CHAPTER 8

Estimating Asymmetric Dynamic Distributions in High Dimensions*,#

Stanislav Anatolyev[a], Renat Khabibullin[b] and Artem Prokhorov[c]

[a]New Economic School and CERGE-EI
[b]Barclays Capital
[c]University of Sydney Business School, St. Petersburg State University and Innopolis University

Abstract
We consider estimation of dynamic joint distributions of large groups of assets. Conventional likelihood functions based on 'off-the-shelf' distributions quickly become inaccurate as the number of parameters grows. Alternatives based on a fixed number of parameters do not permit sufficient flexibility in modelling asymmetry and dependence. This chapter considers a sequential procedure, where the joint patterns of asymmetry and dependence are unrestricted, yet the method does not suffer from the curse of dimensionality encountered in non-parametric estimation. We construct a flexible multivariate distribution using tightly parameterized lower-dimensional distributions coupled by a bivariate copula. This effectively replaces a high-dimensional parameter space with many simple estimations with few parameters. We provide theoretical motivation for this estimator as a pseudo-MLE with known asymptotic properties. In an asymmetric GARCH-type application with regional stock indexes, the procedure provides excellent fit when dimensionality is moderate, and remains operational when the conventional method fails.

8.1 INTRODUCTION

The problem of estimating conditional, or dynamic, distributions for a group of assets is very important to a wide range of practitioners, in particular in the areas of risk management and portfolio optimization. The key problem is how to allow for arbitrary asymmetry and dependence in high dimensions while preserving a feasible parameterization. Traditional multivariate likelihood-based estimators are often impractical in these settings due to high dimensionality or small samples, or both. The existing multivariate densities that allow for asymmetric shapes tend to be tightly parameterized. For example, the multivariate skewed Student-*t* distribution of Bauwens and Laurent (2005) allows for

*Helpful discussions with Eric Renault, Tommaso Proietti, Dmytro Matsypura and Oleg Prokopyev are gratefully acknowledged.
#Research for this paper was supported by a grant from the Russian Science Foundation (Project No. 16-18-10432).

Asymmetric Dependence in Finance: Diversification, Correlation and Portfolio Management in Market Downturns, First Edition. Edited by Jamie Alcock and Stephen Satchell.

different degrees of asymmetry along each dimension, but the degrees of freedom are constrained to be the same along all dimensions. Thus, a more natural benchmark is provided by the copula approach, which allows for greater flexibility as far as the density shape is concerned.

Now consider the problem of constructing a high-dimensional distribution using the copula approach. Suppose we wish to estimate a d-dimensional Student-t distribution. This is equivalent to estimating d univariate Student-t marginals and a Student-t d-copula. The problem has at least $d(d-1)/2$ parameters. The conventional approach is to construct a joint log-density from this d-dimensional distribution and use it in a maximum likelihood (ML) routine. However, for large d and moderate sample sizes, the likelihood is highly unstable, Hessians are near singular, estimates are inaccurate and global convergence is hard to achieve.

One solution is to use copulas which have tighter parameterizations. However, the functional form of such copulas limits the nature of dependence they can accommodate (Nelsen, 2006, Section 4.6). Another solution is to use 'vine copulas' (Aas *et al.*, 2009) when the d-variate density is decomposed into a product of up to $d(d-1)/2$ bivariate densities. However, there are still $O(d^2)$ parameters in the joint likelihood; in addition, the required ordering of components is rarely available, especially in the time-series context. Yet another alternative is to use the factor copula approach (Oh and Patton, 2013). However, the joint density obtained lacks a close form; in addition, it is unclear whether the convolution of distributions imposed by the factor copula covers all classes of joint distributions one may wish to model.

The method we describe in this chapter replaces the initial estimation problem with a sequence of bivariate problems. This procedure was first outlined by Anatolyev *et al.* (2014) and can be thought of as recovering the joint distribution from the distributions of all lower-dimensional sub-vectors comprising the original random vector. We start with univariate distributions and estimate all copula-based bivariate distributions that can be constructed from them. Then we couple each univariate marginal with one of the bivariate distributions to get all possible trivariate distributions involving three given marginals. Then we model the average. At every subsequent step we couple each univariate marginal with a lower-dimensional distribution from the previous step and average over these combinations. This provides sufficient flexibility as we can model asymmetry and dependence differently in each step.

Theoretical justification for this procedure comes from the theory of composite and quasi-likelihoods (see, e.g., Varin *et al.*, 2011 for a review) and the theory of compatible copulas (see, e.g., Nelsen, 2006). The averaging over combinations comes from the theory of model average estimators (see, e.g., Clemen, 1989 for an early review). The procedure is related to the work by Sharakhmetov and Ibragimov (2002) and de la Pena *et al.* (2006), who provide a representation of multivariate distributions in terms of sums of U-statistics of independent random variables, where the U-statistics are based on functions defined over all subvectors of the original random vector. The procedure is also somewhat similar in spirit to Engle's (2009) approach of estimating a vast-dimensional DCC model by merging estimates of many pairwise models, either all of them or a number of selected ones. In contrast to Engle (2009), we reconstruct the dynamics of the entire multivariate distribution, rather than focusing on the dynamics of the conditional second moment.

Our method uses many individual optimization problems at each step. However, each such problem involves substantially fewer parameters than the conventional estimation problem where the entire dependence structure is parameterized. For the Student-t example above, we will show in the application section that the conventional MLE will have difficulty even when dimensions are moderate (i.e., when d is between 5 and 10). In contrast, our procedure requires running MLEs for only bivariate Student-t's.

In parametric estimation, fewer parameters means functional form biases. If we take an 'off-the-shelf' distribution with a high d and a tight parameterization, it will typically be tied to a convenient functional form indexed by a handful of parameters and the patterns of asymmetry and dependence it can accommodate will be limited. As an example, consider a one-parameter Gumbel–Hougaard d-copula. This is a multivariate Archimedean copula with Kendall's τ in the range $[0, 1)$, incapable of capturing discordance or lower-tail dependence (see, e.g., Nelsen, 2006, Section 4.6). An important advantage of our approach is greater flexibility in modelling asymmetry and dependence – because of the many steps, we have more degrees of freedom in choosing parameterizations.

In this chapter we look in more detail at the asymptotic properties of our estimator. We show that this estimator can be viewed as a traditional pseudo-maximum likelihood estimator (PMLE). We also look at the estimator in the framework of the generalized method of moments (GMM) estimation in order to study the consequences of multiple-stage estimation on asymptotic efficiency and standard error construction.

The chapter proceeds by describing the algorithm in Section 8.2. Section 8.3 considers theoretical properties of our estimator. Section 8.4 describes a typical parameterization that arises in a multivariate setting with dynamic and skewed distributions. It also gives details on the compounding functions and goodness-of-fit tests that seem appropriate in this setting. Section 8.5 presents an empirical application of the sequential method. Section 8.6 concludes.

8.2 SEQUENTIAL PROCEDURE

Suppose the group contains d assets. The new methodology can be described in the following sequence of steps.

Step 1. Estimate the univariate marginals by fitting suitable parametric distributions

$$\widehat{F}_1, \widehat{F}_2, \dots, \widehat{F}_d,$$

where $\widehat{F}_j = F(\widehat{\theta}_j)$ for each $j = 1, \dots, d$. This step involves d MLE problems and is standard for parametric modelling of dynamic multivariate distributions using copulas. The conventional next step would be to apply a d-copula to the marginals but, as discussed in the Introduction, this often results in an intractable likelihood.

Step 2. Using $\widehat{F}_j$'s, estimate bivariate distributions for all distinct pairs (i, j)

$$\widehat{F}_{12}, \widehat{F}_{13}, \dots, \widehat{F}_{1K}, \widehat{F}_{23}, \dots, \widehat{F}_{d-1,d}$$

using a suitable parametric copula family as follows:

$$\widehat{F}_{ij} = C^{(2)}\left(\widehat{F}_i, \widehat{F}_j; \widehat{\theta}_{ij}\right),$$

where $C^{(2)}(\cdot, \cdot; \cdot)$ is a bivariate symmetric copula.[1]

In bivariate settings, this step is also standard, and final. For $d > 2$, we repeat it for all asset pairs, effectively obtaining all possible contributions of the pairwise composite likelihood. There are $d(d-1)/2$ distinct pairs, as a symmetric copula for (i, j) is identical to the copula of (j, i). Hence, this step involves at least that many ML estimations.

An alternative that provides even more flexibility is to use an asymmetric copula, in which case $C^{(2)}\left(\widehat{F}_i, \widehat{F}_j; \widehat{\theta}_{ij}\right) \neq C^{(2)}\left(\widehat{F}_j, \widehat{F}_i; \widehat{\theta}_{ji}\right)$. Then, the number of estimations increases

[1]Here and further a *symmetric* copula $C(.,.)$ means that $C(u, v) = C(v, u)$. This is the terminology used, for example, in Nelsen (2006). With reference to copulas it is sometimes equivalent to *exchangeable* copulas and not to be confused with *radially symmetric* copulas, which means that $C(u, v) = u + v - 1 + C(1 - u, 1 - v)$. All of the copulas we used, including symmetric in the sense above, are not radially symmetric so they allow for asymmetry in the joint distribution.

to $d(d-1)$ and in order to obtain a single distribution involving, say, index i and index j, we use the average of the two copulas as follows:

$$\widehat{F}_{ij} = \frac{C^{(2)}\left(\widehat{F}_i, \widehat{F}_j; \widehat{\theta}_{ij}\right) + C^{(2)}\left(\widehat{F}_j, \widehat{F}_i; \widehat{\theta}_{ji}\right)}{2}.$$

Aside from simple averaging, data-driven weighting schemes (e.g., based on information criteria) exist (see, e.g., Burnham and Anderson, 2002, Chapter 4), but we leave this aside for the moment.

It is easy to see that steps 1 and 2 produce pseudo-ML estimators of $\widehat{\theta}_j$'s and $\widehat{\theta}_{ij}$'s. So, for each pair (i, j) we have a pseudo-ML estimator of their distribution. It could be tempting to stop here and to construct the joint distribution over all marginals by aggregating $\widehat{F}_{ij}$ over $i, j = 1, \ldots, d$ (e.g., by averaging). This would be similar to other types of composite likelihood-based estimators (see, e.g., Cox and Reid, 2004; Varin and Vidoni, 2005, 2006, 2008). However, it will impose a restrictive dependence structure on the joint distribution and result in a poorer fit. We return to this point in Section 8.3.1.

Step 3. Using $\widehat{F}_j$ and $\widehat{F}_{ij}$, estimate trivariate distributions for each combination of i and (j, k)

$$C^{(3)}\left(\widehat{F}_i, \widehat{F}_{jk}; \widehat{\theta}_{ijk}\right),$$

where $C^{(3)}\left(\widehat{F}_i, \widehat{F}_{jk}; \widehat{\theta}_{ijk}\right)$ is a suitable copula-type *compounding function* (not necessarily symmetric) that captures dependence between the ith asset and the (j, k)th pair of assets. There are $d(d-1)(d-2)/2$ possible combinations of $\widehat{F}_i$'s with disjoint pairs $\widehat{F}_{jk}$, so this is the number of estimations involved in this step.

Similar to step 2, we construct a single distribution for triplet (i, j, k) by averaging over the three available estimates as follows:

$$\widehat{F}_{ijk} = \frac{C^{(3)}\left(\widehat{F}_i, \widehat{F}_{jk}; \widehat{\theta}_{ijk}\right) + C^{(3)}\left(\widehat{F}_j, \widehat{F}_{ik}; \widehat{\theta}_{jik}\right) + C^{(3)}\left(\widehat{F}_k, \widehat{F}_{ij}; \widehat{\theta}_{kij}\right)}{3}.$$

This formula is an extension of that for $\widehat{F}_{ij}$. It uses triplets of observations to construct the composite likelihood contributions and it applies equal weights when averaging since we have no information-theoretic argument to prefer one estimated distribution over another.

Step *m*, *m* < *d*. Using the $\widehat{F}_j$'s and $\widehat{F}_{i_1,\ldots,j-1,j+1,\ldots,i_m}$, estimate an m-dimensional distribution for each m-tuple. There are $d!/(d-m)!(m-1)!$ possible combinations of $\widehat{F}_i$'s with disjoint $(m-1)$-variate marginals. Let $i_1 < i_2 < \ldots < i_m$. Obtain a model average estimate of the distribution for the $(i_1, i_2, \ldots, i_m)$th m-tuple:

$$\widehat{F}_{i_1 i_2 \ldots i_m} = \frac{1}{m} \sum_{l=1}^{m} C^{(m)}\left(\widehat{F}_l, \widehat{F}_{i_1,\ldots,l-1,l+1,\ldots,i_m}; \widehat{\theta}_{l,i_1,\ldots,l-1,l+1,\ldots,i_m}\right),$$

where $C^{(m)}$ is an mth-order compounding function which is set to be a suitable asymmetric bivariate copula.

Step *d*. Estimate the d-variate distribution:

$$\widehat{F}_{12\ldots d} = \frac{1}{d} \sum_{l=1}^{d} C^{(d)}\left(\widehat{F}_l, \widehat{F}_{1,\ldots,l-1,l+1,\ldots,d}; \widehat{\theta}_{l,1,\ldots,l-1,l+1,\ldots,d}\right),$$

where $C^{(d)}$ is a dth-order compounding function. There are d such functions to be estimated.

As the compounding functions are regular bivariate copulas, it follows that, by construction, $\widehat{F}_{1,2,\ldots,d}$ is non-decreasing on its support, bounded and ranges between 0 and 1. Hence, $\widehat{F}_{1,2,\ldots,d}$ can be viewed as an estimate of the joint cumulative distribution function obtained using sequential composite likelihoods. In essence this cdf is a result of sequential applications of bivariate copulas to univariate cdf's and bivariate copulas.

Nothing guarantees that such a sequential use of copulas preserves the copula properties, that is, nothing guarantees that the mth-order compounding functions are also m-copulas, $m = 3, \ldots, d$. In fact, there are several well-known impossibility results concerning construction of high-dimensional copulas by using lower-dimensional copulas as arguments of bivariate copulas (see, e.g., Quesada-Molina and Rodriguez-Lallena, 1994; Genest *et al.*, 1995b). Basically, the results suggest that copulas are rarely compatible, that is, if one uses a k-copula and an l-copula as arguments of a bivariate copula, the resulting $(k + l)$-variate object does not generally meet all the requirements for being a copula (see, e.g., Nelsen, 2006, Section 3.5).

Strictly speaking, the compounding functions constructed in steps 3 to d may fail to be m-copulas unless we use a compatible copula family. However, the resulting estimator $\widehat{F}_{12\ldots d}$ *is* a distribution and thus implies a d-copula. Therefore we do not require the compounding functions to qualify for being m-copulas as long as they can provide a valid pseudo-likelihood. In the theory section, we discuss the assumptions underlying this estimator. In practice, in order to ensure that we use a valid pseudo-likelihood we choose in steps 3 to d a flexible asymmetric bivariate copula family that passes goodness-of-fit diagnostics.

As an alternative we could use copula functions which *are* compatible. Consider, for example, the Archimedean copulas. These copulas have the form $C(u_1, \ldots, u_d) = \psi^{[-1]}(\psi(u_1) + \ldots + \psi(u_d))$, where $\psi(\cdot)$ is a function with certain properties and $\psi^{[-1]}(\cdot)$ is its inverse. Under a certain monotonicity condition on ψ sometimes referred to as a nesting condition (see, e.g., Theorem 4.6.2 of Nelsen, 2006) this functional form allows us to go from $C(u_1, u_2)$ to the d-copula by repeatedly replacing one of the two arguments with $u_m = C(u_{m+1}, u_{m+2})$, $m = 2, \ldots, d-1$. However, as discussed in the Introduction, the range of dependence such d-copulas can capture is limited and hence we do not use it in the empirical section.[2]

In each step of the procedure we operate only with two types of objects: a multivariate distribution of a smaller (by one) dimension and a univariate distribution. This allows for the number of parameters used in each compounding function to be really small, while the total number of parameters in the joint distribution remains rather large and ensures the flexibility needed to model general asymmetry and dependence. This clarifies the claims made in the Introduction about the advantages of this procedure over the standard single-copula or full-likelihood-based estimation. The conventional methods often produce intractable likelihoods due to dimensionality, or they may be overly restrictive due to a tight parameterization. Our procedure allows us to maintain a high degree of flexibility while trading the dimensionality of the parameter space for numerous simpler estimations.

Finally, if we are faced with an extremely large number of assets our method permits a reduction of the number of estimations by following the approach of Engle *et al.* (2008) and considering *random* pairs, triples, etc., instead of *all* possible pairs, triples, etc., as proposed here.

8.3 THEORETICAL MOTIVATION

8.3.1 Composite Pseudo-likelihood and Model Averaging

Fundamentally, our method of obtaining $\widehat{F}_{12\ldots d}$ falls within a subcategory of sequential pseudo-MLE known as composite likelihood methods (see, e.g., Cox and Reid, 2004; Varin and Vidoni, 2005). Composite likelihood estimators construct joint pseudo-likelihoods using components of the true data

[2] In the application, we have considered using the Clayton copula as an Archimedean alternative to Student's t copula. However, in spite of being a comprehensive copula, it did not pass our goodness-of-fit diagnostics and so we do not report these results here.

generating process such as all pairs (see, e.g., Caragea and Smith, 2007; Varin, 2008) or pairwise differences (see, e.g., Lele and Taper, 2002), and sometimes employing weights on the likelihood components to improve efficiency (see, e.g., Heagerty and Lele, 1998).

Unlike existing composite likelihood approaches, we estimate components of the composite likelihood sequentially, for all possible multivariate marginals of the joint distribution, and employ weighting to combine alternative composite densities. So our estimator is related to the literature on sequential copula-based pseudo-MLE (see, e.g., Joe, 2005; Prokhorov and Schmidt, 2009b) and to the literature on Bayesian model averaging and optimal forecast combination (see, e.g., Clemen, 1989; Geweke and Amisano, 2011).

Consider the sequential procedure for $d = 3$ and ignore for the moment the combinatorics and the weighting. Let $H(x_1, x_2, x_3)$ and $h(x_1, x_2, x_3)$ denote the joint distribution and density, respectively. We wish to estimate these objects. Let $F_j = F(x_j)$ and $f_m = f(x_j), j = 1, 2, 3$, denote the univariate marginal cdf's and pdf's. Note that the conventional 3-copula factorization would lead to the following expression for the log joint density:

$$\ln h(x_1, x_2, x_3) = \sum_{j=1}^{3} \ln f_j + \ln c(F_1, F_2, F_3), \tag{8.1}$$

where $c(u_1, u_2, u_3)$ is a 3-copula density.

Now let $C^{(3)}(u_1, u_2)$ denote the copula function used in step 3 of our procedure, where u_2 is set equal to the copula obtained in step 2, and let $c^{(3)}(u_1, u_2)$ denote the copula density corresponding to $C^{(3)}(u_1, u_2)$. The following result shows that the log joint density (without the weighting) has a useful factorization, analogous to Equation (8.1).

Proposition 8.3.1 *Suppose* $H(x_1, x_2, x_3) = C^{(3)}(F_3(x_3), C^{(2)}(F_2(x_2), F_1(x_1)))$. *Assume* $\ln c^{(2)}(u_1, u_2)$ *is Lipschitz continuous. Then,*

$$\begin{aligned}\ln h(x_1, x_2, x_d) = \sum_{j=1}^{3} \ln f_j + \ln c^{(2)}(F_2, F_1) + \ln c^{(3)}\left(F_3, C^{(2)}(F_2, F_1)\right)\\ + O\left(c^{(2)}(F_2, F_1)^{-1}\right)\end{aligned} \tag{8.2}$$

Proof: see Appendix 8.A for proofs. ∎

In essence, Proposition 8.3.1 shows that under a standard continuity condition, one can reconstruct a trivariate log density, up to an approximation error, by combining likelihood contributions obtained from individual marginals using bivariate copulas as in our algorithm. The approximation error is inversely related to $c^{(2)}(F_2, F_1)$, so it is small in areas of the support where $c^{(2)}$ concentrates a lot of mass and is big in flat areas of the copula density.

Now suppose we stop at step 2, as discussed in Section 8.2. Effectively, this means we omit the third term on the right-hand side of Equation (8.2). The approximation error is now larger and its magnitude is no longer inversely related to values of copula densities. We also omit a valid contribution to the likelihood in the form of the log-copula density from step 3. This may have efficiency implications even if we use model averaging.

Clearly there are many possible combinations of marginals that can be used to form a joint distribution $H(x_1, x_2, x_3)$. For example, $C^{(3)}$ can also be formed as $C^{(3)}\left(F_1, C^{(2)}(F_2, F_3)\right)$, or as $C^{(3)}\left(F_2, C^{(2)}(F_1, F_3)\right)$. Each such combination of marginals will produce a different log-density so it is important to pool them optimally. This question of density pooling is central in the literature on combining multiple prediction densities (see, e.g., Hall and Mitchell, 2007; Geweke and Amisano, 2011), where optimal weights, also known as scoring rules, are worked out in the context of information theory. As an example, define $c_j^{(3)}$ as follows:

$$c_j^{(3)} \equiv c^{(3)}\left(F_j, C_k^{(2)}\right),$$

where $j, k = 1, 2, 3, j \neq k$ and $C_k^{(2)} \equiv C^{(2)}(F_k, F_l), l \neq k, l \neq j$. Then, it is possible in principle to obtain the optimal weights ω_j as solutions to the following problem:

$$\max_{\omega_l : \sum \omega_j = 1} \sum_{\text{sample}} \ln \sum_j \omega_j c_j^{(3)}. \tag{8.3}$$

Such scoring rules make the ω_j's a function of the $c_j^{(3)}$'s and may be worth pursuing in large samples. However, it has been noted in this literature that often a simple averaging performs better due to the estimation error in ω's (see, e.g., Stock and Watson, 2004; Elliot, 2011). Moreover, in our setting, problem (8.3) would need to be solved at each step, imposing a heavy computational burden. Therefore, in our procedure we use a simple average of $C^{(m)}$'s, or equivalently, a simple average of $c^{(m)}$'s.

8.3.2 Asymptotics

We now turn to the asymptotic properties of our estimator. Let $\hat{\boldsymbol{\theta}}$ contain all $\hat{\theta}$'s from the steps described in Section 8.2. Assume that $\hat{F}_{12\ldots d}(x_1, \ldots, x_d)$ is a proper distribution. Then, by the celebrated Sklar (1959) theorem, the distribution $\hat{F}_{12\ldots d}(x_1, \ldots, x_d)$ implies a d-copula $K(u_1, \ldots, u_d; \hat{\boldsymbol{\theta}})$ and the corresponding estimator of density $\hat{f}_{12\ldots d}(x_1, \ldots, x_d)$ implies a d-copula density $k(u_1, \ldots, u_d; \hat{\boldsymbol{\theta}})$. (We denote the implied copula distribution and density functions by K and k, respectively, to distinguish them from the true copula distribution $C(u_1, \ldots, u_d)$ and true copula density $c(u_1, \ldots, u_d)$.) The following result gives explicit formulas for the implied copula (density).

Proposition 8.3.2 *Let $\hat{F}_m^{-1}(u_m), m = 1, \ldots, d$, denote the inverse of the marginal cdf $\hat{F}_m$ from step 1 and let $\hat{f}_m$ denote the pdf corresponding to $\hat{F}_m$. Then, the copula implied by $\hat{F}_{12\ldots d}$ can be written as follows:*

$$K(u_1, \ldots, u_d; \hat{\boldsymbol{\theta}}) = \hat{F}_{12\ldots d}(\hat{F}_1^{-1}(u_1), \ldots, \hat{F}_d^{-1}(u_d))$$

$$k(u_1, \ldots, u_d; \hat{\boldsymbol{\theta}}) = \frac{\hat{f}_{12\ldots d}(\hat{F}_1^{-1}(u_1), \ldots, \hat{F}_d^{-1}(u_d))}{\prod_{m=1}^{d} \hat{f}_m(\hat{F}_m^{-1}(u_m))}.$$

Proposition 8.3.2 gives the form of the flexible parametric d-variate *pseudo*-copula implied by $\hat{F}_{12\ldots d}(x_1, \ldots, x_d)$.[3] So if we estimated $\boldsymbol{\theta}$ using the conventional one-step MLE rather than the sequential MLE algorithm of Section 8.2, the asymptotic properties of our estimator would be the well-studied properties of copula-based *pseudo*- or *quasi*-MLE (see, e.g., Genest *et al.*, 1995a; Joe, 2005; Zhao and Joe, 2005; Prokhorov and Schmidt, 2009b). Sequential estimation only affects the asymptotic variance of $\boldsymbol{\theta}$. The following proposition summarizes these results.

[3]Here, by *pseudo*-copula we mean a possibly misspecified copula function. The same term is sometimes used in reference to an empirical copula obtained using univariate empirical cdf's and to a copula-type function that satisfies most but not all copula properties (see, e.g., Fermanian and Wegkamp, 2012; Fang and Madsen, 2013).

Proposition 8.3.3 *The MLE estimator $\widehat{\boldsymbol{\theta}}$ minimizes the Kullback–Leibler divergence criterion,*

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\theta} \mathbb{E} \ln \frac{c(u_1, \dots, u_d)}{k(u_1, \dots, u_d; \boldsymbol{\theta})},$$

where c is the true copula density and expectation is with respect to the true distribution. Furthermore, under standard regularity conditions, $\widehat{\boldsymbol{\theta}}$ is consistent and asymptotically normal. If the true copula belongs to the family $k(u_1, \dots, u_d; \boldsymbol{\theta})$, it is consistent for the true value of $\boldsymbol{\theta}$. If the copula family is misspecified, the convergence occurs to a pseudo-true value of $\boldsymbol{\theta}$, which minimizes the Kullback–Leibler distance.

It is worth noting that it is still possible in principle to follow the conventional MLE approach here. That is, we can still attempt to find $\widehat{\boldsymbol{\theta}}$ by maximizing the log-likelihood based on the following joint log-density:

$$\ln h(x_1, \dots, x_d) = \sum_{j=1}^{d} \ln f_j(\theta_j) + \ln k(F_1(\theta_1), \dots, F_d(\theta_d); \boldsymbol{\theta}), \tag{8.4}$$

where θ_j's denote parameters of the univariate marginals. However, the dimension of $\boldsymbol{\theta}$ in this problem is greater than for the initial problem in Equation (8.1) and so if the initial problem is intractable, this method will be as well.

Proposition 8.3.3 outlines the asymptotic properties of $\widehat{\boldsymbol{\theta}}$ and thus of $\widehat{F}_{12\dots d}$. However, it does not provide the asymptotic variance of $\widehat{\boldsymbol{\theta}}$. In order to address the issue of the relative efficiency of our procedure, we rewrite our problem in the GMM framework.

It is well known that the MLE can quite generally be written as a method of moments problem based on the relevant score functions. As an example, we look at the ingredients of our procedure for $d = 3$. The first step is the MLE for $F_j \equiv F_j(\theta_j), j = 1, 2, 3$; the second step is the MLE for $c^{(2)}(\widehat{F}_2, \widehat{F}_3; \theta_{23})$, where $\widehat{F}_j \equiv F_j(\widehat{\theta}_j)$; the third step is the MLE for $c^{(3)}(\widehat{F}_1, \widehat{C}^{(2)}; \theta_{123})$, where $\widehat{C}^{(2)} \equiv C^{(2)}(\widehat{F}_2, \widehat{F}_3; \widehat{\theta}_{23})$. The corresponding GMM problems can be written as follows:

1. $\mathbb{E}\begin{bmatrix} \nabla_{\theta_1} \ln f_1(\theta_1) \\ \nabla_{\theta_2} \ln f_2(\theta_2) \\ \nabla_{\theta_3} \ln f_3(\theta_3) \end{bmatrix} = 0,$

2. $\mathbb{E}[\nabla_{\theta_{23}} \ln c^{(2)}(\widehat{F}_2, \widehat{F}_3; \theta_{23})] = 0,$

3. $\mathbb{E}[\nabla_{\theta_{123}} \ln c^{(3)}(\widehat{F}_1, \widehat{C}^{(2)}(\widehat{F}_2, \widehat{F}_3); \theta_{123})] = 0,$

where ∇ denotes the gradient of the score function.

The GMM representation provides several important insights. First, it shows that at steps 2 and 3 we treat the quantities estimated in the previous step as if we knew them. The fact that we estimate them affects the asymptotic variance of $\widehat{\theta}_{23}$ and $\widehat{\theta}_{123}$, and the correct form of the variance should account for that. The appropriate correction and simulation evidence of its effect are provided, for example, by Joe (2005) and Zhao and Joe (2005).

Second, it shows that each estimation in the sequence is an exactly identified GMM problem. That is, each step introduces as many new parameters as new moment conditions. One important implication of this is that the (appropriately corrected for the preceding steps) asymptotic variance of the sequential estimator is identical to the asymptotic variance of the one-step estimator, which is obtained by solving the optimal GMM problem based on all moment conditions at once (see, e.g., Prokhorov and Schmidt, 2009a). Such an optimal GMM estimator may be difficult to obtain in practice due to the large number of moment conditions, but this efficiency bound is the best we can do in terms of relative efficiency with the moment conditions implied by our sequential MLE problems.

Finally, it is worth noting that this efficiency bound does not coincide with the Fisher bound, implied by the MLE based on the full likelihood in Equation (8.4), even if the copula k is correctly specified. The corresponding GMM problem for that likelihood includes moment conditions of the form

$$\mathbb{E}\left[\nabla_{\theta_j} \ln k(F_1(\theta_1), \ldots, F_d(\theta_d); \boldsymbol{\theta})\right] = 0, \quad j = 1, \ldots d,$$

which are not used in the sequential procedure. Therefore, the sequential procedure cannot be expected to be fully efficient.

8.4 PARAMETERIZATIONS

This section describes the models and distributions used in the empirical implementation of our sequential procedure. The particular choices of parameterizations are tied to our data set and should be perceived as suggestive. However, we believe that they are flexible enough to produce good fits when applied to log-returns data – these are models and distributions characterized by asymmetry, skewness, fat tails and dynamic dependence.

Assume that the following sample of log-returns is available: $\left\{\mathbf{y}_t = \left\{y_{it}\right\}_{i=1}^{d}\right\}_{t=1}^{T}$, where y_{it} is the individual log-return of the ith asset at time t, d is the total number of assets and T is the length of the sample.

8.4.1 Univariate Distributions

We would like to use skewed and thick-tailed distributions to model univariate marginals. Azzalini and Capitanio (2003) propose one possible generalization of Student's t-distribution which is able to capture both these features. Moreover, their transformation does not restrict the smoothness properties of the density function, which is useful for quasi-maximum likelihood optimization. The pdf of their skew t-distribution is

$$f_Y(y) = 2\, t_\nu(y)\, T_{\nu+1}\left(\gamma \frac{y-\xi}{\omega}\left(\frac{\nu+1}{\nu+Q_y}\right)^{1/2}\right),$$

where

$$Q_y = \left(\frac{y-\xi}{\omega}\right)^2,$$

$$t_\nu(y) = \frac{\Gamma((\nu+1)/2)}{\omega\,(\pi\nu)^{1/2}\Gamma(\nu/2)}(1+Q_y/\nu)^{-(\nu+1)/2}$$

and $T_{\nu+1}(x)$ denotes the cdf of the standard t-distribution with $\nu+1$ degrees of freedom. The parameter γ reflects the skewness of the distribution. Equivalently, denote

$$Y \sim \mathrm{St}_1(\xi, \omega, \gamma, \nu).$$

It is worth noting the first three moments of the distribution when $\xi = 0$,

$$E(Y) = \omega\overline{\mu},$$
$$E(Y^2) = \overline{\sigma^2} = \omega^2 \frac{\nu}{\nu - 2},$$
$$E(Y^3) = \overline{\lambda} = \omega^3 \overline{\mu}\, \frac{3 + 2\gamma^2}{1 + \gamma^2}\, \frac{\nu}{\nu - 3},$$

where

$$\overline{\mu} := \frac{\gamma}{\sqrt{1+\gamma^2}} \left(\frac{\nu}{\pi}\right)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}.$$

The last moment equation indicates that by varying γ one can change the skewness of the distribution. Also, it follows from the first two equations that the first moment of Y is different from 0 and its second central moment is not equal to 1. It is therefore useful to define the standardized skew t-distribution by adjusting $St_1(\xi, \omega, \gamma, \nu)$ for zero expectation and unit variance through setting ξ and ω in the following way:

$$\omega = \left(\frac{\nu}{\nu - 2} - \overline{\mu}^2\right)^{-1/2},$$
$$\xi = -\omega\overline{\mu}.$$

Denote the standardized skew t-distribution by $St(\gamma, \nu)$, its cdf by $F^{St}_{\gamma,\nu}$ and pdf by $f^{St}_{\gamma,\nu}$. We augment this distribution with the NAGARCH structure for the conditional variance equation (see, e.g., Engle and Ng, 1993):

$$y_{it} = \mu_i + \sqrt{h_{it}}\varepsilon_{it}, \qquad \varepsilon_{it} \sim \text{ i.i.d. } St(\gamma_i, \nu_i),$$
$$h_{it} = \omega_i + \alpha_i \left(y_{i,t-1} - \mu_i + \kappa_i \sqrt{h_{i,t-1}}\right)^2 + \beta_i h_{i,t-1},$$

where the h_{it}'s are the conditional variances of the y_{it}'s and $(\mu_i, \gamma_i, \nu_i, \omega_i, \alpha_i, \beta_i, \kappa_i)$ is the set of parameters. It is worth noting that the parameter κ_i reflects the leverage effect and is expected to be negative.

Using this structure we can write the cdf of y_{it} as follows:

$$F_i(y_{it}) = F^{St}_{\gamma_i,\nu_i}\left(\frac{y_{it} - \mu_i}{h_{it}^{1/2}}\right).$$

Then, the log-likelihood function for each univariate marginal will have the following form:

$$\ln L_i = \sum_{t=2}^{T} \left\{ \ln f^{St}_{\gamma_i,\nu_i}\left(\frac{y_{it} - \mu_i}{h_{it}^{1/2}}\right) - \frac{1}{2}\ln h_{it} \right\}.$$

There are seven parameters in this likelihood function for each i.

8.4.2 Bivariate Copulas

Here we chose the following p-copula adapted from Ausin and Lopes (2010):

$$C_{\eta,R}(u_1, \dots, u_p) = \int_{-\infty}^{T_\eta^{-1}(u_1)} \dots \int_{-\infty}^{T_\eta^{-1}(u_p)} \frac{\Gamma\left(\frac{\eta+p}{2}\right)\left(1 + \frac{\mathbf{v}'R^{-1}\mathbf{v}}{\eta-2}\right)^{-\frac{\eta+p}{2}}}{\Gamma\left(\frac{\eta}{2}\right)\sqrt{(\pi(\eta-2))^p|R|}} d\mathbf{v},$$

where $T_\eta^{-1}(\cdot)$ is the inverse of the standardized Student t cdf, p is number of assets in the group under consideration, η is the number of degrees of freedom and R is the correlation matrix. Denote the expression under the integral by $f_{\eta,R}(\mathbf{v})$ – it is the pdf of the standardized multivariate Student t-distribution. Except for comparison with the benchmark involving $p = d$, we will use only the bivariate version of this copula ($p = 2$).

Following Ausin and Lopes (2010), we assume that the dynamic nature of the correlation matrix R is captured by the following equation:

$$R_t = (1 - a - b)\overline{R} + a\Psi_{t-1} + bR_{t-1},$$

where $a \geq 0,\ b \geq 0,\ a + b \leq 1$, $\overline{R}$ is a positive definite constant matrix with ones on the main diagonal and Ψ_{t-1} is such a matrix whose elements have the following form:

$$\Psi_{ij,t-1} = \frac{\sum_{h=1}^{m} x_{it-h}x_{jt-h}}{\sqrt{\sum_{h=1}^{m} x_{it-h}^2 \sum_{h=1}^{m} x_{jt-h}^2}},$$

where

$$x_{it} = T_\eta^{-1}\left(F_{\gamma_i,\nu_i}^{\mathrm{St}}\left(\frac{y_{it} - \mu_i}{h_{it}^{1/2}}\right)\right).$$

The advantage of defining R_t in this way is that it guarantees positive definiteness. This circumvents the need to use additional transformations (see, e.g., Patton, 2006, who uses the logistic transformation).

Substituting the marginal distributions into the assumed copula function, we obtain the following model for the joint cdf of a vector of financial log-returns $\mathbf{y}_t = (y_{1t}, \ldots, y_{pt})$:

$$F(\mathbf{y}_t) = C_{\eta,R_t}(F_1(y_{1t}), \ldots, F_p(y_{pt})). \tag{8.5}$$

We also derive the joint pdf by differentiating Equation (8.5):

$$f(\mathbf{y}_t) = f_{\eta,R_t}(T_\eta^{-1}(F_1(y_{1t})), \ldots, T_\eta^{-1}(F_p(y_{pt}))) \times \prod_{i=1}^{p}\left\{\frac{1}{t_\eta\left(T_\eta^{-1}(F_i(y_{it}))\right)} f_{\gamma_i,\nu_i}^{\mathrm{St}}\left(\frac{y_{it} - \mu_i}{h_{it}^{1/2}}\right)\frac{1}{h_{it}^{1/2}}\right\}.$$

Then, the log-likelihood function for the conventional full ML estimation can be written as follows:

$$\ln L = \sum_{t=m+1}^{T} \ln f_{\eta,R_t}\left(T_\eta^{-1}(F_1(y_{1t})), \ldots, T_\eta^{-1}(F_p(y_{pt}))\right) + \sum_{t=m+1}^{T}\sum_{i=1}^{p}\left\{-\ln t_\eta\left(T_\eta^{-1}(F_i(y_{it}))\right) + \ln f_{\gamma_i,\nu_i}^{\mathrm{St}}\left(\frac{y_{it} - \mu_i}{h_{it}^{1/2}}\right) - \frac{1}{2}\ln h_{it}\right\}.$$

In the conventional one-step full MLE, $p = d$ and the number of parameters in this likelihood is $d(d - 1)/2 + 3$ from the copula part plus $7d$ from the marginal parts.

In our sequential alternative to the FMLE, we first estimate the skew t-marginal distributions using only the last two terms in the likelihood – they do not depend on the copula parameters. Then, the likelihoods we use in steps 2 to d are based on the bivariate version of this log-likelihood with given $\widehat{F}_i(y_{it})$'s. This version is simpler; it can be written as follows:

$$\ln L_{ij} = \sum_{t=m+1}^{T} \ln f_{\eta,R_t}\left(T_\eta^{-1}\left(\widehat{F}_i(y_{it})\right), T_\eta^{-1}\left(\widehat{F}_j(y_{jt})\right)\right)$$
$$- \sum_{t=m+1}^{T} \left\{\ln t_\eta\left(T_\eta^{-1}\left(\widehat{F}_i(y_{it})\right)\right) + \ln t_\eta\left(T_\eta^{-1}\left(\widehat{F}_j(y_{jt})\right)\right)\right\}.$$

It has only four parameters.

8.4.3 Compounding Functions

The two arguments of the bivariate copula in step 2 are similar objects – they are the marginal distributions of two assets. This is the reason why we use a symmetric copula for the bivariate modeling in step 2. In contrast, the compounding functions in steps 3 to d operate with two objects of different nature: one is a marginal distribution of one asset and the other is a joint distribution of a group of assets. Thus, in general it makes sense to use asymmetric copulas as compounding functions in these steps.

Khoudraji (1995) proposes a method of constructing asymmetric bivariate copulas from symmetric bivariate copulas using the following transformation:

$$C^{(asym)}(u, v) = u^\alpha v^\beta C^{(sym)}\left(u^{1-\alpha}, v^{1-\beta}\right), \quad 0 \le \alpha, \beta \le 1,$$

where $C^{(sym)}(u, v)$ is a generic symmetric copula and $C^{(asym)}(u, v)$ is the corresponding asymmetric copula. We utilize this result to obtain what we call the asymmetrized bivariate standardized t-copula:

$$C_{\eta,\rho}^{(t,asym)}(u, v) = u^\alpha\ v^\beta \int_{-\infty}^{T_\eta^{-1}(u^{1-\alpha})} \int_{-\infty}^{T_\eta^{-1}(v^{1-\beta})} \frac{\Gamma\left(\frac{\eta+2}{2}\right)\left(1 + \frac{x^2+y^2-2\ \rho\ x\ y}{(\eta-2)(1-\rho^2)}\right)^{-\frac{\eta+2}{2}}}{\Gamma\left(\frac{\eta}{2}\right)\pi(\eta-2)\sqrt{1-\rho^2}} dx\ dy,$$

where u denotes the marginal distribution of an asset, v denotes the distribution of a group of assets and we assume a similar time-varying structure on the correlation coefficient as in Section 8.4.2. The form of the compounding function in the mth step will then be

$$C^{(m)}\left(\widehat{F}_l, \widehat{F}_{i_1,\ldots,l-1,l+1,\ldots,i_m}; \theta_{l,i_1,\ldots,l-1,l+1,\ldots,i_m}\right) = C_{\eta,\rho}^{(t,asym)}\left(\widehat{F}_l, \widehat{F}_{i_1,\ldots,l-1,l+1,\ldots,i_m}\right),$$

where $\theta_{l,i_1,\ldots,l-1,l+1,\ldots,i_m} = (\alpha, \beta, \eta, \rho, a, b)'$ is the parameter set (the last three parameters come from the time-varying structure of the correlation matrix R containing in the bivariate case only one correlation coefficient ρ). Correspondingly, there are only six parameters to estimate in each optimization problem of the sequential procedure, regardless of the dimensionality of the original problem.

8.4.4 Goodness-of-Fit Testing

In order to assess the adequacy of distributional specifications, we conduct goodness-of-fit (GoF) testing. For this purpose we use the conventional approach based on probability integral transforms (PIT) first proposed by Rosenblatt (1952). The approach is based on transforming the time series of log-returns into a series that should have a known pivotal distribution in the case of correct specification and then testing the hypothesis that the transformed series indeed has that known distribution.

To assess the quality of fit of marginals, we use the approach of Diebold *et al.* (1998), who exploit the following observation. Suppose there is a sequence $\{y_t\}_{t=1}^T$ which is generated from distributions $\{F_t(y_t|\Omega_t)\}_{t=1}^T$, where $\Omega_t = \{y_{t-1}, y_{t-2}, \dots\}$. Then, under the usual condition of a non-zero Jacobian with continuous partial derivatives, the sequence of probability integral transforms $\{F_t(y_t|\Omega_t)\}_{t=1}^T$ is i.i.d. $U(0,1)$. Diebold *et al.* (1998) propose testing the uniformity property and the independence property separately by investigating the histogram and correlograms of the moments up to order 4. We follow this approach, with the exception that the statistical tests rather than the graphical analyses are conducted in order to separately test the uniformity and independence properties. In particular, we run Kolmogorov–Smirnov tests of uniformity and F-tests of serial uncorrelatedness.

The goodness-of-fit tests for bivariate copulas are based on a similar approach proposed by Breymann *et al.* (2003), which also relies on PIT. Let $\mathbf{X} = (X_1, \dots, X_d)$ denote a random vector with marginal distributions $F_i(x_i)$ and conditional distributions $F_{i|i-1\dots1}(x_i|x_{i-1}, \dots, x_1)$ for $i = 1, \dots, d$. The PIT of vector $\mathbf{x}$ is defined as $T(\mathbf{x}) = T(x_1, \dots, x_d) = (T_1, \dots, T_d)$ such that $T_1 = F_1(x_1)$, $T_p = F_{p|p-1\dots1}(x_p|x_{p-1}, \dots, x_1)$, $p = 2, \dots, d$. One can show that $T(\mathbf{X})$ is uniformly distributed on the p-dimensional hypercube (Rosenblatt, 1952). This implies that $T_1, \dots, T_p$ are uniformly and independently distributed on $[0, 1]$. This approach has been extended to the time-series setting (see, e.g., Patton, 2006). Again, we exploit the Kolmogorov–Smirnov tests for uniformity and F-tests for serial uncorrelatedness. Note, however, that there exist $p!$ ways of choosing conditional distributions of a p-variate vector. For pairwise copulas, this means two such ways: $X_2 \mid X_1$ and $X_1 \mid X_2$. We examine both of them for all pairs.

8.5 EMPIRICAL APPLICATION

This section demonstrates how to apply the new sequential technique to model a joint distribution of DJIA constituents. We have considered larger numbers of stocks but to illustrate the advantage of our method over conventional ones (and to save space) we start with $d = 5$. For univariate marginals, we exploit skewed t-distributions with a NAGARCH structure for conditional variance; for bivariate distributions, we exploit the asymmetrized time-varying t-copula, which is also the copula we use for benchmark comparisons when estimating p-variate distributions with $p > 2$. We have considered other copulas but found this copula to produce the best fit. The Kolmogorov–Smirnov goodness-of-fit tests conducted at each step of the procedure show that these parametric distributions provide a good fit for individual asset returns as well as jointly for their combinations. Eventually, we compare our new methodology with the conventional benchmark – a single five-dimensional time-varying t-copula-based estimation.

8.5.1 Data

We choose the following five stocks from among DJIA constituents (as of 8 June 2009): GE – General Electric Co.; MCD – McDonald's Corp.; MSFT – Microsoft Corp.; KO – Coca-Cola Co.; PG – Procter & Gamble Co. The selection is based on a high level of liquidity and availability of historical prices. Daily data from 3 January 2007 to 31 December 2007 are collected; we focus on this period to avoid dealing with the turbulence that followed. The stock prices are adjusted for splits and dividends, and then the log-returns are constructed and used in estimation. The plots of relative price dynamics, histograms of log-returns and sample statistics of log-returns for the five stocks are presented in Figures 8.1 and 8.2 and in Table 8.1. One can see that the unconditional sample distributions in some cases demonstrate skewness and heavy tails, which justifies the selection of the skew t-distribution for modelling marginals.

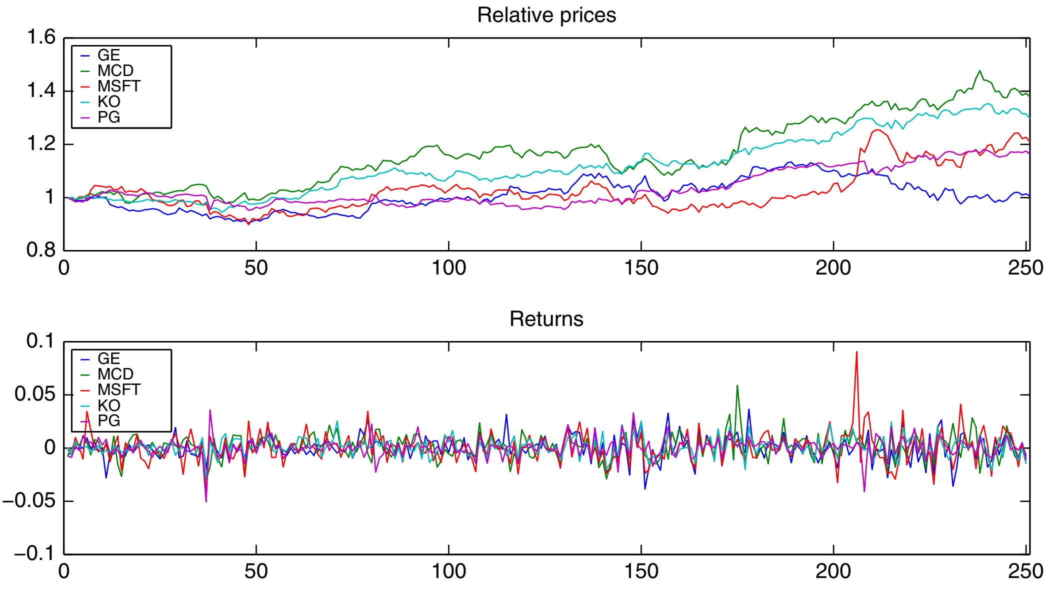

FIGURE 8.1 Relative prices and returns dynamics for GE, MCD, MSFT, KO and PG from 3 January to 31 December 2007.

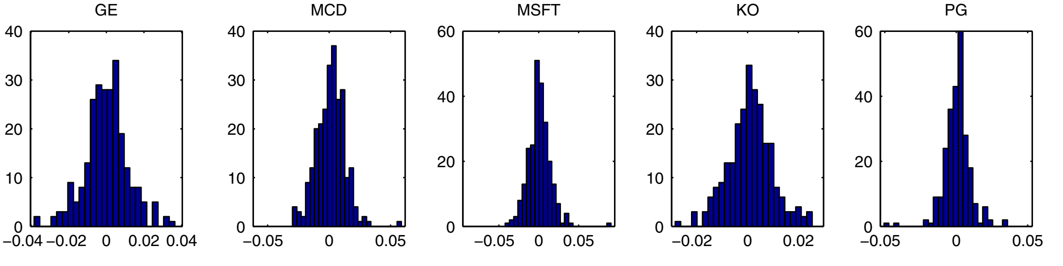

FIGURE 8.2 Histograms of the returns.

TABLE 8.1 Summary statistics of the returns

	GE	MCD	MSFT	KO	PG
Minimum	−0.0384	−0.0298	−0.0421	−0.0285	−0.0506
Maximum	0.0364	0.0589	0.0907	0.0254	0.0359
Mean $\times 10^{-3}$	0.0248	1.2831	0.7575	1.0363	0.5987
Standard deviation	0.0115	0.0116	0.0143	0.0087	0.0091
Skewness (zero-skewness:	−0.0349	0.2617	0.9461	0.0512	−0.6106
p-value)	(0.8216)	(0.0912)	(0.0000)	(0.7408)	(0.0001)
Kurtosis (zero-ex. kurtosis:	3.9742	4.8977	8.7270	3.6313	9.2954
p-value)	(0.0017)	(0.0000)	(0.0000)	(0.0416)	(0.0000)

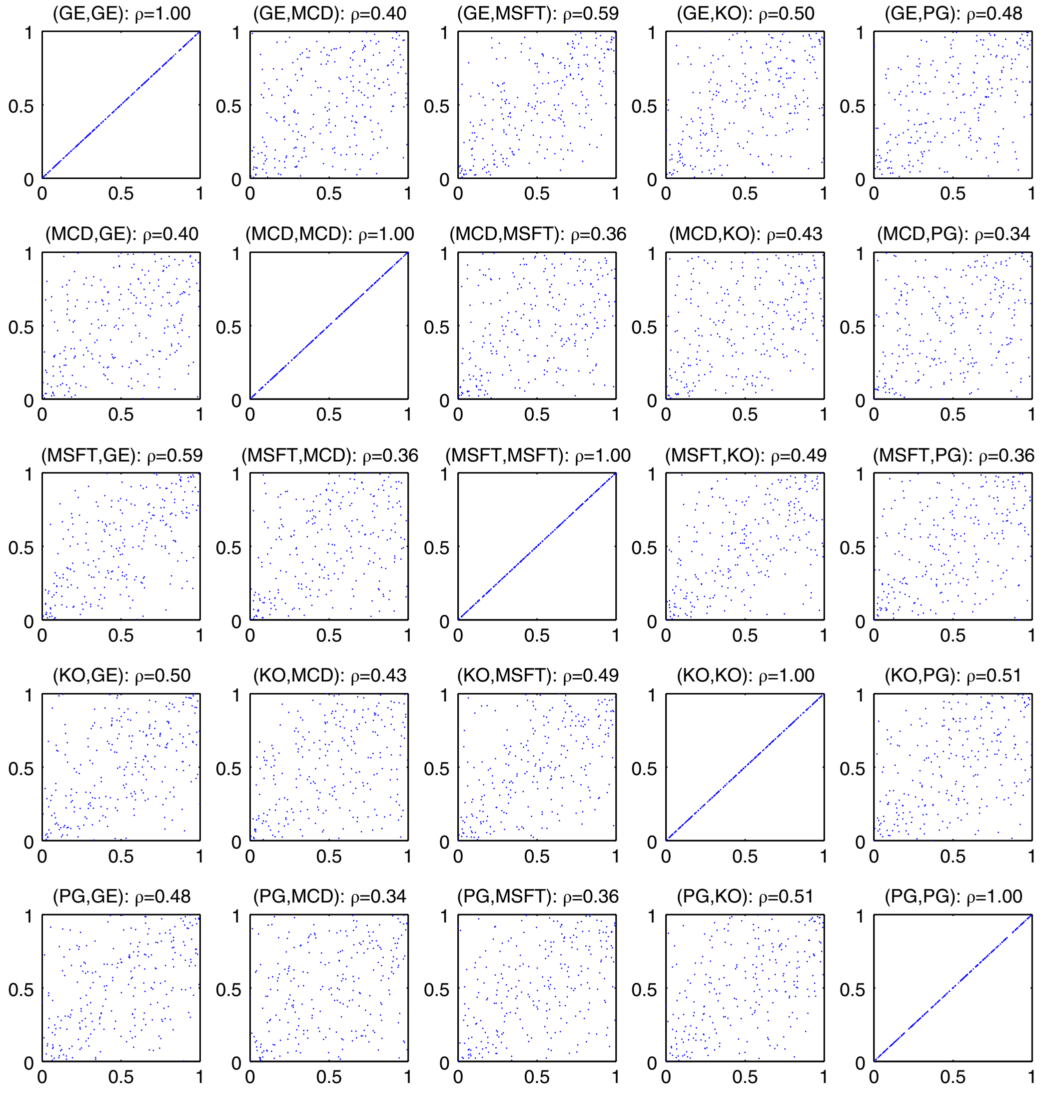

FIGURE 8.3 Pairwise scatter plots of marginal distributions and sample correlations.

The scatter plots and correlations on Figure 8.3 show that, as expected, all of the stocks are positively correlated. The correlation between MSFT and PG is smaller than for most of the other stocks – these two stocks belong to different sectors (Technology and Consumer Goods).[4] At the same time, the correlation between KO and PG is greater than between the other stocks, due to the fact that they both belong to the Consumer Goods sector.

[4]See http://finance.yahoo.com for a classification.

8.5.2 Estimates of Univariate Distributions

We use the skew-t-NAGARCH model for the marginals in order to accommodate the asymmetries, heavy tails, leverage effects and volatility clustering that are observed in stock log-returns. For marginal distributions, we choose the initial value of the conditional variance h_{i1} in the GARCH process to be equal to the sample unconditional variance of log-returns.

The estimates of the parameters of marginal distributions are summarized in Table 8.2. As before, μ denotes mean return, ω is unconditional variance, α reflects the ability to predict conditional variance using current innovations, β is a measure of persistence of conditional variance, κ is leverage effect, the reciprocal of ν captures heavy tails and γ represents skewness. The mean return μ is fairly close to the sample mean. There is a substantial degree of persistence in the conditional variance process for four out of the five series. There is excess kurtosis in all series. The skewness parameter γ and the leverage effects are largely insignificant; however, we keep them in the model because of the non-zero skewness found (see Table 8.1) and because it is now standard in the literature to account for these stylized facts.

The Kolmogorov–Smirnov tests of uniformity applied to the transformed series show that at the 95% confidence level the hypothesis of uniformity is not rejected. The quantitative results along with the diagrams are presented in Figure 8.4. The model passes these tests.

Next, we conduct the tests for serial correlation of the transformed series. Diebold *et al.* (1998) recommend that it is sufficient in practice to investigate the moments up to order 4. We follow this suggestion and test the hypothesis about the joint insignificance of coefficients in the regression of each

TABLE 8.2 Maximum likelihood parameter estimates for marginal distributions (robust standard errors are in parentheses)

	GE	MCD	MSFT	KO	PG
$\mu \times 10^{-3}$	−0.032	1.340	0.660	0.750	0.574
	(0.615)	(0.709)	(0.886)	(0.673)	(0.645)
$\omega \times 10^{-5}$	0.569	5.845	0.852	0.667	0.608
	(0.581)	(1.893)	(0.809)	(0.740)	(0.523)
α	0.106	0.153	0.041	0.142	0.107
	(0.062)	(0.090)	(0.024)	(0.082)	(0.052)
β	0.861	0.379	0.915	0.787	0.837
	(0.084)	(0.130)	(0.055)	(0.139)	(0.063)
κ	−0.074	−0.530	−0.174	−0.032	−0.168
	(0.364)	(0.394)	(0.796)	(0.740)	(0.606)
ν	6.482	9.672	5.898	8.098	3.305
	(2.325)	(4.976)	(2.360)	(4.046)	(0.734)
γ	−0.014	−0.431	0.176	−0.236	−0.106
	(0.077)	(0.706)	(0.533)	(0.950)	(0.482)

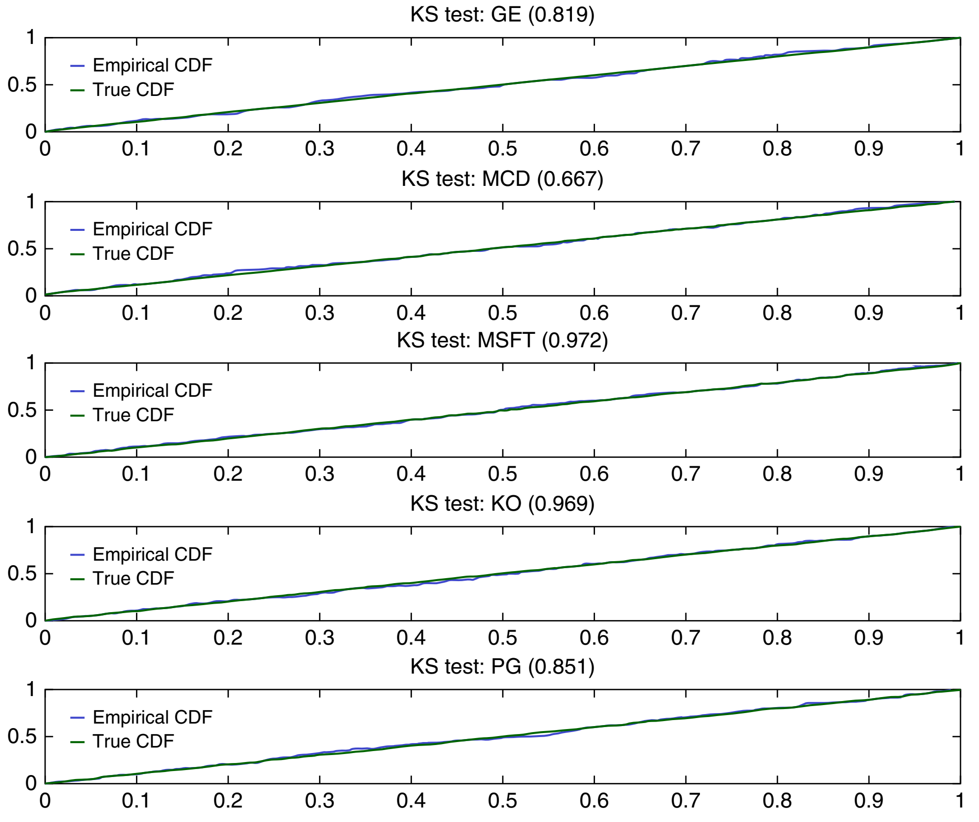

FIGURE 8.4 Kolmogorov–Smirnov tests of marginal distributions (p-values in parentheses).

moment on its 20 lags using the F-test. The results are presented in Table 8.3. The hypotheses of no serial correlation are not rejected at the 95% confidence level in nearly every case; the exception is the fourth central moment of the KO stock, for which the hypothesis is not rejected at the 99% confidence

TABLE 8.3 p-values of F-tests for serial correlation

Central moment	GE	MCD	MSFT	KO	PG
1	0.701	0.454	0.762	0.336	0.310
2	0.763	0.805	0.448	0.070	0.437
3	0.567	0.672	0.763	0.611	0.657
4	0.887	0.774	0.635	0.032	0.172

level. In addition, all the Ljung–Box tests carried out to test for autocorrelation in the residuals of the marginals' specification do not reject the hypothesis of no serial correlation either. This also indicates a good fit of the selected parametric forms of the marginal distributions.

8.5.3 Estimates of Pairwise Copulas

The pairwise copula parameter estimates are summarized in Table 8.4 and the results of pairwise Kolmogorov–Smirnov tests are presented in Figures 8.5 and 8.6. All Kolmogorov–Smirnov tests are passed at any reasonable confidence level. This indicates that the time-varying *t*-copula used for modelling bivariate distributions fits quite well and can be used in step 2 of the sequential approach.

As before, the hypothesis of no serial correlation can be tested by checking the joint insignificance of the coefficients in the regression of each of the first four moments on their 20 lags using the F-test.

TABLE 8.4 Maximum likelihood parameter estimates for pairwise copulas (robust standard errors are in parentheses)

	GE, MCD	GE, MSFT	GE, KO	GE, PG	MCD, MSFT
η	9.627	7.948	6.107	14.236	9.883
	(7.732)	(3.006)	(2.390)	(11.290)	(8.488)
a	0.074	0.089	0.002	0.038	0.159
	(0.075)	(0.113)	(0.002)	(0.023)	(0.096)
b	0.399	0.001	0.486	0.913	0.385
	(0.241)	(0.130)	(0.306)	(0.031)	(0.226)
$\overline{\rho}$	0.418	0.625	0.513	0.557	0.429
	(0.062)	(0.042)	(0.050)	(0.076)	(0.073)
	MCD, KO	**MCD, PG**	**MSFT, KO**	**MSFT, PG**	**KO, PG**
η	11.825	6.011	5.672	6.926	10.760
	(9.397)	(2.476)	(2.226)	(2.968)	(9.417)
a	0.072	0.170	0.031	0.197	0.053
	(0.087)	(0.091)	(0.225)	(0.152)	(0.076)
b	0.447	0.394	0.462	0.000	0.342
	(0.288)	(0.266)	(2.057)	(0.169)	(0.209)
$\overline{\rho}$	0.417	0.368	0.556	0.469	0.504
	(0.059)	(0.080)	(0.056)	(0.065)	(0.050)

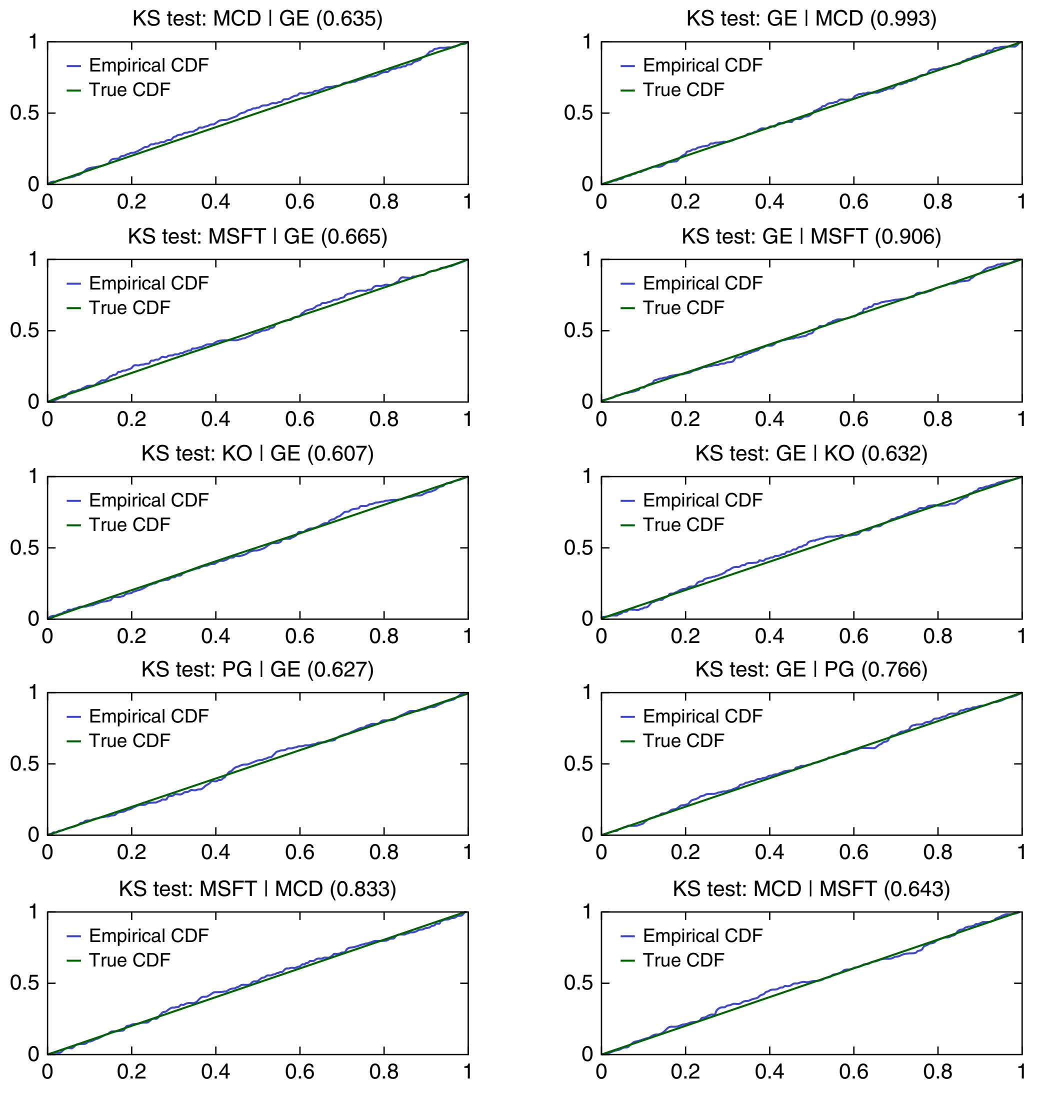

FIGURE 8.5 Pairwise Kolmogorov–Smirnov tests of bivariate copula specification: first five pairs of GE, MCD, MSFT, KO and PG (*p*-values in parentheses).

Additionally, in the bivariate setting we included the lagged moments of the other PIT series in the regression to test for independence. All the results (not shown here to save space) suggest that the hypotheses of no serial correlation cannot be rejected at any reasonable confidence level in every case and that the bivariate specification we chose fits well.

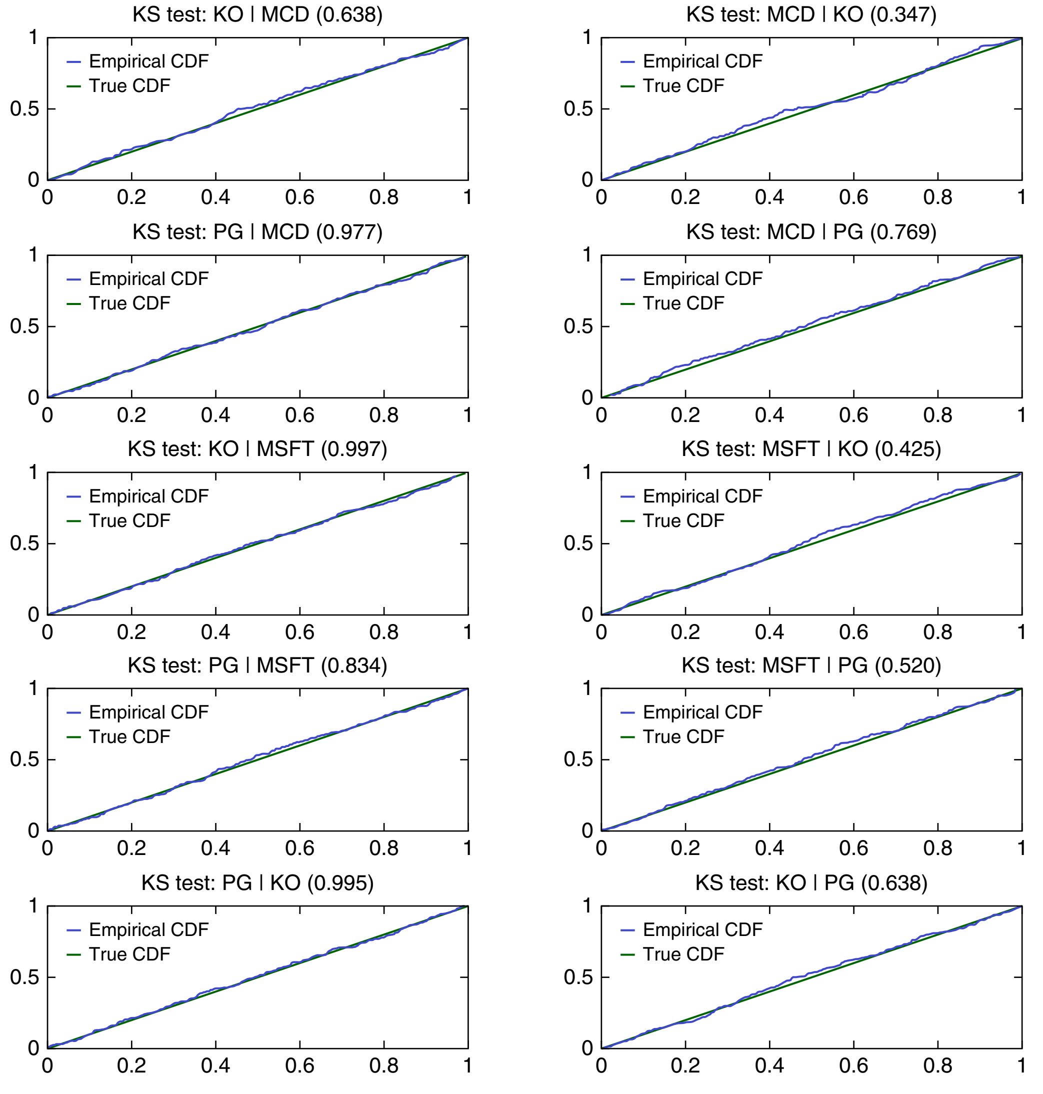

FIGURE 8.6 Pairwise Kolmogorov–Smirnov tests of bivariate copula specification: second five pairs of GE, MCD, MSFT, KO and PG (p-values in parentheses).

8.5.4 Estimates of Compounding Functions

Tables 8.5, 8.6 and 8.7 contain parameter estimates in the t-copula-based approach for groups of assets of different size. We do not present standard errors for the estimates to save space. Also, we omit the plots and p-values for the Kolmogorov–Smirnov tests. The tests support our choice of the asymmetrized bivariate t-copula as a compounding function and show an exceptional fit.

TABLE 8.5 Maximum likelihood parameter estimates of the compounding functions for groups of three assets (standard errors omitted)

$C(\cdot;\cdot)$	α	β	η	$\bar{\rho}$	a	b
(GE; MCD, MSFT)	0.044	0.322	8.612	0.884	0.004	0.992
(MCD; GE, MSFT)	0.007	0.011	8.557	0.408	0.124	0.351
(MSFT; GE, MCD)	0.008	0.001	8.828	0.503	0.143	0.092
(GE; MCD, KO)	0.002	0.251	8.544	0.499	0.001	0.249
(MCD; GE, KO)	0.000	0.112	9.928	0.403	0.059	0.399
(KO; GE, MCD)	0.017	0.008	61.964	0.446	0.048	0.334
(GE; MCD, PG)	0.023	0.114	14.379	0.468	0.109	0.059
(MCD; GE, PG)	0.009	0.028	5.197	0.338	0.114	0.423
(PG; GE, MCD)	0.001	0.194	5.306	0.472	0.106	0.275
(GE; MSFT, KO)	0.076	0.224	8.508	0.670	0.076	0.128
(MSFT; GE, KO)	0.007	0.007	6.867	0.536	0.036	0.232
(KO; GE, MSFT)	0.031	0.162	8.459	0.602	0.004	0.314
(GE; MSFT, PG)	0.002	0.161	10.151	0.638	0.015	0.946
(MSFT; GE, PG)	0.030	0.001	10.296	0.498	0.101	0.034
(PG; GE, MSFT)	0.057	0.251	8.786	0.612	0.060	0.242
(GE; KO, PG)	0.046	0.069	13.185	0.496	0.001	0.033
(KO; GE, PG)	0.009	0.001	23.245	0.497	0.066	0.484
(PG; GE, KO)	0.001	0.234	7.505	0.612	0.136	0.016
(MCD; MSFT, KO)	0.005	0.219	6.018	0.469	0.119	0.362
(MSFT; MCD, KO)	0.003	0.001	9.917	0.436	0.054	0.402
(KO; MCD, MSFT)	0.129	0.025	19.673	0.541	0.026	0.286
(MCD; MSFT, PG)	0.133	0.315	8.481	0.588	0.358	0.136
(MSFT; MCD, PG)	0.001	0.000	6.562	0.407	0.078	0.606
(PG; MCD, MSFT)	0.001	0.128	5.165	0.425	0.120	0.561
(MCD; KO, PG)	0.000	0.004	6.686	0.320	0.154	0.324
(KO; MCD, PG)	0.231	0.020	19.575	0.577	0.066	0.605
(PG; MCD, KO)	0.000	0.405	4.553	0.587	0.164	0.365
(MSFT; KO, PG)	0.002	0.001	6.930	0.473	0.125	0.015
(KO; MSFT, PG)	0.002	0.001	23.721	0.530	0.067	0.251
(PG; MSFT, KO)	0.022	0.143	8.385	0.544	0.159	0.221

TABLE 8.6 Maximum likelihood parameter estimates of the compounding functions for groups of four assets (standard errors omitted)

$C(\cdot;\cdot)$	α	β	η	$\bar{\rho}$	a	b
(GE; MCD, MSFT, KO)	0.057	0.326	8.580	0.630	0.007	0.646
(MCD; GE, MSFT, KO)	0.005	0.180	8.326	0.440	0.137	0.244
(MSFT; GE, MCD, KO)	0.022	0.110	8.443	0.508	0.058	0.238
(KO; GE, MCD, MSFT)	0.015	0.062	8.540	0.466	0.023	0.477
(GE; MCD, MSFT, PG)	0.026	0.137	8.868	0.523	0.096	0.103
(MCD; GE, MSFT, PG)	0.046	0.201	8.464	0.447	0.189	0.390
(MSFT; GE, MCD, PG)	0.007	0.008	8.666	0.423	0.185	0.047
(PG; GE, MCD, MSFT)	0.025	0.235	6.492	0.511	0.092	0.494
(GE; MCD, KO, PG)	0.026	0.269	10.768	0.447	0.092	0.188
(MCD; GE, KO, PG)	0.023	0.228	8.492	0.400	0.163	0.320
(KO; GE, MCD, PG)	0.140	0.073	8.825	0.440	0.101	0.304
(PG; GE, MCD, KO)	0.048	0.400	8.507	0.675	0.079	0.730
(GE; MSFT, KO, PG)	0.013	0.109	8.563	0.553	0.004	0.706
(MSFT; GE, KO, PG)	0.004	0.004	9.226	0.475	0.105	0.130
(KO; GE, MSFT, PG)	0.022	0.073	9.093	0.506	0.064	0.316
(PG; GE, MSFT, KO)	0.022	0.225	8.615	0.591	0.120	0.186
(MCD; MSFT, KO, PG)	0.022	0.356	8.839	0.481	0.229	0.112
(MSFT; MCD, KO, PG)	0.023	0.056	8.487	0.418	0.100	0.345
(KO; MCD, MSFT, PG)	0.041	0.022	8.652	0.450	0.066	0.757
(PG; MCD, MSFT, KO)	0.035	0.327	8.487	0.564	0.102	0.552

TABLE 8.7 Maximum likelihood parameter estimates of the compounding functions for groups of five assets (standard errors omitted)

$C(\cdot;\cdot)$	α	β	η	$\bar{\rho}$	a	b
(GE; MCD, MSFT, KO, PG)	0.175	0.265	10.818	0.546	0.157	0.227
(MCD; GE, MSFT, KO, PG)	0.189	0.410	9.088	0.578	0.247	0.199
(MSFT; GE, MCD, KO, PG)	0.394	0.301	8.580	0.463	0.309	0.285
(KO; GE, MCD, MSFT, PG)	0.285	0.585	8.680	0.433	0.312	0.321
(PG; GE, MCD, MSFT, KO)	0.045	0.252	8.499	0.520	0.051	0.782

8.5.5 Comparison with the Conventional Copula Approach

Now we compare the proposed approaches with the conventional single copula approach to dynamic modelling of joint distributions. The conventional alternative would be to estimate a time-varying five-dimensional *t*-copula using ML.

The parameter estimates for the conventional benchmark method are summarized in Table 8.8. As before, we have run serial correlation tests – they are not reported due to their large number – and almost all of them passed at the 95% confidence level. This means that, for our five time series, there is no obvious leader in the goodness-of-fit competition. Moreover, the number of estimations in our procedure is much larger than in the conventional method. In this example, a five-dimensional distribution requires solving 80 low-dimensional problems in the sequential procedure.[5] The conventional approach would require solving just one, but it would involve estimating 43 parameters in total.[6] Hence, for moderate dimensions such as $d = 5$, the conventional method may be preferred in terms of computer time, provided it is operational (it *was* quicker in this example). However, this changes as we increase the dimensionality of the problem, which is what we will do next.

When we repeat the above exercise for $d = 6, \ldots, 15$, the number of parameters in the conventional MLE based on the d-copula grows according to $O(d^2)$, while the number of additional parameters in each step of the new sequential procedure remains fixed at 6. The number of estimations in the sequential procedure also grows with d (potentially faster than d^2); however, this number can be made small in steps 3 and above, if we consider a random subset of all available combinations in each step. Table 8.9 contains the number of parameters to be estimated in a single optimization problem when we use the new and the conventional method.

TABLE 8.8 Maximum likelihood parameter estimates of time-varying five-dimensional *t*-copula for the returns (robust standard errors are in parentheses)

η	13.426 (4.380)
a	0.030 (0.035)
b	0.157 (0.308)

$\{\overline{R}_{ij}\}$	1	2	3	4	5
GE	1.000	0.425	0.621	0.502	0.510
	(0.000)	(0.055)	(0.042)	(0.055)	(0.049)
MCD	0.425	1.000	0.415	0.398	0.367
	(0.055)	(0.000)	(0.056)	(0.055)	(0.062)
MSFT	0.621	0.415	1.000	0.539	0.465
	(0.042)	(0.056)	(0.000)	(0.053)	(0.057)
KO	0.502	0.398	0.539	1.000	0.495
	(0.055)	(0.055)	(0.053)	(0.000)	(0.049)
PG	0.510	0.367	0.465	0.495	1.000
	(0.049)	(0.062)	(0.057)	(0.049)	(0.000)

[5]There are 5 estimation problems at step 1, 20 distributions of all possible pairs in step 2, 30 combinations of $\widehat{F}_i$ with $\widehat{F}_{jk}$ in step 3, 20 combinations of $\widehat{F}_i$ with $\widehat{F}_{jkl}$ in step 4, and 5 combinations of $\widehat{F}_i$ with $\widehat{F}_{jklm}$.

[6]Each skew-*t* marginal has 6 parameters and the *t*-copula has 13 distinct parameters.

TABLE 8.9 Growth of the number of parameters in a single optimization problem for the conventional and for the sequential methods based on the t-copula

Dimension	Conventional	Sequential
3	6	6
4	9	6
5	13	6
6	18	6
7	24	6
8	31	6
9	39	6
10	48	6
11	58	6
12	69	6
13	81	6
14	94	6
15	108	6

In our application, we discovered that the conventional approach fails to produce reliable convergence when d reaches and exceeds 10. At the same time, the new approach remains functional. Although there are a lot of optimization problems to solve, each such problem is relatively simple and takes very little time. In this application, each of the sequential estimations took only a few seconds, while the high-dimensional standard estimation with d close to 10 takes minutes and fails if the dimension is greater than 10.[7]

8.6 CONCLUDING REMARKS

We have proposed a sequential MLE procedure which reconstructs a joint distribution by sequentially applying a copula-like compounding function to estimates of marginal distributions. We have discussed the theoretical justification of the use of compounding functions and averaging and outlined the asymptotic properties of our estimator. We have shown in an application that this is a reasonable alternative to the conventional single-copula approach, especially when the dimension is higher than moderate.

The issues with conventional ML are not only computational (Hessian non-invertibility, local maxima, etc.). It is often a problem to find a multivariate distribution that accommodates certain features

[7] A Matlab module handling arbitrary dimension and data sets under both conventional and sequential methodology is available at https://sites.google.com/site/artembprokhorov/papers/reconstruct.zip.

(e.g., asymmetry and extreme dependence in higher dimensions) while remaining tractable. Moreover, finite-sample-based ML estimation of highly parameterized multivariate distribution is inaccurate due to the curse of dimensionality.

The proposed method falls short of solving all the issues. For example, the full version of the algorithm requires more computing time than conventional ML (when it works) and the standard errors of the sequential procedure suffer from the 'generated regressor' problem. However, the new method allows us to estimate distributions with arbitrary patterns of dependence and to parameterize dependence between a scalar and a subvector.

The standard way to study the performance of our algorithm relative to the vine copula and factor copula approaches mentioned in the Introduction is by means of simulations. However, it is unclear what criterion to use for such comparisons. The difficulty is not only in coming up with a feasible version of a MISE-type distance for a d-variate function. The operational version of this measure would need to be applicable to sequential estimators. We leave the development and implementation of such criteria for future research.

Alternative methods of constructing a joint distribution from objects of lower dimensions may come from work by de la Pena *et al.* (2006) and Li *et al.* (1999). De la Pena *et al.* (2006) provide a characterization of arbitrary joint distributions using sums of U-statistics of independent random variables. Their terms in the U-statistic are functions $g(\cdot)$ defined over subvectors of the original multidimensional vector. Li *et al.* (1999) discuss the notion of the linkage function $L(\cdot)$, which is a multidimensional analogue of the copula function. Linkage functions link uniformly distributed random vectors rather than uniformly distributed scalar random variables.

Functions $g(\cdot)$ and $L(\cdot)$ are the lower-dimensional objects that may be used in a similar estimation procedure to ours. However, except for some special cases, the closed-form expressions of these objects are unknown and their properties are not so well studied as the properties of copula functions. For this reason, we leave the study of such alternative methods of modelling joint distributions for future research.

REFERENCES

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44, 182–198.

Anatolyev, S., Khabibullin, R. and Prokhorov, A. (2014). An algorithm for constructing high dimensional distributions from distributions of lower dimension. *Economics Letters*, 123, 257–261.

Ausin, M.C. and Lopes, H.F. (2010). Time-varying joint distribution through copulas. *Computational Statistics and Data Analysis*, 54, 2383–2399.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B*, 65, 367–389.

Bauwens, L. and Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics*, 23, 346–354.

Breymann, W., Dias, A. and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3, 1–14.

Burnham, K.P. and Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.

Caragea, P.C. and Smith, R.L. (2007). Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, 98, 1417–1440.

Clemen, R. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting*, 5, 559–583.

Cox, D. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91, 729–737.

de la Pena, V.H., Ibragimov, R. and Sharakhmetov, S. (2006). Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. *IMS Lecture Notes – Monograph Series*, 49, 183–209.

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39, 863–883.

Elliot, G. (2011). Averaging and the optimal combination of forecasts. UCSD Working Paper.

Engle, R. (2009). High dimension dynamic correlations, in J.L. Castle and N. Shephard (eds), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*. Oxford University Press, Oxford.

Engle, R.F. and Ng, V.K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48, 1749–1778.

Engle, R.F., Shephard, N. and Sheppard, K.K. (2008). Fitting vast dimensional time-varying covariance models. NYU Working Paper FIN-08-009.

Fang, Y. and Madsen, L. (2013). Modified Gaussian pseudo-copula: applications in insurance and finance. *Insurance: Mathematics and Economics*, 53, 292–301.

Fermanian, J.-D. and Wegkamp, M.H. (2012). Time-dependent copulas. *Journal of Multivariate Analysis*, 110, 19–29.

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

Genest, C., Quesada-Molina, J. and Rodriguez-Lallena, J. (1995b). De l'impossibilite de construire des lois a marges multidimensionnelles donnees a partir de copules. Comptes rendus de l'Academie des sciences de Paris Series I, Vol. 320. Elsevier, Paris.

Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164, 130–141.

Hall, S.G. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23, 1–13.

Heagerty, P.J. and Lele, S.R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, 1099–1111.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC, Boca Raton, FL.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94, 401–419.

Khoudraji, A. (1995). Contributions à l'étude des copules et à la modélisation des valeurs extrêmes bivariées. Ph.D. dissertation, Université Laval.

Lele, S. and Taper, M.L. (2002). A composite likelihood approach to (co)variance components estimation. *Journal of Statistical Planning and Inference*, 103, 117–135.

Li, H., Scarsini, M. and Shaked, M. (1999). Dynamic linkages for multivariate distributions with given nonoverlapping multivariate marginal. *Journal of Multivariate Analysis*, 68, 54–77.

Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd edn. Springer Series in Statistics, Vol. 139. Springer, Berlin.

Oh, D.H. and Patton, A.J. (2013). Simulated method of moments estimation for copula-based multivariate models. *Journal of American Statistical Association*, 108, 689–700.

Patton, A.J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47, 527–556.

Prokhorov, A. and Schmidt, P. (2009a). GMM redundancy results for general missing data problems. *Journal of Econometrics*, 151, 47–55.

Prokhorov, A. and Schmidt, P. (2009b). Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas. *Journal of Econometrics*, 153, 93–104.

Quesada-Molina, J. and Rodriguez-Lallena, J. (1994). Some advances in the study of the compatibility of three bivariate copulas. *Statistical Methods and Applications*, 3, 397–417.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23, 470–472.

Sharakhmetov, S. and Ibragimov, R. (2002). A characterization of joint distribution of two-valued random variables and its applications. *Journal of Multivariate Analysis*, 83, 389–408.

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

Stock, J.H. and Watson, M.W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405–430.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, 92, 1–28.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92, 519–528.

Varin, C. and Vidoni, P. (2006). Pairwise likelihood inference for ordinal categorical time series. *Computational Statistics and Data Analysis*, 51, 2365–2373.

Varin, C. and Vidoni, P. (2008). Pairwise likelihood inference for general state space models. *Econometric Reviews*, 28, 170–185.

Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21, 5–42.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33, 335–356.

Appendix 8.A

Proof of Proposition 8.1: We provide the proof for $d = 3$. Arguments for $d > 3$ are analogous.
Lipschitz continuity of $\ln c^{(m)}$ implies

$$\frac{d\, c^{(m)}(u_1, u_2)}{d\, u_j} \leq Bc^{(m)}(u_1, u_2), \quad m = 1, \dots, d, \quad j = 1, 2.$$

Since $H(x_1, x_2, x_3) = C^{(3)}\left(F_1, C^{(2)}(F_2, F_3)\right)$, we have

$$\begin{aligned} h(x_1, x_2, x_3) &\equiv \frac{\partial^3 H(x_1, x_2, x_3)}{\partial x_1 \partial x_2 \partial x_3} \\ &= h_c(x_1.x_2, x_3) + \epsilon(x_1, x_2, x_3), \end{aligned}$$

where

$$\begin{aligned} h_c(x_1.x_2, x_3) &\equiv f_1\, f_2\, f_3 \frac{\partial^2 C^{(3)}\left(F_1, C^{(2)}(F_2, F_3)\right)}{\partial F_1 \partial C^{(2)}} \frac{\partial^2 C^{(2)}(F_2, F_3)}{\partial F_2 \partial F_3} \\ &= f_1\, f_2\, f_3\, c^{(2)}(F_2, F_3)\, c^{(3)}(F_1, C^{(2)}(F_1, F_2)) \\ \epsilon(x_1, x_2, x_3) &\equiv f_1\, f_2\, f_3 \frac{\partial^3 C^{(3)}\left(F_1, C^{(2)}(F_2, F_3)\right)}{\partial F_1 [\partial C^{(2)}]^2} \frac{\partial C^{(2)}(F_2, F_3)}{\partial F_2} \frac{\partial C^{(2)}(F_2, F_3)}{\partial F_3}. \end{aligned}$$

Note that $0 \leq \partial C^{(2)}(F_2, F_3)/\partial F_i \leq 1$, $i = 2, 3$. Therefore,

$$\begin{aligned} \epsilon(x_1, x_2, x_3) \leq f_1\, f_2\, f_3 \frac{\partial^3 C^{(3)}\left(F_1, C^{(2)}(F_2, F_3)\right)}{\partial F_1 [\partial C^{(2)}]^2} &= f_1\, f_2\, f_3\, \frac{\partial c^{(3)}(F_3, C^{(2)})}{\partial C^{(2)}} \\ &\leq B\, f_1\, f_2\, f_3\, c^{(3)}(F_3, C^{(2)}) \\ &= \frac{B}{c^{(2)}(F_2, F_1)}\, h_c(x_1, x_2, x_3), \end{aligned}$$

where the second line follows from $\ln c^{(3)}(u_1, u_2)$ being Lipschitz with constant B.

It follows that

$$\ln h(x_1, x_2, x_3) - \ln h_c(x_1, x_2, x_3) \leq B \frac{1}{c^{(2)}(F_2, F_1)}. \qquad \blacksquare$$

Proof of Proposition 8.2: The result follows trivially from application of Sklar's (1959) theorem to $\widehat{F}_{12\ldots d}$.

Proof of Proposition 8.3: These are standard results cited, for example, in Chapter 10 of Joe (1997) or by Joe (2005).