Centre for Economic and Financial Research at New Economic School

Reconstructing High Dimensional Dynamic Distributions from Distributions of Lower Dimension

Stanislav Anatolyev
Renat Khabibullin
Artem Prokhorov

Reconstructing high dimensional dynamic distributions from distributions of lower dimension[*]

Stanislav Anatolyev[†] **Renat Khabibullin**[‡] **Artem Prokhorov**[§]

New Economic School Barclays Capital University of Sydney, CIREQ

August 2013

Abstract

We propose a new sequential procedure for estimating a dynamic joint distribution of a group of assets. The procedure is motivated by the theory of composite likelihood and by the theory of copula functions. It recovers m-variate distributions by coupling univariate distributions with distributions of dimension $m - 1$. This copula-based method produces pseudo-maximum-likelihood type estimators of the distribution of all pairs, triplets, quadruples, etc, of assets in the group. Eventually the joint distribution of unrestricted dimension can be recovered. We show that the resulting density can be viewed as a flexible factorization of the underlying true distribution, subject to an approximation error. Therefore, it inherits the well known asymptotic properties of the conventional copula-based pseudo-MLE but offers important advantages. Specifically, the proposed procedure trades the dimensionality of the parameter space for numerous simpler estimations, making it feasible when conventional methods fail in finite samples. Even though there are more optimization problems to solve, each is of a much lower dimension. In addition, the parameterization tends to be much more flexible. Using a GARCH-type application from stock returns, we demonstrate how the new procedure provides excellent fit when the dimension is moderate and how it remains operational when the conventional method fails due to high dimensionality.

JEL Classification: C13

Keywords: pseudo-likelihood, composite likelihood, multivariate distribution, copulas.

[†]Corresponding author. Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, office 1721(3), Moscow, 117418 Russia. E-mail: sanatoly@nes.ru

[‡]Address: Four Winds Plaza, Bolshaya Gruzinskaya Street 71, Moscow 123056, Russia. E-mail: sir.renat@gmail.com

[§]Address: University of Sydney Business School, Sydney NSW 2006, Australia. E-mail: artem.prokhorov@sydney.edu.au

1 Introduction

The problem of estimating dynamic joint distributions[1] 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. It is also an interesting theoretical question whether there exist reliable and flexible estimation methods that are, in some sense, equivalent to the traditional multivariate likelihood based estimators when the traditional estimators are impractical due to high dimensionality or small samples, or both. In this paper we propose a new method of constructing multivariate dynamic distributions from lower-dimensional distributions in a sequential manner using copulas. We show that this estimator can be viewed as a traditional pseudo maximum likelihood estimator (PMLE), though it has important advantages over PMLE – it is more flexible and works reasonably well in situations when the traditional MLE fails.

In a nutshell, the sequential algorithm proceeds as follows. As a first step, we estimate univariate (marginal) distributions for each asset. In the second step, we construct bivariate distributions for each pair of assets using the univariate distributions from the first step and bivariate copula functions. In the third step, trivariate distributions are formed for all triplets of assets using the univariate distributions estimated in the first step and the bivariate distributions obtained in the second step. The procedure goes on until the joint distribution of the whole group of assets is constructed. In each step m, $m = 1, \ldots, d$, we use copula functions to combine each of the d univariate distributions with each of the available $(m-1)$-variate distributions constructed in the previous step – we call this *compounding* – and then we average the resulting m such combinations to reflect our uncertainty about the true data generating process.

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 of assets comes from the theory of model average estimator (see, e.g., Clemen, 1989, for an early review).

The procedure can be thought of as recovering the joint distribution (or density) from the distributions (or densities) of all lower-dimensional sub-vectors comprising the original random vector. In this sense, it is related to the work by Sharakhmetov and Ibragimov (2002) and de la

---

[1]All distributions mentioned in this paper are, unless otherwise noted, conditional on the past, and that is what we mean by a "dynamic" distribution.

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 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. As a simple example, consider the case of a d-dimensional vector with Student's t distribution (or, equivalently, with Student's t marginals and Student's t d-copula). The scale matrix, representing dependence for this case, has $\frac{d(d-1)}{2}$ parameters. In principle, we can estimate the parameters of the marginal distributions first and then use them to estimate the parameters of the copula – this is known as the inference function for margin (IFM) method (see, e.g., Joe, 2005). Alternatively, we can use the joint log-density to obtain an MLE of the entire parameter vectors (it will include the $\frac{d(d-1)}{2}$ dependence parameters plus whatever parameters are in the marginals). In either case, the order of the problem is at least $O(d^2)$. As we shall illustrate, this can be difficult to handle even when the dimensions are moderate (i.e., when d is between 5 and 10). In contrast, our procedure requires solving only bivariate problems which are very easy to handle.

An important advantage of our approach is greater parametric flexibility. Because of the many steps, we have more degrees of freedom in choosing parameterizations, while parameterizations used in the conventional procedures are typically quite tight and are tied to convenient functional forms indexed by a handful of parameters. As an example, consider an one-parameter Gumbel-Hougaard d-copula. This is a multivariate Archimedean copula whose functional form limits the nature of dependence it can accommodate (see, e.g., Nelsen, 2006, Section 4.6). This makes the copula insufficiently flexible for modelling data with high degrees of dependence often encountered in time series analysis.

Conventional estimation methods that accommodate larger classes of dependence structures typically suffer from an increased number of parameters. As an example, consider replacing Stu-

dent's t d-copula in the example above with the Eyraud-Farlie-Gumbel-Morgenstern d-copula. This copula has $2^d - d - 1$ parameters so if we use the standard estimators, the order of the problem grows even faster than for the t copula. For $d = 2$, the EFGM copula is known to accommodate only very limited range of dependence as measured by Kendall's τ. In higher dimensions it can be used in constructing a very large class of joint distributions (see, e.g., de la Pena et al., 2006) but the cost of this is the exponential growth of the number of parameters.

One recent solution to the dimensionality problem is to use a sequence of bivariate copulas to construct a multivariate copula, which is known as a 'vine copula' or a pair-copula construction [PCC] (see, e.g., Aas et al., 2009; Min and Czado, 2010). Similar to our procedure, this approach results in a flexible parametrization. It decomposes a joint d-variate density into a product of up to $\frac{d(d-1)}{2}$ bivariate copulas which can accommodate different types and ranges of dependence. However, unless some restrictions are imposed on the bivariate copulas or a sequential estimation procedure is used, this does not really reduce the order of the MLE problem. There are still $O(d^2)$ parameters in the joint likelihood to estimate. Moreover, an ordering of the components is required for the original multivariate vector before a PCC can be constructed. Such an ordering is rarely available, especially in the time series setting, which is the setting we consider.

Another alternative is to use the so called factor copula approach (see, e.g., Oh and Patton, 2013). However, the joint density obtained using this method lacks a close form and it is unclear at the moment whether the convolution of distributions imposed by the factor copula specification covers all possible classes of joint distributions one may wish to model.

The paper proceeds by describing our sequential approach in more detail in Section 2. Section 3 considers the theoretical properties of our estimator. Section 4 describes a typical parameterization that arises in a multivariate setting with dynamic, skewed and asymmetric distributions. It also gives detail on the compounding functions and goodness-of-fit tests that seem appropriate in this setting. Section 5 discusses an empirical application of the sequential method. Section 6 concludes.

2   The 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

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

where $\widehat{F}_j$ is the marginal cdf estimate for the jth asset. This step is standard in parametric modeling of dynamic multivariate distributions using copulas. It involves d estimation problems. 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. Exploiting the estimates $\widehat{F}_j$ from step 1, estimate bivariate distributions for all pairs of assets

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

using a suitable parametric copula family as follows:

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

where $\widehat{F}_{ij}$ is a bivariate cdf estimated for the (i,j)th pair of assets, and $\widetilde{C}^{(2)}(.,.;.)$ is a bivariate symmetric copula[2]. This step is also standard in parametric modeling of dynamic bivariate distributions using copulas. Here we repeat it for all asset pairs, effectively obtaining all possible contributions of the pairwise composite likelihood. There are $d(d-1)$ possible pairs to consider in this step. Similar to other pseudo-likelihood based estimators, the use of this type of composite likelihood has been justified either through unbiasedness of the corresponding score functions or through the Kullback-Leibler divergence (see, e.g., Cox and Reid, 2004; Varin and Vidoni, 2005, 2006, 2008). We discuss this in more detail in Section 3.

We also note that due to symmetry of $\widetilde{C}^{(2)}$ the bivariate distribution $\widehat{F}_{ij}$ can be estimated as a simple average

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

though other, data driven, weighting schemes based on the information criteria have recently been developed in connection with the Bayesian model averaging (see, e.g., Burnham and Anderson,

---

[2]Here and further a *symmetric* function $C(.,.)$ means that $C(u,v) = C(v,u)$. This is the terminology used, e.g., in Nelsen (2006). With reference to copulas it is sometimes equivalently called *exchangeable* and it is not to be confused with *radially symmetric* copulas, which means that $C(u,v) = u + v - 1 + C(1-u, 1-v)$. The copulas we use allow for asymmetry.

2002, Chapter 4).

Step 3. Using the estimates of $\widehat{F}_j$ from step 1, the estimates of $\widehat{F}_{ij}$ from step 2, estimate the trivariate distribution for each triplet by

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

where $\widetilde{C}^{(3)}\left(\widehat{F}_i, \widehat{F}_{jk}; \widehat{\theta}_{ijk}\right)$ is a suitable (not necessarily symmetric) copula-type *compounding function* that captures the 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 $\widehat{F}_{jk}$ to consider in this step.

Similar to step 2, we can average over the three available combinations of $(\widehat{F}_i, \widehat{F}_{jk})$ to obtain a simple model average estimate of the trivariate distribution for triplet (i,j,k) as follows

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

Note that the formula for $\widehat{F}_{ijk}$ is a natural extension of that for $\widehat{F}_{ij}$ from step 2. It uses triplets of observations to construct the composite likelihood contributions and it applies equal weights when averaging since we used no information theoretic argument to prefer one estimated distribution of triplet (i,j,k) to another.

Step m. Using compounding and averaging similarly to step 3, estimate the m-dimensional distributions for all groups of m assets, $m < d$. There are $\frac{d!}{(d-m)!(m-1)!}$ possible combinations of $\widehat{F}_i$'s with disjoint $(m-1)$-variate marginals. Let $i_1 < i_2 < \ldots < i_m$, then

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

where $\widehat{F}_{i_1 i_2 \ldots i_m}$ is the estimated distribution function for the $(i_1, i_2, \ldots, i_m)$-th m-ple of assets, and $\widetilde{C}^{(m)}$ is a m-th order compounding function which is set to be a suitable, possibly asymmetric bivariate copula. As before, the intention is that averaging diminishes estimation and model uncertainties.

Step d. Finally, the joint distribution of all d assets is estimated as

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

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

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

Of course, nothing guarantees that such a sequential use of copulas preserves the copula properties, that is, nothing guarantees that the m-th 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 argument 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 a 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 of our procedure may fail to be m-copulas unless we use a compatible copula family. However, the resulting estimator of $\widehat{F}_{12...d}$ *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 which passes goodness-of-fit diagnostics.

As an alternative we could use copula functions which are usually 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. This functional form allows 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 so we do not use them here.[3]

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 in order to ensure sufficient flexibility of the joint model. 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 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.

3   Theoretical Motivation and Asymptotic Properties

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 generating process such as all pairs (see, e.g., Varin, 2008; Caragea and Smith, 2007) 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 first the sequential procedure itself, ignoring for the moment the combinatorics and

---

[3]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 it here.

the weighting. Let $H(x_1, \ldots, x_d)$ and $h(x_1, \ldots, x_d)$ denote the joint distribution and density, respectively. We wish to estimate these objects. Let $F_j = F(x_j)$ and $f_m = f(x_j)$, $j = 1, \ldots, d$, denote the univariate marginal cdf's and pdf's. Note that the conventional d-copula factorization would lead to the following expression for the log joint density:

$$\ln h(x_1, \ldots, x_d) = \sum_{j=1}^{d} \ln f_j + \ln c(F_1, \ldots, F_d), \tag{2}$$

where $c(u_1, \ldots, u_d)$ is a d-copula.

Now let $C^{(m)}(u_1, u_2)$ denote the copula function used in step m of our procedure, where u_2 is set equal to the copula obtained in step $m-1$, and let $c^{(m)}(u_1, u_2)$ denote the copula density corresponding to $C^{(m)}(u_1, u_2)$. The following result shows that in our sequential procedure (without the weighting), the log joint density has a useful factorization, analogous to (2).

Proposition 1 *Suppose $H(x_1, \ldots, x_d) = C^{(d)}(F_d(x_d), C^{(d-1)}(F_{d-1}(x_{d-1}), \ldots))$. Assume $\ln c^{(m)}(u_1, u_2)$ is Lipschitz continuous. Then, up to an approximation error, the following holds*

$$
\begin{aligned}
\ln h(x_1, \ldots, x_d) &\approx \sum_{j=1}^{d} \ln f_j + \ln c^{(2)}(F_2, F_1) + \ln c^{(3)}\left(F_3, C^{(2)}(F_2, F_1)\right) + \ldots \\
&\quad \ldots + \ln c^{(d)}\left(F_d, C^{(d-1)}\left(F_{d-1}, C^{(d-2)}(F_{d-2}, \ldots)\right)\right)
\end{aligned}
\tag{3}
$$

Proof: see Appendix for proofs.

In essence, Proposition 1 shows that under a standard continuity condition on the copulas, our procedure (without weighting and combinatorics) recovers the joint log density, up to an approximation error, by replacing one high-dimensional copula term with a large number of bivariate copula terms. As a result, we obtain an approximate joint likelihood which is easy to parameterize using bivariate copula families.

Clearly there are many possible combinations of marginals that can be used to form copula terms in (3). For example, $c^{(3)}$ can be formed as $c^{(3)}\left(F_1, C^{(2)}(F_2, F_3)\right)$, or as $c^{(3)}\left(F_2, C^{(2)}(F_1, F_3)\right)$, etc. Each such combination of marginals will produce a different log-density so it is important to pool them optimally. This question 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's as solutions to the following problem

$$\max_{\omega_l : \sum \omega_j = 1} \sum_{\text{sample}} \ln \sum_j \omega_j c_j^{(3)} \tag{4}$$

Such scoring rules make ω_j's a function of $c_j^{(3)}$'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., Watson and Stock, 2004; Elliot, 2011). Moreover, in our setting, the problem in (4) 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.

We now turn to the asymptotic properties of our estimator. Let $\widehat{\boldsymbol{\theta}}$ contain all $\hat{\theta}$'s from the steps described in Section 2. Then, by the celebrated Sklar (1959) theorem, the distribution $\widehat{F}_{12\ldots d}(x_1, \ldots, x_d)$ implies a d-copula $K(u_1, \ldots, u_d; \widehat{\boldsymbol{\theta}})$ and the corresponding estimator of density $\widehat{f}_{12\ldots d}(x_1, \ldots, x_d)$ implies a d-copula density $k(u_1, \ldots, u_d; \widehat{\boldsymbol{\theta}})$. (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 copula (density) implied by our estimator.

Proposition 2 *Let $\widehat{F}_m^{-1}(u_m), m = 1, \ldots, d$, denote the inverse of the marginal cdf $\widehat{F}_m$ from Step 1 and let $\widehat{f}_m$ denote the pdf corresponding to $\widehat{F}_m$. Then, the copula implied by $\widehat{F}_{12\ldots d}$ can be written as follows*

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

It is clear from Proposition 2 that our procedure constructs and estimates a flexible parametric

d-variate *pseudo*-copula[4]. Specifically, a *pseudo*-copula is estimated by MLE in each step of our procedure. So the asymptotic properties of our estimator are basically 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). The following proposition summarizes these results.

Proposition 3 *The estimator $\widehat{\boldsymbol{\theta}}$ minimizes the Kullback-Leibler divergence criterion,*

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

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 $k(u_1, \ldots, u_d; \boldsymbol{\theta})$, it is consistent for the true value of $\boldsymbol{\theta}$. If the copula family is misspecified, the convergence is 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

$$\ln h(x_1, \ldots, x_d) = \sum_{j=1}^{d} \ln f_j(\theta_j) + \ln k(F_1(\theta_1), \ldots, F_d(\theta_d); \boldsymbol{\theta}), \tag{5}$$

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 (2) and so if the initial problem was intractable, this method will be as well.

Proposition 3 outlines the asymptotic properties of $\widehat{\boldsymbol{\theta}}$ and thus of $\widehat{F}_{12\ldots d}$. However, it does not provide the asymptotic variance of $\widehat{\boldsymbol{\theta}}$. In order to address the issue of relative efficiency of our procedure, we re-write our problem in the GMM framework. It is well known that the MLE can quite generally be written as a method of moment 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 $\widehat{F}_j \equiv F_j(\hat{\theta}_j)$, and the third step is the MLE for $c^{(3)}\left(\widehat{F}_1, \widehat{C}^{(2)}; \theta_{123}\right)$, where $\widehat{C}^{(2)} \equiv C^{(2)}(\widehat{F}_2, \widehat{F}_3; \hat{\theta}_{23})$. The corresponding

---

[4]Here by *pseudo*-copula we mean a possibly misspecified copula function. The same term is sometimes used in reference to the empirical copula obtained using univariate empirical cdf's.

GMM problems can be written as follows:

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

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

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

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 $\hat{\theta}_{23}$ and $\hat{\theta}_{123}$ and the correct form of the variance should account for that. The appropriate correction and simulation evidence of its effect are provided, e.g., 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 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 (5), 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 efficient even after adjustments for estimation error from preceding steps.

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 dataset 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 is typical for financial time series.

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

4.1 Univariate distributions

We would like to use skewed and thick-tailed distributions to model univariate marginals because this is observed in financial data. Azzalini and Capitanio (2003) propose one of possible generalizations of Student's t-distribution which is able to capture both these features. Moreover, their transformation does not restrict 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

$$
\begin{aligned}
Q_y &= \left(\frac{y-\xi}{\omega} \right)^2, \\
t_\nu(y) &= \frac{\Gamma((\nu+1)/2)}{\omega\,(\pi\nu)^{1/2}\Gamma(\nu/2)} \left(1 + Q_y/\nu\right)^{-(\nu+1)/2},
\end{aligned}
$$

and $T_{\nu+1}(x)$ denotes the cdf of the standard t-distribution with $\nu+1$ degrees of freedom. The parameter γ reflects 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 $\xi = 0$,

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

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 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 $\mathrm{St}_1(\xi, \omega, \gamma, \nu)$ 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 $\mathrm{St}(\gamma, \nu)$, its cdf by $F^{\mathrm{St}}_{\gamma,\nu}$ and pdf by $f^{\mathrm{St}}_{\gamma,\nu}$. We augment this distribution with the NAGARCH structure for the conditional variance equation (see, e.g., Engle and Ng, 1993):

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

where h_{it}'s are the conditional variances of 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:

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

13

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

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

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

4.2 Bivariate copulas

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

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

where $T_\eta^{-1}(\cdot)$ is the inverse of the standardized Student t cdf, d 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, we will use only the bivariate version of this copula.

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

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

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 $\mathbf{y}_t = (y_{1t}, \ldots, y_{pt})$:

$$F(\mathbf{y}_t) = C_{\eta, R_t}(F_1(y_{1t}), \ldots, F_d(y_{dt})). \tag{6}$$

We also derive the joint pdf by differentiating equation (6):

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

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

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

The number of parameters in this likelihood is $d(d-1)/2 + 3$ from the copula part and $7d$ 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:

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

It has only four parameters.

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, rather than symmetric ones, as compounding functions.

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:

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

where u denotes the marginal distribution of an asset, v denotes the distribution of a group of assets, and where we assume a similar time-varying structure on the correlation coefficient as in subsection 4.2. The form of the compounding function in the m-th step will then be

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

where $\theta_{l,i_1,\dots,l-1,l+1,\dots,i_m} = (\alpha, \beta, \eta, \rho, a, b)'$ is a set of parameters (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 dimension of the original problem.

4.4 Goodness-of-fit testing

In order to assess adequacy of the 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 $\Omega_t = \{y_{t-1}, y_{t-2}, \ldots\}$. 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 an exception that the statistical tests rather than 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 $\mathbf{X} = (X_1, \ldots, X_d)$ denote a random vector with marginal distributions $F_i(x_i)$ and conditional distributions $F_{i|i-1\ldots1}(x_i|x_{i-1}, \ldots, x_1)$ for $i = 1, \ldots, d$. The PIT of vector $\mathbf{x}$ is defined as $T(\mathbf{x}) = T(x_1, \ldots, x_d) = (T_1, \ldots, T_d)$ such that $T_1 = F_1(x_1)$, $T_p = F_{p|p-1\ldots1}(x_p|x_{p-1}, \ldots, x_1)$, $p = 2, \ldots, d$. One can show that $T(\mathbf{X})$ is uniformly distributed on the p-dimensional hyper-cube (Rosenblatt, 1952). This implies that $T_1, \ldots, T_p$ 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 $X_1 \mid X_2$. In this paper we examine both of them for all pairs.

5 Empirical application

This section demonstrates how to apply the new sequential technique to model a joint distribution of five DJIA constituents. We have considered larger numbers of stocks but to illustrate the advantage of our method over the 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.

5.1 Data

We choose the following five stocks from among DJIA-constituents (as of June 8, 2009): GE – General Electric Co, MCD – McDonald's Corp, MSFT – Microsoft Corp, KO – Coca-Cola Co, and PG – Procter & Gamble Co stocks. The selection is based on high level of liquidity and availability of historical prices. Daily data from Jan 03, 2007 to Dec 31, 2007 are collected; we focus on this period to avoid dealing with 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 1-2 and in Table 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 modeling marginals.

The scatter plots and correlations on Figure 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 sectors)[5]. 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.

5.2 Estimates of univariate distributions

We use the Skew-t-NAGARCH model for the marginals in order to accommodate 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.

---

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

The estimates of the parameters of marginal distributions are summarized in Table 2. As before, μ denotes the mean return, ω is the unconditional variance, α reflects the ability to predict conditional variance using current innovations, β is the measure of persistence of conditional variance, κ is the leverage effect, the reciprocal of ν captures heavy tails, and γ represents skewness. The mean return μ is fairly close to the sample mean value. There is a substantial degree if 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 skewness found in Table 1 and because it is now standard in the literature to account for these stylized facts.

The Kolmogorov–Smirnov tests for uniformity of the transformed series show that at the 95% confidence level the hypotheses of uniformity is not rejected. The quantitative results along with the diagrams are presented in Figure 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 moment on its 20 lags using the F-test. The results are presented in Table 3. The hypotheses of no serial correlation are not rejected at the 95% confidence level in nearly every case; the exception is the 4th central moment of the KO stock for which the hypothesis is not rejected at the 99% confidence 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.

5.3 Estimates of pairwise copulas

The pairwise copula parameter estimates are summarized in Table 4 and the results of pairwise Kolmogorov–Smirnov tests are presented in Figures 5 and 6. All Kolmogorov–Smirnov tests are passed at any reasonable confidence level. This indicates that the time-varying t-copula used for modeling 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 joint insignificance of the coefficients in the regression of each of the first four moments on their 20 lags using the F-test. 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.

5.4   Estimates of compounding functions

Tables 5, 6 and 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 again favor the selection of the asymmetrized bivariate t-copula as a compounding function.

5.5   Comparison to the conventional copula approach

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

The parameter estimates for the conventional benchmark method are summarized in Table 8. As before, we have run serial correlation tests – they are not reported due to their large number – 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 comparison. 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.[6] The conventional approach would require solving only 6, one of which is 13-dimensional. So for moderate dimensions, such as $d = 5$, the conventional method may be preferred in terms of computer time (it *was* quicker in this example). We will now increase the dimensionality of the problem and illustrate how the conventional estimator becomes harder obtain and eventually fails while the new procedure remains operational.

When we repeat the above exercise for $d = 6, \ldots, 15$, the number of parameters in the conventional MLE based on 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 $O(d^2)$), however this number can

---

[6]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 $\widehat{F}_{jklm}$.

be made small in steps 3 and above, if we consider a random subset of all available combinations in each step. Table 9 contains the number of parameters to be estimated in a single optimization problem when we use the new and the conventional method. 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]

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 discussed theoretical justification for 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.

Alternative methods of constructing a joint distribution from objects of lower dimension may come from work by de la Pena et al. (2006) and by 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 close form expressions of these objects are unknown and their properties are not as well studied as the properties of copula functions. For this reason, we leave the study of such alternative methods of modeling joint distributions for future research.

---

[7]A Matlab module handling arbitrary dimension and data sets under both conventional and sequential methodology is available at `http://alcor.concordia.ca/~aprokhor/research/reconstruct.zip`

References

AAS, K., C. CZADO, A. FRIGESSI, AND H. BAKKEN (2009): "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, 44, 182–198.

AUSIN, M. C. AND H. F. LOPES (2010): "Time-varying joint distribution through copulas," *Computational Statistics and Data Analysis*, 54, 2383–2399.

AZZALINI, A. AND A. CAPITANIO (2003): "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65, 367–389.

BREYMANN, W., A. DIAS, AND P. EMBRECHTS (2003): "Dependence structures for multivariate high-frequency data in finance," *Quantitative Finance*, 3, 1–14.

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

CARAGEA, P. C. AND R. L. SMITH (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 N. REID (2004): "A note on pseudolikelihood constructed from marginal densities." *Biometrika*, 91, 729–737.

DE LA PENA, V. H., R. IBRAGIMOV, AND S. SHARAKHMETOV (2006): "Characterizations of Joint Distributions, Copulas, Informaiton, Dependence and Decoupling, with Applications to Time Series," in *IMS Lecture Notes-Monograph Series - 2nd Lehmann Symposium : Optimality*, vol. 49, 183–209.

DIEBOLD, F. X., T. A. GUNTHER, AND A. S. TAY (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 *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, ed. by J. L. Castle and N. Shephard, Oxford University Press.

ENGLE, R. F. AND V. K. NG (1993): "Measuring and Testing the Impact of News on Volatility," *The Journal of Finance*, 48, 1749–1778.

ENGLE, R. F., N. SHEPHARD, AND K. K. SHEPPARD (2008): "Fitting Vast Dimensional Time-Varying Covariance Models," *NYU Working Paper No. FIN-08-009*.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995a): "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543–552.

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

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

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

Heagerty, P. J. and S. R. Lele (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.

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

Li, H., M. Scarsini, and M. Shaked (1999): "Dynamic Linkages for Multivariate Distributions with Given Nonoverlapping Multivariate Marginals," *Journal of Multivariate Analysis*, 68, 54–77.

Min, A. and C. Czado (2010): "Bayesian Inference for Multivariate Copulas Using Pair-Copula Constructions," *Journal of Financial Econometrics*, 8, 511–546.

Nelsen, R. B. (2006): *An Introduction to Copulas*, vol. 139 of *Springer Series in Statistics*, Springer, 2 ed.

Oh, D. H. and A. J. Patton (2013): "Simulated Method of Moments Estimation for Copula-Based Multivariate Models," *Journal of the American Statistical Association*.

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

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

——— (2009b): "Likelihood-based estimation in a panel setting: Robustness, redundancy and validity of copulas," *Journal of Econometrics*, 153, 93 – 104.

Quesada-Molina, J. and J. Rodriguez-Lallena (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," *The Annals of Mathematical Statistics*, 23, 470–472.

SHARAKHMETOV, S. AND R. IBRAGIMOV (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.

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

VARIN, C., N. REID, AND D. FIRTH (2011): "An Overview of Composite Likelihood Methods," *Statistica Sinica*, 21, 5–42.

VARIN, C. AND P. VIDONI (2005): "A note on composite likelihood inference and model selection." *Biometrika*, 92, 519–528.

——— (2006): "Pairwise likelihood inference for ordinal categorical time series," *Computational Statistics and Data Analysis*, 51, 23652373.

——— (2008): "Pairwise likelihood inference for general state space models," *Econometric Reviews*, 28, 170–185.

WATSON, M. W. AND J. H. STOCK (2004): "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, 23, 405–430.

ZHAO, Y. AND H. JOE (2005): "Composite likelihood estimation in multivariate data analysis," *Canadian Journal of Statistics*, 33, 335–356.

A   Appendix:

Proof of Proposition 1: We provide the proof for $d = 3$. Arguments for $d > 3$ are analogous.

Let $B \geq 0$ denote the Lipschitz constant. Lipschitz continuity of $\ln c^{(m)}$ means

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

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

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

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

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

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

It follows that

$$
\ln h(x_1, x_2, x_3) - \ln h_c(x_1, x_2, x_3) \quad \approx \quad \frac{B}{c^{(2)}(F_2, F_1)}.
$$

So, the approximation error is particularly small in areas of the support where $c^{(2)}$ concentrates a lot of mass.

Proof of Proposition 2: The result follows trivially from application of Sklar's (1959) theorem to $\widehat{F}_{12\ldots d}$.

Proof of Proposition 3: These are standard results cited, for example, in Chapter 10 of Joe (1997) or by Joe (2005).

	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-val)	(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-val)	(0.0017)	(0.0000)	(0.0000)	(0.0416)	(0.0000)

Table 1: Sample statistics for returns on GE, MCD, MSFT, KO, and PG stocks from Jan 3 to Dec 31, 2007.
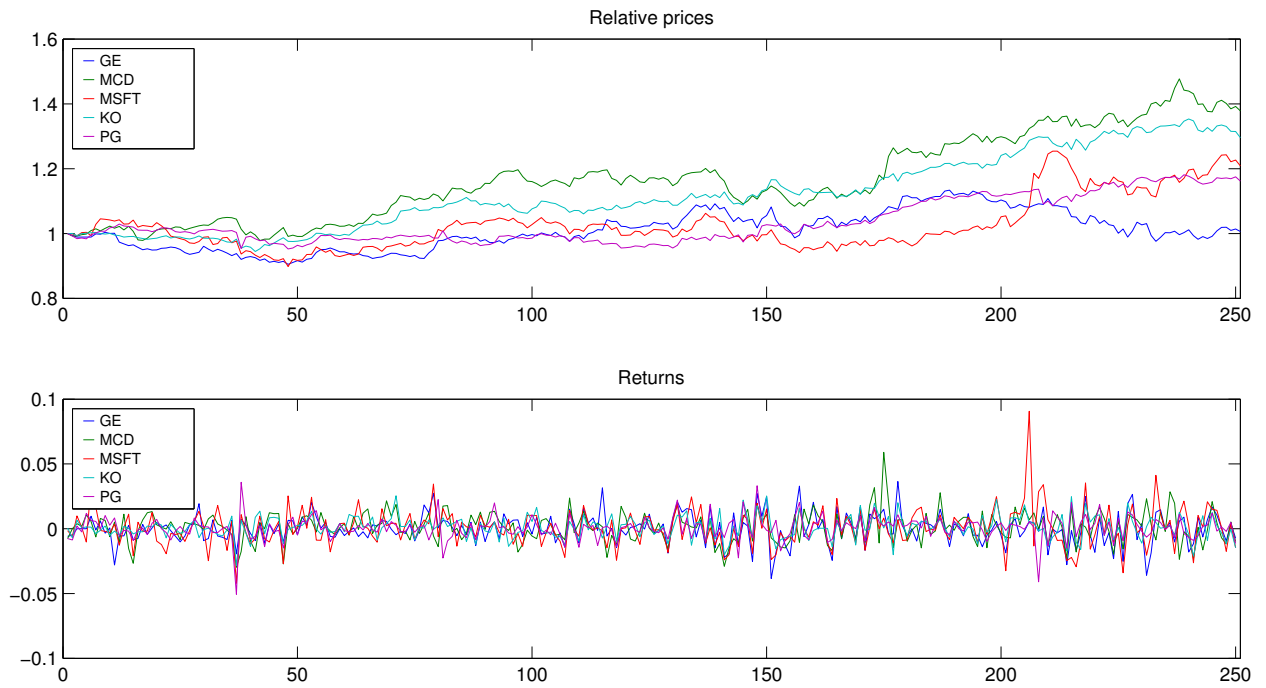

Figure 1: Relative prices and returns dynamics for GE, MCD, MSFT, KO, and PG from Jan 3 to Dec 31, 2007.
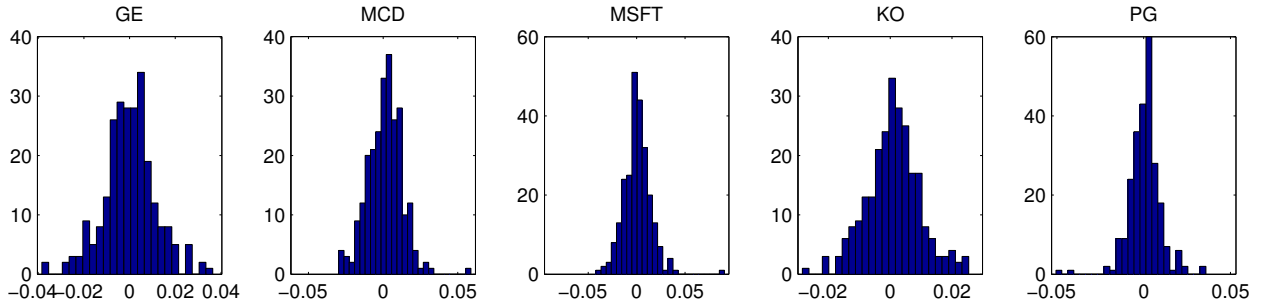
26

Figure 2: Histograms of GE, MCD, MSFT, KO, and PG returns from Jan 3 to Dec 31, 2007.

	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)

Table 2: Maximum likelihood parameter estimates for marginal distributions of GE, MCD, MSFT, KO, and PG returns from Jan 3 to Dec 31, 2007 (robust standard errors are in parentheses).

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

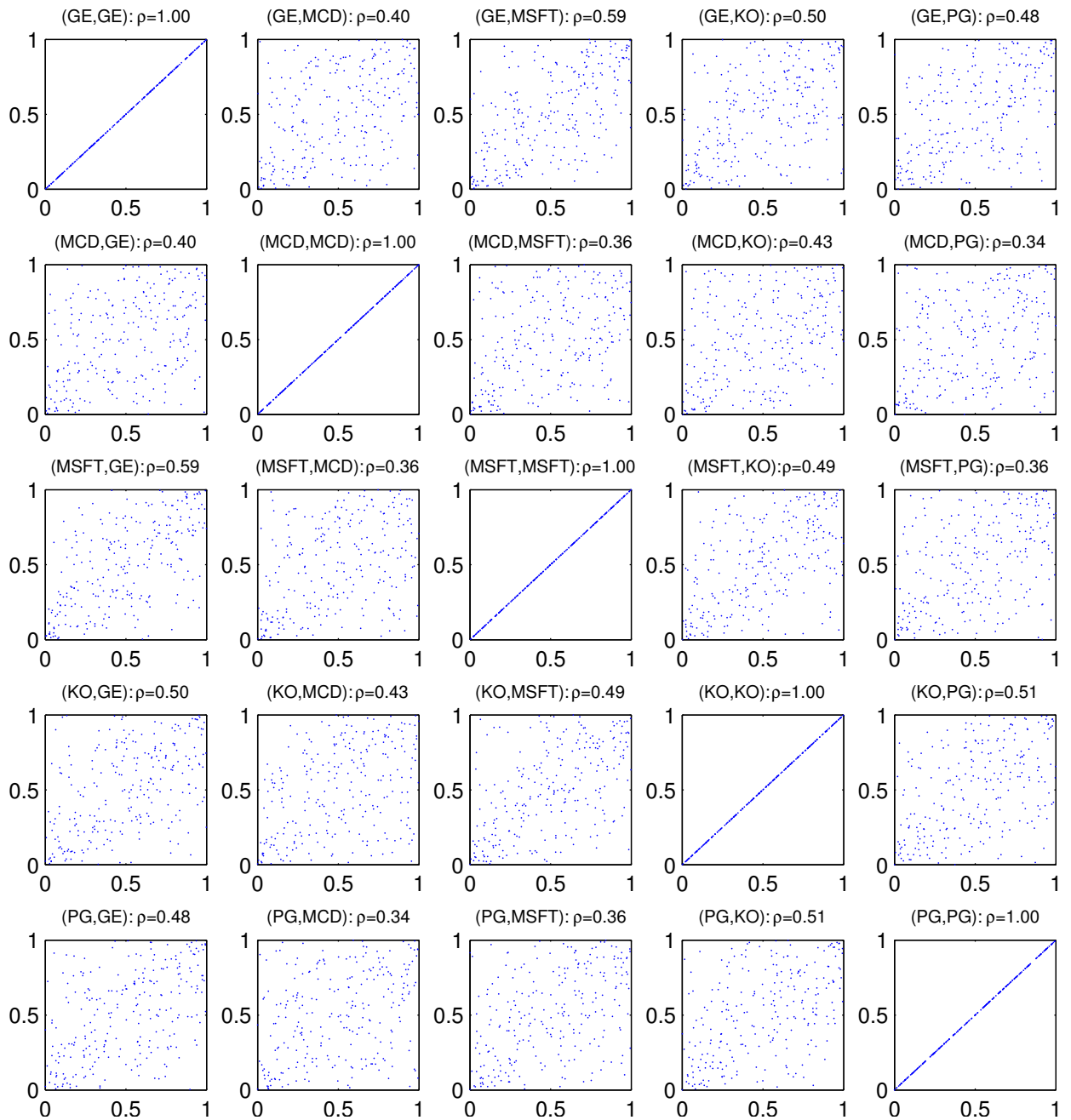Table 3: P-values of F-tests for serial correlation as GoF test of marginals.

Figure 3: Pairwise scatter plots of marginal distributions and sample correlations of GE, MCD, MSFT, KO, and PG returns from Jan 3 to Dec 31, 2007.
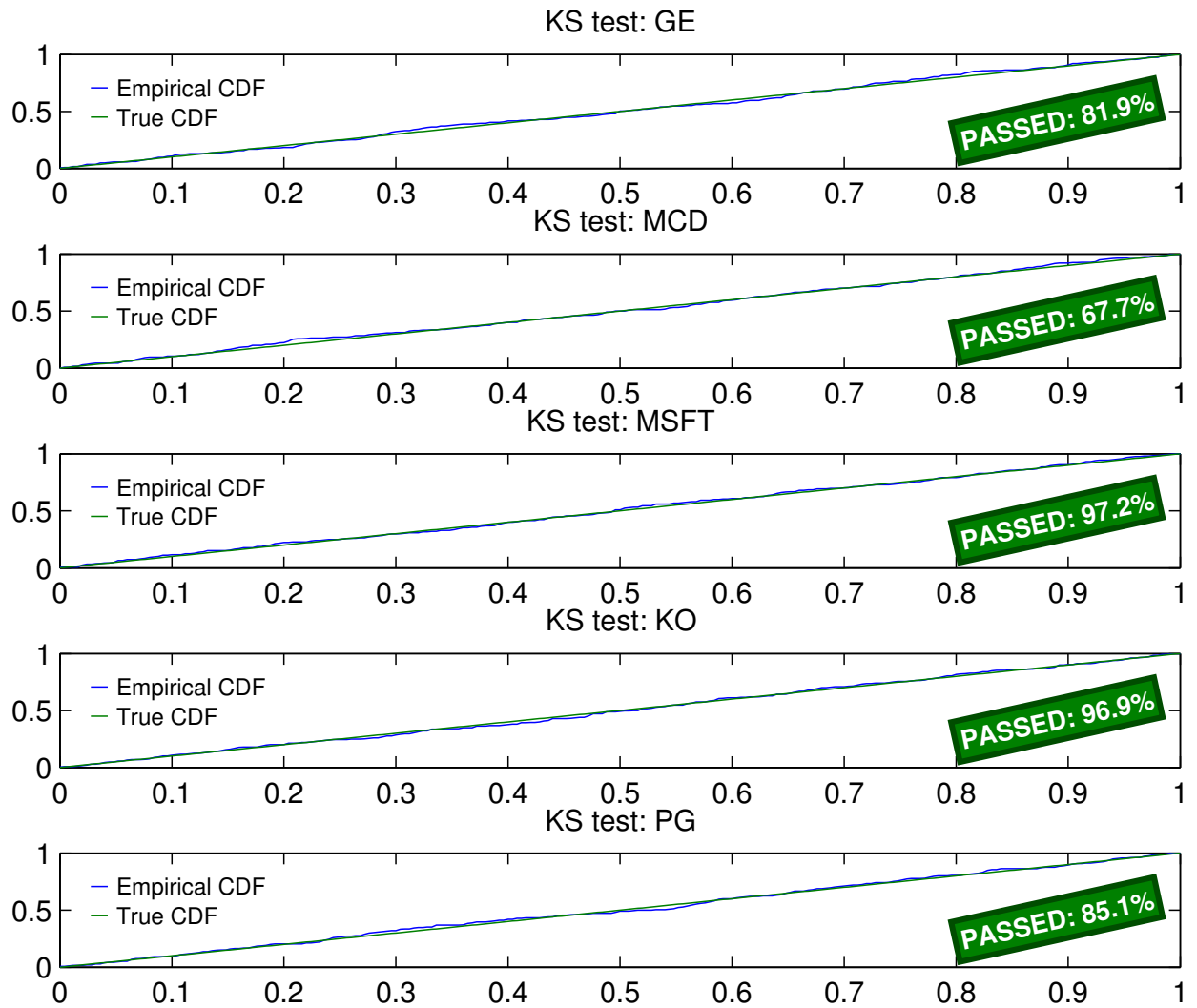
Figure 4: Kolmogorov–Smirnov tests of marginal distributions (p-values are indicated along with passed/not passed)

	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)
$\bar{\rho}$	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)
$\bar{\rho}$	0.417	0.368	0.556	0.469	0.504
	(0.059)	(0.080)	(0.056)	(0.065)	(0.050)

Table 4: Maximum likelihood parameter estimates for pairwise copulas of GE, MCD, MSFT, KO, and PG returns from Jan 3 to Dec 31, 2007 (robust standard errors are in parentheses).

$\widetilde{C}\left(\cdot;\cdot\right)$	α	β	η	$\overline{\rho}$	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 5: Maximum likelihood parameter estimates of the t-based compounding functions for groups of three assets (standard errors omitted).

$\widetilde{C}\left(\cdot;\cdot\right)$	α	β	η	$\overline{\rho}$	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 6: Maximum likelihood parameter estimates of t-copula based compounding functions for groups of four assets (standard errors omitted).

$\widetilde{C}\left(\cdot;\cdot\right)$	α	β	η	$\overline{\rho}$	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

Table 7: Maximum likelihood parameter estimates of t-copula based compounding functions for groups of five assets (standard errors omitted).

		$\{\overline{R}_{ij}\}$	1	2	3	4	5
η	13.426	1	1.000	0.425	0.621	0.502	0.510
	(4.380)		(0.000)	(0.055)	(0.042)	(0.055)	(0.049)
a	0.030	2	0.425	1.000	0.415	0.398	0.367
	(0.035)		(0.055)	(0.000)	(0.056)	(0.055)	(0.062)
b	0.157	3	0.621	0.415	1.000	0.539	0.465
	(0.308)		(0.042)	(0.056)	(0.000)	(0.053)	(0.057)
		4	0.502	0.398	0.539	1.000	0.495
			(0.055)	(0.055)	(0.053)	(0.000)	(0.049)
		5	0.510	0.367	0.465	0.495	1.000
			(0.049)	(0.062)	(0.057)	(0.049)	(0.000)

Table 8: Maximum likelihood parameter estimates of time-varying five-dimensional t-copula for (1) GE, (2) MCD, (3) MSFT, (4) KO, and (5) PG stock returns from Jan 3 to Dec 31, 2007 (robust standard errors are in parentheses).

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

Table 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.
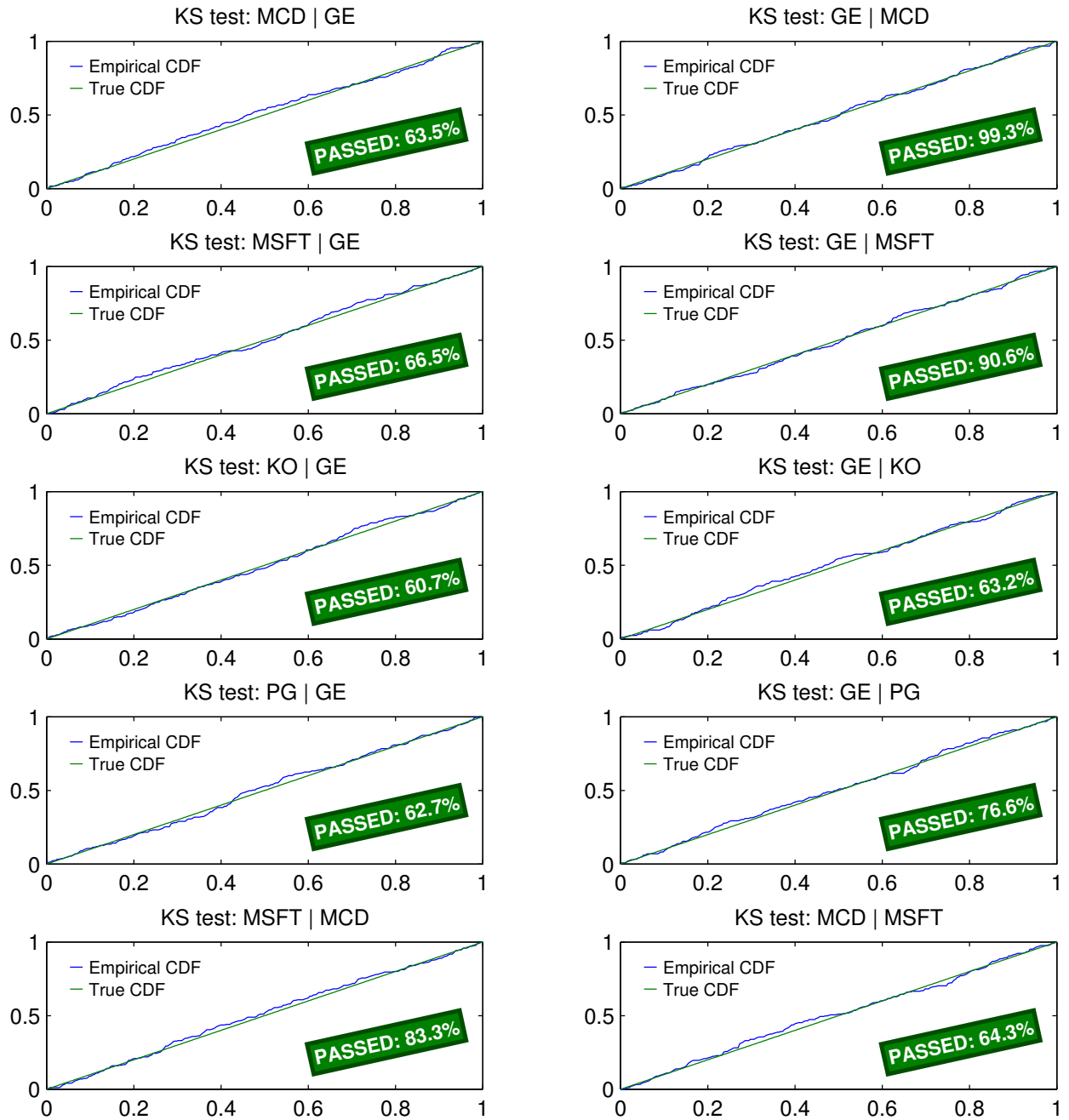
Figure 5: Pairwise Kolmogorov–Smirnov tests of conditional distributions for the first five pairs of GE, MCD, MSFT, KO, and PG – GoF tests of bivariate copula specification (p-values are indicated along with passed/not passed).
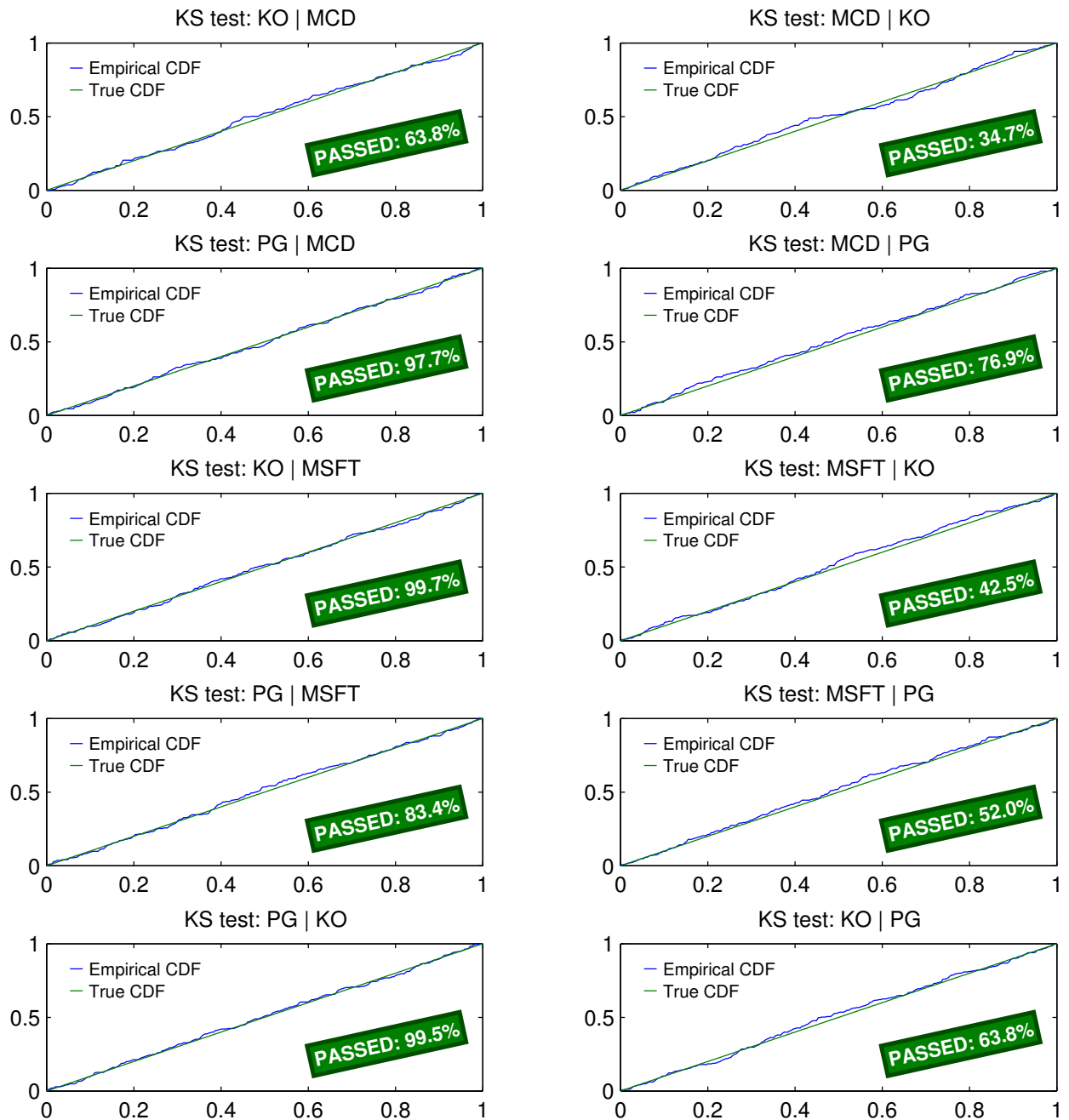
Figure 6: Pairwise Kolmogorov–Smirnov tests of conditional distributions for the second five pairs of GE, MCD, MSFT, KO, and PG – GoF tests of bivariate copula specification (p-values are indicated along with passed/not passed)