

Forecasting Using Time Varying Meta-Elliptical Distributions with a Study of Commodity Prices

Alessio Sancetta and Arina Nikanrova

April 2005

CWPE 0516

Not to be quoted without permission

Forecasting Using Time Varying Meta-Elliptical Distributions with a Study of Commodity Futures Prices*

Alessio Sancetta** and Arina Nikandrova***

**Faculty of Economics, University of Cambridge, UK

***Dharma Holdings, Luxembourg

April 6, 2005

Abstract

We propose a methodological approach to the forecast and evaluation of multivariate distributions with time varying parameters. For reasons related to feasible inference attention is restricted to meta-elliptical distributions. We use our approach for the study of a large data set of 16 commodity prices. Our approach leads to a theory for model validation avoiding common problems caused by discontinuities, time variation of parameters and nuisance parameters.

Keywords: Commodity Prices, Copula Function, Meta-Elliptical Distribution, Non-parametric Estimation, Weibull Distribution.

JEL:C14, C16, C31, C32.

1 Introduction

This paper proposes an iterative methodology for one step ahead forecasting and evaluation of multivariate meta-elliptical distributions. Specifically, we provide a framework for probability forecast of a K dimensional sequence of random variables that are possibly dependent and non-identically distributed. Our motivation is the forecast of the multivariate distribution of a large number of financial returns.

We consider the general case where the parameters of the whole distribution characterising the random variables may change over time, and we use linear filters for forecasting purposes. To obtain forecasts through application of a linear filter to some function of the data is a standard procedure. For example, RiskMetrics estimator of volatility uses an exponential moving average of the squared returns. This forecast is limited to volatility estimation and only assumes that the second moment is changing. The framework we focus on is more general and encompasses well known methods like linear projections estimators and nonparametric estimators.

Generalising GARCH and the approach of Hansen (1994), some studies have considered the case of multivariate time varying distributions, e.g. Jondeau and Rockinger (2005), Patton (2005).

*The first author acknowledges financial support from the ESRC research award RES 000-23-0400. Corresponding author: Alessio Sancetta, Faculty of Economics, Sidgwick Avenue, University of Cambridge, UK; e-mail: Alessio.Sancetta@econ.cam.ac.uk.

The focus of our paper is more on methodology. We consider broader issues related to estimation of time varying parameters that could be nonstationary. Our framework also covers semi and nonparametric techniques that can be applied to a large number of assets.

We are also interested in assessing the proposed model and improving upon it. To this end, we construct inferential procedures imposing no other specific restrictions on the model we use. For example, the model may have discrete components. In this case, standard goodness of fit procedures like the Kolmogorov-Smirnov statistic are not applicable, but our inferential procedure easily circumvents the problem of discontinuity. Also, when parameters are estimated, standard inferential procedures could encounter nuisance parameters problems because the null hypothesis of standard tests might depend on the estimated parameters. This is a serious problem when the dimension of parameters is large, as in the applications we have in mind. Moreover, if the parameters are time varying, it is not clear how to formulate a null hypothesis. Our methodology was developed to deal specifically with these issues and provides simple solutions within the framework of meta-elliptical distributions whose parameters are time varying.

In this paper we do not discuss the specific nature of time variation, which could be due to time dependence, structural breaks, nonstationary, etc. A large literature is devoted to modeling the time dependence of log returns, and GARCH models have proven to be effective in this task. Also, recently interest has been raised about the possibility of heterogeneity in the distribution of returns. A common assumption is local stationarity, which allows us to estimate parameters locally around the time of interest (Fan, 2005). Another possibility is that time heterogeneity is caused by structural breaks. If the time of break can be identified, then different estimations are carried out over intervals of homogeneity (Mikosch and Stărică, 1999, Polzehl and Spokoiny, 2002, 2003, Mercurio and Spokoiny, 2004, and Pesaran et al., 2004 for a Bayesian approach). However, if heterogeneity is caused by time varying parameters and not structural breaks, modelling time heterogeneity and dependence at the same time can be a particularly hard task. In this case, a common approach is the use of unobserved components models, where the time varying parameters are unobserved components. When the dimension of the problem increases, this approach might not be feasible.

The forecasting problem can be easier than the inference problem (particularly in the case of one step ahead forecast). The former does not require a model that can properly differentiate between dependence and heterogeneity and its validity is only subject to its performance. For example, the simplest models usually provide better forecasts without necessarily shedding light on all the driving mechanisms behind the data generating process. This does not mean that we should disregard any attempt to model complex problems in a forecasting framework. In fact some refined modelling approaches can be extremely valuable for forecasting under certain conditions (e.g. Pesaran et al., 2004).

A successful forecasting model can also be considered as a valid model for explaining certain aspects of reality. Hence, the methodology proposed is used to study the behaviour of commodity futures prices. Commodities are usually found to be very volatile (Pindyck, 2004). For many reasons, such as changes in technology affecting the cost of production (e.g. genetically modified agricultural products), cartels among producing countries reducing supply (e.g. OPEC), changes in legislations (e.g. import-export tariffs), international war conflicts (e.g. Iraq war), changes in weather conditions (e.g. global warming), the behaviour of commodity prices can be expected to be nonstationary.

There has been some statistical study in this area (e.g. Deb et al., 1996, Taylor, 1980), though, research is usually concerned with testing some form of rational expectation hypothesis (e.g. Deaton and Laroque, 1992, 1996). In our empirical study we find that commodity log returns can be well described using a simple model based on our approach.

The plan for the paper is as follows. In Section 2, we highlight our general procedure based on dynamic linear filtering. In Section 3, we show how this procedure leads to inferential procedures under minimal conditions naturally avoiding serious problems like nonstationarity and nuisance parameters' under the null hypothesis. Section 4 provides a study of futures prices for 16 commodities. This study is of interest on its own, as to our knowledge, most of the applied work concentrate of stock prices or indices on stock prices. Section 5 provides a further discussion and extensions.

It is convenient to introduce some notation. \mathbb{P}_n stands for the empirical measure that assign weight $1/n$ to each observation. The symbols w and $a.s.$ on top of equality signs means that the equality holds weakly (i.e. in distribution) or almost surely. For typographical reasons, we may put indices in parenthesis instead of using subscripts, e.g. $x(i)$ instead of x_i . When not needed, we may suppress subscripts. Suppose B is a set, then $A \subset\subset B$ is a compact set inside B and $\#A$ is the cardinality of A , when A is a set with countable elements. For any mapping $(x_1, \dots, x_K) \mapsto f$ and $K > 1$, $f_{,k} := \partial_{x(k)} f := \partial f / \partial x_k$. The inner product is denoted by $\langle \dots, \dots \rangle$. Furthermore, for X and Y with joint distribution F , $F(X|Y = y)$ or simply $F(X|y)$ stands for the distribution of X , conditional on $Y = y$.

2 Methodology

Suppose $(\mathbf{X}_s)_{s \in \mathbb{Z}}$ is a sequence of random vectors with values in \mathbb{R}^K . We observe realizations from the segment $(\mathbf{X}_s)_{0 \leq s < t}$, i.e. $\mathbf{x}_0, \dots, \mathbf{x}_{t-1}$. We need to issue a probability forecast for \mathbf{X}_t . In the absence of knowledge about the (conditional) distribution of \mathbf{X}_t , we could postulate a family of joint distributions indexed in some finite parameter space and choose a method to identify the optimal parameter for the forecasting problem. For large K , this is a formidable task, unless we are prepared to impose very restrictive conditions on the family of distributions and the dynamics of the parameters. We propose to consider the problem of providing a good forecast for each marginal distribution first and then to join these marginals using some nonlinear transformations and the copula function.

2.1 First Stage: Describing the Marginals

Define \mathcal{F}_{t-1} to be the sigma algebra generated by the segment of random variables $(X_s)_{0 \leq s < t}$. We postulate a class of marginal distributions $\{P_\theta, \theta \in \Theta \subset \mathbb{R}^d, d \geq 1\}$ for $(X_t)_{t \in \mathbb{Z}}$. We need to choose an \mathcal{F}_{t-1} measurable θ_t such that $P_{\theta(t)}$ is a good distribution forecast for X_t . For the moment, to simplify notation in the discussion, let $d = 1$. Suppose there is some function $g: \mathbb{R} \rightarrow \mathbb{R}$ such that $\theta_t = \mathbb{E}(g(X_t) | \mathcal{F}_{t-1})$. This is often the case.

Example 1 Suppose $P_\theta(x)$ has density, with respect to some dominating measure μ , equal to $p_\theta(x) = \exp\{\langle A(\theta), g(x) \rangle + B(\theta)\} C(x)$, for some positive functions A, B and C , where $\theta = \int g(x) p_\theta(x) d\mu(x)$. In this case, A and g could be $d > 1$ dimensional vector functions. This

density is said to belong to the exponential family model, with natural parameter θ , canonical parameter $A(\theta)$ and canonical statistic $g(x)$. Properties of these models in relation to econometrics can be found in van Garderen (1997). The Gaussian, the Poisson the Binomial distributions all belong to this family.

Then, suppose $g(X_t)$ admits the semimartingale representation

$$g(X_t) = f_t + v_t \varepsilon_t, \quad (1)$$

where f_t and v_t are \mathcal{F}_{t-1} measurable and $\mathbb{E}\varepsilon_t = 0$, $\mathbb{E}\varepsilon_t^2 = 1$. Hence, $\theta_t = f_t$ (by reparametrisation of the marginal distributions, this covers the case $g'(\theta_t) = f_t$, for some function g'). The parameter estimation is equivalent to estimation of the \mathcal{F}_{t-1} measurable trend in $g(X_t)$. We can estimate or at least approximate θ_t by $\hat{\theta}_t = \sum_{s < t} w(s, t) g(x_s)$, where $(w(s, t))_{0 \leq s < t \in \mathbb{N}_+}$ is a linear filter possibly depending on $(X_s)_{s < t}$ so that $(w(s, t))_{0 \leq s < t \in \mathbb{N}_+}$ is \mathcal{F}_{t-1} measurable. This framework encompasses many different methods like averages, moving averages, exponential smoothing, kernel smoothing and linear projections.

Example 2 Suppose $w(s, t) = (1 - h)h^{|t-s-1|}$, which is just exponential smoothing, and $g(x) = x^2$. This is RiskMetrics volatility estimation.

The use of exponential smoothing can be justified under specific assumptions.

Example 3 Suppose $w(s, t) = (1 - h)h^{|t-s-1|}$, and

$$\begin{aligned} g(X_t) &= f_t + v\varepsilon_t \\ &= \beta_t + e_t, \end{aligned}$$

$$\beta_t = \beta_{t-1} + \eta_t, v\varepsilon_t = e_t + \eta_t$$

where η_t and e_t are mutually uncorrelated at all lags, with signal to noise ratio $q = \text{var}(\eta_t) / \text{var}(e_t)$. Under this assumption, the optimal asymptotic one step ahead prediction using the Kalman filter gives $h = \frac{q + (q^2 + 4q)^{1/2}}{2 + q + (q^2 + 4q)^{1/2}}$, which is the steady state Kalman gain (Harvey, 1989, p.175, see also Muth, 1960).

Unfortunately, this does not include the case of parameters being \mathcal{F}_t measurable instead of \mathcal{F}_{t-1} measurable. In this case, we would have $\theta_t = f_t + \text{error}$. If the error is relatively small, the above procedure seems reasonable. More sound, but involved alternatives will not be considered here.

By choice of the linear filter, we can account for time heterogenous parameter if we suspect that this is a problem.

Example 4 Suppose $X_t = \sigma_t \eta_t$ where η_t is standard Gaussian, and σ_t is \mathcal{F}_{t-1} measurable. For volatility estimation, Mercurio and Spokoiny (2004) propose for $\alpha \in \mathbb{R}_+$

$$\mathbb{E}(|X_t|^\alpha | \mathcal{F}_{t-1}) = \mathbb{E}(|X_t|^\alpha | \mathcal{F}_{t-1}) = (\#\mathcal{I})^{-1} \sum_{s \in \mathcal{I}} |X_s|^\alpha$$

over a homogeneity region \mathcal{I} (arguing for $\alpha = 1/2$ as the best choice). The linear filter gives same weight over the homogeneity region and zero elsewhere. Mercurio and Spokoiny (2004) provide inference procedures to identify homogeneity regions.

More generally than (1), we can suppose that there is a function $g : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ such that $\mathbb{E}(g(X_t, \theta_t) | \mathcal{F}_{t-1}) = 0$, and $g(X_t, \theta)$ admits the semimartingale representation

$$g(X_t, \theta) = f_t(\theta) + v_t(\theta) \varepsilon_t,$$

where $f_t(\theta)$ and $v_t(\theta)$ are \mathcal{F}_{t-1} measurable. Then around the optimum, $f_t(\theta)$ admits the usual quadratic representation $f_t(\theta^{i+1}) \simeq f_t(\theta^i) + \partial_\theta f_t(\theta^i) (\theta^{i+1} - \theta^i)$. Hence,

$$\mathbb{E}(g(X_t, \theta^{i+1}) | \mathcal{F}_{t-1}) \simeq f_t(\theta^i) + \partial_\theta f_t(\theta^i) (\theta^{i+1} - \theta^i), \quad (2)$$

where $\mathbb{E}(g(X_t, \theta) | \mathcal{F}_{t-1}) = f_t(\theta)$. Applying a filter $(w(s, t))_{0 \leq s < t \in \mathbb{N}_+}$ such that $\hat{f}_t(\theta) = \sum_{s < t} w(s, t) g(X_s, \theta)$ and substituting $\hat{f}_t(\theta)$ in (2), usual numerical approximation techniques can be employed to find $\hat{\theta}_t$ (e.g. Newton's method), where $\hat{\theta}_t$ is such that $\hat{f}_t(\hat{\theta}_t) = 0$.

The $d > 1$ dimensional case is dealt similarly either by defining a vector of estimating equations or by direct solution if the parameter defining equation admits an explicit solution as a function of X_t . Reparametrisation may be used to simplify the estimation and to imply different dynamics.

Example 5 Suppose X_t is distributed according to the double Weibull density

$$ab|x|^{b-1} \exp(-a|x|^b), \quad (3)$$

where $\theta = (a, b) \in \mathbb{R}_+^2$. It can be shown that $\mathbb{E}|X_t|^b = 1/a$. Suppose a is time varying, i.e. $a = a_t$. If b is unknown, we need to solve a system of non-linear equations (b does not admit an explicit solution), i.e.

$$\hat{a}_t = \left(\sum_{s=0}^{t-1} |X_s|^b w(s, t) \right)^{-1},$$

and \hat{a}_t can be substituted in the loglikelihood of (3) and this loglikelihood be maximized with respect to b only. This can be computationally involved. If we reparametrise $a_t = \gamma_t^{-b}$, then $\mathbb{E}|X_t| \propto \gamma_t$, which implies a solution independent of b , but also different dynamic relations.

2.2 Second Step: Uniform Representation

Once the marginal distributions are defined, the second step for the proposed methodology consists in redefining the random vector $\mathbf{X}_t = (X_{t1}, \dots, X_{tK})$ into a new probability space such that each X_{tk} is transformed into a uniform $[0, 1]$ random variable. To achieve this, under general conditions on the Lebesgue decomposition of the marginal distributions of \mathbf{X}_t , we use the following function

$$\tilde{F}(x, v) = \Pr(X < x) + v \Pr(X = x). \quad (4)$$

Suppose V is a $[0, 1]$ uniform random variable. Then, $\tilde{F}(X, V)$ is a $[0, 1]$ uniform random variable (e.g. Proposition 1 in Rüschendorf and de Volk, 1993). Moreover, define

$$Z = \inf \left\{ x : \Pr(X \leq x) \geq \tilde{F}(X, V) \right\}.$$

Then, $Z \stackrel{a.s.}{=} X$ (e.g. Rüschendorf and de Volk, 1993). We use this transformation to define uniform $[0, 1]$ random variables. Since the true marginal is unknown, the estimated marginals

from the family $\{P_\theta, \theta \in \Theta\}$ will be used as an approximation for $\Pr(X < x)$ and $\Pr(X = x)$ in (4). Hence, no continuity assumption is required for $\{P_\theta, \theta \in \Theta\}$. Therefore, we can define

$$\mathbf{U}_t = (U_{t1}, \dots, U_{tK}) = \left(\tilde{F}_{t1}(X_{t1}, V_{t1}), \dots, \tilde{F}_{tK}(X_{tK}, V_{tK}) \right). \quad (5)$$

Since \mathbf{U}_t is a uniform $[0, 1]^K$ random vector, its joint distribution is the copula of \mathbf{X}_t . If \mathbf{X}_t does not have marginals that are continuous, there is not a unique copula for it (e.g. Sklar, 1973 Corollary to Theorem 1). However, \mathbf{U}_t is derived from (5) and it is a continuous random variable. Hence, there is a unique copula for \mathbf{U}_t which is also its joint distribution. If the copula of \mathbf{X}_t is not unique, the copula of \mathbf{U}_t will be used as a unique choice of copula. This solves any identification issues for the copula of \mathbf{X}_t , and the term copula would only refer to this version with continuous uniform marginals.

2.3 Third Step: Joining the Marginals

As a last step we join the marginals and derive the full joint distribution. For reasons that will become apparent later and in the next Section, we will restrict attention to the class of meta-elliptical copulae and discuss parameters' estimation.

2.3.1 Meta-Elliptical Copulae

Let \mathbf{X} be a random vector in \mathbb{R}^K . Then \mathbf{X} is said to have an elliptic density (e.g. Kano, 1994, Fang et al., 2002) with location parameter μ , scale matrix Σ and function φ , if

$$pdf_\varphi(\mathbf{x}) = \det|\Sigma|^{-1/2} \varphi(\langle (\mathbf{x} - \mu), \Sigma^{-1}(\mathbf{x} - \mu) \rangle). \quad (6)$$

For our purposes, we set $\mu = \mathbf{0}$ and restrict Σ to have diagonal entries equal to one. Moreover, to uniquely identify φ and Σ , we restrict Σ to have all entries in $[-1, 1]$, i.e. $\Sigma_{ij} \in [-1, 1]$. Under this condition, all the marginals of (6) are identical, say F_φ , and only depend on φ and are given by (e.g. Fang et al., 2002)

$$F_\varphi(x) = \frac{1}{2} + \frac{\pi^{(K-1)/2}}{\Gamma((K-1)/2)} \int_{-\infty}^x \int_{z^2}^{\infty} (y - z^2)^{(K-1)/2} \varphi(y) dy dz,$$

where $\Gamma(x)$ is the gamma function. Fang et al. (2002) have introduced the term meta-elliptical distributions to describe distributions related to the density in (6). Using the copula, their definition can be rephrased as follows: X_1, \dots, X_K have meta-elliptical distribution with marginals F_1, \dots, F_K and scaling function φ if their copula density, say c , can be written as

$$\begin{aligned} c_\varphi(\mathbf{u}; \Sigma) &= \det|\Sigma|^{-1/2} \varphi(\langle (\mathcal{Q}_\varphi \mathbf{u}), \Sigma^{-1}(\mathcal{Q}_\varphi \mathbf{u}) \rangle) J(\mathbf{u}), \\ J(\mathbf{u}) &= \left(\prod_{k=1}^K \frac{dF_\varphi(u_k)}{du_k} \right)^{-1}, \end{aligned} \quad (7)$$

where $\mathcal{Q}_\varphi : [0, 1]^K \rightarrow \mathbb{R}^K$ is an operator such that

$$\mathcal{Q}_\varphi \mathbf{u} = (F_\varphi^{-1}(u_1), \dots, F_\varphi^{-1}(u_K))^T. \quad (8)$$

Therefore, we shall call meta-elliptical copulae the class of copulae corresponding to meta-elliptical distributions. One important property of these copulae is that we can write $\mathbf{Z} := \mathcal{Q}_\varphi \mathbf{U}$,

and $\mathbf{EZZ}^T = \Sigma = AA^T$ for some $K \times K$ matrix A (e.g. $A = D\Lambda$, where Λ is a diagonal matrix of eigenvalues and D is an orthonormal matrix of eigenvectors). Then, $\tilde{\mathbf{Z}} := A^{-1}\mathbf{Z}$ is a vector with scaling matrix equal to the identity matrix, and $\tilde{\mathbf{U}} := \mathcal{Q}_\varphi^{-1}\tilde{\mathbf{Z}}$ is a vector with same copula as \mathbf{U} , but with scaling matrix equal to the identity. In certain cases, this implies that $\tilde{\mathbf{U}}$ have independence copula.

Example 6 Define $\varphi(x) := \varphi^K(x) = \exp(-\frac{x}{2}) / (2\pi)^{K/2}$, which is the generator for the Gaussian copula. Then,

$$c_\varphi(\mathbf{u}; I) = \frac{1}{(2\pi)^{K/2}} \exp\left(-\frac{\langle(\mathcal{Q}_\varphi\mathbf{u}), (\mathcal{Q}_\varphi\mathbf{u})\rangle}{2}\right) J(\mathbf{u}),$$

where

$$J(\mathbf{u}) = (2\pi)^{K/2} \exp\left(\frac{\langle(\mathcal{Q}_\varphi\mathbf{u}), (\mathcal{Q}_\varphi\mathbf{u})\rangle}{2}\right).$$

Hence, $c_\varphi(\mathbf{u}; I) = 1$.

Unfortunately, this is not always the case.

Example 7 Define

$$\varphi(x) := \varphi_\gamma^K(x) = \frac{\Gamma((\gamma + K)/2)}{\Gamma(\gamma/2) (\pi\gamma)^{K/2}} (1 + x/\gamma)^{-(\gamma+K)/2},$$

which is the generator for the t -copula with γ degrees of freedom. Then,

$$c_\varphi(\mathbf{u}; I) = \frac{\Gamma((\gamma + K)/2)}{\Gamma(\gamma/2) (\pi\gamma)^{K/2}} \left(1 + \frac{\langle(\mathcal{Q}_\varphi\mathbf{u}), (\mathcal{Q}_\varphi\mathbf{u})\rangle}{\gamma}\right)^{-(\gamma+K)/2} J(\mathbf{u})$$

$$J(\mathbf{u}) = \prod_{k=1}^K \frac{\Gamma(\gamma/2) (\pi\gamma)^{K/2}}{\Gamma((\gamma + 1)/2)} \left(1 + \frac{[F_\varphi^{-1}(u_k)]^2}{\gamma}\right)^{(\gamma+1)/2},$$

and it is simple to see that $c_\varphi(\mathbf{u}; I) \neq 1$.

The dependence properties of elliptical distributions have been studied in details by Hult and Lindskog (2002).

2.3.2 Parameters' Representation

Restricting attention to elliptical copulae with given function φ , the parameter of interest is Σ . Following the same approach as for the marginals, we assume that there exists a function $g: \mathbb{R}^K \rightarrow \mathbb{R}^K \times \mathbb{R}^K$ such that $\Sigma_t = \mathbb{E}[g(\mathbf{U}_t) | \mathcal{F}_{K;t-1}]$, where $\mathcal{F}_{K;t-1}$ is the sigma algebra generated by $(\mathbf{X}_s)_{0 \leq s < t}$. (Notice that we only need to condition on $\mathcal{F}_{k;t-1}$, as the uniform random variables used in (5) are iid copies of a same uniform $[0, 1]$ random variable V .)

Suppose that the following semimartingale representation is valid

$$g(\mathbf{U}_t) = F_t + V_t \boldsymbol{\varepsilon}_t,$$

where, now, F_t and V_t are a $\mathcal{F}_{K;t-1}$ measurable $K \times K$ matrices, and $(\boldsymbol{\varepsilon}_t)_{t \in \mathbb{Z}}$ is a sequence of iid mean zero vectors with variance equal to the K dimensional identity matrix. Then, $\Sigma_t = F_t$. The use of meta-elliptical copulae implies that $g(\mathbf{u}) = (\mathcal{Q}_\varphi\mathbf{u})(\mathcal{Q}_\varphi\mathbf{u})^T$. In many cases, $\varphi = \varphi_\gamma$ where γ is some parameter in a compact parameter space, and this may complicate matters.

Example 8 For the t -copula, the generator φ , as given in Example 7, depends on the degrees of freedom γ , i.e. \mathcal{Q}_φ depends on γ .

In these cases, we may assume a specific value for γ . Alternatively, the scaling matrix can still be obtained from Kendall's tau, say ρ_τ . Kendall's tau is a measure of dependence like the correlation coefficient, but it is independent of the marginal distributions. For the pair of random variables (U_i, U_j)

$$\rho_\tau := 4\mathbb{E}I_{\{U_j \leq U_i\}}I_{\{V_j \leq V_i\}} - 1 = \mathbb{E} \text{sign}(U_i - U_j) \text{sign}(U'_i - U'_j),$$

where (U'_i, U'_j) is an independent copy of (U_i, U_j) (Joe, 1997, for more details). It is well known (Lindskog et al., 2003, Fang et al., 2002) that for a meta-elliptical copula

$$\rho = \sin\left(\frac{\pi\rho_\tau}{2}\right), \quad (9)$$

where ρ is the usual correlation coefficient. Hence, for arbitrary meta-elliptical copulae we can always suppose that $\arcsin(\Sigma_t) = \mathbb{E}[g(\mathbf{U}_t) | \mathcal{F}_{K;t-1}]$, where the (i, j) entry of g is

$$g_{i,j}(u_i, u_j) = \rho_{\tau;i,j,t} + (V_t \boldsymbol{\varepsilon}_t)_{i,j}$$

and $\rho_{\tau;i,j,t}$ is Kendall's tau for U_{ti} and U_{tj} .

In this case, estimators of the trend $\rho_{\tau;i,j,t}$ are given by a second order U-Process

$$\hat{\rho}_{\tau;i,j,t} = \sum_{0 \leq r < s < t} w(r, s, t) \text{sign}(U_{ir} - U_{jr}) \text{sign}(U_{is} - U_{js}), \quad (10)$$

using the $\mathcal{F}_{K;t-1}$ measurable filter $(w(r, s, t))_{0 \leq r < s < t}$.

Example 9 Suppose $w(r, s, t) = \binom{t}{2}^{-1}$. This is the simplest case, where we assume no time dependence and heterogeneity and (10) reduces to an ordinary U-Statistic.

Example 10 Suppose

$$w(r, s, t) = A_t^{-1} \sum_{0 \leq r < s < t} k_h(t-r) k_h(t-s), \quad A_t := \sum_{0 \leq r < s < t} k_h(t-r) k_h(t-s),$$

where $k_h(s) := h^{-1}k(h^{-1}s)$ and $k(s)$ is a decreasing positive function. Hence, the most recent observations are given the largest weight. For $k_h(s) = (1-h)h^{|s|-1}$ with $h \in (0, 1)$ this becomes a double exponential smoothing.

Example 11 Suppose

$$w(r, s, t) = A_t^{-1} \sum_{0 \leq r < s < t} k_h(U_{it-1} - U_{ir}) k_h(U_{jt-1} - U_{is})$$

and

$$A_t := \sum_{0 \leq r < s < t} k_h(U_{it-1} - U_{ir}) k_h(U_{ir-1} - U_{ir}),$$

where k_h is as in the previous example. In this case observations closest in value to (U_{it-1}, U_{jt-1}) are given the highest weight.

The results in Goshal et al. (2000) and Sancetta (2004) can be used to study the properties of local versions of Kendall's tau. However, when φ is fully known, we may avoid working with a second order U-Process and resume the simple averaging approach with an $\mathcal{F}_{K;t-1}$ measurable filter $(w(s,t))_{1 \leq s < t}$, i.e.

$$\hat{\Sigma}_t = \sum_{s \leq t-1} w(s,t) \mathcal{Q}_\varphi \mathbf{U}_i \mathcal{Q}_\varphi \mathbf{U}_i,$$

Whatever the method employed, any additional parameter can be estimated subsequently once $\hat{\Sigma}_t$ has been estimated by direct method or Kendall's tau.

2.4 Optimisation of the Linear Filter

The linear filters $(w(s,t))_{0 \leq s < t}$ and $(w(r,s,t))_{0 \leq r < s < t}$ are \mathcal{F}_{t-1} measurable, but may depend on some parameters.

Example 12 Suppose $w(s,t) = A_t^{-1} k_h(t-s)$ where $A_t := \sum_{0 \leq s < t} k_h(t-s)$, $k_h(s) = h^{-1} k(h^{-1}s)$. Then the filter depends on the smoothing parameter h .

Example 13 Suppose $(w(s,t))_{0 \leq s < t} = (w(s,t))_{h \leq s < t} = \left(\sum_{h \leq r < t} Z_{r,h} Z_{r,h}^T \right)^{-1} \sum_{0 \leq r < t} Z_{r,h} X_r$ where $Z_{r,h} = (X_{r-1}, \dots, X_{r-h})$ (In this case, it is more convenient to give the whole filter). Then, the filter is just a linear projection on the h past values, the autoregressive order.

In the above examples, the filter depends on some unknown parameter h . Ideally, the filtering parameters should minimise the forecast error

$$\mathcal{E}_n(h) := \sum_{1 \leq s \leq n} \mathcal{R}(f_s - \hat{f}_{s,h}),$$

where $\hat{f}_{s,h}$ is the estimated trend at time $s-1$ (where the filter depends on some finite parameter vector h), f_s is the true trend and $\mathcal{R}(x)$ is a convex function of x . Clearly, f_s is unobservable, and has to be replaced by $g(X_s)$ suggesting minimisation of the estimated forecast error

$$\mathcal{E}'_n(h) := \sum_{1 \leq s \leq n} \mathcal{R}(g(X_s) - \hat{f}_{s,h}). \quad (11)$$

Define $\tilde{h} = \arg \min \mathcal{E}_n(h)$ and $\hat{h} = \arg \min \mathcal{E}'_n(h)$. Cheng et al. (2003) show that with probability going to one, $\mathcal{E}_n(\tilde{h})$ performs as well as $\mathcal{E}'_n(\hat{h})$ when $\mathcal{R}(x) = x^2$, or $|x|$, under regularity conditions. Hence the parameters in the filter can be estimated as the ideal choice of filtering parameters. Clearly, this does not say anything about the consistency of the empirical criterion function.

Example 14 Suppose $(X_t)_{t \in \mathbb{Z}}$ is an IGARCH(1,1) process,

$$X_t = \sigma_t \eta_t, \quad \sigma_t = \omega + \alpha X_{t-1}^2 + (1 - \alpha) \sigma_{t-1},$$

with $(\eta_t)_{t \in \mathbb{Z}}$ iid mean zero, variance one and $(\omega, \alpha) \in \mathbb{R}_+^2$. We write

$$\begin{aligned} X_t^2 &= E(X_t^2 | \mathcal{F}_{t-1}) + X_t^2 - E(X_t^2 | \mathcal{F}_{t-1}) \\ &= f_t + v_t \varepsilon_t \end{aligned}$$

and use exponential smoothing as in Example 2 to estimate f_t . Then, for $\mathcal{R}(x) = x^2$, Theorem 1 in Zaffaroni (2004) gives $\sup_{h \in (0,1)} \mathcal{E}'_n(h) / n \xrightarrow{a.s.} 0$, so that the empirical mean square error $\mathcal{E}'_n(h) / n$ provides a meaningless criterion for estimation of h .

Finally, for forecasting purposes, we may want h at time t to be chosen using $\mathcal{E}'_{t-1}(h)$. No new concept is necessary for this and all the previous remarks apply directly in this case as well.

3 Inference

The above hierarchical approach is useful for conducting inference and assessing how appropriate the chosen models are. We can test the validity of the forecasting model in three stages: 1. unconditional marginal validity, 2. marginal dynamics validity, 3. multivariate validity.

For all three stages, we suppose that we have a random sample X_{1k}, \dots, X_{nk} , $k = 1, \dots, K$ and their associated marginals. (In this section, it is more convenient to use a different indexing for the random variables, i.e. not $0, \dots, t-1$, but $1, \dots, n$). Using, (5) we obtain U_{1k}, \dots, U_{nk} , $k = 1, \dots, K$. When the true marginals are unknown we obtain $\hat{U}_{1k}, \dots, \hat{U}_{nk}$, $k = 1, \dots, K$, which uses the estimated marginals in (5).

At all stages both visual inspection and formal inference will have to be carried out. When working with distributions, an important tool is the empirical distribution of the series, which is a benchmark over which to test the validity of a postulated model. By the Kolmogorov-Smirnov result, the empirical distribution converges to the true one exponentially fast (e.g. Vapnik, 1998, for details). The result holds for the multivariate case under stationarity and weak dependence conditions (Kim, 1999, for regularity conditions and details, though it appears that these results can be further weakened).

3.1 The Empirical Copula and the Double Uniform Transform

Suppose $(\mathbf{U}_t)_{t \in \mathbb{Z}}$ is a sequence of stationary uniform $[0, 1]^K$ random vectors with copula C . The empirical copula for a sample of n observations is given by $\mathbb{P}_n I\{\mathbf{U}_t \leq \mathbf{u}\}$, i.e. the empirical expectation of the indicator of the set $\{\mathbf{U}_t \leq \mathbf{u}\}$ for $\mathbf{u} \in [0, 1]^K$. We state a convergence result of Rio (2000, Theorem 7.3) for strongly mixing sequences. Strongly mixing is a weak dependence condition (e.g. Doukhan, 1994, for an exhaustive study).

Theorem 15 *Suppose $(\mathbf{U}_t)_{t \in \mathbb{Z}}$, is a strongly mixing stationary sequence of uniform random variables in $[0, 1]^K$ with copula C . Suppose the strong mixing coefficients are summable. Then,*

$$\xi_n(\mathbf{u}) := \sqrt{n} (\mathbb{P}_n I\{\mathbf{U}_t \leq \mathbf{u}\} - C(\mathbf{u}))$$

converges weakly to a mean zero Gaussian process with a.s. continuous sample paths with covariance function

$$\sum_{t \in \mathbb{Z}} \text{cov}(I\{\mathbf{U}_0 \leq \mathbf{u}\}, I\{\mathbf{U}_t \leq \mathbf{u}'\}).$$

Remark 16 *The result would require C to be continuous, but this is automatically satisfied because copulae are Lipschitz continuous. It can be shown that similar results hold for weaker dependence conditions.*

Remark 17 *The blockwise bootstrap can be applied to improve inference and obtain an estimator of the covariance function (Peligrad, 1998, establishes a.s. consistency of the blockwise bootstrap of Kunsch, 1989, under the conditions of the Theorem above; Radulovic, 2002 for a survey of the bootstrap in these situations).*

The limiting Gaussian distribution allows us to use standard inferential procedures under dependence. However, notice that the Gaussian process is nonstationary.

For testing purposes, we would use $\hat{U}_{1k}, \dots, \hat{U}_{tk}$, which is obtained from (5), replacing the true marginals with the estimated ones. It can be shown that a result similar to Theorem 15 applies when the marginals are estimated, but requires weak convergence of $\hat{U}_{1k}, \dots, \hat{U}_{nk}$ to Gaussian random variables (otherwise the convergence is to the sum of a Gaussian process with some random elements). The statistical properties of (5) when we use estimated marginals are difficult to establish in the general framework considered in this paper. For this reason it can be more sensible to use a further transformation via the empirical distribution, i.e. $\mathbb{P}_n I \left\{ \hat{U}_{tk} \leq u \right\}$. Let \hat{H}_k and $\hat{H}_{k,n}$ be the marginal and empirical distribution of the stationary sequence $\hat{U}_{1k}, \dots, \hat{U}_{nk}$. We suppose that these marginals are time invariant. Define $\mathcal{Q}\mathbf{u} = \left(\hat{H}_1^{-1}(u_1), \dots, \hat{H}_K^{-1}(u_K) \right)$ and $\mathcal{Q}_n\mathbf{u} = \left(\hat{H}_{1,n}^{-1}(u_1), \dots, \hat{H}_{K,n}^{-1}(u_K) \right)$. It is convenient to define $\hat{U}_{1k}, \dots, \hat{U}_{nk}$ as a segment of the sequence

$$\left(\hat{\mathbf{U}}_t \right)_{t \in \mathbb{Z}} := \left(\hat{F}_{t1}(X_{t1}, V_{t1}), \dots, \hat{F}_{tK}(X_{tK}, V_{tK}) \right)_{t \in \mathbb{Z}},$$

where \hat{F}_{tk} is an estimator for \tilde{F}_{tk} in (5). Then we apply the empirical uniform transform to $\left(\hat{\mathbf{U}}_t \right)_{t \in \mathbb{Z}}$, i.e. we replace $\left(\hat{\mathbf{U}}_t \right)_{t \in \{1, \dots, n\}}$ with $\left(\hat{\mathbf{U}}_t^n \right)_{t \in \{1, \dots, n\}}$, where $\hat{\mathbf{U}}_t^n = \mathcal{Q}_n^{-1} \hat{\mathbf{U}}_t$. For an arbitrary sequence of random variables, not necessarily uniform, the empirical copula is defined as $\hat{C}_n(\mathbf{u}) = \mathbb{P}_n I \left\{ \mathcal{Q}_n^{-1} \hat{\mathbf{U}}_t \leq \mathbf{u} \right\}$. Then, we have the following.

Theorem 18 *Under the Conditions of Lemma 24 (in the Appendix),*

$$\xi_n(\mathbf{u}) = \sqrt{n} \left(\hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \right)$$

converges weakly to a Gaussian process $G(\mathbf{u})$ with a.s. continuous sample paths such that

$$G(\mathbf{u}) = B_K(\mathbf{u}) - \sum_{k=1}^K C_{\cdot, k}(\mathcal{Q}^{-1}\mathbf{u}) B(u_k), \quad (12)$$

where $B_K(\mathbf{u})$ is a K dimensional Brownian bridge and $B(u_k)$ is a one dimensional Brownian bridge such that

$$\mathbb{E} B_K(\mathbf{u}) B_K(\mathbf{u}') = \sum_{t \in \mathbb{Z}} \text{cov} \left(I \left\{ \hat{\mathbf{U}}_0 \leq \mathbf{u} \right\}, I \left\{ \hat{\mathbf{U}}_t \leq \mathbf{u}' \right\} \right)$$

$$\mathbb{E} B(u_k) B(u'_k) = \sum_{t \in \mathbb{Z}} \text{cov} \left(I \left\{ \hat{U}_{0k} \leq u_k \right\}, I \left\{ \hat{U}_{tk} \leq u'_k \right\} \right).$$

Several testing procedures based on the empirical copula have been devised, where the marginals are unknown, but estimated by the empirical distribution (e.g. van der Vaart and Wellner, 2000, Genest and Remillard, 2003). These procedures require some continuity for the marginals. The framework we have just outlined allows for situations in which the distributions may change over time and may have discrete components in their Lebesgue decomposition. Hence the empirical distribution might not be directly estimated, and continuity conditions might not be satisfied. The double transform $\mathbf{X}_t \mapsto \hat{\mathbf{U}}_t$ using (5) and $\hat{\mathbf{U}}_t \mapsto \hat{\mathbf{U}}_t^n$ using $\hat{\mathbf{U}}_t^n = \mathcal{Q}_n^{-1} \hat{\mathbf{U}}_t$ solves these problems, and the empirical copula is defined as $\hat{C}_n(\mathbf{u}) = \mathbb{P}_n I \left\{ \hat{\mathbf{U}}_t^n \leq \mathbf{u} \right\}$. Then, we are in the framework of the above authors.

3.2 Test Statistics

Suppose we were assuming that the marginals were known. In practice this assumption means that we would be using $\mathbb{P}_n I \left\{ \hat{\mathbf{U}}_t \leq \mathbf{u} \right\}$ instead of $\mathbb{P}_n I \left\{ \mathbf{U}_t \leq \mathbf{u} \right\}$, i.e. we would be treating our estimated marginals as if they were the true ones. Then, under stationarity of $\left(\hat{\mathbf{U}}_t \right)_{t \in \mathbb{Z}}$ it can be shown that the convergence to (12) will have to be replaced by

$$\sqrt{n} (C_n(\mathbf{u}) - C(\mathbf{u})) - \sqrt{n} \sum_{k=1}^K C_{,k}(\mathbf{u}) \left(\hat{U}_{tk} - u_k \right),$$

where \hat{U}_{tk} is the k^{th} element of $\hat{\mathbf{U}}_t$ (result available from the first author upon request). Then, assuming the marginals to be known when this is not the case implies disregarding the terms in the summation in the above display. That is, we would conduct inference using Theorem 15 on the first term. The practical implication is an increase of the probability of a type one error: rejection of the null of correctly specified copula may imply that the marginals were not correctly specified. This does no harm in our case; therefore, for simplicity (and to ease notation), it will be assumed that the marginals are known, unless we are considering the univariate case.

Theorems 15 and 18 imply ξ_n is asymptotically Gaussian. Then, it is natural to consider the following the test statistics

$$\left[\int_{[0,1]^K} \left| \frac{\xi_n(\mathbf{u})}{\sqrt{\sigma^2(\mathbf{u}, \mathbf{u})}} \right|^p d\mathbf{u} \right]^{\frac{1}{p}}, \quad (13)$$

for some $p \in [1, \infty)$ or

$$\sup_{\mathbf{u} \in [0,1]^K} \left| \frac{\xi_n(\mathbf{u})}{\sqrt{\sigma^2(\mathbf{u}, \mathbf{u})}} \right|, \quad (14)$$

for $p = \infty$. Whatever the value of p , the distribution of the L_p norm of the limiting Gaussian process is unknown, but can be easily simulated or approximated by the bootstrap.

3.3 First Step: Marginal Model Validity

Since many inferential procedures (e.g. tests for goodness of fit) require the true distribution to be continuous, (5) allows us to extend all known inferential procedure for continuous distribution functions to the case of discontinuous ones. We found no mention of this extension in the literature. Even more important, if $(X_t)_{t \in \mathbb{Z}}$ were non-identically distributed, many inferential procedures would fail as all of them require some sort of stationarity. On the other hand, correct specification of $\{P_\theta, \theta \in \Theta\}$ and proper estimation of the parameters θ , allows us to use the uniform transforms based on (5). Then, the empirical distribution of the transformed random variables converge to the uniform distribution under the null of properly specified marginals. A pleasant feature is that under the null the limiting distribution is independent of estimated parameters.

3.4 Second Step: Marginal Dynamics Validity

If the null of properly specified marginal distributions is not rejected, we still need to check that any time dependence has been properly dealt with. In the spirit of inspection of the autocorrelation function for time series, we can estimate the uniform autocorrelation function, i.e. the

autocorrelation function of $(W_t)_{t \in \mathbb{Z}} = (U_t - 1/2)_{t \in \mathbb{Z}}$ and powers of it, where the sequence $(W_t)_{t \in \mathbb{Z}}$ is centered around zero which is both its mean and median under a uniform $[0, 1]$ random variables assumption for $(U_t)_{t \in \mathbb{Z}}$.

Example 19 Suppose $(X_t)_{t \in \mathbb{Z}}$ is a stationary sequence of random variable with time correlated absolute values. Then, inspection of the autocorrelation of $(|X_t|)_{t \in \mathbb{Z}}$ or $(X_t^2)_{t \in \mathbb{Z}}$ is conceptually related to inspection of the autocorrelation of $(|W_t|)_{t \in \mathbb{Z}}$.

Calculation of the autocorrelation function of $(W_t)_{t \in \mathbb{Z}}$ is equivalent to computing Spearman's rho for different lags.

Notice that $(U_t)_{t \in \mathbb{Z}}$ is a sequence of uniform random variables in $[0, 1]$, hence its finite dimensional distributions are completely determined by its moments and comoments (e.g. Theorem 30.1 in Billingsley, 1995). Since these moment must be decreasing to zero exponentially fast (because they are mean values of variables in $(0, 1)$), we can restrict inference to a finite number of powers of W_t . In this case, independence of a sufficiently large number of comoments would imply independence.

Example 20 We can compute the sample autocorrelation of $(W_t^a)_{t \in \mathbb{Z}}$ $a \geq 1$. Under the null of no autocorrelation, the sample autocorrelation function, say $\rho_n(s)$ $s = o(n)$ is asymptotically Gaussian with mean zero and variance $1/n$ for a sample of size n .

Example 21 We can run a regression of W_t on its lagged values and powers of its lagged values. Then an F -test can be used to test for independence.

Some authors (e.g. Deibold et al., 1998, Patton, 2005) proposed similar analyses of an equivalent version of $(W_t^a)_{t \in \mathbb{Z}}$ with $a = 1, 2, 3, 4$.

A formal test for the null of no autocorrelation at all lags can be carried out using the following.

Theorem 22 Suppose $(X_t)_{t \in \mathbb{Z}}$ is a sequence of iid random variables with mean zero and variance one. Suppose $t_n \in \mathbb{N}$, $t_n = O(n^{1/(5+\epsilon)})$, $\forall \epsilon > 0$. Let N be a standard normal random variable. Setting $\lambda = t_n \Pr(|N| \geq z) \leq b < \infty$ (where $b \in \mathbb{R}_+$),

$$\left| \Pr \left(\max_{1 \leq s \leq t(n)} \sqrt{n} |\rho_n(s)| \leq z \right) - \exp\{-\lambda\} \right| = O(t_n^{-1}).$$

More general inference can be carried out as follows. Set $(\mathbf{U}_t)_{t \in \mathbb{Z}} = (U_t, U_{t-1}, \dots, U_{t-l})_{t \in \mathbb{Z}}$. For a test of time series independence, under the null of no independence, we set $C(\mathbf{u}) = \prod_{t=1}^l u_t$, and the limiting Gaussian process, say $G(\mathbf{u})$, has variance function $\sigma^2(\mathbf{u}, \mathbf{u}) = \prod_{t=1}^l u_t (1 - u_t)$. We construct the test statistic

$$\mathcal{M}_{(0,1)^K}^\xi := \sup_{\mathbf{u} \in (0,1)^K} \left| \frac{\xi_n(\mathbf{u})}{\sigma^2(\mathbf{u}, \mathbf{u})} \right|,$$

and use Theorem 15 together with the continuous mapping theorem to infer that

$$\Pr \left(\mathcal{M}_{(0,1)^K}^\xi \leq c \right) \rightarrow \Pr \left(\mathcal{M}_{(0,1)^K}^G \leq c \right),$$

where

$$\mathcal{M}_{(0,1)^K}^G := \sup_{\mathbf{u} \in (0,1)^K} |G(\mathbf{u})|,$$

and $G(\mathbf{u})$ is the limiting Gaussian process. Then, the bootstrap can be used to compute confidence intervals. However, when the null is so simple, simulation from the independence copula $C(\mathbf{u}) = \prod_{t=1}^l u_t$ can be a better alternative.

Notice that $G(\mathbf{u})$ is nonstationary. Direct estimation of the distribution of the supremum of a nonstationary Gaussian problems poses incredible difficulties (see Adler, 2000) and complicated formulae exists only for special cases of nonstationarity (Mikhaleva and Piterbarg, 1996).

Some nonparametric multivariate tests of independence that have been proposed in the literature have simpler asymptotic distribution. For example, we could use (13) with $p = 2$. Under the assumption that the marginals are known, we can use the critical values under the null of independence that have been tabulated by Blum et al. (1961). However, tests based on (14) are more powerful on the tails. Genest and Remillard (2003) provide additional results on tests of independence.

3.5 Third Stage: Multivariate Validity

A meta-elliptical copula implies a special kind of dependence. This suggests a test for meta-elliptical dependence. Use the estimated scaling matrix for the postulated meta-elliptical copula $C_\varphi(\mathbf{u}; \Sigma)$, and from $(\mathbf{U}_t)_{t \in \mathbb{Z}} = (U_{t1}, \dots, U_{tK})_{t \in \mathbb{Z}}$ derive $(\tilde{\mathbf{U}}_t)_{t \in \{1, \dots, n\}} = (\tilde{U}_{t1}, \dots, \tilde{U}_{tK})_{t \in \{1, \dots, n\}}$ using \mathcal{Q}_φ as in (8) and the scaling matrix, as described in Section 2 just below (8). Then, under the null of properly specified dependence structure, $\tilde{\mathbf{U}}_t$ must have meta-elliptical copula with scale matrix equal to the identity (i.e. $C_\varphi(\mathbf{u}; I)$).

There are two advantages in this procedure. The testing procedure does not depend on the estimated scaling matrix under the null. In particular if φ does not depend on any estimated parameter, a test statistic under the null of correctly specified copula does not depend on unknown quantities. This is not the case with other multivariate testing procedure proposed in this context like the hit test (e.g. Diebold et al., 1998, Patton, 2005). Loosely speaking, a hit test computes the relative frequency of observations over some predefined Borel sets and compares it with the frequencies implied by the estimated model in these same sets. Therefore, a hit test requires the null to depend on the estimated parameters (i.e. we may incur a serious nuisance parameters problem, as the scaling matrix is a high dimensional parameter, and the test statistic may not have zero mean under the null). The relative frequencies are based on the construction of a histogram instead of the empirical distribution. This implies that the test depend on the choice of mesh for the histogram (i.e. the Borel sets). (For further details on why procedures based on the empirical distribution should be preferred to procedures based on estimation of probabilities of arbitrary Borel sets, see Vapnik, 1998.) These problems are avoided using the empirical copula and the methods proposed here under meta-elliptical assumption (though, the null might still depend in some cases on a low dimensional estimated parameter, e.g. a one dimensional parameter as in Example 8). However, notice that our approach is valid under meta-ellipticity only, while a hit test is not restricted to such a case..

A second advantage is the following. The cross-sectional dependence structure of $(\mathbf{U}_t)_{t \in \{1, \dots, n\}}$ might change overtime, implying that we should estimate a time varying scaling matrix Σ in (7). In this case, the empirical copula cannot be directly computed for $(\mathbf{U}_t)_{t \in \{1, \dots, n\}}$. (If the dependence structure is stationary, but still time varying, empirical estimators might still perform rather poorly.) On the other hand, $(\tilde{\mathbf{U}}_t)_{t \in \{1, \dots, n\}}$ will be characterised by a stationary depen-

dence structure and empirical testing procedures can be employed. Then

$$\xi(\mathbf{u}) := \frac{\tilde{C}_n(\mathbf{u}) - C_\varphi(\mathbf{u}; I)}{\sqrt{\sigma^2(\mathbf{u})}},$$

$$\sigma^2(\mathbf{u}) = C_\varphi(\mathbf{u}; I) (1 - C_\varphi(\mathbf{u}; I)),$$

where $\tilde{C}_n(\mathbf{u}) = \mathbb{P}_n I \left\{ \tilde{\mathbf{U}}_t \leq \mathbf{u} \right\}$ is the empirical copula of $\left(\tilde{\mathbf{U}}_t \right)_{t \in \{1, \dots, n\}}$. As above, the test statistic is defined as $\mathcal{M}_{(0,1)^\kappa}^\xi$. However, notice that for $\tilde{\mathbf{U}}_t := \left(\tilde{U}_{t1}, \tilde{U}_{t2} \right)^T$, $\text{cov}(\tilde{U}_{t1}, \tilde{U}_{t2})$ is not necessarily zero under the null (e.g. Example 7). As mentioned above, $\text{cov}(\tilde{U}_{t1}, \tilde{U}_{t2}) = 0$ for the Gaussian copula, but not for the t -copula with finite degrees of freedom.

To construct confidence intervals we either need to simulate from the copula $C_\varphi(\mathbf{u}; I)$, or use the bootstrap. The procedure is also valid in the presence of time series dependence of $(\mathbf{U}_t)_{t \in \{1, \dots, n\}}$. In this case, we shall use a block bootstrap approach for deriving the right confidence intervals. Moreover we should include the covariance terms in $\sigma^2(\mathbf{u})$ as done in Theorem 15. Again, the bootstrap can be used for this, but it would be quite cumbersome.

4 Empirical Study of Soft Commodities and Energy Prices

4.1 Data Description

In this study we use futures data on 16 different commodities over the period 3/01/93-28/01/04, which amounts to about 2700 observations excluding festivities. The data were retrieved from Bloomberg using the generic ticker for each contract including festivities. This delivers a concatenated series of front months for each contract series, where the price is the previous period price in the case of a closed market. For each commodity, the front month has a life of about one to three months. We eliminated all festivities and computed log return for all series as $R_t := \ln(P_t/P_{t-1})$. Every time the front month changed at time t , P_t and P_{t-1} referred to two different contracts (in terms of delivery). Unfortunately, contracts on different commodities postulate different delivery periods (for example soybeans and coffee have different settlement dates) so that it is difficult to avoid this problem for a large data set. However, visual inspection of our results did not reveal any clear short term seasonal pattern that might arise from the roll over of contracts.

The commodities studied are crude oil, gas oil (IPE), heating oil, natural gas, propane, unleaded gas, cocoa, coffee, sugar, orange juice, soybean, corn, rice, oats, wheat and cotton. Assuming the data possess suitable ergodic properties, we report sample summary statistics in Table I.

Table I.

Table I. Sample Summary Statistics

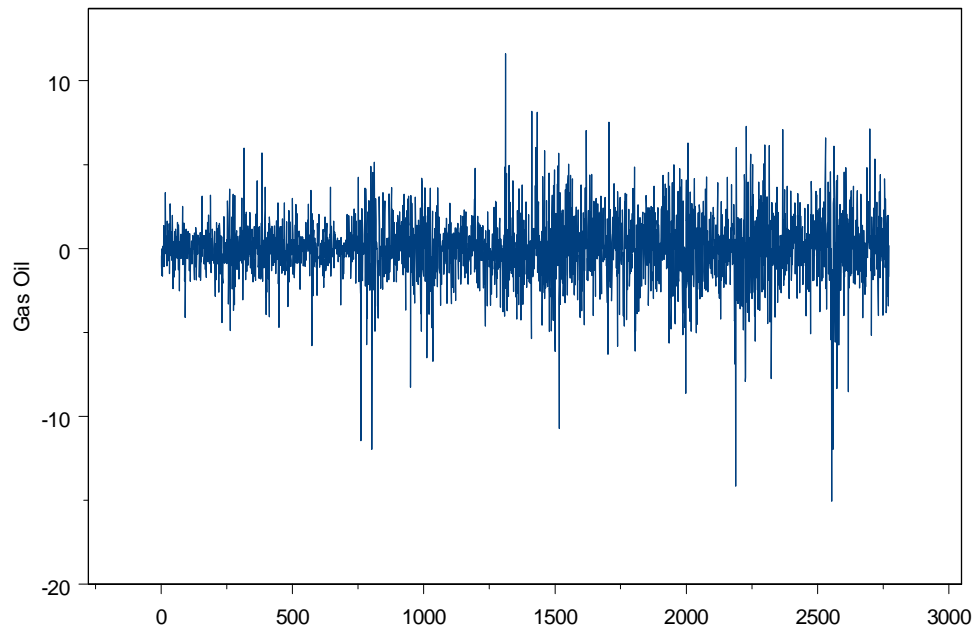
	Mean	Var	Skew	Kurt	Min	1stQu	Median	3rdQu	Max
CRUDE OIL	0.02	5.18	-0.27	4.07	-16.54	-1.15	0.00	1.29	14.23
GAS OIL (IPE)	0.01	3.96	-0.55	5.17	-15.07	-0.98	0.00	1.06	11.62
HEATING OIL	0.02	5.38	-0.95	7.35	-20.97	-1.13	0.03	1.29	10.40
NATURAL GAS	0.04	14.53	-0.21	7.84	-37.57	-1.95	0.05	2.03	32.44
PROPANE	0.03	4.14	-1.26	15.84	-24.78	-0.80	0.00	0.92	12.18
UNLEADED GAS	0.02	5.77	-0.59	7.74	-25.45	-1.27	0.10	1.41	19.49
COCOA	0.02	3.82	0.25	2.72	-10.01	-1.11	0.00	1.08	12.74
COFFEE	0.00	8.06	0.09	7.04	-22.06	-1.40	0.00	1.40	23.77
SUGAR	-0.01	4.76	-0.18	10.76	-18.04	-1.05	0.00	1.07	23.55
ORANGE JUICE	-0.01	4.25	0.72	11.28	-12.91	-0.99	0.00	0.94	22.72
SOYBEANS	0.01	1.86	-1.02	13.28	-17.43	-0.73	0.03	0.76	7.41
CORN	0.01	2.23	-1.84	36.15	-26.12	-0.74	0.00	0.76	9.13
RICE	0.01	3.56	1.03	23.15	-12.97	-0.89	0.00	0.89	28.08
OATS	0.00	4.89	-1.91	20.18	-25.46	-0.92	0.00	1.02	14.54
WHEAT	0.00	3.20	-1.39	41.86	-28.61	-1.00	0.00	0.97	23.30
COTTON	0.01	3.06	-1.45	36.48	-30.44	-0.88	0.00	0.86	13.62

The results are consistent with what is expected from fat-tailed non-Gaussian data sets. Figure I gives the time series plot for the log returns of gas oil, coffee and rice.

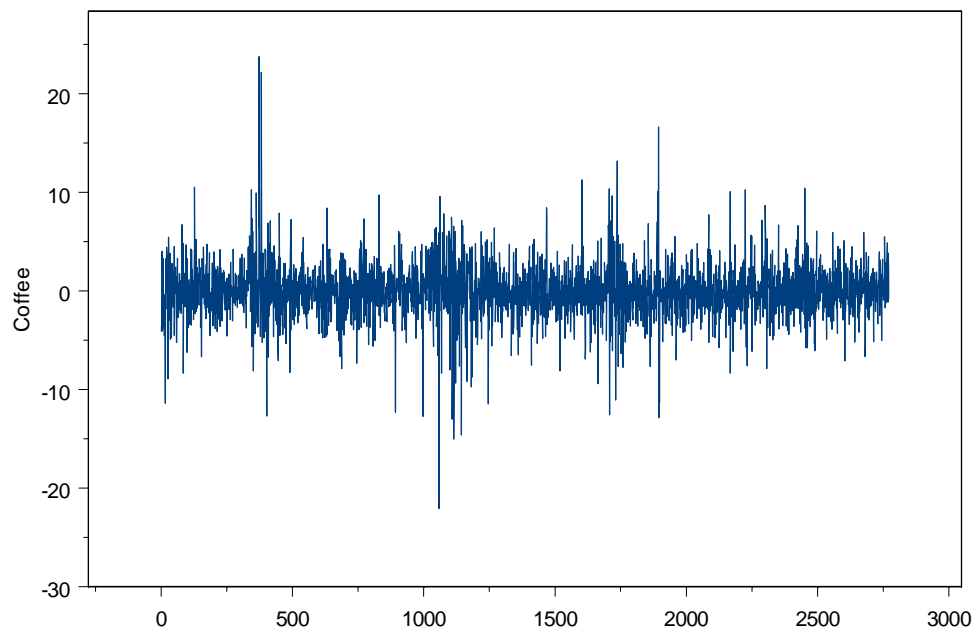
Figure I

Figure I. Time Series Plot.

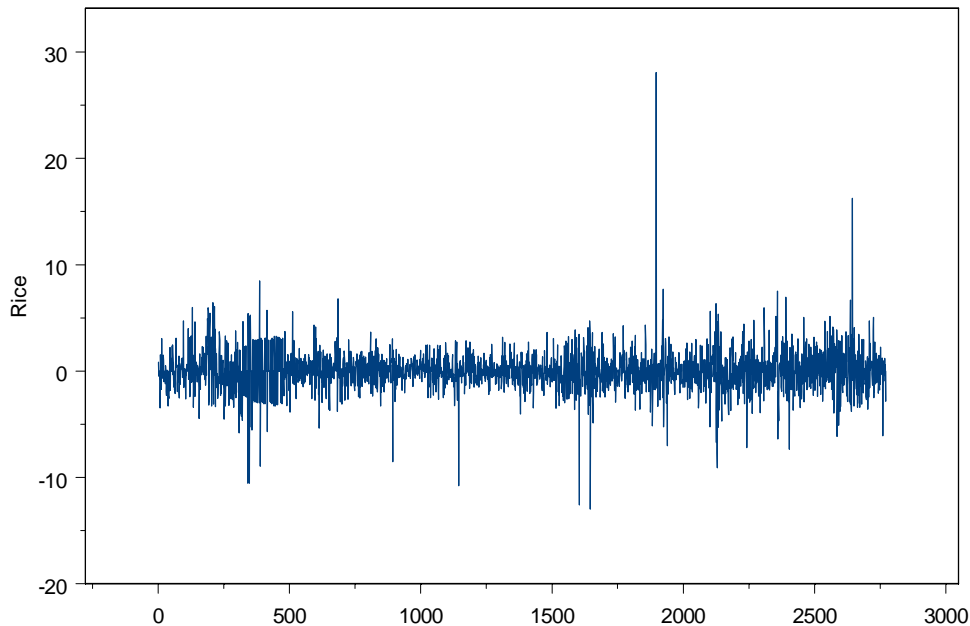
Gas Oil



Coffee



Rice



4.2 Preliminary Marginal Data Analysis

4.2.1 Tails of the absolute values of the data

Let $X := |R|$, where R is a log return. In order to get an idea of the tails of the series under study, we made a preliminary assumption of heavy tails based on the following semiparametric specification

$$\lim_{x \rightarrow \infty} \Pr(X > x) \sim x^{-\alpha} L(x), \quad (15)$$

where $\alpha > 0$ and $L(x)$ is a slowly varying function at infinity, i.e. $L(xt)/L(t) \rightarrow 1$ as $t \rightarrow \infty$ $\forall x > 0$. Under this assumption, X has a tail that is regularly varying of index $-\alpha$. This assumption is consistent with the tails of unconditional returns following a GARCH(1,1) process (Mikosch and Starika, 2000). In particular an IGARCH(1,1) has tails with $\alpha = 2$. (Similar results for more general GARCH processes have been obtained by Basrak et al., 2002)

We used the Hill's plot of the data in order to infer the possible value for α . The Hill's plot is a cross-plot of $(k \in \{1, \dots, n\}, H_k)$ where $H_k := k^{-1} \sum_{i=1}^k \ln(X_{(i)}/X_{(k)})$, $X_{(1)}, \dots, X_{(k)}$ are the k^{th} largest order statistics, $X_{(i+1)} \leq X_{(i)}$. Then, H_k is an estimator for α^{-1} . When X is not Pareto distributed, the alt (alternative) Hill's plot suggested by Stărică is more instructive (Drees et al., 2000). This is just the cross plot of $(s \in (0, 1), H_{n^s})$. Notice that the Hill's plot is consistent under dependence conditions as well (Resnick and Stărică, 1995). Hence it could be used to check the validity of GARCH tails.

By Karamata's theorem (e.g. Proposition 1.5.10 in Bingham et al., 1987) it is simple to show that the variance of X exists if and only if $\alpha > 2$. The plots showed that the more stable regions

always lie above 2. However, in some occasions, the plots are just downward sloping with no flat stable region, thus it is hard to deduce anything. Most likely, this is caused by the fact that (15) is not a good specification for the tails of the data, i.e. tails may actually be thinner than power laws. (This claim can be directly verified by looking at the Hill's plot for random variables simulated from an exponential distribution.) This is consistent with several claims found in the literature. Silvapulle and Granger (2001) argue that exponential tails can describe the tails of financial returns adequately. Laherrère and Sornette (1998) present some evidence that the tails of exchange rates are not power law. Frisch and Sornette (1997) provide some theoretical arguments showing that the tails of financial returns should be asymptotically thinner than power laws. Notice that these claims would rule out the much used t-distribution (which is a regularly varying distribution at infinity).

However, it is important to realise that in such a long series the assumption of stationarity might be rejected. Therefore, any statement based on the Hill's plot has to be taken carefully depending on the degree of the nonstationarity.

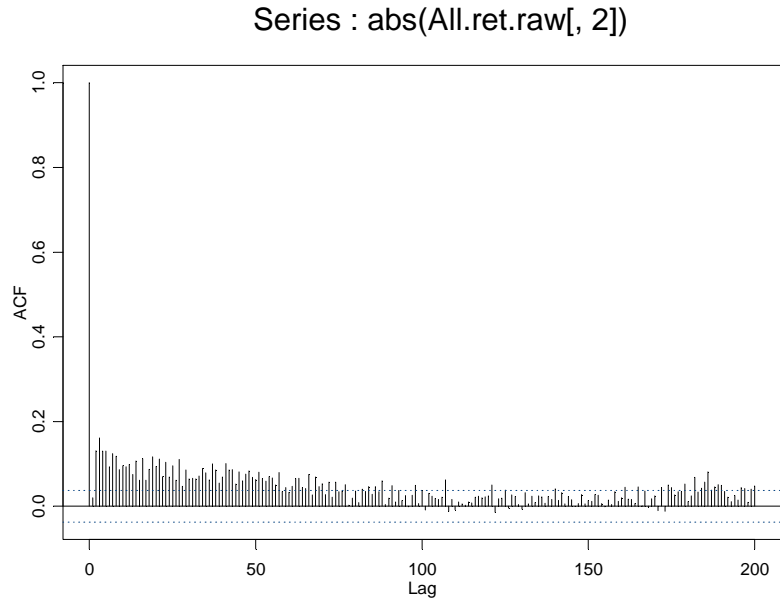
Another tool to verify the assumption of regular variation is the Stărică plot (e.g. Stărică, 1999, and Resnick, 2004, for a simplified version), but we did not pursue this route.

4.2.2 The sample ACF of the absolute values of the data.

Assuming that $\mathbb{E}|R|^2 := \mathbb{E}X^2 < \infty$, we plotted the sample ACF for $(|R_t|)_{t \in \mathbb{Z}}$. We observed a slowly decaying sample ACF for all the series. If $\mathbb{E}X^2 < \infty$ and the data are stationary, then this is a clear sign of strong time dependence for all the series. Figure II plots the sample ACF for Gas Oil. Moreover, a GARCH(1, 1) model was fitted to $(R_t)_{t \in \mathbb{Z}}$, the log returns. It turns out that for all of the commodities the estimated GARCH(1,1) coefficients sum close to 1 (or even more than one in the case of gas oil and propane), which seems to imply that past volatility can explain well current volatility. Nearly integrated GARCH indicates that either the variance or some fractional moment close to the second is not well defined. However, this is inconsistent with the results we obtained from the Hill's plot.

Figure II.

Figure II. Sample ACF Absolute Value of Gas Oil



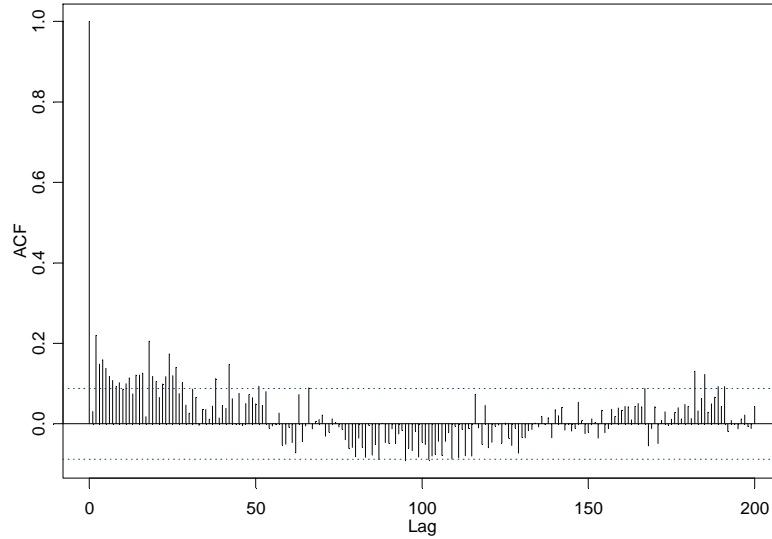
If the data are not stationary, the sample ACF will be misleading, as well as GARCH estimates (e.g. Mikosch and Stărică, 1999). Hence, we looked at subsamples of blocks of observations of sample size 500 to verify whether the sample ACF behaves similarly within each subsample. We observed strong inconsistencies of the sample ACF for all the series in different subperiod. Thus, Figure III illustrates the dramatic change in the sample autocorrelation function of gas oil's absolute returns in two sub-periods of equal length (500 observations). Similar changes were observed in the sample autocorrelations of other commodities.

Figure III.

Figure III. Sample ACF Absolute Value of Gas Oil

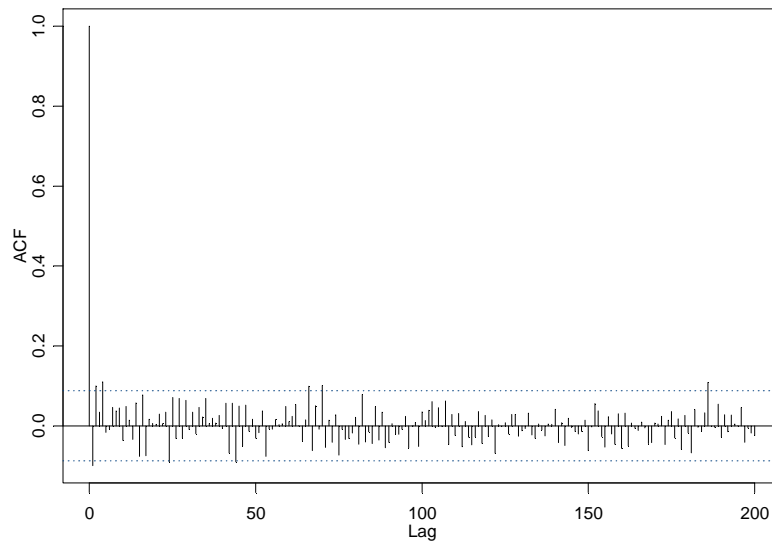
Sub-sample [500, 1000]

Series : `abs(All.ret.raw[500:1000, 2])`



Sub-sample [1500, 2000]

Series : `abs(All.ret.raw[1500:2000, 2])`



From this analysis we concluded that the data might be either nonstationary or exhibit long range dependence in powers of the absolute values. Given the long time span of the time series, we believe nonstationarity is more plausible. Thus it is appropriate to use a linear filter that gives more weight to current observations, i.e. a local time modelling approach.

4.3 A Simple Model

We propose a simple model for the forecast of the distribution of the commodities returns. As done in Granger and Stărică (2005), we decide to model the absolute returns and their sign separately. Suppose $(R_t)_{t \in \mathbb{Z}}$ is a sequence of random variables with values in \mathbb{R} . This sequence is interpreted as log returns. We use the following decomposition $R_t = \varepsilon_t |R_t|$, $\varepsilon_t = \text{sign}(R_t)$. In several occasions, our series have zero returns in a given day. Therefore, $\Pr(\varepsilon_t = 0) > 0$, and the distribution function of R_t has a discrete component at zero in its Lebesgue decomposition. The distribution of ε_t is modelled as $\Pr(\varepsilon_t = 0) = p$, $\Pr(\varepsilon_t > 0) = q_t$, so that q_t can change over time. This allows for trends in prices. Trends for commodities might be the effect of seasonal patterns and other exogenous variables like weather. Moreover, for simplicity we do not model directly the break that occurs every time there is a roll of a contract. (As mentioned above this happens every one or three months depending on the contract specification.)

Based on the argument that tails might be thinner than power laws it is ideal to use a class of marginal distributions that is flexible enough to accommodate tails that can be either thicker than exponential (sub-exponential tails) or thinner than Gaussian, by controlling a shape parameter. To this end, the distribution of $|R_t|$ is restricted to the class of Weibull distributions. Weibull distributions also exhibit the convenient feature of having a closed form analytic expression. Assuming ε_t and $|R_t|$ to be independent, the distribution of $|R_t|$ is given by

$$\begin{aligned} F_{|R(t)|}(x) &= (1-p)(1 - \exp\{-a_t x^b\}), \text{ if } x > 0 \\ &= p, \text{ if } x = 0, \end{aligned}$$

where the scale parameter a_t is allowed to be time varying, while the shape parameter b is constant. Surely, also the shape parameter might vary over time, but for estimation reasons, we prefer b to be constant. The distribution of R_t is given by

$$\begin{aligned} F_{R(t)}(x) &= (1 - q_t - p) \exp\{a_t x^b\}, \text{ if } x < 0 \\ &= p, \text{ if } x = 0 \\ &= 1 - q_t \exp\{a_t x^b\}, \text{ if } x > 0. \end{aligned} \tag{16}$$

We assume that once we model the variation in the parameters, there is no dependence left. This is a convenient assumption. Interesting, Granger and Stărică (2005) find that after accounting for time variation of the unconditional mean and variance of some financial returns, dependence was less of an issue. Our personal experience from working with financial returns suggests similar results.

Finally, to join the marginal distributions, we choose a Gaussian copula where the scale matrix is allowed to vary over time. This choice is based on three arguments. As will be shown in the cross-sectional analysis below, the data exhibit a different range of dependence, going from no dependence to positive dependence. This requires a copula that allows us to capture independence and dependence. Hopefully a time varying scaling matrix might provide a reasonable approximation to the truth. Notice that to conduct proper cross-sectional analysis in the context of time varying parameters for the marginals, it is preferable to analyse the crosscorrelation structure of the time invariant uniform transforms. Hence, preliminary results on cross-sectional analysis cannot not be reported here before estimation of the marginals, which is carried out below.

Second, inferential procedures with the Gaussian copula do not depend on nuisance parameters. Hence it is appropriate to use it as a benchmark model, and to look for better alternatives

if the null of properly specified copula is rejected. Third, this copula is very simple to estimate. Given lack of more detailed information, but only the scaling matrix, it is known that the Gaussian distribution is optimal in an information theoretic sense (i.e. it is maximum entropy).

4.3.1 Estimation

A convenient feature of (16) is that q_t and p can be estimated separately from the other parameters. A time invariant estimator for p is $\mathbb{P}_n I_{\{R(t)=0\}}$. Time varying parameters are obtained as follows. Define $k_h(s) = h^{-1}k(h^{-1}|s|)$, where $k(s)$ is a decreasing function of its argument. Assuming that the parameters of interest admit the representation in (1), we will use a dynamic filter obtained by solving the quadratic function

$$\min_{f(t)} \sum_{0 \leq s < t} |g(x_s) - f_t|^2 k_h(t-s).$$

This implies that the filter at time t is given by

$$w_h(s, t) = k_h(t-s) / \left(\sum_{0 \leq s < t} k_h(t-s) \right), \quad (17)$$

which is a filter for a local constant fit (e.g. Fan and Gijbels, 1996). Notice that the weights in exponential smoothing are a special case of (17). Suppose n is the total number of observations. Then we define

$$k_h(s) = h^{-1} \left[1 - (|t-s|/h)^2 \right] I \{0 < (t-s)/n < h\}. \quad (18)$$

For simplicity, the parameter h in the filter will be chosen using the whole sample of observations, i.e. minimising \mathcal{E}'_n as in (11).

Based on the methods outlined in Section 2,

$$\hat{q}_h(t) := \sum_{0 \leq s < t} w_h(s, t) I_{\{R_s > 0\}}.$$

The parameter h is chosen minimising the absolute prediction error (i.e. (11) with $\mathcal{R}(x) = |x|$). To simplify the estimation of the parameters in $F_{|R(t)|}(x)$, we reparametrise as in Example 5: $a_t = \gamma_t^{-b}$, so that a_t is time varying only through γ_t , which is supposed to be a measurable function with values in \mathbb{R}_+ . By direct calculation, we have that $\gamma_t \kappa = \mathbb{E}(|R_t| | \mathcal{F}_{t-1})$, where

$$\kappa = (1-p) \Gamma \left(\frac{1}{b} + 1 \right), \quad (19)$$

and simply define $m_t := \gamma_t \kappa$, and our estimator

$$\hat{m}_h(t) := \sum_{0 \leq s < t} w_h(s, t) |R_s|, \quad (20)$$

where h minimises the following square prediction error

$$\min_h \sum_{s \leq n} \left| \frac{|R_s|}{\hat{m}_h(t)} - 1 \right|^2. \quad (21)$$

The choice of (21) is based on our desire to avoid large variability in the estimated standardised returns $\left| \tilde{R}_t \right| := |R_t| / \mathbb{E}|R_t|$. By assumption $\left| \tilde{R}_t \right|$ has distribution function

$$\begin{aligned} F_{\left| \tilde{R}(t) \right|}(x) &= (1-p)(1 - \exp\{-\mu x^b\}), \text{ if } x > 0 \\ &= p, \text{ if } x = 0, \end{aligned}$$

where $\mu := \kappa^b$, κ as in (19), and all the parameters are identified. Since \hat{m}_t is a biased estimator for $\mathbb{E}(|R_t| | \mathcal{F}_{t-1})$ unless $h \rightarrow 0$ and $nh \rightarrow \infty$, we do not restrict $\mu = \kappa^b$, but we leave it as a free parameter. The parameters can be estimated by maximum likelihood using estimated $\left| \tilde{R}_t \right|$ from $|R_t| / \hat{m}_t$.

Using (5) and the estimated marginals at time t , we derive $\hat{\mathbf{U}}_t$. Then, the scaling matrix for the Gaussian copula is estimated by

$$\hat{\Sigma}_h(t) := \sum_{0 \leq s < t} w_h(s, t) \mathcal{Q}_\varphi \hat{\mathbf{U}}_s \mathcal{Q}_\varphi \hat{\mathbf{U}}_s^T, \quad (22)$$

where \mathcal{Q}_φ is as in (8) and φ is as in Example 6. The estimator in (22) may give values outside $[-1, 1]$ whenever the h parameter is small. Define $\hat{\sigma}_{kl}$ to be the (k, l) entry in $\hat{\Sigma}_t$. We replace $\hat{\sigma}_{kl}$ with $\hat{\sigma}_{kl} / (\hat{\sigma}_{kk} \hat{\sigma}_{ll})^{1/2}$, so that the diagonal entries of the estimator are all one and the off diagonal entries are between one and minus one. The filter parameter is obtained by likelihood minimisation, i.e. h is chosen to minimize minus the loglikelihood of the copula

$$\min_h \mathbb{P}_n(-\ln c_g(\mathbf{U}_s; \Sigma_h(t))).$$

If h is fixed within groups of commodities, this is a natural approach in high dimensions.

Other approaches for adaptive estimation of h could be used. However, attention will be confined to the above.

4.3.2 Comments Related to the Proposed Model

The model proposed for this empirical study is dictated by the preference for parsimonious modelling, still allowing for some time varying coefficients.

It is reasonable to believe that the series exhibit some nonstationarity (i.e. the parameters are unconditionally time varying). Our approach does not identify regions of homogeneity (as Example 4 suggests), but uses unequal weights in the averaging procedure (in the spirit of Example 2). This choice is based on the following considerations. The series might exhibit both time dependence and nonstationarity. Whether, both time dependence and nonstationarity are fundamental features of the series would require more detailed analysis. In any case, a dynamic filter appears adequate. However, it is plausible to assume several structural breaks on top of trends, due to changing weather conditions, cartels, etc.. If we do not aim at estimating the time of the breaks, a filter that gives more weight to the present than the past allows to adjust quickly to the new structure of the data. This model can be seen as an extension of RiskMetrics.

Filters obtained from local linear estimation are usually preferred to the ones obtained from local constant estimation because they correct at boundary points (e.g. Fan and Gijbels, 1996). However, the remarks in Ruppert and Wand (1994, Remark 4) show that there can be a large increase in the variance of the local linear estimator close to the boundary. When we tried to fit a local linear term instead of a local constant, the estimates were found to be very volatile. For this reason the filter arising from a local constant fit was considered.

4.4 Analysis of the Marginals

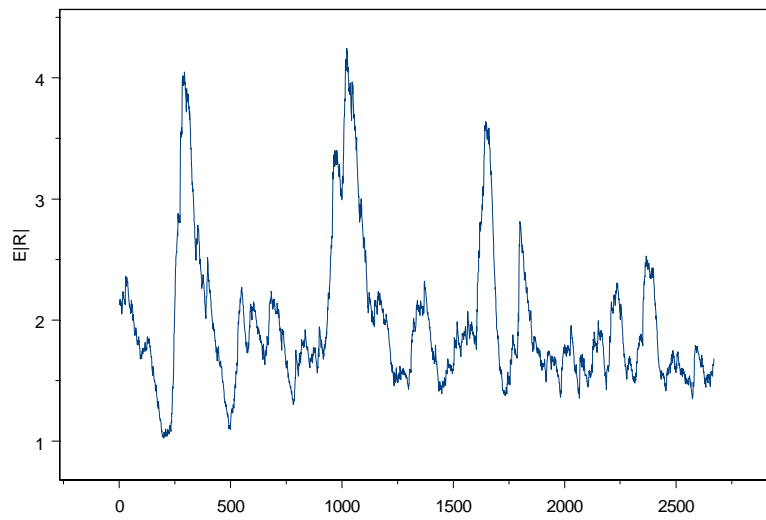
4.4.1 Local Parameter Estimation

For each series, the parameter h was chosen minimising the criterion function over the grid of values $0.01 : 0.145 (.015)$. Figure IV plots the estimated values for m_t and q_t for coffee. There is clear evidence of time varying behaviour. While the plot appears to show nonstationarity in the data, it is fair to say that a similar plot could have been produced by a stationary long memory process. However, Figure V makes the hypothesis of stationarity less plausible. Both figures suggest long term seasonal patterns, but not short term seasonal patterns.

Figure IV.

Figure IV. Local Estimates for Coffee against Time

Estimates for $\mathbb{E}|R_i|$



Estimates for $q_{t(i)}$

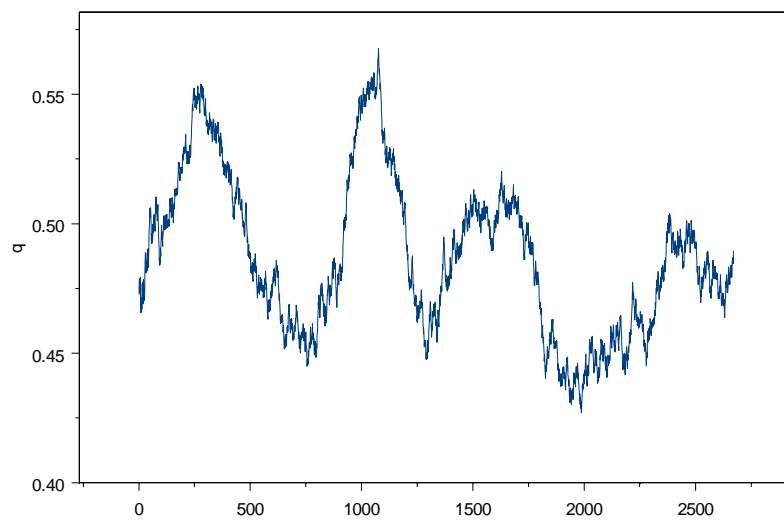
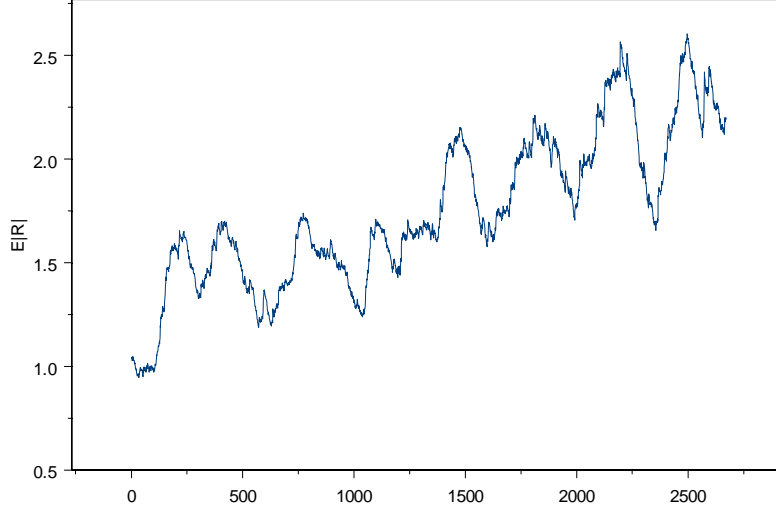


Figure V.

Figure V. Local Estimates for Gas Oil against Time
Estimates for $\mathbb{E}|R_i|$



We also conduct a formal test for unconditional parameter constancy versus unconditional time changing parameters. Define

$$\hat{\mathbf{m}}_n := \left((\hat{m}_t - \mathbb{E}\hat{m}_t) / \sqrt{\text{var}(\hat{m}_t)} \right)_{t \in \{1, 2, \dots, n\}},$$

and

$$\hat{\mathbf{q}}_n := \left((\hat{q}_t - \mathbb{E}\hat{q}_t) / \sqrt{\text{var}(\hat{q}_t)} \right)_{t \in \{1, 2, \dots, n\}}.$$

Under the null $\mathbb{E}\hat{m}_t = m$, $\mathbb{E}\hat{q}_t = q$, we compute the following test statistic

$$\psi_n^m := \max_{t \in \{1, 2, \dots, n\}} |\hat{m}_t - \mathbb{E}\hat{m}_t| / \sqrt{\text{var}(\hat{m}_t)},$$

and similarly for \hat{q}_t . We briefly discuss the statistic for \hat{m}_t , and similar remarks will apply to \hat{q}_t . Recall that \hat{m}_t is a linear combination of $(|X_s|)_{0 \leq s < t}$ with summable absolute weights (the weights sum to one by construction). It is clear that

$$\mathbb{E}\hat{m}_t = \mathbb{E}|X_t|, \text{var}(\hat{m}_t) = \sum_{0 \leq r, s < t} w(r, t) w(s, t) \text{cov}(X_s, X_r).$$

Then, $\mathbb{E}|X_t|$ and $\text{cov}(X_s, X_r)$ are obtained by their empirical counterparts. From (18) we see that there is only a hn number of nonzero weights. Therefore, we only need a consistent estimate of the covariance function of $(|X_t|)_{t \in \mathbb{Z}}$ at hn number of lags. This requires ergodicity assumptions.

Suppose the marginal distributions of $\hat{\mathbf{m}}_n = (\hat{\mathbf{m}}_{n,1}, \dots, \hat{\mathbf{m}}_{n,n})$ to be known, and similarly for $\hat{\mathbf{q}}_n$. Then, for $\lambda = \sum_{t=1}^n \Pr(|\hat{\mathbf{m}}_{n,t}| > z)$,

$$\Pr(\psi_n^m \leq z) \simeq \exp\{-\lambda\},$$

where the error in the approximation is given in Holst and Janson (1990, Theorem 2.1), and decreases exponentially in λ , but it is difficult to precisely compute (except for specific cases).

We assume a Gaussian distribution (i.e. $\lambda = n \Pr(|N| > z) < \infty$, where N is a standard normal random variable) and report results for the above test in Table II. Under the Gaussian assumption, the null of unconditional parameter constancy is rejected for several cases. In the case when the null is not rejected, visual inspection of \hat{m}_t may still reveal a trend. For example, in the case of gas oil the null is not rejected, but Figure IV clearly shows that there is a trend in $(|R_t|)_{t \in \mathbb{Z}}$.

Table II.

Table II. Test Parameter Constancy

	m	q
CRUDE OIL	3.08	2.52
GAS OIL (IPE)	3.50	3.98
HEATING OIL	3.88	3.37
NATURAL GAS	4.91 ***	3.11
PROPANE	4.74 ***	2.42
UNLEADED GAS	2.48	3.42
COCOA	2.89	3.67
COFFEE	4.22 *	3.26
SUGAR	2.63	3.67
ORANGE JUICE	2.96	5.47 ***
SOYBEANS	5.72 ***	3.30
CORN	6.74 ***	4.06
RICE	4.90 ***	3.44
OATS	5.61 ***	2.91
WHEAT	7.39 ***	4.69 ***
COTTON	2.70	4.37 **

* rejected at 90%

** rejected at 95%

*** rejected at 99%

Remark 23 Notice that rejection of the null may also be caused by long range dependence.

4.4.2 Results for local estimation.

We fit the distribution in (16). If the data were normally distributed, we should expect the parameter b to be close to two. Table III shows that this is not the case. For all commodities, b is slightly above one, implying tails fatter than a Gaussian, but thinner than an exponential.

Table III.

Table III. Parameters

Distribution of \tilde{R}_i

	mu	b	p
CRUDE OIL	0.90	1.17	0.01
GAS OIL (IPE)	0.77	1.25	0.10
HEATING OIL	0.91	1.13	0.01
NATURAL GAS	0.89	1.15	0.01
PROPANE	0.71	1.27	0.12
UNLEADED GAS	0.91	1.16	0.01
COCOA	0.86	1.24	0.03
COFFEE	0.92	1.11	0.02
SUGAR	0.91	1.15	0.03
ORANGE JUICE	0.98	1.06	0.02
SOYBEANS	0.87	1.19	0.02
CORN	0.81	1.22	0.06
RICE	0.80	1.19	0.07
OATS	0.80	1.14	0.08
WHEAT	0.85	1.25	0.03
COTTON	0.93	1.11	0.01

We transformed data into uniform using (4) and the estimated marginals. Then, the Kolmogorov Smirnov (K-S) statistic and the Anderson Darling (A-D) statistics were computed for all 16 commodities. Results are in Table IV from which we infer that the model fits the data well.

Table IV.

Table IV. Test for Goodness of Fit

	A-D Stat	K-S Stat
CRUDE OIL	0.99	0.01487
GAS OIL (IPE)	1.42	0.02299
HEATING OIL	1.17	0.01942
NATURAL GAS	0.77	0.01332
PROPANE	1.81	0.02163
UNLEADED GAS	1.11	0.01556
COCOA	0.89	0.01467
COFFEE	0.91	0.01322
SUGAR	1.58	0.01870
ORANGE JUICE	1.50	0.01803
SOYBEANS	1.20	0.01915
CORN	2.10 *	0.01944
RICE	3.05 **	0.03350 ***
OATS	1.58	0.01669
WHEAT	1.78	0.01927
COTTON	0.94	0.01718

* rejected at 90%

** rejected at 95%

*** rejected at 99%

For reasons to be discussed below, we decided to amend the optimal h 's for coffee, sugar, soybeans, corn, rice, oats and wheat. Therefore, we recalculated the same diagnostic tests for these marginals when the new h 's are used. Results were similar, and we recorded a slight improvement for the K-S statistic of rice: the null for rice is rejected at the 5% level, but not at the 1% level of significance.

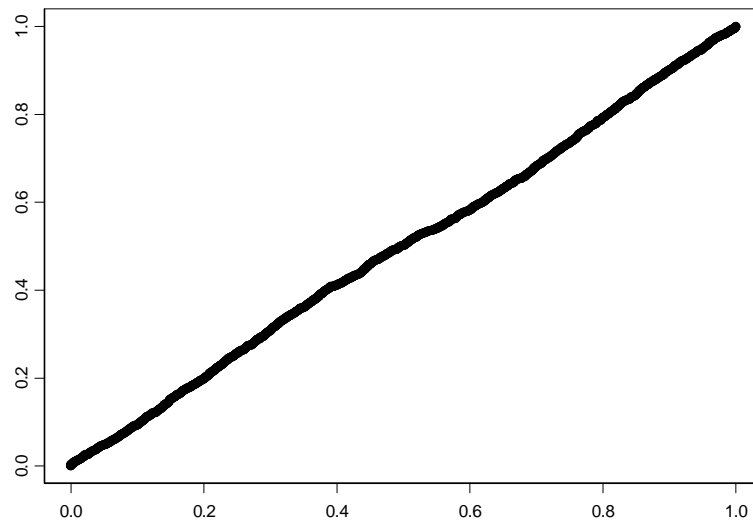
One comment pertains to the results. We evaluated 16 realizations of what asymptotically is the same process. Hence, instead of considering each series separately, we should have considered the distribution of the maximum of the process over 16 realizations. For this reason, a very large number for only one of the statistics (rice) should not be of much concern to us.

To provide visual evidence of the performance of the model, Figure VI reports the QQ-plots for the uniform transform of gas oil, coffee and rice (the worse performing commodity according to our statistics). The graphs show a remarkably good fit for the first two. (This claim can be verified by visual inspection of the QQ-plot of simulated uniform random variables.)

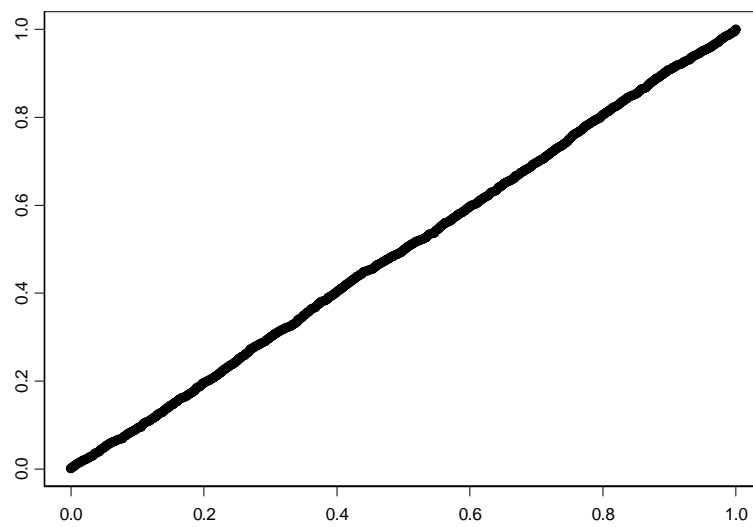
Figure VI.

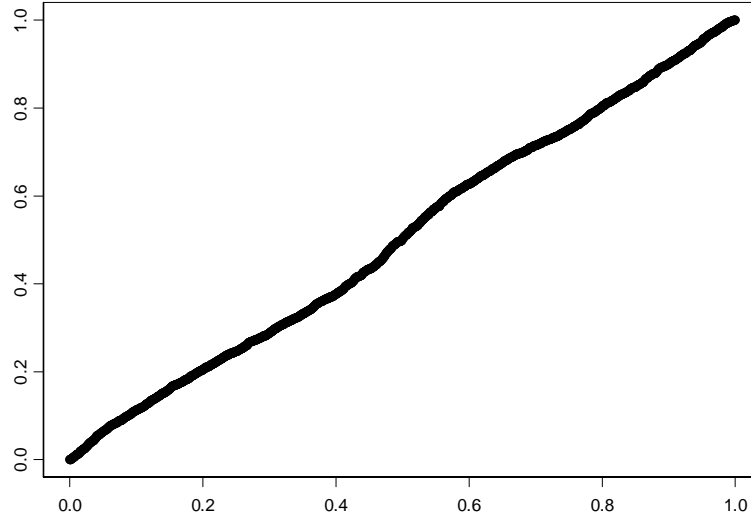
Figure V. QQ-Plot Estimated Distribution.

Gas Oil



Coffee





4.4.3 Time series dependence

Clustering effects. We compute the ACF of $(|R_t|/\hat{m}_h(t))_{t \in \{1, \dots, n\}}$ and $(|\hat{U}_t - 1/2|)_{t \in \mathbb{Z}}$ to check for clustering effects. Apart from coffee, we do not observe any persistency in the sample ACF. In most of the cases the values of the sample ACF's do not exceed the 5% confidence bands. However, for coffee sugar, soybeans, corn, rice oats and wheat some minor dependence seems to appear. (We found no dependence at all in the uniform transforms $(|\hat{U}_t - 1/2|)_{t \in \mathbb{Z}}$ of wheat, but slight dependence at the first lag for the standardised returns $(|R_t|/\hat{m}_h(t))_{t \in \{1, \dots, n\}}$.) Unless we prespecify a lag we wish to test for, these confidence intervals are not appropriate and the ones from Theorem 22 should be used. We investigated subsamples and constructed sample ACF's for the standardised returns and uniform transforms of these last series. The sample ACF's did not appear to be homogeneous across time. We also looked at the time series plot of \hat{U}_t , $|\hat{U}_t - 1/2|$, $R_t/\hat{m}_h(t)$ and $|R_t|/\hat{m}_h(t)$, $t \in \{1, \dots, n\}$. A certain degree of heteroskedasticity and clustering was evident. For this reason using a purely empirical approach we amended h for these series, on the basis of the sample ACF's. To be more precise, we looked at the smallest values of h such that the value of (21) would not be far from the estimated minimum and such that the sample ACF's would exhibit almost no dependence. For coffee and sugar this implied a considerable reduction in the value of h , while for all other series a very minor reduction was sufficient. Moreover, using these new values for h , the time series plots of the uniform transforms and the standardized series looked much more homogeneous. Therefore, we decided to amend the optimal h 's for coffee, sugar, soybeans, corn, rice oats and wheat on the basis of the heuristics just discussed. Table V gives the results for h using (21) and the amended ones. It is evident that for most series, the best results are obtained by retaining only a small sample of observations to compute m_t (recall that h is the proportion of data used to calculate m_t).

Table V.

Table V. Smoothing Parameter

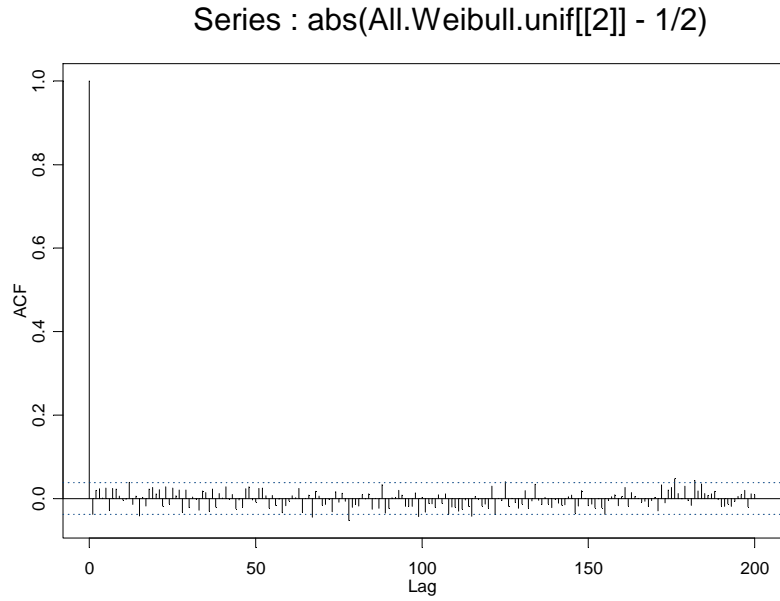
	E R		q
	L ₂ norm	amended	L ₁ norm
CRUDE OIL	0.025	0.025	0.055
GAS OIL (IPE)	0.025	0.025	0.01
HEATING OIL	0.025	0.025	0.07
NATURAL GAS	0.01	0.01	0.085
PROPANE	0.01	0.01	0.085
UNLEADED GAS	0.055	0.055	0.07
COCOA	0.025	0.025	0.145
COFFEE	0.145	0.025	0.115
SUGAR	0.1	0.04	0.01
ORANGE JUICE	0.145	0.145	0.145
SOYBEANS	0.025	0.01	0.145
CORN	0.025	0.01	0.13
RICE	0.025	0.01	0.025
OATS	0.025	0.01	0.04
WHEAT	0.025	0.01	0.025
COTTON	0.04	0.04	0.025

After these changes, we found no evidence of any strong dependence and most of the values were within the nominal 5% level confidence bands. Figure VII plots the sample ACF for the uniform transforms $\left(\left|\hat{U}_t - 1/2\right|\right)_{t \in \{1, \dots, n\}}$ of unleaded gas and the one for rice, which is the worse ACF. Though there is a great improvement, rice might require a more complex modelling approach. Nevertheless, the degree of correlation for rice appears to be very small.

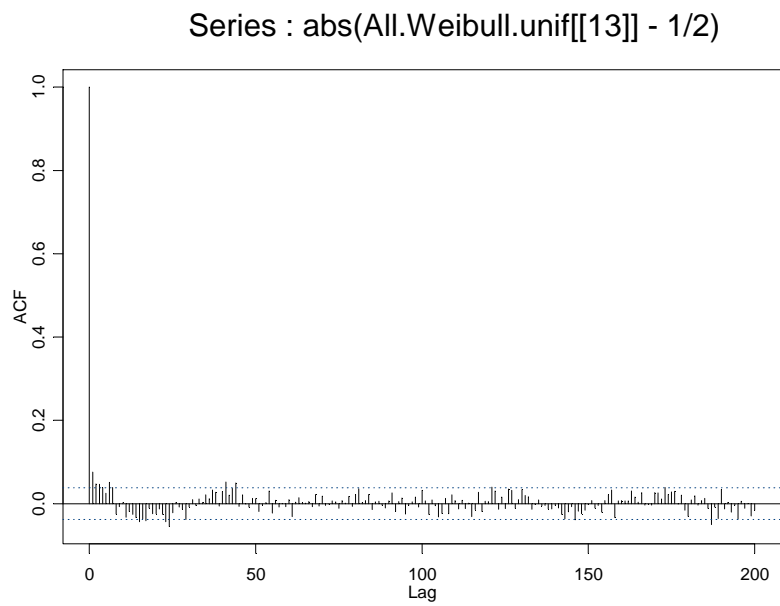
Figure VII.

Figure VII. Sample ACF for $|U_t - 1/2|$

Gas Oil



Rice



Leverage effects. Another measure of non-linear dependence in financial returns is leverage effect, say lev_k , where

$$lev_k = corr(|R_{t+k}|^2, R_t), k \in \mathbb{Z}.$$

We say that there is a leverage effect if $lev_k < 0$ for $k > 0$ and $lev_k \simeq 0$ for $k < 0$. The analysis of commodity data indicate that return-volatility correlation is very small ($|lev_k| \leq 0.05$ for most of the commodities), and the magnitudes for $k > 0$ and $k < 0$ are comparable. Moreover, correlation of returns with subsequent volatility is not necessarily negative. For example, for natural gas, propane, and coffee, $lev_k > 0$ for $k = 1, 2, 3, 4, 5$. Using the same approach as above, we provide a different definition of leverage effect in terms of $corr(|U_{t+k} - 1/2|, U_i)$. We estimated this correlation for all series, but results were very similar. We conclude that these series do not exhibit any substantial leverage effect.

Test of independence. As mentioned above, the ACFs of the residual time series indicate that the adopted detrending procedure removed most of the temporal dependence from the data. To rigorously verify this claim, we conduct the nonparametric test of independence described in Section 3. For the purpose of the test, we only consider one lag.

We computed the critical values using 1000 samples of simulated data from a bivariate product copula. Each sample has the same sample size as the original series ($n = 2670$). We computed the statistics restricting \mathbf{u} in $[.1, .9]$ and $[.05, .95]$. The reason for this is that we want to avoid regions with too few observations. Results are reported in Table VI.

Table VI.

Table VI. Independence Test

\mathbf{u} in $[.1, .9]^2$		\mathbf{u} in $ [.05, .95]^2$	
CRUDE OIL	3.74 ***	CRUDE OIL	3.74 ***
GAS OIL (IPE)	1.82	GAS OIL (IPE)	2.14
HEATING OIL	2.88 **	HEATING OIL	2.88 *
NATURAL GAS	1.84	NATURAL GAS	1.84
PROPANE	4.35 ***	PROPANE	5.02 ***
UNLEADED GAS	3.45 ***	UNLEADED GAS	3.94 ***
COCOA	3.30 ***	COCOA	3.30 **
COFFEE	2.08	COFFEE	2.16
SUGAR	2.78 **	SUGAR	2.92 **
ORANGE JUICE	3.83 ***	ORANGE JUICE	5.41 ***
SOYBEANS	2.37	SOYBEANS	2.59
CORN	2.68 **	CORN	2.68
RICE	4.53 ***	RICE	4.75 ***
OATS	3.50 ***	OATS	3.64 ***
WHEAT	1.83	WHEAT	2.05
COTTON	3.18 ***	COTTON	3.61 ***

* rejected at 90%

** rejected at 95%

*** rejected at 99%

The results show that while the sample ACF improved considerably, we cannot claim that all the series are white noise (though the dependence might be very small).

Asymmetric time dependence. We examine the dependence structure of the time series at the upper and lower tail to detect possible asymmetric time series dependence. The observations were chosen so that the current and previous period observations both lie in the following intervals $[0, .1]$, $[0, .2]$, $[.8, 1]$ and $[.9, 1]$. Then, correlations were calculated. Results are in Table VII, where

the correlation is reported together with the number (n) of observations falling in the conditioning region.

Table VII.

Table VII. Tail Correlation

	correlation in [0,.1]	n	correlation in [0,.2]	n	correlation in [.8,1]	n	correlation in [.9,1]	n
CRUDE OIL	0.26	38	0.04	125	0.04	99	-0.03	23
GAS OIL (IPE)	0.16	26	0.05	111	0.16	112	0.13	31
HEATING OIL	0.41	23	0.10	105	0.03	112	-0.04	30
NATURAL GAS	0.11	24	0.18	101	0.15	115	0.08	27
PROPANE	0.27	28	0.10	120	-0.01	148	0.03	41
UNLEADED GAS	-0.23	37	-0.03	129	-0.05	100	-0.40	24
COCOA	0.11	23	0.05	93	0.26	105	0.28	34
COFFEE	-0.17	24	0.10	102	0.13	113	0.18	41
SUGAR	-0.23	36	0.00	119	0.21	103	0.00	30
ORANGE JUICE	0.05	37	0.22	114	0.11	99	-0.07	19
SOYBEANS	-0.08	24	0.07	112	-0.07	100	0.15	17
CORN	-0.10	25	-0.19	114	0.05	112	0.15	32
RICE	0.25	34	0.09	129	0.13	147	-0.04	48
OATS	0.28	31	-0.04	121	0.02	104	0.26	30
WHEAT	-0.09	24	-0.03	120	0.07	123	0.17	40
COTTON	0.21	38	0.04	123	0.30	127	0.06	36

Some asymmetric dependence might be present. However, differences might also be an artifact of the small number of observations in the tails. Based on the results of Tables VI and VII, we can infer that there is evidence of time series dependence. However, the sample ACF's show that the degree of dependence is very small.

4.5 Preliminary Cross-Sectional Analysis

For the purposes of cross-sectional analysis, commodities (in their uniform representation) were divided in three groups: energy, miscellaneous and grains/oilseeds/fiber. For soft commodities (i.e. non-energy group), this division is based on the nomenclature used by the Economic Research Service tariff database of the United States Department of Agriculture. The energy group comprises of crude oil, gas oil, heating oil, natural gas, propane, unleaded gas; miscellaneous group consists of cocoa, coffee, sugar and orange juice; grains/oilseeds/fiber include soybeans, corn, rice, oats, wheat and cotton. Our subdivision of soft commodities is justified by the fact that correlation of two commodities coming from different groups is usually very low. For this reason, assuming a stationary covariance structure of the transformed uniform data, we report the sample correlation matrix within each group, but not among groups in Table VIII.

The overall correlation among different energy commodities in uniform space is rather strong. However, energy commodities exhibit virtually no correlation with commodities of other groups. Correlation between different commodities within miscellaneous group is rather weak (.3-.11), and there is virtually no correlation between miscellaneous commodities and the last group. The correlation among grains/oilseeds/fiber is lower than correlation among energy commodities, but higher than the correlation within the miscellaneous commodities.

Table VIII.

Table VIII. Correlation Matrices for Uniform Transforms

Energy Group

CRUDE OIL	1.00	0.45	0.79	0.21	0.44	0.74
GAS OIL (IPE)	0.45	1.00	0.52	0.13	0.42	0.41
HEATING OIL	0.79	0.52	1.00	0.24	0.47	0.71
NATURAL GAS	0.21	0.13	0.24	1.00	0.25	0.19
PROPANE	0.44	0.42	0.47	0.25	1.00	0.39
UNLEADED GAS	0.74	0.41	0.71	0.19	0.39	1.00

Miscellaneous Group

COCOA	1.00	0.11	0.08	0.04
COFFEE	0.11	1.00	0.06	0.03
SUGAR	0.08	0.06	1.00	0.01
ORANGE JUICE	0.04	0.03	0.01	1.00

Grains/Oilseeds/Fiber Group

SOYBEANS	1.00	0.57	0.21	0.43	0.38	0.12
CORN	0.57	1.00	0.15	0.52	0.52	0.09
RICE	0.21	0.15	1.00	0.16	0.13	0.05
OATS	0.43	0.52	0.16	1.00	0.38	0.06
WHEAT	0.38	0.52	0.13	0.38	1.00	0.09
COTTON	0.12	0.09	0.05	0.06	0.09	1.00

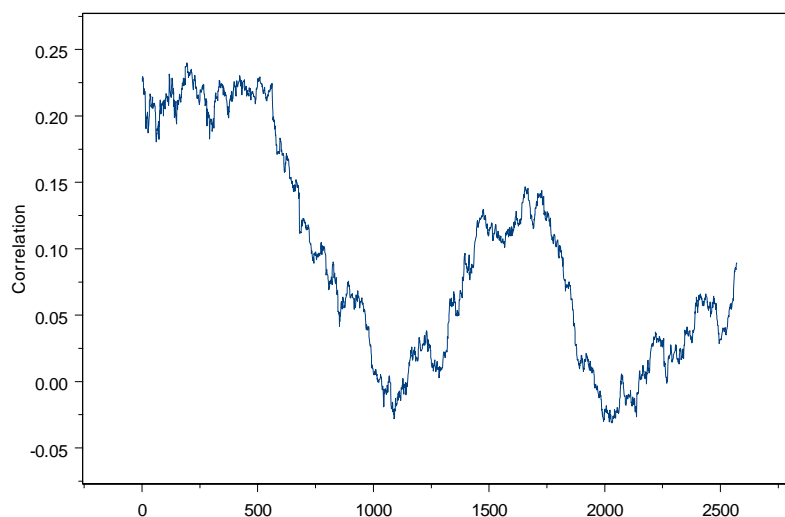
4.5.1 Dynamic Correlation Analysis

We joined the series using the Gaussian copula and estimated (22) as outlined above. Figure VIII plots the estimated time varying correlation for cocoa and sugar in Gaussian space (i.e. after transforming the uniforms into Gaussian random variables). It is clear that correlation does not appear to be homogeneous over time.

Figure VIII.

Figure VIII. Time Varying Correlation

Cocoa with Sugar



We orthogonalized the random variables in Gaussian space and transformed them back into uniform, as described in Section 2. Table IX reports the correlation matrix of the uniform random variables after this nonlinear transformation. The transformed series do not exhibit any substantial correlation within groups. This differs from results in Table VIII.

Table IX.

Table IX. Correlation Matrices for Uniform Transforms
after Transformation

Energy Group

CRUDE OIL	1.00	0.03	0.01	0.04	0.01	0.00
GAS OIL (IPE)	0.03	1.00	0.03	0.00	-0.01	0.00
HEATING OIL	0.01	0.03	1.00	0.00	0.01	-0.03
NATURAL GAS	0.04	0.00	0.00	1.00	-0.07	0.00
PROPANE	0.01	-0.01	0.01	-0.07	1.00	-0.03
UNLEADED GAS	0.00	0.00	-0.03	0.00	-0.03	1.00

Miscellaneous Group

COCOA	1.00	0.02	0.02	-0.01
COFFEE	0.02	1.00	0.00	0.00
SUGAR	0.02	0.00	1.00	0.01
ORANGE JUICE	-0.01	0.00	0.01	1.00

Grains/Oilseeds/Fiber Group

SOYBEANS	1.00	0.01	0.01	-0.02	0.00	0.00
CORN	0.01	1.00	-0.01	0.03	-0.03	-0.01
RICE	0.01	-0.01	1.00	0.03	0.01	-0.04
OATS	-0.02	0.03	0.03	1.00	-0.03	0.00
WHEAT	0.00	-0.03	0.01	-0.03	1.00	-0.06
COTTON	0.00	-0.01	-0.04	0.00	-0.06	1.00

Testing for meta-ellipticity In order to verify that the dependence structure defined by the Gaussian copula is satisfactory, we conducted the test in Section 3. We computed the statistics restricting \mathbf{u} in $[\cdot, .9]$ and $[\cdot, .95]$ and simulating critical values over these regions. Results are reported in Table X. Since we use a one sided filter, the first 100 observations are only used for predictions, hence the sample size is smaller than in the previous test. The results show that the standardization eliminated most of the dependence within each group within the body of the distribution (this is certainly true for \mathbf{u} in $[\cdot, .9]$). The results confirms that meta-ellipticity with inhomogeneous scaling matrix could be a suitable assumption to make, apart from extreme situations.

Table X.

Table XI. Test for Meta Elliptical Dependence

Panel A

Transformed Uniforms, \mathbf{u} in $[.1, .9]^2$

Energy Group

CRUDE OIL	-	2.10	1.68	2.02	1.53	1.64
GAS OIL (IPE)		-	1.53	1.56	1.84	1.42
HEATING OIL			-	1.14	2.19	1.70
NATURAL GAS				-	2.56 *	1.82
PROPANE					-	2.06
UNLEADED GAS						-

Miscellaneous Group

COCOA	-	2.26	2.07	1.71		
COFFEE		-	1.41	1.10		
SUGAR			-	1.74		
ORANGE JUICE				-		

Grains/Oilseeds/Fiber Group

SOYBEANS	-	1.37	1.94	2.11	1.31	1.81
CORN		-	2.23	3.98 ***	1.82	1.17
RICE			-	1.50	2.54 *	2.07
OATS				-	2.26	2.04
WHEAT					-	2.85 **
COTTON						-

Panel B

Transformed Uniforms, \mathbf{u} in $[.05, .95]^2$

Energy Group

CRUDE OIL	-	2.10	1.68	3.15 **	1.53	1.92
GAS OIL (IPE)		-	1.53	2.13	2.10	1.42
HEATING OIL			-	2.12	2.19	1.70
NATURAL GAS				-	2.92 **	2.76
PROPANE					-	3.39 ***
UNLEADED GAS						-

Miscellaneous Group

COCOA	-	2.70	3.01 **	1.94		
COFFEE		-	1.55	2.31		
SUGAR			-	2.31		
ORANGE JUICE				-		

Grains/Oilseeds/Fiber Group

SOYBEANS	-	1.55	2.59	2.55	2.05	2.63
CORN		-	2.23	4.47 ***	2.19	1.23
RICE			-	2.09	3.72 ***	2.27
OATS				-	3.72 ***	2.76 *
WHEAT					-	3.04 **
COTTON						-

Cross-sectional dependence on the tails. In order to further investigate the dependence structure on the tails of the copula, we calculated the correlations of the transformed uniforms when both observations lie in the following intervals $[0, .1]$, $[0, .2]$, $[.8, 1]$ and $[.9, 1]$. There seems to be more dependence as we move to the tails, i.e. in the conditioning regions $[0, .1]$ and $[.9, 1]$. However, the small number of observations make any inference less reliable. For the sake of brevity, we do not report the correlation matrices. The dependence appears to be asymmetric, i.e. the correlations in $[0, .1]$ considerably differ in some cases from the correlations in $[.9, 1]$.

There is no definite pattern, e.g. correlations are not usually higher in $[0, .1]$ than in $[.9, 1]$.

4.6 Summary of Stylized Facts

Some stylised facts emerged from our study. We briefly summarize them.

- The log returns of commodity prices can be characterized by a time varying (symmetric) double Weibull distribution;
- The scale parameter a appears to be time varying, while the shape parameter b might be time homogeneous;
- The shape parameter b is greater than one but less than two, implying tails fatter than a Gaussian, but thinner than an exponential (and clearly thinner than a t-distribution);
- Once we account for time variation in a (i.e. the scaling factor in the Weibull distribution), log-returns do not exhibit any considerable dependence; nevertheless, this does not mean that accounting for variation in the scaling parameter would lead to iid random variables once (5) is used;
- No evidence of leverage effect can be found in the data once we account for time variation;
- Cross-sectional dependence is quite strong in some cases, and appears to be time varying;
- Meta-ellipticity with a Gaussian generator is a good first approximation for non-extreme values.

5 Further Extensions

The methodology we proposed required near independence of the uniform transforms. If our aim is to use the first stage for capturing heterogeneity only, as in Example 4, it may be the case that time dependence is still a problem. In this case, we can use a Markov copula to capture the dependence (Darsow et al., 1992). Here, we only sketch this idea. Suppose the series $(U_i)_{i \in \mathbb{Z}}$ is stationary and can be well described by a Markov process of order l . The copula of U_i, \dots, U_{i-l-1} is given by C^{l+1} and the copula of U_i, \dots, U_{i-l} by C^l . Then, the copula of U_i conditional on $U_{i-1}, \dots, U_{i-l-1}$ is

$$C(u_i | u_{i-1}, \dots, u_{i-l-1}) = \frac{\partial^l C^{l+1}(u_i, u_{i-1}, \dots, u_{i-l-1})}{\partial u_{i-1} \cdots \partial u_{i-l-1}} \frac{\partial u_{i-1} \cdots \partial u_{i-l-1}}{\partial^l C^l(u_{i-1}, \dots, u_{i-l-1})}.$$

Once the marginal dynamics have been modelled using the copula, we can couple the time series copulae using the linkage function (Li et al., 1996).

For brevity, we consider the first order Markovian case; higher order cases can be dealt similarly (see Appendix B). Suppose $(U_{i1}, \dots, U_{iK})_{i \in \mathbb{Z}}$ is a stationary sequence of Markovian random variables uniformly distributed in $[0, 1]^K$. Suppose U_{ik} and $U_{i-1,k}$ have copula C_k ($k = 1, \dots, K$). By Lemma (26) (in Appendix B), $C_k(U_{ik} | U_{i-1,k})$ (the copula of U_{ik} conditional on $U_{i-1,k} = u_{i-1,k}$) and $U_{i-1,k}$ are independent. Suppose $(U_{i1}, \dots, U_{iK})_{i \in \mathbb{Z}}$ have copula C . The linkage function L is the copula of $(U_{i|i-1,1}, \dots, U_{i|i-1,K}, U_{i-1,1}, \dots, U_{i-1,K})_{i \in \mathbb{Z}}$ where $U_{i|i-1,k} := C_k(U_{i,k} | U_{i-1,k})$, and

$$C(U_{i,1}, \dots, U_{i,K}, U_{i-1,1}, \dots, U_{i-1,K}) = L(U_{i|i-1,1}, \dots, U_{i|i-1,K}, U_{i-1,1}, \dots, U_{i-1,K}).$$

Therefore, we may model the time series dependence first using some copula. Then, we can consistently join all the copulae together using the linkage function. The linkage function leads to the following consistency result

$$L(U_{i|i-1,1}, 1, \dots, 1, U_{i-1,1}, 1, \dots, 1) = C_k(U_{i,1}, U_{i-1,1}),$$

and similarly for the other $(U_{i,k}, U_{i-1,k})$ ($k = 2, \dots, K$). The advantage is that we may use a meta-elliptical copula in the second stage (i.e. to model the linkage function), while retaining the flexibility of modelling the time series dependence using a copula that allows for more complex dependence. This allows us to derive the conditional joint distribution. The distribution of $U_{i,1}, \dots, U_{i,K}$ conditional on $U_{i-1,1}, \dots, U_{i-1,K}$ is obtained by differentiating with respect to $(U_{i-1,1}, \dots, U_{i-1,K})$, and dividing everything by the copula density of $(U_{i-1,1}, \dots, U_{i-1,K})$, i.e.

$$\begin{aligned} & \Pr(U_{i,1} \leq u_{i,1}, \dots, U_{i,K} \leq u_{i,K} | U_{i-1,1} = u_{i-1,1}, \dots, U_{i-1,K} = u_{i-1,K}) \\ = & \frac{L_{,K+1..2K}(C_1(u_{i,1}|u_{i-1,1}), \dots, C_k(u_{i,k}|u_{i-1,k}), u_{i-1,1}, \dots, u_{i-1,K})}{C_{,1..K}(u_{i-1,1}, \dots, u_{i-1,K})} \prod_{k=1}^K C_{,2}(u_{i,k}|u_{i-1,k}), \end{aligned}$$

where the last multiplicative term is the Jacobian ($C_{,2}(u_{i,k}|u_{i-1,k}) := C_{,22}(u_{i,k}, u_{i-1,k})$). This shows that a smooth estimator for the conditional copula is required.

Also it is possible that dependence is asymmetric (e.g. Ang and Bekaert, 2002, Ang and Chen, 2002). In such case, a semiparametric approach that retains the simplicity of meta-elliptical copulae plus a correction term that accounts for asymmetric dependence can be used (Sancetta, 2005).

A Proofs of Theorems

The proof of Theorem 18 requires two lemmata.

Lemma 24 *Suppose $(\hat{\mathbf{U}}_t)_{t \in \mathbb{Z}}$ is a strongly mixing stationary sequence of random variables in $[0, 1]^K$ with copula C with continuous partial derivatives. Suppose the strong mixing coefficients are summable. Suppose $\hat{C}_n(\mathbf{u}) = \mathbb{P}_n I \left\{ \mathcal{Q}_n^{-1} \hat{\mathbf{U}}_t \leq \mathbf{u} \right\}$ is the empirical copula of $(\hat{\mathbf{U}}_t)_{t \in \mathbb{Z}}$. Define*

$$Z_n(\mathbf{u}, \mathbf{v}) := \sqrt{n} \left[\left(\hat{C}_n(\mathbf{u}) - C_f(\mathbf{u}) \right) - \left(\hat{C}_n(\mathbf{v}) - C(\mathbf{v}) \right) \right]$$

Then, $\forall \delta$ and $\epsilon > 0 \exists \eta$ such that

$$\Pr \left(\sup_{\|\mathbf{u}-\mathbf{v}\| \leq \delta} Z_n(\mathbf{u}, \mathbf{v}) > \epsilon \right) \leq \eta,$$

i.e. $\sqrt{n} \left(\hat{C}_n(\mathbf{u}) - C_f(\mathbf{u}) \right)$ is stochastically equicontinuous.

Proof. Lemma 10 of Fermanian et al. (2004) establishes the result for the iid case, under the assumption of C having continuous partial derivatives. Their result applies to smooth estimators of the marginal distribution functions of $\hat{\mathbf{U}}_t$. In their result, just use the empirical distribution instead of a smooth estimator. Then, we notice that independence is only used for proving equicontinuity of the multivariate empirical distribution function and the Glivenko-Cantelli Theorem for the marginal distribution functions. Proposition 7.3 in Rio (2000), implies stochastic

equicontinuity of the empirical joint distribution function as soon as the strong mixing coefficients of the series are summable. Corollary 3.1 p. 54 in Rio (2000) implies strong convergence of the empirical distribution function. This together with stochastic equicontinuity implies the Glivenko-Cantelli Theorem for the marginal empirical distribution functions. Then, the result follows as stated. ■

Lemma 25 *Under the conditions of Lemma 24,*

$$\begin{aligned} \sqrt{n} \left(\hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \right) &= \sqrt{n} (\mathbb{P}_n - \mathbb{E}) I_{\{\hat{\mathbf{U}}_t \leq \mathbf{u}\}} \\ &\quad - \sqrt{n} \sum_{k=1}^K C_{f,k}(\mathcal{Q}^{-1}\mathbf{u}) \left(\hat{H}_{k,n}(u_k) - \hat{H}_k(u_k) \right) + o_p(1). \end{aligned} \quad (23)$$

Proof. Notice the following,

$$\hat{C}_n(\mathbf{u}) := \mathbb{P}_n I_{\{\mathcal{Q}_n^{-1}\hat{\mathbf{U}}_t \leq \mathbf{u}\}} = \mathbb{P}_n I_{\{\hat{\mathbf{U}}_t \leq \mathcal{Q}_n \mathbf{u}\}}, C(\mathbf{u}) := \mathbb{E} I_{\{\mathcal{Q}^{-1}\hat{\mathbf{U}}_t \leq \mathbf{u}\}}.$$

Moreover, $\mathcal{Q}_n^{-1}\mathbf{u} \xrightarrow{a.s.} \mathbf{u}$ because $\hat{H}_{k,n} \xrightarrow{a.s.} \hat{H}_k$ ($k = 1, \dots, K$) by the Glivenko-Cantelli Theorem for the empirical distribution (which holds by the remarks in the proof of Lemma 24). By this remark and Lemma 24, we have

$$\begin{aligned} \sqrt{n} \left(\hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \right) &= \sqrt{n} \left(\hat{C}_n(\mathcal{Q}_n^{-1}\mathbf{u}) - C(\mathcal{Q}_n^{-1}\mathbf{u}) \right) + o_p(1) \\ &= \sqrt{n} \left(\mathbb{P}_n I_{\{\hat{\mathbf{U}}_t \leq \mathbf{u}\}} - \mathbb{E} I_{\{\hat{\mathbf{U}}_t \leq \mathbf{u}\}} \right) + \sqrt{n} \left(\mathbb{E} I_{\{\mathcal{Q}^{-1}\hat{\mathbf{U}}_t \leq \mathcal{Q}_n^{-1}\mathbf{u}\}} - C(\mathcal{Q}_n^{-1}\mathbf{u}) \right) + o_p(1) \\ &= \sqrt{n} (\mathbb{P}_n - \mathbb{E}) I_{\{\hat{\mathbf{U}}_t \leq \mathbf{u}\}} - \sqrt{n} \left(C(\mathcal{Q}_n^{-1}\mathbf{u}) - C(\mathcal{Q}^{-1}\mathbf{u}) \right) + o_p(1) \\ &= \sqrt{n} (\mathbb{P}_n - \mathbb{E}) I_{\{\hat{\mathbf{U}}_t \leq \mathbf{u}\}} - \sqrt{n} \sum_{k=1}^K C_{k,k}(\mathcal{Q}_n^{-1}\mathbf{u}) \left(\hat{H}_{k,n}(u_k) - \hat{H}_k(u_k) \right) + o_p(1), \end{aligned}$$

where the last equality follows from

$$\begin{aligned} \left(C(\mathcal{Q}_n^{-1}\mathbf{u}) - C(\mathcal{Q}^{-1}\mathbf{u}) \right) &= \frac{\left(C(\mathcal{Q}_n^{-1}\mathbf{u}) - C(\mathcal{Q}^{-1}\mathbf{u}) \right)}{\left(\hat{H}_{k,n}(u_k) - \hat{H}_k(u_k) \right)} \left(\hat{H}_{k,n}(u_k) - \hat{H}_k(u_k) \right) \\ &\stackrel{a.s.}{=} C_{k,k}(\mathcal{Q}_n^{-1}\mathbf{u}) \left(\hat{H}_{k,n}(u_k) - \hat{H}_k(u_k) \right), \end{aligned}$$

because $\left(\hat{H}_{k,n}(u_k) - \hat{H}_k(u_k) \right) \xrightarrow{a.s.} 0$ uniformly in u_k by the Glivenko-Cantelli Theorem for indicator functions of mixing random variables. ■

Proof of Theorem 18. Lemma 25 gives a probabilistic representation for $\sqrt{n} \left(\hat{C}_n(\mathbf{u}) - C(\mathbf{u}) \right)$. Then, apply Theorem 7.3 in Rio (2000) to this representation to establish weak convergence in the space of continuous bounded functions equipped with the sup norm. ■

Proof of Theorem 22. The correlation function of X_1, \dots, X_n is given by $\rho_n(s) = \sum_{i=1}^{n-s} X_i X_{i+s} / n$, which is a mean zero random variable with variance $(n-s)/n^2$. For $s = o(n)$, $(T-s)/T \rightarrow 1$, so that by Chebyshev's inequality,

$$\rho_n(s) \stackrel{p}{=} \sum_{i=1}^n Y_i^s / n + O((n-s)/n),$$

where $(Y_i^s)_{i \in \mathbb{Z}}$ is a sequence of random variables with same distribution as $(X_i X_{i+s})_{i \in \mathbb{Z}}$ (this is true by almost sure construction). By Theorem 3.2 in Bentkus (2003),

$$\sum_{i=1}^n (Y_i^1, \dots, Y_i^s) / \sqrt{n} \stackrel{w}{=} N_s + O\left(s^{3/2} \ln(s) / \sqrt{n}\right),$$

where N_s is a standard normal random vector with values in \mathbb{R}^s . Choosing $s \leq t_n$ gives the bound $O\left(\ln(t_n)t_n^{3/2}n^{-1/2}\right)$ in the above display. Then, we apply Theorem 3.5 in Holst and Janson (1990) to the Gaussian random vector $N_{t(n)}$ and obtain

$$\left|\Pr\left(\max_s |N_{s,t(n)}| \leq z\right) - \exp\{-\lambda\}\right| = O(t_n^{-1}),$$

where $N_{s,t(n)}$ is the s^{th} entry of the vector $N_{t(n)}$ and $\lambda := t_n \Pr(|N| > z)$ for N standard normal with values in \mathbb{R} . Putting the two bounds together, the bound in the Theorem is $O\left(\ln(t_n)t_n^{3/2}n^{-1/2} + t_n^{-1}\right)$. Equating the two parts, we obtain the bound $O(t_n^{-1})$ for $t_n = O(n^{1/(5+\epsilon)})$ because $\ln(t_n)$ is asymptotically dominated by t_n^ϵ for any $\epsilon > 0$. ■

B The Regression Construction

A reference for almost sure construction of random variables is Rüschemdorf and de Valk (1993). Suppose $\mathbf{X} := (X_1, \dots, X_K)$ is a K dimensional vector of random variables with joint distribution function F . Suppose $\mathbf{V} := (V_1, \dots, V_K)$ is a vector of $[0, 1]$ uniform random variables independent of \mathbf{X} . Define the following transform $\Psi_F : \mathbb{R}^K \times [0, 1]^K \rightarrow [0, 1]^K$, which depends on the joint distribution $F : \mathbb{R}^K \rightarrow [0, 1]$, such that

$$\Psi_F(\mathbf{x}, \mathbf{v}) = \left(\tilde{F}_1(x_1, v_1), \tilde{F}_{2|1}(x_2, v_2|x_1), \dots, \tilde{F}_{K|1, \dots, K-1}(x_K, v_K|x_1, \dots, x_{K-1})\right), \quad (24)$$

where

$$\begin{aligned} \tilde{F}_{k|1, \dots, k-1}(x_k, v_k|x_1, \dots, x_{k-1}) &: = \Pr(X_k < x_k | X_i = x_i, i = 1, \dots, k-1) \\ &+ v_k \Pr(X_k = x_k | X_i = x_i, i = 1, \dots, k-1), \end{aligned} \quad (25)$$

$k = 1, \dots, K$. Then define $\Psi_F^* : [0, 1]^K \rightarrow \mathbb{R}^K$, such that

$$\Psi_F^*(\mathbf{u}) = \mathbf{z} = (z_1, \dots, z_K),$$

where

$$z_k := F_{k|1, \dots, k-1}^{-1}(u_k|z_1, \dots, z_{k-1}) := \inf\{y : F_{k|1, \dots, k-1}(y|z_1, \dots, z_{k-1}) \geq u_k\},$$

and

$$F_{k|1, \dots, k-1}(x_k|x_1, \dots, x_{k-1}) := \Pr(X_k \leq x_k | X_i = x_i, i = 1, \dots, k-1).$$

Using the above notation, we have the following (Theorem 3 in Rüschemdorf and de Valk, 1993).

Lemma 26 (Regression Construction) *Let \mathbf{X} be a K dimensional random vector with distribution function F . Then*

- i. $\mathbf{U} := \Psi_F(\mathbf{X}, \mathbf{V})$ is a K dimensional vector of iid random variables with $[0, 1]$ uniform distribution;*
- ii. $\mathbf{Z} := \Psi_F^*(\mathbf{U})$ is a K dimensional vector of random variables with distribution F ; \mathbf{Z} is called the regression construction of F ;*
- iii. $\mathbf{Z} = \Psi_F^*(\Psi_F(\mathbf{X}, \mathbf{V})) \stackrel{a.s.}{=} \mathbf{X}$.*

References

- [1] Adler, R.J. (2000) On Excursion Sets, Tube Formulae and Maxima of Random Fields. *Annals of Applied Probability* 10, 1-74.
- [2] Ang, A. and J. Chen (2002) Asymmetric Correlations of Equity Portfolios. *Journal of Financial Economics* 63, 443-494.
- [3] Ang, A. and G. Bekaert (2002) International Asset Allocation with Regime Shifts. *Review of Financial Studies* 15, 1137-1187.
- [4] Basrak, B., R.A. Davis and T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and their Applications* 99, 95-115.
- [5] Bentkus, V. (2003) On Normal Approximations, Approximations of Semigroups of Operators, and Approximations by Accompanying Laws. Available at the following URL: <http://www.mathematik.uni-bielefeld.de/fgweb/Preprints/fg03035.pdf>.
- [6] Billingsley, P. (1995) *Probability and Measure*. New York: Wiley.
- [7] Bingham, N.H., C.M. Goldie and J.L. Teugels (1987) *Regular Variation*. Cambridge: Cambridge University Press.
- [8] Blum, J.R., J. Kiefer and M. Rosenblatt (1961) Distribution Free Tests of Independence Based on the Sample Distribution Function. *The Annals of Mathematical Statistics* 32, 485-498.
- [9] Cheng, M.-Y., J. Fan and V. Spokoiny (2003) Dynamic Nonparametric Filtering with Application to Finance. Report 845, Weierstrass Institute for Applied Analysis and Stochastics.
- [10] Cont, R. (2001) Empirical Properties of Asset Returns: Stylised Facts and Statistical Issues. *Quantitative Finance* 1, 223-236.
- [11] Darsow, W.F., B. Nguyen and E.T. Olsen (1992) Copulas and Markov Processes. *Illinois Journal of Mathematics* 36, 600-642.
- [12] Davis, R.A. and T. Mikosch (1998) Limit Theory for the Sample ACF of Stationary Process with Heavy Tails with Application to ARCH. *Annals of Statistics* 26, 2049-2080.
- [13] Deaton A. and G. Laroque (1992) On the Behaviour of Commodity Prices. *The Review of Economic Studies* 59, 1-23.
- [14] Deaton A. and G. Laroque (1996) Competitive Storage and Commodity Price Dynamics. *The Journal of Political Economy* 104, 896-923.
- [15] Deb, P., P.K. Trivedi and P. Varangis (1996) The Excess Co-Movement of Commodity Prices Reconsidered. *Journal of Applied Econometrics* 11, 275-291.
- [16] Diebold, F.X., T. Gunther and A. Tay (1998) Evaluating Density Forecasts, with Applications to Financial Risk Management. *International Economic Review* 39, 863-883.
- [17] Doukhan, P. (1994) *Mixing. Properties and Examples*. New York: Springer-Verlag.

- [18] Drees, H., L. de Haan and S. Resnick (2000) How to Make a Hill's Plot. *The Annals of Statistics* 28, 254-274.
- [19] Fan, J. (2005). A Selective Overview of Nonparametric Methods in Financial Econometrics. *Statistical Science*, forthcoming.
- [20] Fan, J. and I. Gijbels (1996) *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.
- [21] Fang, H.B., K.T. Fang and S. Kotz (2002) The Meta-Elliptical Distributions with Given Marginals. *Journal of Multivariate Analysis* 82, 1-16.
- [22] Fermanian, J.-D., D. Radulovic and M. Wegkamp (2004) Weak convergence of empirical copula processes. *Bernoulli* 10, 847-860. .
- [23] Frisch, U. and D. Sornette (1997) Extreme Deviations and Applications. *Journal de Physique I France* 7, 1155-1171.
- [24] Van Garderen K.J. (1997) Curved Exponential Models in Econometrics. *Econometric Theory* 13, 771-790.
- [25] Genest, C. and B. Rémillard (2004) Tests of independence and randomness based on the empirical copula process. *Test* 13, 335-369.
- [26] Ghosal, S., A. Sen and A.W. van der Vaart (2000) Testing Monotonicity of Regression. *Annals of Statistics* 28, 1054-1082.
- [27] Hansen, B.E. (1994) Autoregressive Conditional Density Estimation. *International Economic Review* 35, 705-730.
- [28] Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- [29] Holst, L. S. Janson (1990) Poisson Approximation Using the Stein-Chen Method and Coupling: Number of Exceedances of Gaussian Random Variables. *The Annals of Probability* 18, 713-723.
- [30] Hult, H. and F. Lindskog (2002) Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions. *Advances in Applied Probability* 34, 587-608.
- [31] Joe, H. (1993) Parametric Families of Multivariate Distributions with Given Marginals. *Journal of Multivariate Analysis* 46, 262-282.
- [32] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman & Hall Ltd.
- [33] Jondeau, E. and M. Rockinger (2005) The Copula-GARCH Model of Conditional Dependencies: An International Stock-Market Application. *Journal of International Money and Finance*, forthcoming.
- [34] Kano, Y. (1994) Consistency Property of Elliptical Probability Density Functions. *Journal of Multivariate Analysis* 51, 139-147.

- [35] Kim, T.Y. (1999) On Tail Probabilities of Kolmogorov-Smirnov Statistics Based on Uniform Mixing Processes. *Statistics and Probability Letters* 43, 217–223.
- [36] Künsch, H.R. (1989) The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* 17, 1217-1241.
- [37] Laherrère, J. and D. Sornette (1998) Stretched Exponential Distributions in Nature and Economy: "Fat Tails" with Characteristic Scales. *The European Physical Journal B* 2, 525-539.
- [38] Li, H., M. Scarsini and M. Shaked (1996) Linkages: a Tool for the Construction of Multivariate Distributions with Given Nonoverlapping Multivariate Marginals. *Journal of Multivariate Analysis* 56, 20-41.
- [39] Lindskog F., A. McNeil and U. Schmock (2003) Kendall's Tau for Elliptical Distributions. In G. Bol, G. Nakhaeizadeh, S.T. Rachev, T. Ridder and K.-H. Vollmer (editors), *Credit Risk: Measurement, Evaluation, and Management*, 149-156. Heidelberg: Springer.
- [40] Mercurio, D. and V. Spokoiny (2004) Statistical Inference for Time-Inhomogeneous Volatility Models. *The Annals of Statistics* 32, 577–602.
- [41] Mikhaleva, T.L. and V.I. Piterbarg (1996). On the Distribution of the Maximum of a Gaussian Field with Constant Variance on a Smooth Manifold. *Theory of Probability and its Applications* 41, 367–379.
- [42] Mikosch, T. and C. Stărică (1999) Change point statistics, long memory and the Garch model. Unpublished manuscript.
- [43] Mikosch, Thomas and C. Stărică (2000) Limit Theory for the Sample Autocorrelations and Extremes of a GARCH (1,1) Process. *The Annals of Statistics* 28, 1427-1451.
- [44] Muth J.F. (1960) Optimal Properties of Exponentially Weighted Forecasts *Journal of the American Statistical Association* 290, 299-306.
- [45] Patton, A. (2005) Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, forthcoming.
- [46] Peligrad, M. (1998) On the Blockwise Bootstrap for Empirical Processes for Stationary Sequences. *The Annals of Probability* 26, 877-901.
- [47] Pesaran, M.H., D. Pettenuzzo and A. Timmermann (2004) Forecasting Time Series Subject to Multiple Structural Breaks. *Cambridge Working Papers in Economics*.
- [48] Pindyck R.S. (2004) Volatility and Commodity Price Dynamics. *Journal of Futures Markets* 24, 1029-1047.
- [49] Polzehl, J. and V. Spokoiny (2002) Local Likelihood Modeling by Adaptive Weights Smoothing. Weierstrass Institute for Applied Analysis and Stochastics. Report 787, Weierstrass Institute for Applied Analysis and Stochastics.
- [50] Polzehl, J. and V. Spokoiny (2003) Varying Coefficient Regression Modeling by Adaptive Weights Smoothing. Report 818, Weierstrass Institute for Applied Analysis and Stochastics.

- [51] Radulovic, D. (2002) On the Bootstrap and Empirical Processes for Dependent Sequences. *Empirical Process Techniques for Dependent Data*, 345-364. Boston: Birkhäuser Boston.
- [52] Resnick, S. (2004) On the Foundations of Multivariate Heavy Tail Analysis. *Journal of Applied Probability* 41, 191-212.
- [53] Resnick, S. and C. Stărică (1995) Consistency of Hill's Estimator for Dependent Data. *Journal of Applied Probability* 32, 139-167.
- [54] Rio, E. (2000) *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. Paris: Springer.
- [55] Rüschemdorf, L. and V. de Valk (1993) On Regression Representation of Stochastic Processes. *Stochastic Processes and their Applications* 46, 183-198.
- [56] Ruppert, D. and M.P. Wand (1997) Multivariate Locally Weighted Least Square Regression. *The Annals of Statistics* 22, 1346-1370.
- [57] Sancetta (2004) Limit Theorems for the Empirical Kendall's Tau Process and its Bootstrap Approximation under Weak Dependence. Submitted.
- [58] Sancetta (2005) Semiparametric Estimation of Copulae. Submitted.
- [59] Silvapulle P. and C.W.J. Granger (2001) Large Returns, Conditional Correlation and Portfolio Diversification: A Value-at-Risk Approach. *Quantitative Finance* 10, 542-551.
- [60] Sklar, A. (1973) Random Variables, Joint Distribution Functions, and Copulas. *Kybernetika* 9, 449-460.
- [61] Stărică, C. (1999) Multivariate Extremes for Models with Constant Conditional Correlation. *Journal of Empirical Finance* 6, 515-553.
- [62] Stărică, C. and C.W.J. Granger (2005) Non-Stationarities in Stock Returns. *Review of Economics and Statistics*, forthcoming.
- [63] Taylor, S.J. (1980) Conjectured Models for Trends in Financial Prices, Tests and Forecasts. *Journal of the Royal Statistical Society, Series A*, 143, 338-362.
- [64] Van der Vaart, A. and J.A. Wellner (2000) *Weak Convergence of Empirical Processes*. Springer Series in Statistics. New York: Springer.
- [65] Vapnik, V. (1998) *Statistical Learning Theory*. New York: Wiley & Sons.