Projektbereich A
Discussion Paper No. A–533

Non-Parametric Volatility Estimation of Exchange Rates and Stock Prices

by

Frank Heid

September 1996

Acknowledgements: Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn is gratefully acknowledged.

SONDERFORSCHUNGSBEREICH 303
Rheinische Friedrich-Wilhelms-Universität Bonn
Lennéstraße 37
D–53113 Bonn
Germany
Telefon (0228) 73 92 42
Telefax (0228) 73 79 40
E–Mail: with2@econ2.uni-bonn.de

Abstract

We model the volatility of financial assets as functions depending on past returns. We apply nonparametric regression techniques to estimate the volatility of daily exchange rates and stock prices. We show that all of the estimated functions have similar shape as they are convex with a minimum close to zero.

JEL-Classification: G20,C14

1 Financial Time Series

Any sequence of prices of financial assets (e.g. stocks, futures, market index values, exchange rates etc.) which is observed on consecutive dates is called a financial time series. A sequence such as $\{x_1, x_2, ..., x_T\}$, is called univariate if it records only one price over time. If x_t contains not only one price but also other relevant factors (which may include prices of other financial assets) it is called a multivariate time series. You can either depict recording times regularly or irregularly and data sets are available for yearly, monthly, weekly, daily or even intradaily recorded prices. In this study we will focus on univariate series of exchange rates and stock prices which were depicted on a daily basis. It is in the nature of financial prices that future prices are uncertain and thus subject to erratic changes. To explain this phenomenon properly it is customary to construct a propabilistic model and view $\{x_1, x_2, ..., x_T\}$ as a realisation of a stochastic process $\{X_1, X_2, ..., X_T\}$. Time series analysis will then select a propabilistic model, i.e. a joint distribution of the $\{X_t\}$, which describes the data best. Because estimating the entire model is beset with many difficulties research is frequently limited to one aspect of the process such as the marginal distribution of the X_t, the expected value of X_t given past prices etc. As stated above our study will concentrate on univariate financial time series only and will exclude other factors which may have an impact on the considered prices. Our primary goal in this study is to describe the data in terms of prediction rather than to offer explanations for the actual price levels. While predictions include price forecasts it will not be our main concern. Instead we want to describe the disribution of prices in more detail which goes beyond merely estimating the conditional mean. It will be shown that the conditional volatility assumes a key rule.

In order to determine the appropriate length of the financial time series in analysis several decisions have to be made. On the one hand a fairly long time span is prerequisite to achieve a high accuracy of the estimates. On the other hand it is unreasonable to expect the propabilistic model to remain constant during the course of a long time span. Some non-random events such as inteventions of central banks certainly have an important impact on prices. Thus, it is better to keep the time series short or to split a long series into several shorter units. Consequently, we have chosen to limit our time span to approximately 3 years. The fact that financial time series are not, in the main, stationary presents another problem for statistical analysis as most estimation procedures are only valid for stationary processes.[1] A remedy is to focus on *returns* which can be shown to be stationary and are in a sense equivalent to prices.

We use two sets of data in this paper. The first one consists of daily noon exchange rates of 19 major currencies vis à vis the Canadian Dollar (cf list in the appendix). It covers a period from January 1993 to June 1996 which amounts to 858 trading days. We have not considered data before October 1992 because the variability bounds imposed by the European Monetary System (EMS) for European currencies were relatively narrow. The data was made available by the Pacific Exchange Rate Service of the University of British Columbia, Vancouver, Canada. The second set of data reports daily closing prices of 175 stocks listed at the New York Stock Exchange provided by the MIT Experimental Stock Market Data Server. The recording time is from September 1993 to May 1996 i.e. 623 trading days. A first impression can be gained by looking at the plots of figure 1. They already reveal some tendencies which are later confirmed in the other series. Most exchange rates are negatively trended while a positive trend can be discerned for most of the stocks. The plots also show that great changes in prices occur frequently. They resemble random walks i.e. cumulative sums of i.i.d. random numbers. The simulation of the TAN series by a random walk demonstrates that nicely (figure 2). The increments we have chosen were i.i.d. normal possess-

[1] A stochastic process $\{X_t\}$ is said to be strictly stationary if for any integers $j_1, j_2, ..., j_n$ the joint distribution of $\{X_{t+j_1}, X_{t+j_2}, ..., X_{t+j_n}\}$ depends only on the intervals seperating the dates $(j_1, j_2, ..., j_n)$ and not the date t itself. The process is called weakly or covariance stationary if the means and autocovariances do not depend on t: $\mathrm{E}[X_t] = \mu$, $\mathrm{E}[(X_t - \mu)(X_{t-j} - \mu)] = \gamma_j$ for all t and any j.

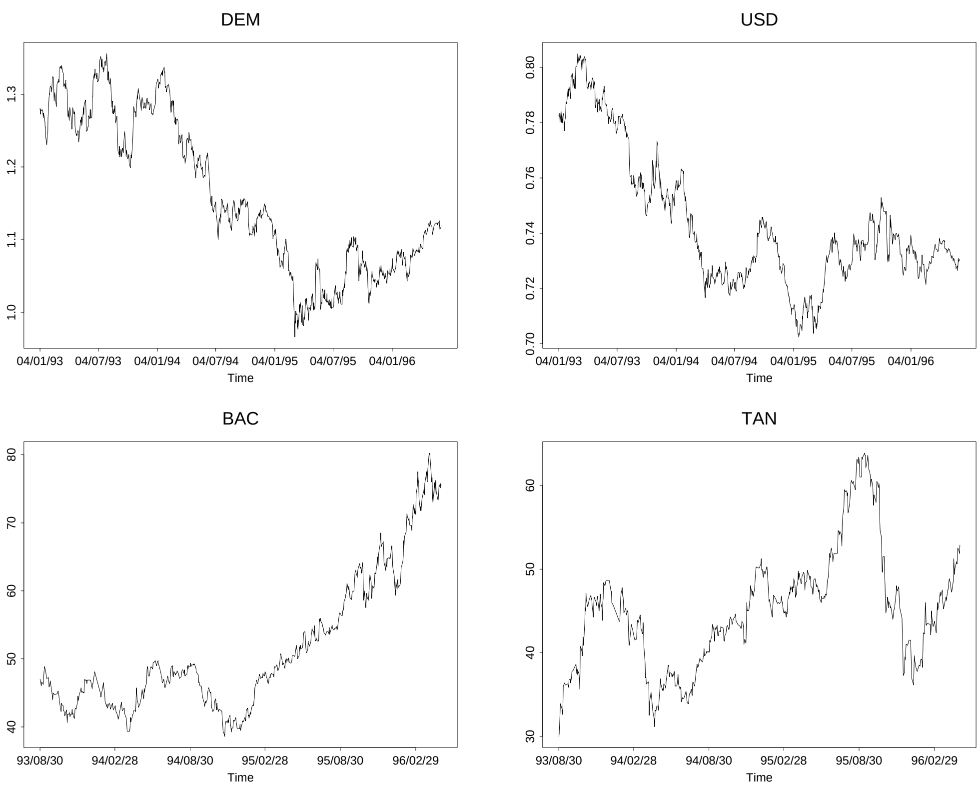


Figure 1: Time series of DEM and USD exchange rates and BAC and TAN stock prices

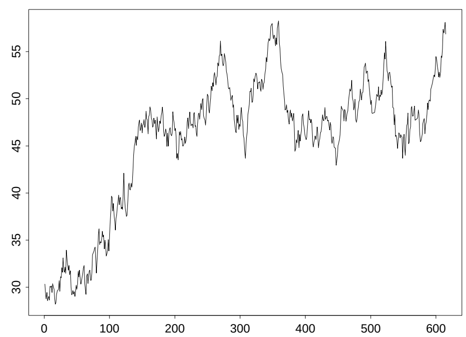


Figure 2: Simulation of TAN series

ing mean and standard deviations as those of the original series. In fact, by a first sight the original and simulated series behave quite similarly. The random walk hypothesis was first introduced by Bachelier (1900) who claimed by pure theoretical reasoning that price changes have independent and identical normal distributions. Later, Fama (1965) came to a similar conclusion by means of an empirical analysis which confirmed that stocks are random walks or something very similar to that. This view has been challenged in manifold papers in the wake of Engle's work on ARCH models (Engle, 1982) and its parametric extensions. They all claim that the innovation process is not white noise because the conditional variance is not constant as were the case for independent innovations. These papers have in common that they specify the conditional variance as pre-defined functions depending on some variables, e.g. past prices, and some unknown parameters. Parametric modelling, however, always bears the risk of miss-specification and, therefore, important characteristics of the regression functions may elude one's observation. In this paper we propose to estimate the conditional volatility by means of nonparametric regression, which does not assume any functional form of the regression function. Thus being more flexible we are able to show that volatilities are *asymmetric* functions of past prices, a feature which is not captured in most of the standard parametric models.[2]

[2]Recently nonparametric estimates of the volatility functions have also been proposed by Bossaerts, Hafner, and Härdle (1996)

2 Daily Returns

Statistical analysis is difficult if consecutive prices are highly correlated. Yet it is even worse, if the whole series is non-stationary. In fact a Dickey-Fuller test applied to our data does not reject the hypothesis of the prevelance of random walks for nearly all exchange rates and stock prices. Consequently, we prefer to analyse price *changes*. Suppose that prices are recorded each trading day at the same time of day and let x_t be the price recorded on date t. Three types of price changes have been considered in previous research:

$$\begin{aligned} r_t^* &= x_t - x_{t-1}, \\ r_t' &= (x_t - x_{t-1})/x_{t-1}, \\ r_t &= \log(x_t) - \log(x_{t-1}). \end{aligned}$$

In these definitions we have excluded dividend payments for stocks which occur so rarely that they are of marginal importance for our analysis. Since r_t^* depends on units of account the comparison of different time series is precluded with the result that most researchers prefer to analyze r_t' and r_t. The relation between the latter is given by

$$x_t/x_{t-1} = 1 + r_t' = e^{r_t} = 1 + r_t + \frac{1}{2}r_t^2 + \frac{1}{6}r_t^3 + ... \tag{1}$$

Thus, r_t' is the simple rate of return and r_t is the rate of return with continuous compounding, for $\exp(r_t) = \lim_{n\to\infty}(1 + r_t/n)^n$. The difference between r_t and r_t' is negligible since from (1) it is of order r_t^2, and r_t is nearly always in the range of -0.1 to 0.1. There are two reasons why we prefer the compound return to the simple return. Firstly, continuous time generalisations of the discrete time series are less complex, and secondly, returns of over more than one day are easily calculated as they are simply the sum of the one-day-returns, i.e.

$$r_{t,k} := \log(x_t) - \log(x_{t-k}) = r_t + r_{t-1} + ... + r_{t-k+1}.$$

Figure 3 shows the equivalent return processes of figure 1. It is impossible not to notice the different stochastic behaviour in each case in comparison to figure 1. While the returns fluctuate around a mean close to zero a Dickey Fuller test no longer rejects the hypothesis of them being random walks. The histograms in figure 4 and figure 5 summarize the first two moments of all the series. As

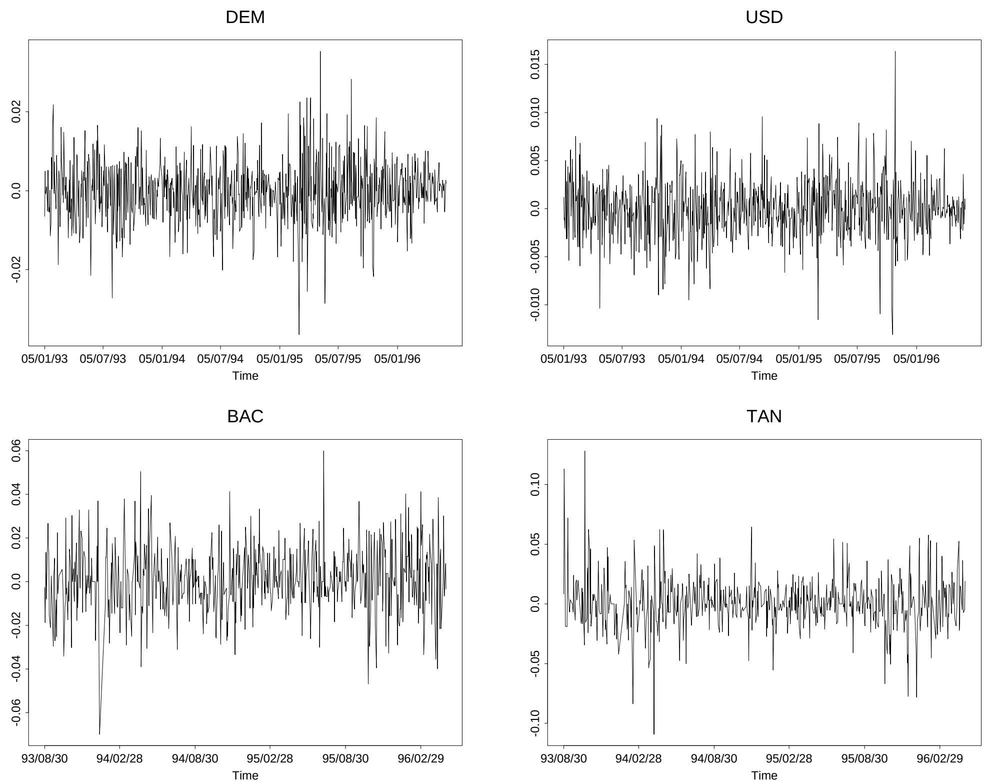


Figure 3: Time series of DEM and USD exchange rates returns and BAC and TAN stock price returns

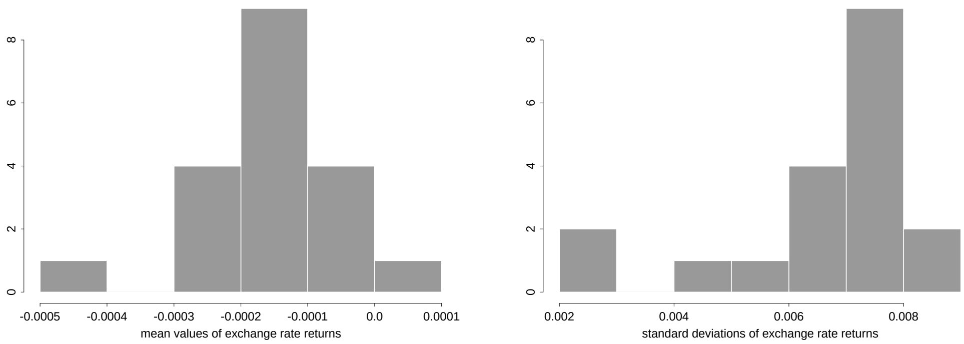


Figure 4: Histograms of the means and the standard deviations of exchange rate returns

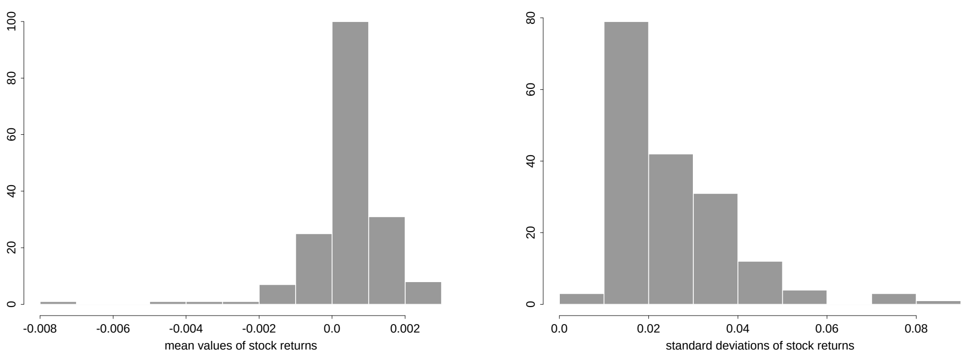


Figure 5: Histograms of the means and the standard deviations of stock returns

already mentioned the exchange rates nearly always have negative mean returns in the considered time span; those of the stocks are mostly positive. Additionally, stock price returns are more volatile - in average by factor 10 - in terms of their standard deviation.

3 Autocorrelation

In the following we will assume that the return processes $\{r_r\}$ are strictly stationary. Since our final goal is to present a stochastic model which describes the data best, it is necessary to unravel the dependence structure of consecutive returns of different lags. If the returns are uncorrelated they are called *white noise*, whilst the term *strict white noise* is applied to independent and identically distributed returns. Thus, the first step in exploring dependence is to look at the autocorrelation function $\rho_{r,\tau} = \text{cov}(r_t, r_{t+\tau})/\text{var}(r_t)$, $\tau > 0$. It is straightforwardly estimated by means of the sample autocorrelation function

$$\hat{\rho}_{r,\tau} = \sum_{t=1}^{T-\tau}(r_t - \bar{r})(r_{t+\tau} - \bar{r}) \Big/ \sum_{t=1}^{T}(r_t - \bar{r})^2, \tau > 0$$

where

$$\bar{r} = \sum_{t=1}^{T} r_t.$$

The convergence of $\hat{\rho}_{r,\tau}$ to $\rho_{r,\tau}$ in probability can be shown for a broad class of stochastic processes. Two important cases are linear processes, strict white noise being a special case, and weakly dependent processes which are discussed in section 5. The estimates $\hat{\rho}_{r,\tau}$ may then be used to test hypotheses about the theoretical correlation coefficients $\rho_{r,\tau}$. The first lag autocorrelation is of particular importance since we may well expect that the current return has the biggest impact on future returns. In our data sets 15 out of 19 exchange rates display positive first lag coefficients. They are very small and range from -0.03 (ATS) to 0.07 (USD) with an average of 0.03 (0.04 for the absolute values). First lag coefficients of the stock returns are mostly negative (100 out of 175) and at the same time very small with an average of -0.01 (0.01 for the absolute values). They are more dispersed with -0.22 as the lowest and 0.27 as the highest value. Special care is recquired in interpreting these results for the results usually depend on common market factors.

The results above suggest to test the hypothesis of the returns being strict white noise. For those processes the following theorem holds.

Theorem 1 *If $\{X_t\}$ is a sequence of independent identically distributed random variables with finite second and fourth moments then $\sqrt{T}\hat{\rho}(h)$ converges in distribution to $N(\rho(h), I)$ where*

$$\hat{\rho}(h) = (\hat{\rho}_{X,1}, \hat{\rho}_{X,2}, ..., \hat{\rho}_{X,h})$$
$$\rho(h) = (\rho_{X,1}, \rho_{X,2}, ..., \rho_{X,h}).$$

	exchange rates	stocks
returns	7/19	102/175
squared returns	17/19	81/175
absolute returns	17/19	133/175

Table 1: Number of rejections of the hypothesis of returns being white noise

Proof: Brockwell and Davis (1991) p. 221

So, if $\{r_t\}$ is strict white noise, then for large T approximately 95% of the sample autocorrelations should fall between the bound $\pm 1.96\, T^{-\frac{1}{2}}$. Similar tests may be based on suitable transformations of r_t for if $\{r_t\}$ is strict white noise, so are $\{|r_t|\}$ and $\{r_t^2\}$. Table 1 summarizes the results of the three tests applied to all exchange rates and stocks. The hypothesis of strict white noise for a particular sequence of returns is rejected if more than 2 out of 30 sample autocorrelations are out of the confidence bounds. As it turns out, the number of rejections is almost always extremely high, being only less significant for the test performed on the exchange rates with non-transformed returns. Note that if the processes were strict white noise and *independent*, we would expect only 5% of rejections - far less than observed. However, we have already pointed out that the processes are probably *not* independent. Nonetheless we may still say that the number of rejections is too high to support the hypothesis of the return processes being strict white noise. The high number of rejections stems from the relatively high lag-1 correlations which is evident from running the the test for the first three lags seperately the results of which are shown in table 2 and table 3.

lag 1

mean	rejections	
returns	-0.01	39/175
squared returns	0.07	64/175
absolute returns	0.1	119/175

lag 2

	mean	rejections
returns	-0.03	35/175
squared returns	0.03	25/175
absolute returns	0.06	50/175

lag 3

	mean	rejections
returns	-0.02	16/175
squared returns	0.02	15/175
absolute returns	0.03	36/175

Table 2: Autocorrelation of stock returns and modified stock returns

As regards the absolute returns as well as the squared returns the lag-1 correlations are significantly different from zero and they are almost always positive which means that high returns are propably to be followed by high returns again (possibly into the opposite direction).

Here we want to emphasize that the tests *do not* reject zero correlations of returns (or absolute returns or squared returns). All we can say is that the processes are not strict white noise. In fact, if the processes are not white noise the variance of $\rho_{r,\tau}$ may exceed $1/T$.

4 A Model of Heteroscedasticity

Why is there much more autocorrelation in the processes $\{|r_t|\}$ and $\{r_t^2\}$ than in the process $\{r_t\}$? One reason may be that high past absolute returns will lead to more trading activity which in turn will increase the volatility of future returns. Hence, changes in conditional variances may explain the observed correlation structure of the return processes. We now want to present a statistical model in which this correlation structure comes out naturally. To keep the setup as simple as possible let the information known to traders at time t be given by a vector of random variables I_t. Then the best predictor of r_t is given by the conditional mean

$$\mu_t := \mathrm{E}[r_{t+1}|I_t].$$

While the unconditional variance of the strictly stationary process stays constant the conditional variance may depend non-trivially on time:

$$\mathrm{var}[r_{t+1}|I_t] = \mathrm{E}[(r_t - \mu_t)^2|I_t] := \sigma_t.$$

Consequently,

$$r_{t+1} = \mu_t + \sigma_t \xi_{t+1} \tag{2}$$

with some $(0,1)$ random variables ξ_t. For the moment let us assume that the conditional mean is always zero. Then it is easily confirmed that there is no correlation between r_t and $r_{t-\tau}$. On the other hand there may be significant correlation between r_t^2 and $r_{t-\tau}^2$.

In this analysis we are particularly interested in the special case if I_{t-1} includes only past returns. Yet to make it statistically managable we have to limit lag 1

	mean	rejections
returns	0.03	1/19
squared returns	0.12	16/19
absolute returns	0.12	16/19

lag 2

	mean	rejections
returns	-0.03	4/19
squared returns	0.06	4/19
absolute returns	0.05	4/19

lag 3

	mean	rejections
returns	-0.04	2/19
squared returns	0.04	2/19
absolute returns	0.07	9/19

Table 3: Autocorrelation of exchange rate returns and modified exchange rate returns

the scope of included lags, otherwise we would run into the curse of dimensionality. To simplify the analysis even further we skip all lags larger than 1. Model (2) then reduces to [3]

$$r_{t+1} = \mu(r_t) + \sigma(r_t)\xi_{t+1}. \tag{3}$$

If, additionally, $\{\xi_t\}$ is assumed to be strict white noise, equation (3) would specify a heteroscedastic nonlinear autoregression (CHARN).

There are some theoretical reasons to believe that $\{r_t\}$ may be a (first order) Markov chain. If financial markets worked efficiently then all relevant information would be included in present returns.[4] In this case a forecast based on all available information is not better than a forecast based solely on today's returns. More precisely, the conditional distribution of future returns based on the whole information set is equal to the conditional distribution given today's return. However, the assumption of efficiency is not convincing if investors have to bear some cost for acquiring information. As in Grossman's and Stiglitz' model (1980) returns may then reveal information rather slowly. Whatever is the case we will state explicitly if $\{\xi_t\}$ is assumed to be strict white noise.

5 Nonparametric Estimation of Time Series

One of the basic assumptions of parametric regression is that the conditional expectation of the dependent variable given the value of the regressor is a known function depending on the regressor and some unknown parameters. In most cases, however, the number of suitable functional forms is large and consequently the statistician very often has to try several alternatives. Yet this always bears the danger of misspecification. In contrast to this, nonparametric regression does not assume any specific functional form of the regression function. In this paper we deal with so-called kernel estimators which were introduced by Nadaraya (1964) and Watson (1964). Given a data set $\{(X_1, Y_1), (X_2, Y_2)...(X_T, Y_T)\}$ a kernel estimator of the regression function

$$r(x) = \mathrm{E}[Y_t | X_t = x]$$

[3]Note that the subscript t can be ommited at the functions μ and σ for we started out from a strictly stationary process $\{r_t\}$.

[4]This notion of efficiency is not to be confused with Pareto-efficiency. A suitable term would be information-efficiency.

is a random function of the form

$$\hat{r}(x) = \sum_{t=1}^{T} Y_t \, k\Big(\frac{x - X_t}{h_T}\Big) \Big/ \sum_{t=1}^{T} k\Big(\frac{x - X_t}{h_T}\Big) \, .$$

k is a function on $\mathbb{R}$, called kernel, and $\{h_T\}$ is a sequence of positive real numbers, called bandwidth paramters which converge to 0 for T converging to infinity. If the data are i.i.d. then, under suitable conditions, it can be shown that $\hat{r}$ converges to r.

Of course the i.i.d. case rules out time series applications. An early reference on dependent situations (for Markov processes) is Roussas (1969). Further important contributions are Yakowitz (1979), Doukhan and Ghindes (1980), Robinson (1980), Bierens (1983) and Härdle and Vieu (1992). Robinson (1983) provided asymptotic normality results. The results of these works are based on the assumption that the underlying process is weakly dependent meaning that it has vanishing memory. We will investigate two concepts of weak dependence in the next chapter

5.1 Weak Dependence

In the sequal denote by $\{X_t\}$ a strictly stationary sequence of random variables on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$. For $a < b$ define by $\mathcal{M}_b^a$ the σ-field generated by $\{X_a, X_{a+1}, ... X_b\}$; define by $\mathcal{M}_\infty^a$ the σ-field generated by $\{..., X_{a-1}, X_a\}$ and by $\mathcal{M}_a^\infty$ the σ-field generated by $\{X_a, X_{a+1}, ...\}$. One of the least restrictive dependence structures discussed in the literature is the strong mixing condition.

Definition: *The process $\{X_t\}$ is called strongly mixing (α-mixing) if there exists a non-negative function α on positive integers such that for each integer k ($-\infty < k < \infty$), for each integer n ($n \geq 1$) and for all events* $\mathrm{A} \in \mathcal{M}_\infty^k$ *and* $\mathrm{B} \in \mathcal{M}_{k+n}^\infty$

$$|\mathrm{P}(\mathrm{A} \cap \mathrm{B}) - \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})| \leq \alpha(n)$$

with

$$\lim_{n \to \infty} \alpha(n) = 0.$$

An assumption which is slightly stronger is that $\{X_t\}$ be uniformly mixing.

Definition: *The process $\{X_t\}$ is called uniformly mixing (ϕ-mixing) if there exists a non-negative function ϕ on positive integers such that for each integer k*

($-\infty < k < \infty$), for each integer n ($n \geq 1$) and for all $\mathrm{A} \in \mathcal{M}_{\infty}^{k}$ *and* $\mathrm{B} \in \mathcal{M}_{k+n}^{\infty}$

$$|\mathrm{P}(\mathrm{A} \cap \mathrm{B}) - \mathrm{P}(\mathrm{A})\mathrm{P}(\mathrm{B})| \leq \phi(n)\mathrm{P}(A)$$

with

$$\lim_{n \to \infty} \phi(n) = 0.$$

It is clear that a uniformly mixing process is strongly mixing (the converse is generally false).

The following lemma states that correlation between functions of X_t and $X_{t+\tau}$ is small if the lag τ is large.

Lemma: *Let $\{X_t\}$ be uniformly mixing. If ξ is measurable with respect to $\mathcal{M}_{-\infty}^{k}$ and η is measurable with respect to $\mathcal{M}_{k+n}^{\infty}$ ($n \geq 0$) then* $\mathrm{E}[|\xi|^r] < \infty$, $\mathrm{E}[|\eta|^s] < \infty$, $r, s > 1$, $\frac{1}{r} + \frac{1}{s} = 1$ *implies*

$$|\mathrm{E}[\xi\eta] - \mathrm{E}[\xi]\mathrm{E}[\eta]| \leq 2\phi(n)^{\frac{1}{r}} E^{\frac{1}{r}}[|\xi|^r] E^{\frac{1}{s}}[|\eta|^s]. \tag{4}$$

Examples:

- *m-dependent processes*
  The process $\{X_t\}$ is called m-dependent if the random vectors $(X_{a-p}, X_{a-p+1}, ..., X_a)$ and $(X_b, X_{b+1}, ..., X_{b+q})$ are independent whenever $b - a > m$, or equivalently if $\mathcal{M}_{-\infty}^{a}$ and $\mathcal{M}_{b}^{\infty}$ are independent whenever $b - a > m$. An m-dependent process is trivially uniformly mixing with $\phi(n) = 0$ for $n > m$.

- *Markov processes*
  If $\{X_t\}$ is a Markov process with finite state space and if the transition matrix is irreducible and aperiodic then $\{X_t\}$ is uniformly mixing. If $\{X_t\}$ has a infinite state space it is uniformly mixing if it satisfies Doeblin's condition, has one ergodic class and is aperiodic (Billingsley, 1968, p. 168). In both cases the mixing conditions satisfy

  $$\phi(n) \leq a\,\rho^n$$

  with positive constants a, ρ, $\rho < 1$.

- *Gaussian processes*
  $\{X_t\}$ is said to be Gaussian if for any $t_1, t_2, ..., t_n$ the random vector $(X_{t_1}, X_{t_2}, ...X_{t_n})$ is Gaussian. A Gaussian sequence satisfies the uniform mixing conditions if and only if the σ-fields $\mathcal{M}_{-\infty}^{k}$ and $\mathcal{M}_{k+n}^{\infty}$ are independent for all sufficiently large n. With the help of this it can be shown that a Gaussian process is uniformly mixing if its spectral density $f(\lambda)$ is continuous and $f(\lambda) \geq m > 0$.

5.2 Consistency and Normality of the Estimator

In this section we will summarize the results of Bierens (1983) and Robinson (1983), two important contributions to the theory of nonparametric time series regression. Bierens (1983) proved uniform consistency of the kernel estimator of uniformly mixing processes. Robinson (1983) showed that the kernel estimator is also consistent, though not uniformly consistent, for strongly mixing processes and delivered asymptotic normality results.

In the sequal let $\{X_t\}$ be a real-valued strictly stationary process. We will focus on estimators of

$$G_T(x) = \mathrm{E}[g(X_{t+k})\,|X_t = x], \tag{5}$$

where g is any real Borel function. In fact, the results of Bierens and Robinson hold for more general regression functions. Yet for our purpose estimation of G_T as in (5) will suffice. Let the estimator of G be given by

$$\widehat{G}_T(x) = \sum_{t=1}^{T} g(X_{t+k})\, k\left(\frac{X_t - x}{h_T}\right) \Bigg/ \sum_{t=1}^{T} k\left(\frac{X_t - x}{h_T}\right). \tag{6}$$

We assume

(U1) $\{X_t\}$ is uniformly mixing with mixing coefficients $\phi(m)$.

(U2) $\mathrm{E}[g(X_t)^2] < \infty$

(U3) The distribution of X_t is absolutely continuous with a continuous density f. The function G_T is continuous on $\mathbb{R}$.

(U4) k is an everywhere positive density on $\mathbb{R}$ with an absolutely integrable characteristic function.

Theorem 2 *(Uniform Consistency): Let* (U1) *to* (U4) *hold and let* h_T *in* (6) *be any sequence of positive numbers satisfying*

$$\lim_{T\to\infty} h_T = 0 \quad ; \quad \lim_{T\to\infty} h_T \rho_T^{-\frac{1}{2}} = \infty$$

where

$$\rho_T = \frac{1}{T}\sum_{t=1} T\phi(t)^{\frac{1}{2}}.$$

Then for every $\epsilon \in (0, \sup_{x\in\mathbb{R}} f(x)]$ *we have*

$$\operatorname{plim} \sup_{x:g(x)\geq\epsilon} |\widehat{G}(x) - G(x)| = 0.$$

If in addition $G(x) f(x)$ *is twice differentiable with continuous and uniformly bounded second derivatives and if* k *has zero mean and finite variance then for any sequence* ζ_T *such that for* $T \to \infty$, $\zeta_T = o(\min(h_T^{-2}, h_T\rho_T^{-\frac{1}{2}})$ *and every* $\epsilon \in \left(0, \sup_{x:f(x)\geq\epsilon} f(x)\right]$ *we have*

$$\operatorname{plim} \zeta_T \sup_{x:f(x)\geq\epsilon} |\widehat{G}_T(x) - G(x)| = 0$$

The sequence $\{\zeta_T\}$ *indicates the rate of uniform convergence of* $\widehat{G}_T$.

Proof Bierens (1983)

Now assume

(**N1**) $\{X_t\}$ is strongly mixing with mixing coefficients $\alpha(j)$ satisfying

$$\sum_{j=N}^{\infty} \alpha(j)^{1-\frac{2}{\theta}} = o\left(N^{-1}\right) \text{ for some } \theta > 2.$$

(**N2**) $G(x)$ has rth degree derivatives satisfying a Lipschitz condition of order 1.

(**N3**) $\mathrm{E}[g(X_t)^{\theta}] < \infty$ for $\theta > 0$.

(**N4**) $G_\gamma(x) := \mathrm{E}[|g(X_{t+k})|^{\gamma}|X_t = x]$ is continuous.

(**N5**) For all s the distribution of (X_t, X_{t+s}) is absolutely continuous with a continuous density.

(**N6**) X_t is absolutely continuous with a positive density f which has rth degree derivatives satisfying a Lipschitz condition of degree 1.

(**N7**) $\int k(u)\,du = 1$. (Note that k need not to be positve.)

(**N8**) $\int u^h k(u)\,du = 0$ for all $0 < h < r+1$

(**N9**) $k(u)$ is bounded with compact support.

Theorem 3 *(Asymptotic Normality): Let* (N1) *to* (N9) *hold and let*

$$S_T = \sqrt{h_T T}\left(\widehat{G}(x) - G(x)\right).$$

Assume that as T converges to infinity

$$h_T^{2(r+1)} T \to 0 \quad , \quad h_T T \to \infty.$$

Then $S_T(x)$ converges to a normal variable with zero mean and variance

$$\int k^2(u)\,du\, \frac{H(x) - G^2(x)}{f(x)} \tag{7}$$

with $H(x) = \mathrm{E}[g^2(X_{t+k})|X_t = x]$

Proof: Robinson (1983)

Theorem 4 *Under the condition of theorem 3* (7) *is consistently estimated, to* $O_p\left((T|h_T|)^{\frac{1}{2}}\right)$*, by*

$$\int k^2(u)\,du\, \frac{\widehat{H}(x) - \widehat{G}(x)^2}{\hat{f}(x)} \tag{8}$$

where $\widehat{H}(x)$ is the kernel estimator of H (replace g by g^2 in (6) *and $\hat{f}$ is the kernel density estimator of f*

$$\hat{f}(x) = \frac{1}{Th_T}\sum_{t=1}^{T} k\left(\frac{x - X_t}{h_T}\right).$$

Proof: Robinson (1983)

5.3 Bandwidth Selection

The choice of an appropriate bandwidth h in (6) plays a prominent role in nonparametric regression. Every selection rule has to face a trade-off between the variance and the bias of the estimator which both constitute the mean squared error. Decreasing h will also decrease the bias on the one hand but, on the other hand, will increase the variance of the estimator and vice versa for increasing h. The right choice of h will balance both effects. If we measure the accuracy of an estimate by a function $d(h)$ a data driven bandwidth choice is said to be asymptotically optimal in the sence of Shibata (1981) if we have

$$\frac{d(\hat{h})}{\inf d(h)} \longrightarrow 1 \quad \text{a.s.} \tag{9}$$

Global optimality as in (9), however, is not always feasable and condition (9) is relaxed by minimizing $d(h)$ in the denominator over some subset of $\mathbb{R}_+$. A frequently used measure of accuracy, which we also consider here, is the *Averaged Squared Error*

$$ASE(h) := \frac{1}{T}\sum_{t=1}^{T}(G(X_t) - \widehat{G}_T(X_T))^2.$$

In the following we will discuss the cross-validation method which can be shown to produce asymptotically the optimal bandwidth (Härdle, Vieu, and Hart 1989). To simplify notation let $Y_t := g(X_{t+k})$. Then the cross-validation method works as follows: Let $\widehat{G}_T^\tau$ be the kernel estimator as in (6) but with some data points in the neighbourhood of τ left out, i.e.

$$\widehat{G}_T^\tau(x) = \sum_{|t-\tau|>l}^{T} Y_t k\Big(\frac{x - X_t}{h}\Big) \Big/ \sum_{|t-\tau|>l}^{T} k\Big(\frac{x - X_t}{h}\Big)\,.$$

Define the cross-validation function

$$CV_l(h) = \frac{1}{T}\sum_{t=1}^{T}\Big(Y_\tau - \widehat{G}_T^\tau(X_\tau)\Big)^2. \tag{10}$$

The bandwidth $\hat{h}$ is chosen as to minimize $CV(h)$. The following theorem states that $\hat{h}$ is asymptotically optimal.

Theorem 5 (Härdle, Vieu and Hart, 1989) *Assume that (X_t, Y_t) is strongly mixing.*[5] *Then, $\hat{h}$ selected as to minimize $CV_l(h)$ over the intervall $H_T = [AT^{-a}, BT^{-b}]$, $0 < b < 1/(2k+1) \leq a < 2/(1+4k)$ is asymptotically optimal, i.e.*

$$ASE(\hat{h}) / \inf_{h\in H_T} ASE(h) \longrightarrow 1 \quad \text{a.s.}$$

[5]For simplicity, we do not state some other technical conditions here

To the practioneer there remain two questions unsolved:

1. What is the proper set H_T?

2. Which leave-out-variable l is to choose?

To tackle the first problem one would usually try some reasonable bandwidths and choose some interval which is likely to contain the optimal bandwidth. The second problem is more delicate and no satisfactory answer can be given. In many cases, however, the estimated bandwidth is not very sensible to the chosen leave-out-parameter.

6 Estimation of the drift and the volatility

Taking up where section 4 left off we turn to the estimation of the first-lag conditional means and variances of the financial time series $\{r_t\}$:

$$\begin{aligned}\mu(r) &:= \mathrm{E}[r_{t+1}|r_t = r] \\ \sigma^2(r) &:= \mathrm{E}[(r_{t+1} - \mu(r))^2|r_t = r] \\ &= \mathrm{E}[r_{t+1}^2|r_t = r] - E^2[r_{t+1}|r_t = r].\end{aligned}$$

We call μ the (conditional) drift and σ the (conditional) volatility of the process r_t. Now let g be any Borel function and $G(r) = \mathrm{E}[g(r_{t+1})|r_t = r]$. We denote by $\widehat{G}$ the kernel estimator of G as it was defined in the previous chapter:

$$\widehat{G} = \sum_{t=1}^{T-1} g(r_{t+1})\, k\left(\frac{r - r_t}{h}\right) \Big/ \sum_{t=1}^{T-1} k\left(\frac{r - r_t}{h}\right). \tag{11}$$

Setting $g(r) = r$ or $g(r) = r^2$ we thus get estimates of $\mu(r)$ and $m_2(r) = \mathrm{E}[r_{t+1}^2|r_t = r]$, $\hat{\mu}(r)$ and $\hat{m}_2(r)$. $\sigma(r)$ is consistently estimated by

$$\hat{\sigma}^2(r) = \hat{m}_2(r) - \hat{\mu}^2(r).$$

We carried out the estimation with two different kernels:

$$\begin{aligned} k_0(x) &= 0.75(1 - x^2)\, \mathrm{I}(|\mathrm{x}| \leq 1) && \text{quadratic kernel,} \\ k_1(x) &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) && \text{Gaussian kernel.}\end{aligned}$$

The quadratic kernel enjoys some optimality properties in case of the standard regression which may carry over to time series regression. The drawback of this kernel is that the denominator of (11) may be zero for some r leaving $\widehat{G}(r)$ undifined. The Gaussian kernel does not suffer from such deficiency for it is positive on the whole real line. Not surprisingly, however, the choice of the kernel did not matter much, a fact which is well known in standard regression. Much more decisive was the choice of the right bandwidth h. In determining the optimal bandwith one can use the method of cross-validation which was introduced in the preceeding chapter. We will slightly modify this method by making use of the fact that we have a whole set of similar time series of equal length at hand. Then averaging over all individual cross-validation functions may significantly reduce the variance of the estimate. In fact, similar the time series are - if they are suitably stanardized. So let $\tilde{r}_{it} = (r_{it} - m_i)/s_i$ be the standardized time series of the financial time series i $(i = 1, ..., I)$ with (unconditional) mean m_i and (unconditional) variance s_i^2. The relation between the conditional moments of the series $\{\tilde{r}_{it}\}$ and the series $\{r_{it}\}$ are given by

$$\mu_i(r) = s_i\, \tilde{\mu}\left(\frac{r - m_i}{s_i}\right) + m_i \tag{12}$$

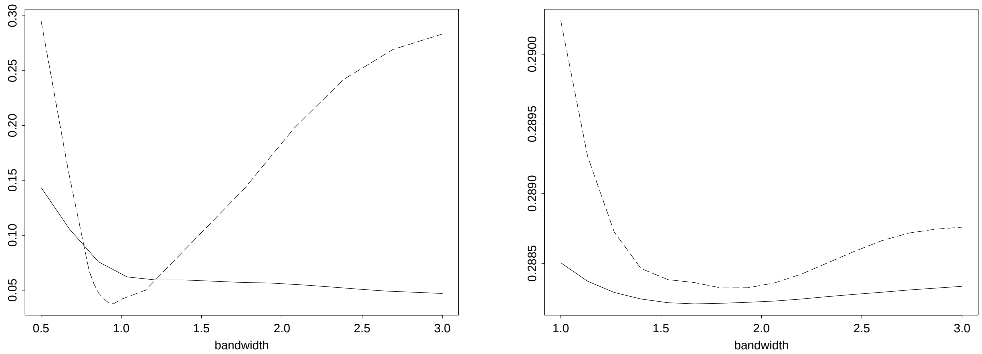
and

$$\sigma_i(r) = s_i\, \tilde{\sigma}_i\left(\frac{r - m_i}{s_i}\right). \tag{13}$$

Then estimation of μ and σ^2 may proceed via estimation of $\tilde{\mu}$ and $\tilde{\sigma}^2$ and use of equation (12) and (13). Assume now that the optimal bandwidth of $\widehat{\tilde{\mu}}_i$ $(\widehat{\tilde{m}}_{2i})$ are equal for all i. Then minimizing the average cross-validation function

$$\overline{\mathrm{CV}}(h) = \frac{1}{I}(\mathrm{CV}_1(h) + \mathrm{CV}_2(h) + ... + \mathrm{CV}_I(h)) \tag{14}$$

will also give assymptotically the optimal bandwidth h_{opt}. In finite samples however, the estimate of h_{opt} with (14) is less volatile then the estimate obtained by minimizing the individual cross-validation functions CV_i. We plotted the mean cross-validation function of $\tilde{\mu}$ and $\tilde{m}_2$ for both data sets in figure 6. Clearly, no minimum for $\tilde{\mu}$ is obtained in case of the exchange rates meaning that the best estimates of the conditional means are just the sample means of the series $\{r_{it}\}$. Though in the case of stocks the cross-validation function adopts a minimum at 1.5 the corresponding estimates of $\mu(r)$ are extremely flat and not signifantly



(a) exchange rates (b) stocks

Figure 6: Cross validation function for μ (solid line) and $\tilde{m}_2$ (dotted line)

different from a constant function. On the other hand the mean cross-validation function for $\tilde{m}_2$ is minimized at 0.95 (exchange rates) and 1.5 (stocks). A sample of the resulting estimates of the $\sigma_i(r)$ as well as their 95%-confidence bounds is shown in figure 7 and figure 8.[6] Apparently the volatility functions are not constant. Most of them are U-shaped with a minimum at 0. This means that very high or very low returns at time t will lead to a higher variance of r_{t+1}. It can also be seen that the effect of returns on the variance is not symmetrical. For the majority of exchange rates the left wings of the volatility functions are flatter than their right wings. This effect is not significant for the stock prices, where frequently the right wing is flatter than the left wing.

[6]Since boundary effects may cheat the eye we confined the plots to the range of $m_i \pm 2s_i$. This was no severe restriction since approximately 95% of the data are contained in this intervall

Figure 7: Volatility functions of exchange rates

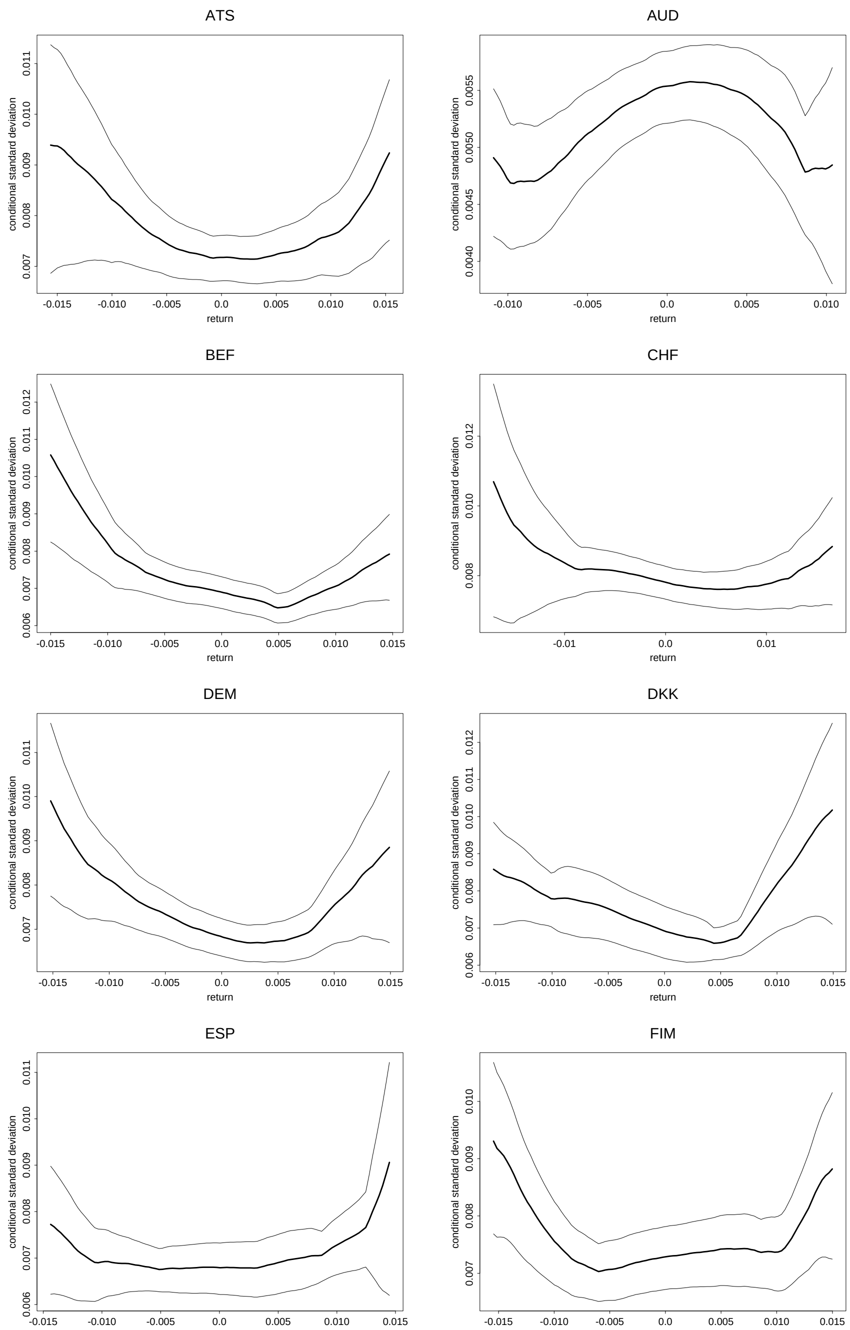


Figure 7 *continued*

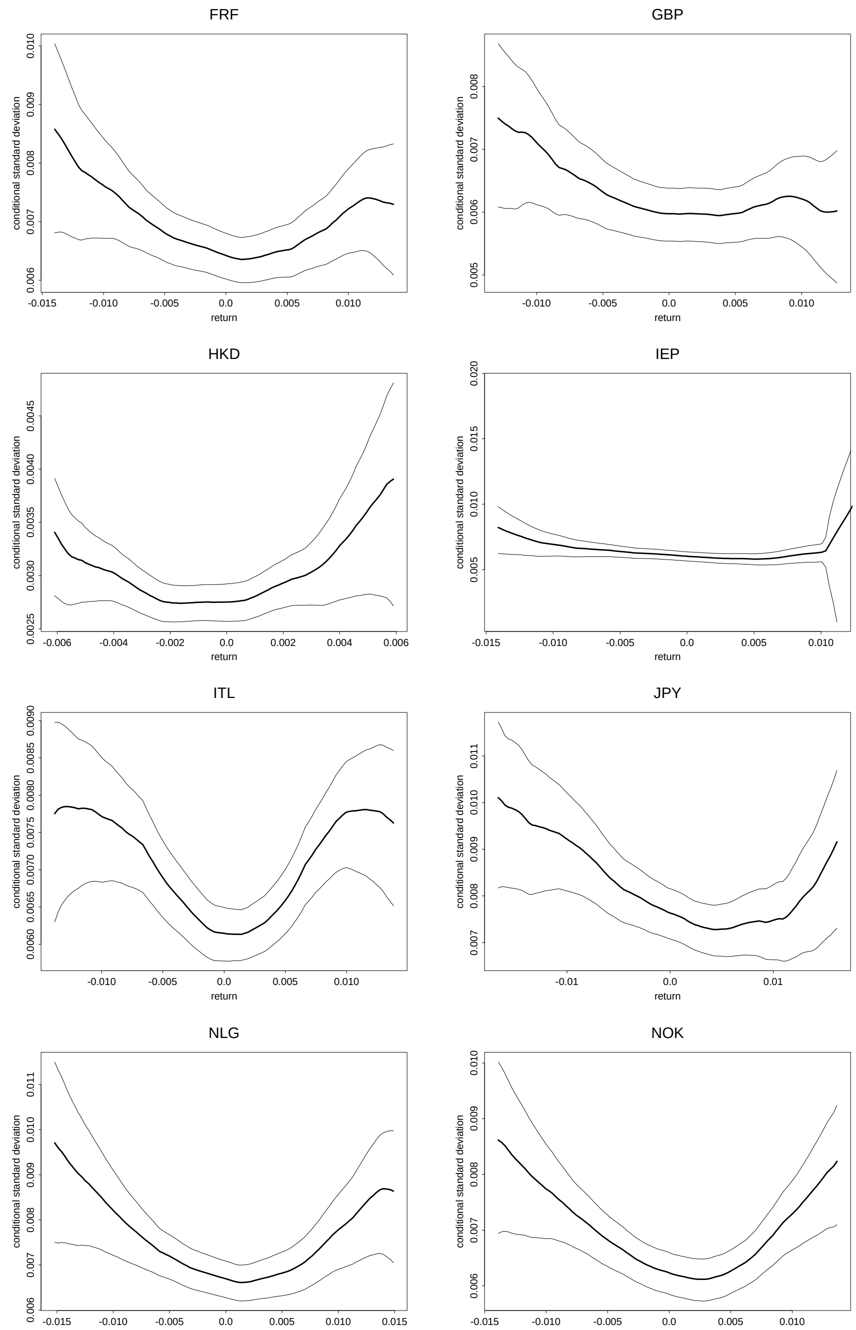


Figure 7 *continued*

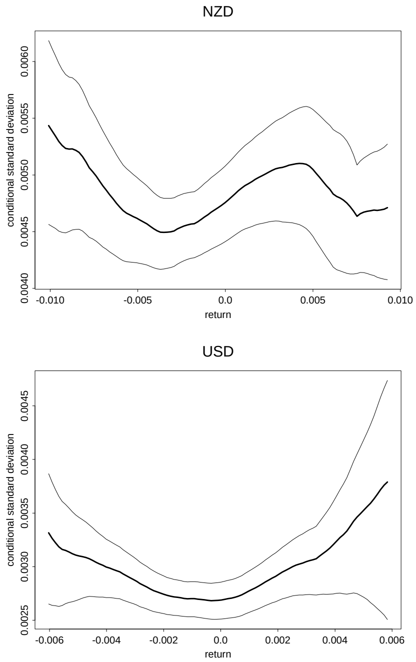


Figure 8: Volatility functions of stock returns

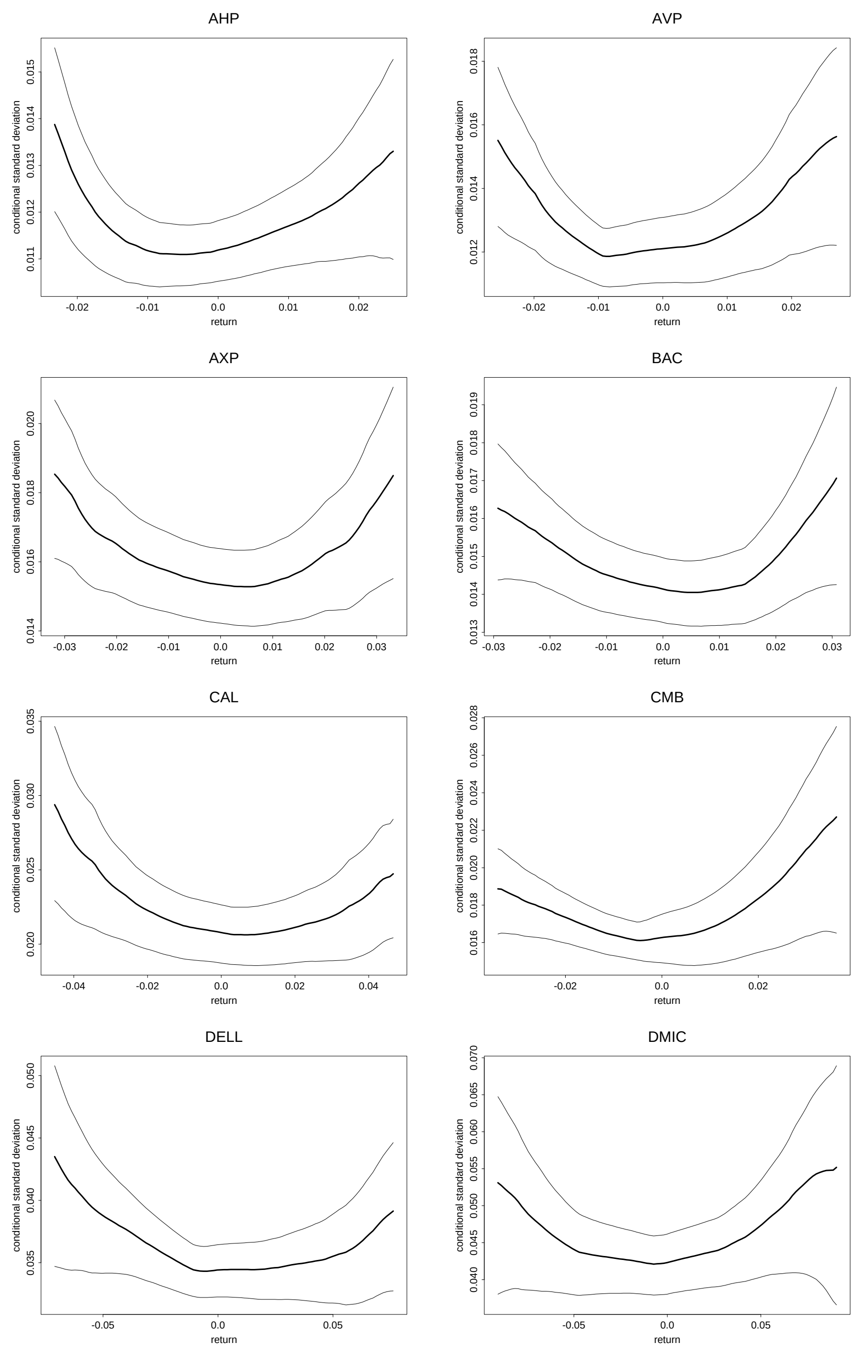


Figure 8 *continued*

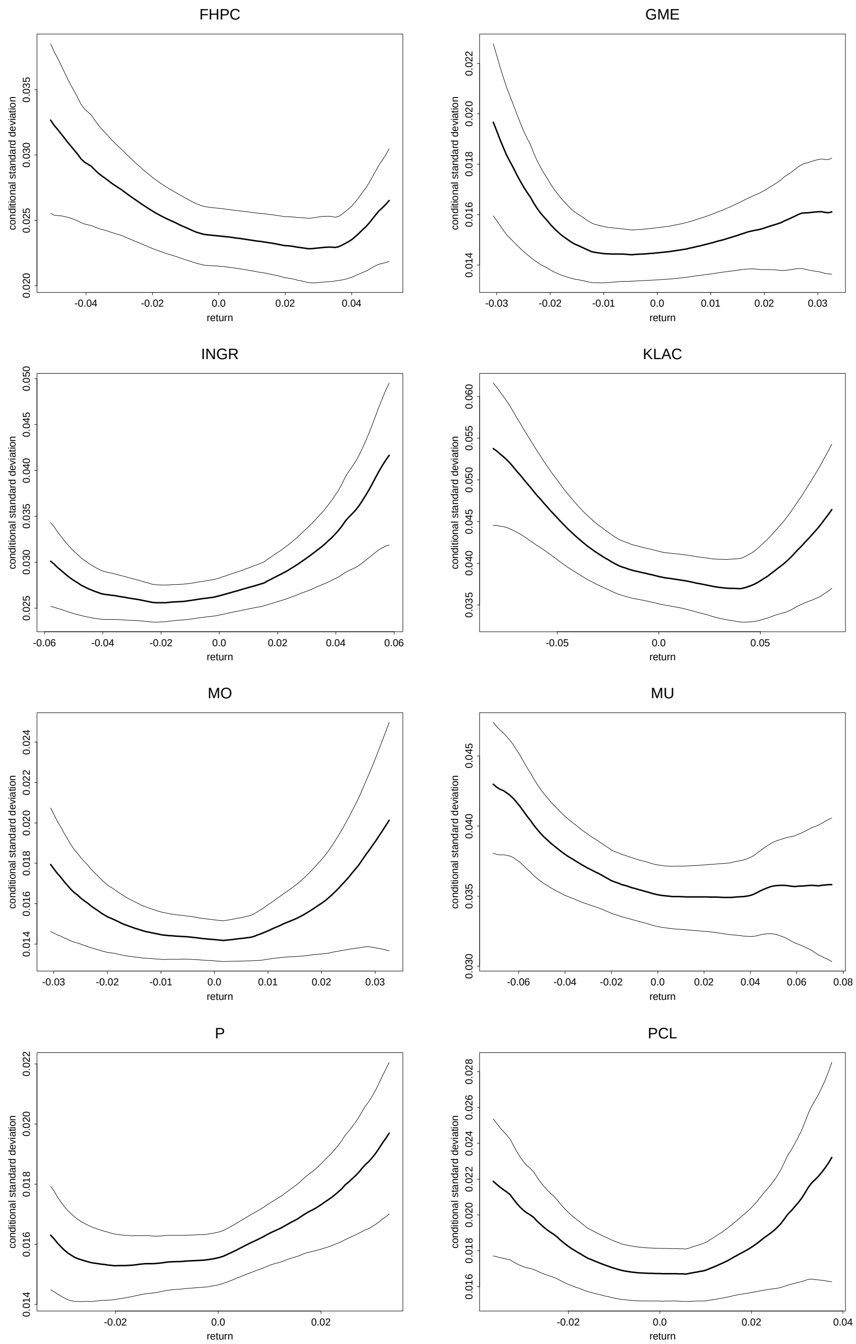
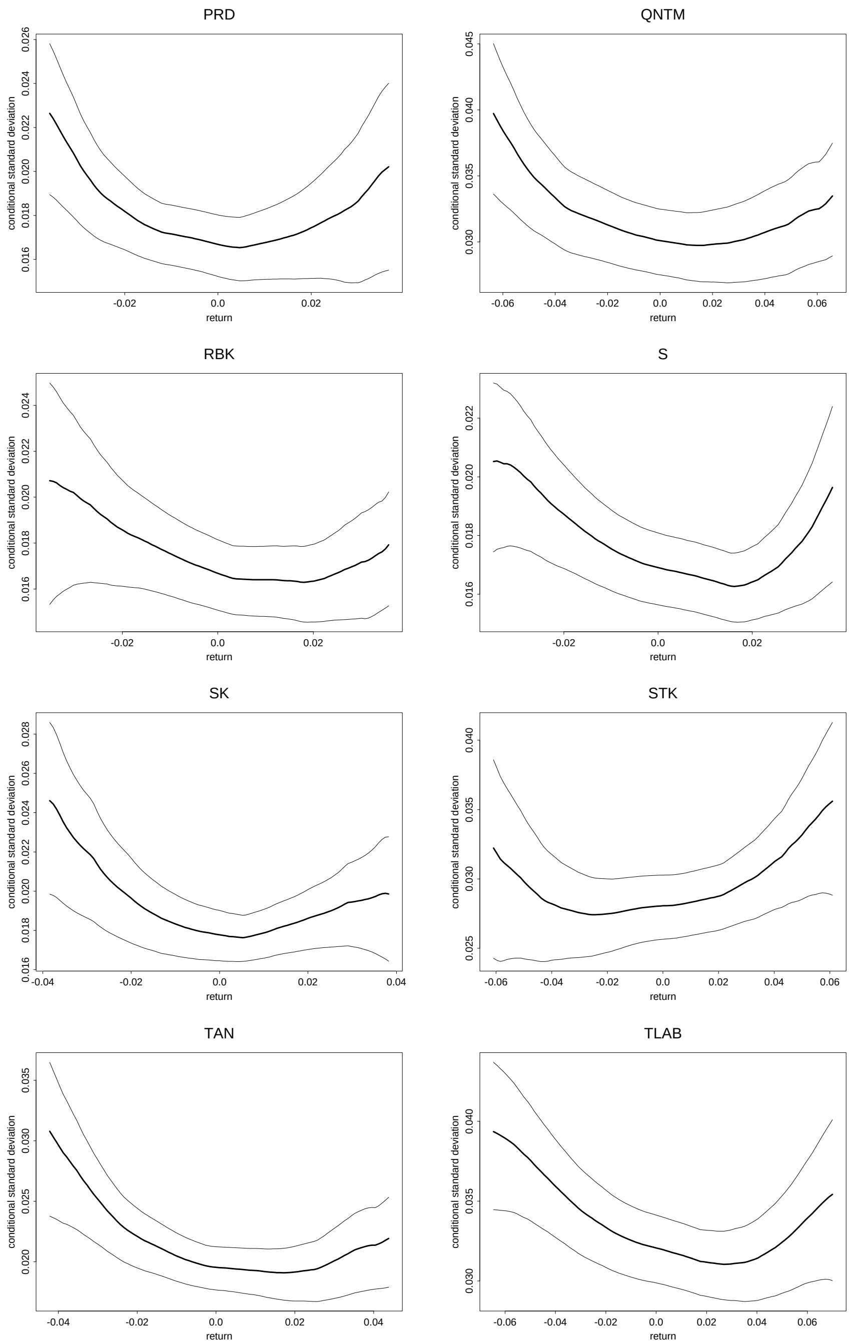


Figure 8 *continued*



7 Analysis of Residuals

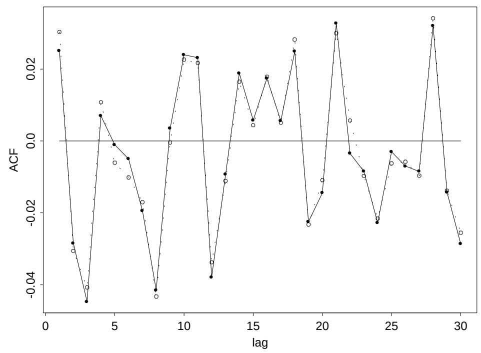
While the best one-day-off-forecast is just given by the unconditional mean its accuracy measured by σ depends significantly on the present return. The question arises if *all* information about future values is captured by μ and σ, i.e. if the process $\{\xi_t\}$ is strict white noise.[7] A first impression can be gained by inspecting figure 9 and 10, where we plotted the average correlations of the simple values (a), the squared values (b), and the absolute values (c) of the residuals (solid line). Additionally we depicted the corresponding values of the return process (dotted line). It comes as no surprise that autocorrelation of ξ_t is small since this was already the case for r_t. There is also a clear reduction of first and second lag correlation for the processes ξ_t^2 and $|\xi_t|$. Unfortunately, the reduction of correlation of higher lags is much less marked. This leads us to doubt the hypothesis of the residuals being white noise. Furthermore the autocorrelations of ξ_t^2 and $|\xi_t|$ are almost always positive[8] - something we would not expect if the residuals were white noise. The conjecture of $\{\xi_t\}$ being no white noise is confirmed by applying the tests of chapter 3 to the residual processes. If we base these tests on the first lag only the hypothesis of white noise is rarely rejected (table 4), confirming that first-lag correlation is low. Yet basing the tests on lags 1 to 30 simultanously tells a different story (table 5). Consequently, we cannot rightfully claim the residuals to be white noise.

8 Conclusion

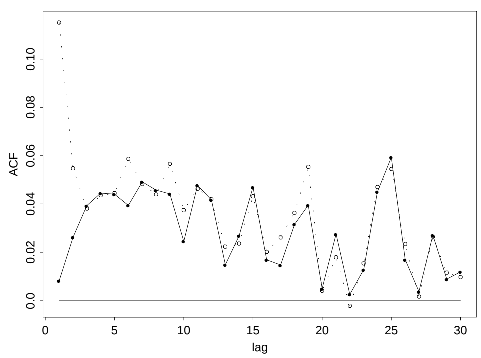
In the first part of the paper we demonstrated that stock prices as well as exchange rates are not pure random walks. Though their one-day returns show very low autocorrelation they are not independent as can be validated by investigating the autocorrelations of the squared or absolute returns. There is, in fact, a tendency of large price changes being again followed by large price changes (possibly into the other direction). This led us to model volatilities of present returns as functions of past returns. Nonparametric regression showed that while the conditional mean is constant the conditional volatility is not. Therefore we

[7]We will call ξ_t in a slight abuse of language the residual process.

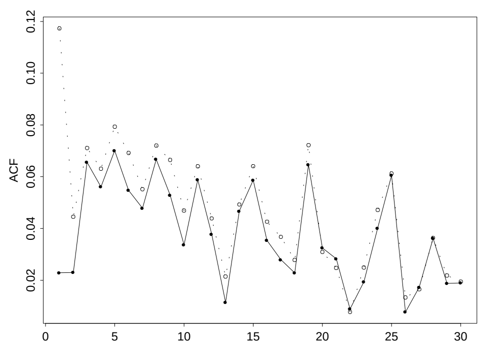
[8]This holds true even by inspecting each single autocorrelation function seperately rather than their mean.



(a) Mean correlation function of ξ_t

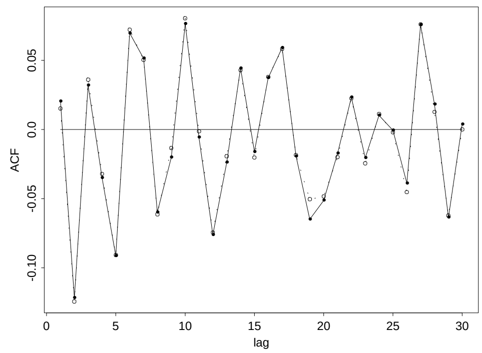


(b) Mean correlation function of ξ_t^2

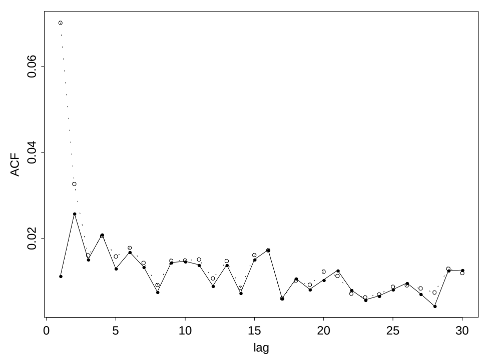


(c) Mean correlation function of $|\xi_t|$

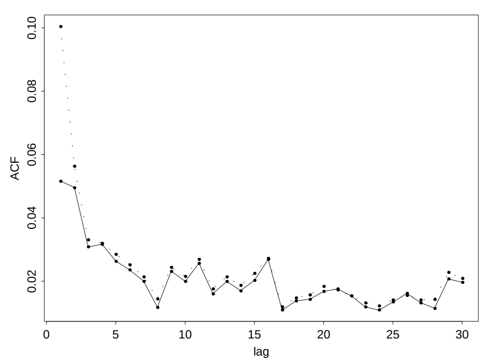
Figure 9: Mean correlation functions of ξ_t, ξ_t^2, $|\xi_t|$ (exchange rates)



(a) Mean correlation function of ξ_t



(b) Mean correlation function of ξ_t^2



(c) Mean correlation function of $|\xi_t|$

Figure 10: Mean correlation functions of ξ_t, ξ_t^2, $|\xi_t|$ (stock returns)

exchange rates	mean	rejections
residuals	0.03 (0.03)	0/19 (1/19)
squared res.	0.01 (0.12)	0/19 (16/19)
absolute res.	0.02 (0.12)	0/19 (16/19)
stocks	mean	rejections
residuals	-0.01 (-0.01)	25/175 (39/175)
squared res.	0.02 (0.07)	0/175 (64/175)
absolute res.	0.04 (0.1)	8/175 (119/175)

Table 4: lag-1-correlation of residuals (of returns in brackets)

can write

$$r_{t+1} = \bar{r} + \sigma(r_t)\xi_{t+1} \tag{15}$$

with r_t being the one-day returns, $\bar{r}$ their unconditional mean (which was shown to be equal to the conditional mean), $\sigma(r_t)$ its conditional volatility and ξ_t being some random numbers. As turned out most of the volatility functions were of similar form as they are U-shaped with a minimum close to zero. However, they are not symmetric since their left wings - at least for the exchange rates - are flatter than their right wings. Hence, high and low returns will influence future volatility differently. The reason for the volatility functions being U-shaped may be sought in the varying trading activities as large price changes may induce traders to re-arrange their portfolios which leads to higher trading activities and thus to increased volatility. This may be put to a formal test by linking returns and trading volume (for which data is also available), but this goes beyond the aim of this paper.

While $\sigma(r_t)$ captures lag-1 correlations nicely it does not explain correlations

	exchange rates	stocks
residuals	7/19 (7/19)	114/175 (102/175)
squared res.	16/19 (17/19)	72/175 (81/175)
absolute res.	17/19 (17/19)	108/175 (133/175)

Table 5: Number of rejections of the hypothesis of residuals being white noise

of higher order. In fact the residuals ξ_t are shown to be not independent, which gives scope for improving model (15). A natural way to do this would be to include more lags into the volatility function. However, this is not that easily done as one may expect, since one soon faces a severe problem which is called the curse of dimensionality. A discussion of this is deferred to another paper.

Appendix

List of exchange rates:

ATS	Austrian Schillings
AUD	Australian Dollars
BEF	Belgian Francs
CHF	Swiss Francs
DEM	German Marks
DKK	Danish Kroner
ESP	Spanish Pesetas
FIM	Finnish Markka
FRF	French Francs
GBP	British Pounds
HKD	Hong Kong Dollars
IEP	Irish Punt
ITL	Italian Lira
JPY	Japanese Yen
NLG	Dutch Guilders
NOK	Norwegian Kroner
NZD	New Zealand Dollars
SEK	Swedish Krona
USD	American Dollars

References

Bachelier, L. (1900): "Théorie de la speculation," *Annales scientifiques de l'école normale supérieure*, 17, 21–88.

Bierens, H. (1983): "Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions," *Journal of the American Statistical Association*, 78, 699–707.

Billengsley, P. (1968): *Convergence of Probability Measures*. John Wiley.

Bossaerts, P., C. Hafner, and W. Härdle (1996): "Foreign Exchange Rates have Surprising Volatility," Discussion paper 68, Humboldt-Universität zu Berlin.

Brockwell, J., and R. Davis (1991): *Time Series: Theory and Methods*. Springer.

Doukham, P., and M. Ghindes (1980): "Estimations dans le Processus $X_{n+1} = f(X_n) + \epsilon_n$," *C.R. de l'Academie des Sciences*, 291, 61–64.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Fama, E. (1965): "The behaviour of stock market prices," *Journal of Business*, 38, 34–105.

Grossman, S., and J. Stiglitz (1980): "On the impossibility of informationally efficient markets," *American Economic Review*, 70, 398–408.

Györfi, L.and Härdle, W., P. Sarda, and P. Vieu (1989): *Nonparametric Curve Estimation from Time Series*, no. 60 in Lecture Notes in Statistics. Springer Verlag.

Härdle, W., and P. Vieu (1992): "Kernel Regression Smoothing of Time Series," *Journal of Time Series Analysis*, 13, 209–232.

Härdle, W., P. Vieu, and J. Hart (1989): *Asymptotic optimal data-driven bandwidths for regression under dependence.*

NADARAYA, E. (1964): "On Estimating Regression," *Theory Prob. Appl.*, 10, 186–190.

ROBINSON, P. (1980): "Discussion of Paper by H.Tong and K.S.Lim," *J.R.Statist.Soc.*, p. 278.

——— (1983): "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185–207.

ROUSSAS, G. (1969): "Nonparametric Estimation of the the Transition Distribution Function of a Markov Process," *Ann. Math. Stat.*, 40, 1386–1400.

SHIBATA, R. (1981): "An optimal selection of regression variables," *Biometrica*, 68, 45–54.

WATSON, G. (1964): "Smooth Regression Analysis," *Sankhya, Series A*, 26, 359–372.

YAKOWITZ, S. (1979): "Nonparametric Estimation of Markov Transition Functions," *Annals of Statistics*, 7, 671–679.