



Duisenberg school of finance – Tinbergen Institute Discussion Paper

TI 12–125/DSF 45/IV

**Stock Market Asymmetries:
A Copula Diffusion**

Denitsa Stefanova

VU University Amsterdam, and Duisenberg School of Finance.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Stock Market Asymmetries: A Copula Diffusion

Denitsa Stefanova

VU University Amsterdam and Duisenberg School of Finance

October, 2011

Abstract

The paper proposes a model for the dynamics of stock prices that incorporates increased asset co-movements during extreme market downturns in a continuous-time setting. The model is based on the construction of a multivariate diffusion with a pre-specified stationary density with tail dependence. I estimate the model with Markov Chain Monte Carlo using a sequential inference procedure that proves to be well-suited for the problem. The model is able to reproduce stylized features of the dependence structure and the dynamic behaviour of asset returns.

JEL Classification: C11, C51, C58

Keywords: tail dependence, multivariate diffusion, Markov Chain Monte Carlo

1 Introduction

There is wide spread evidence that the distribution of financial asset returns deviates from the assumption of normality both in terms of univariate properties of the data such as excess kurtosis or thick tails, as well as the dependence structure: multivariate normality imposes independence between extreme realizations of the variables, whereas returns are known to be highly correlated during large market downturns. In a study of several major international market indices Longin and Solnik (2001) provide strong evidence of the correlation of tail events of asset returns, especially during bear markets.

For risk management applications, multivariate option pricing or portfolio choice decisions it is important to introduce relatively parsimonious models that can capture the above mentioned features of the data. There has been a proliferation of studies that propose models for incorporating the asymmetric response of conditional correlation to returns, mainly building upon the Dynamic Conditional Correlation model of Engle (2002) or the Dynamic Equicorrelation model (DECO) of Engle and Kelly (2009) with tail dependence, applied in Christoffersen et al. (2011) in a study that highlights the diminishing diversification benefits for international investors. Das and Uppal (2004) model high correlation for large drops in asset returns by introducing a systemic jump across all assets. Ang and Chen (2002) compare several discrete time models in terms of their ability to reproduce the asymmetric dependence pattern present in stock return data. None of the models, however, succeeds in either picking up the extremal dependence pattern of the data or explaining the degree of correlation asymmetry.

In this paper we introduce a model that accommodates extreme tail dependencies and nests a variety of dependence structures. We propose a construction of a multivariate diffusion that relates the drift, the diffusion matrix and the stationary density of the process. That is in the spirit of Dupire (1994), who constructs a univariate diffusion process compatible to observed option prices and thus to the risk-neutral conditional distribution. We, however, model the stationary distribution, which relates to the stochastic long-run equilibrium of the process (Philips (1991)). Our model can be extended further to allow for state-dependant extreme co-movements between stock prices, and thus capture clustering of tail events, conditional on exogenous factors, e.g. related to the business cycle.

We model the dependence structure using copula functions, which allows us to separate features of the marginal behaviour of individual assets from their dependence. However, our model is not limited to copula functions alone. The study of the dynamic multivariate spatial dependence structure of stochastic processes has found several model applications in a discrete time setting (Patton (2004), Fermanian and Wegkamp (2004), Christoffersen et al. (2011)). Kunz (2002) proposes a framework for modeling extremes in multivariate diffusions via copula functions, but limits his attention to a specification with a constant diffusion term, or the reducible diffusions in the spirit of Ait-Sahalia (2008). Instead, we propose a more general model for which the above mentioned construction is a special case.

The stochastic process for asset prices that we propose implies a dependence structure that allows

for increased dependence between extreme realizations, but is also flexible enough to include the case of asymptotic independence (as implied by the Gaussian distribution). With this we answer the concern, raised by Poon et al. (2004) in the sense that using a model that precludes independence in the tails may lead to serious overestimation of the joint risks. Based on the copula decomposition between the dependence structure and the marginal distributions, we build a multivariate diffusion with a prespecified stationary density. Its construction relies on restricting the drift for a given specification of the diffusion term and the stationary density via an application of the Fokker-Planck equation (see Hansen and Scheinkman (1995), Chen et al. (2002) for a similar construction). Thus we obtain a flexible process for asset prices that is able to accommodate a wide array of dependence structures.

While accommodating different types of dependence patterns, our model also keeps track of univariate properties of asset returns, such as leptokurtic univariate distributions with semi-heavy tails, or volatility clustering. To this end we select the marginal distributions from the Generalized Hyperbolic (GH) family of distributions. Their ability to replicate the tail behaviour of asset returns has been established in the context of univariate diffusions (Eberlein and Keller (1995), Rydberg (1999), Bibby and Sorensen (2003)). As well, it has been demonstrated (see Jaschke (1997)) that one can obtain a process for returns with Generalized Hyperbolic stationary distribution with stochastic volatility as a weak limit of a GARCH model in the sense of Nelson (1990).

While the stationary distribution of the proposed process is known in closed form, the transition density is not. This raises a serious estimation challenge, as an exact likelihood approach cannot be applied. As well, approximations of the likelihood function in the spirit of Ait-Sahalia (1999) and Ait-Sahalia (2008) prove to be too computationally intensive when explicit solutions for the density approximation coefficients cannot be obtained. Instead, we propose a Markov Chain Monte Carlo (MCMC) method to estimate model parameters, following a sequential inference procedure of Golightly and Wilkinson (2006a), Roberts and Strammer (2001) and Durham and Gallant (2002). It proves to be well suited for the specification we have.

We also address the question of model selection, using the traditional Bayesian approach based on the marginal likelihood functions of alternative models. Results suggest that models that disregard asymmetric dependence between extreme realizations are rejected in favour of those that take these particular features of the dependence structure into account.

The remainder of the paper is organized as follows. Section 2 discusses the issue of modeling dependence through the use of copula functions. Section 3 introduces the process for asset prices, its construction and the particular assumptions on the univariate marginals as well as the dependence structure. Section 4 reviews the estimation methodology of the proposed multivariate diffusion based on copula functions using an MCMC estimation algorithm. Section 5 discusses the estimation results, focusing on the degree of tail dependence that could be achieved under the proposed model specification, and Section

6 concludes.

2 Copula functions and dependence modeling

The pitfalls of using the linear correlation coefficient as a dependence measure have been largely discussed in literature. Linear correlation fully describes the dependence patterns only in the elliptical class of distributions that are inevitably characterized by symmetry. It is also an inadequate tool for discerning dependence when it comes to extreme events, as it is essentially a measure of the central tendency, involving first and second moments. Among the deficiencies of linear correlation comes the fact that second moments have to be finite in order for it to be defined. As well, it is not invariant under non-linear strictly increasing transformations of the variables (a transformation that is known to leave the dependence structure unchanged). In contrast, all concordance measures of dependence are invariant to increasing transformations of the marginals, while the tail dependence coefficient characterizes the extreme dependence using only the dependence function specification.

Thus, copula theory provides a natural environment for the search of dependence measures that are better suited for capturing extreme co-movement asymmetries. The main concept behind copulas is the separation of the distribution structure from the univariate marginals, as they are functions that link marginals to their multivariate distribution, following Sklar's theorem. Their parsimonious nature makes them suitable for high-dimensional models, as the ones encountered in portfolio selection problems, while their functional specification could be flexible enough to allow for asymptotic extreme (in)dependence: dependence structures range from those generated by elliptical copulas that maintain the validity of the mean-variance framework, to copulas that are able to express extreme value dependence (like the Gumbel copula, consistent with multivariate extreme value theory). Various dependence measures useful for financial applications (comonotonicity, concordance, quadrant (orthant) and tail dependence) can be expressed in terms of copulas.

Copula functions are a useful tool to construct multivariate distributions. They are used to disentangle the information contained in the marginal distributions from that pertaining to the dependence structure. As they are defined as multivariate distribution functions, they contain all the relevant information with respect to the dependence structure.

2.1 Copulas and the dependence structure

A standard treatment of copulas can be found in the monographs of Joe (1997), and Nelsen (1999), Embrechts et al. (2002), Frees and Valdez (1998). Cherubini et al. (2004) offer a comprehensive review of the application of copula functions in finance. The main concept behind them is the separation of the distribution structure from the univariate marginals. A copula can be viewed as a multivariate

distribution function on the unit cube, with uniformly distributed marginals. Alternatively, it can be defined as a function $C : [0, 1]^n \rightarrow [0, 1]$ with the following properties:

- (P1) for every u in $[0, 1]^n$, $C(u) = 0$ if at least one coordinate of u is 0; $C(u) = u_k$ if all coordinates of u except u_k equal 1;
- (P2) C is n -increasing if for each $a, b \in [0, 1]^n$ such that $a \leq b$, the volume of the hypercube with corners a and b is positive, that is $V_C([a, b]) = \sum \text{sgn}(c) C(c) \geq 0$ where c are the vertices of $[a, b]$, and $\text{sgn}(c) = 1$ if $c_k = a_k$ for even k , $\text{sgn}(c) = -1$ if $c_k = a_k$ for odd k . For the bi-variate case this translates into $V_C([u_1, u_2] \times [v_1, v_2]) \equiv C(u_1, v_1) + C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) \geq 0$ for all $u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$.

An important result concerning copulas is Sklar's representation theorem (Sklar, 1959):

For a multivariate joint distribution function F with marginals F_1, \dots, F_n , there exists an n -copula C , such that for all x in \mathbb{R}^n we have that:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (2.1)$$

The copula is uniquely determined if all marginal distributions F_1, \dots, F_n are continuous, otherwise C is unique on $\text{Ran}F_1 \times \dots \times \text{Ran}F_n$. The converse statement also holds, i.e. for a given copula C with marginals F_1, \dots, F_n , the function F defined above is an n -dimensional multivariate distribution function. Sklar provides the following corollary: for a multivariate joint distribution function F with continuous marginals F_1, \dots, F_n and copula C , satisfying the above theorem, and for any $u \in [0, 1]^n$, the following holds:

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \quad (2.2)$$

In the subsequent sections we will use the copula density decomposition formula that follows from (2.2):

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i)$$

where $c(\cdot)$ is the copula density and $f_i(\cdot)$ are the univariate PDFs.

A key property of copulas, that makes them particularly well suited for dependence structure modeling, is their invariance under strictly increasing transformations of the marginals. However, this property is true for the linear correlation as a dependence measure only for affine strictly increasing transformations. In particular, if we consider the functions $\alpha(X)$ and $\beta(Y)$ of two random variables X and Y , then the following transformations change the copula functions in a deterministic way (see Nelsen (1999)):

- (i) if α, β are strictly increasing, then $C_{\alpha(X), \beta(Y)}(u, v) = C_{X, Y}(u, v)$;

- (ii) if α is strictly increasing and β is strictly decreasing, then $C_{\alpha(X),\beta(Y)}(u, v) = u - C_{X,Y}(u, 1 - v)$;
- (iii) if α, β are both strictly decreasing, then $C_{\alpha(X),\beta(Y)}(u, v) = u + v - 1 + C_{X,Y}(1 - u, 1 - v)$.

If C is an n -dimensional copula, then it has a known upper and lower bound (the Frechet-Hoeffding bounds):

$$\begin{aligned}
 L_n(u) &\leq C(u) \leq U_n(u) \\
 \text{where } L_n(u) &= \max\left(\sum_{i=1}^n u_i - n + 1, 0\right) \\
 U_n(u) &= \min(u_1, \dots, u_n)
 \end{aligned} \tag{2.3}$$

For $n = 2$ the upper and the lower bound are copulas, but for $n \geq 3$, L_n is the lower bound in the sense that for any $u \in [0, 1]^n$ there exists such a copula C , that $C(u) = L_n(u)$ (Nelsen (1999)).

Following Mari (2002), the continuity of a copula can be established for each $u, v \in [0, 1]^n$, if it satisfies the stronger Lipschitz condition:

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1| \tag{2.4}$$

Further on, as $C(u)$ is increasing and continuous in u , it is differentiable almost everywhere, and the following holds:

$$0 \leq \frac{\partial}{\partial u_i} C(u) \leq 1, \quad i = 1, \dots, n$$

For each copula we can define a survival function: $\bar{C}(u, v) = 1 - u - v + C(u, v)$ for the bi-variate case, and more generally:

$$\bar{C}(u_1, \dots, u_n) = \Pr(U_1 > u_1, \dots, U_n > u_n)$$

Below we discuss briefly several dependence concepts in a copula framework. Following the Frechet-Hoeffding inequality, it was shown that the upper and the lower bound are both copulas in the bi-variate case, and can be thought of as the joint distribution functions of two couples of univariate vectors: $(U, 1 - U)$ for the lower bound and (U, U) for the upper bound. Thus, the lower bound describes the state of perfect negative dependence (two vectors having this copula are said to be countermonotonic), whereas the upper bound corresponds to the state of perfect positive dependence (and the two vectors having this copula are comonotonic).

Following Embrechts et al. (2002), a proper dependence measure δ should have the following properties:

- (i) δ should be defined for every pair X, Y ;

- (ii) $\delta(X, Y) = \delta(Y, X)$;
- (iii) $-1 \leq \delta(X, Y) \leq 1$;
- (iv) $\delta(X, Y) = 1$ iff X and Y are comonotonic, and $\delta(X, Y) = -1$ iff X and Y are counter-monotonic;
- (v) $\delta(\varphi(X), Y) = \delta(X, Y)$ for a strictly increasing function φ , and $\delta(\varphi(X), Y) = -\delta(X, Y)$ for a strictly decreasing function φ .
- (vi) $\delta(X, Y) = 0$ iff X, Y are independent.

As there is no dependence measure that satisfies properties (v) and (vi), then we should modify the following properties if we require (vi):

- (iii-a) $0 \leq \delta(X, Y) \leq 1$;
- (iv-a) $\delta(X, Y) = 1$ iff X and Y are co/counter-monotonic;
- (v-a) $\delta(\varphi(X), Y) = \delta(X, Y)$ for a strictly monotone function φ .

Concordance measures can also be defined in terms of the copula. Following Embrechts et al. (2002), if (X, Y) and (\tilde{X}, \tilde{Y}) are two couples of independent vectors with common marginals, then the difference between the probability of concordance and discordance (Q) can be expressed in terms of their corresponding copulas:

$$\begin{aligned} \text{If } Q &\equiv \Pr \left[(X - \tilde{X})(Y - \tilde{Y}) > 0 \right] - \Pr \left[(X - \tilde{X})(Y - \tilde{Y}) < 0 \right] \\ \text{Then } Q &= Q(C, \tilde{C}) = 4 \int \int_{[0,1]^2} \tilde{C}(u, v) dC(u, v) - 1 \end{aligned}$$

Kendall's tau $\tau(X, Y)$ and Spearman's rho $\rho_S(X, Y)$ are two measures of concordance that also have copula representation:

$$\tau(X, Y) \equiv Q(C, C) = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1 \quad (2.5)$$

$$\rho_S(X, Y) \equiv 3Q(C, \Pi) = 12 \int \int_{[0,1]^2} uv dC(u, v) - 3 \quad (2.6)$$

where $\Pi^n(u) = u_1 u_2 \dots u_n$ is the independence copula.

When both Kendall's tau and Spearman's rho are equal to 1(-1), then the copula of the two vectors is the upper (lower) Frechet bound.

As we are interested in modeling dependence asymmetries in the tails of the distribution, then the tail coefficient, as a measure of dependence in the lower and the upper tail is of particular interest. The

coefficient of upper tail dependence is defined as the probability of an extreme event in Y , conditional on an extreme event in X :

$$\begin{aligned}\tau^U &= \lim_{u \rightarrow 1} \Pr(Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)) \\ &= \lim_{u \rightarrow 1} \frac{\Pr(Y > F_Y^{-1}(u), X > F_X^{-1}(u))}{\Pr(X > F_X^{-1}(u))}\end{aligned}\tag{2.7}$$

provided that the limit exists. If $\tau^U \in (0, 1]$ then the two vectors of random variables are said to be asymptotically dependent in the right tail. Asymptotic independence is reached for the case of $\tau^U = 0$. Joe (1997) shows that the concept of tail dependence can be related to that of the copula by the following alternative definition of the coefficient for upper tail dependence of a bivariate copula, for which the following limit exists:

$$\tau^L = \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u}\tag{2.8}$$

The coefficient of lower tail dependence can be derived in a similar fashion:

$$\begin{aligned}\tau^L &= \lim_{u \rightarrow 0} \Pr(Y \leq F_Y^{-1}(u) \mid X \leq F_X^{-1}(u)) \\ &= \lim_{u \rightarrow 0} \frac{\Pr(Y \leq F_Y^{-1}(u), X \leq F_X^{-1}(u))}{\Pr(X \leq F_X^{-1}(u))} \\ &= \lim_{u \rightarrow 0} \frac{C(u, u)}{u}\end{aligned}\tag{2.9}$$

and the notions of asymptotic dependence and independence are analogous to those in the right tail. Having in mind the relationship between a copula and its survivor copula, it can be shown that the coefficient of upper tail dependence of a copula is in fact the coefficient of lower tail dependence of the survivor copula. We will rely on this property in the subsequent modeling of the extreme-value diffusion process.

Despite these asymptotic measures of dependence, we are interested as well in the behaviour of random variables as they approach the extremes. This ‘near’ tail dependence measure is called quantile dependence and it is defined in the following way for quantiles q :

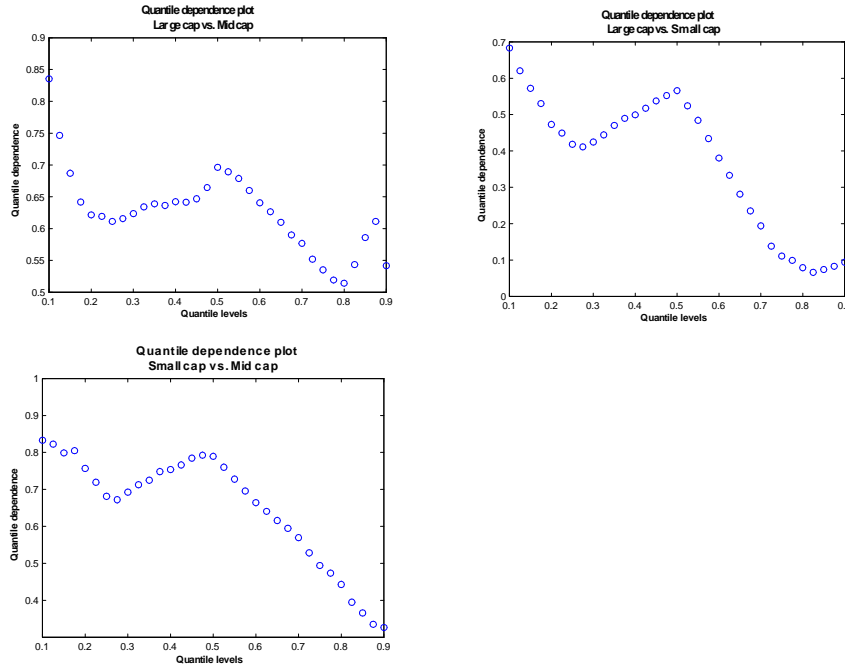
$$\tau(q) = \begin{cases} \Pr[U \leq q \mid V \leq q] & \text{if } q \leq 0.5 \\ \Pr[U > q \mid V > q] & \text{if } q > 0.5 \end{cases}\tag{2.10}$$

2.2 Degree of tail dependence asymmetry in the data

In order to get an impression of the degree of tail dependence asymmetry present in the data, consider daily CRSP US stock capitalization decile indices for the period 1990-2005. These indices represent

Figure 1. Quantile dependence plots

Plots of quantile dependence for all three couples of de-trended log-prices of the three CRSP indices formed on the basis of size deciles for the period 1986-2005 (small-cap (deciles 1-3), mid-cap (deciles 4-7), and large-cap (deciles 8-10)).



yearly rebalanced portfolios based on market capitalization. The stock universe includes stocks listed on NYSE, AMEX, and NASDAQ. All ten capitalization decile indices were grouped in three sub-categories: small-cap (deciles 1-3), mid-cap (deciles 4-7), and large-cap (deciles 8-10).

The degree of ‘near’ tail dependence for all three couples of data is displayed using quantile plots on Fig. 1. The dependence does not decay to zero as we go further in the left tail as it would be the case under bi-variate normality. As well, for the Large-Mid cap couple quantile dependence is high for both tails, while for the other couples of data it tends towards zero for the right tail, pointing towards asymmetric (‘near’) tail dependence.

In order to test the significance in the differences in correlation patterns between the left and the right tail, we use the model-free test of dependence symmetry, developed by Hong et al. (2003). The test statistic under a null hypothesis of symmetry exploits the estimates of the exceedance correlations (ρ_q^-, ρ_q^+) at different quantile levels q and their variance covariance matrix Ω :

$$J = n (\rho^+ - \rho^-) \Omega^{-1} (\rho^+ - \rho^-) \xrightarrow{d} \chi_m^2$$

where n is the sample size and m is the number of quantile levels considered. Table 1 summarizes

Table 1. Test of symmetry in the exceedence correlations

The Hong et al. (2003) test of exceedence correlations symmetry in the lower and upper quartiles for the de-trended log-prices of the three CRSP indices formed on the basis of size deciles for the period 1986-2005 (small-cap (deciles 1-3), mid-cap (deciles 4-7), and large-cap (deciles 8-10)). The test statistic is given by:

$$J = n (\rho^+ - \rho^-) \Omega^{-1} (\rho^+ - \rho^-) \xrightarrow{d} \chi_m^2$$

where ρ^+ and ρ^- are the exceedence correlations calculated at the corresponding quantile levels, n is the sample size and m is the number of quantile levels considered. Results for three for three quantile levels (0.85, 0.90, 0.95) are given below:

	Large vs. Mid cap	Large vs. Small cap	Small vs. Mid cap
Test statistic (J)	1.9351	17.6046	13.3933
p-values	(0.5860)	(5.3065e-004)	(0.0039)

the results of the test, rejecting symmetry for all but the Mid-Large cap couple, for which the quantile dependence plots indicated as well high dependence in both tails.

In the sections that follow we will build a diffusion process that accounts for those dependence features of the data with the help of copula functions. It also accomodates desirable univariate properties of asset returns such as volatility clustering, heavy tails, and slowly decaying autocorrelation function of squared returns, without reverting to a stochastic volatility specification or the introduction of jumps.

3 The multivariate copula diffusion model

In the discrete time literature there exist numerous models that are able to replicate both stylized facts of univariate asset returns series, such as thick-tailed asymmetric marginals, volatility clustering, slowly decaying autocorrelation function of squared returns, and asymmetric dependence structure in the extremes of the multivariate distribution. Copula functions have become increasingly popular in multivariate discrete time models, as in Patton (2004), Jondeau and Rockinger (2002) among others. Astonishingly, much less effort has been spent in this respect in continuous time modeling, except for scalar diffusions. Examples include stochastic volatility models (Heston (1993)) or diffusions with jumps in returns and volatility (Eraker et al. (2003)), hyperbolic diffusions (Bibby and Sorensen (1997)), generalized hyperbolic diffusions (Rydberg (1999)), time-changed Lévy processes (Carr and Wu (2004)). However, the multivariate spatial dependence structure modeling of diffusions has attracted much less attention. Here we propose a construction of a multivariate diffusion with pre-specified stationary density with arbitrary marginals, coupled by a sufficiently parsimonious copula dependence function that avoids the curse of dimensionality problem, normally encountered in modeling multivariate datasets. The aim is to provide a sufficiently flexible treatment of the univariate return series that is able to accommodate the stylized features of the data, as well as to allow for possible asymmetries in the tail dependence of the multivariate

distribution via the copula function.

3.1 Constructing a diffusion with a pre-specified stationary distribution

We assume that uncertainty is driven by a d -dimensional standard Brownian motion and that the price of the risky asset can be expressed as ¹:

$$S_{it} = \exp(\phi_i(t) + X_{it}), i = 1, \dots, d \quad (3.1)$$

for some deterministic function of time $\phi_i(t)$, which we assume to be linear in t , $\phi_i(t) = k_it$ with a linear trend parameter k_i , and where

$$dX_t = \mu(X_t) dt + \Lambda(X_t) dW_t \quad (3.2)$$

Thus, applying Itô's lemma we obtain for the price process for $i = 1, \dots, d$:

$$dS_{it} = S_{it}\mu_i^S(\ln S_{it} - k_it) dt + S_{it} \sum_{j=1}^d \Lambda_{ij} (\ln S_{it} - k_it) dW_{jt} \quad (3.3)$$

$$\text{where } \mu_i^S(X_t) = \mu_i(X_t) + k_i + \frac{1}{2} \sum_{j=1}^d \bar{\sigma}_{ij}(X_t)^2$$

where $\bar{\sigma}_{ij}$ are entries of the matrix Λ in the diffusion term of the process for the de-trended log-price X . As pointed out in Bibby and Sorensen (1997), there is empirical evidence that the increments of the process for the log-price are nearly uncorrelated but not independent, which motivates the specification in 3.1. It is chosen as the most straightforward generalization of the Black Scholes model. The exact parametrization of the drift and the diffusion term will be discussed in the subsequent section, where we present a method to construct a diffusion with a pre-specified stationary distribution.

An application of the Fokker-Planck equation allows us to construct a multivariate stationary diffusion by exploiting the relationship that exists between the invariant density, the drift and the diffusion term for the process in (3.2):

$$\begin{aligned} \mu_j &= \frac{1}{2q} \sum_{i=1}^d \frac{\partial (v_{ij}q)}{\partial x_i} \\ \Sigma &= \Lambda\Lambda^\top \text{ with entries } v_{ij} \end{aligned} \quad (3.4)$$

where Λ is a lower triangular matrix, q is a strictly positive continuously differentiable multivariate density function, and Σ is a continuously differentiable positive definite matrix. Using this construction, q appears to be the stationary density of the Markov process, and the drift vector μ is determined by

¹Following the parametrization of Bibby and Sorensen (1997) and Rydberg (1999)

the choice of q and the volatility matrix Σ . Thus, in order to model the stationary diffusion (3.2), we need to specify its invariant density and its diffusion term. For the diffusion term, we propose a constant conditional correlation specification, given by:

$$\begin{aligned} v_{ij} &= \rho_{ij} \sigma_i^X \sigma_j^X \\ \sigma_i^X &= \sigma_i \left[\tilde{f}^i(x_i) \right]^{-\frac{1}{2} \kappa_i} \end{aligned} \quad (3.5)$$

where $\kappa_i \in [0, 1], i = 1, \dots, d$, and the function $\tilde{f}^i(x_i) \sim f^i(x_i)$, i.e. it is proportional to the i th univariate marginal distribution. This is a multivariate generalization of the diffusion term used in Bibby and Sorensen (2003) for the case of univariate diffusion. We show in a subsequent section that this parameterization of the local volatility matches well its non-parametric estimate.

3.1.1 Choice of the marginal distributions

We choose to model the marginal behaviour with distributions in the family of the Generalized Hyperbolic family. This is a much exploited distribution specification for the univariate return series. Introduced by Barndorff-Nielsen (1977) for studying the particle-size distribution of wind-blown sand, it has consequently found application in numerous fields, including finance. Distributions in that family have been successfully fitted to financial time series, while stochastic processes, built on the basis of generalized hyperbolic laws, have been proposed to model the dynamics of stock returns. Eberlein and Keller (1995) introduce the hyperbolic Levy motion in modeling the dynamic behaviour of asset returns. Their model is further extended in Prause (1999) to the generalized hyperbolic case. Bibby and Sorensen (1997) fit a hyperbolic diffusion model to individual stock price data, while Rydberg (1999) proposes a one-dimensional Normal Inverse Gaussian diffusion that accommodates thick tails in log returns. Bauer (2000) investigates the usefulness of hyperbolic distributions for risk management in the context of VaR modeling. As the family of Generalized Hyperbolic distributions covers a vast spectrum of tail behaviour (from Gaussian to power tails), it is particularly suited in the present context of investigating extreme asset co-covements.

The family of GH distributions is constructed as normal mean-variance mixtures with the Generalized Inverse Gaussian (GIG) as the mixing distribution. Thus, the density function for the GH distribution is expressed as:

$$f_{GH}(x; \alpha, \beta, \delta, \mu) = \int_0^\infty N(x; \mu + \beta s, s) GIG(s; \lambda, \delta^2, \alpha^2 - \beta^2) ds \quad (3.6)$$

where $N(\cdot)$ is the normal density with mean $\mu + \beta s$ and variance s , and the GIG density has the form:

$$\begin{aligned} GIG(x; \lambda, \chi, \psi) &= \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\psi\chi})} x^{\lambda-1} e^{-\frac{1}{2}(\chi x^{-1} + \psi x)} \\ x &> 0, \lambda \in \mathbb{R}, \psi, \chi \in \mathbb{R}_+ \end{aligned} \quad (3.7)$$

where K_λ is the modified Bessel function of the third kind with index λ , whose integral representation is given by $K_\lambda(x) = \frac{1}{2} \int_0^\infty y^{\lambda-1} e^{-\frac{x}{2}(y+y^{-1})} dy$ for $x > 0$. The fact that the GH class of distributions is obtained via this convolution operation is exploited when simulating random GH variables.

Solving this integral form gives the following probability density function of the univariate GH distribution:

$$\begin{aligned} f_{GH}(x; \alpha, \beta, \delta, \mu) &= c(\lambda, \alpha, \beta, \delta) \left(\delta^2 + (x - \mu)^2 \right)^{\frac{\lambda-1/2}{2}} \times \\ &K_{\lambda-\frac{1}{2}} \left(\alpha \sqrt{\delta^2 + (x - \mu)^2} \right) e^{\beta(x-\mu)} \\ \text{where } c(\lambda, \alpha, \beta, \delta) &= \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi} \alpha^{\lambda-\frac{1}{2}} \delta^\lambda K_\lambda \left(\delta \sqrt{\alpha^2 - \beta^2} \right)} \\ x &\in \mathbb{R} \end{aligned} \quad (3.8)$$

$c(\lambda, \alpha, \beta, \delta)$ is the normalizing constant, and the parameters have the following interpretations in terms of the distribution: α determines the shape, β the skewness, μ is a location parameter and δ is a scaling parameter. The parameter domain is:

$$\begin{aligned} \delta &\geq 0, \alpha > |\beta| \text{ for } \lambda > 0 \\ \delta &> 0, \alpha > |\beta| \text{ for } \lambda = 0 \\ \delta &> 0, \alpha \geq |\beta| \text{ for } \lambda < 0 \\ \mu &\in \mathbb{R} \end{aligned}$$

GH distributions have semi-heavy tails, given by $\lim_{x \rightarrow \pm\infty} f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu) \sim |x|^{\lambda-1} \exp\{(\mp\alpha + \beta)x\}$ (Prause (1999), Barndorff-Nielsen and Blaesid (1981)). Thus the class can easily accommodate any tail behaviour ranging from power to exponential decline, and can account for tail asymmetries. A useful reparametrization to display the so-called ‘‘shape triangle’’ of the hyperbolic distribution, is given by $\xi = \left(1 + \delta\sqrt{\alpha^2 - \beta^2}\right)^{-\frac{1}{2}}$ and $\chi = \frac{\beta}{\alpha}\xi$, where ξ and χ vary in $0 \leq |\chi| < \xi < 1$. These parameters are invariant under location and scale transformations and can be interpreted as measures of the asymmetry and kurtosis of the distribution.

The GH family of distributions has the normal distribution as a limiting case for $\delta \rightarrow \infty$, $\delta/\alpha \rightarrow \sigma^2$,

and the Student's t distribution as a limit for $\lambda < 0$, $\alpha = \beta = \mu = 0$ (Barndorf-Nielsen (1978), Prause (1999)). The tail behaviour for those limiting cases is as follows. For the normal distribution we have very thin exponential tails $\lim_{x \rightarrow \pm\infty} f_{Ga}(x) \sim c \exp\left(-\frac{x^2}{2}\right)$, while for the Student's t distribution with ν degrees of freedom we have power tails $\lim_{x \rightarrow \pm\infty} f_t(x) \sim c |x|^{-\nu-1}$.

Various special cases can be obtained for different parametrizations of the GH distribution. For $\lambda = 1$ the hyperbolic distribution is obtained:

$$f_H(x; \alpha, \beta, \delta, \mu) = c(\alpha, \beta, \delta) e^{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)} \quad (3.9)$$

$$\text{where } c(\alpha, \beta, \delta) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}$$

$$x \in \mathbb{R}$$

where $\delta > 0$, $\alpha > |\beta|$, $\mu \in \mathbb{R}$. This parametrization has been widely exploited in literature because of the ease of implementation, as the Bessel function appears only in the normalizing constant. However, it limits the possible tail behaviour cases one could obtain, as the tails are allowed exponential decay: $\lim_{x \rightarrow \pm\infty} f_H(x; \alpha, \beta, \delta, \mu) \sim e^{(\mp\alpha+\beta)x}$, but nevertheless it has proved to be successful in modeling the dynamic behaviour of financial time series.

Another subclass of the GH family is that of the Normal Inverse Gaussian (NIG) distribution, obtained for $\lambda = -1/2$, whose density is given by:

$$f_{NIG}(x; \alpha, \beta, \delta, \mu) = c(\alpha, \delta) \left(\delta^2 + (x - \mu)^2\right)^{\frac{1}{2}} \times \quad (3.10)$$

$$K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right) e^{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)}$$

$$\text{where } c(\alpha, \delta) = \frac{\alpha\delta}{\pi}$$

$$x \in \mathbb{R}$$

where $\delta > 0$, $\alpha \geq |\beta| \geq 0$, $\mu \in \mathbb{R}$. This specification has been successfully used as the stationary measure of a univariate diffusion in Rydberg (1999) for modeling US stock price data. It has a somewhat richer specification for the tail decay as compared to the hyperbolic distribution: $\lim_{x \rightarrow \pm\infty} f_{NIG}(x; \alpha, \beta, \delta, \mu) \sim |x|^{-3/2} e^{(\mp\alpha+\beta)x}$. Also, it is one of the two members of the GH class that are closed under convolution (the other one being the Variance Gamma distribution), so that for the sum of two independent random variables $X_i \sim NIG(x; \alpha, \beta, \delta_i, \mu_i)$, $i = 1, 2$ we have that $X_1 + X_2 \sim NIG(x; \alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2)$. This property is exploited in Rydberg (1999) when modeling log prices as NIG diffusions in that log returns are expected to be also approximately NIG distributed as the time horizon goes to infinity, provided that there is almost no autocorrelation in the increments of log prices.

The moment generating function for the Generalized Hyperbolic distribution is given by:

$$M(u) = e^{u\mu} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + u)^2} \right)^{\frac{\lambda}{2}} \frac{K_{\lambda} \left(\delta \sqrt{\alpha^2 - (\beta + u)^2} \right)}{K_{\lambda} \left(\delta \sqrt{\alpha^2 - \beta^2} \right)} \quad (3.11)$$

$$|\beta + u| < \alpha$$

The characteristic function takes the form:

$$\varphi(u) = e^{i\mu u} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + iu)^2} \right)^{\frac{\lambda}{2}} \frac{K_{\lambda} \left(\delta \sqrt{\alpha^2 - (\beta + iu)^2} \right)}{K_{\lambda} \left(\delta \sqrt{\alpha^2 - \beta^2} \right)} \quad (3.12)$$

The mean and variance in this class of distributions are given by:

$$E[X] = \mu + \frac{\delta\beta K_{\lambda+1}(\delta\gamma)}{\gamma K_{\lambda}(\delta\gamma)} \quad (3.13)$$

$$Var(X) = \frac{\delta K_{\lambda+1}(\delta\gamma)}{\gamma K_{\lambda}(\delta\gamma)} + \frac{\delta^2 \beta^2}{\gamma^2} \left(\frac{K_{\lambda+1}(\delta\gamma)}{K_{\lambda}(\delta\gamma)} - \frac{K_{\lambda+1}^2(\delta\gamma)}{K_{\lambda}^2(\delta\gamma)} \right)$$

where $\gamma^2 = \alpha^2 - (\beta + x)^2$. These expressions have a particularly simple form for the NIG distribution, following the property of the Bessel function that:

$$K_{n+\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}} e^{-x} \left(1 + \sum_{i=1}^n \frac{(n+i)!}{(n-i)!i!} (2x)^{-i} \right), n = 0, 1, 2, \dots$$

So that for NIG we obtain:

$$E[X] = \mu + \frac{\delta\beta}{\gamma}$$

$$Var(X) = \frac{\delta\alpha^2}{\gamma^3}$$

$$Skew(X) = 3\delta\alpha^2\beta\gamma^{-5}$$

$$Kurt(X) = 3\delta\alpha^2(\alpha^2 + 4\beta^2)\gamma^{-7}$$

(Bibby and Sorensen (2003)).

In our empirical application we choose the general form of the GH distribution, or its special case – the NIG distribution, because of the general tail behavior allowed under these specifications.

Univariate diffusion specifications with Generalized Hyperbolic stationary distribution. In order to investigate what stylized facts of asset returns are reproducible with a GH distribution, we also consider a univariate GH diffusion for each of the X_i state variables. The typical construction of a scalar diffusion exploits the relationship between the stationary density and the densities of the speed and the scale measure. We consider a univariate diffusion process for X_i , given by:

$$dX_{it} = \mu_i(X_{it}) dt + \bar{\sigma}_i(X_{it}) dW_{it}$$

Its scale density is defined as $s_i(x) = \exp\left(-\int_x \frac{2\mu_i(u)}{v_i(u)} du\right)$ for $l < x < u$, where $v_i(x) = [\bar{\sigma}_i(x)]^2$, and the speed density is expressed as $m_i(x) = \frac{1}{v_i(x)s_i(x)}$. Under mild conditions, ensuring existence of a solution and stationarity, the relationship between the invariant density $f_i(x)$ and the drift and diffusion coefficients can be shown to verify:

$$2\mu_i(x) - v_i'(x) = v_i(x) \frac{f_i'(x)}{f_i(x)} \quad (3.14)$$

This allows us to construct a stationary univariate diffusion, having $f_i(x)$ as its invariant density. As this construction leaves either the drift or the diffusion coefficient free to be specified, once the form of the stationary density has been chosen, Bibby and Sorensen (2003) suggest the following specification of the drift:

$$\mu_i(x) = \frac{1}{2} v_i(x) \frac{d}{dx} \ln\left(v_i(x) \tilde{f}_i(x)\right) \quad (3.15)$$

where $\tilde{f}_i(x)$, an integrable function on the interval (l, u) , is proportional to the desired stationary density. Notice that the relationship determining the drift of the stationary diffusion depends only on the ratio $\frac{f_i'(x)}{f_i(x)}$, thus it is sufficient to specify the invariant density up to a constant of proportionality. Thus we consider the function $\tilde{f}_i(x) \sim f_i(x)$ that is proportional to the density of the univariate GH distribution (3.8). The volatility term is then given by $\bar{\sigma}_i(x) = \tilde{f}_i(x)^{-\frac{1}{2}\kappa_i}$, and we obtain the general form of a stationary univariate diffusion process for a state variable X_i :

$$dX_{it} = \frac{1}{2} \sigma_i^2 (1 - \kappa_i) \left[\tilde{f}_i(X_{it})\right]^{-\kappa_i - 1} \frac{\partial \tilde{f}_i(X_{it})}{\partial X_{it}} dt + \sigma_i \left[\tilde{f}_i(X_{it})\right]^{-\frac{1}{2}\kappa_i} dW_{it} \quad (3.16)$$

This specification has been exploited in a number of studies², and it nests special cases of a zero drift diffusion (in the case of $\kappa = 1$) or constant diffusion term (in the case of $\kappa = 0$).

The above mentioned models in the family of the GH diffusions have one important advantage over the NIG Levy processes, proposed in Barndorff-Nielsen (1995), that have grown considerably popular in modeling log returns. The latter suffer from the deficiency of being incapable of replicating the persistence

²Bibby and Sorensen (2003), Bibby and Sorensen (1997), K uchler et al. (1999), Rydberg (1999), to cite a few.

in correlation in absolute and squared log returns because of the independent Levy increments. We will demonstrate in a latter section that the NIG (or GH) diffusion can accommodate that.

In order to check the fit of the proposed model, we proceed to a formal validation procedure for the scalar diffusions, proposed in Pedersen (1995) and applied in Rydberg (1999), which is based on the univariate residuals:

$$u_{t_i} = F(t_{i-1}, X_{t_{i-1}} | t_i, X_{t_i}; \psi) \quad (3.17)$$

where $F(\cdot)$ is a transition function $F(x, t | y, s; \psi)$ for a given parameter vector ψ that can be estimated via simulation using the dynamic probability transform for a discretized sample of the process $\{X_{\Delta t}\}_{i=1}^n$ over the period $t = 1, \dots, n$ with a discretization step Δ :

$$\hat{u}_t = \int_{-\infty}^{X_{t\Delta}} F(t\Delta, x | (t-1)\Delta, X_{(t-1)\Delta}) dx \quad (3.18)$$

Under the hypothesis of correct model specification, the series $\{\hat{u}_t\}_{t=1}^n$ is *i.i.d.* $U(0, 1)$.

3.1.2 Choice of the copula and the multivariate stationary diffusion

In this section we specify the spatial dependence structure of the multivariate copula diffusion. Following Sklar's theorem, we define the invariant density as:

$$q(x_1, \dots, x_n) \equiv c(F^1(x_1), \dots, F^n(x_n)) \prod_{i=1}^n \tilde{f}_i(x_i) \quad (3.19)$$

where $\tilde{f}_i(\cdot)$ is proportional to the univariate GH distribution (3.8), and $F^i(\cdot)$ is its corresponding CDF.

In order to account for different degrees of upper and lower tail dependence, we consider several parametric families of copulas that have either no tail dependence (Gaussian copula), symmetric tail dependence (Student's t copula), or that allow for different degrees of dependence in the left and in the right tail (Archimedean copulas). Below we discuss the form and properties of the copula functions that we consider for the stationary distribution of the multivariate diffusion for the state variables.

Elliptic copulas. We consider two elliptical copulas, the Gaussian and the t copula, that are characterized by symmetry in the dependence structure. We choose as a benchmark a diffusion that relies on the Gaussian copula. In this case, dependence is governed by the correlation matrix R_{Ga} . Its CDF is defined as:

$$\begin{aligned} & C^{Ga}(u_1, u_2, \dots, u_d | R_{Ga}) \\ = & \int_{-\infty}^{\Phi^{-1}(u_1)} \dots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{2\pi |R_{Ga}|^{1/2}} \exp\left\{-\frac{1}{2}x^\top R_{Ga}^{-1/2}x\right\} dx_1 \dots dx_d \end{aligned} \quad (3.20)$$

where $\Phi^{-1}(u)$ denotes the inverse of the univariate standard normal CDF. The Gaussian copula generates a multivariate normal distribution iff the marginal distributions are also normal. It has no upper or lower tail dependence for imperfectly correlated random variables: $\tau_{Ga}^U = \tau_{Ga}^L = 0$.

The Student's t copula allows for equal upper and lower tail dependence coefficients. Its CDF is given by:

$$C^t(u_1, u_2, \dots, u_d | R_T, \nu) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \dots \int_{-\infty}^{t_\nu^{-1}(u_d)} \frac{\Gamma(\frac{\nu+d}{2}) |R_T|^{1/2}}{\Gamma(\frac{\nu}{2}) (\nu\pi)^{d/2}} \left(1 + \frac{1}{\nu} x^\top R_T^{-1} x\right)^{-\frac{\nu+d}{2}} dx_1 \dots dx_d \quad (3.21)$$

where ν is the degrees of freedom parameter, R_T is the correlation matrix, and $t_\nu^{-1}(u)$ is the inverse of the univariate CDF of the Student's t distribution with ν degrees of freedom. The tail dependence coefficient is given by $\tau_T^U = \tau_T^L = 2t_{\nu+1}(-\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho})$, where ρ is an off-diagonal element of R_T . Thus, the tail dependence coefficient decreases for higher values of the degrees of freedom parameter and in the limit it goes to zero as $\nu \rightarrow \infty$ (in this case the Student's t copula converges to the Gaussian copula). One interesting property of the t copula is the fact that it can still show tail dependence even if the correlation is zero.

Archimedean copulas. Copulas in this family are constructed using a continuous, decreasing and convex generator function $\varphi(u) : [0, 1] \rightarrow [0, \infty)$ that has a defined pseudo-inverse $\varphi^{[-1]}(\varphi(u)) = u$ for all u in $[0, 1]$:

$$\varphi^{[-1]}(u) = \begin{cases} \varphi^{-1}(u) & \text{for } 0 \leq u \leq \varphi(0) \\ 0 & \text{for } \varphi(0) \leq u \leq \infty \end{cases}$$

The pseudo-inverse is given by the usual inverse for the cases when we have a strict generator function φ . Then the Archimedean copulas are defined in terms of the generator function as follows:

$$C(u_1, u_2, \dots, u_n; \alpha) = \varphi^{-1}(\varphi(u_1; \alpha) + \varphi(u_2; \alpha) + \dots + \varphi(u_n; \alpha)) \quad (3.22)$$

for a given dependence parameter α . The density of Archimedean copulas for the bi-variate case is given by (see Nelsen (1999)):

$$c(u_1, u_2) = \frac{-\varphi'(C(u_1, u_2)) \varphi'(u_1) \varphi'(u_2)}{(\varphi'(C(u_1, u_2)))^3}$$

Archimedean copulas have the useful property that most dependence measures, including the coefficients of upper and lower tail dependence, can be expressed in terms of the generator function. Joe (1997) provides the following result with respect to tail dependence: for a strict generator $\varphi(u)$, if $\varphi'(0)$ is finite and different from zero, then the copula has no tail dependence. The copula has upper tail dependence for $1/\varphi'(0) = -\infty$, given by $\tau^U = 2 - 2 \lim_{z \rightarrow 0^+} \frac{\varphi'(z)}{\varphi'(2z)}$ and lower tail dependence, given

by $\tau^L = 2 \lim_{z \rightarrow +\infty} \frac{\varphi'(z)}{\varphi'(2z)}$. Kendall's tau also has a representation in terms of the generator function:
 $\tau = 4 \int_{[0,1]} \frac{\varphi(z)}{\varphi'(z)} dz + 1$.

We consider the Gumbel copula in the Archimedean class, introduced by Gumbel (1960). It is a parsimonious one-parameter copula, whose generator is given by $\varphi(x) = (-\log(x))^{\frac{1}{\alpha}}$, $\alpha \in (0, 1]$, so that its CDF can be expressed as:

$$C_{\alpha}^G(u_1, u_2, \dots, u_n) = \exp\left(-\left(\sum_{i=1}^n (-\log u_i)^{\frac{1}{\alpha}}\right)^{\alpha}\right), \quad \alpha \in (0, 1] \quad (3.23)$$

Its Kendall's tau is given by $\rho_{\tau} = 1 - \alpha$, and the coefficient of upper tail dependence is given by $\tau_G^U = 2 - 2^{\alpha}$, while the coefficient of lower tail dependence is zero. Independence is achieved for $\alpha = 1$, in this case both tail dependence coefficients are zero.

In order to incorporate lower tail dependence, we use the survival Gumbel copula. For the bivariate case³ it is defined as follows:

$$\overline{C}_{\bar{\alpha}}^G(u, v) = u + v - 1 + \exp\left(-\left[(-\log(1-u))^{\frac{1}{\bar{\alpha}}} + (-\log(1-v))^{\frac{1}{\bar{\alpha}}}\right]^{\bar{\alpha}}\right) \quad (3.24)$$

$$\bar{\alpha} \in (0, 1] \quad (3.25)$$

Its Kendall's tau is given by $\rho_{\tau} = 1 - \bar{\alpha}$, and the coefficient of lower tail dependence is given by $\tau_{SG}^L = 2 - 2^{\bar{\alpha}}$ while the coefficient of upper tail dependence is zero.

The symmetrised Joe-Clayton (SJC) copula. A bi-variate copula function that has both upper and lower tail dependence is the 'BB7' copula of Joe (1997), also known as the Joe-Clayton copula. It is given by:

$$C^{JC}(u_1, u_2 | \tau^L, \tau^U) = 1 - \left\{1 - \left[(1 - (1 - u_1)^{\kappa})^{-\gamma} + (1 - (1 - u_2)^{\kappa})^{-\gamma} - 1\right]^{-\frac{1}{\gamma}}\right\}^{\frac{1}{\kappa}}$$

where $\kappa = \frac{1}{\log_2(2 - \tau^U)}$, $\gamma = -\frac{1}{\log_2(2 - \tau^L)}$, and $\tau^U \in (0, 1)$, $\tau^L \in (0, 1)$

The two parameters of the Joe-Clayton copula are the coefficients of upper (τ^U) and lower (τ^L) tail dependence. As the above parameterization suffers from the drawback that even if both parameters are equal, there is still some residual asymmetry due to the functional form, we consider its 'symmetrised'

³See Theorem 4.7 in Cherubini et al. (2004) for dimensions bigger than 2.

version, proposed by Patton (2004), given by:

$$\begin{aligned} & C^{SJC}(u_1, u_2 | \tau^L, \tau^U) \\ &= \frac{1}{2} [C^{JC}(u_1, u_2 | \tau^L, \tau^U) + C^{JC}(1 - u_1, 1 - u_2 | \tau^L, \tau^U) + u_1 + u_2 - 1] \end{aligned}$$

Nested Archimedean copulas. Applying directly the Archimedean generator function in order to obtain dependence functions for dimensions larger than 2 imposes the potentially implausible restriction of a common dependence parameter across all dimensions. We thus revert instead to a nested version of the dependence function in the family of Archimedean copulas (Whelan (2004), Embrechts et al. (2002)). For our tri-variate application we obtain:

$$C(u_1, u_2, u_3) = \varphi_2^{-1}(\varphi_2(\varphi_1^{-1}(\varphi_1(u_1) + \varphi_1(u_2))) + \varphi_2(u_3)) \quad (3.26)$$

where we repeatedly nest bi-variate functions, and each generating function $\varphi_i(u_i)$ has its own dependence parameter α_i , with $\alpha_1 \leq \alpha_2$, i.e. dependence is higher in the more deeply nested copula⁴.

The parsimonious structure of the Gumbel copula makes it a suitable candidate for a nested copula, so we consider it in our application. We combine it in a mixture copula with its survival counterpart in order to allow for potentially asymmetric extreme behavior.

The mixture copulas. Combining both Gumbel and survival Gumbel copulas in a mixture copula, where each function is assigned a certain weight, is a way to construct a copula that has both lower and upper tail dependence with different tail dependence coefficients. Following the Poon et al. (2004) critique, and in order to allow for asymptotic tail independence, we include the Gaussian copula in this mixture model, to obtain:

$$C_m^{Ga}(u; R_{Ga}, \alpha, \bar{\alpha}, \omega, \bar{\omega}) = \omega C^G(u; \alpha) + \bar{\omega} \bar{C}^G(u; \bar{\alpha}) + (1 - \omega - \bar{\omega}) C^{Ga}(u; R_{Ga}) \quad (3.27)$$

or the Student's t copula:

$$C_m^t(u; R_T, v, \alpha, \bar{\alpha}, \omega, \bar{\omega}) = \omega C^G(u; \alpha) + \bar{\omega} \bar{C}^G(u; \bar{\alpha}) + (1 - \omega - \bar{\omega}) C^T(u; R_T, v) \quad (3.28)$$

where we are mixing the two extreme value copulas: the nested Gumbel copula $C^G(u; \alpha)$, where α is the vector of dependence parameters α_i that determine upper tail dependence, the nested survival Gumbel copula $\bar{C}^G(u; \bar{\alpha})$, where $\bar{\alpha}$ is the vector of dependence parameters $\bar{\alpha}_i$ that determine lower tail dependence, with two elliptic copulas: the Gaussian copula $C^{Ga}(u; R_{Ga})$ with correlation matrix R_{Ga} in (3.27), or the

⁴Usually the Gumbel copula parameter is defined as $\gamma = \frac{1}{\alpha}$, $\gamma \in (1, \infty)$, and higher dependence will translate in higher levels of γ . But for estimation purposes, we chose the alternative parametrization, using $\alpha \in (0, 1]$, so that higher dependence requires a lower level of α .

Student's t copula with a correlation matrix R_T and a degrees of freedom parameter ν in (3.28). The key difference between the two mixture copulas consists in the fact that the one based on the Gaussian copula allows for tail independence by setting the extreme value copula weights to zero, while for the Student's t case there is still some degree of tail dependence, even if the correlation parameter of the Student's t copula is zero. Thus, we achieve varying degrees of tail dependence or asymmetry. Further, $u = (u_1, u_2, \dots, u_n)^\top$ is the vector of marginal CDFs of the random variables, and $\{\omega, \bar{\omega}\} \in [0, 1], \omega + \bar{\omega} \leq 1$ are the corresponding weights for the Gumbel and the survival Gumbel copulas.

4 MCMC estimation of the multivariate copula diffusion

The above construction of a stationary diffusion with a prespecified stationary density (3.1)-(3.5) poses a serious estimation problem. Even though the invariant density is explicitly known, the problem is with the unknown conditional density of the state variables. Thus, exact likelihood estimation cannot be applied in this case. Ait-Sahalia (1999), Ait-Sahalia (2003) proposes closed-form expansions of the likelihood function both for univariate and multivariate discretely sampled diffusions, based on Hermite polynomials and Taylor expansion of some fixed order. While this method seems well suited for the problem at hand, it could become too computationally intensive in the cases where no explicit solutions for the coefficients of the density approximation can be found. B.M.Bibby and Sorensen (1995) and Rydberg (1999) propose another estimation technique that relies on approximating the conditional density by a normal density and applying a martingale estimation technique. However, even though the martingale estimator is consistent and asymptotically normally distributed, it rests inefficient. To solve this problem, Tse et al. (2004) propose an alternative way of dealing with the problem of unknown transition density - the MCMC estimation for a hyperbolic diffusion. Relying on a discretization of the underlying diffusion, they apply a random-walk Metropolis Hastings algorithm in order to estimate parameters. However they assume that the discrete time intervals given by observation times are accurate enough to approximate the transition density. If the available data is not fine enough, this approach would introduce discretization bias. A suitable alternative, much exploited in recent research, is data augmentation, i.e. introducing latent data points between each pair of observations. This technique has been used in Pedersen (1995) for simulated maximum likelihood estimation of diffusions, or in Jones (2003), Elerian et al. (2001), Roberts and Strammer (2001), or Eraker (2001) for MCMC analysis. The simulated maximum likelihood method relies on a discretization scheme such as the Euler scheme to approximate the one-period-ahead transition density. The MCMC approaches on the other hand propose simulated paths of latent data that bridge two consecutive observations, constraining both ends of the simulated path to be equal to the actual data. Thus, conditioning on both the beginning and the end of each observation sub-period reduces the variance of the simulated latent data and augments the efficiency of the algorithm. However, augmentation schemes are susceptible to causing slow rates of convergence of the resulting Markov chain

due to the dependence between the latent data points and the volatility of the diffusion as the degree of augmentation increases (known as the Roberts and Strammer critique). There have been several remedies to this issue proposed in recent literature, as the particular transformation of the diffusion process to one with constant volatility proposed by Roberts and Strammer (2001), the simulation filter for multivariate diffusions of Golightly and Wilkinson (2006a) that builds upon the sequential parameter estimation procedure of Johannes et al. (2004) for discrete-time stochastic volatility models, or the Gibbs sampler of Golightly and Wilkinson (2006b) that iterates between updates of parameter and states and relies upon conditioning on the Brownian increments instead of the underlying latent data in order to overcome the dependence with volatility parameters.

The estimation scheme we propose to apply in the present setup relies on an MCMC estimation algorithm with data augmentation for both the univariate and the multivariate diffusion specifications. It follows the sequential inference procedure of Golightly and Wilkinson (2006a) and is closely related to the work of Roberts and Strammer (2001) and Durham and Gallant (2002). As the augmentation of the parameter and state space with latent data points is the corner stone in each MCMC algorithm for diffusion estimation, we will first discuss the particular scheme that was chosen and the motivation behind it.

4.1 Data augmentation

Let data be observed at times $t_0 < t_1 < \dots < t_{n-1} < t_n$ with a time increment $\Delta\tau = t_{i+1} - t_i$. Due to the one-to-one mapping between each price S_i and its corresponding state variable X_i , the latter are observable. We divide each subinterval between observations in m equidistant points, so that we obtain an augmented data matrix for the state variables X :

$$X^{aug} = [\overline{\overline{X}}_{t_0,0} \quad X_{t_0,1} \quad \dots \quad X_{t_0,m} \quad \overline{\overline{X}}_{t_1,0} \quad \dots \quad \overline{\overline{X}}_{t_{n-1},0} \quad \dots \quad X_{t_{n-1},m} \quad \overline{\overline{X}}_{t_n,1}],$$

where $X_{t_i,j}$ is a d -dimensional vector of latent data points at time $t_i + j\Delta\tau$ and $\overline{\overline{X}}_{t_i,0}$ is the vector of observations at time t_i . Note that the augmented data matrix could also consist of unobservable state variables, whose treatment would be similar to that of the latent data. Thus, the estimation procedure is applicable to a case when the X variables are latent and cannot be obtained directly from the observations of the prices S .

Working with the Euler discretization of the process, the joint posterior of data and model parameters θ is given by:

$$\pi(X; \theta) \propto \pi(\theta) \prod_{t=t_0}^{t_n-1} \left\{ \prod_{j=1}^m \pi(X_{t,j+1} | X_{t,j}; \theta) \right\} \quad (4.1)$$

where $\pi(\theta)$ is the prior density for the parameter vector, and $\pi(X_{t,j+1} | X_{t,j}; \theta) = \phi(X_{t,j+1} - \mu(X_{t,j})\Delta t, \Lambda(X_{t,j})\Lambda(X_{t,j})^\top)$.

comes from the Gaussian transition density implied by the Euler discretization, where $\phi(\tilde{\mu}, \tilde{\sigma})$ denotes the Gaussian density with mean $\tilde{\mu}$ and covariance matrix $\tilde{\sigma}$.

Inference procedures that rely on a Gibbs sampler use the conditional posterior for parameters given data and the conditional posterior of missing data given parameters and observations, rather than the joint posterior (4.1), and iteratively propose parameters and missing data from each one of them, so that the obtained simulated sequence of parameters and missing data (after an initial burn-in stage) forms a Markov chain whose stationary distribution is the posterior in question. An alternative approach is the joint update of parameters and states, which overcomes the problem of increasing correlation between the volatility parameters and latent data as the degree of augmentation becomes large. But as it is virtually infeasible to update all latent data in one single block, this sampling scheme can be applied in a sequential manner, updating parameters and unobserved state variables as each observation becomes available.

A straightforward procedure for sampling the latent data points has been proposed by Eraker (2001). It can easily deal with high-dimensional problems, including unobserved state variables. It consists of designing an Accept-Reject Metropolis Hastings algorithm for updating one column of data at a time, where the conditional posterior of one column of missing data is defined as $\pi(X_i | X_{\setminus i}; \theta) \propto p(X_i | X_{i-1}, X_{i+1}; \theta)$ following the Markov property of the diffusion. At each iteration h the algorithm proposes a latent data point X_i^* from some proposal density (Eraker uses a normal proposal $q(\cdot | X_{i-1}^h, X_{i+1}^{h-1}; \theta^h)$), which is then accepted or not following the acceptance procedure of the Accept-Reject Metropolis Hastings algorithm of Tierney (1994). The sampling scheme, proposed by Elerian et al. (2001), is essentially the same, but instead of updating one column vector at a time, they propose updating blocks of missing data with random size. However, increasing the number of imputed data points m , while reducing the discretization bias of the Euler approximation, seems to adversely affect the mixing properties of the algorithm, (see Roberts and Strammer (2001)), because of the increasing correlation of the diffusion parameters and the simulated path as m increases. In fact, when the number of latent data points tends to infinity, one could very precisely estimate the diffusion term by the quadratic variation, so that when updating the diffusion parameter, its posterior distribution given the simulated latent path tends towards a point mass at its previous iteration value, rendering it impossible to update the parameter. Roberts and Strammer (2001) propose a reparametrization of the missing data that circumvents the problem of reducible data augmentation. The basic idea behind their scheme is a construction of the latent path that does not depend on the diffusion term. They apply the sampling algorithm on a univariate diffusion with constant diffusion term, as well as on a reducible diffusion in the sense of Ait-Sahalia (2003) that has a deterministic time-varying diffusion term, and that could be transformed to a constant volatility diffusion following the Doss transformation. Their methodology could easily be extended to the estimation of a reducible multivariate diffusion, such as the constant volatility specification considered in Kunz (2002), that is a special case of the model we propose, but for a general multivariate diffusion as in (3.5) it is almost

impossible to solve for the volatility transformation. Therefore, a more promising approach that would be applicable for the multivariate specification we are proposing is the joint update of parameters and states following the sequential MCMC method of Golightly and Wilkinson (2006a), as it does not rely on a volatility transformation for the diffusion and at the same time overcomes the Roberts and Strammer critique to data augmentation. As a direct draw from the joint posterior of the model's parameters and the latent state variables is virtually impossible due to the dimension of the state space, a solution to proceed is to revert to Bayesian sequential filtering, devising an MCMC scheme that updates parameters as each new observation becomes available. This idea has been exploited in Stroud et al. (2004), Johannes et al. (2004), Liu and West (2001) among others. In what follows, we will briefly discuss the algorithm that has been applied in Golightly and Wilkinson (2006a) for the estimation of a general multivariate diffusion that was proved to have better convergence properties than the standard Gibbs sampler that iteratively updates parameters and states.

4.2 The sequential parameter and state estimation scheme

Let us consider that we are at time $t_{j+m} = t_M$ and that we observe $\bar{X}_{t_{j+m}} = \bar{X}_{t_M}$, and also suppose that we have a sample of size MC from the marginal parameter posterior distribution $\pi(\theta | \bar{X}_{t_j})$, where \bar{X}_{t_j} denotes all the observed data up to time t_j . As we are interested in sampling the set of parameters from their marginal posterior $\pi(\theta | \bar{X}_{t_M})$, we could do so by formulating the joint posterior for parameters and latent data $\pi(\theta, X_{t_M} | \bar{X}_{t_M})$ and then integrating out the latter, where X_{t_M} denotes all the latent data points up to time t_M . Notice that the marginal parameter posterior at time t_M can be rearranged as follows:

$$\begin{aligned} \pi(\theta | \bar{X}_{t_M}) &= \int_{X_{t_M}^{aug}} \pi(\theta) \prod_{i=0}^{M-1} \pi(X_{t_{i+1}}^{aug} | X_{t_i}^{aug}; \theta) \\ &= \pi(\theta | \bar{X}_{t_j}) \int_{X_{t_M}^{aug} \setminus \{X_{t_j}^{aug}\}} \prod_{i=j}^{M-1} \pi(X_{t_{i+1}}^{aug} | X_{t_i}^{aug}; \theta) \end{aligned} \quad (4.2)$$

So that our target density at time t_M would be

$$\pi(\theta | \bar{X}_{t_M}) = \pi(\theta | \bar{X}_{t_j}) \prod_{i=j}^{M-1} \pi(X_{t_{i+1}}^{aug} | X_{t_i}^{aug}; \theta)$$

with the augmented data for the interval (t_j, t_M) integrated out.

In order to sample from this target density, we need to devise a Metropolis-Hastings algorithm that will propose parameter and latent data points and will accept or reject those proposals given a certain probability.

4.2.1 The parameter proposal

We follow Golightly and Wilkinson (2006a) and Liu and West (2001) and form the proposal for the parameter set θ using a kernel density estimate of the marginal parameter posterior $\pi(\theta | \bar{X}_{t_j})$ with the kernel shrinkage correction of Liu and West (2001) that takes care of the over-dispersion of the kernel density function compared to the posterior sample. Thus, we draw the proposal sample of parameters from the following density:

$$\begin{aligned}\theta^* &\sim \phi(\alpha\theta_u + (1 - \alpha)\bar{\theta}, h^2V) \\ \alpha^2 &= 1 - h^2 \\ h^2 &= 1 - ((3\delta - 1)/2\delta)^2\end{aligned}\tag{4.3}$$

for a discount factor δ , where ϕ denotes the Gaussian density, and u is an integer that has been drawn uniformly from $\{1, 2, \dots, MC\}$. This parameter proposal scheme simplifies considerably the expression for the acceptance probability, as at each observation time t_j we sample from the previous posterior density $\pi(\theta | \bar{X}_{t_j})$, so that it will enter both the target posterior density and the proposal, and thus be cancelled out in the calculation of the acceptance probability.

4.2.2 The latent data points proposal

The idea behind the proposal density q from which the proposal latent data points will be sampled is that it should satisfy $\text{supp}(q) \subseteq \text{supp}(p)$ where p denotes the target density π in its unnormalized form. A good proposal would be one that makes the ratio p/q as close to a constant as possible. This is especially important for independence samplers, as the one used in this setting, as pointed out in Tierny (1994), in order to avoid that the algorithm spends too much time in a certain region of parameter space that it explores.

A proposal for latent data that has been discussed in Durham and Gallant (2002), and implemented in Roberts and Strammer (2001), and Golightly and Wilkinson (2006a) among others, is the Modified Diffusion Bridge proposal, based on an Euler scheme for the transition density. The idea behind it is quite simple: a Brownian bridge is in fact a Brownian motion that is conditioned upon terminating at a specific value within the interval of interest, that is, it bridges the values at each end of the interval. Using such a Brownian bridge is a way to reduce variance in Monte Carlo integration and Durham and Gallant (2002) show that it compares nicely to other transition density approximations like the Milstein

scheme. Thus, the proposal for the latent data points takes the form:

$$\begin{aligned}
q(X_{t_{i+1}} | X_{t_i}, \bar{X}_{t_M}; \theta) &= \phi(X_{t_{i+1}}, X_{t_i} + \tilde{\mu}_i, \tilde{\sigma}_i) \\
\text{where } \tilde{\mu}_i &= \frac{1}{M-i} (\bar{X}_{t_M} - X_{t_i}) \\
\tilde{\sigma}_i &= \Delta t \frac{1}{M-i} (M-i-1) \Lambda(X_{t_i})
\end{aligned} \tag{4.4}$$

where ϕ denotes the Gaussian density and $\Lambda(X_{t_i})$ is the volatility term of the process for X . Thus for each iteration $s = 1, \dots, MC$ we sample a latent data path $X_{t_j}^*, \dots, X_{t_{M-1}}^*$, so we have the joint proposal sample

$$\left(X_{t_j}^*, \dots, X_{t_{M-1}}^*; \theta \right) \sim \pi(\theta | \bar{X}_{t_j}) \prod_{i=j}^{M-2} q\left(X_{t_{i+1}}^* | X_{t_i}^*, X_{t_i}^*, \bar{X}_{t_M}; \theta \right) \tag{4.5}$$

A Metropolis-Hastings algorithm moves as follows: provided that we have obtained the proposed sample at iteration s and that we have a parameter and latent state sample obtained from the previous iteration $s-1$, we decide whether to keep the parameters and latent data from the previous iteration or alternatively replace them with the ones from the proposal. To this end we form the ratio

$$A = \frac{p(X_s^*, \theta_s^*) \tilde{q}(X_{s-1}, \theta_{s-1})}{p(X_{s-1}, \theta_{s-1}) \tilde{q}(X_s^*, \theta_s^*)}$$

where $(X_s^*, \theta_s^*) = (X_{t_j}^*, \dots, X_{t_{M-1}}^*; \theta_s^*)_s$ is the proposed sample at iteration s , $(X_{s-1}, \theta_{s-1}) = (X_{t_j}, \dots, X_{t_{M-1}}; \theta)_{s-1}$ is the previously accepted sample at iteration $s-1$, p denotes the target posterior density in its unnormalized form, and \tilde{q} is the proposal density (4.5). Replacing all terms in the expression, we obtain for the ratio A :

$$A = \frac{\prod_{i=j}^{M-1} \pi\left(X_{t_{i+1}}^* | X_{t_i}^*; \theta_s^* \right) \prod_{i=j}^{M-2} q\left(X_{t_{i+1}} | X_{t_i}, X_{t_i}^*, \bar{X}_{t_M}; \theta \right)}{\prod_{i=j}^{M-1} \pi\left(X_{t_{i+1}} | X_{t_i}; \theta \right) \prod_{i=j}^{M-2} q\left(X_{t_{i+1}}^* | X_{t_i}^*, X_{t_i}^*, \bar{X}_{t_M}; \theta_s^* \right)} \tag{4.6}$$

The standard Metropolis Hastings algorithm then accepts the new draw with probability $\alpha = \min(1, A)$, or else the draw is rejected and the last accepted draw is retained.

4.2.3 The algorithm

The algorithm for carrying out the Metropolis-Hastings scheme for sampling from the conditional posterior of parameters and latent data can be summarized as follows:

Initialization. Set $j = 0$. Initialize the augmented data points for each of the $s = 1, \dots, MC$ iterations by linearly interpolating between observations for the first interval. Initialize the parameter set for all s

by sampling from a prior density $\pi(\theta)$.

1. For each $s = 1, \dots, MC$:
 - Propose the parameters θ^* using (4.3)
 - Propose the latent data Y^* for the interval (t_j, t_{j+m}) using (4.4) for each $i = j + 1, \dots, M - 1$
 - Accept the parameter and latent data proposal with probability $\alpha = \min(1, A)$ with A given by (4.6), and set $(X_s, \theta_s) = (X_s^*, \theta_s^*)$, or else set $(X_s, \theta_s) = (X_{s-1}, \theta_{s-1})$.
2. Set $j = j + m$ and go to (1).

The resulting draws of latent data and parameters form a Markov chain, whose stationary distribution after an initial burn-in period is given by (4.1). The number of imputed data points that are needed could be determined by running the sampler for low values of m and consequently increasing the discretization points until there is no significant change in the posterior parameter samples.

4.2.4 Convergence

In order to assess the accuracy of the parameter estimates obtained as ergodic averages of the form $\hat{\theta}_{MC} = \frac{1}{MC} \sum_{i=1}^{MC} (\theta^i)$ we estimate their variance σ_θ^2 using the batch-mean approach (see Roberts (1996) and Tse et al. (2004)). To this end, we run the MCMC scheme for $MC = m \times n$ iterations with m batches of n draws each. We compute the mean of each batch $k = 1, \dots, m$ with $\hat{\theta}_k = \frac{1}{n} \sum_{i=(k-1)n+1}^{kn} (\theta^i)$. Then we obtain an estimate of σ_θ^2 using:

$$\hat{\sigma}_\theta^2 = \frac{n}{m-1} \sum_{k=1}^m \left(\hat{\theta}_k - \hat{\theta}_{MC} \right)^2 \quad (4.7)$$

and the Monte Carlo standard errors are obtained as $\sqrt{\frac{\hat{\sigma}_\theta^2}{MC}}$.

As well, as a diagnostic tool that allows us to see how well the Markov chain mixes, we compute the simulation inefficiency factor (SIF) (see Kim et al. (1998)), estimated as the variance of the ergodic averages σ_θ^2 , divided by the variance of the sample mean from a hypothetical sampler that draws independent random variables from the parameter posterior. In order to compute the latter variance, we use the output of the MCMC runs, as in Tse et al. (2004), and obtain $\bar{\sigma}_\theta^2 = \frac{1}{MC-1} \sum_{i=1}^{MC} \left(\theta^i - \hat{\theta}_{MC} \right)^2$ so that the SIF is estimated as:

$$SIF = \frac{\hat{\sigma}_\theta^2}{\bar{\sigma}_\theta^2} \quad (4.8)$$

4.2.5 Model comparison through Bayes factors

In order to compare the estimated multivariate diffusion models of asset returns, we follow the traditional Bayesian approach that makes use of the marginal likelihood of each (potentially nonnested) model. The marginal likelihood is obtained by integrating the likelihood function of each model \mathcal{M}_i with respect to the prior density:

$$p(X | \mathcal{M}_i) = \int p(X | \theta_i, \mathcal{M}_i) p(\theta_i | \mathcal{M}_i) d\theta_i$$

where θ_i are the parameters, corresponding to model \mathcal{M}_i . Then the Bayes factors for comparing model \mathcal{M}_i against \mathcal{M}_j are simply the ratio of the marginal likelihoods:

$$B_{ij} = \frac{p(X | \mathcal{M}_i)}{p(X | \mathcal{M}_j)} \quad (4.9)$$

We use the Laplace-Metropolis estimator of the marginal likelihood, proposed by Lewis and Raftery (1997) that relies on the posterior simulation output from the individual estimation of each model and approximates the integral using the Laplace method. Let us denote by θ_i^* the posterior parameter mean (or any other high density point of the parameter posterior). Then the logarithm of the marginal likelihood is estimated as:

$$\log(p(X | \mathcal{M}_i)) \approx \frac{d}{2} \log(2\pi) + \frac{1}{2} \log(|H^*|) + \log(p(\theta_i^*)) + \log(p(X | \theta_i^*, \mathcal{M}_i))$$

where d is the dimension of the diffusion, $p(\theta_i^*)$ is the parameter prior under model \mathcal{M}_i , H^* is the inverse Hessian of $\log(p(\theta_i^*)p(X | \theta_i^*, \mathcal{M}_i))$, $|H^*|$ is its determinant, and $p(X | \theta_i^*, \mathcal{M}_i)$ is the likelihood function, evaluated at θ^* .

Lewis and Raftery (1997) propose to estimate H^* by the sample covariance matrix of parameters from the MCMC output, so the only quantity that is left to be estimated is the likelihood function. The most straightforward estimator would be the one proposed by Pedersen (1995) that consists in averaging over the transition density implied by the Euler discretization. But as estimation was done by exploiting the information in both ends of each observation interval, a more efficient approach would be one that is similar to the Metropolis-Hastings update used for latent data. Elerian et al. (2001) discuss a class of importance sampling estimators of the likelihood function of the form:

$$p(\bar{X}_{t_M} | \bar{X}_{t_j}; \theta) = \int \frac{p(\bar{X}_{t_M}, X_{t_M} | \bar{X}_{t_j}; \theta)}{q(X_{t_M} | \bar{X}_{t_j}, \bar{X}_{t_M}; \theta)} q(X_{t_M} | \bar{X}_{t_j}, \bar{X}_{t_M}; \theta) dX_{t_M}$$

for an interval between two successive observations \bar{X}_{t_M} and \bar{X}_{t_j} . Thus, the modified Brownian bridge proposal density that we used for the Metropolis-Hastings update could be used in this setup as the

importance density q , which leads us to the following estimator of the likelihood function:

$$\hat{p}(\bar{X}_{t_M} | \bar{X}_{t_j}; \theta) = \frac{1}{M} \sum_{k=1}^M \frac{p(\bar{X}_{t_M}, X_{t_j}^k | \bar{X}_{t_j}; \theta)}{q(X_{t_j}^k | \bar{X}_{t_j}, \bar{X}_{t_M}; \theta)}$$

where $X_{t_j}^k, k = 1, \dots, M$ is a set of latent vectors between each pair of observations.

5 Estimation results

Although a joint estimation of each of the multivariate models is feasible, we propose to use a two-step procedure, as this allows us to choose the appropriate marginal distribution for each data series. Such a two-step approach is commonly used in discrete-time copula models (Patton (2004)), as it allows to avoid model misspecification. A two step approach is possible in our continuous time setup as well, as a system of independent univariate diffusions is obtained under the product copula, assuming an identity correlation matrix for the diffusion term.

5.1 Univariate diffusion

The copula construction leaves us the freedom to choose the most appropriate marginal distribution for each univariate series. A Bayes factors comparison selects a NIG distribution for small and large caps, and GH for Mid caps. Table 1 summarizes the estimation results for the parameters specific to each univariate series.

It is interesting to note that the parameter κ for all three series of data is different from 0 or 1, which would correspond to either a constant volatility diffusion for the state variables ($\kappa = 0$) or a zero drift diffusion ($\kappa = 1$). Further analysis of the MCMC output is offered on Fig. 1, Panels A through C, where we present the sample paths of each estimated parameter for the three data series, as well as autocorrelation plots for a lag up to 100, and kernel density estimate of the posterior parameter output. We do not have any significant autocorrelation for any of the parameters, which is a consistent result with Golightly and Wilkinson (2006a), who show a significant reduction in sample autocorrelations of the Simulation Filter as compared to the Gibbs sampler.

In order to examine whether the proposed diffusion replicates certain dynamic properties of the data, we simulate a very long series (of length 100 000) from the univariate NIG diffusion model for log prices X_{it} (3.16) and parameters corresponding to the Large cap series in Table 1, and examine the implied properties of their increments⁵. A stylized fact of asset returns is the persistence in autocorrelation in squared returns in contrast to the lack of autocorrelation in the original return series (except for possibly

⁵Similar results are obtained for the rest of the univariate data series considered.

Table 1. Parameter estimates for the univariate series

The table summarizes the posterior parameter estimates from the MCMC output. Monte Carlo standard errors are reported in paranthesis (multiplied by a factor of 1000) (obtained using the batch-mean approach). SIF refers to the simulation inefficiency factor for each parameter (its integrated autocorrelation time).

	Smallcap	Midcap	Largecap
α	3.0502	18.7839	10.6904
(MC s.e.)	(0.1616)	(0.5220)	(0.2193)
(SIF)	(0.0938)	(0.6694)	(0.6912)
β	-0.5911	0.4476	-1.5737
(MC s.e.)	(0.6329)	(2.9453)	(1.5404)
(SIF)	(0.1104)	(1.5392)	(1.7637)
δ^2	0.0301	0.0721	0.0410
(MC s.e.)	(0.0024)	(0.0011)	(0.0031)
(SIF)	(0.1219)	(1.0535)	(1.8122)
μ	6.7059	6.3101	6.5360
(MC s.e.)	(0.0249)	(0.0129)	(0.0102)
(SIF)	(0.1038)	(0.5407)	(0.4991)
σ^2	0.0406	0.0400	0.0082
(MC s.e.)	(0.0022)	(0.0030)	(0.0006)
(SIF)	(0.1142)	(1.4686)	(1.2930)
κ	0.6490	0.4670	0.5102
(MC s.e.)	(0.0373)	(0.0235)	(0.0850)
(SIF)	(0.0955)	(1.4322)	(1.7551)
λ	0.5	-1.4295	0.5
(MC s.e.)	-	(0.0519)	-
(SIF)	-	(1.1704)	-

Figure 1. Panel a. MCMC estimation output: Large caps, NIG diffusion

The figure displays the output from the MCMC estimation of the NIG diffusion for Large caps. The top figure represents the sample paths of the parameters from 10000 replications; the bottom left – the autocorrelation functions for the sample paths of each parameter, and the bottom right figure – the kernel density estimate of the parameter posterior distributions.

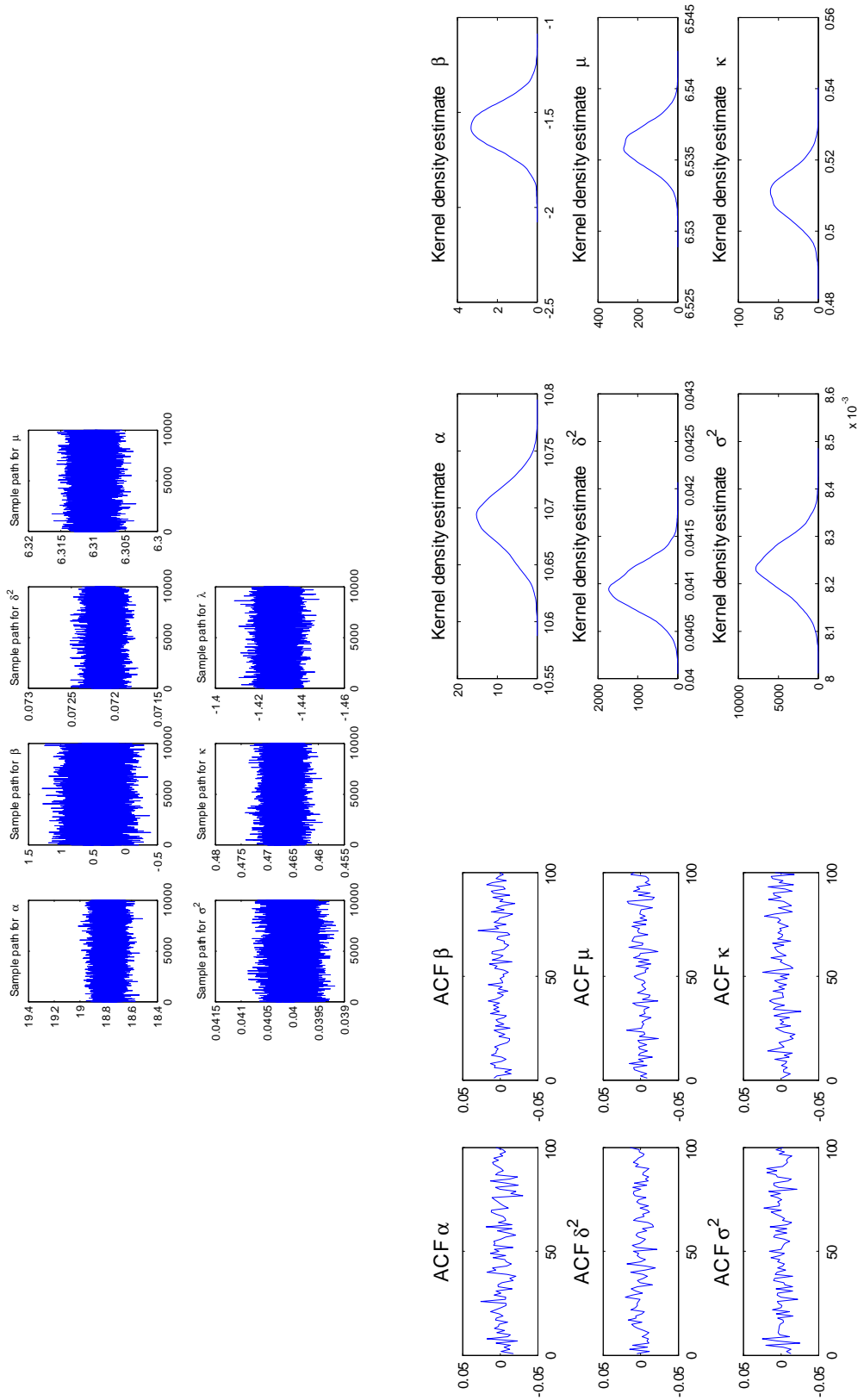


Figure 1, Panel b. MCMC estimation output: Mid caps, GH diffusion. The figure displays the output from the MCMC estimation of the NIG diffusion for Mid caps. The top figure represents the sample paths of the parameters from 10000 replications; the bottom left – the autocorrelation functions for the sample paths of each parameter, and the bottom right figure – the kernel density estimate of the parameter posterior distributions.

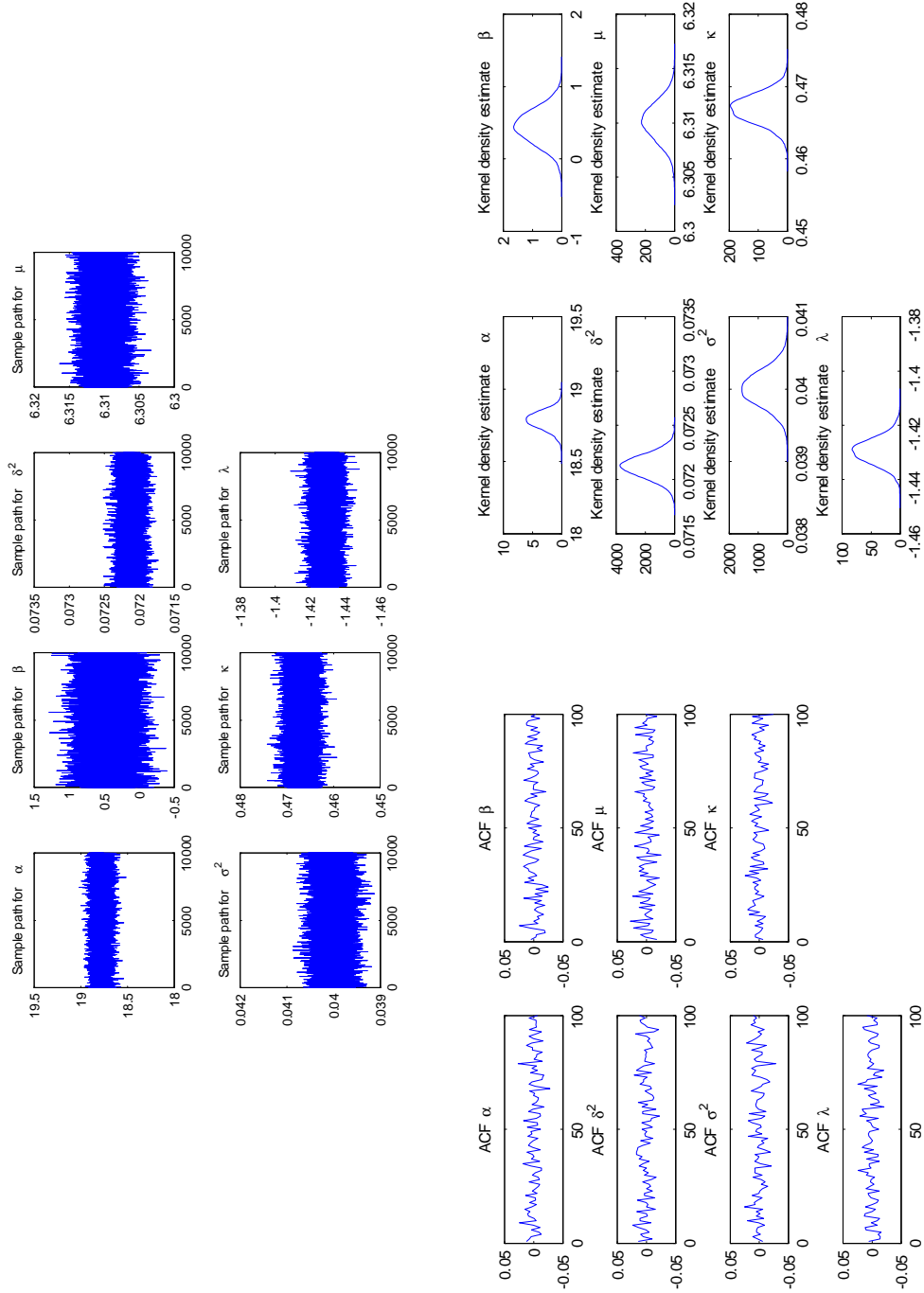
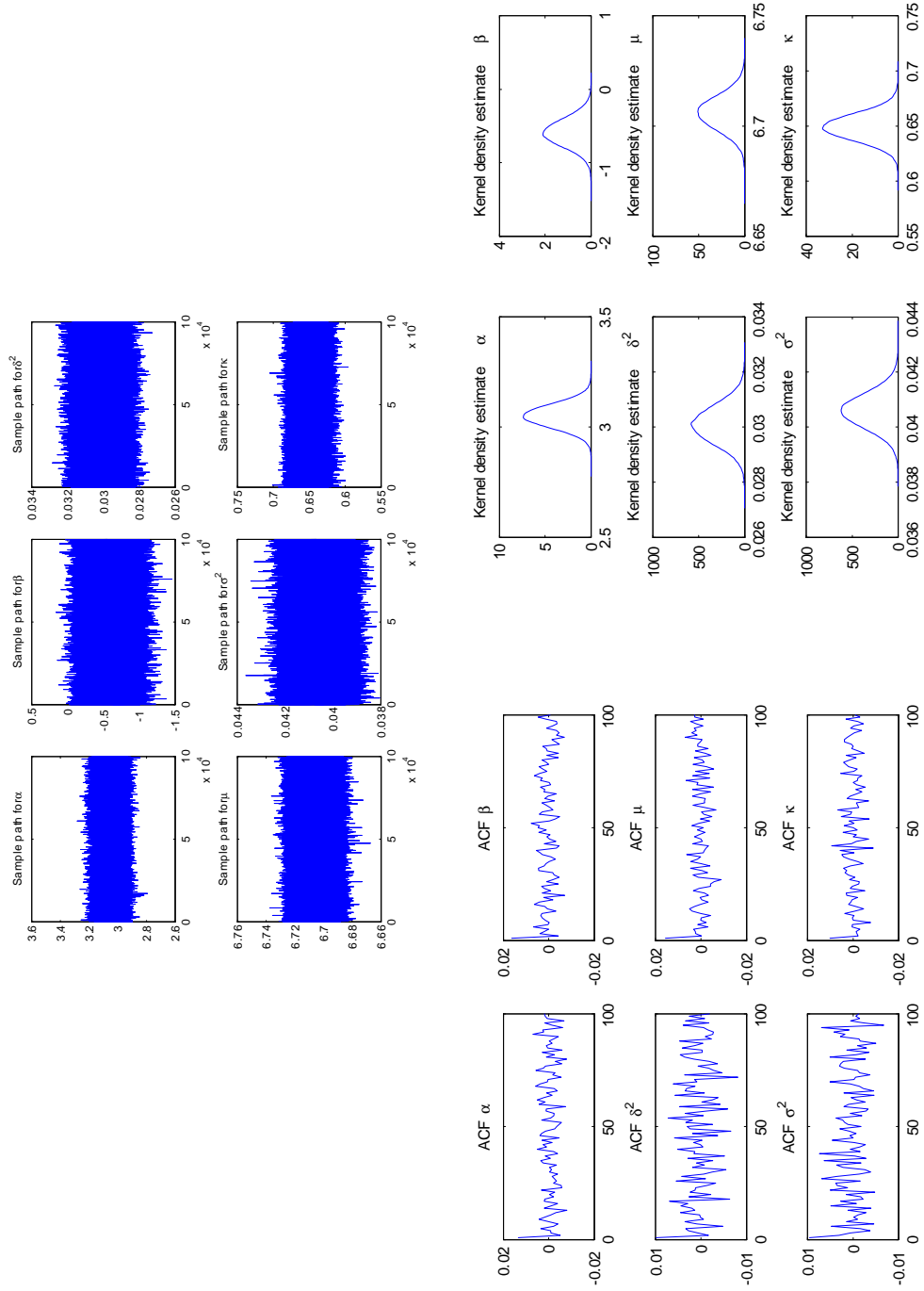


Figure 1, Panel c. MCMC estimation output: Small caps, NIG diffusion. The figure displays the output from the MCMC estimation of the NIG diffusion for Small caps. The top figure represents the sample paths of the parameters from 100000 replications; the bottom left – the autocorrelation functions for the sample paths of each parameter, and the bottom right figure – the kernel density estimate of the parameter posterior distributions.



the first lag). If we examine the autocorrelation patterns in the data and the long simulated series, we find that this property is actually captured by the model, as displayed on Fig. 2.

This finding is not surprising, if we consider the fact that the Euler discretization of a univariate diffusion of the generalized hyperbolic family can be considered as a special case of a nonlinear ARCH model Tse et al. (2004), and thus it can be expected to exhibit volatility clustering and long memory properties. The same behaviour is preserved in the multivariate specification as well.

Another important aspect of our analysis is the fit of each of the univariate diffusions to the empirical distribution of the data, as they provide the inputs for the probability integral transform in the copula construction. Fig. 3 illustrates the close replication of the stationary distribution by the considered marginal processes.

We further test the volatility specification of the model in (3.5) against a non-parametric estimator of the squared diffusion coefficient $V_n(x)$, based on quadratic variation, as proposed in Florens-Zmirou (1993):

$$V_n(x) = \frac{\sum_{j=1}^n 1_{|X_{i,t_j}-x|<h} (X_{i,t_j} - X_{i,t_{j-1}})^2}{\sum_{j=1}^n 1_{|X_{i,t_j}-x|<h} (t_j - t_{j-1})}$$

with a bandwidth parameter h . Fig. 4 displays the fit of the volatility specification for each of the univariate models. The U-shaped parametric volatility form (3.5) matches closely the non-parametric estimator. A constant volatility specification (achieved by setting κ_i to zero) would thus underestimate volatility in the cases when returns are in either tail of the distribution and fail to reproduce the empirical stylized fact that returns are highly volatile in extreme market downturns.

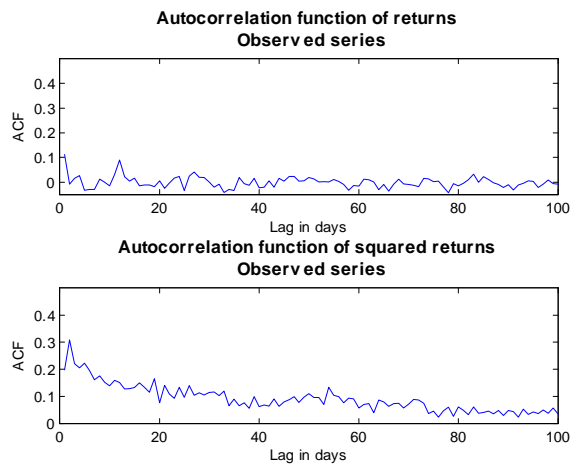
A check of the fit of the univariate models is done via the dynamic probability integral transform that uses the transitional probabilities of the discretized version of the diffusion between two consecutive observations with the Euler discretization scheme. For the model to be well specified, the series of uniform residuals should be *i.i.d.* $U(0,1)$. The residuals could then be analyzed using quantile plots, as illustrated on Fig. 5. A formal test could be conducted using the the statistic $stat = -2 \sum_{i=1}^n \log U_i \sim \chi_{2n}^2$, following Bibby and Sorensen (1997). For 3997 observations, the test statistic for the Small caps is 7.8677e+003, for the Mid cap it is 7.8797e+003, and for the Large cap it is 8.1278e+003, none of which gives reasons to reject the correct model specification.

5.2 Evidence of asymmetric tail dependence captured by a copula diffusion

Having obtained estimates of the univariate marginal distributions for each data series, we now turn to estimating the model parameters that pertain to the dependence structure. The bi-variate quantile plots for all three couples of data on Fig. 1 have shown a substantial degree of quantile ‘near’ tail dependence that does not fade away as we approach the tails of the distribution, especially the left one. As well, the

Figure 2. Autocorrelation plots for simulated and actual return series

Autocorrelation functions for the observed return series (Large cap). The top panel shows autocorrelation in returns, the bottom – autocorrelation in squared returns.



Autocorrelation functions for one of the simulated return series from the univariate diffusion with parameters for Large cap with length 100 000 using the Euler discretisation. The top panel shows autocorrelation in returns(Large cap), the bottom – autocorrelation in squared returns.

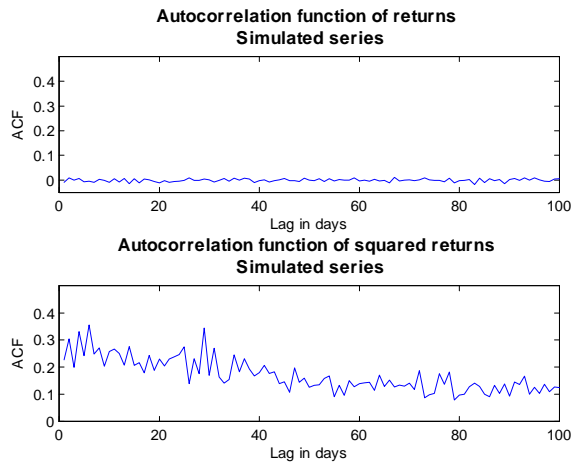


Figure 3. Fitting the marginal distributions

Empirical cumulative distribution functions plotted against the theoretical distribution of each of the three univariate models, with parameters given in Table 1.

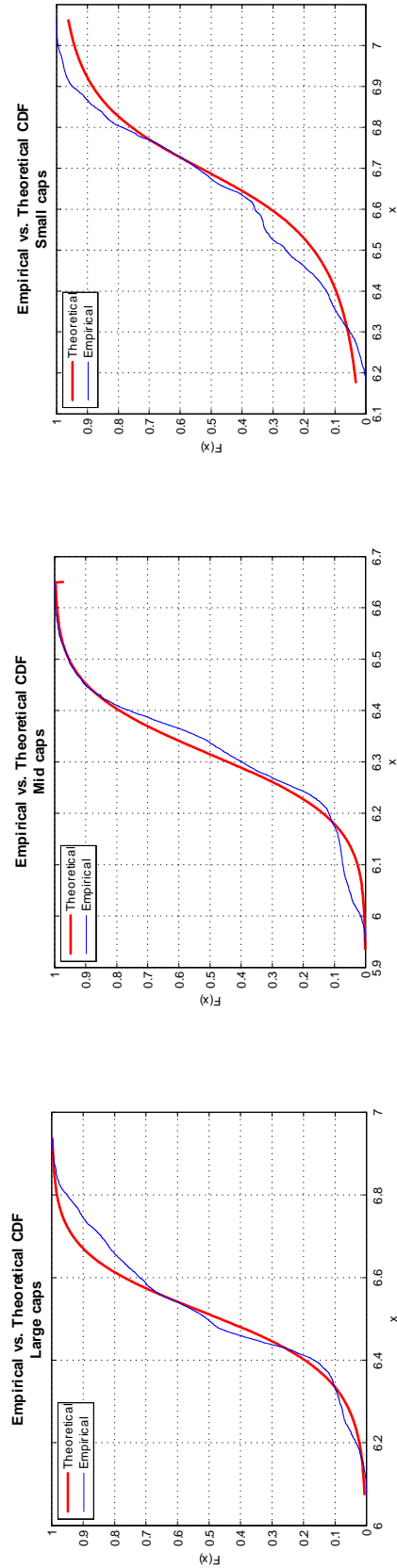


Figure 4. Fitting the volatility pattern

Empirical cumulative distribution functions plotted against the theoretical distribution of each of the three univariate models, with parameters given in Table 1.

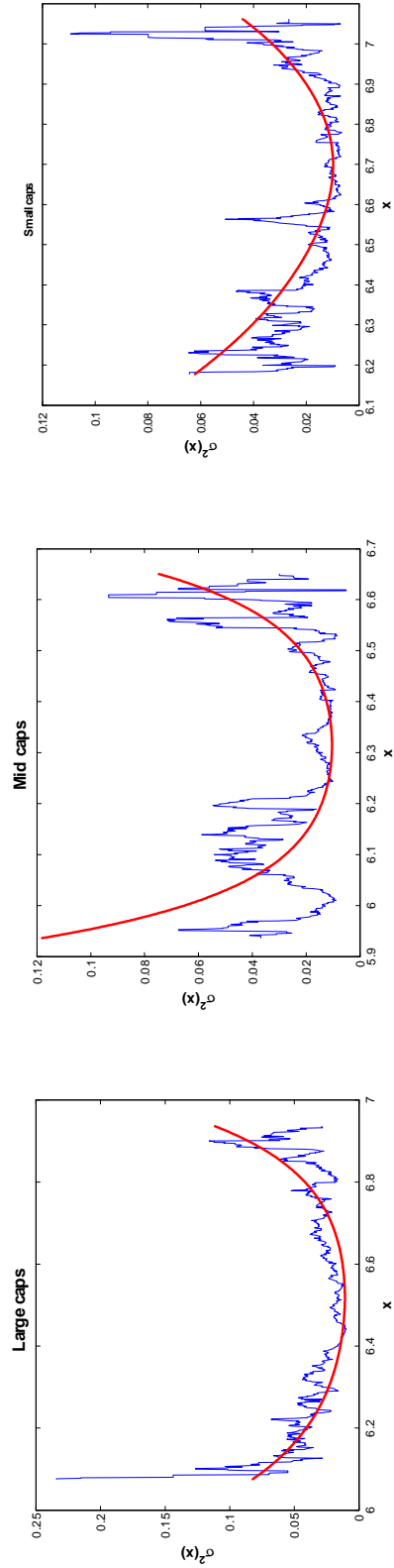
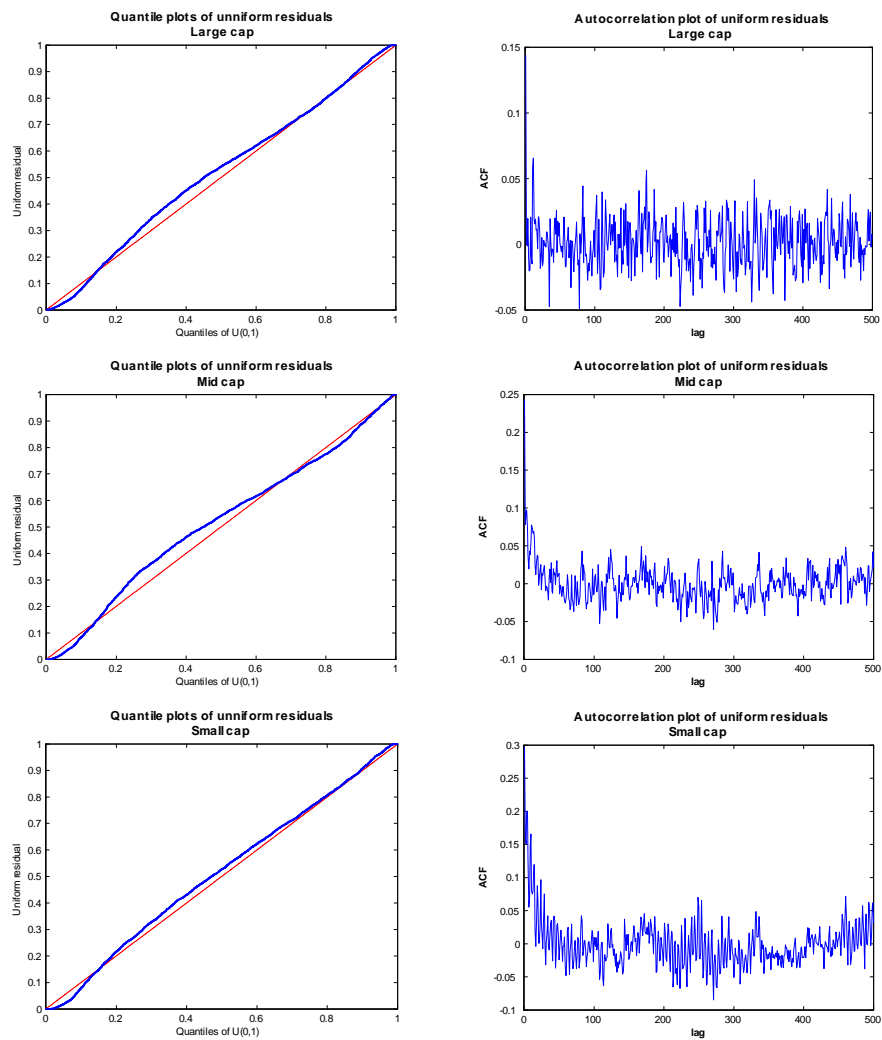


Figure 5. A formal check of the univariate diffusion models

Quantile plots and autocorrelation plots of the uniform residuals for each of the univariate diffusion models.



non-parametric test of exceedence correlations symmetry with exceedence levels chosen close to the tails, whose results are shown in Table 1, rejects symmetry for all couples of data, except the Large cap - Mid cap couple, for which exceedence correlations are high for both tails. In what follows, we verify whether a multivariate copula diffusion model could reproduce these properties of the data.

A good candidate for the purpose of modeling an asymmetric tail behaviour is the bi-variate Symmetrised Joe-Clayton copula, discussed in previous sections. It has two parameters, each one directly linked to the upper or lower tail dependence coefficient. So before we estimate a bi-variate diffusion model based on this copula function, let us first look at the levels of tail dependence that could be achieved through it. In order to do so, we need to obtain the levels of its parameters, implied by the data, so we first estimate the copula parameters from the unconditional distribution of each couple of the CRSP size indices. We apply the Canonical Maximum Likelihood estimation method which consists in first transforming the data into uniform variables using the empirical distribution, that is without imposing any parametric restrictions on the univariate marginals, and then estimating the copula parameters θ with MLE:

$$\hat{\theta} = \arg \max_{\theta} \sum_{t=1}^T \ln c \left(\hat{F}_i(x_i), \hat{F}_j(x_j); \theta \right), \quad i, j = 1, 2$$

where $\hat{F}_i(x_i)$ is the empirical CDF of x_i , and $c(\cdot)$ is the chosen parametric copula function. We estimate the copula parameters for two choices of copulas - the tail independent Gaussian and the asymmetric tail dependent SJC copula. Then for each dependence function we trace quantile plots (Fig. 6), where the levels of quantile dependence are obtained using (2.10), which are then contrasted against the quantile plots for the data itself.

The coefficients of upper and lower tail dependence for the Large cap - Mid cap couple are both high, which corresponds to the symmetric tail behaviour in terms of exceedence correlations that we reported in Table 1. However, the upper tail coefficients for the other two couples of data are low, especially for the Large cap - Small cap couple, where $\tau^U = 0$, while the lower tail dependence coefficients are significantly higher, confirming the evidence of asymmetric tail behaviour. The quantile dependence plots for the SJC copula are closer to the data, while those corresponding to a Gaussian copula deviate from it, especially in the left tail, where Gaussian dependence fades away for decreasing quantile levels.

Using the Simulation MCMC filter, we further estimate the bi-variate diffusion whose stationary distribution has a dependence structure governed by the asymmetric tail SJC copula. We keep the univariate marginal distribution parameteres fixed at their estimated values from the previous section. Results are reported in Table 2.

Note that the estimates of the upper and lower tail dependence parameters for the diffusion models are fairly close to the values obtained for the unconditional distribution, estimated using the Canonical

Figure 6. Quantile dependence plots for varying copula specifications

Plots of quantile dependence for all three couples of de-trended log-prices of the three CRSP indices formed on the basis of size deciles for the period 1986-2005 (small-cap (deciles 1-3), mid-cap (deciles 4-7), and large-cap (deciles 8-10)). Quantile dependence plots of the data are contrasted by quantile dependence that comes from two alternative parametric copula specifications: no-tail dependent Gaussian copula and asymmetric dependent Symmetrised Joe-Clayton (SJC) copula. Fitted parameters for the Gaussian copula are $\rho = 0.6969$ (Large cap vs. Mid cap), $\rho = 0.2754$ (Large cap vs. Small cap), and $\rho = 0.7593$ (Small cap vs. Mid cap). Fitted parameters (upper τ^U and lower τ^L tail dependence) for the SJC copula are $\tau^U = 0.4473, \tau^L = 0.6372$ (Large cap vs. Mid cap), $\tau^U = 0, \tau^L = 0.3230$ (Large cap vs. Small cap), and $\tau^U = 0.1811, \tau^L = 0.7109$ (Small cap vs. Mid cap).

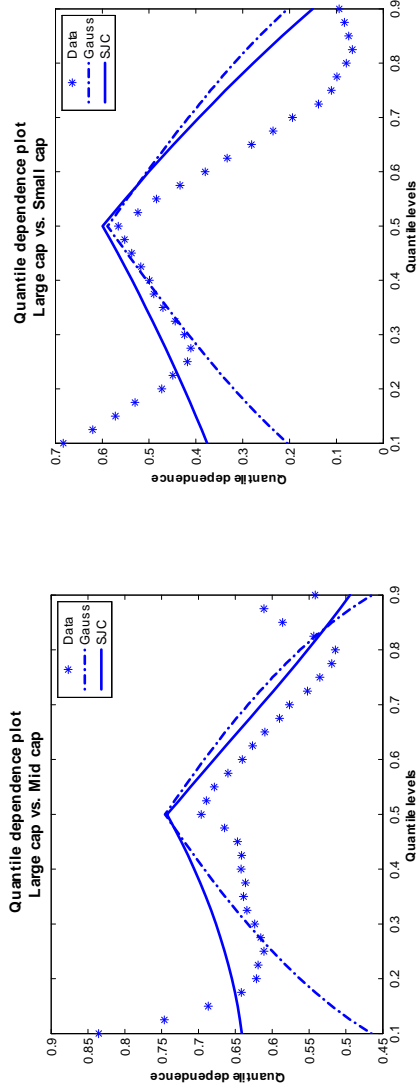


Table 2. Parameter estimates for a bivariate Symmetrised Joe-Clayton copula diffusion

The table summarizes the posterior parameter estimates from the MCMC output. Monte Carlo standard errors are reported in paranthesis (multiplied by a factor of 1000) (obtained using the batch-mean approach). SIF refers to the simulation inefficiency factor for each parameter (its integrated autocorrelation time).

	Large cap - Mid cap	Large cap - Small cap	Small cap - Mid cap
τ^U	0.4171	0.0484	0.1835
(MC s.e.)	(0.1568)	(0.1209)	(0.1934)
(SIF)	(0.2890)	(1.0160)	(0.5710)
τ^L	0.6724	0.2700	0.6602
(MC s.e.)	(0.1585)	(0.2359)	(0.0608)
(SIF)	(1.0040)	(1.0279)	(1.1880)
ρ	0.5968	0.6514	0.6682
(MC s.e.)	(0.0107)	(0.0049)	(0.0050)
(SIF)	(1.4428)	(0.9354)	(2.0860)

Maximum Likelihood with uniform variates from the empirical distribution. Tail asymmetry is found for all couples of data except for the Large cap - Mid cap couple, for which both tails show high dependence.

The obtained parameter estimates are then used to simulate long series from each of the three SJC copula diffusions. Further, we calculate the levels of quantile dependence for each bi-variate series using (2.10). From each bundle of simulated series and their corresponding levels of quantile dependence, we then determine the obtainable degrees of dependence for each quantile level in bands between the 5th and the 95th percentile. Thus, for each quantile level we show the degrees of quantile dependence that can be reached in 90% of the cases with a SJC copula diffusion. Results are presented on Fig. 7.

For the case of the Large cap - Mid cap couple, quantile dependence implied by the data generally falls within the bounds reachable under the estimated parameters for the SJC copula, with the exception of the extreme left tail, which would require an even higher left tail dependence parameter in order to accommodate the dependence found in the data. For the other two couples, the parameters for the SJC diffusion can reasonably well replicate the quantile dependence for the left tail.

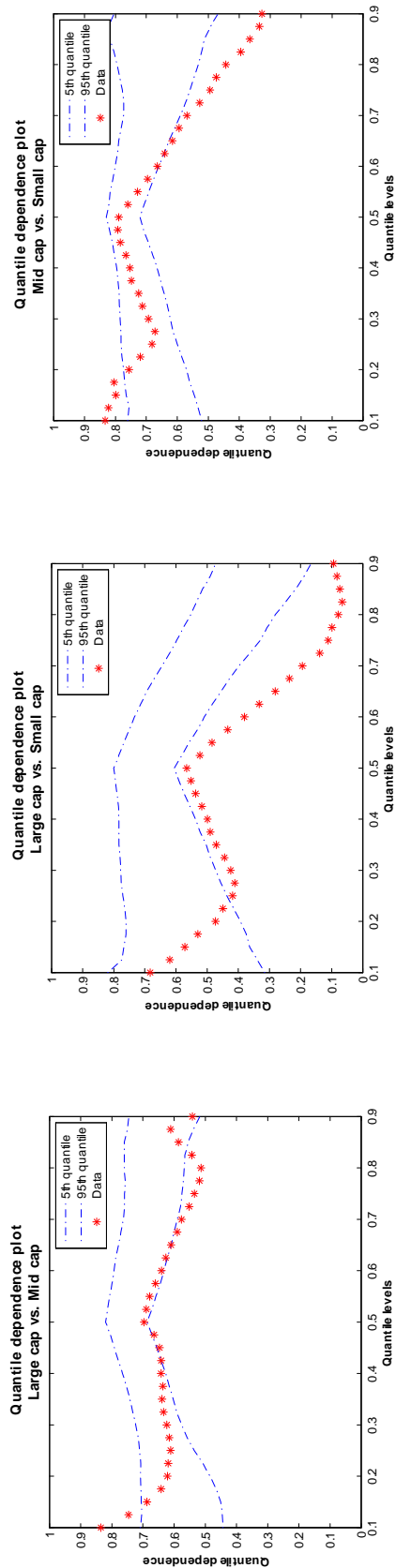
5.3 A generalization to higher dimensions

Even though the SJC copula is intuitively appealing as its parameters are directly linked to the coefficients of upper and lower tail dependence, it cannot be generalized in a straightforward manner to a higher dimension. That is why we turn to copula functions in the Archimedean family that allow an extension to higher dimensions without imposing symmetry. A first specification we consider is that of a nonnested Gumbel copula (3.23) and its survival counterpart (3.24). In this case the same parameter α governs the dependence structure for all n random variables.

Next, we consider a nested specification in order to allow for different dependence parameters among

Figure 7. Quantile dependence plots for simulated bi-variate SJC diffusions

Plots of quantile dependence for all three couples of de-trended log-prices of the three CRSP indices formed on the basis of size deciles for the period 1986-2005 (small-cap (deciles 1-3), mid-cap (deciles 4-7), and large-cap (deciles 8-10)). Quantile dependence plots of the data are contrasted by quantile dependence that comes from simulated series of bi-variate Symmetrised Joe-Clayton (SJC) copula diffusions with parameters taken from Table 2. 950 simulated series of length 50 000, using the Euler discretization, are used to determine the quantile dependence levels, obtainable between the 5th and the 95th percentile bands for a SJC diffusion.



couples with the highest dependence being achievable for the most deeply nested couple. The three-variate nested Archimedean copula, expressed in terms of the copula generator and its inverse is given by (3.26). Thus, we pick up the size decile couple that has the highest dependence and model it as the most deeply nested couple. The generating function for this couple is φ_1 with a dependence parameter α_1 . Thus we obtain the first copula, $C(u_1, u_2; \alpha_1) = \varphi_1^{-1}(\varphi_1(u_1) + \varphi_1(u_2))$. We then couple it with the third remaining data series using a second generating function φ_2 with a dependence parameter α_2 that implies lower dependence than α_1 and obtain the nested copula $C(u_1, u_2, u_3; \alpha_1, \alpha_2) = \varphi_2^{-1}(\varphi_2(C(u_1, u_2; \alpha_1)) + \varphi_2(u_3))$. This subsequent nesting of generating functions requires that they are quite parsimonious in nature in order to keep the resulting copula function tractable. The Gumbel copula that we use is a good candidate for that. In our application we use either of the size decile couples as the most deeply nested one, although the most fitted couple for that is the Large cap - Mid cap one, as it implies high dependence in both tails.

In all cases we consider the mixture copula function as defined in (3.27) which combines the two extreme value Gumbel copulas with the tail independent Gaussian one. If the estimate of the weight for the Gaussian copula goes close to 1, then our series is asymptotically independent. Otherwise there is some degree of dependence in either of the tails, depending on the weighting of the Gumbel copula or its survival counterpart. As well, in order to allow for richer parameterization of the dependence structure, we consider a mixture copula of the two extreme value ones with the Student's t as in (3.28). In this case we always have asymptotic dependence, unless the degrees of freedom parameter of the t -copula goes to infinity.

Estimation results for the multivariate diffusion with a Gaussian dependence structure is given in the first column of Table 3. Then we add the three alternative cases of a diffusion with tail dependence as implied by the nested mixture copula ((3.27) with nested Gumbel and Survival Gumbel), and finally we consider the most parsimonious specification where there is only one parameter that determines upper tail dependence, and one for lower tail dependence ((3.27) with non-nested Gumbel and Survival Gumbel copulas).

The relatively high and symmetric lower and upper tail dependence coefficients for the Mid-Large cap couple that we found earlier are confirmed in the estimation results for the nested Gaussian-Gumbel-Survival Gumbel diffusion for the case where it is most deeply nested in the copula specification (Table 2, second column). The two parameters that determine upper and lower tail dependence for this couple, α_1^G and $\bar{\alpha}_1^G$ respectively, are almost equal, pointing to tail symmetry. As well, for this particular couple we have $\tau_G^U = 0.7712$ and $\tau_G^L = 0.7150$, indicating close symmetry in both tails. Further, for the two alternative cases for which Large-Small or Mid-Small are the most deeply nested couples, the lower tail dependence parameter $\bar{\alpha}_1^G$ is lower than the upper tail dependence parameter α_1^G , indicating higher dependence in the left tail, again confirming the previously found evidence.

Table 3. Parameter estimates for the dependences structure (tri-variate diffusion, Gaussian underlying)

Estimation results for the trivariate diffusions using the Gaussian copula, the nested Gaussian-Gumbel-Survival Gumbel (Ga-G-SG) mixture copula (the most deeply nested couple is given in paranthesis), the nonnested Gaussian-Gumbel-Survival Gumbel (Ga-G-SG) mixture copula. Monte Carlo standard errors (multiplied by a factor of 1000), and Simulation Inefficiency Factors (SIF) are given in paranthesis. The first three parameters (R_{12}, R_{13}, R_{23}) correspond to the off-diagonal entries of the correlation matrix R_{G_a} for the Gaussian copula. The parameters α_1^G and α_2^G are the dependence parameters for the nested Gumbel copula, and the parameters $\bar{\alpha}_1^G$ and $\bar{\alpha}_2^G$ are the dependence parameters for the nested Survival Gumbel copula. For the nonnested case, the relevant parameters are α_1^G for the Gumbel copula and $\bar{\alpha}_1^G$ for the Survival Gumbel copula. ω^G and $\bar{\omega}^G$ are the corresponding weights for the Gumbel and the survival Gumbel copula for the mixture model. The parameters ρ_{12}, ρ_{13} , and ρ_{23} are the off-diagonal entries of the correlation matrix in the diffusion specification. Results are obtained for 50000 Monte Carlo replications with a thinning factor of 5 with 10 latent data points simulated between each pair of observations.

	Gaussian	Ga-G-SG (Large-Mid cap)	Ga-G-SG (Large-Small cap)	Ga-G-SG (Small-Mid cap)	Ga-G-SG (nonnested)
R_{12}	0.5671	0.5347	0.4636	0.6634	0.5758
MC s.e.	0.3701	0.3326	0.7224	0.6114	0.3537
SIF	0.8621	1.0437	2.3408	0.6891	0.9540
R_{13}	0.2723	0.5179	0.7443	0.3907	0.2571
MC s.e.	0.7875	0.4191	0.6868	0.4915	0.5131
SIF	0.7359	0.7188	2.7202	0.8441	0.7251
R_{23}	0.5207	0.4152	0.6110	0.3085	0.4698
MC s.e.	0.4399	0.3302	0.6260	0.5521	1.3536
SIF	0.9162	1.6992	0.8855	1.2236	1.5260
α_1^G	-	0.2972	0.3358	0.3318	0.4494
MC s.e.	-	0.3546	0.8711	0.3463	0.3541
SIF	-	0.5754	1.5005	1.5945	1.2328
α_2^G	-	0.6335	0.6238	0.7235	-
MC s.e.	-	0.1928	0.4072	0.7235	-
SIF	-	0.9156	2.6644	1.0750	-
$\bar{\alpha}_1^G$	-	0.3618	0.1993	0.2613	0.4354
MC s.e.	-	0.1998	0.3006	0.5387	1.0229
SIF	-	0.2375	1.8385	1.3787	1.6558
$\bar{\alpha}_2^G$	-	0.6544	0.6415	0.6107	-
MC s.e.	-	0.4667	0.3408	0.7371	-
SIF	-	0.8040	1.0323	1.1141	-

Table 3 cont.

	Gaussian	Ga-G-SG (Large-Mid cap)	Ga-G-SG (Large-Small cap)	Ga-G-SG (Small-Mid cap)	Ga-G-SG (nonnested)
ω^G	-	0.3321	0.2107	0.2431	0.3832
MC s.e.	-	1.0111	0.5641	0.3543	0.7265
SIF	-	2.0983	1.3229	0.8943	1.0348
$\bar{\omega}^G$	-	0.2853	0.2752	0.2531	0.2324
MC s.e.	-	0.3789	0.2519	0.6596	0.3619
SIF	-	1.4739	0.7950	1.2326	2.1457
ρ_{12}	0.7894	0.7917	0.7795	0.6935	0.8287
MC s.e.	0.0195	0.0086	0.0080	0.0095	0.0104
SIF	1.2371	0.2271	0.7370	0.6121	1.2730
ρ_{13}	0.5078	0.5089	0.5185	0.4685	0.5499
MC s.e.	0.0189	0.0229	0.0291	0.0236	0.0105
SIF	0.8625	0.9588	1.5205	0.5720	0.6771
ρ_{23}	0.7162	0.7158	0.6760	0.5676	0.7366
MC s.e.	0.0209	0.0067	0.0102	0.0159	0.0137
SIF	0.8581	0.7418	1.1571	0.9287	1.1969

When we consider the multivariate diffusion with extreme dependence modeled with the non-nested version of the Gumbel and Survival Gumbel copulas, we find almost symmetric tail dependence (the values for the parameters α_1^G and $\bar{\alpha}_1^G$ imply tail dependence coefficients of $\tau_G^U = 0.6345$ and $\tau_G^L = 0.6477$). As in this case there is only one parameter governing dependence in either the left or the right tail across all data series, extreme dependence for some couples may be over/underestimated.

Table 4 reports the results when we replace the Gaussian copula with a Student’s t one (see (3.28)).

As with the case of the Gaussian copula in the mixture model, here again the parameters, driving upper and lower tail dependence for the most deeply nested couple (the Large-Mid cap one) are very close, and imply tail coefficients of $\tau_G^U = 0.7870$ and $\tau_G^L = 0.7917$.

5.4 Model selection through Bayes factors

The multivariate diffusion models considered above imply different dependence structures through their stationary distributions. Bayes factors provide us with a guideline of how to select a model among the provided alternatives. So far we have seen that the mixture model with either a Gaussian or a t -copula, combined with the nested version of the extreme value Archimedean copulas provide the richest specification in terms of tail dependence modeling. In what follows we verify whether either of those models is selected vs. a more parsimonious alternative on the basis of the Bayes factor criterion.

We compute the log of Bayes factors, following (4.9) as $\log(p(Y | \mathcal{M}_b)) - \log(p(Y | \mathcal{M}_j))$. As a benchmark model (\mathcal{M}_b) we take either the Gaussian or the Student’s t -extreme value nested mixture copula diffusion (with the Large-Mid cap couple being the most deeply nested one). The alternatives

Table 4. Parameter estimates for the dependences structure (tri-variate diffusion, Student's t underlying)

Estimation results for the trivariate diffusions using the Student's t copula, the Student's t – nonnested Survival Gumbel mixture copula, and the Student's t – nested Gumbel - Survival Gumbel mixture copula (the most deeply nested couple is given in paranthesis). Monte Carlo standard errors (multiplied by a factor of 1000), and Simulation Inefficiency Factors (SIF) are given in paranthesis. The first three parameters (R_{12}, R_{13}, R_{23}) correspond to the off-diagonal entries of the correlation matrix R_T for the Student's t copula. The parameters α_1^G and α_2^G are the dependence parameters for the nested Gumbel copula, and the parameters $\bar{\alpha}_1^G$ and $\bar{\alpha}_2^G$ are the dependence parameters for the nested Survival Gumbel copula. For the nonnested case, the relevant parameters are α_1^G for the Gumbel copula and $\bar{\alpha}_1^G$ for the Survival Gumbel copula. ω^G and $\bar{\omega}^G$ are the corresponding weights for the Gumbel and the survival Gumbel copula for the mixture model. γ is the degrees of freedom parameter for the Student's t copula. The parameters ρ_{12}, ρ_{13} , and ρ_{23} are the off-diagonal entries of the correlation matrix in the diffusion specification. Results are obtained for 50000 Monte Carlo replications with a thinning factor of 5 with 10 latent data points simulated between each pair of observations.

	t	t-G-SG (Large - Mid cap)	t-SG (nonnested)
R_{12}	0.4408	0.2574	0.5266
MC s.e.	0.5433	1.4015	0.6040
SIF	1.3619	0.7629	1.3392
R_{13}	0.5273	0.2362	0.4154
MC s.e.	0.6911	0.9873	0.6353
SIF	0.9564	1.0469	0.8209
R_{23}	0.3334	0.3161	0.4461
MC s.e.	0.5146	0.5147	0.9027
SIF	1.1373	1.1320	0.9049
α_1^G	-	0.2786	-
MC s.e.	-	0.2191	-
SIF	-	0.5660	-
α_2^G	-	0.6570	-
MC s.e.	-	0.5395	-
SIF	-	1.0512	-
$\bar{\alpha}_1^G$	-	0.2730	0.3434
MC s.e.	-	0.2961	0.5440
SIF	-	0.6114	0.7326
$\bar{\alpha}_2^G$	-	0.6660	-
MC s.e.	-	0.5939	-
SIF	-	1.3265	-

Table 4 cont.

	t	t-G-SG (Large - Mid cap)	t-SG (nonnested)
ω^G	-	0.5118	-
MC s.e.	-	0.4382	-
SIF	-	0.7870	-
$\overline{\omega}^G$	-	0.1529	0.2829
MC s.e.	-	0.2248	0.9130
SIF	-	1.4495	1.9105
γ	5.4774	3.9575	4.8266
MC s.e.	4.8170	2.4907	5.8874
SIF	0.8904	0.7732	0.9437
ρ_{12}	0.8184	0.7837	0.8166
MC s.e.	0.0074	0.0223	0.0171
SIF	0.3969	1.1166	1.2428
ρ_{13}	0.5113	0.4922	0.5522
MC s.e.	0.0286	0.0296	0.0085
SIF	1.5033	0.9770	0.6370
ρ_{23}	0.7165	0.7045	0.7372
MC s.e.	0.0085	0.0129	0.0092
SIF	0.3875	0.8620	0.6073

considered (\mathcal{M}_j) are the tail independent Gaussian diffusion, the symmetric tail dependent t-copula diffusion, or any of the non-nested specifications considered. Results are provided in Table 5.

The Bayes factor selection criterion suggests that the extreme value nested mixture copulas should be selected. As well, the more parsimonious dependence structure, implied by the nonnested copulas is highly detrimental to the models, at least for the purposes of selection through Bayes factors.

Further, when we compare the two nested mixture models, Bayes factors point in favour of the Student's t mixture copula, with a value for the log of the Bayes factor of 9.06 when the latter is taken as the benchmark \mathcal{M}_b . But still this is far from the significantly higher values of the factors when the other alternative models are considered. This is not surprising, as the two nested mixture models are close in the way they treat the dependence structure, while the model with the Student's t underlying copula provides a more versatile way to account for dependence between extreme realizations.

6 Discussion and concluding remarks

In this paper we introduce a multivariate diffusion model for stock prices based on copula functions that is able to reproduce a number of stylized facts for both the univariate return series and the dependence structure. It extends the univariate stationary diffusion modeling based on the Generalized Hyperbolic family of distributions that has proved successful in replicating dynamic return characteristics as a slowly decay-

Table 5. Bayes factors

Log Bayes factors for tri-variate diffusions with dependence modeled using alternative copula functions. Benchmark models (M_b) are those involving the mixed copula diffusions with an Elliptic copula and the nested version of the extreme value Gumbel - Survival Gumbel copulas (Large-Mid cap being the most deeply nested couple). Two choices for the Ellitic copula are considered: the Gaussian one (Gauss-G-SG), and the Student's t one (t-G-SG). The four alternative diffusions ($M_j, j = 1, \dots, 4$) are a Gaussian, a Student's t (t), and two nonnested versions of the mixture copula diffusion: the Gaussian-Gumbel-Survival Gumbel (Gauss-G-SG (nonnested)) and the Student's t - Survival Gumbel (t-SG (nonnested)).

	Gaussian	t	Gauss-G-SG (nonnested)	t-SG (nonnested)
Gauss-G-SG (Large - Mid cap)				
Bayes factors	206.52	208.67	464.89	386.32
t-G-SG (Large - Mid cap)				
Bayes factors	215.58	217.73	473.95	395.02

ing autocorrelation function of squared returns (or volatility clustering effect, as alternatively modeled under stochastic volatility or an ARCH process), or static properties like thick tails and excess kurtosis. Seeking to reproduce increased dependence when there are extreme market downturns, we extend the copula-GARCH approach to a continuous-time diffusion framework where the stationary distribution of the process is modeled using a copula function that can account for tail dependence.

There are a number of ways in which the model can be extended. There is overwhelming empirical evidence that the correlation of asset returns changes dynamically through time. Popular discrete time approaches include the GARCH-DCC model of Engle (2002), while in continuous time a promising alternative is the Wishart process of Bru (1991). Our present model specification imposes constant conditional correlation for asset returns. It can be extended to a more general model where correlation is modeled as either a function of the state variables of the model itself, or rendered stochastic by being represented as a function of exogenous factors. There is empirical evidence that the dynamics of asset return correlations are linked to the phase of the business cycle and tend to increase in periods of recession (e.g. Ledoit et al. (2003) and Erb et al. (1994)). As well, Longin and Solnik (1995) find that correlations for international stock market indices increase during hectic periods of high volatility.

Another possible extension concerns the dependence structure of the assets, modeled through a copula function. The present specification assumes that the parameters governing dependence are fixed. A number of studies have addressed time variation in dependence through a dynamic copula approach. In the case of modeling asymmetric dependence between exchange rates, Patton (2004) find significant implications of the time variation in the copula dependence parameters, while Goorbergh et al. (2003) find substantial pricing differences for multivariate options when a dynamic copula model is used contrary to one with a fixed dependence structure, especially for market conditions marked with increased volatility.

In our setup, time variation in the dependence parameter could be achieved by modeling it as a function of exogenous factors that are stochastically time varying themselves and that have a potential of explaining increased dependence in extreme down markets.

References

- Ait-Sahalia, Y.: 1999, Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance* **54**, 1361–1395.
- Ait-Sahalia, Y.: 2003, Closed-form likelihood expansions for multivariate diffusions. Working paper.
- Ait-Sahalia, Y.: 2008, Closed-form likelihood expansions for multivariate diffusions, *The Annals of Statistics* **36**, 906–937.
- Ang, A. and Chen, J.: 2002, Asymmetric correlations of equity portfolios, *Journal of Financial Economics* **63**, 443–494.
- Barndorff-Nielsen, O. E.: 1978, Hyperbolic distributions and distributions on hyperbolae, *Scandinavian Journal of Statistics* **5**, 151–157.
- Barndorff-Nielsen, O. E.: 1977, Exponentially decreasing distributions for the logarithm of the particle size, *Proceedings of the Royal Society London A* **353**, 401–419.
- Barndorff-Nielsen, O. E.: 1995, Normal inverse gaussian processes and the modeling of stock returns. Rep. No. 300.
- Barndorff-Nielsen, O. E. and Blaesid, P.: 1981, *Hyperbolic Distributions and Ramifications: Contributions to Theory and Application*, Vol. 4, c. taillie and g. patil and b. baldessari edn, pp. 19–44.
- Bauer, C.: 2000, Value-at-risk using hyperbolic distributions, *Journal of Economics and Business* **52**, 435–453.
- Bibby, B. M. and Sorensen, M.: 1997, A hyperbolic diffusion model for stock prices, *Finance Stoch.* **1**, 25–41.
- Bibby, B. M. and Sorensen, M.: 2003, *Hyperbolic Processes in Finance*, s. rachev edn, Elsevier Science, pp. 211–248.
- B.M.Bibby and Sorensen, M.: 1995, Martingale estimating functions for discretely observed diffusion processes, *Bernoulli* **1**, 17–39.
- Bru, M. F.: 1991, Wishart processes, *Journal of Theoretical and Applied Probability* **4**, 725–743.
- Carr, P. and Wu, L.: 2004, Time-changed levy processes and option pricing, *Journal of Financial Economics* **71**, 113–141.
- Chen, X., Hansen, L. P. and Scheinkman, J. A.: 2002, Principal components and the long run. Working paper.

- Cherubini, U., Luciano, E. and Vecchiato, W.: 2004, *Copula Methods in Finance*, John Wiley and Sons, England.
- Christoffersen, P., Errunza, V. R., Jacobs, K. and Jin, X.: 2011, Is the potential for international diversification disappearing? Working Paper.
- Das, S. and Uppal, R.: 2004, Systemic risk and international portfolio choice, *Journal of Finance* **59**, 2809–2834.
- Dupire, B.: 1994, Pricing with a smile, *Risk* **Jan.**, 18–20.
- Durham, G. and Gallant, A.: 2002, Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, *Journal of Business and Economic Statistics* **20**(3), 297–316.
- Eberlein, E. and Keller, U.: 1995, Hyperbolic distributions in finance, *Bernoulli* **1**, 281–299.
- Elerian, O., Chib, S. and Shephard, N.: 2001, Likelihood inference for discretely observed non-linear diffusions, *Econometrica* **69**, 959–993.
- Embrechts, P., McNeil, A. and Strautman, D.: 2002, *Correlation and Dependence in Risk Management: Properties and Pitfalls*, m. dempster edn, Cambridge University Press, Cambridge, pp. 176–223.
- Engle, R.: 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* **20**, 339–350.
- Engle, R. and Kelly, B.: 2009, Dynamic equicorrelation. Working Paper, New York University.
- Eraker, B.: 2001, MCMC analysis of diffusion models with applications to finance, *Journal of Business and Economic Statistics* **19**(2), 177–191.
- Eraker, B., Johannes, M. and Polson, N.: 2003, The impact of jumps in returns and volatility, *Journal of Finance* **19**(53), 1269–1300.
- Erb, C., Harvey, C. and Viskanta, T.: 1994, Forecasting international equity correlations, *Financial Analyst Journal* **November/December**, 32–45.
- Fermanian, J.-D. and Wegkamp, M.: 2004, Time-dependent copulas. Working paper.
- Florens-Zmirou, D.: 1993, On estimating the diffusion coefficient from discrete observations, *Journal of Applied Probability* **30**, 790–804.
- Frees, E. W. and Valdez, E.: 1998, Understanding relationships using copulas, *North American Actuarial Journal* **2**(1), 1–25.
- Golightly, D. and Wilkinson, J.: 2006a, Bayesian sequential inference for nonlinear multivariate diffusions, *Statistics and Computing* **16**, 323–338.
- Golightly, D. and Wilkinson, J.: 2006b, Bayesian sequential inference for stochastic kinetic biomechanical network models, *Journal of Computational Biology* **3**(13), 838–851.
- Goorbergh, R., Genest, C. and Werker, B.: 2003, Multivariate option pricing using dynamic copula models. CentER Discussion paper No. 2003-122.

- Gumbel, E.: 1960, Bivariate exponential distributions, *Journal of the American Statistical Association* **55**, 698–707.
- Hansen, L. P. and Scheinkman, J. A.: 1995, Back to the future: Generating moment implications for continuous-time markov processes, *Econometrica* **63**(4), 767–804.
- Heston, S. L.: 1993, A closed-form solution for options with stochastic volatility with applications to bonds and currency options, *The Review of Financial Studies* **6**(2), 327–343.
- Hong, Y., Jun, T. and Zhou, G.: 2003, Asymmetric correlations of equity portfolios, *Journal of Financial Economics* **63**, 443–494.
- Jaschke, S. R.: 1997, A note on stockastic volatility, GARCH models, and hyperbolic distributions. Working paper, SFB 373, Humboldt-Universitat Berlin.
- Joe, H.: 1997, *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- Johannes, M., Polson, N. and Stroud, J.: 2004, Sequential parameter estimation in stochastic volatility models with jumps.
- Jondeau, E. and Rockinger, M.: 2002, Conditional dependency of financial series: The copula-garch model. FAME Research Paper No. 69.
- Jones, C.: 2003, Nonlinear mean reversion in the short-term interest rate, *The Review of Financial Studies* **16**, 793–843.
- Kim, S., Shephard, N. and Chib, S.: 1998, Stochastic volatility: Likelihood inference and comparison with arch models, *Review of Economic Studies* **65**, 361–393.
- Kunz, A.: 2002, Extremes in multivariate diffusions in finance: A data analysis. Working paper.
- Ledoit, O., Santa-Clara, P. and Wolf, M.: 2003, Flexible multivariate GARCH modeling with an application to international stock markets, *Review of Economics and Statistics* **85**, 735–747.
- Lewis, M. and Raftery, A.: 1997, Estimating bayes factors via posterior simulation with the laplace-metropolis estimator, *Journal of the American Statistical Association* **92**, 648–655.
- Liu, J. and West, M.: 2001, *Combined Parameter and State Estimation in Simulation-Based Filtering*, a. doucet, n. de freitas and n. gordon edn, Springer, Berlin.
- Longin, F. and Solnik, B.: 1995, Is the correlation in international equity returns constant: 1960-1990?, *Journal of International Money and Finance* **14**, 3–26.
- Longin, F. and Solnik, B.: 2001, Extreme correlation of international equity markets, *Journal of Finance* **56**, 649–675.
- Mari, D. D.: 2002, Some measures of extremal dependence in bivariate distributions.
- Nelsen, R. B.: 1999, *An Introduction to Copulas*, Springer, New York.
- Nelson, D. B.: 1990, ARCH models as diffusion approximations, *Journal of Econometrics* **45**, 7–38.
- Patton, A.: 2004, Modeling asymmetric exchange rate dependence. Working Paper, LSE.

- Pedersen, A. R.: 1995, Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes, *Bernoulli* **1**, 257–279.
- Philips, P.: 1991, Error correction and long-run equilibrium in continuous time, *Econometrica* **59**, 967–980.
- Poon, S. H., Rockinger, M. and Tawn, J.: 2004, Extreme value dependence in financial markets: Diagnostics, models, and financial implications, *Review of Financial Studies* **17**(2), 581–610.
- Prause, K.: 1999, *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*, PhD thesis, University of Freiburg.
- Roberts, G.: 1996, *Markov Chain Concepts Related to Sampling Algorithms*, w. gilks, s. richardson and d.j. spiegelhalter edn, Chapman and Hall, London, pp. 45–57.
- Roberts, G. O. and Strammar, O.: 2001, On inference for partially observed non-linear diffusion models using the metropolis-hastings algorithm, *Biometrika* **88**(3), 603–621.
- Rydberg, T.: 1999, Generalized hyperbolic diffusion processes with applications in finance, *Mathematical Finance* **9**(2), 183–201.
- Stroud, J., Polson, N. and Muller, P.: 2004, *Practical Filtering for Stochastic Volatility Models*, a. harvey, s.j. koopman and n. shephard edn, Cambridge University Press, Cambridge, pp. 236–247.
- Tierny, L.: 1994, Markov chains for exploring posterior distributions. Technical report N0 560 School of Statistics, University of Minnesota.
- Tse, Y., Zhang, X. and Yu, J.: 2004, Estimation of hyperbolic diffusion using mcmc, *Quantitative Finance* **4**, 158–169.
- Whelan, N.: 2004, Sampling from archimedean copulas, *Quantitative Finance* **4**, 339–352.