A mixed distribution approach to copula models of portfolio returns

D. J. Miller\textsuperscript{1} & W.-H. Liu\textsuperscript{2}
\textsuperscript{1}Department of Agricultural Economics, Purdue University, USA
\textsuperscript{2}National Defense Management College, National Defense University, Taiwan

Abstract

We propose a mixed distribution procedure for copula models of the return distribution for hedge portfolios. The marginal distributions of the returns for the spot and futures portfolio components are formed as mixed continuous and discrete distributions based on the empirical distribution function and generalized Pareto tail probability models. We then use a bivariate normal copula model to form the joint distribution of correlated spot and futures returns from the mixed marginal distributions. We apply the proposed method to derive the optimal hedge ratio for equity positions on the Taiwan Stock Exchange (TSE) hedged with the Taiwan Futures Exchange (TAIFEX) contract for the TSE stock index.

Keywords: copula, empirical distribution, extreme value, generalized Pareto, heavy tails, lower partial moments, mixed distribution, optimal hedge ratio.

1 Introduction

Given the recent growth of investment and hedging opportunities in developing markets, international investors have an increasing number of tools for reducing business cycle risk. This is important because returns on newly developed equities markets exhibit many of the same stochastic properties as established markets plus higher conditional volatility and conditional probability of large price changes (De Santis and Imrohoroglu [1]). Thus, the problem of determining optimal hedging strategies for stock portfolios in developing markets is especially important. The purpose of this paper is to develop a model of the joint probability distribution of the component returns in a hedge portfolio. We demonstrate the proposed procedure by considering the optimal hedge strategies of an investor on the Taiwan
Stock Exchange (TSE). The value of the spot position is represented by the TSE capitalization weighted stock index (TAIEX), and the futures hedge is based on the TAIEX futures contracts offered by the Taiwan Futures Exchange (TAIFEX).

For purposes of our analysis, we consider the continuously compounded daily returns or log-returns

\[ R^s_t = \ln (S_t) - \ln (S_{t-1}) \quad R^f_t = \ln (F_t) - \ln (F_{t-1}) , \]

based on the daily closing spot price \((S_t)\) of the stock index and the daily closing price \((F_t)\) of the associated futures contract. The daily log-return of a hedged spot-futures portfolio is \(R^h_t = R^s_t + \theta R^f_t\) where the superscripts \(s\) and \(f\) identify 'spot' or 'futures' log-returns, and the superscript \(h\) on \(R^h_t\) indicates log-returns for the 'hedged' portfolio. The parameter \(\theta\) represents the adjusted hedge ratio between the futures and spot positions, and \(\theta = -1\) if the spot position is fully hedged in the futures market.

Researchers in economics and finance use several criteria to select the optimal \(\theta\). One of the most commonly used tools is the minimum variance hedge ratio \(\tilde{\theta} = -\text{Cov} \left( R^s_t, R^f_t \right) \left[ \text{Var} \left( R^f_t \right) \right]^{-1}\),

which is based on the implicit assumptions of normally distributed returns and a hedging objective that protects equally against upside and downside risk. However, the findings of Adams and Montesi [2] imply that market participants may focus more on downside risk. Accordingly, researchers have considered other hedging criteria that account for downside risk, and one of the most widely used methods is the minimum lower partial moment (LPM) approach. Given some target return \(c\) and the distribution of log-returns from the hedged portfolio \(F_h\), the objective of the optimal hedging problem is to choose \(\theta\) to minimize the LPM criterion

\[ \ell(c, \lambda) = \int_{-\infty}^{c} (c - x)^{\lambda} \, dF_h(x) , \]

for some \(\lambda \geq 0\). We denote the optimal hedge ratio as \(\hat{\theta}\), which implicitly depends on \(c\) and \(\lambda\). Effectively, the minimum LPM criterion assigns losses to returns that fall short of the target while disregarding the impact of higher returns. Bawa [3] shows that \(\ell(c, \lambda)\) may be related to the stochastic dominance criterion and that the minimum LPM criterion is equivalent to maximum expected utility under particular utility functions. For \(\lambda = 0\), \(\ell(c, 0) = F(c)\) is the probability of sub-target returns, and this objective is equivalent to Roy’s safety-first criterion. Fishburn [4] shows that \(\lambda < 1\) is consistent with risk-seeking behavior and \(\lambda > 1\) represent risk-averse behavior.

2 A mixed distribution model of portfolio returns

To determine the optimal hedge ratios under the \(\ell(c, \lambda)\) criterion, we need an estimate of \(F_h\). Although a bivariate normal model of \(R^s_t\) and \(R^f_t\) is feasible, the
observed log-returns data for many financial returns series have non-normal characteristics such as leptokurtosis (heavy tails). To model the potentially heavy-tailed marginal distributions of $R^s_t$ ($F^s_s$) and $R^f_t$ ($F^f_f$), we consider parametric extreme value models from the generalized Pareto (GP) family. Following Reiss and Thomas [5], the $\gamma$-parameterization of the GP distribution function with location and scale parameters $\psi$ and $\omega$ is

$$W_{\gamma,\psi,\omega}(x) = 1 - \left(1 - \gamma \left(\frac{x - \psi}{\omega}\right)\right)^{\gamma^{-1}}. \quad (4)$$

The parametric GP family is highly versatile and can be used to represent a wide range of tail probabilities. Further, Balkema and de Haan [6] and by Pickands [7] present limit theorems that prove convergence of the GP family to a large class of distributions.

The maximum likelihood (ML) estimators of the parameters $(\gamma, \psi, \omega)$ for an upper (lower) tail of the log-return distribution are computed from the $k$ largest (smallest) observations. We use the ML estimator described by Prescott and Walden [8], and Davidson and Smith [9] derive conditions under which the ML estimators of $\gamma$ and $\omega$ are $\sqrt{k}$-consistent with asymptotic distribution

$$\left(\hat{\gamma}, \hat{\omega}\right) \sim N \left(\left(\frac{1+\gamma}{k}, \left(\frac{1+\gamma}{\omega}, 2\omega^2\right)\right)\right). \quad (5)$$

To estimate $F^s_s$ and $F^f_f$ based on the estimated GP tail probability models, we follow McNeil and Fry [10] and form $\hat{F}^s_s$ and $\hat{F}^f_f$ as mixed continuous–discrete distributions. In particular, we combine the continuous GP models of the tail regions with the discrete empirical distribution function (EDF) in the center of the distribution. Given the consistency properties of the GP and EDF estimators, we can show that the mixed marginal distributions converge in probability to the underlying population distributions.

We then use a bivariate copula model to form the joint distribution of $R^s_t$ and $R^f_t$ from these fitted marginal distributions. Following Nelsen [11], a $d$-variate copula is a parametric function that maps from the set of $d$ unit intervals back to the unit interval, $C : [0, 1]^d \rightarrow [0, 1]$. We can view this as a distribution function with $d$ dependent U(0,1) marginal variables, and the degree of dependence among the univariate components is controlled by parameters of the copula. To combine $d$ random variables with non-uniform marginal distributions, we can use the probability integral transformation, $U_j = F_j(X_j) \sim U(0,1)$ for each $j = 1, \ldots, d$. The key characteristics of copula families are that dependence among the component random variables may be induced while retaining all features of the marginal distributions. Although we could use bivariate GP models, Reiss and Thomas [5] note that the bivariate GP distribution does not have marginal distributions with GP character. Given that our mixed model is a consistent estimator of the log-return distributions, we prefer to use the copula approach to retain these marginal characteristic in the joint distribution.
There are several prominent families of copula models, and we use the bivariate normal copula
\[ C(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \Sigma) , \tag{6} \]
in this paper. The uniform components are constructed under the probability integral transformation from the marginal CDF’s, \( U_s^s = F_s(R_s^t) \) and \( U_f^f = F_f(R_f^t) \). The covariance matrix \( \Sigma \) has unitary diagonals and \( \varphi \) on the off-diagonals, and the degree of dependence among \( R_s^t \) and \( R_f^t \) is controlled by \( \varphi \). A consistent estimator of \( \varphi \) is the sample covariance of the transformed observations
\[ \hat{\varphi} = n^{-1} \sum_t \Phi^{-1}(\hat{U}_s^t) \Phi^{-1}(\hat{U}_f^t) , \tag{7} \]
where \( \hat{U}_s^t = \hat{F}_s(R_s^t) \) and \( \hat{U}_f^t = \hat{F}_f(R_f^t) \) are asymptotically \( U(0, 1) \).

3 Optimal hedging of a TSE equity position

The spot prices \( (S_t) \) are daily closing observations for TAIEX, and the futures prices \( (F_t) \) are daily closing observations for the nearby TAIEX futures contract traded on TAIFEX. The sample period covers \( n = 721 \) trading days from 21 July 1998 to 3 April 2001. Although weekend effects are a common concern in the analysis of financial data, TSE and TAIFEX allow Saturday trading so there may be only one inactive day during some weekends. For this reason, we ignore the possible impact of weekend effects. Further, there were four extended trading holidays in the sample period (11–19 February 1999, 21–26 September 1999, 2–8 February 2000, and 19–28 January 2001). Due to the potential impact of these holiday periods, we allocate the holiday returns equally across the period.

A data summary is provided in table 1. The sample means of the log-returns are not statistically different from zero, and the log-returns exhibit strong positive correlation \( \hat{\rho}_{s,f} = 0.8881 \). We also evaluate the skewness and kurtosis of \( R_s^t \) and \( R_f^t \) using the standardized estimators
\[ \hat{\kappa}_1 = \sqrt{\frac{n}{6} \left( \frac{\hat{\mu}_3}{\hat{\sigma}^3} \right)} \sim N(0, 1) \quad \hat{\kappa}_2 = \sqrt{\frac{n}{24} \left( \frac{\hat{\mu}_4}{\hat{\sigma}^4} - 3 \right)} \sim N(0, 1) , \tag{8} \]
where \( \hat{\mu}_j = n^{-1} \sum_{t=1}^n (R_t)^j \) for \( j = 3 \) and \( j = 4 \). The observed outcomes are reported with the p-values for the associated Z-tests. Clearly, the distributions of \( R_s^t \) and \( R_f^t \) are not significantly skewed but do have strong leptokurtic character. Accordingly, we reject the normality hypotheses at all reasonable \( \alpha \) levels.

Based on the observed sample autocorrelation functions (ACF) for the log-return series, \( R_s^t \) and \( R_f^t \), we cannot reject the joint null hypothesis of stationarity for either return series. To check for autoregressive conditional heteroscedasticity (ARCH), we conduct the LM ARCH test and strongly reject the null hypothesis of no ARCH effects in \( R_s^t \) and \( R_f^t \). We also compute the ACF for the squared log-return series, \( (R_s^t)^2 \) and \( (R_f^t)^2 \) and find significant autocorrelation up to lag \( i = 4 \).
Table 1: Data summary (p-values for asymptotic Z-tests in parentheses).

<table>
<thead>
<tr>
<th>Series</th>
<th>Min</th>
<th>Max</th>
<th>Average $\hat{\mu}$</th>
<th>Std. Dev. $\hat{\sigma}$</th>
<th>Skewness $\hat{\kappa}_1$</th>
<th>Kurtosis $\hat{\kappa}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spot $R^s_t$</td>
<td>-0.0993</td>
<td>0.1061</td>
<td>-0.000524 (0.9784)</td>
<td>0.01931</td>
<td>0.0893 (0.9331)</td>
<td>14.7801 (0.0000)</td>
</tr>
<tr>
<td>Futures $R^f_t$</td>
<td>-0.0814</td>
<td>0.0676</td>
<td>-0.000535 (0.9799)</td>
<td>0.02120</td>
<td>-1.2450 (0.2131)</td>
<td>8.9985 (0.0000)</td>
</tr>
</tbody>
</table>

for $(R^2_t)$ and up to lag $i = 8$ for $(R^f_t)^2$, which is also evidence of ARCH-type effects.

Given that the heavy-tailed character of $R^s_t$ and $R^f_t$ may be due to the ARCH effects, we estimate ARCH models of the log-returns

$$R_t = \mu_t + \varepsilon_t = \phi_0 + \phi_1 R_{t-1} + \varepsilon_t,$$

where $\varepsilon_t = \sigma_t v_t$ and $v_t$ is a white noise process with unit variance. The conditional mean $\mu_t$ is AR(1), and the conditional variance $\sigma^2_t$ is specified as an exponential generalized ARCH (EGARCH) process

$$\ln(\sigma^2_t) = \delta_0 + \sum_{i=1}^p \delta_i \ln(\sigma^2_{t-i}) + \sum_{j=1}^q \beta_j \left[|v_{t-j}| - E|v_{t-j}| + \eta v_{t-j}\right].$$

Under EGARCH, the current conditional variance is exponentially related to the previous conditional variances. The parameter $\eta$ allows for asymmetric response in the conditional variance to deviations of $|v_t|$ from $E|v_t|$. The marginal change in $\ln(\sigma^2_t)$ with respect to $v_{t-j}$ is $\beta_j (1 + \eta)$ if $v_{t-j} > 0$ and is $\beta_j (1 - \eta)$ if $v_{t-j} < 0$.

The EGARCH models of $R^s_t$ and $R^f_t$ were estimated by the normal ML procedure in SAS. The choices of $p$ and $q$ were based on the Akaike Information Criterion (AIC), and the resulting models are AR(1)–EGARCH(4,3) for $R^s_t$ and AR(1)–EGARCH(2,2) for $R^f_t$. In both fitted models, nearly all of the EGARCH parameters are statistically significant. The intercept parameter $\beta_0$ is not significantly different from zero in either model, which is consistent with the mean-zero character of $R^s_t$ and $R^f_t$ noted in table 1. To examine the tail character of the AR–EGARCH noise term, we compute the standardized residuals

$$\hat{v}_t = (R_t - \hat{\mu}_t) / \hat{\sigma}_t,$$

where $\hat{\mu}_t = \hat{\phi}_0 + \hat{\phi}_1 R_{t-1}$. Based on normal QQ plots of $\hat{v}_t$ for $R^s_t$ and $R^f_t$, both AR–EGARCH noise processes are strongly leptokurtotic relative to the normal distribution, even after we have accounted for the presence of conditional heteroscedasticity. We also compute the ACF of $\hat{v}_t$ and $\hat{v}^2_t$ for $R^s_t$ and $R^f_t$, and
Table 2: Estimated generalized Pareto (GP) models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \gamma )</th>
<th>( \omega )</th>
<th>( \psi )</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Returns — Upper Tail</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^s_t ) (k = 200)</td>
<td>-0.0572</td>
<td>0.01607</td>
<td>-0.01203</td>
<td>0.00773</td>
</tr>
<tr>
<td>( R^f_t ) (k = 200)</td>
<td>-0.1495</td>
<td>0.0217</td>
<td>-0.01692</td>
<td>0.00838</td>
</tr>
<tr>
<td>Log-Returns — Lower Tail</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^s_t ) (k = 250)</td>
<td>-0.0242</td>
<td>0.01358</td>
<td>0.00761</td>
<td>-0.00659</td>
</tr>
<tr>
<td>( R^f_t ) (k = 150)</td>
<td>-0.0350</td>
<td>0.01613</td>
<td>0.01064</td>
<td>-0.01400</td>
</tr>
</tbody>
</table>

The residuals appear to be serially uncorrelated with no evidence of ARCH-type effects. Thus, the AR–EGARCH models appear to be well specified, but the noise processes are heavy-tailed. To evaluate time-varying hedge strategies, we could estimate the AR–EGARCH models with the non-normal estimation procedures described by Hamilton [12] and by Mittnik, Paolella, and Rachev [13] and use the residuals to form our mixed distribution estimates of the conditional marginal distributions of \( R^s_t \) and \( R^f_t \). However, the objective of this paper is to determine unconditional optimal hedge ratios, and we use the fitted mixed distribution models for \( R^s_t \) and \( R^f_t \) for our analysis.

The estimated GP parameters are presented in table 2, and the estimated asymptotic standard errors of \( \hat{\gamma} \) and \( \hat{\omega} \) appear in parentheses. The reciprocal \( \alpha = \gamma^{-1} \) is known as the tail index and is a measure of heavy-tailed character. Note that the estimated tail indices for the upper and lower tails are clearly different, which implies the marginal distributions are asymmetric. To graphically evaluate the goodness-of-fit of the estimated GP models, we present tail probability plots in fig 1. Note that the lower tail plots are formed from the negative observations (e.g., \(-R^s_t\)), so the log-scaled outcomes of \( 1 - F(x) \) appearing on the left of fig 1 represent the log-scaled outcomes of \( F(x) \). The solid lines are the probabilities of extreme outcomes under the fitted GP models, and the circles represent the probabilities assigned to the observed sample under the empirical distribution function. We find that the estimated GP models provided a very close fit to the observed log-returns.

The fitted bivariate normal copula model of the joint distribution of \( R^s_t \) and \( R^f_t \) is derived from the sample correlation of \( \tilde{U}^s_t \) and \( \tilde{U}^f_t \), \( \varphi = 0.7139 \). We then solve the minimum LPM problem by Monte Carlo integration using \( m = 7500 \) pseudo-random draws \((Z^s_t, Z^f_t)\) from the fitted bivariate normal copula model. The nor-
normal draws are converted to represent log-return outcomes under the fitted quantile functions of $R_s^t$ and $R_f^t$. For example, an outcome for the spot log-return is generated as $\hat{F}_s^{-1}(\Phi(Z_s))$. To guarantee that the generated log-return outcomes mimic the observed sample, we recenter and scale the Monte Carlo draws. We minimize $\ell(c, \lambda)$ with $c \in \{-0.015, -0.0125, \ldots, 0.015\}$ and $\lambda \in \{1, 2, 3\}$. For each candidate $\theta, c,$ and $\lambda,$ we form the set of $m$ log-returns for the hedged portfolio ($R^h_t$).
and evaluate $\ell(c, \lambda)$. The optimal hedge ratio is selected for each $c$ and $\lambda$ by grid search over the set of candidate $\theta$ values.

The estimated optimal hedge ratios are plotted in the upper half of fig 2. As expected, the optimal hedge ratio approaches $\theta = -1$ as the target return ($c$) increases. That is, hedgers increase their protection of the spot position as they become sensitive to losses across a broader range of sub-target returns. In most
cases, the minimum LPM hedge strategies employ a smaller share of futures contracts than the minimum variance hedge strategy (indicated by the horizontal line at $\tilde{\theta} = -0.8808$). Only the 'risk neutral' hedgers ($\lambda = 1$) with high target returns ($c \geq 0.007$) have optimal strategies that exceed the minimum variance hedge ratio. Note that the relationship between $\tilde{\theta}$ and $\lambda$ seems counter-intuitive — the optimal hedge ratios increase from $\theta = -1$ as the risk aversion coefficient ($\lambda$) increases, which is contrary to the theoretical behavior of risk averse hedgers under a maximum expected utility criterion Holthausen [14]. Similar patterns in the optimal hedge ratios have been reported by Lien and Tse [15] and Chen, Lee and Shrestha [16], and Lien and Tse [17] show that the result is true.

For each $(c, \lambda)$ scenario, we also compute the hedging effectiveness score

$$HE(c, \lambda) = 1 - \frac{\ell(c, \lambda)|_{\theta=\tilde{\theta}}}{\ell(c, \lambda)|_{\theta=0}},$$

(12)

as a measure of optimal hedging performance. The score indicates the percentage reduction in risk (expected loss) of the optimal strategy versus the unhedged portfolio ($\theta = 0$) under the lower partial moment criterion. Clearly, minimizing $\ell(c, \lambda)$ is equivalent to maximizing $HE$ by choice of $\theta$, so the hedging effectiveness score may be more appropriately used for out-of-sample evaluations (Chen, Lee, and Shrestha [16]). However, we find that $HE$ is also a useful diagnostic tool for describing our results. The hedging effectiveness scores for the optimal hedge strategies are plotted in the lower half of fig 2. Overall, hedging effectiveness is roughly constant for $c \leq 0$ and declines as the target return increases above zero. Further, the relative risk reduction improves as the degree of risk aversion increases.

References


