Incorporating Estimation Risk in Copula-based
Portfolio Optimization *

Yue Xu†
This Version: December 5, 2004

Abstract

This paper develops a Bayesian Markov Chain Monte Carlo algorithm to jointly estimate
copula-based multivariate models and then uses it to study the issue of estimation risk in portfo-
lio optimization problems. Because of the computational challenges introduced by the presence
of copula functions, the copula-based models are often estimated by two-stage methods. A
Bayesian MCMC algorithm is developed to jointly estimate the models. It is then embedded
in a procedure to draw from the predictive distribution of a pair of asset returns to incorporate
estimation risk in a portfolio optimization problem. The results show that for an investor with
moderate risk aversions the optimal portfolio weights on risky asset are much less than the case
where the estimation risk is ignored.

*I am especially grateful to Keisuke Hirano for his continuous support, advice and patient guidance. I also wish
to thank Adrian Peralta-Alva, Tie Su, David Kelly and those who have assisted me in this effort. This paper is
preliminary and comments are welcome. Any errors are mine.

†Correspondence address: Department of Economics, University of Miami, P. O. Box 248126, Coral Gables, FL
33124-6550. Email: yxu2@umsis.miami.edu.
1 Introduction

A common problem in portfolio optimization is that the parameters in the probabilistic model of asset returns are usually unknown and have to be estimated from data. When we use the estimates to calculate optimal asset allocations, extra risk is introduced into the problem, which is often called estimation risk. For example, within the framework of the mean-variance analysis, portfolio weights are very sensitive to the level of the expected returns and using the corresponding estimates may bring in extra uncertainty in portfolio optimization.

A copula is a special multivariate distribution function with marginal distributions being uniform with support on [0, 1]. It fully captures the dependence structures between two or more random variables. Several recent papers have used the copula theory to consider asymmetric dependence between financial variables. For instance, Patton (2004) uses copula-based models to study the importance of skewness and asymmetric dependence for asset allocation. Hu (2003) uses mixed copula functions to estimate association across financial markets. Van den Goorbergh and Genest (2003) use a copula-based model to study multivariate option pricing. In these studies, two-stage methods is used to estimate the joint models, and the estimation risk is largely ignored. In this paper, we develop a Bayesian MCMC algorithm to jointly estimate the copula-based models. One advantage of the Bayesian method is that we get a simulated sequence of the model parameter values as a byproduct, which can be nested in a sampling algorithm to generate simulations from the predictive distribution. We apply the algorithm to a copula-based portfolio optimization problem and investigate how the optimal portfolio weights are influenced once the investor incorporates estimation risk.

A large number of papers have tried to incorporate estimation risk in portfolio choice, especially focusing on the impact on optimal portfolio weights. For example, Chen and Brown (1983), Alexander and Resnick (1985), and Barberis (2000). They find that after considering the estimation risk a risk-averse investor usually allocates much less wealth on risky assets. A natural way to incorporate estimation risk is the Bayesian approach. The available data are used to get a predictive distribution, which is only conditional on the past asset returns. The optimal portfolio maximizes the expected utility with respect to this predictive distribution. Using the Bayesian approach

---

1 Joe (1997) and Nelsen (1998) provide detailed discussion of copula theory.
2 The correlation between asset returns are higher in downturn markets than in upturn market, as reported in Longin and Solnik (2001), and Ang and Chen (2002).
under the framework of mean-variance analysis with the assumption of multivariate normal and i.i.d. returns, we can get exact analytic solutions for optimal portfolio weights; see Bawa et al. (1979). In other cases, for example, with power utility and non-normal asset returns, we use numerical methods to do posterior computations, similar to Barberis (2000). We first draw from the posterior distribution of the parameters, then simulate asset returns conditional on the simulated parameter values. After we introduce the copula-based models for asset returns, the computation is even more difficult. Conjugate families are not available for the joint models, and the complicated form of the joint distribution makes it hard to sample directly from the posterior distribution. We develop a “Metropolis-within-Gibbs” algorithm to draw from the conditional posterior distributions iteratively.

We apply the Bayesian MCMC algorithm to investigate the estimation risk issue in copula-based portfolio optimization problem. We use the skewed-$t$ distributions to model the marginal processes of the two asset returns and use a mixed copula to combine them to construct a bivariate distribution. The Bayesian MCMC method is used to jointly estimate the model parameters, and then it is embedded in a sampling algorithm to simulate asset returns from the predictive distribution. The optimal portfolio weights can be found by maximizing the expected utility, where the expectation is calculated by numerical integration using the simulations from the predictive distribution.

The paper is organized as follows. Section 2 develops a general Bayesian MCMC algorithm to jointly estimate the copula-based models. Section 3 applied the Bayesian algorithm to a copula-based portfolio optimization problem. The focus is the effects of estimation risk on optimal portfolio weights. Section 4 concludes and provides some suggestions for future research.

2 A Bayesian Method to Jointly Estimate Copula-based Models

A copula is a special multivariate distribution whose marginal distributions are $Unif(0,1)$. It is often used to combine two (or more) marginal distributions to form a new joint distribution.

2.1 Copulas and Sklar’s Theorem

For two random variables$^3$ $X$ and $Y$, with marginal distribution functions $F(x) = \Pr(X < x)$ and $G(y) = \Pr(Y < y)$, let their joint distribution function be $H(x, y) = \Pr(X < x, Y < y)$.

$^3$In this paper, we focus on two-dimensional copulas, but the ideas can be generalized to $n$-dimensional cases.
**Definition 1 (Copula)** The copula of \((X,Y)\), where \(X \sim F\), \(Y \sim G\), \(F\) and \(G\) are continuous distribution functions, is the joint distribution function of \(U \equiv F(X)\) and \(V \equiv G(Y)\), denoted as \(C(u,v)\). The function \(c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}\) is the density function of the copula \(C(u,v)\).

It can be shown that the random variables \(U\) and \(V\) have \(Unif(0,1)\) distributions regardless of the original distributions. Thus, a copula \(C(u,v)\) is simply a multivariate joint distribution with support on \([0,1] \times [0,1]\) and uniform marginal distributions. The concept of copula is very useful in describing multivariate dependence. This can be seen from the following theorem.

**Theorem 1 (Sklar’s Theorem)** Let \(H\) be the joint distribution function of \((X,Y)\) and \(F,G\) be the continuous marginal distribution functions of \(X\) and \(Y\). Then there exists a unique copula \(C\) such that

\[
H(x,y) = C(F(x),G(y)).
\]

(1)

Conversely, let \(F\) and \(G\) be continuous distribution functions and \(C\) be a copula, then the function \(H\) defined by equation (1) is a bivariate distribution function with marginal distributions \(F\) and \(G\).

This theorem first appeared in Sklar (1959). It shows that we can decompose a bivariate distribution into three parts: two marginal distributions \(F\) and \(G\), and a copula. Since \(F\) and \(G\) contain all the information about marginal distributions, what is left is the dependence structure, which is completely captured by the copula. Sklar’s theorem also provides a scheme to extract the copula from a known multivariate distribution. Given \(H(x,y)\) with marginal distribution functions \(F(x)\) and \(G(y)\), the function

\[
C(u,v) = H(F^{-1}(u),G^{-1}(v))
\]

is a two-dimensional copula.

For a bivariate distribution function \(H(x,y)\), it can be shown that

\[
h(x,y) \equiv \frac{\partial^2 H(x,y)}{\partial x \partial y}
= \frac{\partial^2 C(F(x),G(y))}{\partial u \partial v} \frac{\partial F(x)}{\partial x} \frac{\partial G(y)}{\partial y}
= c(F(x),G(y))f(x)g(y),
\]

where \(h(x,y)\) is the density function. This is very useful for likelihood-based analysis. Suppose the parameter vector \(\theta = (\theta_X, \theta_Y, \theta_C)\), where \(\theta_X, \theta_Y\) are the parameters in the marginal distributions,
and $\theta_C$ is the parameter in the copula function. The joint density of the two random variables can be decomposed as follows:

$$h(x, y|\theta) = c\left(F(x|\theta_X), G(y|\theta_Y)|\theta_C\right)f(x|\theta_X)g(y|\theta_Y).$$

Then, if the random vector $(x_t, y_t)'$ is serially independent, the log likelihood function is

$$L_{XY} = L_X + L_Y + L_C,$$

where

$$L_{XY} = \sum_{t=1}^{T} \log h(x_t, y_t|\theta),$$

$$L_X = \sum_{t=1}^{T} \log f(x_t|\theta_X),$$

$$L_Y = \sum_{t=1}^{T} \log g(y_t|\theta_Y),$$

$$L_C = \sum_{t=1}^{T} \log c\left(F(x_t|\theta_X), G(y_t|\theta_Y)|\theta_C\right).$$

The separability of the log likelihood function naturally gives rise to a two-stage estimation method. In the first stage, the parameters in the two marginal models are estimated by maximizing the log likelihood functions $L_1$, $L_2$, and we get estimates $\hat{\theta}_1$ and $\hat{\theta}_2$. In the second stage, we take the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ as given and maximize $L_C$ with respect to $\delta$ to get the estimate $\hat{\delta}$. We also take advantage of the separability property when we carry out the joint Bayesian estimation: we use the Gibbs sampling method to draw from the conditional distributions of $\theta_X$, $\theta_Y$, $\theta_C$ iteratively, and we may use the posterior distributions associate with the univariate models as proposal distributions in the Metropolis-Hastings step.

There are many copula functions. We list in Appendix A the functional forms of those used in this paper. We choose two standard copula functions: the Gaussian copula and the Clayton copula. The Gaussian copula is extracted from the multivariate normal distribution. In the case where we assume multivariate normal returns, we assume the dependence is symmetric implicitly and use the Gaussian copula to model it. The Clayton copula can capture asymmetric dependence. It
exhibits lower tail dependence but no upper tail dependence\textsuperscript{4}. To take into account the uncertainty about the parametric form of the copula, we use the following mixture to model dependence,

\[
MC(u, v) = \lambda C_C(u, v) + (1 - \lambda) C_{Ga}(u, v),
\]

where \(\lambda \in (0, 1)\), and \(C_C, C_{Ga}\) denote the Clayton copula and the Gaussian copula. We note that a mixture of copulas is also a valid copula function. A similar mixed copula approach is also used in Hu (2003).

In this paper, we use simulation methods to solve the portfolio optimization problem, which involve generating random numbers from copula-based multivariate distribution \(H(x, y) = C(F(x), G(y))\). We first simulate from the copula \(C(u, v)\) using the conditional distribution \(C_{2|1} = \frac{\partial C}{\partial u}(u, v):\) if \((U, Q)\) are independent random numbers from \(Unif(0, 1)\), then \((U, V) = (U, C_{2|1}^{-1}(Q|U))\) has the distribution \(C\). If there is no closed form for \(C_{2|1}^{-1}\) we may use numerical methods to find the root of \(q = C_{2|1}(v|u)\). Let \(X = F^{-1}(U), Y = G^{-1}(V)\), then \((X, Y)\) has the joint distribution \(C(F(x), G(y))\).

2.2 Using Copulas to Model Co-movement between Asset Returns

We model asset returns in the following way,

\[
r_t = f(X_t) + \sigma_t \varepsilon_t,
\]

where \(f(\cdot)\) is some function. \(X_t\) may include previous asset returns and other predictor variables such as dividend yield. \(\varepsilon_t\) is disturbance term with zero mean and unit variance, and \(\sigma_t\) is the standard deviation. \(\sigma_t\) may be a constant, leading to a homoscedastic model, or time varying, leading to a heteroscedastic model. In this paper, we assume the asset returns are i.i.d. over time, thus \(\sigma_t\) does not change over time.

To model the co-movements between two asset returns \(r_{1t}\) and \(r_{2t}\), we assume the disturbances

\textsuperscript{4}If the limit

\[
\lim_{\varepsilon \to 0} \Pr(U \leq \varepsilon | V \leq \varepsilon) = \lim_{\varepsilon \to 0} \frac{\Pr(U \leq \varepsilon, V \leq \varepsilon)}{\Pr(V \leq \varepsilon)} = \lim_{\varepsilon \to 0} \frac{C(\varepsilon, \varepsilon)}{\varepsilon} = \tau^L
\]

exists, then the copula \(C\) exhibits lower tail dependence if \(\tau^L \in (0, 1]\) and no lower tail dependence if \(\tau^L = 0\). The lower tail dependence measures the possibility that one variable takes extremely small values given that the other variable took extremely small values. The upper tail dependence can be defined in a similar way.
are from some joint distribution, which is formed by using some copula to combine two marginal
distributions, and the joint model for \((r_{1t}, r_{2t})\) is

\[
\begin{pmatrix} r_{1t} \\ r_{2t} \end{pmatrix} \sim C(F(r_{1t}|\theta_1), G(r_{2t}|\theta_2)|\delta),
\]

where \(F(\cdot)\) and \(G(\cdot)\) are marginal distribution functions with parameter vector \(\theta_1\) and \(\theta_2\), and \(\delta\) is the parameter vector in the copula function.

### 2.3 A Bayesian MCMC Algorithm to Jointly Estimate Copula-based Models

For parameters \(\{\theta_1, \theta_2, \delta\}\) in the general copula-based model (3), with the observations \(z = \{r_{1i}, r_{2i}\}_{i=1}^T\), the posterior has the following property,

\[
p(\theta_1, \theta_2, \delta|z) \propto p(z|\theta_1, \theta_2, \delta)p(\theta_1, \theta_2, \delta),
\]

where

\[
p(z|\theta_1, \theta_2, \delta) = \prod_{i=1}^T c[F(r_{1i}|\theta_1), G(r_{2i}|\theta_2)|\delta] \cdot f(r_{1i}|\theta_1) \cdot g(r_{2i}|\theta_2),
\]

and \(p(\theta_1, \theta_2, \delta)\) is the prior distribution. The analytic form of the posterior is complicated because of the presence of the copula function. Our strategy is to use the hybrid Gibbs sampling method to draw from the posterior (joint) distribution: group the parameters into three blocks \(\{\theta_1\}\), \(\{\theta_2\}\), \(\{\delta\}\) and draw from the posterior distributions for each of them conditional on the other two blocks and observations. The algorithm of drawing \(\theta_1|z, \theta_2, \delta\) is similar to that of drawing \(\theta_2|z, \theta_1, \delta\).

In some steps of the Gibbs sampler, it may be difficult to sample directly from the conditional distributions and we use a Metropolis-Hastings step to make the simulation. This method is first suggested by Müller (1991, 1993), and it sometimes called “Metropolis-within-Gibbs”.

#### 2.3.1 Draw \(\theta_1|z, \theta_2, \delta\)

Conditional on \(\{z, \theta_2, \delta\}\), \(g(r_{2i}|\theta_2)\) is a constant. Thus we have

\[
p(\theta_1|z, \theta_2, \delta) \propto \left\{ \prod_{i=1}^T c[F(r_{1i}|\theta_1), G(r_{2i}|\theta_2)|\delta] \right\} \cdot \left\{ \prod_{i=1}^T f(r_{1i}|\theta_1) \right\} \cdot p_{\theta_1}(\theta_1),
\]
where \( p_{\theta_1}(\theta_1) \) is the prior for \( \theta_1 \). The product of the last two terms turns out to be proportional to the posterior density of \( \theta_1 \) that resulted from using the univariate marginal distribution \( F \) to model the return series \( \{r_{1i}\}_{i=1}^T \). We have well-developed theory on the Bayesian analysis of such univariate models and it is not hard to simulate from its posterior distribution. This suggests a Metropolis-Hastings algorithm using the “posterior” of \( \theta_1 \) as the independent proposal distribution\(^5\).

Define two functions as follows,

\[
\begin{align*}
h(x) &= \left\{ \prod_{i=1}^T c[F(r_{1i}|x), G(r_{2i}|\theta_2)|\delta] \right\} \cdot \left\{ \prod_{i=1}^T f(r_{1i}|x) \right\} \cdot p_{\theta_1}(x), \\
q(x) &= \left\{ \prod_{i=1}^T f(r_{1i}|x) \right\} \cdot p_{\theta_1}(x),
\end{align*}
\]

thus, \( h(x) \) is proportional to the entire conditional posterior density of \( \theta_1 \), and \( q(x) \) is proportional to the “posterior” density under univariate models. The Metropolis-Hastings algorithm is as follows,

1. Generate \( \tilde{\theta}_1 \sim q(x) \),
2. Take

\[
\theta_1^{(j+1)} = \begin{cases} \tilde{\theta}_1 & \text{with probability } \min\left\{ \frac{h(\tilde{\theta}_1)q(\theta_1^{(j)})}{q(\tilde{\theta}_1)h(\theta_1^{(j)})}, 1 \right\} \\ \theta_1^{(j)} & \text{otherwise.} \end{cases}
\]

Notice that the acceptance probability can be further simplified as follows,

\[
\frac{h(\tilde{\theta}_1)q(\theta_1^{(j)})}{q(\tilde{\theta}_1)h(\theta_1^{(j)})} = \frac{\prod_{i=1}^T c[F(r_{1i}|\tilde{\theta}_1), G(r_{2i}|\theta_2)|\delta]}{\prod_{i=1}^T c[F(r_{1i}|\theta_1^{(j)}), G(r_{2i}|\theta_2)|\delta]}.
\]

Therefore, the algorithm only involves simulating from the “posterior” of \( \theta_1 \) and evaluating the

\(^5\)The proposal distribution is the posterior of \( \theta_1 \) given the univariate marginal model for the first return series. We use quotation marks to distinguish this posterior from the posterior of \( \theta_1 \) under the joint copula-based models.
copula p.d.f. functions using two different marginal parameter values. These two steps are not difficult to carry out. The algorithm also suggests that the strength of dependence between the two random variables will be the key factor in how much a joint analysis affects the inference for $\theta_1$ and $\theta_2$: the candidate of simulation is generated from the “posterior” distribution, and the acceptance probability depends on the copula functions.

This algorithm is an application of independent Metropolis-Hastings, and the proposal distribution $q(x)$ for $\theta^{(j+1)}$ does not depends on $\theta^{(j)}$. In a general Metropolis-Hastings algorithm the proposal distribution would take the form of $q(x|\theta^{(j)})$. To allow the proposal distribution to depend on previously simulated value, we may use random walks: simulate the candidate $\bar{\theta}$ according to $\bar{\theta} = \theta^{(j)} + \epsilon$, where $\epsilon$ is a random perturbation with distribution $g$. If $g$ is a symmetric function (that is, $g(-t) = g(t)$), the acceptance probability will reduce to

$$\min\left\{1, \frac{h(\bar{\theta})}{h(\theta^{(j)})}\right\}.$$ 

The algorithm outlined in Appendix B uses this random walk Metropolis-Hastings method.

It is necessary to impose some regularity conditions on both the target distribution $h$ and the proposal distribution $q$. It turns out that the minimal regularity condition is very general. For every distribution $q$, whose support includes the support of $h$, the Metropolis-Hastings algorithm, using $q$ as a proposal distribution and defining the acceptance probability as in the above algorithm, will generate a Markov chain with $h$ as its stationary distribution\(^6\). In the above algorithm, the proposal distribution density $q$ is a part of the target distribution density $h$, hence the support of $h$ is included in the support of $q$. The minimal regularity condition is satisfied. However, this regularity condition is only for the convergence of the Metropolis-Hastings algorithm, not for the Metropolis-within-Gibbs.

2.3.2 Draw $\delta | z, \theta_1, \theta_2$

Conditional on $\{z, \theta_1, \theta_2\}$, both $f(r_{1i}|\theta_1)$ and $g(r_{2i}|\theta_2)$ are constants, thus we have

$$p(\delta|z, \theta_1, \theta_2) \propto \left\{ \prod_{i=1}^{T} c[F(r_{1i}|\theta_1), G(r_{2i}|\theta_2)|\delta] \right\} \cdot p_\delta(\delta),$$

\(^6\)Robert and Casella (1999) provide proofs of related theorems.
where \( p_\delta(\delta) \) is the prior for \( \delta \). We outline the Metropolis-Hastings algorithm to draw \( \delta \) as follows. Define two functions,

\[
\begin{align*}
    h(x) &= \left\{ \prod_{i=1}^{T} c[F(r_{1i}|\theta_1), G(r_{2i}|\theta_2)|x] \right\} \cdot p_\delta(x), \\
    q(x) &= p_\delta(x).
\end{align*}
\]

Then, \( h(x) \) is proportional to the posterior density of \( \delta \), and \( q(x) \) is just the prior density. The Metropolis-Hastings algorithm is as follows,

1. Given \( \delta^{(j)} \),

2. Generate \( \tilde{\delta} \sim q(x) \),

3. Take

\[
\delta^{(j+1)} = \begin{cases} 
\tilde{\delta} & \text{with probability } \min\left\{ \frac{h(\tilde{\delta})q(\delta^{(j)})}{q(\delta)h(\delta^{(j)})}, 1 \right\} \\
\delta^{(j)} & \text{otherwise.}
\end{cases}
\]

Notice that the acceptance probability can be further simplified as follows,

\[
\frac{h(\tilde{\delta})q(\delta^{(j)})}{q(\delta)h(\delta^{(j)})} = \frac{\prod_{i=1}^{T} c[F(r_{1i}|\theta_1), G(r_{2i}|\theta_2)|\tilde{\delta}]}{\prod_{i=1}^{T} c[F(r_{1i}|\theta_1), G(r_{2i}|\theta_2)|\delta^{(j)}]}
\]

Therefore, the algorithm only involves simulating from the prior distribution of \( \delta \) and evaluating the copula p.d.f. functions using two different parameter values. As before, it is straightforward to show that the minimal regularity condition for Metropolis-Hastings is satisfied. One possible problem with this algorithm is the inefficiency because of using the prior as the proposal distribution. An alternative is to construct an over-dispersed distribution to approximate the target distribution and use the approximation as the proposal distribution. Or, we may use an adaptive Metropolis-Hastings algorithm.
2.3.3 Monitoring Convergence

In this paper, we use the following three methods to monitor the convergence of the algorithm. A natural way is to draw graphs of the simulated chains. As Robert and Casella (1999) state, this method is only useful to test strong nonstationarities.

We also use a nonparametric test, the Kolmogorov-Smirnov test, to assess stationarity. The idea is straightforward: if the chain \( \{ \theta(t) \} \) is stationary, then \( \theta(t_1) \) and \( \theta(t_2) \) should have the same marginal distribution for arbitrary times \( t_1 \) and \( t_2 \). From an MCMC sample \( \{ \theta(t) \}_{t=1}^{T} \), we can form two subsamples \( \{ \theta(t) \}_{t=1}^{T/2} \) and \( \{ \theta(t) \}_{t=T/2+1}^{T} \). To make approximately i.i.d. samples, we may choose the first point in every three points in the original sample to form a subsample, i.e., \( \{ \theta(1), \theta(4), \ldots \} \).

Suppose the two subsamples are \( \{ \theta_1(k) \}_{k=1}^{m} \), \( \{ \theta_2(t) \}_{t=1}^{n} \), where \( m \) and \( n \) may not be equal. Let \( F_m(z) \) and \( G_n(z) \) denote the two empirical distributions determined by the two subsamples. Then, the Kolmogorov-Smirnov statistic is

\[
K = \sup_{-\infty < z < \infty} \left| F_m(z) - G_n(z) \right|
\]

which is the largest difference between the two empirical distributions. Under the null hypothesis that

\[
H_0 : F_m(z) = G_n(z) \text{ for all } z
\]

it can be shown that

\[
\lim_{m,n \to \infty} P \left( \sqrt{\frac{mn}{m+n}} \cdot K < x \right) = Q(x),
\]

where

\[
Q(x) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2}.
\]

Thus, the statistic \( K \) has asymptotic Kolmogorov-Smirnov distribution, and we may use it to perform the test.

Gelman and Rubin (1992) develop an alternative method for assessing convergence, which is based on multiple chains of the iterative simulation with starting values sampled from some over-dispersed distribution. At each step of the simulation of each scalar parameter we calculate an estimate of potential scale reduction factor (PSRF) which indicates how much better the distributional estimate would be if the simulation were continued indefinitely. As the number of
simulations goes to infinity, this factor declines to 1. In practice, when the factor is near 1 for all scalar parameters, we may summarize the target distribution by a set of simulations. This method is fully quantitative in monitoring convergence, and can be easily embedded in the simulation algorithm. There are several other tests and monitoring methods to assess the convergence of averages and convergence to i.i.d. sampling; see Robert and Casella (1999).

3 Incorporating Estimation Risk in Portfolio Optimization

In portfolio optimization problem, the model parameters are unknown and we usually plug in their estimates. This practice brings in extra risk, which is often called estimation risk. In this section, we follow Kandel and Stambaugh (1996) to show how the estimation risk can be incorporated by a Bayesian method, and then apply the framework to a copula-based portfolio optimization problem.

3.1 General Idea

We consider a risk-averse investor at time $T$. She allocates her initial wealth between two assets: a low risk - low return (L-L asset, hereafter) asset and a high risk - high return (H-H asset, hereafter) asset, to maximize one-period-ahead expected utility. Let $\omega$ denote the fraction of her wealth allocated to H-H asset. We set $0 \leq \omega \leq 1$, and the portfolio is short sales constrained. At the end of the next period the investor’s wealth is

$$W_{T+1} = W_T \left[ (1 - \omega) \exp(r_{1,T+1}) + \omega \exp(r_{2,T+1}) \right],$$

where $W_T$ is the initial wealth at time $T$, $r_{1,T+1}$ is the continuously compounded return of the L-L asset during the next period and the $r_{2,T+1}$ is that of the H-H asset.

The investor’s preference over the wealth is represented by the constant relative risk aversion power utility function

$$v(W) = \begin{cases} 
\frac{1}{1 - A} W^{1-A} & \text{if } A > 0 \text{ and } A \neq 1, \\
\log W & \text{if } A = 1.
\end{cases}$$

Let $z$ denote the past returns of the assets available at time $T$, and $R$ denote $(r_{1,T+1}, r_{2,T+1})'$. Suppose we use a probabilistic model to describe the movement of asset returns and the parameter
vector is \( \theta \). The value of the parameter is unknown, and the past returns could be used to get an estimate, say \( \hat{\theta} \). Ignoring estimation risk, the investor solves the following problem

\[
\max_\omega \int_R v(W_{T+1}) p(R|z, \hat{\theta}) dR. \tag{4}
\]

If we use the Bayesian approach to incorporate uncertainty over \( \theta \), the expected utility is defined in terms of the predictive density function \( p(R|z) \) instead of the conditional density \( p(R|z, \hat{\theta}) \). The predictive density function is obtained by integrating \( \theta \) out of the posterior joint density of \( R \) and \( \theta \). Since \( p(R, \theta|z) = p(R|\theta)p(\theta|z) \), the predictive density function can be obtained from

\[
p(R|z) = \int p(R|\theta)p(\theta|z)d\theta, \tag{5}
\]

where \( p(\theta|z) \) is the posterior density of \( \theta \) given the observation \( z \). Now, the investor’s problem is

\[
\max_\omega \int_R \int_\theta v(W_{T+1}) p(R|z, \theta)p(\theta|z)d\theta dR. \tag{6}
\]

The Bayesian approach requires a prior distribution for the parameters, and generates a posterior distribution according to Bayes’ rule. For a general copula-based model, conjugate families are not available. We have to use numerical method to do the posterior computation when solving the problem (6). As equation (5) shows, to sample from the predictive distribution, we may first use the Bayesian MCMC algorithm developed in the previous section to sample from the posterior distribution \( p(\theta|z) \) and get a long sequence of \( \{\theta^{(j)}\}_{j=1}^N \). Then, we sample from the conditional distribution \( p(R|z, \theta) \) for each simulated parameter value \( \theta^{(j)} \). The resulting sequence of asset returns is a simulation from the predictive distribution.

### 3.2 The Model

By Sklar’s theorem, we may model marginal distributions and dependence separately. Modeling of univariate asset returns has been well developed. However, to keep the problem tractable, we assume the returns are i.i.d. and

\[
\begin{align*}
r_{1,t} &= \mu_1 + \sigma_1 \varepsilon_{1,t}, \\
r_{2,t} &= \mu_2 + \sigma_2 \varepsilon_{2,t},
\end{align*}
\]
where \( r_{1,t} \) and \( r_{2,t} \) are the continuously compounded returns of the two assets over month \( t \), \( (\varepsilon_{1,t}, \varepsilon_{2,t})' \) are i.i.d. over time, and \( (\sigma_1, \sigma_2) \) are standard deviations. We assume that the disturbances follow skew-t distributions. The skewed-t distribution was developed by Hansen (1994). It has zero mean, unit variance, and parameters \( (\nu, \tau) \) that control kurtosis and skewness, respectively. Thus, it is a suitable model for the standardized residuals. If a copula \( C(u, v|\delta) \) is used to combine the marginal processes, the joint distribution is

\[
\begin{pmatrix}
  r_{1,t} \\
  r_{2,t}
\end{pmatrix}
\overset{i.i.d.}{\sim} C\left( G(r_{1,t}|\mu_1, \sigma_1, \nu_1, \tau_1), G(r_{2,t}|\mu_2, \sigma_2, \nu_2, \tau_2)|\delta \right),
\]

where \( G(\cdot|\mu, \sigma, \nu, \tau) \) is the CDF of a skewed-t distribution with mean \( \mu \), standard deviation \( \sigma \) and shape parameters \( (\nu, \tau) \). \( \delta \) is the parameter in the copula function. To model the dependence, we consider a mixture of the Gaussian copula and the Clayton copula. The Bayesian MCMC algorithm for this mixed model is outlined in Appendix B.

### 3.3 Data

We analyze the return of portfolios composed of two market indices: the first and the tenth Stock Capitalization Decile Indices\(^7\). Patton (2004) also uses these indices in his analysis of the importance of asymmetric dependence in portfolio optimization. Other researchers have used them to investigate the time series properties of stock returns, e.g., Perez-Quiro and Timmermann (2001).

The data set in this paper is obtained from the Center for Research in Security Prices (CRSP). We use monthly returns of the two indices from January 1960 to December 1993, yielding 408 observations. These two series represent a high risk - high return asset and a low risk - low return asset respectively, as shown in Table 1. The average monthly returns of the two indices are 1.18% and 0.75%, and their standard deviations are 6.87% and 4.15%. The Jarque-Bera statistics indicate that the hypothesis of normal distributions is rejected by both series. The correlation coefficient indicates a statistically significant contemporaneous dependence between the two indices.

\(^7\) The securities are ranked according to capitalization and then divided into ten equal parts. CRSP Stock File Capitalization Decile Indices are calculated for each of these ten market groups.
Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>1st Decile</th>
<th>10th Decile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.1778</td>
<td>0.7521</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>6.8685</td>
<td>4.1502</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.2072</td>
<td>-0.4288</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>10.3347</td>
<td>5.6152</td>
</tr>
<tr>
<td>Min</td>
<td>-35.7526</td>
<td>-21.7661</td>
</tr>
<tr>
<td>Max</td>
<td>46.0267</td>
<td>16.5937</td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>917.4691</td>
<td>128.7702</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.6571</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Monthly returns of 1st and 10th decile indices.

3.4 Monitoring Convergence of the Gibbs Sampling Algorithm

In this paper, to monitor convergence, we use 10 over-dispersed starting values and run 30,000 iterations. In calculating the K-S statistics, we discard the first 10,000 iterations. To correct the correlation between iterations, in each chain, we choose the first and the eleventh points in every 20 points to get two subsamples, both of which have 1,000 points. Table 2 shows the K-S tests for the simulation of $\mu_1$. The K-S statistics and the corresponding $p$-values indicate that the null hypothesis cannot be rejected. Figure 2 shows the cumulative average and the potential scale reduction factor. The horizontal axis represents iterations from 1 to 30,000 in all three plots. After the first 20,000 iterations, the first plot shows that the cumulative averages from different chains seem to converge to the same value, and the second plot shows that PSRF is very close
to 1. The between-sequence and the within-sequence variances, which are used in calculation of
PSRF, also converge to a constant value. When we examine the K-S statistics and PSRFs for all
the other parameters, we get similar results. Therefore, we conclude that after 30,000 iterations
the simulated sequences seem to converge to the target distribution.

Table 2: Kolmogorov-Smirnov Test

<table>
<thead>
<tr>
<th>Chain 1</th>
<th>Chain 2</th>
<th>Chain 3</th>
<th>Chain 4</th>
<th>Chain 5</th>
<th>Chain 6</th>
<th>Chain 7</th>
<th>Chain 8</th>
<th>Chain 9</th>
<th>Chain 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-S Statistic</td>
<td>0.0330</td>
<td>0.0430</td>
<td>0.0300</td>
<td>0.0290</td>
<td>0.0260</td>
<td>0.0530</td>
<td>0.0330</td>
<td>0.0420</td>
<td>0.0360</td>
</tr>
<tr>
<td>p-value</td>
<td>0.6406</td>
<td>0.3072</td>
<td>0.7530</td>
<td>0.7888</td>
<td>0.8840</td>
<td>0.1168</td>
<td>0.6406</td>
<td>0.3344</td>
<td>0.5288</td>
</tr>
</tbody>
</table>

Figure 2: The cumulative average and the potential scale reduction factor

3.5 Estimation Results and Optimal Portfolio Weights

Since we assume the asset returns are i.i.d. over time, the joint model is relatively simple and we
may use the MLE method to jointly estimate the parameters. We also use the Bayesian MCMC
algorithm to jointly estimate the model. We generate a sample of size 130,000 from the posterior
distribution and use the last 100,000 to make inference. Table (3) presents the results. In the
MLE rows, the numbers in the parentheses are standard errors, while in the Bayesian rows, the
numbers are standard deviations of the posterior distributions.

Table (4) shows the optimal portfolio weights of the two cases: ignoring estimation risk and
considering estimation risk, as illustrated by equations (4) and (6). Under risk aversion levels
$A = 3, 5, 7$, the investor allocates less wealth on the H-H asset after she takes estimation risk into
consideration.
Table 3: *Estimates of the model parameters*

<table>
<thead>
<tr>
<th>Estimation Methods</th>
<th>Marginal Distribution Parameters</th>
<th>Copula Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_1$</td>
<td>$\sigma_1$</td>
</tr>
<tr>
<td>MLE</td>
<td>1.2380</td>
<td>7.1952</td>
</tr>
<tr>
<td></td>
<td>(0.2608)</td>
<td>(0.6744)</td>
</tr>
<tr>
<td>Bayesian</td>
<td>1.1527</td>
<td>6.9247</td>
</tr>
<tr>
<td></td>
<td>(0.3278)</td>
<td>(0.5142)</td>
</tr>
</tbody>
</table>

Table 4: *Optimal Portfolio Weights on High Risk - High Return Asset*

<table>
<thead>
<tr>
<th>Portfolios</th>
<th>A=1</th>
<th>A=3</th>
<th>A=5</th>
<th>A=7</th>
<th>A=9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignoring Estimation Risk</td>
<td>1.0000</td>
<td>0.6469</td>
<td>0.3312</td>
<td>0.1752</td>
<td>0.0655</td>
</tr>
<tr>
<td>Considering Estimation Risk</td>
<td>1.0000</td>
<td>0.5627</td>
<td>0.2919</td>
<td>0.1629</td>
<td>0.0739</td>
</tr>
</tbody>
</table>

4 Conclusion

Recent studies in the empirical financial literature have used copula theory to describe dependence in asset returns; see Embrechts (2001), Hu (2003), etc. Because of the computational challenges introduced by the presence of copula functions, the copula-based models are often estimated by two-stage methods. This paper develops a Bayesian MCMC algorithm to jointly estimate the copula-based models. The use of the algorithm is illustrated by an application to a simple copula-based portfolio optimization problem, where estimation risk is taken into consideration. The results of the application show that for an investor with moderate risk aversions the optimal portfolio weights on risky asset are much less than the case where the estimation risk is ignored.

In the application, we assume the asset returns are i.i.d. over time. However, the overall framework developed in this paper can be applied to more complicated cases, for example, a joint model that uses the GARCH type models to capture marginal processes.
Appendix

A Some Copula Functions

We list the functional forms of the copulas used in this paper.

Gaussian copula

\[
C_{Ga}(u, v; \rho) = \Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v)) \\
= \Phi^{-1}(u) \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp \left\{ -\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)} \right\} dt ds,
\]

\[
c_{Ga}(u, v; \rho) = \frac{1}{(1-\rho^2)^{1/2}} \exp \left\{ -\frac{\Phi^{-1}(u)^2 - 2\rho\Phi^{-1}(u)\Phi^{-1}(v) + \Phi^{-1}(v)^2}{2(1-\rho^2)} + \frac{\Phi^{-1}(u)^2 + \Phi^{-1}(v)^2}{2} \right\},
\]

\[
\rho \in (-1, 1).
\]

Clayton copula

\[
C(u, v; \gamma) = (u^{-\gamma} + v^{-\gamma} - 1)^{-\frac{1}{\gamma}},
\]

\[
c(u, v; \gamma) = (1 + \gamma)[uv]^{-1-\gamma}(u^{-\gamma} + v^{-\gamma} - 1)^{-2-1/\gamma},
\]

\[
C_{2|1}(v|u; \gamma) = [1 + u^{-\gamma}(v^{-\gamma} - 1)]^{-1-1/\gamma},
\]

\[
C_{2|1}^{-1}(q|u; \gamma) = \left[q^{-\gamma/(1+\gamma)} - 1\right] u^{-\gamma} + 1 \right]^{-1/\gamma},
\]

\[
\gamma \in [-1, \infty) \backslash \{0\}.
\]
B Gibbs Sampling from Posterior Distribution of a Copula-based Model

In this section, we outline the Gibbs sampling algorithm to draw from the posterior distribution of the copula-based model, where the marginal distributions are skewed-$t$ and the copula function is a mixture of the Clayton copula and the Gaussian copula. We show the blocks of parameters and how we sample at each block.

To simplify the simulation of the posterior distribution, we introduce a discrete latent variable in the mixture, $w_i$, with the following probability mass function,

$$\Pr(w_i = 1) = \lambda,$$
$$\Pr(w_i = 0) = 1 - \lambda,$$

where $\lambda \in [0, 1]$. By the construction, the joint density of $(r_{1i}, r_{2i}, w_i)$ is

$$h(r_{1i}, r_{2i}, w_i) = \left[ \lambda c_1(u_i, v_i | \delta_1) \right]^{1(w_i=1)} \times \left[ (1 - \lambda) c_2(u_i, v_i | \delta_2) \right]^{1(w_i=0)} \times f(r_{1i} | \theta_1) \times g(r_{2i} | \theta_2)$$

where $u_i = F(r_{1i} | \theta_1), v_i = G(r_{2i} | \theta_2)$. We group the parameters and latent variable into 5 blocks, $\theta_1, \theta_2, \lambda, \delta s$ and $\{w_i\}_{i=1}^n$ and sample each of them conditional on the other blocks and the observations $\{r_{1i}, r_{2i}\}_{i=1}^n$. We use $z$ to denote $\{r_{1i}, r_{2i}, w_i\}_{i=1}^n$.

B.1 Draw $\lambda|\theta_1, \theta_2, \delta, s, z$

Conditional on $\{\theta_1, \theta_2, \delta, z\}, f(\cdot), g(\cdot)$ and $c_j(\cdot), j = 1, 2$ are constant, then we have

$$p(\lambda|\theta_1, \theta_2, \delta, z) \propto \prod_{i=1}^n h(r_{1i}, r_{2i}, w_i) \times p(\lambda)$$
$$\propto \lambda^{\sum 1(w_i=1)} \times (1 - \lambda)^{\sum 1(w_i=0)} p(\lambda),$$

19
where \( p(\lambda) \) is the prior density of \( \lambda \) and it is uniform on \([0, 1]\). Thus, the resulting posterior distribution for \( \lambda \) is \( \text{Beta}(\alpha_n, \beta_n) \), where

\[
\alpha_n = \sum_{i=1}^{n} 1(w_i = 1),
\]

\[
\beta_n = n - \sum_{i=1}^{n} 1(w_i = 1).
\]

To simulate \( \lambda \) from a Beta distribution, first simulate \( x_\alpha \) and \( x_\beta \) from \( \chi^2_{2\alpha} \) and \( \chi^2_{2\beta} \) distributions, respectively, then let

\[
\lambda = \frac{x_\alpha}{x_\alpha + x_\beta}.
\]

**B.2 Draw \( \theta_1 | \lambda, \theta_2, \delta s, z \)**

Conditional on \( \{\lambda, \theta_2, \delta s, z\} \), \( g(\cdot) \) is constant, then we have

\[
p(\theta_1 | \lambda, \theta_2, \delta s, z) \propto \prod_{i=1}^{n} h(r_{1i}, r_{2i}, w_i) \times p(\theta_1 | \theta_2, \lambda s, \delta s),
\]

\[
\propto \prod_{i=1}^{n} \left\{ \left[ c_1(u_i, v_i | \delta_1) \right] \mathbf{1}(w_i = 1) \times \left[ c_2(u_i, v_i | \delta_2) \right] \mathbf{1}(w_i = 0) \times f(r_{1i} | \theta_1) \right\}
\]

\[
\times p(\theta_1),
\]

where \( p(\theta_1) \) is the prior for \( \theta_1 \). We assume the prior of \( \theta_1 \) is independent of \( (\theta_2, \lambda, \delta s) \) and has the following forms,

\[
\mu_1 \sim \mathcal{N}(0, 10)
\]

\[
\sigma^2_1 \sim \mathcal{IG}(3, 2 \cdot \sigma^2_0)
\]

\[
\nu_1 \sim \mathcal{G}(2, 0.25) + 2
\]

\[
\tau_1 \sim \mathcal{U}[-1, 1]
\]

where \( \sigma^2_0 \) is chosen such that the mean of the inverse Gamma distribution is equal to the MLE of \( \sigma^2_1 \). In our study, the estimate is more than 10, thus the prior distribution is very flat.

To simulate \( \theta_1 \) we use the random walk Metropolis-Hastings algorithm, where the random disturbances come from \( t \) distributions whose scale parameters are equal to the standard errors of
MLE. Specifically, for the \((j + 1)\)th iteration, the candidates have the following distributions

\[
\begin{align*}
\mu_1 & \sim t_v(0, s_{\mu_1}) + \mu_1^{(j)}, \\
\sigma_1^2 & \sim 1(\sigma_1^2 > 0) \cdot [t_v(0, s_{\sigma_1^2}) + \sigma_1^{2(j)}], \\
\nu_1 & \sim 1(\nu_1 > 2) \cdot [t_v(0, s_{\nu_1}) + \nu_1^{(j)}], \\
\tau_1 & \sim 1(-1 < \tau_1 < 1) \cdot [t_v(0, s_{\tau_1}) + \tau_1^{(j)}].
\end{align*}
\]

We use the same algorithm to draw from the conditional posterior distributions of \(\theta_2\).

**B.3 Draw \(\delta_1|\theta_1, \theta_2, \delta_2, \lambda, z\)**

Conditional on \(\{\theta_1, \theta_2, \delta_2, \lambda, z\}, f(\cdot), g(\cdot), c_2(\cdot)\), are constant. Then we have

\[
p(\delta_1|\theta_1, \theta_2, \delta_2, \lambda, z) \propto \prod_{i=1}^n h(r_{1i}, r_{2i}, w_i) \times p(\delta_1|\theta_1, \theta_2, \delta_2, \lambda) \propto \prod_{i=1}^N \left\{c_1(u_i, v_i|\delta_1)^{1(w_i=1)} \right\} \times p(\delta_1)
\]

where \(p(\delta_1)\) is the prior for \(\delta_1\). We also use the random walk Metropolis-Hastings algorithm to draw \(\delta_1\). To draw \(\delta_2\), the algorithm is the same but with different prior and proposal distributions.

For the Clayton copula, we assume a Gamma prior, the parameters values \(a_0 = 2, b_0 = 0.25\),

\[
\delta_C \sim G(2, 0.25),
\]

and in the random walk Metropolis-Hastings algorithm, the random disturbance is from \(t\) distribution whose scale parameter is equal to the standard error of the MLE of \(\delta_C\). Specifically, for the \((j + 1)\)th iteration, the candidate has the following distribution,

\[
\delta_C \sim 1(\delta_C > 0) \cdot [t_v(0, s_{\delta_C}) + \delta_C^{(j)}].
\]

For the Gaussian copula, we assume a uniform prior on \((-1,1)\), and in the random walk Metropolis-Hastings algorithm, the random disturbance is from \(t\) distribution whose scale parameter is equal to the standard error of MLE of \(\rho\). Specifically, for the \((j + 1)\)th iteration, the candidate has the
following distribution,
\[ \rho \sim 1(-1 < \rho < 1)[t_u(0, s_\rho) + \rho^{(j)}] \]

B.4 Impute \( w_i | \theta s, \lambda, \delta s, r_1, r_2 \)

Conditional on \( \{ \theta s, \lambda, \delta s, r_1, r_2 \} \), \( f(\cdot), g(\cdot), c_j(\cdot), j = 1, 2 \) are constant, thus

\[
p(w_i | \theta s, \lambda, \delta s, x_i, y_i) \propto h(r_{1i}, r_{2i} | w_i) \times p(w_i | \lambda) \\
\propto [c_1(u_i, v_i | \delta_1) \cdot \lambda]^1(w_i = 1) \times [c_2(u_i, v_i | \delta_2) \cdot (1 - \lambda)]^1(w_i = 0),
\]

the prior of \( w_i \) is adjusted by the copula functions. It is straightforward to draw \( w_i \).
References


