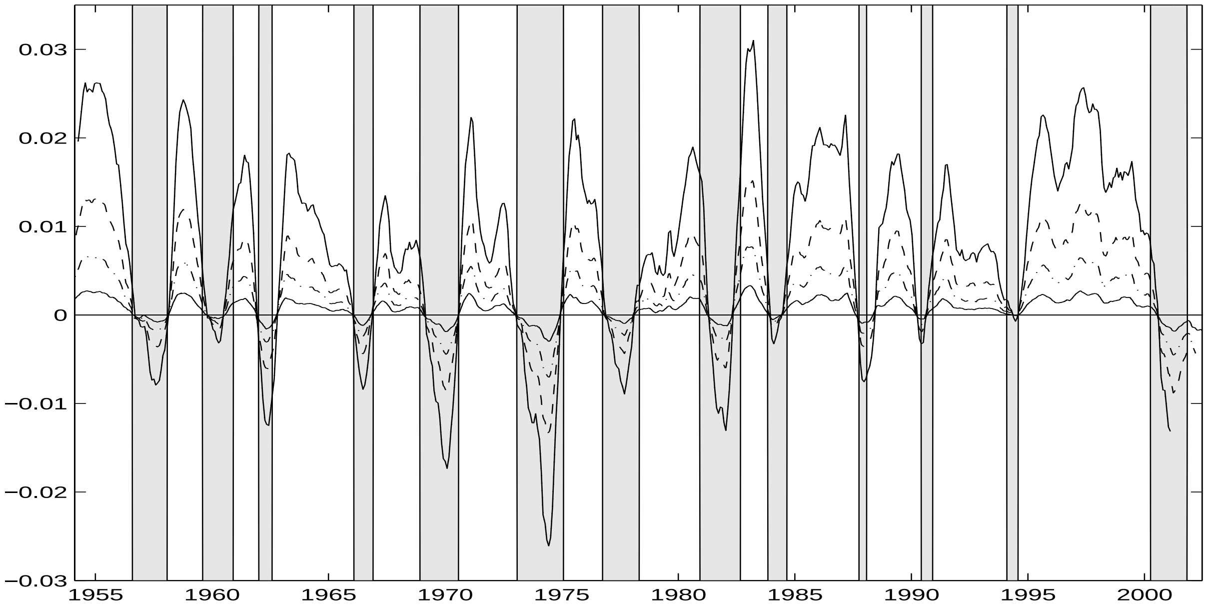

Figure IV.6. *The estimator* $\widehat{\mu}_{\mathrm{He}}(t; h_\mu^{(l)})$ *as a function of the return period.* From low to high: *two-day, weekly, two-week, monthly returns. While the level of the estimates varies according to the frequency of the observations the overall shape remains the same.*

of risk. As a consequence, the assumption of constant expected return or constant market price of risk cannot be rejected for periods as long as ten to fifteen years, depending on the estimate one uses. On the other hand, the graphs individuate *statistically significant* variations of the market price of risk and of the expected return. In particular, all methods document the existence of periods of significantly positive (at 95% level) price of risk. The estimated market price of risk in Figure IV.2 as well as the heteroscedastic estimate of the expected return in Figure IV.4 display also short periods of significantly negative (at 95% level) estimates. Shortly, we find that both the estimated expected return and the estimated market price of risk seem to change level and sign *significantly* over time.

Finally, Figure IV.6 displays the kernel estimates of expected returns for various sampling frequencies. The bandwith used was the locally optimal one given by the iterative method of Gasser et al. [17]. While the level of the expected return varies according to the frequency of the observations the overall shape remains the same.

V. Statistical analysis of the S&P 500 index: goodness of fit

In this section we are concerned with the goodness of fit of our model. Our first step is to show that the marginal distribution of the residuals $\widehat{\varepsilon}_t$ defined in (10) is nicely fitted by an asymmetric Pearson type VII distribution F^{VII}; see (12). Assuming the $\widehat{\varepsilon}_t$ are iid, maximum likelihood point estimation of its four parameters (asymptotic standard deviation in parentheses) yields:

$$(17)\quad \widehat{m}_- = 7.91 \quad (1.42), \quad \widehat{c}_- = 3.86 \quad (0.42), \quad \widehat{m}_+ = 16.79 \quad (2.68), \quad \widehat{c}_+ = 5.39 \quad (0.80)\,.$$

The estimated values $\widehat{m}_-$ and $\widehat{m}_+$ imply that the left tail of the underlying noise distribution has tail index 14.82 while the right one has tail index 32.59. Hence the distribution is asymmetric with moderately heavy tails. The asymmetry confirms the empirical observation that extreme negative stock returns are usually larger in absolute value than the largest positive return.

Next, we check the goodness of fit of the asymmetric Pearson type VII distribution F^{VII} with parameters (17) to the residuals. Assuming that $\widehat{\varepsilon}_t$ has exactly this distribution, $F^{\mathrm{VII}}(\widehat{\varepsilon}_t)$ has uniform distribution on $(0,1)$. Hence, writing Φ for the standard normal distribution function, $\Phi^{-1}(F^{\mathrm{VII}}(\widehat{\epsilon}_t))$ is standard normally distributed. The left-hand graph of Figure V.1 displays the normal probability plot of the transformed data $\Phi^{-1}(F^{\mathrm{VII}}(\widehat{\epsilon}_t))$. The resulting plot is very close to a straight line providing evidence that the parametric family of distributions with density (12) gives a nice fit to the noise (ε_t) in the heteroscedastic regression model (2).

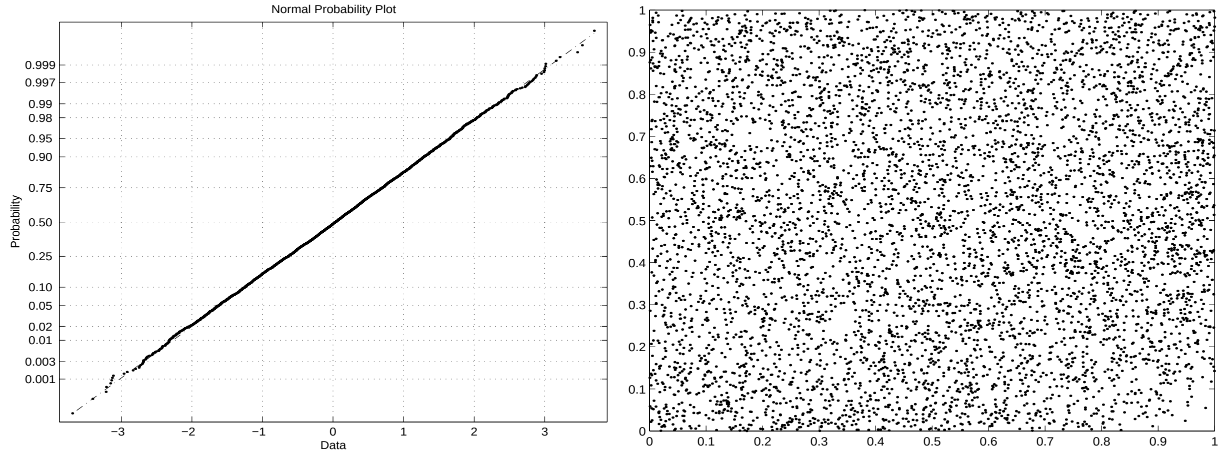

Figure V.1. Left: *The normal probability plot of the transformed noise* $\Phi^{-1}(F^{\mathrm{VII}}(\widehat{\varepsilon}_t))$ *is close to a straight line. This is a clear indication of the fact that the noise sequence in the heteroscedastic regression model is nicely fitted by an asymmetric Pearson type* VII *distribution.* Right: *Scatter plot of the pairs* $(F^{\mathrm{VII}}(\widehat{\varepsilon}_t), F^{\mathrm{VII}}(\widehat{\varepsilon}_{t+1}))$ *of transformed residuals. No clusters or patterns are visible in this plot. This is an indication of independence of the residuals.*

To check the appropriateness of the assumption of independence, i.e. to search for possible patterns of non-linear dependence, it is most useful to have a look at copulas of the pair $(\widehat{\epsilon}_t, \widehat{\epsilon}_{t+1})$. More concretely, the joint distribution of a pair of r.v. (U, V) is uniquely determined by the marginal distribution of the coordinates F_U and F_V and by their copula, i.e. the distribution on the unit square of $(F_U(U), F_V(V))$. Hence, it is the copula that provides the complete description of the dependency structure between the marginal random variables (see Nelsen [38]). Moreover, U and V are independent if and only if their copula is the uniform copula. Graphically, this corresponds to an uniform filling of the unit square by the pairs $(F_U(U), F_V(V))$. Hence, a simple but very informative way of assessing the independence of the coordinates of a bivariate random vector is looking at realizations of its copula. The appearance of an uniformly covered unit square supports the assumption of independence while the presence of patterns indicates dependency.

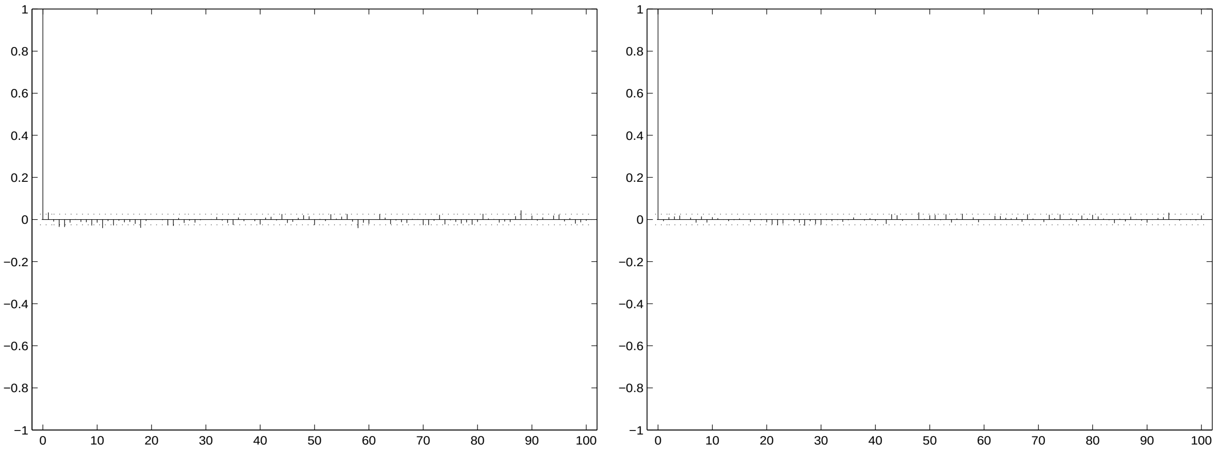

Figure V.2. *The sample ACFs of the transformed residuals* $\Phi^{-1}(F^{\mathrm{VII}}(\widehat{\epsilon}_t))$ (left) *and their absolute values* (right). *The sample ACFs vanish at all lags. This is a strong indication of independence of the noise* (ε_t).

To obtain the copula the pair $(\widehat{\epsilon}_t, \widehat{\epsilon}_{t+1})$, we transformed first the residuals into uniforms[23] and produced the scatter plot $(F^{\mathrm{VII}}(\widehat{\epsilon}_t), F^{\mathrm{VII}}(\widehat{\epsilon}_{t+1}))$ in the right-hand graph of Figure V.1. As mentioned, an uniform filling of the unit square is interpreted as evidence of *independent* components. The graph reveals only a very slight disinclination for particularly large values of residuals to be followed by particularly small values. This seems to indicate that the assumption of independent innovations provides a reasonable approximation for the dynamics of the data.

Figure V.2 displays the sample ACF of the residuals transformed into normal random variables $\Phi^{-1}(F^{\mathrm{VII}}(\widehat{\epsilon}_t))$ and of its absolute values. The ACFs plots in Figure V.2 vanish at

[23]The estimated asymmetric Pearson type VII distribution F^{VII} with parameters (17) were used to produce two samples of uniform rv, $(F^{\mathrm{VII}}(\widehat{\epsilon}_t), F^{\mathrm{VII}}(\widehat{\epsilon}_{t+1}))$.

all lags indicating that there is practically no linear dependence structure left in the residuals. The marginals being normal, the absence of correlations supports the modeling hypothesis of independence.

All the measures of goodness of fit that we analyzed show that the assumptions of the model 1 seem to fit the data and support the results of the non-stationary, non-parametric analysis presented in Section III and IV.

VI. Conclusions

In this paper we have tried to argue that

- Non-parametric curve estimation is a feasible technique for simultaneous estimation of expected return and volatility.
- A simple heteroscedastic regression (2) with iid noise is a suitable model for returns. This claim is supported by the properties of the residuals which mimic the behavior of an iid sample.
- In this context, the estimation of the time-varying expected return is subject to a high level of statistical uncertainty. Nevertheless, unconditional expected return and volatility seem to *significantly* change over time.
- Periods of negative point estimate of expected returns (negative market price of risk) in the S&P 500 index can be identified with the periods of bear market as defined in Klein and Niemira [29] and Pagan and Sossounov [41].
- Periods of high volatility in the S&P 500 do not necessarily coincide with those of bear markets.

A major task of future research is to investigate the feasibility of improving the accuracy of estimating expected return, i.e., the possibility of producing smaller confidence bands for μ

[24]. Ongoing research focuses on the estimation of $\mu(t)$ and $\sigma(t)$ based only on observations up to time t [25] as well as on extensions to the multivariate set-up.

Appendix

VI.A. **Statistical properties of the volatility estimator $\widehat{\sigma}^2(t)$.** In the sequel, we assume that σ^2 is twice differentiable with a continuous second derivative, μ is Lipschitz continuous of order $\alpha \geq 0.25$ and $E|\varepsilon_1|^{5+\epsilon} < \infty$ for some $\epsilon > 0$. Then the following statements can be derived from Theorem 3.1 and Remark at the bottom of p. 622 in Müller and Stadtmüller [37]:

(1) The estimated squared volatility $\widehat{\sigma}^2(t)$ satisfies

$$\left|\widehat{\sigma}^2(t) - \sigma^2(t)\right| \leq c\,(\log n/n)^{2/5},$$

almost surely, for some unspecified positive constant c, uniformly for $t \in [\delta, 1-\delta]$, any fixed $\delta \in (0,1)$, and the bandwidth is chosen as $h_{\sigma^2} \sim (\log n/n)^{1/5}$.

(2) The expected value $E\widehat{\sigma}^2(t)$ satisfies

$$|E\widehat{\sigma}^2(t) - \sigma^2(t)| \leq c\,(h_{\sigma^2}^2 + n^{-1})$$

[24]The results obtained by using local polynomial regression or splines techniques were not more encouraging than the ones presented which are based on the Müller and Stadtmüller [37] results.

[25]Estimates based only on past observations would provide sensible indicators of the current state of the market.

for some unspecified positive constant c, uniformly for $t \in [\delta, 1-\delta]$, any fixed $\delta \in (0,1)$.

(3) The variance $\mathrm{var}(\widehat{\sigma}^2(t))$ satisfies

$$
\begin{aligned}
\mathrm{var}(\widehat{\sigma}^2(t)) \quad &\sim \quad \frac{\sigma^4(t)}{nh}\, V \left(2 + (E\varepsilon_1^4 - 3) \sum_{j=-m_1}^{m_2} w_j^4\right) \\
&+ \frac{2\,V\,\sigma^4(t)}{n\,h}\Big((E\varepsilon_1^4 - 2) \sum_{i=1}^{m_1+m_2} \sum_{j,j-i\in[-m_1,m_2]} w_j^2 w_{j-i}^2 - (m_1+m_2) \sum_{j\in[-m_1,m_2]} w_j^4 \\
&+ 2 \sum_{i=1}^{m_1+m_2} \left(\sum_{j,j-i\in[-m_1,m_2]} w_j\, w_{j-i}\right)^2 \Big) \qquad (18)
\end{aligned}
$$

where $V = \int K^2(x)\,dx = 0.6$ for the Epanechnikov kernel used in our analysis. The derivation of the variance follows.

VI.B. **The variance of the estimator $\widehat{\sigma}^2(t)$.** Here we derive the asymptotic order of the variance in (18) under the standard conditions $n \to \infty$, $h = h_n \to 0$ and $n\,h \to \infty$. We have by the $m_1 + m_2$-dependence of the sequence $(\widetilde{\sigma}^2(t_k))$,

$$
\begin{aligned}
\mathrm{var}(\widehat{\sigma}^2(t)) \quad &= \quad \sum_{k=1}^{n} W_k^2(t)\, \mathrm{var}(\widetilde{\sigma}^2(t_k)) + \sum_{1\le k\ne l\le n} W_k(t)\, W_l(t)\, \mathrm{cov}(\widetilde{\sigma}^2(t_k), \widetilde{\sigma}^2(t_l)) \\
&= \quad \sum_{k=1}^{n} W_k^2(t)\, \mathrm{var}(\widetilde{\sigma}^2(t_k)) + 2\sum_{k=1}^{n} \sum_{1\le i\le \min(m_1+m_2,n-k)} W_k(t)\, W_{k+i}(t)\, \mathrm{cov}(\widetilde{\sigma}^2(t_k), \widetilde{\sigma}^2(t_{k+i})) \\
&= \quad I_1 + I_2\,.
\end{aligned}
$$

By virtue of Lemma 5.3(iii) in Müller and Stadtmüller [37], using the continuity of $\mathrm{var}(\widehat{\sigma}(t))$ and the fact that $(nh)\sum_{k=1}^{n} W_k^2(t) \sim \int K^2(x)\,dx = V$,

$$
I_1 \sim \frac{\sigma^4(t)}{nh}\, V \left(2 + (E\varepsilon_1^4 - 3) \sum_{j=-m_1}^{m_2} w_j^4\right).
$$

Using the continuity of the function $W_k(t)$, we obtain

$$I_2 \sim 2\sum_{k=1}^{n} W_k^2(t) \sum_{1\le i\le \min(m_1+m_2,n-k)} \operatorname{cov}(\widetilde{\sigma}^2(t_k), \widetilde{\sigma}^2(t_{k+i}))$$

Therefore we have to evaluate the quantities $\operatorname{cov}(\widetilde{\sigma}^2(t_k), \widetilde{\sigma}^2(t_{k+i}))$. First observe that

$$\begin{aligned} E\widetilde{\sigma}^2(t_k) &= E\left(\sum_{j=-m_1}^{m_2} w_j\mu(t_{k+j}) + \sum_{j=-m_1}^{m_2} w_j\sigma(t_{k+j})\varepsilon_{k+j}\right)^2 \\ &= \left(\sum_{j=-m_1}^{m_2} w_j\mu(t_{k+j})\right)^2 + \sum_{j=-m_1}^{m_2} w_j^2\,\sigma^2(t_{k+j}) \\ &\sim \sigma^2(t_k)\sum_{j=-m_1}^{m_2} w_j^2 = \sigma^2(t_k)\,. \end{aligned}$$

where we used the facts that $\sum_{j=-m_1}^{m_2} w_j = 0$ and $\sum_{j=m_1}^{m_2} w_j^2 = 1$, and the uniform continuity of μ and σ^2 on $[0,1]$. Therefore, by similar calculations as for Lemma 5.3(iii) in Müller and Stadtmüller [37], uniformly on compact intervals $[\delta, 1-\delta]$ for $\delta < 1$,

$$\begin{aligned} I_2 &\sim 2\sum_{k=1}^{n} W_k^2(t) \sum_{1\le i\le \min(m_1+m_2,n-k)} E(\widetilde{\sigma}^2(t_k)\widetilde{\sigma}^2(t_{k+i})) - 2\sum_{k=1}^{n} W_k^2(t)\sigma^4(t_k)\,(m_1+m_2) \\ &\sim 2\sum_{k=1}^{n} W_k^2(t) \sum_{1\le i\le \min(m_1+m_2,n-k)} E(\widetilde{\sigma}^2(t_k)\widetilde{\sigma}^2(t_{k+i})) - 2\,\frac{V}{n\,h}\,\sigma^4(t)\,(m_1+m_2)\,. \end{aligned}$$

Furthermore, direct calculation and similar arguments as above yield for[26] $1 \leq k \leq n$, $1 \leq k+i \leq n$,

$$
\begin{aligned}
& E(\widetilde{\sigma}^2(t_k), \widetilde{\sigma}^2(t_{k+i})) \\
= \; & \left(\sum_{j_1=-m_1}^{m_2} w_{j_1}\mu(t_{k+j_1})\right)^2 \left(\sum_{j_2=-m_1}^{m_2} w_{j_2}\mu(t_{k+i+j_2})\right)^2 \\
& + \left(\sum_{j_1=-m_1}^{m_2} w_{j_1}\mu(t_{k+j_1})\right)^2 \sum_{j_2=-m_1}^{m_2} w_{j_2}^2\sigma^2(t_{k+i+j_2}) \\
& + \left(\sum_{j_2=-m_1}^{m_2} w_{j_2}\mu(t_{k+i+j_2})\right)^2 \sum_{j_1=-m_1}^{m_2} w_{j_1}^2\sigma^2(t_{k+j_1}) \\
& + \sum_{j,j-i\in[-m_1,m_2]} w_j^2 w_{j-i}^2 \sigma^4(t_{k+j}) E\varepsilon_1^4 \\
& + \sum_{j\in[-m_1,m_2]} w_j^2\sigma^2(t_{k+j}) \sum_{l\in[-m_1,m_2]} w_l^2\sigma^2(t_{k+i+l}) - \sum_{j\in[-m_1,m_2]} w_j^4\sigma^2(t_{k+j})\sigma^2(t_{k+i+j}) \\
& +2 \sum_{j,j-i\in[-m_1,m_2]} \sum_{l,l-i\in[-m_1,m_2]\,,j\neq l} w_j w_l w_{j-i} w_{l-i} \sigma^2(t_{k+j})\sigma^2(t_{k+l}) \\
& +2 \left(\sum_{j_2=-m_1}^{m_2} w_{j_2}\mu(t_{k+i+j_2})\right) \sum_{j,j-i\in[-m_1,m_2]} w_j^2 w_{j-i}\sigma^3(t_{k+j}) E\varepsilon_1^3 \\
& +2 \left(\sum_{j_1=-m_1}^{m_2} w_{j_1}\mu(t_{k+j_1})\right) \sum_{j,j+i\in[-m_1,m_2]} w_j^2 w_{j+i}\sigma^3(t_{k+i+j}) E\varepsilon_1^3 \\
& +4 \left(\sum_{j_2=-m_1}^{m_2} w_{j_2}\mu(t_{k+i+j_2})\right) \left(\sum_{j_1=-m_1}^{m_2} w_{j_1}\mu(t_{k+j_1})\right) \sum_{j,j-i\in[-m_1,m_2]} w_j w_{j-i}\sigma^2(t_{j+k+i}) \\
\sim \; & \sigma^4(t_k)\, E\varepsilon_1^4 \sum_{j,j-i\in[-m_1,m_2]} w_j^2 w_{j-i}^2 + \sigma^4(t_k) - \sigma^4(t_k) \sum_{j\in[-m_1,m_2]} w_j^4 \\
& +2\,\sigma^4(t_k) \left[\left(\sum_{j,j-i\in[-m_1,m_2]} w_j\, w_{j-i} \right)^2 - \sum_{j,j-i\in[-m_1,m_2]} w_j^2 w_{j-i}^2 \right]
\end{aligned}
$$

[26]If one of the indices of the summands in these sums is outside the natural domain, we interpret the summand as zero.

$$= \sigma^4(t_k)\left[1+(E\varepsilon_1^4-2)\sum_{j,j-i\in[-m_1,m_2]} w_j^2 w_{j-i}^2 - \sum_{j\in[-m_1,m_2]} w_j^4 + 2\left(\sum_{j,j-i\in[-m_1,m_2]} w_j\, w_{j-i}\right)^2\right].$$

Therefore we finally have

$$I_2 \sim \frac{2\,V\,\sigma^4(t)}{n\,h}\Big((E\varepsilon_1^4-2)\sum_{i=1}^{m_1+m_2}\sum_{j,j-i\in[-m_1,m_2]} w_j^2 w_{j-i}^2 - (m_1+m_2)\sum_{j\in[-m_1,m_2]} w_j^4$$
$$+2\sum_{i=1}^{m_1+m_2}\left(\sum_{j,j-i\in[-m_1,m_2]} w_j\, w_{j-i}\right)^2\Big).$$

This formula together with the one for I_1 give the asymptotic expression for the variance $\mathrm{var}(\widehat{\sigma}^2(t))$ given in (18).

VI.C. **Statistical properties of the estimator of expected return $\widehat{\mu}_{\mathrm{He}}$ of in the heteroscedastic regression model.** We assume that μ is twice differentiable with continuous second derivative. Then Lemma 5.3 of Müller and Stadtmüller [37] gives the following results for $\widehat{\mu}_{\mathrm{He}}(t)=\widehat{\mu}_{\mathrm{He}}(t;h_\mu)$, the kernel estimator (3) of μ with specification $U_k=R_k$:

(1) The expected value $E\widehat{\mu}_{\mathrm{He}}(t;h_\mu)$ satisfies, as $n\to\infty$, $h_\mu=h_{\mu,n}\to 0$, $nh_\mu\to\infty$,

$$E\widehat{\mu}_{\mathrm{He}}(t)-\mu(t)=\mu''(t)h_\mu^2\,B+o(h_\mu^2)+O(n^{-1}), \tag{19}$$

where $B=\int K(u)u^2du/2$, and

$$|E\widehat{\mu}_{\mathrm{He}}(t)-\mu(t)|\le c\,(h_\mu^2+n^{-1}),$$

for some unspecified positive constant c, uniformly for $t\in[\delta,1-\delta]$, any fixed $\delta\in(0,1)$.

(2) The variance of $\widehat{\mu}_{\mathrm{He}}(t)$ satisfies for every t, as $n \to \infty$, $h_\mu = h_{\mu,n} \to 0$, $nh_\mu \to \infty$,

$$\operatorname{var}(\widehat{\mu}_{\mathrm{He}}(t)) = \frac{\sigma^2(t)}{nh_\mu} V \left(1 + o(1)\right), \tag{20}$$

where $V = \int K^2(u)du = 0.6$ for the Epanechnikov kernel used in our analysis.

We mention that the bandwidths h_μ for the estimation of μ and h_{σ^2} for σ^2 are in general very different.

References

[1] Andersen, T., Bollerslev, T. and Diebold, F. (2004) Parametric and non-parametric volatility measurement. *Handbook of Financial Econometrics*, forthcoming.

[2] Berkes, I., Horvath, L. and Kokoszka, P. (2003) GARCH processes: structure and estimation, *Bernoulli* **9** , 201–207.
Probabilistic and statistical properties of GARCH processes. Preprint. Available at http://www.math.usu.edu/∼piotr.

[3] Blume, M. (1966) *The Assessment of Portfolio Performance.* Ph.D. Dissertation, University of Chicago.

[4] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

[5] Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994) GARCH models. In: Engle, R.F. and McFadden, D.L. (Eds.) *Handbook of Econometrics.* Vol. **4**, pp. 2961–3038. Elsevier, Amsterdam.

[6] Brandt, M. (1999) Estimating portfolio and consumption choice: a conditional Euler equations approach. *J. Finance* **54**, 1609–1646.

[7] Campbell, J. (1987) Stock returns and the term structure. *J. Finan. Economics* **18**, 373–399.

[8] Campbell, J. and Hentschel, L. (1992) No news is good news: an asymmetric model of changing volatility in stock returns. *J. Financial Economics* **31**, 281-318.

[9] Campbell, J., Lo A. and MacKinlay, A. (1996) *The Econometrics of Financial Markets.* Princeton University Press.

[10] Chan, K., Karolyi, A. and Stulz, R. (1992) Global financial markets and the risk premium on U.S. equity. *J. Finan. Economics*, **32**, 137–168

[11] Diggle, P. and Hutchinson, M. (1989) On spline smoothing with autocorrelated errors. *Austral. J. Statist.* **31**, 166–182.

[12] Drees, H. and Stărică, C. (2002) A simple non-stationary model for stock returns. Available at www.math.chalmers.se/∼starica.

[13] Fama, E. F., and K. R. French (1989) Business conditions and expected returns on stocks and bonds. *J. Financial Economics* **25**, 23–49.

[14] Foster, D. F., T. Smith, and R. E. Whaley (1997) Assessing goodness-of-fit of asset pricing models: the distribution of the maximal R-squared. *J. Finance* **52**, 591–607.

[15] French, K., Schwert, W. and Stambaugh, R. (1987) Expected stock returns and volatility. *J. Financial Economics* **19**, 3–29.

[16] Gasser, T., Engel, J. and Seifert, B. (1993) Nonparametric function estimation. In: *Computational Statistics. Handbook of Statistics.* **Vol. 9**, pp. 423–465. North-Holland, Amsterdam.

[17] Gasser, T., Kneip, A. and Köhler, W. (1991) A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* **86**, 643–652.

[18] Gasser, T., Müller, H.-G. and Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc., Ser. B* **47**, 238–252.

[19] Glosten, C.R., Jagannathan, R. and Runkle, D.E. (1993) On the relation between the expected value and the volatility of the nominal excess returns on stocks. *J. Finance* **48**, 1779-1802.

[20] Hall, P. and T.E. Wherly (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *J. Amer. Statist. Assoc.* **86**, 665–672.

[21] Harrison, P. and Zhang, H.H. (1998) An investigation of the risk and return relation at long horizons. *Rev. Econom. Statist.* **81**, 399-408.

[22] Hart, J. (1991) Kernel regression estimation with time series errors. *J. Roy. Statist. Soc., Ser. B* **53**, 173–187.

[23] Hart, J. and Wehrly, T. (1986) Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.* **81**, 1080–1088.

[24] Harvey, C. (1989) Time-varying conditional covariances in tests of asset pricing models. *J. Financial Economics* **24**, 289–317.

[25] Herzel, S., Nord, T. and Stărică , C. (2004) The GARCH(1,1) model often fails to provide sensible longer-horizon volatility forecasts. Technical report. Available at www.math.chalmers.se/∼starica.

[26] Herzel, S., Stărică, C. and Tütüncu, R. (2002) A non-stationary multivariate model for financial returns. Available at www.math.chalmers.se/∼starica.

[27] Kandel, S. and Stambaugh, R. (1990) Expectations and volatility of consumption and asset returns. *Review of Financial Studies* **3**, 207–232.

[28] King, B. (1966) Market and industry factors in stock price behavior. *J. Business* **39**, 139–190.

[29] Klein, P. and Niemira, M. (1994) *Forecasting Financial and Economic Cycles.* Wiley, New York.

[30] Lo, A. W., and MacKinlay, A. C. (1990) Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies* **3**, 431–467

[31] Markowitz, H.M. and Usmen, N. (1996) The likelihood of various stock market return distributions, part I: principles of inference. *J. Risk Uncertainty* **13**, 207–219.

[32] Markowitz, H.M. and Usmen, N. (1996) The likelihood of various stock market return distributions, part II: empirical results. *J. Risk Uncertainty* **13**, 221–247.

[33] Merton, R. (1980) On estimating the expected return on the market: an exploratory investigation. *J. Financial Economics* **8**, 323-361.

[34] Mikosch, T. (2003) Modeling dependence and tails of financial time series. In: Finkenstädt, B. and Rootzén, H. (Eds.) *Extreme Values in Finance, Telecommunications and the Environment.* Chapman and Hall, London.

[35] Mikosch, T. and Stărică, C. (2004) Changes of structure in financial time series and the GARCH model. *Revstat Statistical Journal* **2**, 41-73. Available under www.math.ku.dk/∼mikosch.

[36] Mikosch, T. and Stărică, C. (2003) Non-stationarities in financial time series, the long-range dependence and IGARCH effects. *The Review of Economics and Statistics* **86**, 378-390.

[37] Müller, H.-G. and Stadtmüller, U. (1987) Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 610–625.

[38] Nelsen, R.B. (1999) An Introduction to Copulas. *Lecture Notes in Statistics*, 139, Springer, NY.

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

[40] Officer, R. (1976) The variability of the market factor of the New York Stock Exchange. *J. Business* **46**, 434–453

[41] Pagan, A. and Sossounov, K. (2003) A simple framework for analyzing bull and bear markets. *J. of Applied Econometrics* **18**, 23–46.

[42] Rodrguez-Poo, J. and Linton, O. (2001) Nonparametric factor analysis of residual time series. *TEST* **10**, 161–182.

[43] Samuelson, P. (1965) Proof that property anticipated prices fluctuate randomly. *Industrial Management Review* **VI**, 41–49.

[44] Scruggs, J. (1998) Resolving the puzzling inter-temporal relation between the market risk premium and conditional market variance: A two-factor approach. *J. Finance* **52**, 575-603.

[45] Stărică C. (2003) Is Garch(1,1) as good a model as the accolades of the Noble prize would imply? Technical report. Available at www.math.chalmers.se/∼starica.

[46] Stărică C. and Granger, C. (2004) Non-stationarities in stock returns, *Review of Economics and Statistics*, forthcoming. Available at www.math.chalmers.se/∼starica.

[47] Taylor, S.J. (1986) *Modelling Financial Time Series.* Wiley, Chichester.

[48] Wand, M.P. and Jones, M.C. (1995) *Kernel Smoothing.* Chapman and Hall, London.

[49] Whitelaw, R. (1994) Time variations and covariations in the expectation and volatility of stock market returns. *J. Finance* **49**, 515–541.

[50] Whitelaw, R. (1997) Time-varying sharpe ratios and market timing.
Available at http://pages.stern.nyu.edu/ rwhitela/research.html.