Nonparametric Methods in Economics and Finance: Dependence, Causality and Prediction

This book is no. 386 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Nonparametric Methods in Economics and Finance: Dependence, Causality and Prediction

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. mr. P.F. van der Heijden
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 11 oktober 2006, te 10.00 uur

door

Valentyn Panchenko

geboren te Lviv, Oekraïne

Promotiecommissie:

Promotor:	Prof. dr. C.H. Hommes
Co-promotor:	Dr. C.G.H. Diks
Overige leden:	Prof. dr. H.P. Boswijk
	Dr. D.J.C. van Dijk
	Dr. K.J. van Garderen
	Prof. dr. ir. J.G. de Gooijer
	Prof. dr. E. Maasoumi

Faculteit Economie en Bedrijfskunde

Acknowledgements

This thesis is the result of a fruitful co-operation with my supervisor (co-promotor) Cees Diks, whom I would like to thank first and foremost. His deep multidisciplinary knowledge combined with immense energy and enthusiasm, friendly attitude and constant support were the main ingredients of my academic development and successful graduation. Furthermore, I would like to thank my promotor Cars Hommes for creating a unique working environment in the Centre for Nonlinear Dynamics in Economics and Finance (CeNDEF) and encouraging my research. Informal meetings, group trips and barbecues contributed greatly to a friendly atmosphere within CeNDEF. I also express my gratitude to the other members of my thesis committee, Peter Boswijk, Dick van Dijk, Kees Jan van Garderen, Jan de Gooijer and Essie Maasoumi for reading the manuscript and providing valuable comments.

I also thank to the Tinbergen Institute, the University of Amsterdam and the Netherlands Organisation for Scientific Research (NWO) for providing their generous financial support.

It is been a great pleasure to work in CeNDEF and the Department of Quantitative Economics. I am grateful to all my colleagues and staff members for creating a pleasant atmosphere. Thanks to Sebastiano Manzan, Sander van der Hoog, Florian Wagener, Jan Tuinstra, Roald Ramer, Maurice Bun, Frank Kleibergen and many others. I would like to express my special gratitude to my colleague, co-author and close friend Misha Anufriev, who has greatly contributed to my personal and professional development. Special thanks to my office-mate and friend Pietro Dindo, who has been dealing with me for these four years and helped me out every time I needed it.

I also would like to thank my friends and acquaintances for making my stay in Amsterdam a pleasurable experience. Thank you for taking care of my leisure time. I appreciate

your friendly support and promise to keep in touch.

And the last, but the most important acknowledgement. All of these would not have been possible without love, warm support and encouragement of my family. I am grateful to my mother Iryna, my father Aleksander, my sister Anyuta and my grandparents. They are always close to my heart.

Valentyn Panchenko
Amsterdam, August 2006

Contents

Chapter 1

General Introduction and Overview

1.1 Introduction

Time series analysis is an essential part of financial economics. It is largely applied in asset pricing, risk management, financial planning and actuarial sciences. Stimulated by freely-available vast datasets of financial time series and fast computational capabilities, the popularity of time series analysis in academic research as well as in business practice has grown over the years. See Campbell, Lo, and MacKinlay (1997) and Tsay (2002) for a comprehensive overview of recent techniques in financial econometrics. Most of the techniques heavily rely on assumptions of linear relationships, normality and/or stationarity in financial time series.

It is now widely acknowledged that many economic and financial time series exhibit nonlinear relations and/or non-Gaussian underlying processes (see e.g. Bollerslev, Engle, and Nelson, 1994 for stylised facts on assets returns). Recently suggested nonlinear models (see Franses and Van Dijk, 2000, for an overview) attempted to model nonlinear features of financial time series. Despite greater generality compared to linear models, the application of nonlinear models still requires imposing distributional and relational (functional form) conditions. At the same time it is not always possible to verify whether these assumptions hold. Often, the inference procedure itself, heavily relies on the underlying conditions.

Nonparametric methods, which relax distributional and functional form assumptions, foster even greater flexibility in modelling financial time series. The first nonparametric

techniques date back to the end of 19th century. The term "histogram", a basic nonparametric technique, was introduced by Karl Pearson in 1891 (Stigler, 1986). The concept of rank correlation was suggested in the works of Spearman (1904) and Kendall (1938). Hoeffding (1948a) continued this work in the area of nonparametric dependence measures. Kernel smoothing, a popular nonparametric technique for density estimation, was established by Rosenblatt (1956a) and Parzen (1962). Nonparametric methods are generally computationally intensive. Therefore, practical applications of these methods were not feasible until recently. With the rapid development in computer power, the introduction of supercomputer systems and parallel computing, nonparametric methods achieved wider acknowledgement in applied econometrics. The interested reader is referred to Fan (2000, 2005) for extensive overviews of recent nonparametric methods.

This thesis covers a range of topics in nonparametric and semiparametric methods. All methods discussed here are illustrated using financial time series. However, the methods are general enough to be applied in many fields where data analysis is required. The thesis is organised in increasing order of the information revealed from the observed data. In the first two chapters we discuss the issues of dependence, i.e. nonparametric testing for serial correlations and goodness-of-fit of copulas. Next, we try to reveal the direction of dependence, by focusing on causality. The corresponding chapters discuss issues related to testing for Granger non-causality and improvements in the methodology. The last chapter of the thesis is concerned with specifying multivariate models for predicting future observations conditional on those observed in the recent past. Here we try to obtain information not only about the presence of dependence and its direction, but also the ability to predict the future.

While covering the above topics we make use of various statistical methods and concepts, including Monte Carlo tests, bootstraps, U-statistics, asymptotic theory, among others. We recommend the interested reader the following sources for background information on Monte Carlo tests – Diks and Tong (1999) and Kiviet and Dufour (1997), bootstrap – Efron and Tibshirani (1993) and Horowitz (2001), U-statistics and asymptotic theory – Serfling (1980) and Van der Vaart (1998).

1.2 Review of major concepts

As an introduction, we here provide a short review of some major concepts which are used in several chapters. The concept of weak dependence is important throughout the entire thesis, since we are working with time-series data. The notions of U- and V-statistics are used to derive consistent estimators of certain parameters of interest in Chapters $2-4$. The theory of copulas is employed in Chapters 3 and 6.

1.2.1 Weak dependence and mixing

The cornerstones of asymptotic theory, the laws of large numbers and the central limit theorem, have been originally established under the assumption of independent identically distributed (i.i.d.) random variables. In the context of time series this assumption is practically always violated. Therefore, there is a continuing search for conditions weaker than independence, under which the above theorems, or variants thereof, hold. Here, we provide a brief review of the developments in this area, mainly based on the survey of Bradley (1986, 2005). We refer the interested reader to this source for a more detailed exposition.

Let $\{X_t\},\ t \in \mathbb{Z}$ denote a strictly stationary process with values generated in the probability space $(\Omega, \mathcal{F}, P)$. Define a σ-field $\mathcal{F}_J^L := \sigma(X_t, J \leq t \leq L) \subset \mathcal{F}$ for $J, L \in \mathbb{Z}, -\infty \leq J \leq L \leq \infty$. Intuitively, $\mathcal{F}_{-\infty}^0$ represents the past and $\mathcal{F}_n^\infty$ the n-step ahead future. For each $n \geq 1$, define the following dependence coefficients (below an expression $\sup_{x \in X} g(x)$ is written as $\sup g(x),\ x \in X$):

$$\alpha(n) := \sup |P(A \cap B) - P(A)P(B)|, \quad A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty; \tag{1.1}$$

$$\beta(n) := \sup \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)|, \quad A_i \in \mathcal{F}_{-\infty}^0, B_j \in \mathcal{F}_n^\infty; \tag{1.2}$$

$$\rho(n) := \sup |\mathrm{Corr}(X, Y)|, \quad X \in \mathcal{L}^2_{\mathbb{R}}(\mathcal{F}_{-\infty}^0), Y \in \mathcal{L}^2_{\mathbb{R}}(\mathcal{F}_n^\infty); \tag{1.3}$$

$$\phi(n) := \sup |P(B|A) - P(B)|, \quad A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty, P(A) > 0; \tag{1.4}$$

$$\psi(n) := \sup \left| \frac{P(A \cap B)}{P(A)P(B)} - 1 \right|, \quad A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{n}^{\infty}, P(A) > 0, P(B) > 0; \quad (1.5)$$

$$\psi^{*}(n) := \sup \frac{P(A \cap B)}{P(A)P(B)}, \quad A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{n}^{\infty}, P(A) > 0, P(B) > 0; \quad (1.6)$$

$$\psi^{'}(n) := \inf \frac{P(A \cap B)}{P(A)P(B)}, \quad A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{n}^{\infty}, P(A) > 0, P(B) > 0; \quad (1.7)$$

$$I(n) := \sup \sum_{i=1}^{I} \sum_{j=1}^{J} P(A_i \cap B_j) \log \left(\frac{P(A_i \cap B_j)}{P(A_i)P(B_j)} \right), \; A_i \in \mathcal{F}_{-\infty}^{0}, B_j \in \mathcal{F}_{n}^{\infty}. \quad (1.8)$$

In Eq. (1.3), $\mathcal{L}^2_{\mathbb{R}}(\mathcal{A})$ denotes the space of square-integrable, $\mathcal{A}$-measurable real-value random variables. In Eqs. (1.2) and (1.8) the supremum is taken over all pairs of partitions $A_1, \ldots, A_I$ and $B_1, \ldots, B_J$ of Ω such that $A_i \in \mathcal{F}_{-\infty}^{0}$ for all i and $B_j \in \mathcal{F}_{n}^{\infty}$ for all j and in Eq. (1.8) we assume that $0/0 := 0, 0 \log 0 := 0$.

Eqs. (1.1–1.8) define measures of dependence between the corresponding σ-fields. Intuitively, they measure dependence between the past and future observations. To derive asymptotic results for mixing processes, we are interested in the rate at which the past observations become independent of the future series as we increase the gap n between them.

Definition 1.1 (Mixing processes) *The stochastic process* $\{X_t\}, \; t \in \mathbb{Z}$ *is said to be*
"strongly mixing" (or "α-mixing") if $\alpha(n) \to 0$ *as* $n \to \infty$,
"absolutely regular" (or "β-mixing") if $\beta(n) \to 0$ *as* $n \to \infty$,
"ρ-mixing" if $\rho(n) \to 0$ *as* $n \to \infty$,
"uniform mixing" (or "ϕ-mixing") if $\phi(n) \to 0$ *as* $n \to \infty$,
"ψ-mixing" if $\psi(n) \to 0$ *as* $n \to \infty$,
"ψ^*-mixing" if $\psi^*(n) \to 1$ *as* $n \to \infty$,
"$\psi^{'}$-mixing" if $\psi^{'}(n) \to 1$ *as* $n \to \infty$,
"information regular" if $I(n) \to 0$ *as* $n \to \infty$.

Many processes used in recent time series modelling satisfy one or more of the mixing conditions above. For example, strictly stationary auto-regressive processes of order one (AR(1) processes), with innovations drawn from a continuous distribution satisfy

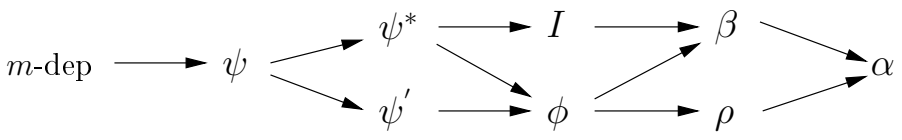

Figure 1.1: *The hierarchy of mixing conditions. An arrow pointing, say, from a to b indicates that condition a implies condition b.*

strong mixing. The popular auto-regressive heteroskedastic processes, ARCH, and their generalisations, GARCH-type processes, also satisfy strong mixing under some regularity conditions (see e.g. Carrasco and Xiaohong, 2002).

Definition 1.2 (m-dependence) *The stochastic process $\{X_t\}$, $t \in \mathbb{Z}$ is said to be m-dependent if the σ-fields $\sigma(\{X_t\}_{t=-\infty}^{0})$ and $\sigma(\{X_t\}_{t=m}^{\infty})$ are independent.*

An obvious example of an m-dependent process is the moving average process of order m, or shortly, the MA(m) process.

Figure 1.1 illustrates the hierarchy of mixing conditions. The strong mixing (or α-mixing) is the weakest condition among all mixing conditions presented above. Therefore, jointly they are often referred to as strong mixing conditions. Strong mixing (α-mixing) was used by Rosenblatt (1956b) to prove the central limit theorem for weakly dependent data. Most of the results of modern time series analysis rely on this result and its variants, see Doukhan (1994) and Doukhan and Louhichi (1999).

1.2.2 U- and V-statistics

The term U-statistic (U stands for unbiased), was introduced by Hoeffding (1948b), while Von Mises (1947) suggested a related statistic, called V-statistic. They are important statistical concepts which are used to construct estimators of population parameters. Let X denote a (possibly multivariate) random variable taking values in some space $\mathbb{M}$ (typically $\mathbb{M}$ will represent the m-dimensional Euclidean space $\mathbb{R}^m$) with cumulative distribution function (CDF) $F(\cdot)$. The starting point is a measurable kernel function $h : \mathbb{M}^r \to \mathbb{R}$, that is symmetric in its arguments. We are interested in estimating the

population parameter

$$\theta(F) \;=\; E[h(X_1,\ldots,X_r)] = \int_{\mathbb{R}^r} h(x_1,\ldots,x_r)\,dF(x_1)\cdots dF(x_r).$$

Suppose we observe a (possibly multivariate) sample $\{X_t\}_{t=1}^n$. Then, a U-statistic estimator, U_n, is an unbiased estimator of the population parameter θ and is given by

$$U_n \;=\; \binom{n}{r}^{-1} \sum_{1\le t_1<t_2<\cdots<t_r\le n} \cdots \sum h(X_{t_1},\ldots,X_{t_r}), \qquad n \ge r.$$

The sample mean is the simplest example of a U-statistic estimator, that is $U_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. The kernel function is $h(x) = x$ in this case. A more involved commonly known unbiased estimator of sample variance is also a U-statistic with the kernel $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$, i.e.

$$U_n = S_n^2 = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} h(X_i, X_j) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

A V-statistic estimator, V_n is defined as

$$V_n \;=\; n^{-r} \sum_{t_1=1}^{n} \cdots \sum_{t_r=1}^{n} h(X_{t_1},\ldots,X_{t_r}).$$

Next, we briefly discuss the asymptotic distribution of the U- and V-statistic estimators. Define an array of functions $h_c,\ c = 0,\ldots,r$ related to h, i.e.

$$h_c(x_1,\ldots,x_c) \;=\; E[h(x_1,\ldots,x_c,X_{c+1},\ldots,X_r)].$$

Note that $h_0 = \theta$ and $h_r(x_1,\ldots,x_r) = h(x_1,\ldots,x_r)$.

It can be shown (see e.g. Hoeffding, 1948b) that the variance of a U-statistic U_n depends on the variances σ_c^2 of the functions h_c. A kernel h is called degenerate (of the degree k) for a given CDF F, if $\sigma_1^2 = 0$ ($\sigma_1^2 = \cdots = \sigma_k^2 = 0$, while $\sigma_{k+1}^2 \neq 0$) and it is

nondegenerate otherwise.

To derive the asymptotic distribution of the U-statistic Hoeffding (1948b, Theorem 7.1) assumes independent identically distributed (i.i.d.) random variables X_i, while Yoshihara (1976, Theorem 1) and Denker and Keller (1983, Theorem 1(c)) relax the assumption of independence to accommodate weakly dependent processes, in particular they assume strictly stationary absolutely regular processes, or β-mixing (see Subsection 1.2.1).

Theorem 1.1 *For a nondegenerate kernel $h : \mathbb{M}^r \to \mathbb{R}$ under the assumptions of i.i.d. or stationary absolutely regular process, the asymptotic distribution of U_n is defined by*

$$\sqrt{n}(U_n - \theta) \xrightarrow{d} N(0, r^2\sigma_1^2)$$

A sample estimate of σ_1^2 can be computed relatively easy.

In the case of a degenerate kernel h, Theorem 1.1 does not apply. In this case the asymptotic distribution of U_n becomes more complicated and is given by

$$n(U_n - \theta) \xrightarrow{d} \binom{r}{c} \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1),$$

where λ_j are the eigenvalues for the kernel $h_2(x_1, x_2) - \theta$ and the Z_j are independent standard normal.

It can be shown that U- and V-statistics based on the same kernel h are asymptotically equivalent, i.e. $\sqrt{n}(U_n - V_n) \xrightarrow{p} 0$ and therefore all results discussed above apply to the V-statistic. We refer the interested reader to Serfling (1980, chapters 5 and 6) for this result, as well as a more detailed exposition of this topic.

1.2.3 Copula Theory

Here we present a general overview of copula theory and provide some basic definitions, major theorems and essential properties. For more information we refer the interested reader to the monograph by Nelsen (1999).

We start with a simple bivariate example that is generalised further in this section. Let $F(x_1, x_2)$ denote the joint CDF of a bivariate random variable $X = (X_1, X_2)'$. The corresponding marginal CDFs are denoted by $F_1(x_1)$ and $F_2(x_2)$. Since it is usually easier to handle marginal distributions separately, we are interested in a function that can reconstruct the joint distribution with CDF $F(x_1, x_2)$ from the two marginals $F_1(x_1)$ and $F_2(x_2)$. Actually, there exist such a function, which is called copula. This function fully describes the dependence between the marginal random variables X_1 and X_2. For instance, if we start from a bivariate Gaussian distribution with the correlation coefficient ρ, the corresponding copula will be a function of ρ and two marginal normal distributions.

The term *copula* was first introduced by Sklar (1959) to describe a function that "couples" univariate marginal distributions to a multivariate joint distribution.

Definition 1.3 (Copula) *A copula function $C(u)$ is a CDF of a probability measure with support in $[0,1]^m$ and uniform marginal CDFs, i.e.*

$$C(u_1, u_2, \ldots, u_m) \;=\; P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_m \leq u_m),$$

for some random variable $U = (U_1, U_2, \ldots, U_m)'$ with uniform marginals $U_i,\ i = 1, \ldots, m$.

The following result is cardinal to the theory of copulas.

Theorem 1.2 (Sklar's theorem) *Given an m-dimensional joint CDF, F, on $\mathbb{R}^m$ with marginal distribution functions $F_1, F_2, \ldots, F_m$, there exists an m-copula C such that for all $x \in \mathbb{R}^m$*

$$F(x_1, x_2, \ldots, x_m) \;=\; C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)). \tag{1.9}$$

Moreover, if the marginal distributions are continuous the copula C is unique.

For the proof of Theorem 1.2 we refer the reader to Sklar (1996). From Theorem 1.2 one can derive a method to specify a parametric copula, known as the inversion method.

Corollary 1.1 (Inversion method) *Let X be a multivariate continuous random variable with joint CDF, F, and let $F_i^{-1}(u) = \min\{x|u \leq F_i(x)\}$ denote the (quasi)-inverse of the marginal CDF, F_i. Then the copula of X is given by*

$$C(u_1, u_2, \ldots, u_m) \quad = \quad F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_m^{-1}(u_m)).$$

The inversion method allows for constructing a copula from a known multivariate distribution. By taking F to be a multivariate Gaussian CDF, or a multivariate Student t CDF, one may construct a Gaussian or Student t copula respectively (examples of these copulas are shown in terms of copula densities below). The inversion method is not the only method of copula construction. Copulas of lower dimensions (e.g. 2 or 3) can be constructed using algebraic or geometric methods. An extension of these copulas, with exception of Archimedean copulas, to higher dimensions is not trivial though. See Nelsen (1999) for more results.

One of the essential properties of copulas is their invariance under strictly increasing transformations of X_i.

Theorem 1.3 (Invariance) *Let $X = (X_1, X_2, \ldots, X_m)'$ denote a continuous multivariate random variable with copula C and let $T_1, T_2, \ldots, T_m$ be strictly increasing functions. Then the transformed random variable $T(X) = (T_1(X_1), T_2(X_2), \ldots, T_m(X_m))'$ has exactly the same copula C as X.*

Theorem 1.3 follows from Corollary 1.1 and the properties of the distribution function. As a consequence of Theorem 1.3, the dependence between the random variables $X_1, X_2, \ldots, X_m$ is completely captured by the copula regardless of the scale in which each random variable X_n is measured. Therefore any dependence measure may be based directly on the copula.

Definition 1.4 (Copula density) *Suppose $C(u_1, u_2, \ldots, u_m)$ is a copula function of a continuous random variable, then the copula density $c(u_1, u_2, \ldots, u_m)$ is defined as follows*

$$c(u_1, u_2, \ldots, u_m) \quad = \quad \frac{\partial^m C(u_1, u_2, \ldots, u_m)}{\partial u_1 \cdots \partial u_m}.$$

By taking partial derivatives of the distribution function $F(x)$ and the corresponding copula representation (1.9) with respect to $x_i,\ i = 1, \ldots, m$, we derive the following decomposition of the multivariate density function:

$$f(x) \quad = \quad c(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)) \prod_{i=1}^{m} f_i(x_i), \tag{1.10}$$

where $f_i(x_i)$ is the density associated with the marginal CDF $F_i(x_i)$. This decomposition is important for copula parameter estimation using the maximum likelihood approach, because it provides an explicit expression for the likelihood function in terms of the copula density and the product marginal densities. It is also useful for deriving the copula density from known multivariate densities. For instance, the expression for the multivariate Gaussian density with zero mean and the variance-covariance matrix Ω is given by $f(x, \Omega) = |\Omega|^{-1/2} (2\pi)^{-m/2} \exp(-\frac{1}{2}(x'\Omega^{-1}x))$. According to Eq. (1.10) one can derive the Gaussian copula density by dividing the multivariate Gaussian density with *dependent marginals* and the variance-covariance matrix Ω by the multivariate Gaussian density with *independent marginals* and, hence, the diagonal variance part, Ω_0, of the variance-covariance matrix Ω. The *Gaussian copula density* then turns out to be given by

$$c(u, \Sigma) = |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} (\Phi_m^{-1}(u))' (\Sigma^{-1} - \mathbb{I}) \Phi_m^{-1}(u) \right),$$

where $\mathbb{I}$ is the identity matrix, Σ is the correlation matrix, $\Phi_m^{-1}(u) = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m))'$, and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal CDF.

Similarly one can derive an expression for the Student t copula. The starting point is the m-dimensional joint Student t density with zero mean, correlation matrix Σ and number of degrees of freedom $\nu > 2$ (equal for all marginal densities), which is given by

$$f(x, \Sigma, \nu) = |\Sigma|^{-1/2} (\pi\nu)^{-m/2} \frac{\Gamma([\nu + m]/2)}{\Gamma(\nu/2)} \left(1 + \frac{x'\Sigma^{-1}x}{\nu} \right)^{-(\nu+m)/2}.$$

Performing the same manipulations one can compute the *Student t copula density*, which takes the form

$$c(u,\Sigma,\nu) = |\Sigma|^{-1/2}\,\frac{\Gamma([\nu+m]/2)\Gamma^{m-1}(\nu/2)}{\Gamma^{m}((\nu+1)/2)}\,\frac{\left(1+\dfrac{(T_{m,\nu}^{-1}(u))'\Sigma^{-1}T_{m,\nu}^{-1}(u)}{\nu}\right)^{-(\nu+m)/2}}{\prod_{i=1}^{m}\left(1+\dfrac{(T_{\nu}^{-1}(u_i))^2}{\nu}\right)^{-(\nu+1)/2}},$$

where $T_{m,\nu}^{-1}(u) = (T_{\nu}^{-1}(u_1),\ldots,T_{\nu}^{-1}(u_m))'$, and $T_{\nu}^{-1}(\cdot)$ is the inverse of the univariate Student t CDF with number of degrees of freedom ν.

1.3 Chapter overview

The sequel of this thesis contains a collection of papers organised in five chapters. Each chapter is based on a separate paper and is mostly self-contained. Therefore some basic definitions may be briefly restated in subsequent chapters. We tried to keep the notation uniform throughout the thesis, although some deviations between chapters are possible.

Chapter 2 is based on the paper by Diks and Panchenko (2006a). The aim of this chapter is to introduce tests for serial independence using kernel-based quadratic forms to measure the divergence between the observed and the hypothetical distribution. This separates the problem of consistently estimating a divergence measure from that of consistently estimating the underlying joint densities, the existence of which is no longer required. We obtain exact level tests by implementing a Monte Carlo procedure using permutations of the original observations. We also address the bandwidth selection problem by introducing a multiple bandwidth procedure based on a range of different bandwidth values. After numerically establishing that the tests perform well compared to existing nonparametric tests for serial independence, we discuss applications to residuals of time series fitted to parametric models. We illustrate the approach with an application to financial returns data.

Chapter 3 is an adaptation of the papers by Panchenko (2005a) and Panchenko (2005b). This chapter develops a goodness-of-fit test for copulas based on quadratic forms. The suggested test avoids the use of plug-in estimators that is the common practice in the literature. The test is applied to datasets of US large cap stocks and DAX

index constituencies to assess the performance of the Gaussian copula for modelling the dependence between the returns of multiple assets. We find that the Gaussian copula is inadequate to characterise the dependence between asset returns.

Chapter 4 is based on the paper by Diks and Panchenko (2005). We address a consistency problem in the commonly used nonparametric test for Granger causality developed by Hiemstra and Jones (1994). In this chapter we show that the relationship tested is not implied by the null hypothesis of Granger non-causality. Monte Carlo simulations using processes satisfying the null hypothesis show that, for a given nominal size, the actual rejection rate may tend to one as the sample size increases. Our results imply that evidence for nonlinear Granger causality reported in the applied empirical literature should be re-interpreted.

Chapter 5 is based on the paper by Diks and Panchenko (2006b). In this chapter we introduce a new nonparametric test for Granger non-causality which avoids the over-rejection of the Hiemstra and Jones (1994) test observed in Chapter 4. We study the reason behind this observed phenomenon analytically. It turns out that the Hiemstra-Jones test for the null of Granger non-causality, which can be rephrased in terms of conditional independence of two vectors X and Z given a third vector Y, is sensitive to variations in the conditional distributions of X and Z given Y that may be present under the null. To overcome this problem we replace the global test statistic by an average of local conditional dependence measures. By letting the bandwidth tend to zero at appropriate rates, the variations in the conditional distributions are accounted for automatically. Based on asymptotic theory we formulate practical guidelines for selecting the bandwidth depending on the sample size. We conclude with an application to historical returns and trading volumes of the Standard and Poor's index which indicates that the evidence for volume Granger-causing returns is weaker than suggested by the Hiemstra-Jones test.

In Chapter 6, we suggest a new semiparametric procedure for estimating multivariate models with conditioning variables. The semiparametric model is based on the parametric conditional copula of Patton (2005a) and nonparametric conditional marginals. To avoid the curse of dimensionality in the estimation of the latter, we propose a dimension reduction technique. The marginals are estimated using conditional kernel smoothers based on

a local linear estimator. To model the dependence between multivariate time series we use the Gaussian copula with a time-varying correlation matrix. This semiparametric copula model is compared with the parametric dynamic conditional correlations (DCC) model of Engle and Sheppard (2001) using predictive likelihood as a criterion. The comparison is based on the recent conditional test for predictive abilities of Giacomini and White (2006). We use various simulations and financial time series to compare the methods and indicate when the proposed semiparametric model is expected to be superior to the fully parametric DCC model.

Finally, we provide a summary of our findings and suggest directions for future research.

Chapter 2

Nonparametric Tests for Serial Independence based on Quadratic Forms

2.1 Introduction

Tests for serial independence are important diagnostic tools for time series modelling. A classical example is the Durbin and Watson (1950, 1951) test for serial correlation, while Li (2004) describes a range of more recent diagnostic checks in time series. Because in many applied fields it has been realised that the time series processes encountered are nonlinear and non-Gaussian, nonparametric measures of dependence are currently receiving much interest. In fact nonparametric tests for independence date back to Hoeffding (1948a), followed by the work of, among others, Heer (1991), Rosenblatt and Wahlen (1992) and Ahmad and Li (1997). However, the methods considered there do not directly provide tests for serial independence in time series. In the time series context several different nonparametric approaches have been taken. For instance, Delgado and Mora (2000) and some references therein use the empirical distribution function, while Hong (2000) suggested a frequency domain approach for testing serial independence. The approach we consider in this chapter is more closely related to the nonparametric kernel-based BDS test of Brock, Dechert, Scheinkman, and LeBaron (1996). Before going into details, the

hypothesis of serial independence in time series contexts is made explicit.

We consider a strictly stationary real-valued time series process $\{X_t\}$, $t \in \mathbb{Z}$. We introduce the m-dimensional lag ℓ delay vectors $X_t^{m,\ell} = (X_t, X_{t+\ell}, X_{t+2\ell}, \ldots, X_{t+(m-1)\ell})'$ and denote the time invariant probability measure of $X_t^{m,\ell}$ by ν_m, suppressing the dependence on ℓ, which we think of as being fixed. The null hypothesis of interest is that $\{X_t\}$ is serially independent, i.e. its elements are independent and identically distributed (i.i.d.). Under the null hypothesis the m-dimensional delay vector measure ν_m is equal to the product measure $\nu_1^m = \nu_1 \times \cdots \times \nu_1$ (m terms) of marginal probability measures. In cases where the probability density functions (pdfs) of $X_t^{m,\ell}$ exist, these are denoted by $f_m(x)$, where $x = (x_1, \ldots, x_m)'$, and the null hypothesis can be expressed as $f_m(x) = f_1(x_1) \times \cdots \times f_1(x_m)$.

Recently, information theoretic measures of divergence such as the Kullback-Leibler divergence and the Hellinger distance gained much attention in the literature, see Granger and Lin (1994), Hong and White (2005), Granger, Maasoumi, and Racine (2004). These measures of divergence, being defined in terms of the joint and marginal pdfs, are usually estimated on the basis of plug-in kernel estimators of the joint and marginal densities. To establish consistency of the test statistics thus obtained, it suffices to take the bandwidth according to the optimal value for kernel density estimation as in Silverman (1986), although it has been recognised that this choice need not be optimal in terms of the power of the tests. Along the same lines, Feuerverger (1993) reaches the conclusion that the consistency of the associated density estimator is not required for the consistency of the quadratic measure of Rosenblatt (1975), given by $T = \int_{\mathbb{R}^m} (\widehat{f}(x) - \widehat{g}(x))^2 dx$, where $\widehat{f}$ and $\widehat{g}$ are kernel density estimators. Anderson, Hall, and Titterington (1994) indicate that relative oversmoothing is appropriate for this type of statistic in a two-sample test. A problem related to using the consistency of the plug-in estimators is the difficulty to produce efficient kernel estimates of multivariate densities due to the curse of dimensionality.

In this chapter a different perspective on the above issues is offered by defining divergence measures between distributions using kernel-based quadratic forms. These divergence measures naturally lead to U- and V-statistic estimators (see Chapter 1), which are closely related to the statistic T. However, it becomes apparent that the bandwidth plays

an entirely different role here than in nonparametric density estimation, where it controls the trade-off between the bias and variance of the density estimators $\widehat{f}_i$. Starting from quadratic forms, a different divergence measure is associated with each fixed bandwidth value. Each of the members of this family of divergence measures, parametrised by the bandwidth, can be estimated consistently using U- or V-statistics based on a kernel function with that same particular bandwidth, even if the distributions are not continuous (and hence no underlying densities exist). Rather than acting as a bias-variance trade-off parameter, the bandwidth merely controls the length scale in the sample space at which two probability measures are compared.

It follows from the results by Denker and Keller (1983) that under strict stationarity and absolute regularity (see Chapter 1), the proposed tests for serial independence are consistent against all fixed alternatives. In contrast to some other tests based on quadratic functionals, such as that of Székely and Rizzo (2005), there is no need to impose any conditions on the moments of the time series. Although consistency does not require the bandwidth to vanish with the sample size, the approach still faces the common bandwidth selection problem. This is addressed by implementing a multiple bandwidth procedure along the lines of the approach of Horowitz and Spokoiny (2001).

In Section 2.2 we introduce the notion of the squared distance measures between probability distributions in terms of kernel-based quadratic forms and derive some of their properties. Section 2.3 describes a single bandwidth permutation test for serial independence based on squared distances and investigates how the power depends on different values of the bandwidth. Subsequently the multiple bandwidth procedure to deal with the bandwidth selection problem is described. In Section 2.4 the finite-sample performance against fixed and local alternatives is compared with some other nonparametric tests for serial independence. There we also investigate the behaviour of the test when applied to estimated residuals. After an application to the log-return series of S&P 500 stock index in Section 2.5, Section 2.6 summarises and concludes.

2.2 Quadratic forms and their estimators

In this section we briefly review the distance notions between probability measures in $\mathbb{R}^m$ which will serve as the divergence between the joint probability measure ν_m and its counterpart ν_1^m. These functionals were first introduced by Diks, Van Zwet, Takens, and DeGoede (1996) in the context of measuring the divergence between chaotic time series, and later applied in a test for symmetry Diks and Tong (1999). For two m-dimensional probability measures μ_1 and μ_2, consider a quadratic form of the type:

$$Q = \|\mu_1 - \mu_2\|^2 = (\mu_1 - \mu_2, \mu_1 - \mu_2) = (\mu_1, \mu_1) - 2(\mu_1, \mu_2) + (\mu_2, \mu_2),$$

where $(\mu_i, \mu_j) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(x-y)\mu_i(dx)\mu_j(dy)$ is a bilinear form, which can be concisely written as $(\mu_i, \mu_j) = E(K_h(X-Y))$ where X and Y are two independent m-dimensional vectors, distributed according to μ_i and μ_j, respectively. Whenever $K_h(\cdot)$ is a positive definite kernel function this bilinear form defines an inner product, and the squared distance Q defines a metric on the space of probability measures on $\mathbb{R}^m$. We typically consider kernels that factorise as $K_h(x) = \prod_{i=1}^m \kappa(x_i/h)$ where x_i refers to the the i'th element of vector x, $\kappa(\cdot)$ is a one-dimensional kernel function, which is symmetric around zero, and h is a bandwidth parameter.

Because Fourier transforms leave the L_2 norm invariant by Parseval's identity, and convolution amounts to multiplication in Fourier space, the quadratic form can be expressed as

$$\begin{aligned} Q &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_h(x-y)(\mu_1 - \mu_2)(dx)(\mu_1 - \mu_2)(dy) \\ &= \int_{\mathbb{R}^m} \tilde{K}_h(k)|(\tilde{\mu}_1 - \tilde{\mu}_2)(k)|^2 dk, \end{aligned}$$

where $\tilde{K}_h(\cdot)$ is the Fourier transform of $K_h(\cdot)$, $\tilde{\mu}_i$ the characteristic function of μ_i, and $|\cdot|$ the modulus. It follows that if $\tilde{K}_h(\cdot)$ is an integrable real-valued positive function, $K_h(\cdot)$ is positive definite, and $Q = 0$ if and only if μ_1 and μ_2 are identical probability measures, and is strictly positive otherwise. Here we focus on three specific cases of positive definite kernels: the Gaussian kernel $\kappa(x) = \exp(-x^2/4)$, as in Diks and Tong (1999), the double exponential kernel $\kappa(x) = \exp(-|x|/4)$, and the Cauchy kernel $\kappa(x) = 1/(1+x^2)$. The

factor 4 in the Gaussian and double exponential kernels is chosen for convenience as it simplifies some of the derivations discussed below.

The squared distance Q satisfies all the essential "ideal" properties of a dependence measure summarised by Granger, Maasoumi, and Racine (2004). It is well-defined for continuous as well as discrete random variables. It is nonnegative, equal to zero only in the case of independence, and can be related to the correlation coefficient ρ in the case of a bivariate normal distribution, as shown in Appendix 2.A. Since $(\cdot,\cdot)$ defines an inner product on the space of measures on $\mathbb{R}^m$, $Q^{\frac{1}{2}}$ is a real distance notion between probability measures with the usual properties, such as the triangular inequality. Although Q is not invariant under monotonic transformations of marginals, if desired, invariance of estimators can always be achieved by transforming the data to a known marginal distribution, e.g. by using empirical probability integral transforms. Moreover, in Appendix 2.B we establish the equivalence of the quadratic form Q and the quadratic measure of Rosenblatt (1975).

For convenience we introduce the short-hand notation $Q^{ij} = (\mu_i, \mu_j)$. As shown above, Q^{ij} can be expressed in terms of averages of the kernel function: $Q^{ij} = E(K_h(X-Y))$ for independent vectors $X \sim \mu_i$ and $Y \sim \mu_j$. This suggests estimating Q^{ij} using empirical averages of the values of the kernel function obtained from the data, thus leading naturally to the use of U- and V-statistics as discussed in detail in Chapter 1. For example, given an observed time series $\{X_t\}_{t=1}^T$, from which $n = T-(m-1)\ell$ delay vectors $X_t^{m,\ell}$, $t = 1,\ldots,n$ of dimension m can be constructed, for the first term Q^{11} this leads to the U-statistic estimator

$$\begin{aligned}\widehat{Q}^{11} &= \frac{2}{n(n-1)}\sum_{t=2}^{n}\sum_{s=1}^{t-1} K_h(X_t^{m,\ell} - X_s^{m,\ell}) \\ &= \frac{2}{n(n-1)}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\prod_{k=0}^{m-1}\kappa\left((X_{t+k\ell} - X_{s+k\ell})/h\right).\end{aligned}$$

For the bounded kernel functions considered here, it follows from the work of Denker and Keller (1983), Theorem 1, part (c), that under strict stationarity and absolute regularity of the time series, both U- and V-statistics are consistent and asymptotically normal. In particular this implies $\widehat{Q}^{11} \xrightarrow{p} Q^{11}$. Similarly one can construct a consistent U-statistic estimator $\widehat{C}_h(x) = \frac{1}{n}\sum_{t=1}^{n}\kappa\left((x-X_t)/h\right)$ for $E[\kappa((x-X)/h)]$ and use this to obtain

consistent estimators for Q^{12} and Q^{22} after writing these in terms of $E[\kappa((x-X)/h)]$:

$$\begin{aligned}\widehat{Q}^{12} &= \frac{1}{n}\sum_{t=1}^{n}\prod_{k=0}^{m-1}\widehat{C}_h(X_{t+k\ell}),\\ \widehat{Q}^{22} &= \frac{1}{n^m}\prod_{k=0}^{m-1}\left(\sum_{t=1}^{n}\widehat{C}_h(X_{t+k\ell})\right).\end{aligned}$$

Taken together, $\widehat{Q} = \widehat{Q}^{11} - 2\widehat{Q}^{12} + \widehat{Q}^{22}$ is a consistent estimator of Q.

Note that there is a connection with the BDS test for serial independence by Brock, Dechert, Scheinkman, and LeBaron (1996). Using the functional $Q^{11} - Q^{22}$ with kernel function $\kappa(x) = I_{(-h,h)}(x)$, which is 1 if $x \in (-h, h)$ and 0 otherwise, will lead to the BDS test, with Q^{11} playing the role of the correlation integral and Q^{22} of its value under the null hypothesis of serial independence.

Based on the theory of U-statistics one might develop asymptotic theory for the functional Q, possibly with a suitably chosen rate at which h tends to zero as $n \to \infty$ (cf. Wolff, 1994). However, as reported by Skaug and Tjøstheim (1993a), Granger, Maasoumi and Racine (2004) and Hong and White (2005) in similar testing contexts, asymptotic theory provides rather poor finite sample approximations to the null distributions of the test statistics, and inference based on such tests becomes unreliable. To avoid this problem we proceed with a permutation procedure.

2.3 Permutation test

The idea to use a permutation test in the context of serial independence dates back to Pitman (1937). Due to the decreasing cost of computing power permutation tests have gained increasing attention (for a practical exposition see Good, 2000). Under the condition of exchangeability of the observations a permutation test is exact for any sample size n, i.e. the rejection rate under the null hypothesis is equal to the nominal size α. Moreover, Hoeffding (1952) shows that under general conditions permutation tests are asymptotically as powerful as certain related parametric tests.

2.3.1 Single bandwidth

First we consider a standard procedure using a single fixed bandwidth h. Since deviations from the null lead to positive values of Q, a test based on this squared distance would reject whenever the estimate $\widehat{Q}$ is too large. Thus, a one-sided test is appropriate in this context. Conditional on the observed values of the data under the null hypothesis of serial independence, each permutation of the observed data is equally likely. We denote the estimate $\widehat{Q}$ based on the original data as $\widehat{Q}_0$. Under the null the values of $\widehat{Q}_i$, $i = 0, \ldots, B$, computed using the original data and B permutations, respectively, are exchangeable. An exact p-value (in that it is uniformly distributed on $1/(B+1), \ldots, 1$ under the null) is calculated as

$$\widehat{p} = \frac{\sum_{i=0}^{B} I(\widehat{Q}_i > \widehat{Q}_0) + L}{B+1}, \tag{2.1}$$

where $I(\cdot)$ denotes the indicator function taking the value 1 if the condition in brackets is true and 0 otherwise. Let $Z = \sum_{i=0}^{B} I(\widehat{Q}_i = \widehat{Q}_0) \geq 1$ denote the number of ties plus one. In case $Z = 1$, $L = 1$, while for $Z > 1$, for L we take a random variable, uniformly distributed on $1, \ldots, Z$. That is, each rank of $\widehat{Q}_0$ among the $\widehat{Q}_i$ that happen to be equal to $\widehat{Q}_0$, is taken to be equally probable. This is equivalent to adding a very small amount of noise to each of the $\widehat{Q}_i$'s before determining their ranks, thus making the rank of $\widehat{Q}_0$ among the $\widehat{Q}_i$ unique. If $0 < \alpha = k/(B+1) < 1$ for some integer k, rejecting whenever $\widehat{p} \leq \alpha$ yields an exact level-α test. Generally, the power of a permutation test decreases if the number of permutations B decreases. The results by Marriott (1979) indicate that little power is lost by taking $B + 1 = 5/\alpha$.

Notice that the term $\widehat{Q}^{22}$ is constant under permutations, and hence can be left out of consideration while determining the significance of $\widehat{Q}$. This reflects the fact that Q^{22} is a functional of the marginal distribution, which plays a role here as an infinite dimensional nuisance parameter.

So far we have only considered the calculation of p-values for a fixed bandwidth. To deal with the problem of bandwidth selection, Subsection 2.3.3 describes a method for determining a single p-value over a range of different bandwidth. However, we first motivate the multiple bandwidth procedure by presenting some bandwidth-related simulation

results.

2.3.2 Bandwidth-related simulations

Hereafter we refer to the bandwidth that yields the highest empirical power for a fixed size α as the optimal bandwidth h^*. We investigate the dependence of the optimal bandwidth on three parameters, namely the data generating process (DGP), the delay vector dimension m and the sample size n. A description of the DGPs used, along with broader simulation results, are presented in Section 2.4. Here we only display bandwidth-related simulations. We consider $d = 30$ different bandwidth values h_i ranging from 0.01 to 3.0, equidistant on a logarithmic scale:

$$h_i \;=\; h_{\max}(h_{\min}/h_{\max})^{\frac{d-i}{d-1}}, \qquad i = 1, \ldots, d. \tag{2.2}$$

The number of permutations was set to $B + 1 = 100$, including the original series and the number of simulations was set to $1,000$. Since the Cauchy and double exponential kernels gave similar results, we here only discuss the results for the Gaussian kernel.

Figure 2.1 shows the power as a function of the bandwidth for series of various lengths n, (left panel, DGP 1, $m = 2$, $\ell = 1$), and for various DGPs, (right panel, $n = 100$, $m = 2$, $\ell = 1$). The left panel shows no clear shift in the optimal bandwidth h^* as n increases. Similar results were observed for other DGPs. Intuitively, the reason is that the optimal bandwidth depends on the typical length scale of the differences between the joint delay vector measure ν_m and the product measure ν_1^m. As long as this length scale is not taken to decrease with n, the optimal bandwidth may asymptotically tend to some finite positive value. Analytical support for a fixed optimal bandwidth was reported by Anderson, Hall, and Titterington (1994) in a two-sample test based on a statistic of the type $T = \int_{\mathbb{R}^m}(\widehat{f}(x) - \widehat{g}(x))^2 dx$. However, the right panel of Figure 2.1 illustrates that the optimal bandwidth h^* depends on the particular DGP, e.g. for the nonlinear MA(1) process (DGP 1), $h^* \simeq 0.7$, and for the bilinear process (DGP 7), $h^* \simeq 1.2$. This suggests that using a single bandwidth value in a practical situation, when the underlying DGP is

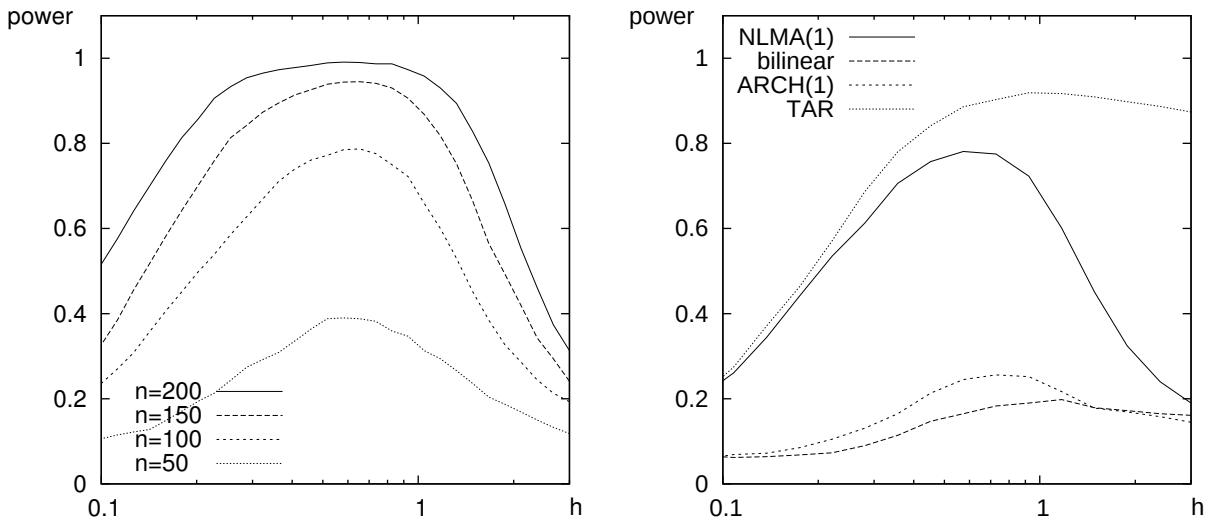

Figure 2.1: *Observed power as a function of bandwidth h. The left panel shows results for various series lengths n, for a nonlinear MA(1) process (NLMA(1), DGP 1); the right panel for various DGPs for $n = 100$. In all cases: dimension $m = 1$, lag $\ell = 1$, nominal size $\alpha = 0.05$, number of permutations $B + 1 = 100$ and number of simulations $1,000$.*

not known, may not be optimal.

2.3.3 Multiple bandwidth procedure

Motivated by the above findings, we require a procedure that produces a single test statistic (p-value) incorporating a range of bandwidth values. Horowitz and Spokoiny (2001) suggest an adaptive rate-optimal test that uses many different bandwidths. Since the theoretical distribution of their test statistic under the null is not known, they find critical values by simulation. We develop a similar procedure in the Monte Carlo context and implement it in the form of a multiple bandwidth permutation test. The procedure is based on determining the significance of the smallest single-bandwidth p-value over a range of different bandwidths, and can be summarised as follows:

1. Calculate the vector of $\widehat{Q}_{h,0}$-values for a range of bandwidths: $h \in H = \{h_1, \ldots, h_d\}$. We define h on a geometric grid as in Eq. (2.2).

2. Randomly permute the data and calculate a bootstrap vector $\widehat{Q}_{h,1}$. Repeat this B times, to obtain $\widehat{Q}_{h,i}$ for $h \in H$, and $i = 1, \ldots, B$.

3. Transform $\widehat{Q}_{h,i}$ into a p-value: $\widehat{p}_{h,i} = (\sum_{j=0}^{B} I(\widehat{Q}_{h,j} > \widehat{Q}_{h,i}) + L)/(B+1)$, with L defined similarly to Eq. (2.1).

4. Select the smallest p-value among all bandwidths and call it $\widehat{T}_i$: $\widehat{T}_i = \inf_{h \in H} \widehat{p}_{h,i}$.

5. Calculate an overall p-value on the basis of the rank of $\widehat{T}_0$ among the $\widehat{T}_i$, i.e. $\widehat{p} = (\sum_{i=0}^{B} I(\widehat{T}_i < \widehat{T}_0) + L)/(B+1)$ using a ties randomisation procedure as in Eq. (2.1).

In step 3 we pretend each of the permuted series to be the originally observed series and determine the corresponding p-values $\widehat{p}_{h,i}$ that would have been obtained for series i for each of the different bandwidths. In step 4, for each series the smallest p-value over the different bandwidths is selected (denoted by $\widehat{T}_i$, $i = 0, \ldots, B$). We finally use the exchangeability of the B series under the null to calculate an overall p-value by establishing the significance of $\widehat{T}_0$ for the actually observed data (step 5). As in the single bandwidth case, the multiple bandwidth procedure yields an exact α-level test if the null hypothesis is rejected whenever $\widehat{p} \leq \alpha$. The power of this multiple-bandwidth procedure depends on the range $R = [h_{\min}, h_{\max}]$, the number d of elements in the bandwidth set H and the number of permutations B. The range R should be wide enough to contain h^* for various DGPs. The number of bandwidths d chosen in R is important for the power. Taking d too small we risk losing the optimal bandwidth h^* through the grid. Our simulations suggest that the empirical power of the multiple bandwidth procedure reduces as the bandwidth range R becomes wider. Therefore, in practice we suggest taking $R = [0.5, 2.0]$ which includes h^* for all considered DGPs. For this range reasonable power is achieved using $d = 5$ bandwidths.

Also the number of permutations $B+1$ has an important impact on the power of our multiple bandwidth procedure. Figure 2.2 shows the power as a function of the single bandwidth in contrast to the power under the multiple bandwidth procedure for the range $R = [0.5, 2.0]$ with $d = 5$ for various numbers of permutations $B+1 = 20, 100, 500$ for the nonlinear MA(1) process (DGP 1) of length $n = 100$, for embedding dimension $m = 2$

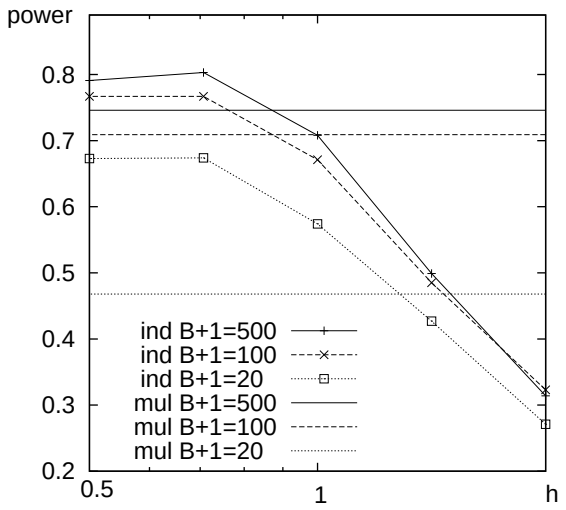

Figure 2.2: *Observed power as a function of the single bandwidth in contrast to the power under the multiple bandwidth procedure for the range* $R = [0.5, 2.0]$ *with* $d = 5$ *for various numbers of permutations* $B + 1 = 20, 100, 500$ *for the nonlinear MA(1) series (DGP 1) of length* $n = 100$*, dimension* $m = 2$*, and lag* $\ell = 1$*, nominal size* $\alpha = 0.05$ *and number of simulations* $1,000$*.*

and lag $\ell = 1$. The same process (single bandwidth) is illustrated on Figure 2.1 for a wider bandwidth range and $B + 1 = 100$. We observe that the power for the multiple bandwidth procedure is more sensitive to the number of permutations $B + 1$ than for the single bandwidth procedure. This has been observed for other DGPs. The reason for the higher sensitivity to B is that the $\widehat{T}_i$ are discrete multiples of $1/(B + 1)$, which for small B leads to many identical $\widehat{T}_i$-values (ties) which reduces the power. We find that for the considered range $R = [0.5, 2.0]$ with $d = 5$, taking $B + 1 = 100$ produces good results. These are the parameter values we recommend in practical applications of the test.

2.4 Test performance

We next investigate the power of the proposed test, hereafter called Q-test, and compare it with that of similar nonparametric tests such as the BDS test and the recent test of Granger, Maasoumi, and Racine (2004), which we refer to as the GMR test. Permutation

tests differ from asymptotic tests (based on the derived asymptotic distribution of test statistic) in that the critical value in the former is a random variable. This fact makes the analytic evaluation of its power function difficult. Hoeffding (1952) has shown that under certain conditions the random critical value of the permutation test converges in probability to a constant if the number of permutations B tends to infinity as $n \to \infty$. Relying on this fact Hoeffding (1952) investigated the large-sample power properties of permutation tests based on a relatively simple test statistic and demonstrated that under general conditions the permutation tests are asymptotically as powerful as the corresponding parametric tests. In the present context the test statistic is much more complex and the results of Hoeffding (1952) may not be directly applicable. Therefore, we rely heavily on simulations.

2.4.1 Fixed alternatives

We compare the rejection rates of the tests against fixed alternatives for the following stationary DGPs, where $\{\varepsilon_t\}$ is an i.i.d. sequence of $N(0,1)$ random variables:

DGP 0. $Y_t = \varepsilon_t$

DGP 1. $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$

DGP 2. $Y_t = \varepsilon_t + 0.6\varepsilon_{t-1}^2 + 0.6\varepsilon_{t-2}^2$

DGP 3. $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}\varepsilon_{t-2}$

DGP 4. $Y_t = 0.3Y_{t-1} + \varepsilon_t$

DGP 5. $Y_t = 0.8|Y_{t-1}|^{0.5} + \varepsilon_t$

DGP 6. $Y_t = \text{sign}(Y_{t-1}) + \varepsilon_t$

DGP 7. $Y_t = 0.6\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$

DGP 8. $Y_t = \sqrt{h_t}\varepsilon_t, \qquad h_t = 1 + 0.4Y_{t-1}^2$

DGP 9. $Y_t = \sqrt{h_t}\varepsilon_t, \qquad h_t = 0.01 + 0.80h_{t-1} + 0.15Y_{t-1}^2$

DGP 10. $$Y_t = \begin{cases} -0.5Y_{t-1} + \varepsilon_t, & Y_{t-1} < 1 \\ 0.4Y_{t-1} + \varepsilon_t, & \text{else} \end{cases}$$

DGP 11. $Y_t = 4Y_{t-1}(1 - Y_{t-1}), \qquad 0 < Y_t < 1$

DGP 12. $Y_t = 1 + 0.3Y_{t-2} - 1.4Y_{t-1}^2$

DGP 13. $Y_t = Z_t + \sigma\varepsilon_t, \qquad Z_t = 1 + 0.3Z_{t-2} - 1.4Z_{t-1}^2.$

The above DGPs or slight modifications of these were previously considered by Granger, Maasoumi, and Racine (2004), Granger and Lin (1994), Hong and White (2005), Brock, Dechert, Scheinkman, and LeBaron (1996) and others. DGP 0 satisfies the null hypothesis and is included to assess the empirical size of the tests. DGPs $1-3$ are nonlinear MA processes of order 1, 2 and 2 respectively. Granger, Maasoumi, and Racine (2004) suggested that a good measure of dependence should reflect the theoretical properties of these MA processes, i.e. zero dependence at lags beyond their nominal lags. DGP 4 is a linear AR(1) process. DGPs 5 and 6 are nonlinear AR(1) processes. The properties of DGP 6 were investigated by Granger and Teräsvirta (1999). DGP 7 is a bilinear process introduced by Granger and Andersen (1978). DGPs 8 and 9 are instances of ARCH(1) and GARCH$(1,1)$ processes proposed by Engle (1982) and Bollerslev (1986) respectively. The coefficients of the GARCH$(1,1)$ process are taken close to the corresponding estimates of Bollerslev (1986). DGP 10 is a threshold auto-regressive process (TAR) proposed by Tong (1978). DGPs 11 and 12 are the logistic map and the Hénon map respectively, generating deterministic chaotic time series, while DGP 13 is the Hénon map with additive Gaussian observational noise $\sigma\varepsilon_t$ where σ equals 20 percent of the standard deviation of the clean Hénon process. We used series of length $n = 100$ (except $n = 50$ for DGP 6 and $n = 20$ for DGPs $11 - 13$), and the total number of permutations, including the original series, was set to $B + 1 = 100$. The bandwidth set H included $d = 5$ different values in the range $R = [0.5, 2.0]$ after normalising the series to unit variance. The three different kernels mentioned earlier were used for comparison: the Gaussian, double exponential and Cauchy kernels. We considered different lags $\ell = 1, 2, 3$ for a delay vector dimension $m = 2$, and extended the delay vector dimension to $m = 3, 4, 5, 10$ for lag $\ell = 1$. All tests were conducted at a nominal size of $\alpha = 0.05$, and the number of simulations was set to $1,000$. To decrease the standard error of the estimated size, we increased the number of simulations to $5,000$ for DGP 0, which is true under the null.

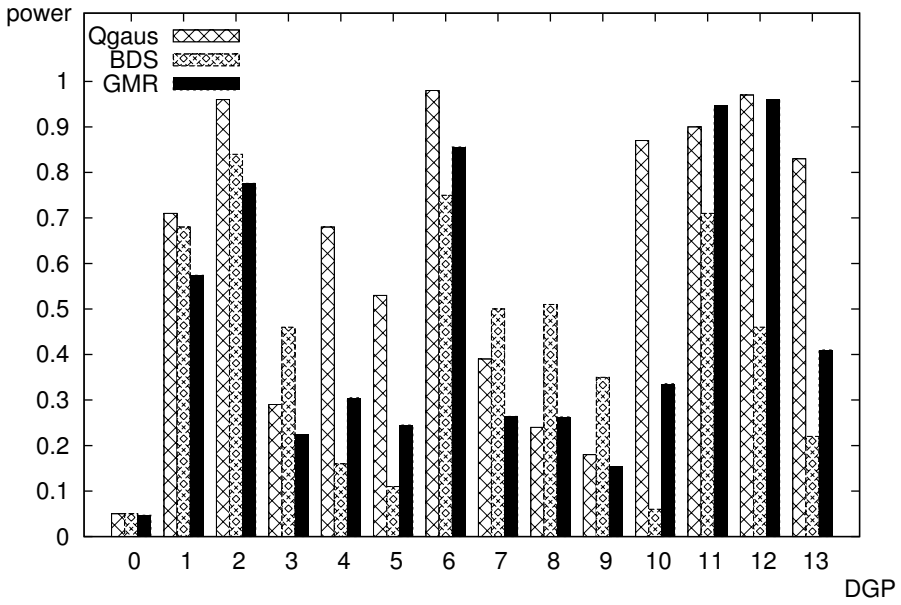

Figure 2.3: *Observed rejection rates (size/power) for various DGPs. Nominal size* $\alpha = 0.05$*, sample size* $n = 100$*, lag* $\ell = 1$*, dimension* $m = 3$*, number of permutations* $B + 1 = 100$ *(*200 *for BDS), number of simulations* $1,000$*.*

Generally, the expectation of the BDS test statistic under the alternative can lie to the left or right relative to that under the null. This was confirmed by simulations for certain alternatives, e.g. the logistic map (DGP 11). Therefore, we implemented the BDS test as a two-sided test. To make it comparable with the Q-test we applied a similar multiple bandwidth permutation procedure and doubled the number of permutations to $B + 1 = 200$ to take into account the two-sidedness. The bandwidth range $R = [0.5, 2.0]$, which is typical for the BDS test, coincides exactly with that used in the Q-test. We set the number of bandwidths to $d = 5$ also for the BDS test.

We used the original routine for the GMR test to compute rejection rates for the considered DGPs. Since their test embeds likelihood cross validation of Silverman (1986, Sec. 3.4.4) to select optimal bandwidths (determining separate optimal bandwidth values under the null and the alternative), no bandwidth selection was required. For dimensions higher than two we used their "portmanteau" version of the test.

Figure 2.3 reports the observed rejection rates (at size $\alpha = 0.05$, $\ell = 1$, $m = 3$) for the considered processes for the introduced Q-test based on the Gaussian kernel, the BDS

test and the GMR test. See Appendix 2.C for the numerical values and extended results (higher lags ℓ and dimensions m) of these tests and the Q-tests based on other kernels. As expected, for all tests the nominal size of 0.05 is within the 95% confidence interval of the actual size estimate. The Q-test yields powers comparable to those obtained using the BDS and GMR procedures and in 8 out of 13 cases outperforms them, i.e. for the nonlinear MA(1) – MA(2), linear, fractional and sign function AR(1) and TAR processes and the Hénon map without and with the observational noise (DGPs $1, 2, 4-6, 10, 12, 13$). In absolute terms the power of the Q-test is smaller for the nonlinear MA(2) and bilinear processes and the logistic map (DGPs $3, 7, 11$), but still comparable to that obtained by the best performing test (for a particular DGP). In comparison with the BDS test, the Q-test shows less power for the ARCH(1) and GARCH$(1, 1)$ processes (DGPs 8 and 9). The GMR test behaves similar to the Q-test in this situation. Comparing the performance of the Q-test based on the Gaussian, double exponential and Cauchy kernels we do not observe large differences (see Appendix 2.C). Therefore, we proceed with the analysis based on the Gaussian kernel only.

The ARCH(1) process (DGP 8) and its generalisation, the GARCH$(1, 1)$ process (DGP 9), are used in financial econometrics to model periods of consecutive large deviations from the mean, interchanged by periods of moderate deviations, mimicking observed behaviour of stock returns. Since the GARCH$(1, 1)$ process is of special interest in financial econometrics we undertake a more detailed analysis of this process. The power of the Q-test increases if we consider higher delay vector dimensions m for this DGP. To obtain an even further increase in power against the GARCH$(1, 1)$ process we can adopt a semi-parametric approach and transform the data to their absolute values before testing. Table 2.1 shows the rejection rates obtained with the test for GARCH$(1, 1)$ using this transformation in contrast to no transformation. After this transformation the Q-test becomes more powerful than the BDS and the GMR test conducted on the transformed and original data. The intuition behind this increase in power lies in the local nature of the kernel. Initially distant delay vectors with differently signed elements can be mapped locally close to each other upon replacing the vector elements by their absolute values, enabling the test to capture more of the dependence. We conclude from this that applying

DGP	Qgaus		BDS		GMR	
m	abs	orig	abs	orig	abs	orig
2	0.29	0.13	0.25	0.26	0.12	0.13
3	0.39	0.18	0.31	0.35	0.15	0.15
5	0.46	0.26	0.40	0.43	0.18	0.18
10	0.53	0.38	0.42	0.48	0.15	0.17

Table 2.1: *Observed power against GARCH*$(1,1)$ *(DGP* 9*) after (abs) and before (orig) transforming the data to absolute values, nominal size* $\alpha = 0.05$*, sample size* $n = 100$*, lag* $\ell = 1$*, number of permutations* $B + 1 = 100$ *(*200 *for BDS) and number of simulations* $1,000$*.*

the Q-test to the absolute values of the data is preferable when structure in volatility is to be detected.

2.4.2 Local alternatives

We next consider power against local alternatives. For a test similar to that of GMR, Hong and White (2005) found nontrivial power as the distance between the null distribution and a local alternative reduces at the rate $n^{-1/2}h^{-1/2}$ with $h \to 0$, which is required for consistent kernel estimation of the density. The test statistic for the Q-test is estimated using U-statistics, which in the non-degenerate case converge at the parametric rate $n^{-1/2}$. Moreover, the consistency of the Q-test does not require the bandwidth to diminish with the sample size. Therefore, we may expect the test to have nontrivial asymptotic power at the rate $n^{-1/2}$ and illustrate this via simulations. For the same reasons a similar rate is expected for the BDS test. Following Hong and White (2005) we consider a sequence of processes with lag j dependence with the following joint probability function:

$$f_{jn}(y_t, y_{t+j}) = f(y_t)f(y_{t+j})[1 - a_n q_j(y_t, y_{t+j}) + r_{jn}(y_t, y_{t+j})], \tag{2.3}$$

where $q_j(y_t, y_{t+j})$ is a function characterising the deviation from the null hypothesis, a_n governs the rate of convergence to the null as $n \to \infty$, and $r_{jn}(y_t, y_{t+j})$ is a higher order term obtained from the Taylor series expansion of $f_{jn}(y_t, y_{t+j})$ around the point $a_n = 0$. See Hong and White (2005) for assumptions on $q_j(\cdot, \cdot)$ and $r_{jn}(\cdot, \cdot)$ which ensure that

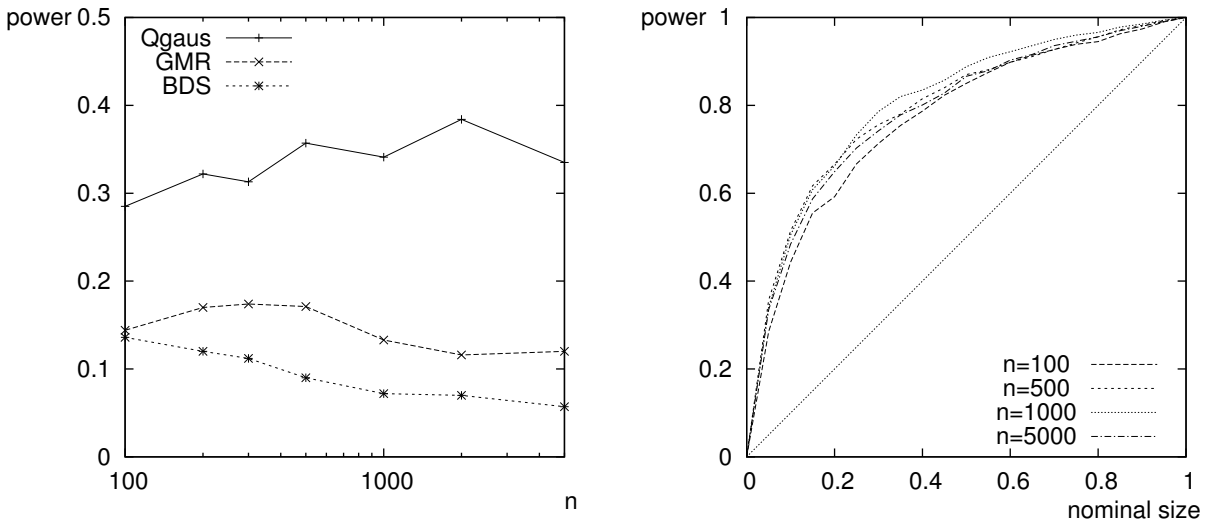

Figure 2.4: *Observed power against local alternatives converging to the null at rate* $n^{-1/2}$*, as a function of sample size* $n = 100, \ldots, 5,000$ *at nominal size* $\alpha = 0.05$ *(left panel); as a function of nominal size for the Q-test (right panel). Lag* $\ell = 1$*, dimension* $m = 2$*, number of permutations* $B + 1 = 100$ *(*$B + 1 = 200$ *for BDS), number of simulations* $1,000$.

$f_{jn}(\cdot, \cdot)$ is a proper density function.

The simulations are based on an MA(1) process $Y_t = \varepsilon_t + a_n \varepsilon_{t-1}$ where $\{\varepsilon_t\}$ is a sequence of independent standard normal random variables. The joint density of (Y_t, Y_{t+1}) can be represented in the form (2.3) with $q_j(y_t, y_{t+j}) = y_t y_{t+j}$. Figure 2.4 (left panel) shows the rejection rates (powers) of the considered test against a sequence of local alternatives which converges to the null at the usual parametric rate $a_n = Cn^{-1/2}$, where C is a constant and $n = 100, \ldots, 5,000$. A horizontal line in the graph would indicate the parametric rate. After an initial transient period for small n, the curves level out, suggesting that all tests asymptotically approach the parametric rate. The Q-test has a substantially larger nontrivial asymptotic power at this rate than the two other tests. The nontrivial asymptotic power for the Q-test against this sequence of local alternatives can also be observed for other values of the nominal size, as illustrated by the power-size plots for increasing sample sizes n shown in the right panel of Figure 2.4.

2.4.3 Application to residuals

So far our theory and simulations were concerned only with the independence hypothesis for raw data. However, in practice the tests of independence are often used as specification tests while applied to the estimated residuals of some parametric model. Generally, estimated residuals are not independent and thus not exchangeable, even if they are based on i.i.d. innovations. The main question which determines the validity of the tests based on residuals is whether the dependence in the residuals introduced by parameter estimation affects the test statistic. A test employing parametrically estimated residuals will in general remain consistent if its rate is slower than the parametric rate, which is the case in the asymptotic test of Hong and White (2005), which is similar to the GMR test. Brock, Dechert, Scheinkman, and LeBaron (1996) show that the presence of the estimated parameters does not affect the asymptotic distribution of their test statistic. Our simulations on estimated residuals show that the GMR and the BDS tests remain correct in terms of size. This is not the case, however, for the Q-test, at least for moderate sample sizes. We filtered DGP 0 and DGP 4 through the AR(1) model using ordinary least squares (OLS) regression. Since the assumed model is correct, the corresponding residuals are asymptotically i.i.d. and satisfy the null hypothesis. Our simulations on the corresponding residuals showed that for $n = 100, 500, 1000$ the actual size of the Q-test was around 0.01 with nominal size $\alpha = 0.05$. This indicates a bias in the estimated p-values, which does not vanish with increasing sample size. In order to use the Q-test as a specification test on the estimated residuals we employ a parametric bootstrap (Efron, 1979). In this procedure we condition on a number of original observations, equal to the order of the model, and the marginal distribution of the original residuals. The BDS and GMR permutation tests were applied directly to the residuals.

Table 2.2 shows rejection rates of the tests applied to residuals of the AR(1) model estimated by OLS from previously considered DGPs. Under the null, that is, for DGP 4, the observed size of all tests is close to the nominal level 0.05. The power of all tests drops compared to the tests of Subsection 2.4.1 based on raw data, which indicates that indeed some of the dependence structure is captured by the AR(1) model. The power of the Q-test on estimated residuals is comparable with that of the other tests, i.e. its power

DGP	1	2	3	4	5	6	7	8	9	10	11	12	13
Q	0.61	0.63	0.12	0.04	0.13	0.08	0.12	0.21	0.11	0.37	0.79	0.66	0.54
BDS	0.68	0.70	0.27	0.06	0.06	0.06	0.21	0.47	0.21	0.13	0.52	0.26	0.13
GMR	0.70	0.72	0.15	0.04	0.11	0.07	0.12	0.24	0.07	0.33	0.68	0.63	0.42

Table 2.2: *Observed rejection rates (size/power) for estimated residuals of the parametric AR* (1) *model (DGP* 4*), nominal size* $\alpha = 0.05$*, sample size* $n = 100$*, lag* $\ell = 1$*, dimension* $m = 2$*, number of permutations* $B + 1 = 100$ *(*200 *for BDS) and number of simulations* $1,000$.

is lower for the nonlinear MA and bilinear processes (DGPs $1 - 3, 7$), but it performs slightly better for the TAR model (DGP 10) and the logistic map (DGP 11).

2.5 Application to financial time series

We consider an application to the Standard and Poor's 500 Stock Index daily log-returns $X_t = \ln(P_t/P_{t-1})$, where P_t is the dividend-adjusted closing price index on day t, in the period 06/2001–05/2005 (source DATASTREAM). The sample was divided into two subsamples: period 1 (06/2001–03/2003) and period 2 (03/2003–05/2005), each having 500 observations.

Figure 2.5 shows the daily time series in levels of the S&P500 Stock Index as well as the partial autocorrelation function (PACF) plots of the log-returns and absolute log-returns series for the two periods. The sample division was made on the basis of visual inspection and basic statistics: period 1 corresponds to a downward trend and exhibits strong volatility while period 2 corresponds to an upward trend with moderate volatility. First, we test for a geometric random walk hypothesis, which is equivalent to the null hypothesis of serial independence of the log-returns, using the Q-test for lags $\ell = 1, \ldots, 10$ and dimensions $m = 2, 5$. The results (Table 2.3, columns "orig") suggest that H_0 is rejected for most of the lags for both periods. The evidence is stronger in the downward period and for the higher dimension ($m = 5$). Next, we apply the test to the absolute values of the log-returns in search for a structure in volatility and detect a stronger structure in volatility in the downward period (Table 2.3, columns "abs"). Comparing the results of the Q-test ($m = 2$) with the PACFs in Figure 2.5 we notice that both tests

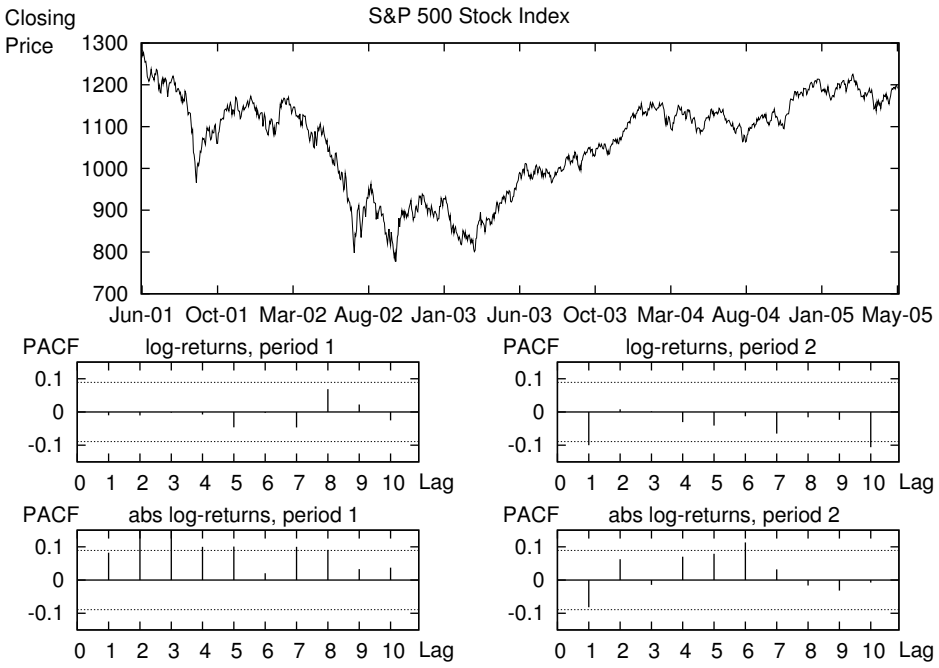

Figure 2.5: *Daily series of the S&P500 Stock Index and PACF plots of log-return and absolute log-return series ($n = 500$) for the two periods.*

reject the null at the same lags. This shows that the fully nonparametric Q-test is able to detect the same structure as a commonly used parametric test. In an attempt to model the detected volatility structure we use an ARCH(1) specification and apply the Q-test on the absolute values of estimated residuals as a model specification test. Table 2.3, columns "ARCH" shows that the ARCH(1) filter is indeed able to capture the volatility structure for most of the lags and embedding dimensions in the two periods.

2.6 Concluding remarks

We introduced a new nonparametric test for serial independence based on quadratic forms. The test does not require the use of plug-in density estimators and remains consistent without letting the bandwidth diminish with sample size. We showed that the dependence measure used has desirable theoretical properties and several connections with other dependence measures. In particular we noticed that the test statistics are closely related to the statistics introduced by Rosenblatt (1975). Our findings imply that the latter statis-

		Period 1			Period 2		
m	ℓ	orig	abs	ARCH	orig	abs	ARCH
2	1	0.09	0.13	0.24	0.01	0.06	0.15
2	2	0.04	0.01	0.18	0.14	0.26	0.39
2	3	0.15	0.01	0.14	0.93	0.45	0.58
2	5	0.13	0.01	0.27	0.64	0.40	0.56
2	10	0.36	0.06	0.41	0.04	0.73	0.16
5	1	0.02	0.01	0.03	0.02	0.08	0.21
5	2	0.01	0.01	0.05	0.08	0.03	0.09
5	3	0.05	0.01	0.25	0.54	0.04	0.19
5	5	0.01	0.01	0.14	0.04	0.60	0.19
5	10	0.81	0.13	0.16	0.02	0.05	0.06

Table 2.3: *P-values based on the series of S&P 500 log-returns, their absolute values, and ARCH(1) filtered series for two periods. Nominal size* $\alpha = 0.05$*, sample size* $n = 500$*, number of permutations* $B + 1 = 100$ *and number of simulations* 1 000*.*

tics for fixed bandwidths have an interpretation as quadratic forms, so that they can be meaningfully used even if the underlying distributions are discontinuous.

We suggested a multiple bandwidth procedure to avoid the problem of optimal bandwidth selection while providing good power for various DGPs. Numerous simulations showed that the Q-test implemented on the basis of the exact permutation procedure has good finite sample performance against local and fixed alternatives in comparison with two other recent nonparametric tests: the BDS and GMR tests. The Q-test showed remarkably better power against TAR models. Further, we addressed the issue of using the Q-test as a parametric model specification test while applying it to residuals series and compared its performance in this situation with the BDS and the GMR tests. Finally, the test was applied to recent S&P 500 log-return series in downward- and upward-trend periods. The hypothesis of serial independence of the log-returns was rejected, with stronger rejection in the downward period. An application to residuals indicated that much of the structure in the volatility could be successfully accounted for by an ARCH(1) model.

Appendix

2.A Relation between Q and correlation coefficient

Our aim here is to find an analytic expression for the introduced distance measure, the quadratic form Q between the time series $\{X_t\}$ of the above structure and a time series $\{Y_t\}$ independently sampled from a multivariate normal distribution. The expression will be derived for the Gaussian product kernel $K_h(x-y) = \prod_{i=1}^{m} \exp\left(-(x_i-y_i)^2/(4h^2)\right)$. We consider a strictly stationary and weakly dependent time series $\{X_t\}$ generated by a Gaussian process such that the m-dimensional delay vectors $X_t^{m,\ell} = (X_t, X_{t+\ell}, \ldots, X_{t+(m-1)\ell})'$ are multivariate normal random variables (standardised to unit variances) with correlation matrix Ω. In the case of independence the correlation matrix reduces to the identity matrix. To simplify the integration we transform the multivariate normal pdf of the form $f(x) = |\Omega|^{-1/2}(2\pi)^{-m/2}\exp(-\frac{1}{2}(x'\Omega^{-1}x))$ to z coordinates defined by $z = Vx$, where V is an orthogonal matrix and $\Omega = VDV'$ by the spectral theorem, where $D = \mathrm{diag}(\eta_1^2, \ldots, \eta_m^2)$: $f^*(z) = (2\pi)^{-m/2}\prod_{i=0}^{m}\eta_i^{-1}\exp(-z_i^2/(2\eta_i^2))$. The absolute value of the determinant of the Jacobian is one (property of an orthogonal matrix). Using the above transformation we can compute the elements of Q, letting $f_0(\cdot)$ denote the product of marginal pdfs:

$$
\begin{aligned}
Q^{11} &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K_h(r-s)\, f^*(r) f^*(s)\, dr ds = h^m \prod_{i=1}^{m} \frac{1}{\sqrt{h^2+\eta_i^2}}, \\
Q^{12} &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K_h(r-s)\, f^*(r) f_0(s)\, dr ds = h^m \prod_{i=1}^{m} \frac{1}{\sqrt{h^2+(\eta_i^2+1)/2}}, \\
Q^{22} &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K_h(r-s)\, f_0(r) f_0(s)\, dr ds = h^m \prod_{i=1}^{m} \frac{1}{\sqrt{h^2+1}}.
\end{aligned}
$$

Combining terms we can express Q as a function of the eigenvalues η_i^2 which are determined by the autocorrelations ρ_i, the bandwidth h, and the delay vector dimension m:

$$
Q = h^m \left(\prod_{i=1}^{m} \frac{1}{\sqrt{h^2+\eta_i^2}} - 2\prod_{i=1}^{m} \frac{1}{\sqrt{h^2+(\eta_i^2+1)/2}} + \prod_{i=1}^{m} \frac{1}{\sqrt{h^2+1}} \right).
$$

In the case of a bivariate standard normal distribution with a correlation coefficient ρ, the eigenvalues are simply expressed as $\eta_1^2 = 1+\rho$, $\eta_2^2 = 1-\rho$ and one obtains a direct correspondence between Q and ρ^2.

2.B Equivalence of Q and quadratic distance

We establish the equivalence of the quadratic distance of Rosenblatt (1975) $T = \int_{\mathbb{R}^m}(\widehat{f}(x) - \widehat{g}_2(x))^2 dx$ and the U-statistics estimator of quadratic form $\widehat{Q}$. For simplicity we consider a Gaussian kernel for density estimation. Rewrite T explicitly in terms of the kernel density estimators:

$$T = \int_{\mathbb{R}^m} \left(\frac{1}{n_1} \sum_{t=1}^{n_1} \left(\frac{1}{\sqrt{2\pi}h} \right)^m e^{\frac{-\|X_t - x\|^2}{2h^2}} - \frac{1}{n_2} \sum_{s=1}^{n_2} \left(\frac{1}{\sqrt{2\pi}h} \right)^m e^{\frac{-\|Y_s - x\|^2}{2h^2}} \right)^2 dx.$$

Expanding the square, one arrives at the form $T = T^{11} - 2T^{12} + T^{22}$. For brevity we will derive the T^{11} term only, derivations for T^{12} and T^{22} being similar:

$$\begin{aligned} T^{11} &= \frac{1}{n_1^2} \left(\frac{1}{\sqrt{2\pi}h} \right)^{2m} \int_{\mathbb{R}^m} \sum_{t=1}^{n_1} \sum_{s=1}^{n_1} e^{\frac{-\|X_t - x\|^2 - \|X_s - x\|^2}{2h^2}} dx \\ &= \frac{1}{n_1^2} \left(\frac{1}{\sqrt{2\pi}h} \right)^{2m} \int_{\mathbb{R}^m} \sum_{t=1}^{n_1} \sum_{s=1}^{n_1} e^{\frac{-\|X_t - X_s\|^2}{4h^2}} e^{\frac{-2\|x - (X_t + X_s)/2\|^2}{2h^2}} dx \\ &= \frac{1}{n_1^2} \left(\frac{1}{2\sqrt{\pi}h} \right)^{m} \sum_{t=1}^{n_1} \sum_{s=1}^{n_1} e^{\frac{-\|X_t - X_s\|^2}{4h^2}}. \end{aligned}$$

Above we used the Gaussian kernel factorisation that allows to reduce the analysis of the m-dimensional norm $\|\cdot\|^2$ to one dimension $(\cdot)^2$. In this form T^{11} is exactly the same as the V-statistic estimator of Q^{11} times a factor $\left(\frac{1}{2\sqrt{\pi}h}\right)^m$ which does not depend on the data. Analogously, one can establish equivalence of T^{12}, T^{22} and the V-statistic estimators of Q^{12} and Q^{22} respectively. Given the asymptotic equivalence of V-statistics and U-statistics we establish that $T \approx \left(\frac{1}{2\sqrt{\pi}h}\right)^m \widehat{Q}$.

2.C Performance for various kernels

The tables below report the rejection rates for five nonparametric tests for serial independence. The Q-columns correspond to the tests based on quadratic forms with Gaussian, double exponential and Cauchy kernels respectively, and the remaining two columns to the BDS test and the GMR test, respectively. The nominal size of the tests was set to 0.05. We consider three lags $\ell = 1, 2, 3$ for delay vector dimension $m = 2$, and only one lag $\ell = 1$ for higher dimensions ($m = 3, 4, 5, 10$). The bandwidth set H included $d = 5$ different values in the range $R = [0.5, 2.0]$ (after normalisation of the series to the unit variance). We used series of length $n = 100$ ($n = 50$ for DGP 6 and $n = 20$ for DGP $11 - 13$), the total number of permutations was set to $B + 1 = 100$ (for BDS test $B + 1 = 200$). The number of simulations was set to $1,000$ ($5,000$ for DGP 0).

0. $Y_t = \varepsilon_t$ — 1. $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$

m	ℓ	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.054	0.059	0.051	0.057	0.052	0.71	0.81	0.76	0.76	0.78
2	2	0.051	0.048	0.050	0.049	0.054	0.06	0.06	0.07	0.06	0.04
3	1	0.050	0.047	0.050	0.052	0.047	0.71	0.78	0.77	0.68	0.57
5	1	0.050	0.052	0.049	0.052	0.056	0.57	0.64	0.62	0.46	0.39
10	1	0.051	0.049	0.052	0.055	0.049	0.32	0.37	0.35	0.22	0.20

2. $Y_t = \varepsilon_t + 0.6\varepsilon_{t-1}^2 + 0.6\varepsilon_{t-2}^2$ — 3. $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}\varepsilon_{t-2}$

m	ℓ	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.94	0.94	0.93	0.83	0.83	0.14	0.13	0.12	0.34	0.21
2	2	0.26	0.30	0.28	0.25	0.25	0.12	0.11	0.11	0.20	0.11
3	1	0.96	0.96	0.95	0.84	0.78	0.29	0.23	0.25	0.46	0.22
5	1	0.93	0.93	0.91	0.76	0.60	0.35	0.26	0.31	0.39	0.14
10	1	0.76	0.80	0.75	0.46	0.39	0.26	0.20	0.26	0.24	0.07

4. $Y_t = 0.3Y_{t-1} + \varepsilon_t$ — 5. $Y_t = 0.8|Y_{t-1}|^{0.5} + \varepsilon_t$

m	ℓ	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.70	0.67	0.59	0.21	0.40	0.55	0.60	0.53	0.15	0.34
2	2	0.12	0.11	0.09	0.05	0.07	0.07	0.07	0.07	0.05	0.06
3	1	0.68	0.64	0.56	0.16	0.31	0.53	0.57	0.50	0.11	0.25
5	1	0.61	0.58	0.46	0.13	0.20	0.43	0.48	0.40	0.09	0.18
10	1	0.45	0.45	0.32	0.07	0.15	0.29	0.35	0.25	0.06	0.12

6. $Y_t = \text{sign}(Y_{t-1}) + \varepsilon_t$ 7. $Y_t = 0.6\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$

m	ℓ	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.98	0.98	0.98	0.77	0.90	0.18	0.17	0.16	0.32	0.17
2	2	0.70	0.72	0.68	0.33	0.37	0.20	0.18	0.18	0.41	0.22
3	1	0.98	0.98	0.97	0.75	0.86	0.39	0.31	0.33	0.50	0.26
5	1	0.96	0.97	0.95	0.64	0.74	0.48	0.38	0.43	0.51	0.21
10	1	0.90	0.92	0.87	0.40	0.54	0.38	0.33	0.35	0.34	0.14

8. $Y_t = \varepsilon_t\sqrt{1 + 0.4Y_{t-1}^2}$ 9. $Y_t = \sqrt{h_t}\varepsilon_t$, $h_t = 0.01 + 0.80h_{t-1} + 0.15Y_{t-1}^2$

m	ℓ	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.25	0.22	0.23	0.55	0.32	0.13	0.12	0.12	0.26	0.13
2	2	0.07	0.07	0.06	0.12	0.09	0.12	0.12	0.11	0.23	0.11
3	1	0.24	0.22	0.24	0.51	0.26	0.18	0.15	0.16	0.35	0.15
5	1	0.24	0.20	0.21	0.39	0.18	0.26	0.23	0.24	0.43	0.18
10	1	0.20	0.16	0.18	0.24	0.11	0.38	0.32	0.36	0.48	0.17

10. $Y_t = \begin{cases} -0.5Y_{t-1} + \varepsilon_t, Y_{t-1} < 1 \\ 0.4Y_{t-1} + \varepsilon_t, \text{else} \end{cases}$ 11. $Y_t = 4Y_{t-1}(1 - Y_{t-1}),\ 0 < Y_t < 1$

m	ℓ	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.91	0.94	0.90	0.07	0.49	0.98	0.97	1.00	0.85	0.96
2	2	0.10	0.09	0.10	0.06	0.05	0.08	0.12	0.23	0.16	0.74
3	1	0.87	0.91	0.87	0.06	0.34	0.90	0.92	1.00	0.71	0.95
5	1	0.77	0.83	0.76	0.04	0.23	0.54	0.59	0.79	0.35	0.88
10	1	0.49	0.61	0.47	0.03	0.14	0.17	0.24	0.31	0.11	0.59

12. $Y_t = 1 + 0.3Y_{t-2} - 1.4Y_{t-1}^2$ 13. $Y_t = Z_t + \sigma\varepsilon_t$, $Z_t = 1 + 0.3Z_{t-2} - 1.4Z_{t-1}^2$

m	ℓ	Q_g	Q_d	Q_c	BDS	GMR	Q_g	Q_d	Q_c	BDS	GMR
2	1	0.99	0.99	1.00	0.46	0.99	0.93	0.91	0.95	0.30	0.59
2	2	0.16	0.22	0.50	0.13	0.55	0.15	0.15	0.26	0.06	0.13
3	1	0.97	0.97	1.00	0.46	0.96	0.83	0.84	0.91	0.22	0.41
5	1	0.86	0.89	0.96	0.35	0.87	0.74	0.73	0.81	0.10	0.25
10	1	0.40	0.39	0.51	0.12	0.46	0.37	0.34	0.45	0.39	0.12

Chapter 3

Goodness-of-fit test for copulas

3.1 Introduction

The copula introduced in Chapter 1 proved to be a handy instrument in the analysis of multivariate time series. It allows to capture the full dependence within multivariate time series without specifying the shape of the marginal distributions. This result is due to Sklar's theorem: any multivariate distribution can be decomposed into a copula and its marginals; if the marginal distributions are continuous the copula is unique. Moreover, the copula is invariant under strictly increasing transformations.

Due to these favourable properties copulas proved to be useful in financial applications, e.g. risk management, portfolio aggregation, spillover effects. For a review we refer the interested reader to Bouyé, Durrleman, Nikeghbali, Riboulet, and Roncalli (2000).

The major drawback in the copula approach is that in order to proceed with a traditional parametric analysis a specific functional form has to be assumed for the copula, but there is no indication of a suitable parametric form for the copula. Though many functional forms have been suggested (see e.g. Nelsen, 1999), there are no general guidelines for optimal parametric copula selection.

Up to now there have been a few studies trying to tackle the problem. Durrleman, Nikeghbali, and Roncalli (2000) constructed the Deheuvels or empirical copula and compared it with various parametric copulas based on bivariate time series. The discrete L^2 norm between the empirical copula and candidate parametric copulas was chosen as a

criterion of fit. Malevergne and Sornette (2003) investigated whether bivariate dependence can be described by the Gaussian copula. Their tests are based on Kolmogorov and Anderson-Darling distances and their modifications. Chen, Fan, and Patton (2004) considered multivariate analysis to test the hypothesis of the Gaussian and Student t copula. However, according to these authors their first test suffers from the curse of dimensionality. The second test does not have this problem, but may be inconsistent.

This study develops an alternative goodness-of-fit test for bivariate and multivariate copulas. The test is based on a divergence measure first introduced by Diks, Van Zwet, Takens, and DeGoede (1996). This measure, a kernel-based positive definite bilinear form, can be consistently estimated using V-statistics. It does not require the use of plug-in estimators (a common practice in the field, see e.g. Granger, Maasoumi, and Racine, 2004) and separates the problem of inference from consistent estimation of multivariate densities. The proposed test is nonparametric and may be applied to any functional form of the copula.

As an example of an empirical application, the fit of the Gaussian copula is evaluated on US large cap stock returns data. Both bivariate and multivariate portfolios of assets are considered in the analysis.

3.2 Estimation of copula parameters

Consider the case where a copula and marginals distributions are continuous. According to Sklar's theorem, the joint CDF $F(x),\ x \in \mathbb{R}^m$ can be represented as

$$F(x) = C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)), \tag{3.1}$$

where $C(u),\ u \in [0,1]^m$ is a copula and $F_n(x_n),\ n = 1, \ldots, N$ are marginal CDFs.

The corresponding density function is

$$f(x) = c(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)) \prod_{n=1}^{m} f_n(x_n), \tag{3.2}$$

where $f_n(x_n)$ is the density of the marginal $F_n(x_n)$ and $c(u)$ is the density of the copula

$C(u)$

$$c(u_1, \ldots, u_m) = \frac{\partial^m C(u_1, \ldots, u_m)}{\partial u_1 \cdots \partial u_m}. \tag{3.3}$$

The canonical maximum likelihood (CML) method Bouyé, Durrleman, Nikeghbali, Riboulet, and Roncalli (2000) is used to estimate the vector of parameters α of the copula. First the data $\{X_{t1}, X_{t2}, \ldots, X_{tm}\}_{t=1}^{T}$ are transformed into the corresponding empirical CDFs $\widehat{F}_i(x_i)$ through

$$\widehat{F}_i(x) = \frac{1}{T} \sum_{t=1}^{T} I(X_{tn} \leq x). \tag{3.4}$$

The vector of parameters α is estimated semi-parametrically maximising log-likelihood for the copula density c, given the empirical marginal CDFs $\hat{F}_i(x)$

$$\hat{\alpha} = \arg\max_{\alpha} \sum_{t=1}^{T} \ln c(\widehat{F}_1(X_{t1}), \ldots, \widehat{F}_m(X_{tm}); \alpha). \tag{3.5}$$

3.3 A notion of distance between probability distributions

Following Diks, Van Zwet, Takens, and DeGoede (1996) for integrable functions f and g define the bilinear form

$$(f, g) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_m(x, y) f(x) g(y)\, dxdy, \tag{3.6}$$

where $x, y \in \mathbb{R}^m$ and $K_m(\cdot, \cdot)$ is a positive definite symmetric kernel such as

$$K_m(x, y) = e^{-\|x-y\|^2/(2mh^2)}, \tag{3.7}$$

where $\|\cdot\|$ denotes Euclidean norm in $\mathbb{R}^m$ and $h > 0$ is a smoothing parameter, or bandwidth. This kernel allows factorisation

$$K_m(x, y) = \prod_{i=1}^{m} \kappa\left(\frac{x_i - y_i}{2mh^2}\right), \tag{3.8}$$

where $\kappa(\cdot)$ is the Gaussian kernel. The Gaussian kernel is chosen for convenience. In general, other positive definite kernel functions can be used (c.f. chapter 2).

According to Proposition 4 of Diks and Tong (1999) for integrable functions f, the quadratic form $(f, f) \geq 0$ and $(f, f) = 0$ if and only if $f = 0$ almost everywhere. In fact, (f, g) is an inner product of f and g, which can be used as a distance measure distance between f and g. Note that the defined bilinear form (3.6) is an expectation of the kernel K_m taken with respect to the independent random vectors X with the pdf $f(x)$ and Y with the pdf $g(y)$, that is

$$E[K_m(X, Y)] = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K_m(x, y) f(x) g(y)\, dxdy. \tag{3.9}$$

Define a squared distance Q between f and g as

$$Q = (f - g, f - g). \tag{3.10}$$

It is obvious that Q becomes zero only when f and g are equal. Following the properties of the inner product, Q can be decomposed as follows

$$Q = Q^{11} - 2Q^{12} + Q^{22}, \tag{3.11}$$

where $Q^{ij} = \langle f_i | K_m | f_j \rangle$. Each term of the above decomposition can be consistently estimated using V-statistics, so that

$$\widehat{Q}^{ij} = \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} K_m(X^i_{t_1}, X^j_{t_2}), \tag{3.12}$$

where X^i_t denotes a realisation of the random vector X^i at a time t.

Denker and Keller (1983) use the fact that the consistency of a V-statistic estimator is

established for stationary absolutely regular processes and develop asymptotic distribution for V-statistic under these conditions. The same can be done for the functional of interest Q.

3.4 Testing procedure

Our goal is to test whether a specific functional copula can adequately describe the dependence between components of a given multivariate time series. Serial independence of the multivariate series is assumed throughout this section. First, the data series are transformed to the empirical marginal cumulative distributions according to Eq. (3.4). Jointly, this transformed data can be viewed as a sample from the true copula. In the second stage, the parameters of an assumed parametric copula are estimated as described in Section 3.2. Next, we sample from the candidate parametric copula with the estimated parameters. Under the null hypothesis of the true copula being the assumed parametric copula, both series, the transformed empirical series $\{X_t\}$ (parallel to the notation of Section 3.3) and the sampled series $\{Y_t\}$, originate from the same copula. The squared distance Q is used to compare the series. Its estimate denoted by $\widehat{Q}$ is computed element-wise according to Eqs. (3.11) and (3.12).

Since the asymptotic theory for the statistics Q is still under development, the parametric bootstrap (Efron and Tibshirani, 1993) is proposed to determine p-values of the test. We repeatedly (B times) resample series $\{X_t^j\}$, $1 \leq j \leq B$ from the assumed parametric copula with estimated parameters using different seeds of the random number generator and compare them with the initially sampled series $\{Y_t\}$. The corresponding distances are denoted by Q_j. P-values are generated following the standard procedure of comparing $\widehat{Q}$, an estimated distance between the transformed empirical data $\{X_t\}$ and the initially sampled series $\{Y_t\}$, with values of $\widehat{Q}_j$, an estimated distance between $\{X_t^j\}$ and $\{Y_t\}$, i.e.

$$\widehat{p} = \frac{\sum_{j=1}^{B} I(\widehat{Q} \leq \widehat{Q}_j) + 1}{B + 1}, \tag{3.13}$$

where I is indicator function. A test of nominal size α is obtained if the null hypothesis is rejected whenever $\widehat{p} \leq \alpha$. The number of replications for a test of a nominal size α

is found from the relation $B + 1 = \frac{k}{\alpha}$, where k is a positive integer constant. For a 5%-test, the minimal value of B is 19 ($k = 1$). Our simulations showed that increasing the constant k (i.e. $k = 2, 5$) practically does not change the outcome of this test. Besides, by holding the number of replications B relatively small, we save a lot computational time. Note that $\widehat{Q}^{22}$, an estimated element of the decomposition (3.11) depends only on $\{Y_t\}$. Therefore, it takes the same values in $\widehat{Q}$ and $\widehat{Q}_j$ and may be abandoned without any effect on p-values.

The optimal value of the bandwidth parameter d is determined via simulation. The power of the test is used as a criterion for selection. The highest power is achieved for the value of the bandwidth $h = 0.05$ (time series are standardised to unit variance).

3.5 Performance of the test and comparison with other tests

Recently Berg and Bakken (2006) performed a comparative analysis of the different goodness-of-fit tests for copulas. Here we shortly summarise their findings. Four tests were considered in their study, i.e. the considered test based on squared distance Q (Q-test), and the tests proposed by Chen, Fan, and Patton (2004), Berg and Bakken (2005) and Genest, Quessy, and Remillard (2006). The latter three tests are based on probability integral transforms while the Q-test does not require this transform.

The size and power of the tests were assessed using mixed copula C_{mix}:

$$C_{\text{mix}} = (1 - \gamma)C_{\text{Gaus}} + \gamma C_{\text{Alt}}, \qquad \gamma \in [0, 1], \tag{3.14}$$

where C_{Gaus} denotes Gaussian copula described in details in the next section, C_{Alt} denotes the alternative copula. Three alternative copulas were considered in the study, i.e.

- C_t – Student t copula ($\nu = 4$)
- C_{Cl} – one parameter Clayton copula ($\delta = 1.0$)
- $C_{\text{sur}-\text{Cl}}$ – one parameter survival Clayton copula ($\delta = 1.0$)

For coefficient $\gamma = 0$ the null hypothesis holds. By increasing β one diverges from the null. Three different dimensions were considered in the test, i.e. $m = 2$, $m = 5$ and $m = 10$. The number of simulations was set to 500.

The study concluded that the Q-test performed relatively well in detecting the Clayton copula mixture and the survival Clayton copula mixture. The power of the test was higher in the case when $m = 2$. The performance of the Q-test in the case of the Student t copula mixture was inferior to some, but not all tests considered in the study. For more details as well as graphical analysis we refer the interested reader to the original source, Berg and Bakken (2006).

3.6 Testing the Gaussian copula hypothesis

In this section we apply the previously described procedure to test the Gaussian copula hypothesis. The m-variate Gaussian copula with the correlation matrix R is defined as

$$C^m_{\text{Gaus},R}(u) = \Phi^m_R(\Phi^{-1}(u_1), ..., (\Phi^{-1}(u_m)), \tag{3.15}$$

where Φ^m_R denotes the joint distribution function of $N-$variate standard normal distribution with the linear correlation matrix R, and Φ denotes the cumulative distribution function of the univariate standard normal distribution. Due to the relatively simple estimation procedure and the possibility to incorporate multivariate series, the Gaussian copula is now widely applied in finance. However, this copula has several symmetries and zero tail dependence (Nelsen, 1999). Thus, it may be inadequate to characterise the dependence between financial series, which often exhibit asymmetries and high dependence in the case of extreme events.

The proposed goodness-of-fit test is used to determine whether the Gaussian copula can adequately describe the dependence between daily series of (log)-returns of US large cap stocks. The dataset consists of US traded securities with a volume over 10 million per day and a market capitalisation over one billion US dollars as of the 1/04/2003. The dataset comprises 33 stocks from different sectors. The selection of the large cap stocks ensures considerable interest from the side of investors. The length of the time series

covering the period 01/1997–03/2003 is $1,607$.

Collections of the assets of dimension 2, 5 and 10 were randomly selected 100 times, while the full set of 33 stocks was also included in the analysis. In most of the applications individual financial series were filtered with a (G)ARCH model to remove persistence in the variance (see e.g. Chen, Fan, and Patton, 2004). This is often done in practice, because most of the procedures require serial independence of the individual series. This is also the case for our test. One should be careful with such a filtering, since it may destroy some of the dependence structure between assets (Malevergne and Sornette, 2003). Therefore, we consider both the raw data and the GARCH(1,1) filtered series and compare the corresponding outcomes of the test. The presence of serial dependence in the raw data may weaken the power of the test.

The CML parameter estimator (3.5) in case of the Gaussian copula reduces to

$$\hat{R} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t^{'}, \tag{3.16}$$

where $\mathbf{y}_t = (\Phi^{-1}(\hat{F}_1(x_1^t)), ..., \Phi^{-1}(\hat{F}_m(x_m^t)))^{'}$ and $\hat{F}_i(x_i^t)$ is computed according to Eq. (3.4). In practice, to sample from the Gaussian copula with correlation matrix R, we first obtain a sample from a multivariate normal distribution, standardised to unit marginal variances, with correlation matrix R, $z \sim N(0, R)$. Next, we use the CDF transform, and obtain $(\Phi(z_1), ..., \Phi(z_m))$, which can be view as a sample from the Gaussian copula with correlation matrix R. Next, the testing procedure described in Section 3.4 is applied. Table 3.1 presents the rejection rates with the nominal size set to 0.05. The number of replications B was set to 19 and the bandwidth is $h = 0.05$.

Table 3.1: Rejection rates for the null hypothesis of the Gaussian copula for US large cap stock daily returns

procedure\dimension	2	5	10	33
raw returns	14/100	32/100	52/100	1/1
GARCH(1,1) filtered returns	10/100	19/100	45/100	1/1

The analysis reveals that the rejection rates are notably larger than the nominal

size. The number of rejections increases with the dimension of the asset collections. GARCH(1,1) filtered data indicate less deviations from the Gaussian copula possibly because of the change in the dependence structure. Consequently, the analysis suggests that the Gaussian copula is inadequate in characterising the dependence between US large cap stocks, especially for multivariate collections of assets. Similar rejection rates were found by Malevergne and Sornette (2003) for bivariate collections of stocks from the dataset similar to the one applied in this paper. Chen, Fan, and Patton (2004) report lower than nominal rejection rates for bivariate collections and rejection probabilities close to one for 5- and higher dimensional collections. The difference may be attributed to the difference in data selection and testing procedures.

Panchenko (2005b) considered the same testing procedure on the (log)-returns of AEX constituencies; 24 stocks in total. The dataset was divided into two periods: (1) a volatile downside, 10/2001–03/2003 and (2) a steady upside, 03/2003–07/2005. Each period comprised 630 observations. Portfolios of the stocks of the dimension 2, 5 and 10 were randomly selected 100 times, the full set of 24 stocks was also included in the analysis.

Table 3.2 presents the rejection rates (nominal size 0.05) for two-dimensional portfolios. For dimensions $5, 10$ and the full set of 24, the test always rejects.

Table 3.2: Rejection rates of the null hypothesis of the Gaussian copula for two-dimensional portfolios of the AEX constituencies daily returns.

procedure\period	Oct 00 - Mar 03	Mar 03 - Jul 05
raw returns	61/100	33/100
GARCH(1,1) filtered returns	54/100	18/100

The rejection rates are much larger then the nominal size. The Gaussian copula is rejected more often during period (1), the volatile downside. GARCH(1,1) filtered data indicate less deviations from the Gaussian copula, possibly because of the change in the dependence structure. The rejection rates are much higher than in the case of US large cap stock portfolios. As a possible explanation for this phenomenon, we suggest overreaction of agents trading on the AEX to market news, which can be observed by comparing the time series of the AEX index and the major US indices.

3.7 Conclusions

In line with the current increasing application of nonlinear time series analysis a goodness-of-fit test for copulas is suggested. The test procedure remains consistent and applicable even in the case of higher dimensions. The asymptotic theory for the test is still under development and a bootstrap procedure is used instead. The application of the test to the US large cap stock returns and AEX constituencies returns showed inadequacy of the Gaussian copula. As an alternative a more flexible Student t copula may be used for modelling the dependence of multivariate collections of stocks. To reflect dynamical changes in the dependence structure more flexible copula forms are to be developed.

Chapter 4

A cautionary note on the Hiemstra-Jones test for Granger non-causality

4.1 Introduction

Consider a strictly stationary and weakly dependent bivariate time series process $\{(X_t, Y_t)\}$, $t \in Z$. By definition, X is strictly Granger (1969) causing Y if the conditional distribution of Y_t, given the past observations $Y_{t-1}, Y_{t-2}, \ldots$ and $X_{t-1}, X_{t-2}, \ldots$, differs from the conditional distribution of Y_t, given the past observations $Y_{t-1}, Y_{t-2}, \ldots$ only. Intuitively, X is a Granger cause of Y if adding past observations of X to the information set increases the knowledge on the distribution of current values of Y. Note that although the definition concerns conditional distributions given an infinite number of past observations, in practice tests are usually confined to finite orders in X and Y.

Several recent empirical studies report results for the Hiemstra and Jones (1994) test; a non-parametric test for Granger non-causality against general (linear and non-linear) alternatives. Evidence for causality is reported by e.g. Abhyankar (1998), Silvapulla and Moosa (1999), and Asimakopoulos, Ayling, and Mahmood (2000). Okunev, Wilson, and Zurbruegg (2002) report inefficiencies in Australian real estate and stock market prices based on the presence of Granger causality. In this chapter we argue that the Hiemstra-

Null Hypothesis:	Linear test		H-J test	
	F-stat	P-value	Z-stat	P-value
X does not Granger Cause Y	176.34	0.0000	-0.05	0.5204
Y does not Granger Cause X	2.64	0.1043	-6.79	1.0000

Table 4.1: Example test results for a simulated bivariate time series of length $10,000$.

Jones test does not provide a solid basis for conclusions of this type.

Initially, our interest was raised by the fact that, for some data sets, counter-intuitive results are obtained from the Hiemstra-Jones test and conventional tests of the same null hypothesis against linear Granger causality. Even if there is strong evidence for linear Granger causality, the Hiemstra-Jones test can fail to detect causality, or suggest that there is more causality than under the null hypothesis of no Granger causality (c.f. large negative values of test statistics reported in Brooks and Henry, 2000).

As an illustration we consider the test results presented in Table 4.1. We simulated $10,000$ observations from a bivariate data generating process (to be described in more detail later) with strong linear Granger causality from X to Y and applied linear and nonlinear Granger causality tests in both directions for one lag. The linear test for Granger causality indicates strong evidence in support of the hypothesis that $\{X_t\}$ Granger causes $\{Y_t\}$. However, the Hiemstra-Jones test fails to detect the evident linear causal relation. Although this result may be counter-intuitive, it is not alarming in the sense that it merely indicates a lack of power of the Hiemstra-Jones test. However, the analysis presented in the following sections shows that this lack of power is in fact caused by an inconsistency which can lead to the much more severe problem of over-rejection, i.e. rejecting the null, when it is true, at a higher than nominal rate.

4.2 The Hiemstra-Jones test

Following Hiemstra and Jones (1994), we let $F(Y_t|I_{t-1})$ denote the conditional probability distribution of Y_t given the information set I_{t-1}, which consists of an ℓ_Y-length lagged vector of Y_t, say $Y_{t-\ell_Y}^{\ell_Y} \equiv (Y_{t-\ell_Y}, Y_{t-\ell_Y+1}, \ldots, Y_{t-1})$, and an ℓ_X-length lagged vector of X_t, say $X_{t-\ell_X}^{\ell_X} \equiv (X_{t-\ell_X}, X_{t-\ell_X+1}, \ldots, X_{t-1})$. Hiemstra and Jones consider testing, for a given

pair of lags ℓ_Y and ℓ_X, the following relationship:

$$H_0 : \qquad F(Y_t|I_{t-1}) = F(Y_t|I_{t-1} - X_{t-\ell_X}^{\ell_X}). \tag{4.1}$$

That is, the null hypothesis of interest states that taking the vector of past X-values out of the information set does not affect the distribution of current Y-values, or in other words, that X is not Granger-causing Y.

Adopting the notation used by Hiemstra and Jones, we denote the m-length lead vector of Y_t by Y_t^m, so that we can summarise the vectors defined so far, for $t \in Z$, as:

$$\begin{aligned} Y_t^m &= (Y_t, Y_{t+1}, \ldots, Y_{t+m-1}), \qquad m = 1, 2, \ldots \\ Y_{t-\ell_Y}^{\ell_Y} &= (Y_{t-\ell_Y}, Y_{t-\ell_Y+1}, \ldots, Y_{t-1}), \qquad \ell_Y = 1, 2, \ldots \\ X_{t-\ell_X}^{\ell_X} &= (X_{t-\ell_X}, X_{t-\ell_X+1}, \ldots, X_{t-1}), \qquad \ell_X = 1, 2, \ldots. \end{aligned} \tag{4.2}$$

A crucial claim made by Hiemstra and Jones (1994) without proof, states that the null hypothesis given in Eq. (4.1) implies, for all $h > 0$:

$$\begin{aligned} P\left(\|Y_t^m - Y_s^m\| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h, \|X_{t-\ell_X}^{\ell_X} - X_{s-\ell_X}^{\ell_X}\| < h\right) \\ = P\left(\|Y_t^m - Y_s^m\| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right), \end{aligned} \tag{4.3}$$

where $P(A|B)$ denotes the conditional probability of A given B, and $\|\cdot\|$ the supremum norm, which for a d-dimensional vector $x = (x_1, \ldots, x_d)^T$ is given by $\|x\| = \sup_{i=1}^d |x_i|$. Eq. (4.3) states that the conditional probability that two arbitrary m-length lead vectors of $\{Y_t\}$ are within distance h, given that the corresponding lagged ℓ_Y-length lag vectors of $\{Y_t\}$ are h-close, is the same as when in addition one also conditions on the ℓ_X-length lag vectors of $\{X_t\}$ being h-close.

For an observed bivariate time series $\{(Y_t, X_t)\}$, $t = 1, \ldots, T$, the Hiemstra-Jones test consists of choosing a value for h (typical values are between 0.5 and 1.5 after normalising the time series to unit variance), and testing (4.3) by estimating the conditional

probabilities as ratios of unconditional probabilities. Using the notation

$$\begin{aligned} C1(m+\ell_Y,\ell_X,h) &= P(\|Y_{t-\ell_Y}^{m+\ell_Y} - Y_{s-\ell_Y}^{m+\ell_Y}\| < h, \|X_{t-\ell_X}^{\ell_X} - X_{s-\ell_X}^{\ell_X}\| < h), \\ C2(\ell_Y,\ell_X,h) &= P(\|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h, \|X_{t-\ell_X}^{\ell_X} - X_{s-\ell_X}^{\ell_X}\| < h), \\ C3(m+\ell_Y,h) &= P(\|Y_{t-\ell_Y}^{m+\ell_Y} - Y_{s-\ell_Y}^{m+\ell_Y}\| < h), \\ C4(\ell_Y,h) &= P(\|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h), \end{aligned} \tag{4.4}$$

Eq. (4.3) can be expressed as

$$\frac{C1(m+\ell_Y,\ell_X,h)}{C2(\ell_Y,\ell_X,h)} = \frac{C3(m+\ell_Y,h)}{C4(\ell_Y,h)}. \tag{4.5}$$

Upon denoting the U-statistics estimators of the probabilities in (4.4) with an additional index n, where $n = T+1-m-\max(\ell_Y,\ell_X)$, Hiemstra and Jones show that, under (4.3):

$$\sqrt{n}\left(\frac{C1(m+\ell_Y,\ell_X,h,n)}{C_2(\ell_Y,\ell_X,h,n)} - \frac{C3(m+\ell_Y,h,n)}{C4(\ell_Y,h,n)}\right) \overset{\text{a}}{\sim} N(0,\sigma^2(m,\ell_Y,\ell_X,h)), \tag{4.6}$$

with $\sigma^2(m,\ell_Y,\ell_X,h)$ as given in their appendix. One-sided critical values are used, based on this asymptotic result, rejecting when the observed value of the test statistic in (4.6) is too large.

4.3 A counter-example

In this section we show that Eq. (4.3) is not, in general, implied by the null hypothesis given in (4.1). The left hand side of Eq. (4.3) under the null can be either smaller or larger than the right hand side. To emphasise the relevance to econometrics and financial time series analysis, we focus on a class of bivariate time series processes with conditional heteroskedasticity. The starting point is a first order process $\{Y_t\}$ with conditional heteroskedasticity

$$Y_t \sim N(0, g(Y_{t-1})), \tag{4.7}$$

where $g(\cdot)$ is some positive function, such that $\{Y_t\}$ is stationary and ergodic (e.g. if $g(Y_{t-1}) = a + bY_{t-1}^2$ with $a > 0$ and $0 < b < 1$, Y_t is a stable ARCH(1) process). Next

consider a process $\{X_t\}$, which is instantaneously driven by $\{Y_t\}$ via:

$$X_{t-1} \sim N(0, f(Y_{t-1})), \tag{4.8}$$

where $f(\cdot)$ is again a non-negative function, and where X_{t-1} is taken to be conditionally independent of Y_t given Y_{t-1}. In the particular case where $f(s) = g(s)$ for all s, Y_t and X_{t-1} are identically distributed, conditionally on Y_{t-1}. Clearly, X is not Granger-causing Y. A single lagged value Y_{t-1} fully determines the distribution of Y_t, so that Eq. (4.1) holds for all $\ell_Y, \ell_X = 1, 2, \ldots$.

Next we show that $g(\cdot)$ and $f(\cdot)$ can be chosen in such a way that Eq. (4.3) does not hold for any positive value of h. A convenient first step is to rephrase Eq. (4.3) as

$$\begin{aligned} P\left(\|Y_t^m - Y_s^m\| < h, \|X_{t-\ell_X}^{\ell_X} - X_{s-\ell_X}^{\ell_X}\| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right) \\ = P\left(\|Y_t^m - Y_s^m\| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right) \\ \times P\left(\|X_{t-\ell_X}^{\ell_X} - X_{s-\ell_X}^{\ell_X}\| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right), \end{aligned} \tag{4.9}$$

obtained by rewriting Eq. (4.3), which is of the form $P(A|B,C) = P(A|B)$, in the form $P(A,C|B) = P(A|B)P(C|B)$. As Hiemstra and Jones we focus on the case $m = 1$ (one-step lead vectors of $\{Y_t\}$). In that case we find for $\ell_X = 1$:

$$\begin{aligned} P\left(|Y_t - Y_s| < h, |X_{t-1} - X_{s-1}| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right) \\ = P\left(|Y_t - Y_s| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right) \\ \times P\left(|X_{t-1} - X_{s-1}| < h \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right). \end{aligned} \tag{4.10}$$

The difference between the left- and right-hand sides can be expressed as a conditional covariance

$$D = \mathrm{Cov}\left(I(|Y_t - Y_s| < h), I(|X_{t-1} - X_{s-1}| < h) \middle| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h\right), \tag{4.11}$$

where $I(\cdot)$ denotes the indicator function which is one if the inequality in its argument holds and zero otherwise.

Although for given values of Y_{t-1} and Y_{s-1}, $Y_t - Y_s$ and $X_{t-1} - X_{s-1}$ are independent

normally distributed random variables (with variances $V = g(Y_{t-1}) + g(Y_{s-1})$ and $W = f(Y_{t-1}) + f(Y_{s-1})$, respectively) the condition in Eq. (4.11) does not fix Y_{t-1} and Y_{s-1}, so that the covariance need not be zero. We obtain

$$D = \mathrm{Cov}\left(I\left(|Z_1|\sqrt{V} < h\right), I\left(|Z_2|\sqrt{W} < h\right)\right), \tag{4.12}$$

where Z_1 and Z_2 are standard normal random variables, independent of each other and (V, W), and V and W are the conditional variances of $Y_t - Y_s$ and $X_{t-1} - X_{s-1}$, given $\|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h$, which are jointly distributed according to

$$(V, W) \sim (g(Y_{t-1}) + g(Y_{s-1}), f(Y_{t-1}) + f(Y_{s-1})) \left| \|Y_{t-\ell_Y}^{\ell_Y} - Y_{s-\ell_Y}^{\ell_Y}\| < h. \right. \tag{4.13}$$

Since Z_1 and Z_2 are independent, and independent of (V, W), the expectations with respect to those variables can be taken. If we define

$$r(s) = E\left(I\left(|Z| < s\right)\right) = P(|Z| < s), \qquad \text{for } Z \sim N(0,1), \tag{4.14}$$

we obtain

$$D = \mathrm{Cov}\left(r(h/\sqrt{V}), r(h/\sqrt{W})\right). \tag{4.15}$$

Depending on the joint distribution of V and W, D can be either negative, zero, or positive. The most problematic case is $D > 0$, since the one-sided Hiemstra-Jones test will then tend to over-reject. Clearly, if either V or W is degenerate (i.e. with probability one takes only one specific value), the covariance is zero and $D = 0$. The case $D > 0$ thus requires V and W to be non-degenerate random variables. Let us focus on V first. The fact that V can have a non-degenerate distribution follows from the existence of stationary ARCH(1) processes with time varying conditional variance. If for such non-degenerate V we define W in such a way that it is positively correlated with V, then $D > 0$. An obvious example would be to take $g(s) = f(s)$ for all s, which implies $W = V$. In that case one finds $D = \mathrm{Var}(r(h/\sqrt{V})) > 0$.

Further analytic results presented in the next chapter, indicate that also for processes of a different form than that in Eq. (4.7) and (4.8), but which also satisfy the null hy-

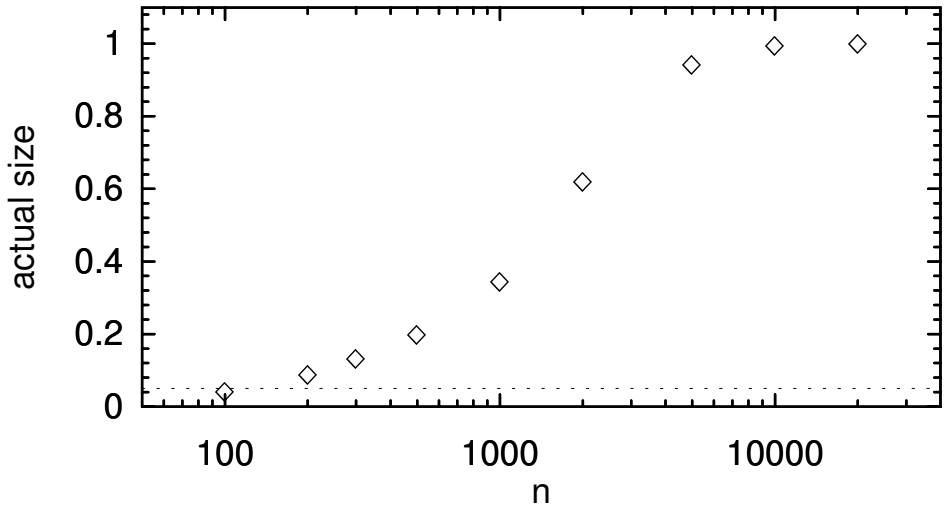

Figure 4.1: Simulated size of the Hiemstra-Jones test ($h = 1$, $\ell_Y = \ell_X = 1$) for the bivariate ARCH process given in Eq. (4.17) as a function of the time series length (nominal size 0.05). Number of realisations: $1,000$ for $n < 10,000$, and 500, 100 for $n = 10,000$ and $20,000$ respectively.

pothesis, D typically is nonzero. These results also suggest some ways of reducing the bias, which we hope to prove useful in future work on alternative tests. The fact that the sizes reported in the bootstrap study by Diks and DeGoede (2001) were close to nominal is related to the relatively small sample sizes used there.

The data generating process for the time series used to generate Table 4.1 in Section 4.1 was:

$$\begin{aligned} Y_t &\sim N\left(0.2X_{t-1}, 1 + 0.4Y_{t-1}^2\right) \\ X_{t-1} &\sim N\left(0, (1 + 0.4Y_{t-1}^2)^{-1}\right). \end{aligned} \tag{4.16}$$

The resulting time series processes are stationary and exhibit conditional heteroskedasticity. Clearly, $\{X_t\}$ linearly Granger causes $\{Y_t\}$. At the same time there is contemporaneous nonlinear dependence between $\{X_t\}$ and $\{Y_t\}$. There is a negative bias in the Hiemstra-Jones test statistic introduced by the negative dependence between the conditional variance of X_{t-1} and Y_t given Y_{t-1}. This destroys power in the Granger non-causality test from X to Y and leads to smaller test statistics than expected under the null in the test of Granger non-causality from Y to X.

Simulations

To illustrate the effect of a positive covariance in Eq. (4.15) we imposed V and W to be equal (see Eq. 4.13) by taking $g(Y_{t-1}) = f(Y_{t-1}) = 1 + 0.4Y_{t-1}^2$. We simulated the resulting bivariate ARCH process,

$$\begin{aligned} Y_t &\sim N(0, 1 + 0.4Y_{t-1}^2) \\ X_{t-1} &\sim N(0, 1 + 0.4Y_{t-1}^2), \end{aligned} \tag{4.17}$$

which is true under the null and calculated the rejection rates (at nominal size 0.05) of the Hiemstra-Jones test for the null hypothesis that X does not Granger cause Y.

Figure 4.1 shows the rejection rates found as a function of the time series length n. The actual size of the test is close to the nominal size of 0.05 only for the short time series of 100. The size increases with the length n of the time series, and is close to one already for time series of length $5,000$.

4.4 Case study

In practice the test is usually applied after filtering out seasonalities, linear structure, and (G)ARCH structure. Although this may lead to smaller rejection rates due to whitening of the data, it does not affect our conclusion in typical cases where the model specification is not known to be correct. To illustrate this point we mimic a typical empirical study relying on the Hiemstra-Jones test by investigating an artificial bivariate process of the form

$$\begin{aligned} Y_t &\sim N(-2Y_{t-1}e^{-Y_{t-1}^2}, 1 + 0.4Y_{t-1}^2) \\ X_{t-1} &\sim N(-2Y_{t-1}e^{-Y_{t-1}^2}, 1 + 0.4Y_{t-1}^2). \end{aligned} \tag{4.18}$$

The process (4.18) satisfies the null hypothesis that X does not Granger cause Y. The Y-series exhibits nonlinear AR(1) dependence in the mean and ARCH(1) structure, while X is instantaneously driven by Y through the mean and variance. Time series of $1,000$ and $10,000$ observations were considered for the study.

As mentioned above, a researcher usually filters the data to correct for some structure.

Procedure	$n = 1,000$		$n = 10,000$	
	Z-stat	P-value	Z-stat	P-value
Raw data	3.183796	0.0007	7.893391	0.0000
GARCH(1,1)	1.689628	0.0455	5.823490	2.9E-9
GARCH(1,1) & AR(1)	2.362784	0.0091	4.932293	4.1E-7
Correct model	1.323609	0.0928	0.082326	0.4672

Table 4.2: Results of the case study with different filters

The univariate GARCH(1,1) model is a popular choice in financial time series. For both series we apply the GARCH(1,1) filter with two different mean specifications. First, we consider a simple model with constant mean. Thereafter, also AR(1) structure typical in financial studies is included in the mean equation. Table 4.4 summarises the results of the Hiemstra-Jones tests after the above procedures. The results for the raw data and the residuals of the correctly specified model are included for reference. The test on the raw data strongly rejects the null hypothesis while it holds, as a result of the bias. Filtering with a GARCH(1,1) model with constant mean reduces the bias, but because of misspecification does not remove it completely. Adding AR(1) to the mean equation worsens the bias compared to the former procedure when the series length is 1000. This suggests that removing (G)ARCH and AR structure without knowledge of the correct model class may not correct the test and consequently, produce unreliable results. The test on the residuals of the correctly specified model leads to an anticipated result since the residual series by construction are practically independent, in which case Eq. (4.3) holds. In that sense the Hiemstra-Jones test performed on the residuals may be considered as a model specification test.

4.5 Concluding remarks

The analytic and numerical evidence presented in this chapter clearly shows that Eq. (4.3), which is the relationship tested in the Hiemstra-Jones test, is not generally compatible with the null hypothesis stated in Eq. (4.1). This indicates that rejections of the null hypothesis reported in the empirical literature may be spurious. A simulated empirical study shows that the main conclusion remains valid also for studies which corrected for

AR and (G)ARCH structure.

One might still argue, correctly, that the Hiemstra-Jones test is a valid test of the relationship given in Eq. (4.3). In fact one might even go one step further and take Eq. (4.3) as a definition of Granger causality, which is exactly the approach taken in the original test by Baek and Brock (1992). Although one can indeed test (4.3) using the Hiemstra-Jones test, the interpretation involves some subtleties. A problem with this approach is that it is hard to find out in detail exactly which subclass of data generating processes satisfy the null hypothesis. Although it is easy to give some sufficient conditions for (4.3) to hold for all h (for example, $\{Y_t\}$ and $\{X_t\}$ being independent) it is surprisingly difficult to formulate necessary conditions in terms of the data generating process.

We finally note that, since for $\{X_t\}$ and $\{Y_t\}$ independent (4.3) holds for all h, the Hiemstra-Jones test might still be used as a model specification test by applying it to the residuals of an estimated model for the data generating process. However, in that case the Hiemstra-Jones test is used as a test for independence rather than conditional independence.

Chapter 5

A new statistic and practical guidelines for nonparametric Granger non-causality testing

5.1 Introduction

Granger (1969) causality has turned out to be a useful notion for characterising dependence relations between time series in economics and econometrics. Intuitively, for a strictly stationary bivariate process $\{(X_t, Y_t)\}$, $\{X_t\}$ is a Granger cause of $\{Y_t\}$ if past and current values of X contain additional information on future values of Y that is not contained in past and current Y-values alone. If we denote the information contained in past observations X_s and Y_s, $s \leq t$, by $\mathcal{F}_{X,t}$ and $\mathcal{F}_{Y,t}$, respectively, and let '$\sim$' denote equivalence in distribution, the formal definition is:

Definition 5.1 *For a strictly stationary bivariate time series process* $\{(X_t, Y_t)\}$, $t \in \mathbb{Z}$, $\{X_t\}$ *is a Granger cause of* $\{Y_t\}$ *if, for some* $m \geq 1$,

$$(Y_{t+1}, \ldots, Y_{t+m})|(\mathcal{F}_{X,t}, \mathcal{F}_{Y,t}) \not\sim (Y_{t+1}, \ldots, Y_{t+m})|\mathcal{F}_{Y,t}.$$

Since this definition is general and does not involve any modelling assumptions, such as a linear autoregressive model, it is often referred to as general or, by a slight abuse of

language, nonlinear Granger causality.

Traditional parametric tests for Granger non-causality within linear autoregressive model classes have reached a mature status, and have become part of the standard toolbox of economists. The recent literature, due to the availability of ever cheaper computational power, has shown an increasing interest in nonparametric versions of the Granger non-causality hypothesis against general (linear as well as nonlinear) Granger causality. Among the various nonparametric tests for the Granger non-causality hypothesis, the Hiemstra and Jones (1994) test (hereafter HJ test) is the most frequently used among practitioners in economics and finance. Although alternative tests, such as that proposed by Bell, Kay, and Malley (1996), and by Su and White (2003), may also be applied in economics and finance, we limit ourselves to a discussion of the HJ test and our proposed modification of it.

The reason for considering the HJ test here in detail is our earlier finding (c.f. Chapter 4) that this commonly used test can severely over-reject if the null hypothesis of Granger non-causality is true. The aim of the present chapter is two-fold. First, we derive the exact conditions under which the HJ test over-rejects, and secondly we propose a new test statistic which does not suffer from this serious limitation. We will show that the reason for over-rejection of the HJ test is that the test statistic, due to its global nature, ignores the possible variation in conditional distributions that may be present under the null hypothesis. Our new test statistic, provided that the bandwidth tends to zero at an appropriate rate, automatically takes into account such variation under the null hypothesis while obtaining an asymptotically correct size.

The practical implication of our findings is far-reaching: all cases for which evidence for Granger causality was reported based on the HJ test may be caused by the tendency of the HJ test to over-reject. Reports of such evidence are numerous in the economics and finance literature. For instance, Brooks (1998) finds evidence for Granger causality between volume and volatility on the New York Stock Exchange, Abhyankar (1998) and Silvapulla and Moosa (1999) in futures markets, and Ma and Kanas (2000) in exchange rates. Further evidence for causality is reported in stock markets (Ciner, 2001), among real estate prices and stock markets (Okunev, Wilson, and Zurbruegg, 2000, 2002) and

between London Metal Exchange cash prices and some of its possible predictors (Chen and Lin, 2004). Although we do not claim that the reported Granger causality is absent in all these cases, we do state that the statistical justification is not warranted.

This chapter is organised as follows. In Section 5.2 we show that the HJ test statistic can give rise to rejection probabilities that tend to one with increasing sample size under the null hypothesis. In Section 5.3 the reason behind this phenomenon is studied analytically and found to be related to a bias in the test statistic due to variations in conditional distributions. The analytic results suggest an alternative test statistic, described in Section 5.4, which automatically takes these variations into account, and can be shown to give asymptotic rejection rates equal to the nominal size for bandwidths tending to zero at appropriate rates. The theory is confirmed by the simulation results presented at the end of the section. In Section 5.5 we consider an application to S&P500 volumes and returns for which the HJ test indicates volume Granger-causing returns, while our test indicates that the evidence for volume causing returns is considerably weaker. Section 5.6 summarises and concludes.

5.2 The Hiemstra-Jones test

In testing for Granger non-causality, the aim is to detect evidence against the null hypothesis

$$H_0: \qquad \{X_t\} \text{ is not Granger causing } \{Y_t\},$$

with Granger causality defined according to Definition 5.1. We limit ourselves to tests for detecting Granger causality for $m = 1$, which is the case considered most often in practice. Under the null hypothesis Y_{t+1} is conditionally independent of $X_t, X_{t-1}, \ldots$, given $Y_t, Y_{t-1}, \ldots$. In a nonparametric setting, conditioning on the infinite past is impossible without a model restriction, such as an assumption that the order of the process is finite. Therefore, in practice conditional independence is tested using finite lags ℓ_X and ℓ_Y:

$$Y_{t+1}|(X_t^{\ell_X}; Y_t^{\ell_Y}) \sim Y_{t+1}|Y_t^{\ell_Y},$$

where $X_t^{\ell_X} = (X_{t-\ell_X+1}, \ldots, X_t)$ and $Y_t^{\ell_Y} = (Y_{t-\ell_Y+1}, \ldots, Y_t)$. For a strictly stationary bivariate time series $\{(X_t, Y_t)\}$ this is a statement about the invariant distribution of the $\ell_X+\ell_Y+1$-dimensional vector $W_t = (X_t^{\ell_X}, Y_t^{\ell_Y}, Z_t)$, where $Z_t = Y_{t+1}$. To keep the notation compact, and to bring about the fact that the null hypothesis is a statement about the invariant distribution of W_t, we often drop the time index and just write $W = (X, Y, Z)$, where the latter is a random vector with the invariant distribution of $(X_t^{\ell_X}, Y_t^{\ell_Y}, Y_{t+1})$. Here we only consider the choice $\ell_X = \ell_Y = 1$, in which case $W = (X, Y, Z)$ denotes a three-variate random variable, distributed as $W_t = (X_t, Y_t, Y_{t+1})$. Throughout we will assume that W is a continuous random variable.

The HJ test is a modified version of the Baek and Brock (1992) test for conditional independence, with critical values based on asymptotic theory. To motivate the test statistic it is convenient to restate the null hypothesis in terms of ratios of joint distributions. Under the null the conditional distribution of Z given $(X, Y) = (x, y)$ is the same as that of Z given $Y = y$ only, so that the joint probability density function $f_{X,Y,Z}(x, y, z)$ and its marginals must satisfy

$$\frac{f_{X,Y,Z}(x,y,z)}{f_{X,Y}(x,y)} = \frac{f_{Y,Z}(y,z)}{f_Y(y)}, \tag{5.1}$$

or equivalently

$$\frac{f_{X,Y,Z}(x,y,z)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}\frac{f_{Y,Z}(y,z)}{f_Y(y)} \tag{5.2}$$

for each vector (x, y, z) in the support of (X, Y, Z). The last equation is identical to $f_{X,Z|Y}(x, z|y) = f_{X|Y}(x|y)f_{Z|Y}(z|y)$, which explicitly states that X and Z are independent conditionally on $Y = y$, for each fixed value of y.

The Hiemstra-Jones test employs ratios of correlation integrals to measure the discrepancy between the left- and right-hand-sides of (5.1). For a multivariate random vector V taking values in $\mathbb{R}^{d_V}$ the associated correlation integral $C_V(h)$ is the probability of finding two independent realisations of the vector at a distance smaller than or equal to h:

$$\begin{aligned} C_V(h) &= P[\|V_1 - V_2\| \leq h], \qquad V_1, V_2 \text{ indep.} \sim V \\ &= \int\int I(\|s_1 - s_2\| \leq h) f_V(s_1) f_V(s_2)\, ds_2\, ds_1 \end{aligned}$$

where $I(\|s_1 - s_2\| \leq h)$ is the indicator function, which is one if $\|s_1 - s_2\| \leq h$ and zero

otherwise, and $\|x\| = \sup_{i=1,\ldots,d_V} |x_i|$ denotes the supremum norm. Hiemstra and Jones (1994) argue that Eq. (5.1) implies for any $h > 0$:

$$\frac{C_{X,Y,Z}(h)}{C_{X,Y}(h)} = \frac{C_{Y,Z}(h)}{C_Y(h)} \tag{5.3}$$

or equivalently

$$\frac{C_{X,Y,Z}(h)}{C_Y(h)} = \frac{C_{X,Y}(h)}{C_Y(h)} \frac{C_{Y,Z}(h)}{C_Y(h)}. \tag{5.4}$$

Note that Eqs. (5.3–5.4) correspond to Eq. (4.3) in the case $\ell_X = \ell_Y = 1$.

The HJ test consists of calculating sample versions of the correlation integrals in Eq. (5.3), and then testing whether the left-hand- and right-hand-side ratios differ significantly or not. The estimators for each of the correlation integrals take the form

$$C_{W,n}(h) = \frac{2}{n(n-1)} \sum \sum_{i<j} I_{ij}^W,$$

where $I_{ij}^W = I(\|W_i - W_j\| \leq h)$. For the asymptotic theory we refer to Hiemstra and Jones (1994).

As stated in the introduction, the main motivation for the present chapter is that in certain situations the HJ test rejects too often under the null, and we wish to formulate an alternative procedure to avoid this. Before investigating the reasons for over-rejection analytically, we use a simple example to illustrate the over-rejection numerically, and to show that simple remedies such as transforming the data to uniform marginals and filtering out GARCH structure do not work. In Chapter 4, we demonstrated that for a process with instantaneous dependence in conditional variance the actual size of the HJ test was severely distorted. Here we illustrate the same point for the similar process, but without instantaneous dependence:

$$\begin{aligned} X_t &\sim N\left(0, c + aY_{t-1}^2\right) \\ Y_t &\sim N\left(0, c + aY_{t-1}^2\right). \end{aligned} \tag{5.5}$$

This process satisfies the null hypothesis; $\{X_t\}$ is not Granger causing $\{Y_t\}$. The values for the coefficients a and c are chosen in such a way that the process remains stationary

and ergodic ($c > 0$, $0 < a < 1$).

We performed some Monte Carlo simulations to obtain the empirical size of the HJ test for the ARCH process (5.5) with coefficients $c = 1$, $a = 0.4$. For various sample sizes, we generated $1,000$ independent realisations of the bivariate process and determined the observed fraction of rejections of the null at a nominal size of 0.05. The solid line in Figure 5.1 shows the rejection rates found as a function of the time series length n. The simulated data were normalised to unit variance before the test was applied, and the bandwidth was set to $h = 1$, which is within the common range $(0.5, 1.5)$ used in practice. For time series length $n < 500$ the test based on the original series under-rejects. Its size is close to nominal for series length $n = 500$. For longer series the actual size increases and becomes close to one when $n = 60,000$. The reason that the observed size increases with the series length n is that, as detailed in the next section, the test statistic is biased in that it does not converge in probability to zero under the null as the sample size increases. As the sample size increases the biases converges to a nonzero limit while the variance decreases to zero, giving rise to apparently significant values of the test statistic. In comparison with the process with instantaneous dependence considered in Chapter 4 the current process indicates less size distortion. This is due to the weaker covariance between the concentration measures H_X and H_Z, formally defined further in Section 5.3, for the current process, which is the main cause of the bias.

As suggested by Pompe (1993) in the context of testing for serial independence, transforming the time series to a uniform marginal distribution by using ranks, may improve the performance of the test. Here we investigate if it reduces the bias of the HJ test. The long-dashed line in Figure 5.1 shows that the uniform transform improves the size for time series of length $n = 1,000$, but magnifies the size distortion for time series length $n > 2,000$.

As another solution one might argue that it is possible to filter out the conditional heteroskedasticity using a univariate (G)ARCH specification. This would remove the bias caused by the conditional heteroskedasticity in the HJ test. However such a filtering procedure has several drawbacks. First, it may affect the dependence structure and consequently the power of the test. Second, a (G)ARCH filter may not fully remove the

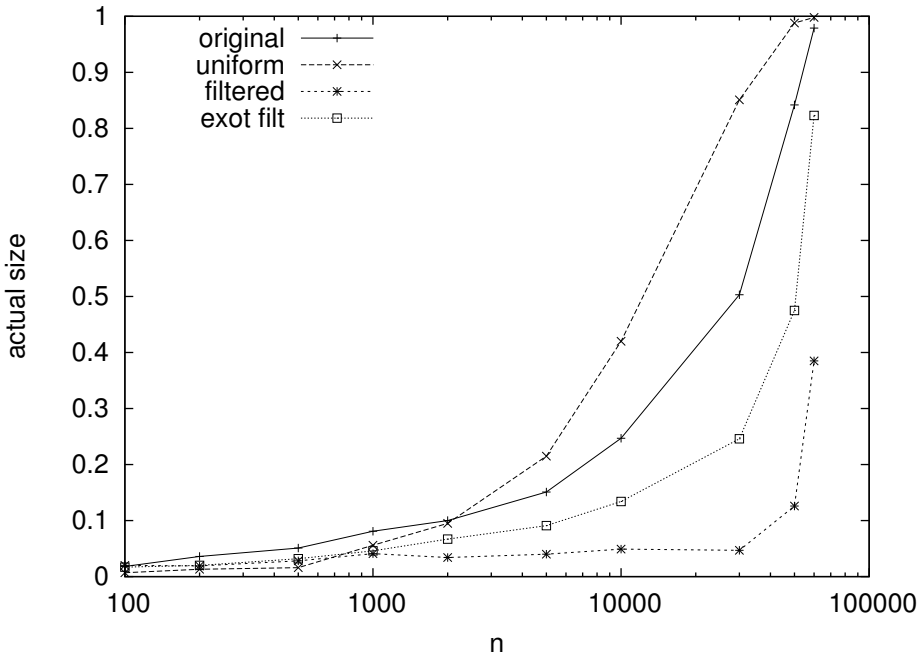

Figure 5.1: Observed rejection rates (empirical size, number of realisations: $1,000$) of the HJ test ($h = 1$) for the bivariate ARCH process (5.5) as a function of the time series length n (nominal size 0.05) for: original data (solid line), uniformly transformed data (long-dashed line), ARCH filtered data (dashed line) and for data generated with "exotic" model (5.6) and filtered with a misspecified ARCH(1) model (dotted line).

conditional heteroskedasticity in the residuals. To illustrate the latter point we filtered the original series considered before by univariate ARCH(1) model. The parameters of the model were estimated for every realisation using the asymptotically efficient two stage procedure of Engle (1982). Figure 5.1 (dashed line) shows that the filtering removes the bias for time series length $n < 30,000$, however the actual size remains distorted for longer series.

It is important to mention that in the previous case the correct model for the conditional variance of series Y_t was used and, as the next section clarifies, most of the source of the bias was removed. In practice the correct model is not known and the model used to filter out the heteroskedasticity is likely to be misspecified. To show the effect of model misspecification we generated data according to the following "exotic" ARCH model:

$$\begin{aligned} X_t &\sim N\left(0, c + aY_{t-1}^2 \exp(-bY_{t-1}^2))\right) \\ Y_t &\sim N\left(0, c + aY_{t-1}^2 \exp(-bY_{t-1}^2))\right). \end{aligned} \tag{5.6}$$

With parameters $c = 1$, $a = 2$ and $b = 0.4$ the process (5.6) is stationary and the fluctuations in the conditional variance are similar in magnitude as for the ARCH process (5.5) with the coefficients considered before. Instead of using a correctly specified filter we proceeded as before, calculating the size using a conventional ARCH(1) filter prior to application of the HJ test. The results represented by the dotted line in Figure 5.1 indicate that the misspecified ARCH(1) filter is not able to remove a large part of the source of the bias, and the sensitivity of the HJ test to dependence in the conditional variance leads to over-rejection, even for shorter time series.

5.3 Bias from correlations in conditional concentrations

In this section we show that the reason that the HJ test is inconsistent is that the assumption made by HJ that Eq. (5.1) implies Eq. (5.3) does not hold in general. In fact Eq. (5.3) follows from Eq. (5.1) only in specific cases, e.g. when the conditional distributions of Z and X given $Y = y$ do not depend on y. To see this, note that under the null hypothesis

$$\begin{aligned} &P\left(\|X_1 - X_2\| < h, \|Z_1 - Z_2\| < h \middle| Y_1 = Y_2 = y\right) \\ &\quad = P\left(\|X_1 - X_2\| < h \middle| Y_1 = Y_2 = y\right) P\left(\|Z_1 - Z_2\| < h \middle| Y_1 = Y_2 = y\right), \end{aligned} \tag{5.7}$$

whereas Eq. (5.4) states

$$\begin{aligned} &P\left(\|X_1 - X_2\| < h, \|Z_1 - Z_2\| < h \middle| \|Y_1 - Y_2\| < h\right) \\ &\quad = P\left(\|X_1 - X_2\| < h \middle| \|Y_1 - Y_2\| < h\right) P\left(\|Z_1 - Z_2\| < h \middle| \|Y_1 - Y_2\| < h\right). \end{aligned} \tag{5.8}$$

In general these conditions are not equivalent. In both equations a statement regarding the factorisation of probabilities is made, but the events on which the conditioning takes place differ. In general, under the null the conditional distributions of X and Z are allowed to depend on Y. Therefore, the distributions of $X_1 - X_2$ and $Z_1 - Z_2$ will generally depend, under the null, on Y_1 and Y_2. Even for small h the condition in Eq. (5.8) holds for many close but very different Y_1, Y_2 pairs. Therefore, for small h the left-hand-side of

Eq. (5.8) behaves as an average of that of Eq. (5.7) over all possible values of y. Because factorisation of densities is not preserved under averaging — $af_1(x)g_1(z)+(1-a)f_2(x)f_2(z)$ typically cannot be written as the product of a function of x and of z — the average probability on the left-hand side of Eq. (5.8) will typically not factorise in the form on the right-hand side.

Although this argument shows that the relationship tested in the HJ test is generally inconsistent with the null hypothesis, one might argue that the test could still be asymptotically valid if appropriate measures are taken to eliminate the 'bias' in Eq. (5.3) asymptotically, for example by allowing for the bandwidth h to tend to zero at an appropriate rate with increasing sample size.

To see whether such an approach might work we examine the behaviour of the fractions in (5.3) for small values of the bandwidth h. For continuous distributions the following small h approximation is useful:

$$\begin{aligned} C_V(h) &= \int\int I(\|s_1 - s_2\| \le h) f_V(s_1) f_V(s_2)\, ds_1\, ds_2 \\ &= \int\int_{B_h(s_1)} f_V(s_2)\, ds_2\, f_V(s_1)\, ds_1 + \mathrm{o}(h^{d_V}) \\ &= (2h)^{d_V} \int f_V^2(s)\, ds + \mathrm{o}(h^{d_V}) \\ &= (2h)^{d_V} H_V + \mathrm{o}(h^{d_V}), \end{aligned} \tag{5.9}$$

where $B_h(s_1)$ denotes a ball (or, since we use the supremum norm, a hypercube) with radius h centred at s_1. The constant $H_V \equiv \int f_V^2(s)\, ds = E\left[f_V(V)\right]$ can be considered as a *concentration measure* of V. To illustrate this, consider a family of univariate pdfs with scale parameter θ, that is, $f_V(v;\theta) = \theta^{-1} g(\theta^{-1} v)$ for some pdf $g(\cdot)$. One readily finds $\int f_V^2(s;\theta)\, ds = \frac{1}{\theta}\int g^2(s)\, ds = \frac{\text{cnst.}}{\theta}$, which shows that, in the univariate case, the concentration measure is inversely proportional to the scale parameter θ. For later convenience, for a pair of vector-valued random variables (V, Y) of possibly different dimensions, we also introduce the *conditional concentration* of the random variable V given $Y = y$, as $H_V(y) = \int f_{V|Y}^2(v|y)\, dv = (\int f_{V,Y}^2(v,y)\, dv)/f_Y^2(y)$.

By comparing the leading terms of the expansion in powers of h in Eqs. (5.4) and (5.9),

we find that

$$\frac{E[f_{X,Y,Z}(X,Y,Z)]}{E[f_Y(Y)]} = \frac{E[f_{X,Y}(X,Y)]}{E[f_Y(Y)]}\frac{E[f_{Y,Z}(Y,Z)]}{E[f_Y(Y)]}. \tag{5.10}$$

That is, for h small, testing the equivalence of the ratios in (5.3) amounts to testing (5.10) instead of the null hypothesis. Unless some additional conditions hold, this will typically not be equivalent to testing the null hypothesis. To see what these additional conditions are, it is useful to rewrite (5.10) as follows. For the left-hand-side one can write

$$\begin{aligned}\frac{E\left[f_{X,Y,Z}(X,Y,Z)\right]}{E[f_Y(Y)]} &= \frac{E_Y\left[E_{X,Z|Y}[f_{X,Z|Y}(X,Z|Y)f(Y)]\right]}{E[f_Y(Y)]} \\ &= \int E_{X,Z|Y=y}[f_{X,Z|Y}(X,Z|y)]w(y)\,dy \\ &= \int H_{X,Z}(y)w(y)\,dy,\end{aligned}$$

where $w(y)$ is a weight function given by $w(y) = f_Y^2(y)/\int f_Y^2(s)\,ds$. This brings about the fact that the ratio on the left-hand-side of (5.10) for small h is proportional to a weighted average of the conditional concentration $H_{X,Z}(y)$, with weight function $w(y)$. In a similar fashion, for the terms on the right-hand-side one derives

$$\frac{E\left[f_{X,Y}(X,Y)\right]}{E[f_Y(Y)]} = \int H_X(y)w(y)\,dy, \qquad \text{and} \qquad \frac{E\left[f_{Y,Z}(Y,Z)\right]}{E[f_Y(Y)]} = \int H_Z(y)w(y)\,dy.$$

Under the null hypothesis, Z is conditionally independent of X given $Y = y$, so that $H_{X,Z}(y)$ is equal to $H_X(y)H_Z(y)$, for all y. It follows that the left- and right-hand-sides of (5.10) coincide under the null if and only if

$$\int H_X(y)H_Z(y)w(y)\,dy - \int H_X(y)w(y)\,dy \int H_Z(y)w(y)\,dy = 0,$$

or

$$\mathrm{Cov}(H_X(S), H_Z(S)) = 0, \tag{5.11}$$

where S is a random variable with pdf $w(y)$. Only under specific conditions, such as either $H_X(y)$ or $H_Z(y)$ being independent on y, (5.11) holds under the null, and hence (5.3) as h tends to zero. Also if $H_X(y)$ and $H_Z(y)$ depend on y, (5.11) may hold, but this is an exception rather than the rule. Typically the covariance between the conditional

concentrations of X and Z given Y will not vanish, inducing a bias in the HJ test for small h.

Therefore, letting the bandwidth tend to zero with increasing sample size in the HJ test would not provide a theoretical solution to the problem of over- or under-rejection caused by positive or negative covariance of the concentration measures, respectively. In simulations for a particular process and small to moderate sample sizes one can often identify a seemingly adequate rate for bandwidths vanishing according to $h_n = Cn^{-\beta}$, for which the size of the HJ test remains close to nominal. However, this does not imply that using the HJ test with such a sample size dependent bandwidth is advisable in practice. The optimal choices for C and β may depend strongly on the data generating process, and our results show that asymptotically the HJ test for typical processes (those with non-vanishing covariance of concentrations of X and Y) is inconsistent.

The fact that the conditional concentration measures of $X_t^{\ell_X}$ and Y_{t+1} given $Y_t^{\ell_Y}$ affect the leading bias term poses severe restrictions on applicability to economic and financial time series in which conditional heteroskedasticity is usually present. Consequently there is a risk of over-rejection by the HJ test which can not be easily eliminated either by using (G)ARCH filtering, or by using a bandwidth that decreases with the sample size. To avoid this problem, in the next section we suggest a new test statistic for which a consistent test is obtained as h tends to zero at the appropriate rate. The idea is to measure the dependence between X and Z given $Y = y_i$ locally for each y_i. By allowing for the bandwidth to decrease with the sample size, variations in the local (fixed Y) distributions of X and Z given Y are automatically taken into account by the test statistic.

5.4 A modified test statistic

In comparing Eqs. (5.2) and (5.10) it can be noticed that although Eq. (5.2) holds pointwise for any triple (x, y, z) in the support of $f_{X,Y,Z}(x, y, z)$, Eq. (5.10) contains separate averages for the nominator and the denominator of Eq. (5.2), which do not respect the fact that the y-values on the right-hand side of Eq. (5.2) should be identical. Because

Eq. (5.2) holds point-wise, rather than Eq. (5.10), the null hypothesis implies

$$q_g \equiv E\left[\left(\frac{f_{X,Y,Z}(X,Y,Z)}{f_Y(Y)} - \frac{f_{X,Y}(X,Y)}{f_Y(Y)}\frac{f_{Y,Z}(Y,Z)}{f_Y(Y)}\right) g(X,Y,Z)\right] = 0$$

where $g(x,y,z)$ is a positive weight function Under the null hypothesis the term within the round brackets vanishes, so that the expectation is zero. Although q_g is not positive definite, a one-sided test, rejecting when its estimated value is too large, in practice is often found to have larger power than a two-sided test. In tests for serial dependence Skaug and Tjøstheim (1993b) report good performance of a closely related unconditional test statistic (their dependence measure I_4 is an unconditional version of our term in round brackets).

We have considered several possible choices of the weight function g, being
(i) $g_1(x,y,z) = f_Y(y)$,
(ii) $g_2(x,y,z) = f_Y^2(y)$ and
(iii) $g_3(x,y,z) = f_Y(y)/f_{X,Y}(x,y)$.
Monte Carlo simulations using the stationary bootstrap (Politis and Romano, 1994) indicated that g_1 and g_2 behave similarly and are more stable than g_3. We will focus on g_2 in this chapter, as its main advantage over g_1 is that the corresponding estimator has a representation as a U-statistic, allowing the asymptotic distribution to be derived analytically for weakly dependent data, thus eliminating the need of the computationally more requiring bootstrap procedure. For the choice $g(x,y,z) = f_Y^2(y)$, we refer to the corresponding functional simply as q:

$$q = E\left[f_{X,Y,Z}(X,Y,Z)f_Y(Y) - f_{X,Y}(X,Y)f_{Y,Z}(Y,Z)\right].$$

A natural estimator of q based on indicator functions is:

$$T_n(h) = \frac{(2h)^{-d_X-2d_Y-d_Z}}{n(n-1)(n-2)}\sum_i\left[\sum_{k,k\neq i}\sum_{j,j\neq i}\left(I_{ik}^{XYZ}I_{ij}^{Y} - I_{ik}^{XY}I_{ij}^{YZ}\right)\right],$$

where $I_{ij}^W = I(\|W_i - W_j\| < h)$. Note that the terms with $k = j$ need not be excluded explicitly as these each contribute zero to the test statistic. The test statistic can be

interpreted as an average over local BDS test statistics (see Brock, Dechert, Scheinkman, and LeBaron, 1996), for the conditional distribution of X and Z, given $Y = y_i$.

If we denote local density estimators of a d_W-variate random vector W at W_i by

$$\widehat{f}_W(W_i) = \frac{(2h)^{-d_W}}{n-1} \sum_{j,j\neq i} I_{ij}^W,$$

the test statistic simplifies to

$$T_n(h) = \frac{(n-1)}{n(n-2)} \sum_i (\widehat{f}_{X,Y,Z}(X_i, Y_i, Z_i)\widehat{f}_Y(Y_i) - \widehat{f}_{X,Y}(X_i, Y_i)\widehat{f}_{Y,Z}(Y_i, Z_i)).$$

For an appropriate sequence h_n of bandwidth values these estimators are consistent and the test statistic consists of a weighted average of local contributions $\widehat{f}_{X,Y,Z}(x,y,z)\widehat{f}_Y(y) - \widehat{f}_{X,Y}(x,y)\widehat{f}_{Y,Z}(y,z)$ which tends to zero in probability under the null hypothesis.

In Appendix 5.A, using the approach proposed by Powell and Stoker (1996), we show that for $d_X = d_Y = d_Z = 1$ the test is consistent if we let the bandwidth depend on the sample size as

$$h_n = Cn^{-\beta} \tag{5.12}$$

for any positive constant C and $\beta \in (\frac{1}{4}, \frac{1}{3})$. In that case the test statistic is asymptotically normally distributed in the absence of dependence between the vectors W_i. Under suitable mixing conditions (Denker and Keller, 1983) this can be extended to a time series context provided that covariances between the local density estimators are taken into account, giving:

Theorem 5.1 *For a sequence of bandwidths h_n given by (5.12) with $C > 0$ and $\beta \in (\frac{1}{4}, \frac{1}{3})$ the test statistic T_n satisfies:*

$$\sqrt{n}\frac{(T_n(h_n) - q)}{S_n} \xrightarrow{d} N(0,1).$$

In Appendix 5.A the asymptotic normality of T_n is shown under a decreasing bandwidth, while Appendix 5.C considers the autocorrelation robust estimation of the asymptotic variance σ^2 by S_n^2.

5.4.1 Bandwidth choice

In the typical case where the local bias tends to zero at the rate h^2 as in Condition 1 in Appendix 5.A, the bandwidth choice which is optimal in that it asymptotically gives the estimator T_n with the smallest mean squared error (MSE) is given by

$$h_n^* = C^* n^{-\frac{2}{7}}$$

with

$$C^* = \left(\frac{18 \cdot 3 q_2}{4(E[s(W)])^2} \right)^{\frac{1}{7}} \tag{5.13}$$

as derived in Appendix 5.B.

To gain some insights into the order of magnitude of C^* it is helpful to calculate its value for some processes. Here we consider the ARCH process given in (5.5). The optimal C-value derived in the appendix is analytically hard to track since it involves the marginal distribution of the process. However, we can derive an approximate optimal value of C analytically by ignoring the deviation from normality of Y (an assumption which is reasonable for small a). Taking $Y \sim N(0,1)$ and X, Z independent and $N(0, 1 + aY^2)$ conditional on Y, we find

$$q_2 = \frac{\mathrm{e}^{2/a} \mathrm{erfc}(\sqrt{2/a})}{1152\pi^2\sqrt{a}}, \tag{5.14}$$

where $\mathrm{erfc}(s) = 1 - \mathrm{erf}(s)$ and

$$E[s(W)] = \frac{\sqrt{6a/\pi}(3+a) + (a(a-6)-9)\mathrm{e}^{3/(2a)}\mathrm{erfc}(\sqrt{3/(2a)})}{768\sqrt{2}a^{3/2}\pi^{3/2}}. \tag{5.15}$$

To investigate the behaviour of the bandwidth for small a, one may use the fact that

$$q_2 = \frac{1}{1152\sqrt{2}\pi^{3/2}} + \mathrm{o}(a) \qquad \text{and} \qquad E[s(W)] = a^2 \left(\frac{1}{288\sqrt{3}\pi^2} + \mathrm{o}(a) \right).$$

This suggests that as a tends to zero the (asymptotically) optimal bandwidth diverges at the rate $a^{-4/7}$. This is consistent with the fact that larger bandwidths are optimal for a smaller correlation between the conditional concentrations of X and Z given Y.

The optimal bandwidth for (G)ARCH filtered data depends on the correlation of the

n	100	200	500	1,000	2,000	5,000	10,000	20,000	60,000
h	1.50	1.50	1.50	1.20	1.00	0.76	0.62	0.51	0.37
size	0.022	0.033	0.052	0.052	0.051	0.050	0.050	0.052	0.053
power	0.073	0.155	0.411	0.661	0.900	0.998	1.000	1.000	1.000

Table 5.1: Observed rejection rates (size and power) of the T_n test for bivariate ARCH process (5.5) as a function of the time series length n and decreasing bandwidth h according to (5.16) (nominal size 0.05). Number of realisations: $10,000$ for $n < 60,000$, and $3,000$ for $n = 60,000$.

conditional concentrations after filtering, which may depend strongly on the underlying data generating process. However, the consistency of the test does not require filtering prior to testing, and it is possible to obtain a rough indication of the optimal bandwidth for raw returns. Since the covariance between conditional concentrations for bivariate financial time series are mainly due to ARCH/GARCH effects, Eqs. (5.14) and (5.15) can be used together with an estimate of the ARCH coefficient a to obtain a rough indication of the optimal constant C^* for applications to unfiltered financial returns data. To provide a feel for the order of magnitude: for $a = 0.4$ one finds $C^* \simeq 8$. Note that this value is asymptotically optimal and may lead to unrealistically large bandwidths for small n. In applications we therefore truncate the bandwidth by taking

$$h_n = \max(Cn^{-2/7}, 1.5). \tag{5.16}$$

5.4.2 Simulations

We use numerical simulations to investigate the behaviour of the proposed T_n test with the shrinking bandwidth given by (5.12). As the underlying process for the simulations we choose the process (5.5) considered before, a bivariate conditional heteroskedastic process with lag one dependence. The interest in this process is stipulated by its relevance to econometrics and financial time series. The null hypothesis $\{X_t\}$ is not Granger causing $\{Y_t\}$ is satisfied.

Table 5.1 reports the T_n test rejection rates (both size and power) for increasing series length n with n-dependent bandwidths h_n given by (5.12), for a nominal size of 0.05. The size computations were based on the ARCH process (5.5) with coefficients $c = 1$,

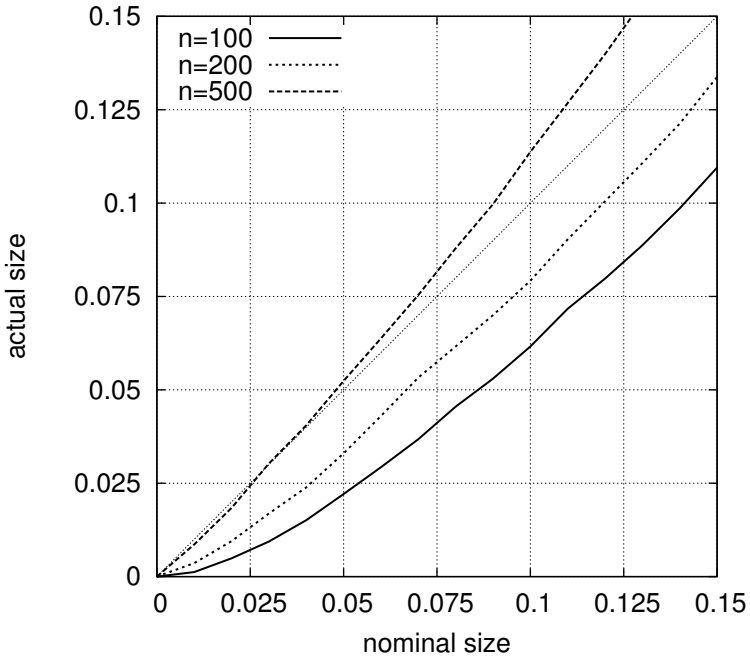

Figure 5.2: Size-size plot of T_n test for process (5.5) with shrinking bandwidth for time series lengths $n = 100$ (solid line), 200 (dashed line), 500 (long-dashed line). The number of realisations is $10,000$. The dotted line along the diagonal represents the ideal situation where the actual size and the nominal size coincide.

$a = 0.4$. For β we used the theoretically optimal rate of $\frac{2}{7}$, and we chose $C = 8.62$ which empirically turned out to give fast convergence of the size to the nominal value 0.05. This C-value is close to the approximate optimal asymptotic value $C^* \simeq 8$ for $a = 0.4$ reported above.

To compute the power we took the same process and reversed the roles of $\{X_t\}$ and $\{Y_t\}$, so that the relation tested became: $\{Y_t\}$ is not Granger causing $\{X_t\}$. For the power calculations the coefficient a was reduced to 0.1 to make the simulations more informative (for higher a the power was one in nearly all cases). The power of the test increases with n, in accordance with the consistency of the test under the decreasing bandwidth procedure.

To provide some guidance for choosing critical p-values in practice for small sample sizes, Figure 5.2 shows some size-size plots for small n ranging over nominal sizes between 0 and 0.15.

Finally, we present some simulations for lags $\ell_X = \ell_Y$ larger than one, since these are

used often for the HJ test. In the applications presented in the next section we compare both tests for larger values of ℓ_X and ℓ_Y as well, and to motivate this we should check if the empirical size of our new test does not exceed the nominal size for larger lags. Table 5.2 gives the empirical rejection rates for the bivariate ARCH process (5.5), again with $c = 1$ and $a = 0.4$, under the null hypothesis (that is, testing $\{X_t\}$ does not Granger cause $\{Y_t\}$) for lag lengths $\ell_X = \ell_Y$ ranging from 1 to 5. The results indicate that the rejection rate decreases with $\ell_X = \ell_Y$, and hence that the T_n test is progressively conservative for increasing lag lengths, so that the risk of rejecting under the null becomes small.

$\ell_X = \ell_Y$	$n = 1,000$	$n = 10,000$
1	0.0517	0.0502
2	0.0391	0.0316
3	0.0318	0.0197
4	0.0243	0.0112
5	0.0187	0.0099

Table 5.2: Observed rejection rates (empirical size) of T_n test for bivariate ARCH process (5.5) as a function of number of lags $\ell_X = \ell_Y$ for time series length $n = 1,000$ and $n = 10,000$ with optimal bandwidth $h = 1.2$ and $h = 0.62$ respectively (nominal size 0.05, number of realisations $10,000$)

5.5 Applications

We consider an application to daily volume and returns data for the Standard and Poor's 500 index in the period between 01/1950–12/1990. We deliberately have chosen this period to roughly correspond to the period for which Hiemstra and Jones (1994) found strong evidence for volume Granger-causing returns (1947 – 1990) for the Dow Jones index. To keep our results comparable with those of Hiemstra and Jones, we closely followed their procedure. That is, we adjusted for day-of-the-week and month-of-the-year effects on returns and percentage volume changes, using a two-step procedure in which we first adjust for effects in the mean, and subsequently in the variance. The calendar adjusted, standardised, returns and percentage volume change data were used to estimate a linear bivariate vector auto-regressive (VAR) model, the residuals of which are considered in the application below.

We applied the HJ and T_n test to the residuals of the VAR model, before as well as after EGARCH(1,1) filtering the VAR residuals of the returns data. Table 5.3 shows the resulting T-values for the HJ and T_n test in both directions, for $\ell_X = \ell_Y = 1, \ldots, 8$ and for two different values of h: 1.5, the value used by Hiemstra and Jones (1994) for the Dow Jones data, and 0.6, which is roughly the optimal value ($h^* \simeq 0.57$) we found from Eqs. (5.13–5.15) for the ARCH coefficient a, estimated from the data as 0.27.

	returns ⇒ volume				volume ⇒ returns			
	$h = 1.5$		$h = 0.6$		$h = 1.5$		$h = 0.6$	
$\ell_X = \ell_Y$	HJ	T_n	HJ	T_n	HJ	T_n	HJ	T_n
	before filtering							
1	9.476**	9.415**	10.298**	8.850**	5.351**	5.106**	5.736**	4.893**
2	10.989**	11.076**	10.616**	8.182**	6.671**	6.447**	6.818**	5.396**
3	10.909**	10.662**	9.112**	6.425**	6.026**	5.683**	5.717**	3.948**
4	10.758**	9.823**	7.934**	5.121**	6.029**	5.552**	4.692**	2.887**
5	10.118**	8.856**	5.821**	3.540**	5.695**	5.191**	2.837**	1.234
6	9.428**	7.903**	4.391**	2.603**	5.935**	5.338**	3.314**	1.604
7	8.959**	7.4215**	3.102**	2.085*	5.194**	4.706**	1.327	0.248
8	8.494**	6.577**	1.649*	0.701	4.484**	4.085**	0.418	0.567
	after EGARCH filtering							
1	7.461**	7.429**	7.946**	6.781**	1.532	1.481	1.628	1.529
2	8.444**	8.600**	8.012**	6.493**	3.022**	3.091**	3.251**	2.825**
3	7.537**	7.788**	6.381**	5.109**	1.894*	1.982*	2.534**	2.023*
4	7.257**	7.198**	5.169**	3.900**	2.141*	2.225*	1.964*	0.989
5	6.125**	6.107**	2.686**	2.023*	2.095*	2.142*	1.160	0.853
6	5.582**	5.445**	2.136*	1.477	2.969**	2.965**	1.411	1.129
7	5.028**	4.873**	1.192	0.532	2.278*	2.285*	1.414	0.943
8	4.495**	4.249**	0.779	0.253	1.754*	1.725*	0.398	0.860

Table 5.3: T-ratios for the S&P500 returns and volume data. Results are shown for the HJ test and T_n for bandwidth values of 1.5, the value used by Hiemstra and Jones (1994) and 0.6, corresponding to the optimal bandwidth for T_n (based on an estimated ARCH parameter 0.27). T-ratios before and after EGARCH filtering the returns are given, for $\ell_X = \ell_Y = 1, \ldots, 8$. The asterisks indicate significance at the 5% (*) and 1% (**) levels.

The results obtained with both tests strongly indicate evidence for returns affecting future volume changes, for nearly all lags and both bandwidths. Only for large values of the lags $\ell_X = \ell_Y$ the evidence is somewhat weaker. Although both tests point in the same direction, when comparing the overall results for equal bandwidths and lags $\ell_X = \ell_Y$ the T-values are somewhat smaller for the T_n test than for the HJ test. As argued in the previous sections, the HJ test may be inconsistent due to a bias which cannot be

removed simply by choosing a smaller bandwidth. To investigate the possible effects of this bias one should contrast the HJ test with our new test with an appropriately scaled bandwidth, which we have shown to be consistent asymptotically. That is, at least for the unfiltered data, one should actually compare the HJ test for $h = 1.5$ with the T_n test for the adaptive bandwidth 0.6. In that case the table shows even larger differences between the T-values of the HJ test and the T_n test.

For the other causal direction — volume changes affecting future returns — the different results obtained for the HJ test with $h = 1.5$ and the T_n test with $h = 0.6$, for the filtered data is large enough to make a difference for obtaining significance at the 5% and 1% nominal level for several lags. Overall, the evidence for volume changes affecting future returns, although still present after filtering for lag $\ell_X = \ell_Y = 2$ and arguably 3, is much weaker for T_n with $h = 0.6$ than for the HJ test with $h = 1.5$.

In summary, our findings on the basis of the Standard and Poor's data indicate that the strong evidence for volume Granger causing returns obtained with the HJ test may be partly due to the bias we identified in the HJ test statistic. If the test is performed with the consistent T_n statistic with a near-optimal bandwidth, for which theory and simulations indicate that the actual size is close to nominal, the evidence for volume Granger causing returns tends to become weaker. Finally, since the T-values can be seen to decrease for smaller h in most cases, the results also suggest that, when in doubt, it is better to use a smaller bandwidth. Intuitively this is related to the fact that it reduces the bias and increases the variance of the test statistic relative to the bias, so that the risk of over-rejection becomes smaller.

5.6 Concluding remarks

Motivated by the fact that the HJ test can over-reject, as demonstrated in simulations, our aim was to construct a new test for Granger non-causality. By analysing the HJ test analytically we found it to be biased even if the bandwidth tends to zero. Based on the analytic results, which indicated that the bias is caused by covariances in conditional concentrations, we proposed a new test statistic T_n that automatically takes the variation

in concentrations into account.

By symmetrising the new test statistic, we expressed it as a U-statistic for which we developed asymptotic theory under bandwidth values that tend to zero with the sample size at appropriate rates. The theory allowed us to derive the optimal rate as well as the asymptotically optimal multiplicative factor for the bandwidth. For ARCH type processes the optimal bandwidth can be expressed in terms of the ARCH coefficient, which is useful for getting an indication of the order of bandwidth magnitude to be used in practice for financial returns data. Simulations for the new test confirmed that the size converges to the nominal size fast as the sample size increases. Additional simulations indicated that the test becomes conservative for larger lags taken into account by the test.

In an application to relative volume changes and returns for historic Standard and Poor's index data we found that some of the strong evidence for relative volume changes Granger causing returns obtained with the HJ test may be related to its bias, since use of the new test, which is shown to be consistent, strongly weakens the evidence against the null hypothesis. This result suggests that some of the rejections of the Granger non-causality hypothesis reported in the literature may be spurious.

Appendix

5.A Asymptotic distribution of T_n

The test statistic T_n can be written in terms of a U-statistic by symmetrisation with respect to the three different indices. This gives

$$T_n(h) = \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k \neq i} K(W_i, W_j, W_k)$$

with $W_i = (X_i^{\ell_X}, Y_i^{\ell_Y}, Z_i)$, $i = 1, \ldots, n$ and

$$K(W_i, W_j, W_k) = \frac{(2h)^{-d_X - 2d_Y - d_Z}}{6} \left(\begin{array}{l} \left(I_{ik}^{XYZ} I_{ij}^{Y} - I_{ik}^{XY} I_{ij}^{YZ}\right) + \left(I_{ij}^{XYZ} I_{ik}^{Y} - I_{ij}^{XY} I_{ik}^{YZ}\right) + \\ \left(I_{jk}^{XYZ} I_{ji}^{Y} - I_{jk}^{XY} I_{ji}^{YZ}\right) + \left(I_{ji}^{XYZ} I_{jk}^{Y} - I_{ji}^{XY} I_{jk}^{YZ}\right) + \\ \left(I_{ki}^{XYZ} I_{kj}^{Y} - I_{ki}^{XY} I_{kj}^{YZ}\right) + \left(I_{kj}^{XYZ} I_{ki}^{Y} - I_{kj}^{XY} I_{ki}^{YZ}\right) \end{array} \right) \tag{5.17}$$

For a given bandwidth h the test statistic T_n is a third order U-statistic. To develop asymptotic distribution theory under a shrinking bandwidth h_n we closely follow the methodology proposed by Powell and Stoker (1996). Although their main goal was to derive MSE (mean squared error) optimal bandwidths for point estimators, it turns out that similar considerations can be used to derive rates for the bandwidth that provide consistency and asymptotic normality of T_n. We first treat the analytically simplest case of a random sample $\{W_i\}_{i=1}^n$, and deal with dependence later.

Because T_n is a U-statistic, its finite sample variance is given by (see Chapter 1):

$$\mathrm{Var}\left(T_n\right) = \frac{9}{n}\zeta_1 + \frac{18}{n^2}\zeta_2 + \frac{6}{n^3}\zeta_3 + \mathrm{o}\left(\frac{\zeta_1}{n} + \frac{\zeta_2}{n^2} + \frac{\zeta_3}{n^3}\right),$$

where

$$\begin{aligned} \zeta_1 &= \mathrm{Cov}(K(W_1, W_2, W_3), K(W_1, W_2', W_3')) = \mathrm{Var}(K_1(W_1)) \\ \zeta_2 &= \mathrm{Cov}(K(W_1, W_2, W_3), K(W_1, W_2, W_3') = \mathrm{Var}(K_2(W_1, W_2)) \\ \zeta_3 &= \mathrm{Var}(K(W_1, W_2, W_3)), \end{aligned}$$

with W_1, W_2, W_3, W_2' and W_3' all independent and identically distributed according to

W. The functions $K_1(w_1)$ and $K_2(w_1, w_2)$ are given by $K_1(w_1) = E[K(w_1, W_2, W_3)]$ and $K_2(w_1, w_2) = E[K(w_1, w_2, W_3)]$.

Following Powell and Stoker (1996), we define the functions $r(w, h) = K_1(w, h)$ and $r_0(w) = \lim_{h \to 0} r(w, h)$. It can be verified that

$$\begin{aligned} r_0(w) &= \tfrac{2}{3} f_{X,Y,Z}(x,y,z) f_Y(y) + \tfrac{1}{3} f_Y^2(y) H_{X,Z}(y) - \tfrac{1}{3} f_{X,Y}(x,y) f_{Y,Z}(y,z) \\ &\quad -\tfrac{1}{3} f_{Y,Z}(y,z) f_Y(y) \textstyle\int f_{X,Y}(x',y) f_{X,Y,Z}(x',y,z) dx' \\ &\quad -\tfrac{1}{3} f_{X,Y}(x,y) f_Y(y) \textstyle\int f_{Y,Z}(y,z') f_{X,Y,Z}(x,y,z') dz'. \end{aligned}$$

For example, the fourth term on the right-hand-side follows from:

$$\begin{aligned} (2h)^{-d_X - 2d_Y - d_Z} E_{W_k}[I_{jk}^{XY} I_{ji}^{YZ}] &= \textstyle\int f_{X,Y}(x_k, y_k) \delta_{x_j, y_j}(x_k, y_k) I_{ij}^{YZ} \, dx_k \, dy_k (2h)^{-d_Y - d_Z} \\ &\quad + \mathrm{o}(1) \\ &= f_{X,Y}(x_j, y_j) I_{ij}^{YZ} (2h)^{-d_Y - d_Z} + \mathrm{o}(1), \end{aligned}$$

where $\delta_{v_0}(v)$ stands for the Kronecker delta function, which can be thought of as the limiting pdf of a random variable with all mass at the point v_0, and

$$\begin{aligned} (2h)^{-d_Y - d_Z} E_{W_j}[f_{X,Y}(x_j, y_j) I_{ij}^{YZ}] &= \textstyle\int f_{X,Y}(x_j, y_j) \delta_{y_j, z_j}(y_i, z_i) f_{X,Y,Z}(x_j, y_j, z_j) \, dx_j \, dy_j \, \mathrm{d}z_j \\ &\quad + \mathrm{o}(1) \\ &= \textstyle\int f_{X,Y}(x_j, y_i) f_{X,Y,Z}(x_j, y_i, z_i) dx_j + \mathrm{o}(1). \end{aligned}$$

Adapting from Powell and Stoker (1996), we assume the following three conditions:

Condition 1*: (rate of convergence of pointwise bias of $r(w_i, h)$). The functions $r(w_i, h)$ satisfy*

$$r(w_i, h) - r_0(w_i) = s(w_i) h^\alpha + s^*(w_i, h),$$

for some $\alpha > 0$, and the remainder term $s^*(\cdot)$ satisfies $E\|s^*(W_i, h)\|^2 = \mathrm{o}(h^{2\alpha})$.

For our kernel the bias in each of the contributions to the kernel converges to zero at rate $\alpha = 2$. Therefore Condition 1 holds with $\alpha = 2$. In fact it might be possible to replace the local bias Condition 1 by a global version, involving $E[r(W_i, h) - r_0(W_i)]$, which may tend to zero faster than the local bias. However, for our purposes the local assumption with $\alpha = 2$ suffices.

Condition 2: *(series expansion for second moment of $K_2(W_1, W_2)$). The function $K_2(w_1, w_2)$ satisfies*

$$E\left[(K_2(W_1, W_2))^2\right] = q_2 h^{-\gamma} + q_2^*(h)$$

for some $\gamma > 0$, where the remainder term q_2^* satisfies $(q_2^*(h))^2 = \mathrm{o}(h^{-\gamma})$.

This is a weaker version of Powell and Stoker's (1996) Assumption 2, which required a series expansion locally. For our purposes the weaker assumption suffices, since T_n is a global functional of the distribution of W.

Condition 3: *(series expansion for second moment of $K(W_1, W_2, W_3)$). The function $K(w_1, w_2, w_3)$ satisfies*

$$E\left[(K(W_1, W_2, W_3))^2\right] = q_3 h^{-\delta} + q_3^*(h)$$

for some $\delta > 0$, where the remainder term q_3^* satisfies $(q_3^*(h))^2 = \mathrm{o}(h^{-\delta})$.

For our kernel Condition 3 is satisfied with $\delta = d_X + 2d_Y + d_Z$, since none of the contributions to the kernel have a variance increasing faster in h than at the rate $h^{d_X+2d_Y+d_Z}$. Finding an appropriate value for γ in Condition 2 is somewhat more involved. We examine the rate at which each of the contributions to the kernel function depend on h. For example, for the term $(2h)^{-d_X-2d_Y-d_Z} I_{ik}^{XYZ} I_{ij}^{Y}$ we find $E_{W_k}[(2h)^{-d_X-2d_Y-d_Z} I_{ik}^{XYZ} I_{ij}^{Y}] = (2h)^{-d_Y} f_{X,Y,Z}(X_i, Y_i, Z_i) I_{ij}^{Y} + \mathrm{o}(1)$ from which one obtains

$$\begin{aligned} E\left[\left((2h)^{-d_Y} E_{W_k}\left[I_{ik}^{XYZ} I_{ij}^{Y}\right]\right)^2\right] &= (2h)^{-2d_Y} E\left[f_{X,Y,Z}^2(X_i, Y_i, Z_i) I_{ij}^{Y} + \mathrm{o}(h^{d_Y})\right] \\ &= (2h)^{-d_Y} E\left[f_{X,Y,Z}^2(X_i, Y_i, Z_i) f_Y(Y_i)\right] + \mathrm{o}(h^{-d_Y}). \end{aligned}$$

Proceeding in this way for each of the terms in the kernel, one finds that the dominant contributions are given by the terms $(2h)^{-d_X-2d_Y-d_Z} I_{ij}^{XYZ} I_{ik}^{Y}$ and $(2h)^{-d_X-2d_Y-d_Z} I_{ji}^{XYZ} I_{jk}^{Y}$. For the first of these one finds $E_{W_k}[(2h)^{-d_X-2d_Y-d_Z} I_{ij}^{XYZ} I_{ik}^{Y}] = (2h)^{-d_X-d_Y-d_Z} I_{ij}^{XYZ} f_Y(Y_i) +$

o(1), giving

$$\begin{aligned} E\left[\left((2h)^{-d_X-d_Y-d_Z}E_{W_k}\left[I_{ij}^{XYZ}I_{ik}^{Y}\right]\right)^2\right] &= (2h)^{-2d_X-2d_Y-2d_Z}E\left[I_{ij}^{XYZ}f_Y^2(Y_i)\right] \\ &\quad +\mathrm{o}(h^{-d_X-d_Y-d_Z}) \\ &= (2h)^{-d_X-d_Y-d_Z}E\left[f_{X,Y,Z}(X_i,Y_i,Z_i)f_Y^2(Y_i)\right] \\ &\quad +\mathrm{o}(h^{-d_X-d_Y-d_Z}). \end{aligned}$$

All other terms increase with vanishing h slower, which demonstrates that Condition 2 holds with $\gamma = d_X + d_Y + d_Z$ and a constant q_2 given by

$$q_2 = \frac{4}{36} \times 2^{-d_X-d_Y-d_Z}E[f_{X,Y,Z}(X_i,Y_i,Z_i)f_Y^2(Y_i)].$$

The factor 4 enters due to the fact that there are two terms, $E_{W_k}[(2h)^{-d_X-2d_Y-d_Z}I_{ij}^{XYZ}I_{ik}^{Y}]$ and $E_{W_k}[(2h)^{-d_X-2d_Y-d_Z}I_{ij}^{XYZ}I_{jk}^{Y}]$, which are asymptotically perfectly correlated if h tends to zero sufficiently slowly with the sample size.

It follows from Condition 1 that

$$\mathrm{Var}\left[r(W_i,h)\right] = \mathrm{Var}\left[r_0(W_i)\right] + C_0h^{\alpha} + \mathrm{o}(h^{\alpha}),$$

where $C_0 = 2\mathrm{Cov}\left[r_0(W_i), s(W_i)\right]$. We can thus express the mean squared error of T_n as

$$\mathrm{MSE}[T_n] = \left(E[s(W_i)]\right)^2 h^{2\alpha} + \frac{9}{n}C_0h^{\alpha} + \frac{9}{n}\mathrm{Var}\left[r_0(W_i)\right] + \frac{18}{n^2}q_2h^{-\gamma} + \frac{6}{n^3}q_3h^{-\delta}. \qquad (5.18)$$

T_n is asymptotically $N(0,\sigma^2/n)$ distributed with $\sigma^2 = 9\mathrm{Var}[r_0(W_i)]$, provided that each of the h-dependent terms in the MSE of T_n are $\mathrm{o}(n^{-1})$. If we let $h \sim n^{-\beta}$, this implies the following four conditions should hold:

$$-2\alpha\beta < -1, \qquad -\alpha\beta < 0, \qquad \gamma\beta < 1, \qquad \delta\beta < 2.$$

The first two of these imply $\beta > \frac{1}{2\alpha} = \frac{1}{4}$ and $\beta > 0$, respectively, while the last two imply $\beta < \frac{1}{\gamma} = \frac{1}{d_X+d_Y+d_Z}$ and $\beta < \frac{2}{\delta} = \frac{2}{d_X+2d_Y+d_Z}$. Because $\frac{1}{d_X+d_Y+d_Z} < \frac{2}{d_X+2d_Y+d_Z}$, the conditions can be summarised as: $\frac{1}{4} < \beta < \frac{1}{d_X+d_Y+d_Z}$. Therefore, for the case $d_X = d_Y =$

$d_Z = 1$, and a sequence of bandwidths $h_n \sim n^{-\beta}$ for some $\beta \in (\frac{1}{4}, \frac{1}{3})$, the test statistic is asymptotically normal:

$$\sqrt{n}\frac{T_n(h_n) - q}{\sigma} \xrightarrow{d} N(0,1)$$

with $\sigma^2 = 9\mathrm{Var}[r_0(W_i)]$.

Note that it might also be possible to derive appropriate values for the rate β for $d_X + d_Y + d_Z > 3$, but only provided that the overall bias $E[s(W_i)]$ tends to zero faster than h^2.

5.B Optimal bandwidth

The MSE optimal bandwidth balances the dominating squared bias and variance terms (the first and fourth term on the right-hand-side of Eq. (5.18)), the other bandwidth dependent terms being of smaller order. The optimal bandwidth which asymptotically minimises the sum of these terms is given by

$$h^* = \left(\frac{18 \cdot 3q_2}{4(E[s(W)])^2}\right)^{\frac{1}{7}} n^{-\frac{2}{7}}. \tag{5.19}$$

To guide the choice of the multiplicative factor C in $h = Cn^{-\frac{2}{7}}$, it is illustrative to examine the optimal choice $C^* = \left(\frac{18 \cdot 3q_2}{4(E[s(W)])^2}\right)^{\frac{1}{7}}$ in specific cases. Above an expression for q_2 was found already in terms of the joint density of W. A similar expression for $E[s(W_i)]$ can be found by using local Taylor expansions of the density of w, locally near w_k. As each of the 6 terms in T_n have the same expectation, to determine the bias we consider the first of these only:

$$(2h)^{-d_X - 2d_Y - d_Z}(I_{ik}^{XYZ} I_{ij}^{Y} - I_{ik}^{XY} I_{ij}^{YZ}).$$

Taking averages over j and k for a fixed vector w_i leads to an expression involving plug-in estimators of local densities

$$\begin{aligned}\frac{(2h)^{-d_X - 2d_Y - d_Z}}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} (I_{ik}^{XYZ} I_{ij}^{Y} - I_{ik}^{XY} I_{ij}^{YZ}) \;&=\; \widehat{f}_{X,Y,Z}(x_i, y_i, z_i)\widehat{f}_Y(y_i) \\ &\quad - \widehat{f}_{X,Y}(x_i, y_i)\widehat{f}_{Y,Z}(y_i, z_i).\end{aligned}$$

An expression for the local bias can be obtained by examining the bias of each of the estimated densities in this expression.

For a general density $f_V(v)$ of a random vector $V = (V^1, \ldots, V^m)$, of which a sample $\{V_i\}_{i=1}^n$ is available, the bias of $\widehat{f}(\tilde{v}) = (2h)^m \frac{1}{n} \sum_{i=1}^n I(\|V_i - \tilde{v}\| \leq h)$ locally at $\tilde{v}$ can be found from a Taylor expansion of the density of $f_V(v)$ around $\tilde{v}$:

$$f(v) - f(\tilde{v}) = \sum_{i=1}^m a_i(\tilde{v})(v^i - \tilde{v}^i) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m b_{ij}(\tilde{v})(v^i - \tilde{v}^i)(v^j - \tilde{v}^j) + O(\|v - \tilde{v}\|^3),$$

with $a_i(\tilde{v}) = \frac{\partial}{\partial v^i}\big|_{v=\tilde{v}} f(v)$ and $b_{ij}(\tilde{v}) = \frac{\partial^2}{\partial v^i \partial v^j} |_{v=\tilde{v}} f(v)$. The local bias of $\widehat{f}(\tilde{v})$ is given by

$$\begin{aligned} E[\widehat{f}(\tilde{v})] - f(\tilde{v}) &= \frac{1}{2}(2h)^{-m} \sum_{i=1}^m \sum_{j=1}^m \int_{\tilde{v}^1-h}^{\tilde{v}^1+h} \cdots \int_{\tilde{v}^m-h}^{\tilde{v}^m+h} b_{ij}(\tilde{v})(v^i - \tilde{v}^i)(v^j - \tilde{v}^j)\, dv^1 \ldots dv^m \\ &\quad + \mathrm{o}(h^2) \\ &= \frac{1}{2}(2h)^{-1} \sum_{i=1}^m \int_{\tilde{v}^i-h}^{\tilde{v}^i+h} b_{ii}(\tilde{v})(v^i - \tilde{v}^i)^2 dv^i + \mathrm{o}(h^2) \\ &= (2h)^{-1} \frac{1}{3} h^3 \sum_{i=1}^m b_{ii}(\tilde{v}) + \mathrm{o}(h^2) \\ &= \frac{1}{6} h^2 \nabla^2 f(\tilde{v}) + \mathrm{o}(h^2). \end{aligned}$$

Up to leading order in h, the bias of products of estimated densities follows from identities such as $E[\widehat{f}_V \widehat{f}_W] = E[(f_V + (\widehat{f}_V - f_V))(f_W + (\widehat{f}_W - f_W))] = f_V f_W + f_V E[\widehat{f}_W - f_W] + f_W E[\widehat{f}_V - f_V] + \mathrm{o}(h^2)$. In this way the local bias of $\widehat{f}_{X,Y,Z}(x_i, y_i, z_i)\widehat{f}_Y(y_i) - \widehat{f}_{X,Y}(x_i, y_i)\widehat{f}_{Y,Z}(y_i, z_i)$ can be written as

$$\begin{aligned} r(w_i, h) - r_0(w_i) &= \tfrac{1}{6} h^2 \left[f_Y(y_i) \nabla^2 f_{X,Y,Z}(x_i, y_i, z_i) - f_{X,Y}(x_i, y_i) \nabla^2 f_{Y,Z}(y_i, z_i) \right. \\ &\quad \left. + f_{X,Y,Z}(x_i, y_i, z_i) \nabla^2 f_Y(y_i) - f_{Y,Z}(y_i, z_i) \nabla^2 f_{X,Y}(x_i, y_i) \right] \\ &\quad + \mathrm{o}(h^2), \end{aligned} \tag{5.20}$$

which shows that Condition 1 holds for $\alpha = 2$ and $s(w)$ equal to one 6th of the term

between square brackets. Suppressing the subscripts for convenience, one may write

$$\begin{aligned} s(w) &= \tfrac{1}{6}f(y)[\nabla_x^2 f(x,y,z) + \nabla_y^2 f(x,y,z) + \nabla_z^2 f(x,y,z)] \\ &\quad -\tfrac{1}{6}f(x,y)[\nabla_y^2 f(y,z) + \nabla_z^2 f(y,z)] \\ &\quad -\tfrac{1}{6}f(y,z)[\nabla_x^2 f(x,y) + \nabla_y^2 f(x,y)] \\ &\quad +\tfrac{1}{6}f(x,y,z)\nabla_y^2 f(y), \end{aligned}$$

where $\nabla_x^2 = \sum_{j=1}^{d_X} \frac{\partial^2}{\partial x^{j2}}$ and ∇_y^2 and ∇_z^2 are defined analogously. Upon taking expectations with respect to W one obtains the coefficient of the leading bias term $E[s(W)]$, which enters expression (5.19) for the optimal bandwidth.

Under the null hypothesis the leading bias term can be simplified by rewriting it in terms of conditional densities:

$$\begin{aligned} s(w) &= \tfrac{1}{6}f(y)f(x,z)\nabla_y^2 f(y|x,z) - \tfrac{1}{6}f(x,y)f(z)\nabla_y^2 f(y|z) \\ &\quad -\tfrac{1}{6}f(y,z)f(x)\nabla_y^2 f(y|x) + \tfrac{1}{6}f(x,y,z)\nabla_y^2 f(y) \\ &\quad +\tfrac{1}{6}f(y)f(y,z)\nabla_x^2[f(x|y,z) - f(x|y)] \\ &\quad +\tfrac{1}{6}f(x,y)f(y)\nabla_z^2[f(z|x,y) - f(z|y)]. \end{aligned}$$

The terms within square brackets are zero if the null hypothesis holds. The remaining terms can be expressed as:

$$\begin{aligned} s(w) &= \tfrac{1}{6}\nabla_y^2[f(y)f(x,y,z) - f(x,y)f(y,z)] \\ &\quad -\tfrac{1}{3}\nabla_y f(y)\cdot\nabla_y f(x,y,z) + \tfrac{1}{3}\nabla_y f(x,y)\cdot\nabla_y f(y,z), \end{aligned}$$

where ∇_y is the gradient operator, and the dot denotes the usual vector inner product. Again the term in square brackets vanishes under the null, and the remaining terms reduce to

$$s(w) = \frac{1}{3}f^2(y)\nabla_y f(x|y)\cdot\nabla_y f(z|y).$$

Finally, the following expression for $E[s(W)]$ under the null is obtained by taking expectations of this local expression with respect to the random vector W:

$$E[s(W)] = \frac{1}{3}E_Y\left[f_Y^2(Y)\nabla H_X(Y)\cdot\nabla H_Z(Y)\right].$$

5.C Dependence

According to Denker and Keller (1983), for weakly dependent data T_n is still asymptotically $N(q, \frac{\sigma^2}{n})$ distributed, provided that the covariance among the $r_0(W_i)$ is taken into account in the asymptotic variance σ^2:

$$\sigma^2 = 9\left[\mathrm{Var}(r_0(W_1)) + 2\sum_{k\geq 2}\mathrm{Cov}(r_0(W_1), r_0(W_{1+k}))\right].$$

If we estimate $r_0(W_i)$ as

$$\widehat{r}_0(W_i) = \frac{(2h)^{-d_X-2d_Y-d_Z}}{(n-1)(n-2)}\sum_{j,j\neq i}\sum_{k,k\neq i}K(W_i, W_j, W_k),$$

an autocorrelation consistent estimator for σ^2 is given by Newey and West (1987):

$$S_n^2 = \sum_{k=1}^{K} R_k\omega_k,$$

where $R_k = \frac{1}{n-k}\sum_{i=1}^{n-k}(\widehat{r}_0(W_i) - T_n)(\widehat{r}_0(W_{i+k}) - T_n)$ is the sample autocovariance function of $\widehat{r}_0(W_i)$, and ω_k a decreasing weight function as in Hiemstra and Jones (1994). It follows that

$$\sqrt{n}\frac{(T_n - q)}{S_n} \xrightarrow{d} N(0,1),$$

which proves Theorem 5.1.

Although T_n is a third order U-statistic, both T_n and the asymptotic variance S_n^2 can be determined in $\mathcal{O}(n^2)$ computational time. For each i, the calculation of $\widehat{f}_W(W_i)$ and the I_{ij}^W is $\mathcal{O}(n)$. A second $\mathcal{O}(n)$ calculation then provides $\widehat{r}_0(W_i)$ through

$$\begin{aligned}\widehat{r}_0(W_i) &= \tfrac{1}{3}\left(\widehat{f}_{X,Y,Z}(X_i,Y_i,Z_i)\widehat{f}_Y(Y_i) - \widehat{f}_{X,Y}(X_i,Y_i)\widehat{f}_{Y,Z}(Y_i,Z_i)\right)\\ &\quad +\tfrac{1}{3n}\sum_j\Big(\widehat{f}_{X,Y,Z}(X_j,Y_j,Z_j)I_{ij}^Y(2h)^{-d_Y} + I_{ij}^{XYZ}\widehat{f}_Y(Y_j)(2h)^{-d_X-d_Y-d_Z}\\ &\qquad -\widehat{f}_{X,Y}(X_j,Y_j)I_{ij}^{YZ}(2h)^{-d_Y-d_Z} - I_{ij}^{XY}\widehat{f}_{Y,Z}(Y_j,Z_j)(2h)^{-d_X-d_Y}\Big),\end{aligned}$$

a result which follows from straightforward calculation from the definition of $\widehat{r}_0(W_i)$.

Chapter 6

Estimating and Evaluating the Predictive Abilities of Semiparametric Multivariate Models

6.1 Introduction

Financial institutions face the important problem of quantifying and hedging their exposure to market risk. This risk arises from a volatile market environment, i.e. changing prices of equities, commodities, exchange rates and interest rates. A popular measure for quantifying market risk with a single number is Value at Risk (VaR). VaR is the maximum expected loss over a period of time, at a given confidence level. Statistically, VaR of a portfolio is a specific conditional quantile of the distribution function of the portfolio's value over a given future time period, see Engle and Manganelli (2004) for a review and recent methodology. In this chapter, we reduce the methodological question of the VaR estimation to the prediction of the portfolio value distribution. The latter is derived from the forecast of the multivariate asset return distribution. We specify a flexible semiparametric model for predicting the future multivariate asset return distribution in terms of recent past returns. The major question addressed here is whether the semiparametric model is able to produce a more accurate forecast than a fully parametric multivariate model.

To specify a multivariate model we use the concept of copula. Copulas can fully characterise the dependence structure of multivariate distributions. They have proven to be a useful device in many financial applications; e.g. risk management, portfolio aggregation, spillover effects (for an extensive review of applications see Bouyé, Durrleman, Nikeghbali, Riboulet, and Roncalli, 2000, Embrechts, Lindskog, and McNeil, 2003). The Basel Committee on Banking and Supervision recognised copulas as one of the important risk management techniques, as reflected in the New Basel Capital Accord (Basel II).

The copula approach allows for flexible modelling of marginal distributions and their mapping into a joint distribution through a copula function. The marginals may be specified either parametrically (e.g. Patton, 2005b) or nonparametrically (e.g. Genest, Ghoudi, and Rivest, 1995). The latter approach is commonly used in situations where the focus is on exploiting a dependence structure per se. Copulas themselves can be modelled nonparametrically via the empirical copula of Deheuvels (1979), however the parametric approach to copula functions is more common. The combination of nonparametric estimation of the marginals and a parametric copula leads to a semiparametric procedure for estimating dependence parameters. This procedure provides a good balance between fully nonparametric and fully parametric models. The former may suffer from the curse of dimensionality in multivariate settings, while the latter has to impose distributional assumptions on marginals as well as on joint distributions.

Until recently copulas have been applied to model contemporaneous dependence structure and, therefore, were of limited use in the case of multivariate processes with serial dependence. Ignoring serial dependence can lead to misspecification of marginal distributions and poor estimation of copula parameters. Stylised facts of return series such as conditional heteroskedasticity, long memory and others (see e.g. Tsay, 2002) suggest that processes with serial dependence are of special interest in financial econometrics. Fan and Chen (2002) started to fill this gap by applying copulas to modelling serial dependence in a univariate time series setting. Later, Patton (2005a) introduced the notion of conditional copula, by allowing for conditioning variables in marginal distributions. This opened the way to an even broader range of copula applications in finance, including prediction based on the recent past.

All known applications of conditional copulas, we are aware of, (e.g. Patton, 2005a, Patton, 2005b and Van den Goorbergh, 2004) use a parametric specification for their marginal distributions, conditional on past observations. To develop a flexible semiparametric approach in the context of conditional copulas, we suggest a nonparametric procedure for estimating conditional marginal distributions. The major problem with the application of this procedure to conditional copulas is the requirement of a single set of conditioning variables for all marginals. Conditioning on a large set of variables inevitably will lead to the curse of dimensionality in the context of the nonparametric estimation of conditional distribution functions. To overcome this problem, we apply a dimension reduction technique to multiple conditioning variables. After reducing the dimension of the set of conditioning variables to one, we make use of the nonparametric procedure of Hansen (2004) for estimating a smooth conditional cumulative distribution function (CDF). The parametric copula is modelled using a Gaussian specification with time-varying parameters. The semiparametric procedure may be summarised as follows:

1. Dimension reduction to one conditioning variable

2. Nonparametric estimation of univariate conditional marginal CDFs

3. Parameter estimation of a conditional copula

4. Predictive ability evaluation

Once a model is specified and estimated, the next important task is model evaluation and comparison with a benchmark model. The benchmark is specified as a parametric copula model with parametric marginals. We will adopt the recent methodology of Giacomini and White (2006) for testing conditional predictive abilities. This methodology is specifically suitable in our case since it allows for comparison of parametric and semiparametric techniques and acknowledges estimation uncertainty and possible nonstationarities in data.

The rest of the chapter is organised as follows. In Section 6.2 we provide an outline of conditional copula theory. In Section 6.3 we discuss the specification of conditional

marginals, concentrating on nonparametric estimation and the dimension reduction technique. Section 6.4 describes the estimation procedure for the copula parameters. Further, in Section 6.5 we discuss the conditional predictive ability test and its application for selecting the best performing model. Section 6.6 investigates the performance of the suggested semiparametric and parametric models under various data generating processes. In Section 6.7 we apply the described methodology to multivariate collections of major US stocks, UK stocks and international stock indexes, while Section 6.8 concludes.

6.2 Conditional copula

A general overview of copula theory is provided in Subsection 1.2.3 of Chapter 1. Here we describe an extension of the copula concept to the conditional copula.

Most variables of interest in economics and finance exhibit dependence on past observations. Therefore, it is important to condition on the lagged variables in a quantitative analysis of economic and financial data. Until recently, copulas have been used to model contemporaneous dependence structure. To extend the application of copulas to the case of conditioning variables, Patton (2005a) introduced the notion of conditional copula.

Let $F_{X,Y}(x, y)$ denote the joint CDF of a multivariate variable (X, Y), where $X = (X_1, \ldots, X_k)'$ and $Y = (Y_1, \ldots, Y_m)'$, $F_{Y|X}(y|x)$ denote the corresponding joint distribution of Y conditional on X and $F_{Y_i|X}(y_i|x)$ denote the marginal CDF of Y_i, a univariate component of Y, conditional on X.

Definition 6.1 (Conditional Copula) *The conditional copula $C_{Y|X}(u)$ of the multivariate random variable Y conditional on the multivariate random variable X, that is $Y|(X = x) \sim F_{Y|X}(y|x)$, is the conditional multivariate CDF of the marginally uniform random variable $U = (U_1, \ldots, U_m)'$, such that $U_i \equiv F_{Y_i|X}(y_i|x)$ given $X = x$.*

Sklar's theorem is then, $F_{Y|X}(y_1, \ldots, y_m|x) = C_{Y|X}(F_{Y_1|X}(y_1|x), \ldots, F_{Y_m|X}(y_m|x))$. The index $Y|X$ in the notation of copula function is dropped hereafter for brevity. It is important to clarify that conditioning enters the copula function $C(u)$ through the corresponding marginal distributions. If the dependence structure between the components

of Y given X depends on X, the latter will enter the corresponding copula function explicitly and not only through the marginals. Hereafter, by referring to U_i (or U) we will mean conditionally uniform random variable(s) without explicitly specifying the set of conditioning variables, when the latter is clear from the context. Patton (2005a) noted that the conditional copula satisfies the properties of the (unconditional) copula (see Chapter 1), Sklar's theorem (Theorem 1.2), its inverse (Corollary 1.1) and the invariance property (Theorem 1.3) hold in the case of conditioning. In this case, the (unconditional) multivariate distributions and marginal distributions are replaced by conditional multivariate distributions and conditional univariate distributions respectively. To draw an analogy with the (unconditional) copula we refer to the latter as *conditional marginal distributions* (slightly abusing terminology). An important consequence in the framework of conditional copulas is that the set of conditioning variables X is the same for every conditional marginal distribution and the conditional copula. Otherwise, Sklar's theorem may fail for conditional copulas. It may be the case that the set of conditioning variables may be reduced for some conditional marginal distributions, i.e. $F_{Y_i|X_1}(y_i|x_1) = F_{Y_i|X_1,X_2}(y_i|x_1,x_2)$ for all y_i, x_1, x_2, where X_1 and X_2 denote conditioning random variables. In this situation Y_i and X_2 are conditionally independent given X_1. This would reduce the dimensionality of the conditioning set and ease the estimation procedure.

For simplicity, we assume that the stationary multivariate time series process $\{Y_t\}$ satisfies the Markovian property. Thus, the conditioning set X_t includes only the recent past Y_{t-1}. This assumption is often made in empirical applications, however it is not crucial for our approach, and could be easily abandoned. We could also extend the conditioning set to higher lags in the past and include exogenous factors such as interest rates, growth rates, etc. For brevity we will not pursue this approach here.

6.3 Conditional marginal distributions

A joint conditional distribution is determined by a conditional copula and conditional marginals which we shall specify in this section. Conditional marginal distributions can be modelled either parametrically or nonparametrically. The former approach imposes

restrictive distributional assumptions, while the latter attempts to deduce conditional marginal distributions entirely from the data. The price to pay for relaxing the assumptions is the amount of data necessary for estimation, at a reasonable precision level, and the speed of the estimation procedure. The parametric approach has been widely used in modelling asset returns, e.g. ARCH (Engle, 1982) and GARCH (Bollerslev, 1986) models in the univariate case, and dynamic conditional correlations (DCC) of Engle (2002) in the multivariate case. Despite the variety of parametric specifications there is no general agreement on which particular model is to be chosen in practice. To avoid this ambiguity, nonparametric specifications of marginals prevailed in the (unconditional) copula approach (see e.g. Genest, Ghoudi, and Rivest, 1995, Embrechts, Lindskog, and McNeil, 2003). However, this advantage of copula modelling has not been exploited yet in the recent conditional copula approach of Patton (2005a). With this in mind, we wish to fill the gap and use a nonparametric specification for the conditional marginal distributions in the conditional copula framework.

6.3.1 Dimension reduction

The major technical difficulty in implementing a nonparametric specification of conditional marginal distributions for the conditional copula model is conditioning on a high dimensional set of variables. In this case a direct nonparametric approach is practically inapplicable mainly due to the bandwidth choice problem and the curse of dimensionality. Therefore, we suggest a technique to reduce the dimension of the conditioning set.

Below we focus on marginal distributions of the univariate random variables Y_i which are components of the multivariate random variable Y. For brevity the index i in Y_i is dropped in this section. The conditioning set is represented by the multivariate random variable X.

Hall and Yao (2005) suggest to approximate the conditional distribution function $F_{Y|X}(y|x)$ by $F_{Y|\beta' X}(y|\beta' x)$ using $\beta' x$ as a one dimensional projection of the m-dimensional conditioning variable X. The parameter vector β is estimated by minimising the mean

integrated square error (MISE) over b, i.e.

$$\widehat{\beta} = \underset{b}{\operatorname{argmin}} \int\int \left(\widehat{F}_{Y,b'X}(y, b'x) - \widehat{F}_{Y,X}(y,x) \right)^2 w(x,y)\, dxdy,$$

where $w(y,x)$ is a weighting function. The estimation procedure is based on a local linear regression, which requires bandwidth selection and is computationally intensive.

We suggest an alternative procedure that is based on an ordinary linear regression and therefore does not require bandwidth selection and is computationally fast. The suggested estimator of β enjoys the same properties as the estimator of Hall and Yao (2005), i.e. it is consistent (in a sense that the estimator converges in probability to a pseudo-true value β) and normally distributed with $\sqrt{T}$-convergence rate, where T denotes the sample size. The technique we suggest is closely related to the sufficient dimension reduction of Cook (2004).

Suppose $\{(X_t, Y_t)\}$ is a weakly dependent time series process. In order to satisfy Condition (3) of Theorem 6.1 below (subject to mild regularity conditions discussed in Hall and Li, 1993) we shall standardise X to zero mean and identity variance-covariance matrix, i.e. $\widetilde{X} = \Sigma_X^{-1/2}(X - E(X))$, where Σ_X is the variance-covariance matrix of X. This linear transformation does not affect the conditional distribution $F_{Y|X}(y|x)$. In practice we replace the true variance-covariance matrix Σ_X by its sample estimate. To ensure that the sample estimate $\widehat{\Sigma}_X$ is a positive definite matrix, which may fail due to computational issues, we employ the eigenvalue method of Rousseeuw and Molenberghs (1993). Similarly to Hall and Yao (2005) our central assumption is that the true conditional distribution function $F_{Y|X}(y|x)$ can be represented by $F_{Y|Z}(y|z)$, where $Z = \beta'\widetilde{X}$ is a one-dimensional projection of the standardised m-dimensional conditioning variable $\widetilde{X}$. It follows from the definition of the distribution function that the parameter β is identified up to a multiplicative scalar. The following theorem due to Li and Duan (1989, Theorem 2.1) establishes an estimation procedure for β.

Theorem 6.1 *Suppose that $L(y, \zeta)$ is an arbitrary function and consider the minimisation problem:*

$$(\alpha^*, \beta^*) = \underset{(a,b)}{\operatorname{argmin}}\, E_{Y,\widetilde{X}} \left[L(Y, a + b'\widetilde{X}) \right]. \tag{6.1}$$

The parameter β^* is proportional to the parameter β under the following conditions:

(1) The criterion function $L(y, \zeta)$ is convex in ζ.

(2) There is a unique solution for the minimisation problem (6.1).

(3) The conditional expectation $E\left[b'\widetilde{X}|\beta'\widetilde{X}\right]$ exists and is linear in $\beta'\widetilde{X}$.

A class of estimators defined by Eq. (6.1) includes ordinary least squares (OLS), for which Condition (1) is satisfied. The existence Condition (2) will hold for OLS if the standardised predictor $\widetilde{X}$ is well-defined (non-multicollinearity). Condition (3) is extensively discussed in Hall and Li (1993). They claim that this condition always holds for any β for spherically symmetric data. Moreover, the authors suggest that for low-dimensional projections of any high-dimensional data in many cases Condition (3) holds provided the standardised predictor $\widetilde{X}$ is well-defined. Strong consistency and asymptotic normality with $\sqrt{T}$-convergence rate of a sample estimator of β^* is established by Theorem 5.1 of Li and Duan (1989) for the case of i.i.d. data. To establish the same results for weakly dependent time series, one has to apply the central limit theorem (Doukhan and Louhichi, 1999) and the law of large numbers (Birkel, 1992) for weakly dependent processes. We are going to assume that we may reasonably summarise the conditioning information in one variable $Z = \beta'X$.

6.3.2 Smooth estimation of conditional distribution function

Hall, Wolff, and Yao (1999) present an extensive overview of recent techniques for estimating a conditional distribution function. All the methods described there are based on unsmoothed estimators. Hansen (2004) shows that it is possible (at least asymptotically) to increase the efficiency of the nonparametric procedure by using smoothed estimators. The analytical derivation there was supported by simulations for various data generating processes. An analogous conclusion has been drawn by Chen, Fan, and Tsyrennikov (2004) in the context of unconditional nonparametric marginals. Considering the possible efficiency gains we will follow the smooth method of Hansen (2004). We outline this method and refer the reader to the work of Hansen (2004) and the references therein for a more detailed exposition.

Suppose that $\{(Y_{i,t}, Z_t)\}$ is a strictly stationary time series process with CDF $F_{Y_i,Z}(y_i, z)$. The process $\{Y_{i,t}\}$ is one of the elements of the m-variate time series process $\{Y_t\}$, the conditional marginal distribution $U_i \equiv F_{Y_i|Z}(y_i|z)$ of which we are after. The one-dimensional variable Z summarises the information about Y_{t-1}. For brevity, we drop the subscripts of F and the index i, i.e. by writing $F(y|z)$ in this section we refer to the conditional CDF of the univariate random variable Y_i given Z.

In the context of (non-smooth) conditional distribution function estimation, a local linear estimator may achieve lower bias in comparison with the Nadaraya-Watson estimator. The same is true in the context of smooth estimation. Therefore, the smooth estimator will be of the form of a modified local linear estimator ensuring that it remains a valid conditional distribution function:

$$\widehat{U} \equiv \widehat{F}_{h,r}(y|z) \;=\; \frac{\sum_{t=1}^{T} w_t^* K_h(y - Y_t)}{\sum_{t=1}^{T} w_t^*}, \tag{6.2}$$

where $K_h(s) = K(\frac{s}{h})$ and $K(s) = \int_{-\infty}^{s} \kappa(u)du$ is an integrated smooth kernel with bandwidth parameter h, $\kappa(u)$ is a symmetric second-order kernel which is normalised to unit variance and has a finite sixth moment (here we use the Gaussian kernel, which is a standard normal probability distribution function), and w_t^* denotes modified weights from local linear estimation, i.e.

$$w_t^* \;=\; \begin{cases} 0, & \text{for } \widehat{\gamma}(z - Z_t) > 1, \\ w_t(1 - \widehat{\gamma}(z - Z_t)), & \text{for } \widehat{\gamma}(z - Z_t) \leq 1, \end{cases}$$

where $w_t = \frac{1}{r} w(\frac{z - Z_t}{r})$ are kernel weights with kernel w (the Gaussian kernel here) with the same conditions as on κ and bandwidth r and $\widehat{\gamma}$ is the slope from the weighted least squares, that is

$$\widehat{\gamma} \;=\; \left(\sum_{t=1}^{T} w_t (z - Z_t)^2 \right)^{-1} \left(\sum_{s=1}^{T} w_s (z - Z_s) \right).$$

Nonnegative weights w_t^* ensure that $\widehat{F}$ is a proper distribution function.

As in any kernel smoothing problem the bandwidth choice is the crucial part of the analysis. In the present situation the bandwidths (h^*, r^*) are selected by a plug-in method. The plug-in bandwidths minimise the estimated asymptotic mean integrated square error (MISE). The following theorem of Hansen (2004, Theorem 1) defines the MISE for the smooth local linear estimator.

Theorem 6.2 *Assume that $F(y|z)$ and the marginal density of Y, $f(y)$, are continuously differentiable up to fourth order in both y and z. If $r = cT^{-1/5}$ and $h = O(r)$ as $T \to \infty$, then*

$$\int_{-\infty}^{\infty} E\left(\widehat{F}_{h,r}(y|z) - F(y|z)\right)^2 dy = \frac{R}{f(y)Tr}(V - h\psi) + \frac{r^4 V_1}{4} - \frac{h^2 r^2 V_2}{2} + \frac{h^4 V_3}{4} + O(T^{-6/5}), \tag{6.3}$$

where $R = \int_{-\infty}^{\infty}(w(s))^2 ds$ denotes the roughness of the weighting kernel w and

$$\begin{aligned}
\psi &= 2\int_{-\infty}^{\infty} sK(s)\kappa(s)ds > 0 \\
V &= \int_{-\infty}^{\infty} F(y|z)(1 - F(y|z))dy \\
V_1 &= \int_{-\infty}^{\infty} (F^{(2)}(y|z))^2 dy \\
V_2 &= \int_{-\infty}^{\infty} f(y|z)f^{(2)}(y|z)dy \\
V_3 &= \int_{-\infty}^{\infty} (f^{'}(y|z))dy.
\end{aligned}$$

The notation $f^{(2)}(y|z)$ (or $F^{(2)}(y|z)$) refers to the second-order partial derivative of $f(y|z)$ (or $F(y|z)$) with respect to y. The parameters of the MISE equation (Eq. 6.3), $f(y), V, V_1, V_2, V_3$ are estimated using kernel smoothing and local polynomial regression techniques (see Hansen, 2004 for details). Plug-in bandwidths $(\widehat{h}^*, \widehat{r}^*)$ are found by numerical minimisation of the estimated MISE.

It can be seen that estimators of the MISE parameters and, thus, bandwidths $(\widehat{h}^*, \widehat{r}^*)$

depend on the conditioning variable Z. Since the bandwidth selection procedure is computationally involved, selection of bandwidths $(\widehat{h}^*, \widehat{r}^*)$ for every Z_t appears to be an infeasible task for reasonable sample sizes ($T \geq 500$). To reduce the number of computations, the conditioning observations Z_t were grouped according to their ranks into I groups G_i:

$$G_i = \left\{ Z_t : \ \Phi\left(2.5\left[-1 + \frac{2i-2}{I}\right]\right) < \frac{\text{rank}(Z_t)}{T} \leq \Phi\left(2.5\left[-1 + \frac{2i}{I}\right]\right) \right\},$$

where Φ denotes standard normal CDF. The plug-in bandwidth for each group $(\widehat{h}_i^*, \widehat{r}_i^*)$ is determined on the basis of observation Z_t with $\text{rank}(Z_t)/T = \Phi(2.5[-1 + (2i-1)/I])$. Intuitively, the proposed grouping puts observations Z_t with similar values into the same group. Our simulations showed that for sample size $T = 1000$, $I = 13$ provides a good approximation, i.e. the corresponding results are hardly distinguishable from the results of the full procedure, when the bandwidths are computed individually for every observation.

We follow the above procedure and produce the CDF-transform of each realisation of the multivariate time series process $\{Y_t\}$ using the estimated conditional CDF, i.e. $\widehat{U}_{t,i} = \widehat{F}_{Y|Z}(\widehat{Y}_{t,i}, \widehat{Z}_t)$ and join them in the series of CDF-transforms $\widehat{U}_t$.

6.4 Specification and estimation of copula

Similarly to the case of marginals, it is possible to specify a copula either nonparametrically or parametrically. A nonparametric approach will lead to the empirical copula of Deheuvels (1979). It is more flexible than the parametric copula, but the estimation procedure is more involved. For simplicity, here we will pursue the parametric copula approach.

Many copula specifications have been suggested for the bivariate case (see Nelsen, 1999 for extensive review). However, for most of the copulas, extensions to higher dimensions are not feasible. Therefore, most of the copulas considered for multivariate modelling are based on the inverses of the Gaussian distribution or, more general, the Student t distribution. The former will yield the Gaussian copula where the correlation matrix Σ

acts as a matrix of parameters, while the latter will result in the Student t copula with correlation matrix Σ and an additional parameter, the number of degrees of freedom ν. Several studies (e.g. Chen, Fan, and Patton, 2004 and Chapter 3) reported a poor fit of the Gaussian copula with constant correlations to the multivariate asset returns data, while the Student t copula provided a slightly better fit to the data in the unconditional copula framework. In the recent past Engle and Sheppard (2001) and Engle (2002) suggested dynamic conditional correlations (DCC) models for a parsimonious specification of the time-varying correlation matrix Σ for multivariate GARCH models. Chen, Fan, and Patton (2004) showed that the Gaussian DCC model provided a good fit to the data in the context of the unconditional copula. Relying on these results, we suggest using the DCC model to specify the time evolution of correlation matrix Σ. It is important to notice that there are other, possibly better, ways to model the time evolution of Σ. Recent work of Van der Weide (2002), Pelletier (2005) and Hafner, Van Dijk, and Franses (2005) suggest new promising nonlinear and semiparametric techniques for modelling the correlation dynamics. We adopt the DCC of Engle (2002) for simplicity and wider acknowledgement in empirical research.

The Gaussian DCC (1,1) copula model is defined as follows. Using the inversion method (Corollary 1.1 in Chapter 1), we derive the Gaussian copula

$$C_{\Sigma}(u_1, \dots u_m) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_m))), \tag{6.4}$$

where Φ_{Σ} denotes the multivariate Gaussian CDF with linear correlation matrix Σ and $\Phi^{-1}(\cdot)$ is the inverse of the univariate standard normal CDF. The time-varying conditional correlation matrix evolves according to the following specification:

$$\Sigma_t = (1 - \varphi_1 - \varphi_2)\overline{\Sigma} + \alpha\Phi_m^{-1}(u_{t-1})(\Phi_m^{-1}(u_{t-1}))' + \varphi_2\Sigma_{t-1}, \tag{6.5}$$

where $\overline{\Sigma}$ is the unconditional correlation matrix and $\Phi_m^{-1}(u_t) = (\Phi^{-1}(u_{t1}), \dots, \Phi^{-1}(u_{tm}))'$. To ensure positive definiteness of Σ_t, we require $0 \leq \varphi_1 \leq 1$, $0 \leq \varphi_2 \leq 1$ and $\varphi_1 + \varphi_2 \leq 1$. The correlation matrix $\overline{\Sigma}$ is computed by compounding pairwise correlation coefficients $\rho_{ij} = \text{Corr}[\Phi^{-1}(U_i), \Phi^{-1}(U_j)]$. It may happen due to numerical problems that $\overline{\Sigma}$ fails to

be positive definite. To avoid this situation, we use the eigenvalue method of Rousseeuw and Molenberghs (1993). We estimate the parameters of the Gaussian DCC copula model using the semiparametric maximum likelihood method introduced by Genest, Ghoudi, and Rivest (1995). The log likelihood function for the conditional copula and marginals can be derived from the decomposition of the multivariate conditional density (see Eq. (1.10) in Chapter 1 for the unconditional case):

$$L(y|z) = \log c(F_1(y_1|z), \ldots, F_m(x_m|z)) + \sum_{i=1}^{m} \log f_m(x_m|z) = L_c(u) + \sum_{i=1}^{m} L_i(x_i|z). \quad (6.6)$$

The estimation procedure consists of two stages. First we perform a conditional marginal CDF transformation and construct the marginally uniform series of transformations $\widehat{U}_t$. In the next stage, we estimate the parameters of the copula by maximising the log likelihood function of the copula using the marginal transforms obtained from the first stage:

$$(\widehat{\varphi_1}, \widehat{\varphi_2}) = \underset{(\phi_1,\phi_2)}{\operatorname{argmax}} \sum_{t=1}^{T} L_c(\widehat{U}_{t1}, \ldots, \widehat{U}_{tm}, (\phi_1, \phi_2)). \quad (6.7)$$

The following theorem, due to Genest, Ghoudi, and Rivest (1995, Proposition 2.1), establishes desirable properties of the semiparametric estimator $\widehat{\varphi}$.

Theorem 6.3 *Under suitable regularity condition, the semiparametric estimator ($\widehat{\varphi}$ is $\sqrt{T}$-consistent and asymptotically normal.*

The Gaussian copula density is obtained by taking partial derivatives in Eq. (6.4):

$$c(u) \;=\; |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\Phi^{-1}(u))'(\Sigma^{-1} - \mathbb{I})\Phi^{-1}(u)\right),$$

where $\mathbb{I}$ is the identity matrix. Using the time-varying conditional correlation Σ_t, we may specify the log likelihood function:

$$L = -\frac{1}{2}\left(\sum_{t=1}^{T}\log|\Sigma_t| + (\Phi^{-1}(U_t))'(\Sigma_t^{-1} - \mathbb{I})\Phi^{-1}(U_t)\right).$$

Since there is no closed form expression for the derivatives of the log-likelihood L, the optimisation is conducted numerically. The optimisation is performed using the simplex method of Nelder and Mead (1965). This method is suitable for our purposes since it does not require evaluation of the derivatives.

6.5 Predictive abilities of candidate models

The major question of this paper is whether using the conditional semiparametric approach, which is based on a parametric copula and nonparametric conditional marginals, can improve the accuracy of the forecast distribution of future returns. Therefore, we are interested in comparing the semiparametric candidate model specified in Section 6.3 (nonparametric conditional marginals) and in Section 6.4 (parametric copula) with a similar parametric benchmark, i.e. a parametric copula model with parametric marginals.

6.5.1 The benchmark model

We use a slight modification of the commonly used DCC model as a benchmark. This model can be viewed as a parametric copula model with conditional marginals specified parametrically according to the GARCH(1,1) model:

$$\begin{aligned} Y_t|(Z_t, h_t) &\sim N(\phi Z_t, h_t), \\ h_t &= \gamma_0 + \gamma_1 Y_{t-1}^2 + \gamma_2 h_{t-1}. \end{aligned} \tag{6.8}$$

The multivariate variable Z_t compounds all conditioning variables of the copula. Here, we assume that Z_t enters only the mean and does not affect the variance in Eq. (6.8). Then, the copula takes the form of the Gaussian DCC copula described in the previous section, see Eqs. (6.4, 6.5). The model is estimated using a two stage maximum likelihood: first the GARCH(1,1) models for marginals and then, the Gaussian DCC copula model.

For the properties of the estimators derived from this two-stage procedure we refer the reader to Engle and Sheppard (2001). The optimisation is performed numerically using the Nelder and Mead (1965) algorithm.

6.5.2 Conditional test for predictive abilities

Once a benchmark model and a candidate model are specified and estimated, we wish to determine whether the latter is superior to the former. The accuracy of the out-of-sample prediction of the distribution of future returns is used as a criterion. For brevity, we will focus on a one-step-ahead forecast at time $t+1$, though an extension is possible. The major problem in testing density forecasts arises from the following fact. While producing a forecast of the whole future distribution, we observe only one (possibly multivariate) outcome from the true distribution. Various techniques have been suggested in the forecast evaluation literature to overcome this problem (see Diebold, Gunther, and Tay, 1998 for review). We will focus on the so-called predictive log likelihood or scoring rule (Diebold and Lopez, 1996) defined as

$$L_{t+1}(\widehat{f}_t, Y_{t+1}) \;=\; \log \widehat{f}_t(Y_{t+1}), \tag{6.9}$$

where $\widehat{f}_t(\cdot)$ is the density forecast evaluated at point Y_{t+1}. In the case of the copula, we observe the decomposition of L into a copula part L_c and the sum of marginal log likelihoods L_n according to Eq. (1.10). While the former is a part of the estimation process, the latter is readily available by differentiating the smooth conditional distribution function (6.2) at the point Y_{t+1}. Similarly to the idea of maximum likelihood, the model producing the best density forecast is expected to return the highest average score. Predictive log likelihood is especially convenient in the case of multivariate forecasting.

Further, we adopt the recent conditional predictive ability test of Giacomini and White (2006). Their test is based on the out-of-sample evaluation using a rolling window scheme. The in-sample size T used for estimation remains constant, while the sample itself and the points at which the forecast is evaluated, move with time. It is assumed that the number of out-of-sample forecasts N tends to infinity while the in-sample size T remains finite.

The advantage of the conditional test vs. the traditional unconditional methodology of West (1996) is that it can be applied in a more general setting. The test evaluates not only the model itself, but the whole forecasting method, which includes the choice of the in-sample size T. The conditional methodology can be applied for the comparison of a wide range of models, such as parametric, semiparametric, nonparametric and Bayesian models. Non-nested and nested models can be treated in a similar way, while the test of West (1996) is limited to the case of non-nested models. The conditional methodology of Giacomini and White (2006) is specifically applicable for our situation since we wish to compare semiparametric and parametric models. In theory we would expect that the semiparametric procedure will perform at least as good as the parametric one, provided (a) the DGP is stationary and (b) the sample size T is infinite. However, in practice for financial time series achieving both (a) and (b) is not realistic, and we have to find the sample interval of the size T, on which the sample is approximately stationary and T is large enough to ensure precision of the estimates. Although being asymptotically superior, a semiparametric model can actually be outperformed by a parametric model for a fixed T, since the latter may require less data to achieve the same precision as the former.

Applying the conditional testing methodology Amisano and Giacomini (2006) suggest a weighted likelihood ratio test. Their weighting scheme is an attempt to concentrate the test on the region of the distribution one is specifically interested in. In the context of VaR estimation, one would focus on the accuracy of the forecast of the lower tail of the return distribution. They define the weighted likelihood ratio as

$$R_{t+1} = w(Y_{t+1})(L_{t+1}(\widehat{f}_t^{cd}, Y_{t+1}) - L_{t+1}(\widehat{f}_t^{bm}, Y_{t+1})),$$

where $\widehat{f}_t^{cd}$ and $\widehat{f}_t^{bm}$ are predictive log likelihoods (Eq. 6.9) of the candidate model and the benchmark respectively, Y_{t+1} is an out-of-sample observation, and $w(Y_{t+1})$ is a weighting function.

The null hypothesis is stated in terms of the expectation of the weighted likelihood ratio R over the number of evaluations N using the rolling forecast scheme:

$$H_0 \quad : \qquad E[R_{t+1}] \leq 0,$$
$$H_A \quad : \qquad E[R_{t+1}] > 0.$$

The null hypothesis states that on average the forecast produced by the candidate method is not better than the forecast produced by the benchmark. The test statistic S takes the form:

$$S = \frac{\frac{1}{N}\sum_{t=T}^{T+N} R_{t+1}}{\widehat{\sigma}_N/\sqrt{N}},$$

where $\widehat{\sigma}_N^2$ is a heteroskedasticity and autocorrelation consistent Newey and West (1987) estimator of the asymptotic variance:

$$\widehat{\sigma}_N^2 = \frac{1}{N}\sum_{t=T}^{T+N} R_{t+1}^2 + 2\left[\sum_{j=1}^{p_N}\left(1 - \frac{j}{p_N+1}\right)\frac{1}{N-j}\sum_{t=T+j}^{T+N} R_{t+1}R_{t+1-j}\right], \tag{6.10}$$

where p_N is the number of nonzero autocorrelations taken into account, which is a parameter of choice. It is required that $p_N = o(N^{1/4})$ for the consistency of the estimator (Eq. 6.10). As a rule of thumb, we choose $p_N = \lfloor N^{1/5} \rfloor$, i.e. the largest integer smaller than or equal to $N^{1/5}$.

Amisano and Giacomini (2006) show that under the null test statistic S converges in distribution to $N(0,1)$ for a fixed estimation window as the number of evaluations $N \to \infty$, subject to mild regularity conditions.

Despite the general attractiveness of the weighting scheme, one should be cautious about applying it in practice. In some cases, the test based on the weighting scheme may lead to spurious results. Consider a simple example to illustrate this point. Assume that the true data are generated from a Gaussian distribution. Since the true distribution is unknown in practice, we can try various distributions to fit the data. Suppose the choice lies between a Gaussian and Student t distribution. We estimate the parameters of the corresponding distributions, using e.g. maximum likelihood, and apply the weighted likelihood ratio test. We use the Student t distribution as the alternative, and the Gaussian

as the benchmark. Putting higher weights on tails or on the centre of the distribution, may cause the test to overreject the null hypothesis of no better predictive performance. The reason is that the Student t distribution has a higher peak and fatter tail density than the Gaussian, naturally leading to higher scores in these regions. Our simulations confirmed these concerns. For data originating from the Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1.5$, the test with the centre weight $w = \phi(x/\sigma)$ always rejects while the null hypothesis is true (1,000 simulations). In the case of equal weighting along the whole density function, i.e. $w = 1$, the rejection rate of the test remained close to the nominal size. A detailed study of the non-equal density weighting is left for future research. The conclusion we draw at this moment is that no weighting should be applied while using this test. We proceed by putting $w = 1$ in the conditional likelihood test.

6.6 Simulations

We next investigate the performance of the proposed semiparametric model (see Section 6.3 for the specification of nonparametric conditional marginals and Section 6.4 for the parametric copula) vs. the parametric DCC model (Subsection 6.5.1) using various data generating processes (DGPs). Since these two models differ in the specification of the marginals, we will focus on the multivariate DGPs with the identical correlation structure, but various specifications for the marginals. The correlation matrix is specified as in Eq. (6.5) with $\alpha = 0.1$ and $\beta = 0.7$. First, we sample uniform random variates $u_{t,i}$ from the Gaussian copula (Eq. 6.4) and, using the corresponding inverse CDF, transform them into innovations $\{\varepsilon_{t,i}\}$ which follow the standard normal distribution, the Student t distribution with three degrees of freedom rescaled to unit variance and the double exponential distribution rescaled to unit variance. For brevity, the same specification is applied for each marginal. Then, we use the following DGPs for specifying conditional variance $h_{t,i}$ of the random variables $Y_{t,i} = \sqrt{h_{t,i}}\varepsilon_{t,i}$, indices $(i, j, k) \in \{1, 2, 3\}$ and $i \neq j \neq k \neq i$:

DGP 0. $h_{t,i} = 0.05 + 0.1Y_{t-1,i}^2 + 0.75h_{t-1,i}$

DGP 1. $h_{t,i} = 0.05 + 0.4Y_{t-1,i}^2 + 0.1Y_{t-1,j}^2 + 0.1Y_{t-1,k}^2$

ε-dist.\ DGP	0. GARCH		1. ARCH		2. GARCH		3. TARCH		4. EGARCH	
	d.r.	ign.	d.r.	ign.	d.r.	ign.	d.r.	ign.	d.r.	ign.
st. normal	.041	.035	.055	.043	.036	.041	.710	.652	.642	.571
Student t	.621	.642	.651	.662	.631	.635	.842	.856	.789	.736
doub. exp.	.384	.390	.561	.559	.381	.380	.752	.756	.697	.678

Table 6.1: *Rejection rates of the predictive abilities test under various innovation distributions and DGPs for conditional variance. In column "d.r." results when the dimension reduction is used, in column "ign." results when the Granger causality is ignored. Nominal size* $\alpha = .05$*, dimension* $m = 3$*, in-sample size* $T = 1,000$*, out-of-sample number of evaluations* $N = 3,000$*, number of simulations* $1,000$.

DGP 2. $h_{t,i} = 0.05 + 0.1Y_{t-1,i}^2 + 0.75h_{t-1,i} + 0.05Y_{t-1,j}^2 + 0.05Y_{t-1,k}^2$

DGP 3. $h_{t,i} = 0.05 + 0.05Y_{t-1,i}^2 + 0.075Y_{t-1,i}^2\mathbb{I}(Y_{t-1,i} < 0) + 0.75h_{t-1,i} + 0.05Y_{t-1,j}^2 + 0.05Y_{t-1,k}^2$

DGP 4. $\ln(h_{t,i}) = -1.00 + 0.1|Y_{t-1,i}| - 0.05Y_{t-1,i} + 0.75\ln(h_{t-1,i}) + 0.05|Y_{t-1,j}| + 0.05|Y_{t-1,k}|$.

The considered DGPs are modifications of models commonly applied in financial econometrics, i.e. DGP 1 corresponds to the ARCH(1) model of Engle (1982), DGP 0 and DGP 2 to its generalisation, the GARCH(1,1) of Bollerslev (1986), DGP 3 to the threshold TARCH(1,1) model of Zakoian (1994) and DGP 4 to the exponential EGARCH(1,0) model of Nelson (1991). DGP 1 and DGP 2 are symmetric in lagged returns, while DGP 3 and DGP 4 allow for asymmetries. The last two terms in each of the DGPs (1–4) introduce nonlinear Granger causality into the corresponding model by making the conditional variance for the process $\{Y_{t,i}\}$ dependent on the lagged values of $\{Y_{t,j}\}$ and $\{Y_{t,k}\}$. We set the dimension of the series to $m = 3$, in-sample size $T = 1,000$ and number of out-of-sample evaluations $N = 3,000$.

In Table 6.1 we report the rejection rates of the test for predictive abilities of the semiparametric model against the parametric benchmark. In column "d.r." we show the results of the test when the dimension reduction procedure is used for the semiparametric estimation, while in column "ign." the results for the estimation procedure when the Granger causality is ignored and for estimating each process $\{Y_{t,i}\}$ we condition only on its own past. For DGP 0 (GARCH(1,1) without Granger causality terms) under

the standard normal distribution of the innovations $\{\varepsilon_{t,i}\}$, the functional form and the distribution of the marginals are equivalent to the parametric DCC model, that is, the benchmark model is correct. Under these conditions the semiparametric model is expected to show no better performance than the parametric DCC benchmark, i.e. the null hypothesis is true. The rejection rates are slightly lower than the nominal size $\alpha = 0.05$ in this case, which means that the test is conservative. For the ARCH and GARCH specifications with the Granger causality terms (DGPs 1 and 2, respectively) under the standard normal distribution of the innovations $\{\varepsilon_{t,i}\}$, the functional form and the distribution of the marginals are not equivalent to the parametric DCC model due to the Granger causality terms. The semiparametric model model is expected to perform better than the parametric benchmark when Granger causality is not ignored ("d.r."-column). Howerer, we observe that the rejection rates in these cases are not always higher than those under the null. The null hypothesis of no better performance of the semiparametric model would ideally be rejected in all other cases, where the underlying distribution of the innovations and/or the functional form of the conditional variance is different from the considered parametric DCC specification. The rejection rates (power) are the highest for Student t innovations, while for the innovations distributed according to the double exponential distribution the rejections rates are lower. As anticipated the lowest rejection rates are observed for the standard normal distribution, since the DCC model uses the latter as the underlying distribution for the innovations. It is important to note that the innovations of the DCC model could be adapted. However, since the true underlying process is not known in practice, we kept the innovations process fixed to the one most often used in practice. With respect to the functional form of the conditional variance, the highest rejection rates are observed for the TARCH process (DGP 3) followed by the EGARCH specification (DGP 4). This is due to the fact that the DCC benchmark uses the symmetric GARCH(1,1) specification for the conditional variance , while the above processes (DGPs 3 and 4) are asymmetric. Comparing the power of the test for different conditioning procedures, i.e. using the dimension reduction ("d.r."-column) and conditioning on the own past ("ign."-column), it is difficult to draw a definitive conclusion. In most of the cases the differences in rejection rates are not substantial, only for

the EGARCH model (DGP 4) the power is slightly higher for the procedure based on the dimension reduction. The major reason for this fact is that the dimension reduction procedure employed assumes linear approximation, i.e. $F_{Y|X}(y|x)$ is approximated by $F_{Y|\beta' X}(y|\beta' x)$. At the same time, the DPGs used, being closely related to empirical financial process, i.e. contain nonlinear structure in the conditional dependence. We leave the improvement in the dimension reduction methodology for future research.

6.7 Application to multivariate asset returns

Next, we evaluate the predictive ability of the semiparametric model on daily log-return series of three asset collections. Log-returns are defined as $X_t = \ln(P_t/P_{t-1})$, where P_t is the dividend-adjusted price on day t. The full sample covers the period 08/1990–07/2005. For the sake of robustness a few returns that were extremely abnormal (about 30 observations in total) were eliminated from the sample. The first collection consists of five major US traded stocks from different sectors (IBM, Citigroup, Wal-Mart, Chevron and Verizon); the second is comprised of five major UK traded stocks (Barclays, BP, British Airways, Reuters and Vodaphone). For these collections we use closing prices P_t to compute log returns. The third collection consists of four international indexes (S&P 500, FTSE 100, DAX and CAC 40) measured simultaneously at 16:00 hours London time. This helps to avoid the problem of non-synchronous returns (Martens and Poon, 2001) for the third collection. The in-sample size is set to $T = 1,000$. A smaller sample size is generally not feasible either for the parametric or for the semiparametric model, while a larger sample size might provide worse forecasts due to nonstationarities. Given the available data-set and in-sample size T, we set the number of out-of-sample evaluations to $M = 2,900$. The dimension reduction procedure discussed in Subsection 6.3.1 was embedded into the procedure. Table 6.2 reports the values of the test statistic S and the corresponding p-values.

The test fails to reject the null of no better predictive performance of the semiparametric model compared to the parametric DCC model for the collections of US and UK stocks, while the test on the collection of four international indexes clearly rejects the

Collection	5 US stocks	5 UK stocks	4 Int Indexes
S	-0.45	-0.93	2.29
p-value	(0.673)	(0.824)	(0.011)

Table 6.2: *Test statistic S and the corresponding p-values of the predictive abilities test on asset return data.*

null. A possible explanation is that the GARCH(1,1) model can adequately explain the dynamics of the considered individual US and UK stocks, but is, however, not fully applicable to the dynamics of international stock indexes. Though, more evidence is needed to draw stronger conclusions about the applicability of semiparametric models, we may expect them to predict better in settings, where a collection includes a range of various assets with qualitatively different dynamics, e.g. stocks, bonds, foreign currencies, stock indexes.

6.8 Conclusions

In line with current developments in quantitative risk management we suggested a flexible multivariate semiparametric model to forecast future asset returns. The procedure heavily relies on the concept of conditional copula, which was shortly reviewed. Motivated by the fact that until now conditional copulas had been used in a fully parametric setting, our aim was to construct a semiparametric copula model with nonparametric marginals. We used conditional kernel smoothers based on local linear estimators for conditional marginal CDFs. The major difficulty in using a conditional copula is the requirement of the same high-dimensional conditioning set X for each marginal of Y. This would inevitably lead to the curse of dimensionality in the estimation of the conditional marginals. To avoid this we suggested a dimension reduction technique. For simplicity, we applied the Gaussian copula with a time-varying correlation matrix specified as in the DCC model. The copula specifications may be extended to include other multivariate copulas and richer correlation dynamics. Moreover, a nonparametric copula can be used as well.

To assess the predictive abilities of the semiparametric model we confronted this model with the widely used fully parametric DCC model. We used the conditional predictive abilities framework in conjunction with a predictive likelihood as a loss function. Simu-

lations on various DGPs with different underlying distributions suggested that the test for predictive abilities is able to detect the better predictive performance of the semiparametric model, when the marginals are misspecified relatively to the parametric model. In an application to three asset collections, we found that the suggested semiparametric model outperformed the fully parametric DCC model in the case of the collection of four international stock indexes and did not show evidence of better performance in the cases of the collection of five US stocks and the collection of five UK stocks. Jointly with the simulations, this shed light on the setting where the semiparametric model is expected to forecast better.

Further improvements in nonparametric marginals estimation along with a more flexible model for the conditional copula are necessary for practical applicability of the method. This is left for future research. Moreover, it would be interesting to extend the conditioning set by including important macroeconomic variables, e.g. interest rates, growth rates, etc. The efficiency of the dimension reduction procedure could be improved by incorporating the information on the functional form of conditional dependence, e.g. conditional heteroskedasticity. To allow for more flexibility in modelling correlation dynamics in the copula function while maintaining parsimony, we could change the DCC model for more recent specifications (e.g. Pelletier, 2005 or Hafner, Van Dijk, and Franses, 2005).

Chapter 7

Summary

In this thesis we investigate a number of issues related to nonparametric and semiparametric methods in financial econometrics. Compared to parametric methods, nonparametric and semiparametric methods require fewer assumptions on distributions and/or functional forms and allow for more flexibility in financial modelling and prediction. On the other hand, nonparametric estimators require more observations since their asymptotic convergence rates are typically slower than those of parametric ones. In a multivariate setting this may lead to the curse of dimensionality. That is why often a semiparametric model is used to combine both approaches. In most chapters of this thesis we pursue the nonparametric approach. The last chapter, which deals with multivariate time series, employs a semiparametric approach. The thesis is divided into three major themes: dependence (Chapters 2 and 3), causality (Chapters 4 and 5) and prediction (Chapter 6).

In an introductory Chapter 1 we review the major theoretical notions used throughout this thesis, in particular the notion of weakly dependent processes and mixing, U- and V-statistics and the copula concept. Discussing various notions of weak dependence, we provide a classification of strong mixing conditions and explain their role in time series analysis. Next, we shortly describe U- and V-statistics, important classes of estimators, which allow for developing asymptotic theory in a nonparametric setting. In this chapter we also discuss copulas, which are functions fully characterising the dependence between multivariate random variables. Copulas allow for decomposing the multivariate modelling problems into univariate components and dependence structure.

In Chapter 2 we propose new tests for serial independence that are based on quadratic forms, and confront them with related commonly used nonparametric tests. To obtain exact level tests we implement a Monte Carlo procedure using permutations of the original observations. To address the bandwidth selection problem, we introduce a novel multiple bandwidth procedure based on a number of different bandwidth values. Numerical simulations show that the tests perform well compared to existing nonparametric tests. However, we find that the permutation test is conservative when applied to residuals of estimated time series models. We overcome this problem by using a parametric bootstrap. The practical application of the new tests is illustrated using financial returns data. In the future we plan to investigate analytically the sensitivity of the test to residuals and generalise the notion of quadratic forms to the context of one- and two-sample testing.

In Chapter 3 we use quadratic forms to develop a goodness-of-fit test for copulas. The distance between the observed and the hypothetical copula is measured with the help of kernel-based quadratic forms. In this way, we avoid using plug-in density estimators which do not guaranty optimal testing and may suffer the curse of dimensionality. The tests based on quadrartic forms show good finite sample properties (at least for low dimensional datasets) in comparison to other goodness-of-fit tests. We apply the tests to datasets of US large cap stocks and DAX constituencies and reveal that the Gaussian copula is statistically inadequate to characterise the dependence between asset returns. In future research we hope to improve the performance of the test for higher dimensions.

Chapter 4 addresses a consistency problem in the commonly used nonparametric test for Granger causality developed by Hiemstra and Jones (1994). Intuitively, X is a Granger cause of Y if adding past observations of X to the information set increases the knowledge on the distribution of current values of Y. We demonstrate analytically that the relationship tested in Hiemstra and Jones (1994) is not implied by the null hypothesis of Granger non-causality. Simulations based on processes satisfying the null hypothesis show that the test heavily over-rejects and, hence, may lead to spurious test results.

In Chapter 5 we study the reasons behind the observed over-rejection analytically. We find that the Hiemstra-Jones test for the null of Granger non-causality is sensitive to conditional heteroskedasticity, which is usually present in financial time series. We

suggest a new statistic, which replaces the global test statistic by an average of local conditional dependence measures. Moreover, we derive appropriate rates at which the bandwidth tends to zero to overcome bias problems in the test. We apply the new test to historical returns and trading volumes of the S&P index and find less evidence for volume Granger-causing returns than suggested by the Hiemstra-Jones test.

In Chapter 6 we suggest a new semiparametric procedure for estimating multivariate autoregressive models. The conditional copula, which summarises cross-sectional dependence between the elements of the time series, is modelled parametrically, while conditional marginal distributions are estimated nonparametrically. We propose a dimension reduction technique to avoid the curse of dimensionality in the estimation of the latter. We confront the semiparametric copula model with the commonly used fully parametric DCC model and compare the predictive abilities of these models. It turns out that semiparametric models are able to produce statistically better one-step-ahead forecasts of returns of a collection based on four international indices. For the collections based on US and UK stocks the DCC model was performing better. We suggest a number of future improvements in the semiparametric procedure to increase its viability.

In this thesis we demonstrate that there is a wide range of applicability of nonparametric and semiparametric techniques in financial econometrics. Given the fast developments in computer technology and improvements in simulation algorithms, we believe that these techniques will receive an even greater recognition in the future. One of the major problems to be solved in the nonparameteric analysis of multivariate data is the curse of dimensionality. In the future, we hope to contribute to the solution of this problem by using dimension reduction techniques based on quadratic forms.

Bibliography

Abhyankar, A. (1998). Linear and non-linear Granger causality: Evidence from the U.K. stock index futures market. *Journal of Futures Markets* **18**, 519–540.

Ahmad, I. A. and Li, Q. (1997). Testing independence by nonparametric kernel method. *Statistics & Probability Letters* **34**, 201–210.

Amisano, G. and Giacomini, R. (2006). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*. Forthcoming.

Anderson, N. H., Hall, P. and Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis* **50**, 41–54.

Asimakopoulos, I., Ayling, D. and Mahmood, W. M. (2000). Non-linear Granger causality in the currency futures returns. *Economics Letters* **68**, 25–30.

Baek, E. and Brock, W. (1992). A general test for Granger causality: bivariate model. Working paper, Iowa State University and University of Wisconsin, Madison.

Bell, D., Kay, J. and Malley, J. (1996). A non-parametric approach to non-linear causality testing. *Economics Letters* **51**, 7–18.

Berg, D. and Bakken, H. (2005). A goodness-of-fit test for copulae based on the probability integral transform. Mimeo, The Norwegian Computing Center and University of Oslo, Oslo.

Berg, D. and Bakken, H. (2006). Copula goodness-of-fit tests: a comparative study. Mimeo, The Norwegian Computing Center and University of Oslo, Oslo.

Birkel, T. (1992). Laws of large numbers under dependence assumptions. *Statistics and Probability Letters* **14**, 355–362.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994). ARCH models. In *Handbook of Econometrics*, R. F. Engle and D. McFadden, eds., vol. 4, chap. 49. Elsevier, pp. 2959–3038.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G. and Roncalli, T. (2000). Copulas for finance: a reading guide and some applications. Working paper, Groupe de Recherche Operationnelle, Credit Lyonnais.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability ans Statistics*, E. Eberlein and M. S. Taqqu, eds. Boston: Birkhäuser, pp. 165–192.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2**, 107–144.

Brock, W. A., Dechert, W. D., Scheinkman, J. A. and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews* **15**, 197–235.

Brooks, C. (1998). Predicting stock index volatility: Can market volume help? *Journal of Forecasting* **17**, 59–80.

Brooks, C. and Henry, O. T. (2000). Linear and non-linear transmission of equity return volatility: evidence from the US, Japan and Australia. *Economic Modelling* **17**, 497–513.

Campbell, J. Y., Lo, A. and MacKinlay, C. (1997). *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Carrasco, M. and Xiaohong, C. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18**, 17–39.

Chen, A.-S. and Lin, J. W. (2004). Cointegration and detectable linear and nonlinear causality: analysis using the London Metal Exchange lead contract. *Applied Economics* **36**, 1157–1167.

Chen, X., Fan, Y. and Patton, A. (2004). Simple tests for models of dependence between multiple financial time series, with applications to U.S. equity returns and exchange rates. Discussion Paper 483, Financial Markets Group, London School of Economics.

Chen, X., Fan, Y. and Tsyrennikov, V. (2004). Efficient estimation of semiparametric multivariate copula models. Working paper 04-W20, Department of Economics, Vanderbilt University.

Ciner, C. (2001). Energy shocks and financial markets: Nonlinear linkages. *Studies in Nonlinear Dynamics and Econometrics* **5**, 203–212.

Cook, D. R. (2004). Testing predictor contributions in sufficient dimension reduction. *Annals of Statistics* **32**, 1062–1099.

Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. *Académie Royale de Belgique, Bulletin de la Classe des Sciences* **5**, 274–292.

Delgado, M. and Mora, J. (2000). A nonparametric test for serial independence of regression errors. *Biometrika* **87**, 228–234.

Denker, M. and Keller, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64**, 505–522.

Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* **39**, 863–83.

Diebold, F. X. and Lopez, J. A. (1996). Forecast evaluation and combination. In *Handbook of Statistics*, G. S. Maddala and C. R. Rao, eds., vol. 14 of *Statistical Methods in Finance*. North Holland, Amsterdam, pp. 241–268.

Diks, C. and DeGoede, J. (2001). A general nonparametric bootstrap test for Granger causality. In *Global Analysis of Dynamical Systems*, H. W. Broer, B. Krauskopf and G. Vegter, eds. Bristol: IoP Publishing, pp. 391–403.

Diks, C. and Panchenko, V. (2005). A note on the Hiemstra-Jones test for Granger non-causality. *Studies in Nonlinear Dynamics and Econometrics* **9**, Art. 4, 1–7.

Diks, C. and Panchenko, V. (2006a). Tests for serial independence based on quadratic forms. *Statistica Sinica*. Forthcoming.

Diks, C. and Panchenko, V. (2006b). A new statistic and practical guidelines for non-parametric Granger causality testing. *Journal of Economic Dynamics and Control*. Forthcoming.

Diks, C. and Tong, H. (1999). A test for symmetry of multivariate probability distributions. *Biometrika* **86**, 605–614.

Diks, C., Van Zwet, W. R., Takens, F. and DeGoede, J. (1996). Detecting differences between delay vector distributions. *Physical Review E* **53**, 2169–2176.

Doukhan, P. (1994). *Mixing: Properties and Examples*, vol. 85 of *Lecture note in statistics*. Berlin: Springer Verlag.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* **84**, 313–342.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least-squares regression. I. *Biometrika* **37**, 409–428.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least-squares regression. II. *Biometrika* **38**, 159–177.

Durrleman, V., Nikeghbali, A. and Roncalli, T. (2000). Which copula is the right one? Working paper, Groupe de Recherche Operationnelle, Credit Lyonnais.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**, 1–26.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance*, S. Rachev, ed., Handbooks in Finance, 1. Elsevier, pp. 329–384.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987–1008.

Engle, R. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* **20**, 339–350.

Engle, R. F. and Manganelli, S. (2004). CAViaR: conditional value at risk by regression quantiles. *Journal of Business & Economic Statistics* **22**, 367–381.

Engle, R. F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working paper 2001-15, University of California, San Diego.

Fan, J. (2000). Prospects of nonparametric modeling. *Journal of the American Statistical Association* **95**, 1269–1300.

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statistical Science* **20**, 317–337.

Fan, Y. and Chen, X. (2002). Estimation of copula-based semiparametric time series models. Department of Economics, Vanderbilt University, Working paper, 02-W26R.

Feuerverger, A. (1993). A consistent test for bivariate dependence. *International Statistical Review* **61**, 419–433.

Franses, P. H. and Van Dijk, D. (2000). *Non-Linear Time Series Models in Empirical Finance*. Cambridge University Press.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.

Genest, C., Quessy, J.-F. and Remillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transform. *Scandinavian Journal of Statistics* **33**, 337–366.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*. Forthcoming.

Good, P. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics. New York: Springer-Verlag.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438.

Granger, C. W. J. and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Models*. Göttingen: Vandenhoeck and Ruprecht.

Granger, C. W. J. and Lin, J. (1994). Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis* **15**, 371–384.

Granger, C. W. J., Maasoumi, E. and Racine, J. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis* **25**, 649–669.

Granger, C. W. J. and Teräsvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters* **62**, 161–165.

Hafner, C. M., Van Dijk, D. and Franses, P. H. (2005). Semi-parametric modelling of correlation dynamics. Econometric Institute Research Report 2005-26, Erasmus University Rotterdam.

Hall, P. and Li, K.-C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics* **21**, 867–889.

Hall, P., Wolff, R. C. L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of American Statistical Association* **94**, 154–163.

Hall, P. and Yao, Q. (2005). Approximating conditional distribution functions using dimension reduction. *The Annals of Statistics* **33**, 1404–1421.

Hansen, B. E. (2004). Nonparametric Estimation of Smooth Conditional Distributions. Working paper, University of Wisconsin-Madison.

Heer, G. R. (1991). Testing independence in high dimensions. *Statistics & Probability Letters* **12**, 73–81.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance* **49**, 1639–1664.

Hoeffding, W. (1948a). A non-parametric test of independence. *The Annals of Mathematical Statistics* **19**, 546–557.

Hoeffding, W. (1948b). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**, 293–325.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *Annals of Mathematical Statistics* **23**, 169–192.

Hong, Y. (2000). Generalized spectral tests for serial dependence. *Journal of the Royal Statistical Society Series B* **62**, 557–574.

Hong, Y. and White, H. (2005). Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* **73**, 837–901.

Horowitz, J. L. (2001). The bootstrap. In *Handbook of Econometrics*, J. J. Heckman and E. E. Leamer, eds., vol. 5 of *Handbook of Econometrics*, chap. 52. Elsevier, pp. 3159–3228.

Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika* **30**, 81–93.

Kiviet, J. F. and Dufour, J.-M. (1997). Exact tests in single equation autoregressive distributed lag models. *Journal of Econometrics* **80**, 325–353.

Li, K.-C. and Duan, N. (1989). Regression analysis under link violation. *The Annals of Statistics* **17**, 1009–1052.

Li, W. K. (2004). *Diagnostic Checks in Time Series*. Chapman & Hall/CRC.

Ma, Y. and Kanas, A. (2000). Testing for a nonlinear relationship among fundamentals and the exchange rates in the ERM. *Journal of International Money and Finance* **19**, 135–152.

Malevergne, Y. and Sornette, D. (2003). Testing the Gaussian copula hypothesis for financial assets dependences. *Quantitative Finance* **3**, 231–250.

Marriott, F. (1979). Barnard's Monte Carlo tests: How many simulations? *Applied Statistics* **28**, 75–77.

Martens, M. and Poon, S.-H. (2001). Returns synchronization and daily correlation dynamics between international stock markets. *Journal of Banking & Finance* **25**, 1805–1827.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–315.

Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer Verlag.

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

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.

Okunev, J., Wilson, P. and Zurbruegg, R. (2000). The causal relationship between real estate and stock markets. *Journal of Real Estate Finance and Economics* **21**, 251–261.

Okunev, J., Wilson, P. and Zurbruegg, R. (2002). Relationships between Australian real estate and stock market prices – a case of market inefficiency. *Journal of Forecasting* **21**, 181–192.

Panchenko, V. (2005a). Goodness-of-fit test for copula. *Physica A* **355**, 176–182.

Panchenko, V. (2005b). Copulas for finance: estimation and evaluation. *Medium Econometrische Toepassingen* **13**, 10–12.

Parzen, E. (1962). On estimation of probability density and mode. *Annals of Mathematical statistics* **33**, 1065–1076.

Patton, A. (2005a). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics*. Forthcoming.

Patton, A. (2005b). Modelling asymmetric exchange rate dependence. *International Economic Review*. Forthcoming.

Pelletier, D. (2005). Regime switching for dynamical correlations. *Journal of Econometrics*. Forthcoming.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any population. II. The correlation coefficient test. *Supplement to the Journal of the Royal Statistical Society* **4**, 225–232.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association* **89**, 1303–1313.

Pompe, B. (1993). Measuring statistical dependences in time series. *Journal of Statistical Physics* **73**, 587–610.

Powell, J. L. and Stoker, T. M. (1996). Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics* **75**, 219–316.

Rosenblatt, M. (1956a). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical statistics* **27**, 642–669.

Rosenblatt, M. (1956b). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the USA* **42**, 43–47.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *The Annals of Statistics* **3**, 1–14.

Rosenblatt, M. and Wahlen, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statistics & Probability Letters* **15**, 245–252.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlations matrices. *Communications in statistics: theory and methods* **22**, 965–984.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Silvapulla, P. and Moosa, I. A. (1999). The relationship between spot and futures prices: Evidence from the crude oil market. *Journal of Futures Markets* **19**, 157–193.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

Skaug, H. J. and Tjøstheim, D. (1993a). A nonparametric test of serial independence based on the empirical distribution function. *Biometrika* **80**, 591–602.

Skaug, H. J. and Tjøstheim, D. (1993b). Nonparametric tests of serial independence. In *Developments in Time Series Analysis*, T. Subba Rao, ed., chap. 15. London: Chapman and Hall, pp. 207–230.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris* **8**, 229–231.

Sklar, A. (1996). Random variables, distribution functions, and copulas - a personal look backward and forward. In *Distributions with Fixed Marginals and Related Topics*, L. Rüschendorf, B. Schweizer and M. D. Taylor, eds. Hayward, CA: Institute of Mathematical Statistics, pp. 1–14.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology* **15**, 72–101.

Stigler, S. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press.

Su, L. and White, H. (2003). A nonparametric Hellinger metric test for conditional independence. Working paper, Department of Economics, UCSD.

Székely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis* **93**, 58–80.

Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing*, C. H. Chen, ed. Amsterdam: Sijthoff and Noordhoff, pp. 101–141.

Tsay, R. S. (2002). *Analysis of Financial Time Series.* Wiley series in probability and statistics. Wiley-Interscience.

Van den Goorbergh, R. (2004). A copula-based autoregressive conditional dependence model of international stock markets. Working Paper 022, Research Department, Netherlands Central Bank.

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge, New York: Cambridge Unversity Press.

Van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* **17**, 549–564.

Von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *The Annals of Mathematical Statistics* **18**, 309–348.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* **64**, 1067–1084.

Wolff, R. C. (1994). Independence in time series: another look at the BDS test. *Philosophical Transactions of the Royal Society Series A* **348**, 383–395.

Yoshihara, K. (1976). Limiting behavior of U-statistics for stationary absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **35**, 237–252.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* **18**, 931–955.

Samenvatting (Summary in Dutch)

Dit proefschrift behandelt een aantal onderwerpen gerelateerd aan niet-parametrische en semi-parametrische methoden in de financiële econometrie. Vergeleken met parametrische methoden vereisen niet-parametrische en semi-parametrische methoden minder aannamen omtrent verdelingen en/of functionele vormen en laten ze meer flexibiliteit toe in financiëel modelleren en voorspellen. Anderzijds vereisen niet-parametrische schatters meer waarnemingen omdat hun asymptotische convergentie typisch trager is dan die van parametrische schatters. In een multivariate context kan dit leiden tot de zogenaamde 'curse of dimensionality'. Dit is de reden dat vaak voor een semi-parametrisch model wordt gekozen waarin beide benaderingen gecombineerd worden. In de meeste hoofdstukken van dit proefschrift streven we de niet-parametrische benadering na. Het laatste hoofdstuk, dat multivariate tijdreeksen behandelt, volgt een semi-parametrische benadering. Het proefschrift is onderverdeeld in drie hoofdthemas: afhankelijkheid (Hoofdstukken 2 en 3), causaliteit (Hoofdstukken 4 en 5) en voorspelling (Hoofdstuk 6).

In het inleidende Hoofdstuk 1 beschrijven we de belangrijkste theoretische noties die in het proefschrift gebruikt worden, in het bijzonder zwakke afhankelijkheid en menging in processen, statistische grootheden die bekend staan als U- en V-statistics, en het begrip copula. In de behandeling van zwakke afhankelijkheid komen we tot een classificatie van voorwaarden van sterke menging, en lichten hun rol in de tijdreeksanalyse toe. Daarna beschrijven we U- en V-statistics. Dit zijn belangrijke klassen van schatters met behulp waarvan asymptotische theorie in niet-parametrische contexten ontwikkeld kan worden. In dit hoofdstuk introduceren we ook copulas, functies die de afhankelijkheid tussen variabelen volledig beschrijven. Copulas laten een decompositie toe van multivariate modelleerproblemen in univariate componenten enerzijds, en de afhankeli-

jkheidsstructuur anderzijds.

In Hoofdstuk 2 stellen we nieuwe toetsen voor seriële afhankelijkheid voor die gebaseerd zijn op kwadratische vormen, en confronteren deze met veel gebruikte gerelateerde niet-parametrische toetsen. Om toetsen met een exact significantieniveau te verkrijgen, implementeren we een Monte Carlo procedure die gebruik maakt van permutaties van de oorspronkelijke waarnemingen. Het bandbreedte selectie probleem wordt behandeld door een nieuwe meervoudige bandbreedte methode te ontwikkelen, gebaseerd op meerdere waarden van de bandbreedte. Door middel van numerieke simulaties wordt aangetoond dat de toetsen goed presteren ten opzichte van bestaande niet-parametrische toetsen. De permutatie toets heeft echter een blijkt conservatief te zijn wanneer deze wordt toegepast op de residuen van geschatte tijdreeksmodellen. Het blijkt dat dit verholpen kan worden met een parametrische bootstrap. De praktische toepassing van de toets wordt geïllustreerd met behulp van financiële tijdreeksen. In toekomstig onderzoek willen we de gevoeligheid van de toets voor residuen analytisch onderzoeken, en het concept van de kwadratische vorm generaliseren naar de contexten van één- en twee-steekproef toetsen.

In Hoofdstuk 3 gebruiken we kwadratische vormen om een goodness-of-fit toets voor copulas te ontwikkelen. De afstand tussen de waargenomen en de hypothetische copula wordt gemeten met behulp van op kernels gebaseerde kwadratische vormen. Op deze manier vermijden we het gebruik van zogenaamde plug-in dichtheidsschatters, die niet optimaal voor het doel van toetsen hoeven te zijn, en waarvoor de 'curse of dimensionality' gemakkelijk op kan treden. De toets gebaseerd op de kwadratische vorm blijkt goede eindige steekproefeigenschappen te hebben (tenminste voor laag-dimensionale data) in vergelijking met andere goodness-of-fit toetsen. We passen de toets toe op 'US large cap' data en DAX aandelen, en komen tot de conclusie dat de afhankelijkheden in deze data niet goed beschreven kunnen worden met behulp van een Gaussische copula. In toekomstig onderzoek hopen we de toepasbaarheid van de toets op hoger dimensionale problemen te verbeteren.

Hoofdstuk 4 behandelt een consistentieprobleem in de veelgebruikte niet-parametrische toets voor Granger causaliteit ontwikkeld door Hiemstra en Jones (1994). Intuïtief is er een Granger causaal verband van X naar Y als het toevoegen van verleden observaties

van X aan de informatieset (die de verleden observaties van Y al bevat) de kennis omtrent de verdeling van de huidige waarde van Y beïnvloedt. We laten analytisch zien dat de relatie die getoetst wordt door Hiemstra en Jones (1994) niet noodzakelijk consistent is met de nulhypothese (het ontbreken van Granger causaliteit). Simulaties gebaseerd op processen die aan de nulhypothese voldoen laten inderdaad zien dat de toets zwaar kan over-verwerpen, en dus onterecht kan leiden tot de conclusie dat Granger causaliteit aanwezig is.

In Hoofdstuk 5 bestuderen we de oorzaken van deze over-verwerping analytisch. We komen tot de conclusie dat de Hiemstra-Jones toets voor de nulhypothese van afwezigheid van Granger causaliteit gevoelig is voor voorwaardelijke heteroskedasticiteit, die vaak aanwezig is in financiële tijdreeksen. We stellen een nieuwe toetsgrootheid voor die de globale toetsgrootheid vervangt door een gemiddelde van locale afhankelijkheidsmaten. Verder leiden we de juiste snelheden af waarmee de bandbreedte naar nul moet gaan opdat de toets asymptotisch de bias problemen overwint. We passen de toets toe op historische prijzen en handelsvolumes van de S&P index en vinden beduidend minder statistisch bewijs voor de hypothese dat handelsvolume prijzen zou beïnvloeden dan gesuggereerd wordt door de Hiemstra-Jones toets.

In Hoofdstuk 6 suggereren we een nieuwe semi-parametrische procedure voor het schatten van multivariate autoregressieve modellen. De voorwaardelijke copula, die de kruis-afhankelijkheid tussen de tijdreekselementen samenvat, wordt parametrisch gemodelleerd, terwijl de voorwaardelijke marginale verdelingen niet-parametrisch geschat worden. Voor het schatten van de voorwaardelijke marginale verdeling stellen we een dimensiereductie techniek voor om de 'curse of dimensionality' te vermijden. We confronteren dit semi-parametrische copula model met het veelgebruikte, volledig parametrische DCC model, en vergelijken de voorspelkracht van beide modelvormen. De semi-parametrische modellen blijken betere één-stap-vooruit voorspellingen te genereren voor een collectie van vier internationale indices. We suggereren een aantal toekomstige verbeteringen voor de semi-parametrische procedure om de praktische bruikbaarheid te verhogen.

In dit proefschrift tonen we aan dat er een brede toepasbaarheid is voor niet-parametrische en semi-parametrische methoden in financiële economie. Gegeven de snelle on-

twikkelingen op het gebied van computertechnologie en verbeteringen van simulatiealgoritmen, menen we dat deze technieken in toenemende mate toegepast zullen worden in de toekomst. Eén van de belangrijkste onopgeloste problemen in de niet-parametrische analyse van multivariate data is de 'curse of dimensionality'. In de toekomst hopen we te kunnen bijdragen aan de oplossing van dit probleem door middel van dimensiereductie technieken gebareerd op kwadratische vormen.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

335. A.H. NÖTEBERG, *The medium matters: The impact of electronic communication media and evidence strength on belief revision during auditor-client inquiry.*
336. M. MASTROGIACOMO, *Retirement, expectations and realizations. Essays on the Netherlands and Italy.*
337. E. KENJOH, *Balancing work and family life in Japan and four European countries: Econometric analyses on mothers' employment and timing of maternity.*
338. A.H. BRUMMANS, *Adoption and diffusion of EDI in multilateral networks of organizations.*
339. K. STAAL, *Voting, public goods and violence.*
340. R.H.J. MOSCH, *The economic effects of trust. Theory and empirical evidence.*
341. F. ESCHENBACH, *The impact of banks and asset markets on economic growth and fiscal stability.*
342. D. LI, *On extreme value approximation to tails of distribution functions.*
343. S. VAN DER HOOG, *Micro-economic disequilibrium dynamics.*
344. B. BRYS, *Tax-arbitrage in the Netherlands evaluation of the capital income tax reform of January 1, 2001.*
345. V. PRUZHANSKY, *Topics in game theory.*
346. P.D.M.L. CARDOSO, *The future of old-age pensions: Its implosion and explosion.*
347. C.J.H. BOSSINK, *To go or not to go…? International relocation willingness of dual-career couples.*
348. R.D. VAN OEST, *Essays on quantitative marketing models and Monte Carlo integration methods.*
349. H.A. ROJAS-ROMAGOSA, *Essays on trade and equity.*
350. A.J. VAN STEL, *Entrepreneurship and economic growth: Some empirical studies.*
351. R. ANGLINGKUSUMO, *Preparatory studies for inflation targeting in post crisis Indonesia.*
352. A. GALEOTTI, *On social and economic networks.*
353. Y.C. CHEUNG, *Essays on European bond markets.*
354. A. ULE, *Exclusion and cooperation in networks.*
355. I.S. SCHINDELE, *Three essays on venture capital contracting.*
356. C.M. VAN DER HEIDE, *An economic analysis of nature policy.*
357. Y. HU, *Essays on labour economics: Empirical studies on wage differentials across categories of working hours, employment contracts, gender and cohorts.*
358. S. LONGHI, *Open regional labour markets and socio-economic developments: Studies on adjustment and spatial interaction.*
359. K.J. BENIERS, *The quality of political decision making: Information and motivation.*

360. R.J.A. LAEVEN, *Essays on risk measures and stochastic dependence: With applications to insurance and finance.*
361. N. VAN HOREN, *Economic effects of financial integration for developing countries.*
362. J.J.A. KAMPHORST, *Networks and learning.*
363. E. PORRAS MUSALEM, *Inventory theory in practice: Joint replenishments and spare parts control.*
364. M. ABREU, *Spatial determinants of economic growth and technology diffusion.*
365. S.M. BAJDECHI-RAITA, *The risk of investment in human capital.*
366. A.P.C. VAN DER PLOEG, *Stochastic volatility and the pricing of financial derivatives.*
367. R. VAN DER KRUK, *Hedonic valuation of Dutch Wetlands.*
368. P. WRASAI, *Agency problems in political decision making.*
369. B.K. BIERUT, *Essays on the making and implementation of monetary policy decisions.*
370. E. REUBEN, *Fairness in the lab: The effects of norm enforcement in economic decisions.*
371. G.J.M. LINDERS, *Intangible barriers to trade: The impact of institutions, culture, and distance on patterns of trade.*
372. A. HOPFENSITZ, *The role of affect in reciprocity and risk taking: Experimental studies of economic behavior.*
373. R.A. SPARROW, *Health, education and economic crisis: Protecting the poor in Indonesia.*
374. M.J. KOETSE, *Determinants of investment behaviour: Methods and applications of meta-analysis.*
375. G. MÜLLER, *On the role of personality traits and social skills in adult economic attainment.*
376. E.H.B. FEIJEN, *The influence of powerful firms on financial markets.*
377. J.W. GROSSER, *Voting in the laboratory.*
378. M.R.E. BRONS, M*eta-analytical studies in transport economics: Methodology and applications.*
379. L.F. HOOGERHEIDE, *Essays on neural network sampling methods and instrumental variables.*
380. M. DE GRAAF-ZIJL, *Economic and social consequences of temporary employment.*
381. O.A.C. VAN HEMERT, *Dynamic investor decisions.*
382. Z. ŠAŠOVOVÁ, *Liking and disliking: The dynamic effects of social networks during a large-scale information system implementation.*
383. P. RODENBURG, *The construction of instruments for measuring unemployment.*
384. M.J. VAN DER LEIJ, *The economics of networks: Theory and empirics.*
385. R. VAN DER NOLL, *Essays on internet and information economics.*