Density and Risk Forecast of Financial Returns Using Decomposition and Maximum Entropy

Tae-Hwy Lee,* Zhou Xi,† and Ru Zhang‡
Department of Economics
University of California, Riverside

March 2014 (preliminary)

Abstract

In this paper we consider a multiplicative decomposition of the financial returns to improve the density forecasts of financial returns. The multiplicative decomposition is based on the identity that financial return is the product of its absolute value and its sign. Advantages of modeling the two components are discussed. To reduce the effect of the estimation error in the decomposition density forecast model, we impose a moment constraint that the conditional mean forecast is set to match with the martingale difference model or the constant mean model. Imposing such a simple moment constraint operates a shrinkage and tilts the density forecast of the decomposition model to produce the maximum entropy density forecast. We also show that the risk forecasts produced from the density forecast using the decomposition and maximum entropy is superior to Riskmetrics and the normal density forecast especially in extreme tail events of large loss.

Key words: decomposition, copula, moment constraint, shrinkage, maximum entropy, density forecast, MCMC, Value-at-Risk, out-of-sample prediction.

*E-mail: taelee@ucr.edu.
†E-mail: zhou.xi@email.ucr.edu.
‡E-mail: ru.zhang@email.ucr.edu.
1 Introduction

In this paper the density forecast model is based on a multiplicative decomposition of a financial return series into a product of the absolute return and the sign of the return. The joint density forecast is constructed from the margins of the two components and their copula function. A copula function incorporates the possible dependence between the absolute return and the sign of the return. It is well documented empirically that each of the two components are easier to predict than the return. That does not necessarily imply we can predict the financial returns in the conditional mean using the decomposition. Our interest is not forecasting the mean returns, but forecasting the conditional density of financial returns using the decomposition. A better density forecast model can produce a better risk forecast in terms of Value-at-Risk (VaR) forecast. For example, in financial risk management, we are more interested in the density (especially in tails) than the mean, and the density forecast can be better predicted using the decomposition model. While the decomposition allows us to model two components in much richer specifications, the multiplication of the two components to produce the return makes the conditional moment forecasts be subjected to multiplicative estimation errors which are of higher order of magnitude. In other words, while disaggregation by decomposition provides richer information (signal), the multiplication amplifies the magnitude order of the estimation errors (noise). To control the latter would improve the density forecast from the decomposition model.

To improve the density forecast we consider imposing some sensible moment constraints. A trick is to find the maximum entropy density that satisfies such moment constraints (particularly in the conditional mean). Noting that a simple mean forecast model such as zero mean (ZM) and constant mean (or historical mean, HM) will give less estimation error than the more complex decomposition model, we use the maximum entropy principle for the out-of-sample density forecast subject to the moment constraint that the mean forecasts from the decomposition model and the simple model are equal. When the mean forecast from the decomposition model deviates from the simple models, imposing the conditional mean constraint tilts the density forecast of the decomposition model to produce the maximum
entropy density forecast. The tilted density forecast would improve over the original density forecast of the decomposition model if the constraint is correct. The underlying reason for the benefit of imposing the constraint is the “shrinkage” principle, which we show how it works in the density forecast.

Traditionally econometric modeling has been focused on the conditional moments of variables of interest, particularly mean and variance. Some recent research has shifted from the conditional moments to the conditional density. The common density forecast models for financial returns assume a particular distribution such as Gaussian, Student $t$, Log-Normal, Generalized Pareto distributions or some variants to capture fat-tails or skewness. In particular, Granger and Ding (1995a, 1995b, 1996) and Rydén, Teräsvirta and Åsbrink (1998) provide several stylized facts about the financial returns. Let $r_t$ be the return on a financial asset at time $t$, $|r_t|$ denote the absolute value, and $\text{sign}(r_t) = 1 (r_t > 0) - 1 (r_t < 0)$. The following three distributional properties (DP) have been stylized in these papers.

DP1: $|r_t|$ and $\text{sign}(r_t)$ are independent.

DP2: $|r_t|$ has the same mean and standard deviation.

DP3: The marginal distribution for $|r_t|$ is exponential.

The multiplicative decomposition, $r_{t+1} = |r_{t+1}| \times \text{sign}(r_{t+1})$, treats the two components $|r_{t+1}|$ and $\text{sign}(r_{t+1})$ separately for their marginal densities and then links them by a copula to obtain the joint density of $|r_{t+1}|$ and $\text{sign}(r_{t+1})$. A goal is to obtain the one-month ahead return density forecast $f_t (|r_{t+1}|, \text{sign}(r_{t+1}))$. While the conditional mean of the return $r_{t+1}$ may be a martingale difference, the conditional means of the margins for $|r_t|$ and $\text{sign}(r_t)$ are not martingale difference. That is, the conditional means of $|r_{t+1}|$ and $\text{sign}(r_{t+1})$ can be dynamically modelled unlike that of the returns $r_t$. Ding, Granger, and Engle (1993) show that $|r_{t+1}|$ is easily predictable, while Korkie, Sivakumar, and Turtle (2002) and Christoffersen and Diebold (2006) show that $\text{sign}(r_{t+1})$ is predictable as well. Let $I_t$ be the information set at time $t$. If the indicator series $1 (r_{t+1} < 0)$ displays conditional mean serial dependence, namely, if $E [1 (r_{t+1} < 0) | I_t]$ is a nonconstant function of $I_t$, then
the signs can be predicted. Further, let \( \mu_{t+1} = E(r_{t+1}|I_t) \) be the conditional mean and \( \sigma_{t+1}^2 = E[(r_{t+1} - \mu_{t+1})^2|I_t] \) be the conditional variance. Then Christoffersen and Diebold (2006) note that

\[
E[1(r_{t+1} < 0)|I_t] = \Pr(r_{t+1} < 0|I_t) = \Pr\left(\frac{r_{t+1} - \mu_{t+1}}{\sigma_{t+1}} < \frac{-\mu_{t+1}}{\sigma_{t+1}} | I_t\right) = F_t\left(\frac{-\mu_{t+1}}{\sigma_{t+1}}\right),
\]

which shows the sign is predictable if \( \sigma_{t+1} \) is predictable and \( \mu_{t+1} \) is not zero. Using a series expansion of the conditional distribution \( F_t(\cdot) \), such as the Gram-Charlier expansion or the Edgeworth expansion, it can be seen that \( 1(r_{t+1} < 0) \) can be predictable if the conditional higher moments (skewness and kurtosis) are predictable. Because the absolute return and the sign of the stock return are predictable, the margin density forecast models of \( |r_t| \) and \( \text{sign}(r_t) \) can be specified such that these serial dependence properties associated with the predictability are incorporated. It can eventually yield a more precise density forecast model for \( r_{t+1} \).

Some studies support DP1 that the sign and absolute value of the return are independent. However, it seems that the evidence of financial returns exhibiting negative conditional skewness indicates the possibility that the sign and the absolute return are not independent. As a result, we consider both cases in this paper. When the dependence is weak, shrinkage toward the independence can benefit the density forecasts. In that case we use independent copula for the joint density of the sign of return and absolute value of the return. Comparing a dependent copula with the independent copula would form a case of comparing nested density forecast models whose formal test statistics can be developed in line with Amisano and Giacomini (2007), Bao, Lee, and Saltoğlu (2007), and Granziera, Hubrich, and Moon (2013). We test for the independence with applying Hong and Li (2005) to the copula density function.

The decomposition model may not generate satisfactory moment predictions even for the first moment (mean). The mean of density forecast from the decomposition model may be unreasonably far from zero which contradicts the fact that stock returns are close to zero in mean especially in high frequency. Second, in our empirical applications we find that when applying the decomposition model to the annual returns at monthly frequency (i.e.,
the returns over 12 months, January to January, February to February, etc.), the predicted conditional mean fluctuates rather excessively, often deviating unreasonably too far from the historical mean (HM) predictions. The plot of HMs of the annual returns at monthly frequency over rolling windows (shifting one month forward with dropping the oldest month in the estimation window) exhibits quite smooth and stable values around a constant near zero. We think this would make a classic environment to apply the shrinkage principal by imposing a moment constraint. Since our criterion function will be the logarithmic scoring rule in evaluating the density forecast, the improvement can be achieved by solving the constrained maximum entropy problem. Jaynes (1968) provides the solution for a discrete density. A general solution for any distribution can be seen in Csiszár (1975). Robertson, Tallman, and Whiteman (2005) and Giacomini and Ragusa (2013) provide excellent applications of the constrained maximum entropy problem to macroeconomic models. Encouraged from these studies we consider the shrinkage of imposing a smooth moment constraint when the decomposition model produces unstable mean forecasts. Indeed, we find that imposing a sensible moment (mean) constraint the decomposition model could improve the density forecast of the financial returns $r_{t+1}$.

A major issue in density forecasting is to make better risk forecasts. If the density forecast can be improved from using the decomposition and the maximum entropy, then the improvement should come from some parts of the support of the density. We know it is not from the middle part of the support as we have replaced the noisy mean forecast by the simpler ZM or HM. It has to be some other parts of the support. If it is from the tails, then it can produce better risk forecasts. To be specific, we can use the new density forecast to compute improved risk forecasts in tail quantiles (VaR) or risk spectrum which is the weighted average of return quantiles with weighs reflecting different risk aversion; we can also apply density forecasts to calculate the expected short fall, loss given default, unexpected loss, tail Value-at-Risk, or expected default frequency, etc. In this paper, we will focus on the VaR forecasts produced from inverting different density (distribution) forecast models.

The rest of the paper is structured as follows. Section 2 introduces the benchmark normal
density forecast model and the decomposition density forecast model which is obtained from
the margins of the absolute return and the sign of the return together with their copula func-
tion. In Section 3, we consider the maximum entropy decomposition density forecast with
a mean moment constraint imposed. Section 4 discusses the scoring rules to evaluate the
density forecast models. Section 5 explains how to make risk forecast using the density fore-
cast and how to evaluate risk forecasts. Section 6 includes empirical results which compare
the benchmark density forecasts, the decomposition density forecast and the decomposition
density forecasts with moment constraints, as well as the risk forecasts from these density
forecast models. Section 7 offers some concluding remarks.

2 Density Forecast Models

2.1 Normal Density Forecast Model

A benchmark density forecast of \( r_{t+1} \) at time \( t \) is the normal density forecast model
\[
    f_{t+1}(r) = \frac{1}{\sqrt{2\pi \sigma^2_{t+1}}} \exp \left\{ -\frac{(r - \mu_{t+1})^2}{2\sigma^2_{t+1}} \right\}, \quad r \in (-\infty, +\infty), \tag{2}
\]
where \( \mu_{t+1} := E(r_{t+1}|I_t) \) and \( \sigma^2_{t+1} := E[(r_{t+1} - \mu_{t+1})^2|I_t] = \gamma_0 + \gamma_1 (r_t - \mu_t)^2 + \gamma_2 \sigma^2_t \). For
the conditional mean specification, it is common to assume that \( \mu_{t+1} \) is zero mean (ZM) or
equal to the historical mean (HM). According to the weak form of efficient market hypothesis,
the best forecast of the conditional mean is ZM. Consider a linear predictive regression for
the conditional mean
\[
    r_{t+1} = \alpha + x'_t \beta + \varepsilon_{t+1}. \tag{3}
\]
If \( \alpha = 0 \) and \( \beta = 0 \), then \( \mu_{t+1} = 0 \) (zero mean, ZM). If \( \beta = 0 \), then \( \mu_{t+1} = \alpha \), which is
estimated by HM at time \( t \), \( \hat{\mu}_{t+1} = \bar{r}_t = \frac{1}{R} \sum_{s=t-R+1}^t r_s \), using the rolling estimation window
of \( R \) observations. In general we can use a set of covariates to forecast the conditional mean,
where \( \mu_{t+1} = \alpha + x'_t \beta \) as in Goyal and Welch (2008). Goyal and Welch (2008) find that none
of their 17 predictors can make a better mean forecast than HM. Their results demonstrate
that it is very difficult to outperform the historical mean specification. We confirmed that the
conditional mean forecast using a covariate performs worse than simpler model of ZM or HM
in the density forecast. Therefore in this paper we do not consider using covariates for the
density forecast and set $\beta = 0$. The conditional variance forecast $\hat{\sigma}_{t+1}^2 = \gamma_0 + \gamma_1 \epsilon_t^2 + \gamma_2 \epsilon_t^2$ is the predicted according to the GARCH(1,1) model using the estimated parameter values at time $t$. This normal density forecast model will be refereed as Model 1 or M1. In particular, M1 with $\mu_{t+1} = 0$ is labeled as M1-ZM, and M1 with $\mu_{t+1} = r_t$ is labeled as M1-HM.

2.2 The Independent Decomposition Model

Since the pioneering work by Granger and Ding (1995a, 1995b), the decomposition model for financial returns has been studied in many papers, such as Korkie, Sivakumar, and Turtle (2002), Rydberg and Shephard (2003) and Anatolyev and Gospodinov (2010), among others. The financial return $r_{t+1}$ can be decomposed as the product of its sign and absolute value:

$$r_{t+1} - c = |r_{t+1} - c| \times \text{sign}(r_{t+1} - c) =: U_{t+1} V_{t+1},$$

where $U_{t+1} = |r_{t+1} - c|$ and $V_{t+1} = \text{sign}(r_{t+1} - c)$ with a constant $c$. In this paper, we set $c = 0$.

To get the density forecast of $r_{t+1}$, several assumption about the joint and marginal distribution of $U_{t+1}$ and $V_{t+1}$ are needed. Just for simplicity, let us first assume DP1 of Rydén, Teräsvirta and Åsbrink (1998) that the absolute return $U_{t+1}$ and the sign of the return $V_{t+1}$ are independent. Then the joint density of $U_{t+1}$ and $V_{t+1}$ can be written as:

$$f_{t+1}(u,v) = f_{t+1}^U(u) \times f_{t+1}^V(v).$$

The marginal density of the absolute return $U_{t+1}$ takes the positive support (like duration), and we take a model similar to the autoregressive duration (ACD) model of Engle and Russell (1998), that is,

$$U_{t+1} = \psi_{t+1} e_{t+1},$$

where $\psi_{t+1} = E(U_{t+1}|I_t)$ is the conditional mean of the absolute return and $\{e_{t+1}\}$ is an i.i.d. positive random variable with $E(e_{t+1}|I_t) = 1$. To model the density of $e_{t+1}$, Engle and Russell (1998) consider exponential and Weibull distributions, Grammig and Maurer (2000) consider the Burr distribution, and Lunde (1999) proposes a generalized gamma distribution. Both of the Burr and gamma distribution nest the exponential distribution. Based on the stylized
facts DP2 and DP3 of the absolute returns, we consider only the exponential distribution. Further, unlike duration, the absolute return has a density function with strictly decreasing in $u$. We observe from the data that absolute return has a strictly decreasing density, yet if we use more complicated distributions, they cannot guarantee this property. In our empirical experiments, we also find that the Weibull distribution gives a much worse result.

The conditional mean $\psi_{t+1}$ is modeled using an ACD-like model:

$$
\psi_{t+1} = \delta_0 + \delta_1 U_t + \delta_2 \psi_t,
$$

(7)

While other (nonlinear) specifications such as a logarithmic model of Bauwens and Giot (2000) and a threshold model of Zhang, Russell and Tsay (2001) are possible, the above simple linear model is sufficient and a higher order specification is not necessary to make the density forecast more accurate. The (conditional) marginal density forecast of $U_{t+1}|I_t$ is then an exponential distribution with the mean equal to the conditional mean forecast $\psi_{t+1}$ from the above ACD like linear model. That is,

$$
f_{U_{t+1}}(u) = \frac{1}{\psi_{t+1}} \exp \left( - \frac{1}{\psi_{t+1}} u \right), \quad u > 0.
$$

(8)

Once we get $\psi_{t+1}$, the density forecast of the absolute return will be $f_{U_{t+1}}(u) = \frac{1}{\psi_{t+1}} \exp \left( - \frac{1}{\psi_{t+1}} u \right)$.

Next, the marginal density of the sign of the return $V_{t+1} = 1 (r_{t+1} \geq 0) - 1 (r_{t+1} < 0)$ can be modeled using a Bernoulli-type density since the event is binary. Let $v_{t+1} = 1$ when the sign of the actual stock return at $t + 1$ is positive and otherwise $v_{t+1} = -1$. Then the sign forecast density function can be written as:

$$
f_{V_{t+1}}(v) = \begin{cases} 
    p_{t+1} & \text{if } v = 1 \\
    1 - p_{t+1} & \text{if } v = -1
\end{cases}
$$

(9)

where $p_{t+1} := \Pr (V_{t+1} = 1|I_t)$.

To predict $p_{t+1}$, the simplest way is to use the historical percentage of positive returns, that is, $\hat{p}_{t+1} = \frac{1}{R} \sum_{s=-R+1}^{t} \mathbf{1} (r_s > 0)$. This is a special case of the generalized linear model (GLM), in which $p_{t+1} = G (a + x'_t b)$, where $G(\cdot)$ is a link function. More complicated models can be used to estimate $p_{t+1}$. For instance, If $G(\cdot)$ is the identity function, we
have the ordinary least square estimator. $G(\cdot)$ can be also the standard normal cumulative distribution function (CDF) or the logistic function. In these two cases, we have the probit model or the logit model. However, these more complicated models may not work better due to the parameter estimation uncertainty. So in this paper, we only consider the historical percentage estimator of $p_{t+1}$. Once we get $\hat{p}_{t+1}$, the density forecast of the sign of the return will be $f_{t+1}^V(v) = \frac{\hat{p}_{t+1}}{1 - \hat{p}_{t+1}} \frac{1}{\pi v}$.

The joint density forecast $f_{t+1}^{UV}(u,v)$ of $U_{t+1}$ and $V_{t+1}$ will be called the decomposition model, and will be referred to as Model 2 or M2. This decomposition model assuming independence between $U_{t+1}$ and $V_{t+1}$ will be referred to as M2-I, which is a special case of the decomposition model incorporating dependent copula functions in the next section.

### 2.3 Decomposition Model using Copulas

#### 2.3.1 Testing for Independence

Although stated in DP1 as one of the stylized facts in some studies, a model which incorporates the possible dependence between the absolute return $U_{t+1}$ and the sign $V_{t+1}$ may improve the density forecast of the return. We begin by testing for DP1. Under DP1, the joint density is equal to the product of their marginal densities:

$$H_0^1 : f_{t+1}^{UV}(u,v) = f_{t+1}^U(u) \times f_{t+1}^V(v).$$

Since the joint density function of $U_{t+1}$ and $V_{t+1}$ can be written in their margins and the conditional copula function

$$f_{t+1}^{UV}(u,v) = f_{t+1}^U(u) \times f_{t+1}^V(v) \times c\left(F_{t+1}^U(u), F_{t+1}^V(v)\right),$$

where $c\left(F_{t+1}^U(u), F_{t+1}^V(v)\right)$ is the conditional copula density function such that $c(w_1, w_2) = \frac{\partial^2 C(w_1, w_2)}{\partial w_1 \partial w_2}$ and $C(\cdot, \cdot)$ is the conditional copula function. The copula is defined as a CDF function of $F_{t+1}^U(u)$ and $F_{t+1}^V(v)$ such that

$$F_{t+1}^{UV}(u,v) = C(F_{t+1}^U(u), F_{t+1}^V(v)),$$

where $F_{t+1}^U(u) = \Pr(U \leq u | I_t)$, $F_{t+1}^V(v) = \Pr(V \leq v | I_t)$. Moreover, the conditional copula function can be rewritten as:

$$C(w_1, w_2) = F_{t+1}^{UV}\left((F_{t+1}^U)^{-1}(w_1), (F_{t+1}^V)^{-1}(w_2)\right)$$
where \( w_1 = F_{t+1}^U(u) \) and \( w_2 = F_{t+1}^V(v) \), while \((F_{t+1}^U)^{-1}(w_1)\) and \((F_{t+1}^V)^{-1}(w_2)\) denote the inverse functions of the CDFs of \( U_{t+1} \) and \( V_{t+1} \).

From (10) and (11), to test the null hypothesis that the sign and absolute value of the return are independent, we just need to test that the conditional copula function is equal to 1

\[
H_0^2 : c \left( F_{t+1}^U(u), F_{t+1}^V(v) \right) = 1. \tag{14}
\]

We use Hong and Li (2005), modified for the copula density as in Lee and Yang (2013). We estimate \( c \left( F_{t+1}^U(u), F_{t+1}^V(v) \right) \) from a nonparametric predictive copula density:

\[
\hat{c}_p(F_{t+1}^U(u), F_{t+1}^V(v)) = \frac{1}{P} \sum_{t=R}^{T-1} K_h(w_1, \hat{w}_{1,t+1})K_h(w_2, \hat{w}_{2,t+1}), \tag{15}
\]

where \( w_1 = F_{t+1}^U(u), w_2 = F_{t+1}^V(v), K_h(\cdot) \) is a kernel function, and

\[
\hat{w}_{1,t+1} = \hat{F}_{t+1}^U(u_{t+1}) = \frac{1}{R} \sum_{s=t-R+1}^{t} 1(u_s \leq u_{t+1}), \tag{16}
\]

\[
\hat{w}_{2,t+1} = \hat{F}_{t+1}^V(v_{t+1}) = \frac{1}{R} \sum_{s=t-R+1}^{t} 1(v_s \leq v_{t+1}), \tag{17}
\]

are the marginal empirical distribution functions (EDF) estimated using the rolling windows of the most recent \( R \) observations at each time \( t (t = R, \ldots, T - 1) \). As \( w_1, w_2 \in [0, 1] \) are probabilities, the boundary-modified kernel of Rice (1984) is used following Hong and Li (2005):

\[
K_h(a, a') = \begin{cases} 
    h^{-1}k \left( \frac{a - a'}{h} \right) \int_{(a/h)}^{1} k(u)du & \text{if } a \in [0, h) \\
    h^{-1}k \left( \frac{a - a'}{h} \right) & \text{if } a \in [h, 1 - h] \\
    h^{-1}k \left( \frac{a - a'}{h} \right) \int_{1-(1-a)/h}^{1} k(u)du & \text{if } a \in (1 - h, 1] 
\end{cases} \tag{18}
\]

where \( k(\cdot) \) is a symmetric kernel function and \( h \) is the bandwidth such that \( h \to 0, nh \to \infty \) as \( n \to \infty \). We will use the quadratic kernel function \( k(z) = \frac{15}{16} (1 - u^2)^2 1(|z| \leq 1) \).

Then the test statistic for the null hypothesis \( H_0^2 \) in (14), based on a quadratic form, is given by:

\[
\hat{M}_P = \int_{0}^{1} \int_{0}^{1} \left[ \hat{c}_p(w_1, w_2) - 1 \right]^2 dw_1 dw_2, \tag{19}
\]

which can be pivotalized as

\[
\hat{T}_P = \left[ P h \hat{M}_P - A_h^0 \right] / V_h^{1/2},
\]
where $A^0_h$ is the nonstochastic centering factor and $V_0$ is the nonstochastic scale factor

$$A^0_h \equiv \left[ (h - 1)^2 \int_{-1}^{1} k^2(w_1)dw_1 + 2 \int_{0}^{1} \int_{-1}^{b} k^2_b(w_1)dw_1db \right] - 1,$$

$$V_0 \equiv 2 \left[ \int_{-1}^{1} \left( \int_{-1}^{1} k(w_1 + w_2)k(w_2)dw_2 \right)^2 dw_1 \right],$$

and $k_b(\cdot) = k(\cdot)/ \int_{-1}^{b} k(z)dz$. Hong and Li (2005) show that, under some regularity conditions, $\hat{T}_P$ follows the standard normal distribution as $P \to \infty$ under $H^2_0$ in (14). Larger values of $\hat{T}_P$ or smaller asymptotic p-values suggest rejection of $H^2_0$ in (14), indicating that the absolute value of the return $U_{t+1}$ and sign of the return $V_{t+1}$ are not independent.

### 2.3.2 The Decomposition Model

In our empirical results reported later, the independence between the sign and absolute value of the return is clearly rejected, which indicates the need to incorporate the dependence between them. One way to incorporate dependence between $U_{t+1}$ and $V_{t+1}$ is to include copula in the joint density function $f_{UV}^{t+1}(u,v)$, as in Equation (11). However, due to the binary (discrete with the bounded support) property of the Bernoulli-type distribution of the sign of the return $V_{t+1}$, not all copula functions can be used for our decomposition model. We follow Anatolyev and Gospodinov (2010) and consider the following representation of the joint density function

$$f_{UV}^{t+1}(u,v) = f_{t+1}^U(u)\rho_{t+1}^{\frac{u+1}{T}}(1 - \rho_{t+1})^{\frac{1-u}{T}}$$

where $\rho_{t+1} = \rho_{t+1}(F^U_{t+1}(u)) = 1 - \partial C(F^U_{t+1}(u), 1 - p_{t+1})/\partial w_1$. This can be proved following Anatolyev and Gospodinov (2010) with minor changes. Since the sign variable $V$ takes only two discrete values of 1 and $-1$ while the absolute returns $U$ is continuous,

$$f_{UV}^{t+1}(u,v) = \frac{\partial F_{t+1}^{UV}(u,v)}{\partial u} - \frac{\partial F_{t+1}^{UV}(u,v-2)}{\partial u}$$

$$= \frac{\partial C(F^U_{t+1}(u), F^V_{t+1}(v))}{\partial u} - \frac{\partial C(F^U_{t+1}(u), F^V_{t+1}(v-2))}{\partial u}$$

$$= f_{t+1}^U(u) \left[ \frac{\partial C(F^U_{t+1}(u), F^V_{t+1}(v))}{\partial w_1} - \frac{\partial C(F^U_{t+1}(u), F^V_{t+1}(v-2))}{\partial w_1} \right].$$
Since $C(w_1, 1) \equiv w_1$ and $C(w_1, 0) \equiv 0$, we have $\frac{\partial C(w_1, 1)}{\partial w_1} = 1$ and $\frac{\partial C(w_1, 0)}{\partial w_1} = 0$. Therefore, if $v = -1$, then $F_{t+1}^V(v) = 1 - p_{t+1}$, $F_{t+1}^V(v - 2) = 0$, and

$$f_{t+1}^{UV}(u, v) = f_{t+1}^U(u) \left[ \frac{\partial C(F_{t+1}^U(u), 1 - p_{t+1})}{\partial w_1} - 0 \right] = f_{t+1}^U(u)(1 - \rho_{t+1}).$$

If $v = 1$, then $F_{t+1}^V(v) = 1$, $F_{t+1}^V(v - 2) = 1 - p_{t+1}$, and

$$f_{t+1}^{UV}(u, v) = f_{t+1}^U(u) \left[ 1 - \frac{\partial C(F_{t+1}^U(u), 1 - p_{t+1})}{\partial w_1} \right] = f_{t+1}^U(u)\rho_{t+1}.$$

Putting these together yields the expression in (22).

We consider the Independent copula, Frank copula, Clayton copula and Farlie-Gumbel-Morgenstern copula. Their conditional copula function, conditional copula density function and the $\rho$-function are given as follows.

(1) Independent copula

$$C_{\text{Indep}}(w_1, w_2) = w_1 w_2,$$
$$c_{\text{Indep}}(w_1, w_2) = 1,$$
$$\rho_{t+1}^{\text{Indep}}(w_1, p_{t+1}, \theta) = p_{t+1}.$$

(2) Frank copula

$$C_{\text{Frank}}(w_1, w_2, \theta) = -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta w_1} - 1)(e^{-\theta w_2} - 1)}{(e^{-\theta} - 1)} \right), \theta \in (-\infty, +\infty) \setminus \{0\},$$
$$c_{\text{Frank}}(w_1, w_2, \theta) = \frac{\theta(1 - e^{-\theta})e^{-\theta(w_1+w_2)}}{[(1-e^{-\theta}) - (1-e^{-\theta w_1})(1-e^{-\theta w_2})]^2},$$
$$\rho_{t+1}^{\text{Frank}}(w_1, p_{t+1}, \theta) = \left( 1 - \frac{1 - e^{-\theta(1-p_{t+1})}}{1 - e^{\theta p_{t+1}}}e^{\theta(1-w_1)} \right)^{-1}.$$

(3) Clayton copula

$$C_{\text{Clayton}}(w_1, w_2, \theta) = (w_1^{-\theta} + w_2^{-\theta} - 1)^{-1/\theta}, \theta \in [-1, +\infty) \setminus \{0\},$$
$$c_{\text{Clayton}}(w_1, w_2, \theta) = \frac{(1+\theta)(w_1^{-\theta} + w_2^{-\theta} - 1)^{-1/\theta - 2}}{(w_1 w_2)^{\theta+1}},$$
$$\rho_{t+1}^{\text{Clayton}}(w_1, p_{t+1}, \theta) = 1 - \left( 1 + \frac{(1-p_{t+1})^{-\theta} - 1}{w_1^{-\theta}} \right)^{-1/\theta - 1}.$$
(4) Farlie-Gumbel-Morgenstern copula (FGM)

\[
C_{\text{FGM}}(w_1, w_2, \theta) = w_1 w_2 (1 + \theta(1 - w_1)(1 - w_2)), \quad \theta \in [-1, 1],
\]

\[
c_{\text{FGM}}(w_1, w_2, \theta) = 1 + \theta - 2\theta w_1 - 2\theta w_2 + 4\theta w_1 w_2,
\]

\[
\rho_{t+1}^{\text{FGM}}(w_1, p_{t+1}, \theta) = 1 - (1 - p_{t+1})(1 + \theta p_{t+1}(1 - 2w_1)).
\]

Notice that Frank copula and FGM copula are symmetric copula, while Clayton copula is asymmetric copula which shows lower tail dependence. For Frank and FGM copula, \(\theta < 0\) implies negative dependence and \(\theta > 0\) implies positive dependence. For Clayton copula, \(\theta \to 0\) leads to independence between \(w_1\) and \(w_2\) and in this case \(\rho_{t+1} \to p_{t+1}\). For Frank copula, \(\theta \to 0\) and \(\rho_{t+1} \to p_{t+1}\) implies independence. For FGM copula, \(\theta = 0\) and \(\rho_{t+1} = p_{t+1}\) implies independence.

While all the parameters including the parameter in the copula function as well as the parameters of the marginal densities can be estimated all at once, we do it in two steps, estimating the marginal densities first and then the copula density. Noting that the copula parameter \(\theta\) goes into \(\rho_{t+1}(F_{t+1}^U(u))\), rewrite it as \(\rho_{t+1}(F_{t+1}^U(u), \theta)\). The maximum likelihood estimator (MLE) is given by:

\[
\hat{\theta} = \arg \max_\theta \sum_{t=1}^{n} \log \left( f_{t+1}^U(u) \left[ \rho_{t+1}(F_{t+1}^U(u), \theta) \right]^{\frac{\nu_{t+1}}{2}} \left[ 1 - \rho_{t+1}(F_{t+1}^U(u), \theta) \right]^{\frac{1-\nu_{t+1}}{2}} \right)
\]

\[
= \arg \max_\theta \sum_{t=1}^{n} \log f_{t+1}^U(u) + \frac{\nu_{t+1}}{2} \log \rho_{t+1}(F_{t+1}^U(u), \theta) + \frac{1-\nu_{t+1}}{2} \log \left[ 1 - \rho_{t+1}(F_{t+1}^U(u), \theta) \right].
\]

Since the marginal density \(f_{t+1}^U(u)\) does not depend on the copula parameter \(\theta\), we can maximize the likelihood function in two steps. First we obtain the marginal density of \(U_{t+1}\) and its distribution \(\hat{F}_{t+1}^U(u)\), and then in the second step we get the MLE of \(\theta\) by

\[
\hat{\theta} = \arg \max_\theta \sum_{t=1}^{n} \frac{\nu_{t+1}}{2} \log(\rho_{t+1}(\hat{F}_{t+1}^U(u), \theta)) + \frac{1-\nu_{t+1}}{2} \log \left[ 1 - \rho_{t+1}(\hat{F}_{t+1}^U(u), \theta) \right].
\]

Shih and Louis (1995) show that this two step estimation is consistent although it may not be efficient. Also see Song, Fan, and Kalbfleisch (2005) and Chen, Fan and Tsyrennikov (2006).

In the rest of the paper the decomposition model using different copula functions is referred to as Model 2 or M2. In particular, M2 with Independent copula is referred to as
M2-I (the model in the previous section), M2 with Frank copula is referred to as M2-F, M2 with Clayton copula as M2-C, and M2 with Farlie-Gumbel-Morgenstern copula as M2-FGM.

3 Decomposition with Moment Constraints

A possible problem with the joint density forecast model using decomposition is that the mean prediction \( E(U_{t+1}V_{t+1}|I_t) \) from the joint density forecast function \( f_{t+1}^{UV}(u,v) \) of the decomposition model (Model 2) may not be equal to \( E(r_{t+1}|I_t) \) from Model 1, as discussed in Section 1. The mean prediction \( E(U_{t+1}V_{t+1}|I_t) \) can deviate from the mean forecast of Model 1 for which ZM or HM is used for \( E(r_{t+1}|I_t) \). In other words, the estimated decomposition model may not satisfy the mean moment condition \( E(U_{t+1}V_{t+1} + r_{t+1}|I_t) = 0 \). With this in mind, we want to impose the (conditional) moment constraint that

\[
E(U_{t+1}V_{t+1}|I_t) = \mu_{t+1}. \tag{23}
\]

We consider two values for \( \mu_{t+1} \), ZM or HM.

Since we will use the “logarithmic score” to evaluate and compare different density forecast models, we impose the moment constraint by solving the following constrained maximization problem of the cross-entropy of the new density forecast \( h_{t+1}^{UV}(u,v) \) with respect to the original density forecast \( f_{t+1}^{UV}(u,v) \):

\[
\begin{align*}
\max_{h_{t+1}^{UV}(u,v)} & = - \int \int \left( \log \frac{h_{t+1}^{UV}(u,v)}{f_{t+1}^{UV}(u,v)} \right) h_{t+1}^{UV}(u,v)du dv \\
\text{subject to} & \int \int m_t(u,v)h_{t+1}^{UV}(u,v)du dv = 0, \tag{25} \\
\text{and} & \int \int h_{t+1}^{UV}(u,v)du dv = 1, \tag{26}
\end{align*}
\]

where \( f_{t+1}^{UV}(u,v) \) is the density forecast from the decomposition model and \( h_{t+1}^{UV}(u,v) \) is a new density forecast satisfying the moment constraint of (25). The moment constraint in (23) is rewritten as (25) with

\[
m_t(u,v) = uv - \mu_{t+1}. \tag{27}
\]

Note that the expectation in (23) is evaluated using the new density forecast \( h_{t+1}^{UV}(u,v) \). Note that the moment constraint function \( m_t(u,v) \) is denoted with the subscript \( t \) as it is
measurable with respect to the information $I_t$ at time $t$ (as $\mu_{t+1}$ is $I_t$-measurable). The moment condition (25) with (27) will make the new joint density forecast $h_{t+1}^{UV}(u, v)$ have the same mean forecast $\mu_{t+1}$ as the benchmark Model 1.

The maximization of (24) subject to the moment constraint (25) is well established in the literature. Jaynes (1957) was the pioneer to consider this problem. Jaynes (1957, 1968) provides a solution for discrete density, while a general solution for any type of density can be found in Csiszár (1975). Also see Maasoumi (1993), Zellner (1994), Golan, Judge, and Miller (1996), Ullah (1996), Bera and Bilias (2002), among others.

The solution to the above maximization problem, if exists, is given by

$$h_{t+1}^{UV}(u, v) = f_{t+1}^{UV}(u, v) \exp \left[ \eta_t^* + \lambda_t^* m_t(u, v) \right],$$

where

$$\lambda_t^* = \arg \min_{\lambda_t} I_t(\lambda_t),$$

$$\eta_t^* = - \log I_t(\lambda_t^*),$$

$$I_t(\lambda_t) = \int \int \exp[\lambda_t m_t(u, v)] f_{t+1}^{UV}(u, v) dudv.$$

That is, to get the solution $h_{t+1}^{UV}(u, v)$, we constructed a new density forecast by exponentially tilting through $\lambda_t^*$ and normalizing it through $\eta_t^*$. This derivation can also be found in recent econometric applications of the maximum entropy, as in Kitamura and Stutzer (1997), Imbens, Spady and Johnson (1998), Bera and Bilias (2002), Kitamura, Tripathi and Ahn (2004), Robertson, Tallman and Whiteman (2005), Park and Bera (2006), Bera and Park (2008), Stengos and Wu (2010), and Giacomini and Ragusa (2013), among others.

Note that the objective function of (24) is the (negative) conditional Kullback-Leibler (1951) information criterion (KLIC) divergence measure between the new conditional density and the original conditional density. If the (conditional) moment constraint is true, the difference of expected (conditional) logarithmic scores between $h_{t+1}^{UV}(u, v)$ and $f_{t+1}^{UV}(u, v)$ is
nonnegative. To be specific, if the conditional moment constraint is true,

\[
KLIC(h_{t+1}^{UV}, f_{t+1}^{UV}) = \int \int \left( \log \frac{h_{t+1}^{UV}(u, v)}{f_{t+1}^{UV}(u, v)} \right) h_{t+1}^{UV}(u, v) dudv
\]

\[
= \int \int \log \exp [\eta_t + \lambda_t m_t(u, v)] h_{t+1}^{UV}(u, v) dudv
\]

\[
= \eta_t \int \int h_{t+1}^{UV}(u, v) dudv + \lambda_t \int \int m_t(u, v) h_{t+1}^{UV}(u, v) dudv
\]

\[
= \eta_t.
\]

Since \( \eta_t = KLIC(h_{t+1}^{UV}, f_{t+1}^{UV}) \geq 0 \), we have \( \eta_t^* = -\log I_t(\lambda_t^*) \geq 0 \) and therefore \( 0 < I_t(\lambda_t^*) \leq 1 \).

To find \( \lambda_t^* \), we need first to find the function of \( I_t(\lambda_t) \), which is the integral of the joint density function of \( U_{t+1} \) and \( V_{t+1} \) times an exponential function of the moment constraint. We will use the numerical integration in the empirical section, since the analytical solution to \( I_t(\lambda_t) \) does not have an explicit expression under the historical mean constraint as well as for some copula functions. To implement the numerical integral, we note from (31)

\[
I_t(\lambda_t) = E_t \exp [\lambda_t m_t(u, v)],
\]

(32)

where the expectation is taken over the joint density forecast \( f_{t+1}^{UV}(u, v) \). Thus we generate \( S \) random draws \( \{u_t^s, v_t^s\}_{s=1}^S \) from \( f_{t+1}^{UV}(u, v) \) and calculate \( I_t(\lambda_t) = \frac{1}{S} \sum_{s=1}^S \exp [\lambda_t m(u_t^s, v_t^s)] \).

Then \( \lambda_t^* \) can be solved by minimizing \( I_t(\lambda_t) \), and then \( \eta_t^* \) is obtained by \( \eta_t^* = -\log I_t(\lambda_t^*) \).

A possible problem with using the numerical integration is that, as we discuss below, \( I(\lambda_t) \) is flat for a wide range of \( \lambda_t \), so that the numerical integration may not well behave and the algorithm may stop before reaching \( \lambda_t^* \). Therefore the nonnegativity of \( \eta_t^* \) may not be guaranteed.

To generate the random draws of \( \{u_t^s, v_t^s\}_{s=1}^S \) from \( f_{t+1}^{UV}(u, v) \) under independent copula, we just need to generate \( U_{t+1} \) and \( V_{t+1} \) separately according to their marginal density functions since the joint density \( f_{t+1}^{UV}(u, v) \) is just equal to the product of the two marginal density functions under independence. For the random draws under dependence, an easy way to generate \( \{u_t^s, v_t^s\}_{s=1}^S \) from \( f_{t+1}^{UV}(u, v) \) is to first generate \( U_{t+1} \) based on its exponential marginal density function \( f_{t+1}^{U}(u) = \frac{1}{\psi_{t+1}} \exp \left( -\frac{1}{\psi_{t+1}} u \right) \), and then generate \( V_{t+1} \) based on the
conditional density of $f^V_U(v|u)$, and since $V_{t+1}$ is binary, the conditional density should be Bernoulli-type. From (22), the conditional density of $V_{t+1}$ conditioning on $U_{t+1}$ is given by:

$$f^V_{t+1}(v|u) = \frac{f^{UV}_{t+1}(u, v)}{f^U_{t+1}(u)} = \rho_{t+1}^{1-v}(1-\rho_{t+1})^{1-u}, \quad v \in \{-1, 1\}.$$  

The decomposition model (Model 2) with the moment constraint imposed will be called Model 3 or M3. In particular, we will call M3 imposing ZM moment constraint $m_t(u, v) = uv - \mu_{t+1} = 0$ with $\mu_{t+1} = 0$ as M3-ZM, and imposing the HM constraint where $\bar{\mu}_{t+1} = \bar{r}_t$ as M3-HM. And for each copula function, we will denote the model using the name of copula such as M3-ZM-I to denote M3 with the zero mean constraint and Independent copula. The labels for other copula functions are made similarly. Model 2 is the original density forecast model $f^{UV}_{t+1}(u, v)$, while Model 3 is the tilted maximum entropy density forecast model $h^{UV}_{t+1}(u, v)$. We will evaluate Model 2 and Model 3 to examine if imposing the moment constraint can improve the density forecast with different copula functions.

To better understand when the constraint can improve the density forecast, let us consider a simple case when the maximization problem can be solved analytically. Solving it analytically will give the most accurate results of $\lambda_t^{*}$ and $\eta_t^{*}$ and ensure that $\eta_t^{*}$ is nonnegative. To illustrate, consider a simple case under DP1 with ZM ($\bar{\mu}_{t+1} = 0$). In this case, the analytical expression of $I_t(\lambda_t)$ is obtained as follows:

$$I_t(\lambda_t) = \int \int \exp[\lambda_t m(u, v)] f^{UV}_{t+1}(u, v) du dv = \int \int f^{UV}_{t+1}(u, v) \exp[\lambda_t m_t(u, v)] dv du + \int \int f^{UV}_{t+1}(u, v) \exp[\lambda_t m_t(u, v)] dv du = \int f(u, -1) \exp[\lambda_t m(u, -1)] du + \int f(u, 1) \exp[\lambda_t m(u, 1)] du = \int_0^\infty \frac{1}{\psi_{t+1}} \exp(-\frac{1}{\psi_{t+1}} u) (1-p_{t+1}) \exp(-\lambda_t u) du + \int_0^\infty \frac{1}{\psi_{t+1}} \exp(-\frac{1}{\psi_{t+1}} u) p_{t+1} \exp(\lambda_t u) du = \frac{(1-p_{t+1})}{1+\bar{\lambda}_t} + \frac{P_{t+1}}{1+\bar{\lambda}_t}$$  

A plot of $I_t(\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_t = 0.55$ is given in Figure 1a. These values of $\frac{1}{\psi_{t+1}}$ and $p_t$ are estimated from actual annualized monthly equity premium series. A plot of $I_t(\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_t = 0.65$ is given in Figure 1c, with the values of $\frac{1}{\psi_{t+1}}$ and $p_t$ estimated from actual annualized monthly stock return series. We can see that for a wide range of $\lambda_t$, $I_t(\lambda_t)$
is near flat. $I_t(\lambda_t)$ appears to change little over the flat area especially when $p_t$ is nearer to 0.50. Therefore when using numerical integration $I_t(\lambda_t) = \frac{1}{S} \sum_{s=1}^{S} \exp [\lambda_t m_t (u_s^t, v_s^t)]$ to find the optimal value, it can easily stop somewhere in the flat area where $I_t(\lambda_t)$ may be above 1 (and thus $\eta^*$ may be less than 0).

To find $\lambda_t^* = \arg \min_{\lambda_t} I_t(\lambda_t)$, one can use the analytical solution for $\lambda_t^*$ from solving the first order condition

$$\frac{dI_t(\lambda_t)}{d\lambda_t} = -\frac{1 - p_{t+1}}{(\psi_{t+1} + \lambda_t)^2} + \frac{p_{t+1}}{(\frac{1}{\psi_{t+1}} - \lambda_t)^2} = 0.$$ \hspace{1cm} (34)

Solving for $\lambda_t$ and choosing the solution whose absolute value is less than $\frac{1}{\psi_{t+1}}$, we get:

$$\lambda_t^* = \frac{-1 + 2\sqrt{p_{t+1}(1 - p_{t+1})}}{2p_{t+1} - 1},$$ \hspace{1cm} (35)

if $p_t \neq \frac{1}{2}$. Note that $I_t(0) = 1$ and in fact $I_t(\lambda_t) < 1$ for some $\lambda_t$. To look more closely at the bottom of Figure 1a and Figure 1c, these figures are magnified into Figure 1b and Figure 1d for a narrower domain of $-2 < \lambda_t < 2$ that includes the optimal value $\lambda_t^*$. It can be seen that $I_t(\lambda_t^*) < 1$ at the optimal value of $\lambda_t$.

Plug $\lambda_t^*$ back into $I_t(\lambda_t)$, we can find the minimized value of the integral

$$I_t(\lambda_t^*) = \frac{(1 - p_{t+1})(2p_{t+1} - 1)}{2p_{t+1} - 2 + 2\sqrt{p_{t+1}(1 - p_{t+1})}} + \frac{p_{t+1}(2p_{t+1} - 1)}{2p_{t+1} - 2\sqrt{p_{t+1}(1 - p_{t+1})}}.$$ \hspace{1cm} (36)

It is interesting to see that $I_t(\lambda_t^*)$ does not depend on $\psi_{t+1}$ but depends only on $p_{t+1}$.

Figure 1e is the plot $I_t(\lambda_t^*)$ as a function of $p_{t+1} \in (0 \ 1) \setminus \{0.5\}$. Note that the optimal value of $I_t(\lambda_t^*)$ is smaller than 1 for all values of $p_{t+1}$ on $(0 \ 1) \setminus \{0.5\}$, which means that $\eta_t^*$ is greater than 0 for all $p_{t+1}$ (except for 0.50). While $\lambda_t^*$ in (35) is not defined for $p_{t+1} = 0.5$, the limiting value $\lim_{p_t \to 0.5} I_t(\lambda_t^*) = 1$ as shown in Figure 1e, in which case $\eta_t^* = -\log I_t(\lambda_t^*) \to 0$ indicating that the moment constraint would not improve the density forecast when $p_{t+1} \to 0.5$. It is important to note that, the further $p_{t+1}$ deviates from 0.5, the more room we can have for improvement from imposing the moment constraint. This is because $\eta_t^*$ can be substantially less than 1 when $p_{t+1}$ deviates from 0.5. See Figure 1(e,f).

---

1This is due to the particular moment condition $m_t(u, v) = uv - \mu_{t+1}$ and $\mu_{t+1} = 0$ (27) in deriving this. It is not true in general with a different moment condition. For example, if the moment condition with $\mu_{t+1} \neq 0$, $I_t(\lambda_t^*)$ will depend on $\psi_{t+1}$ and $p_{t+1}$.
4 Density Forecast Evaluation

To evaluate density forecasts from different models, we use a scoring rule. A scoring rule is a positive-oriented criterion to evaluate density forecasts. A larger expected score usually means that the associated density forecast is better. Formally, a score function or scoring rule $S(f, y)$ of the density forecast $f$ is a real value evaluated at the realized value $y$ of a random variable. Let $E_h S(f, y) = \int S(f, y)h(y)dy$ be the expected score value of $S(f, y)$ under the density function $h(\cdot)$. A scoring rule is said to be proper if $E_h S(h, y) \geq E_h S(f, y)$ for all density functions $f(\cdot)$ and $h(\cdot)$. If the equality holds only if $f(\cdot) = h(\cdot)$, then the score function is strictly proper. See Gneiting and Raftery (2007) and Gneiting and Ranjan (2011). For a proper scoring rule, the expected score of the true density is always greater than the expected score of any other density.

One of the most popular scoring rules is the likelihood-based scoring rule (also known as the logarithmic score):

$$S(f, y) = \log f(y).$$

(37)

The difference of the expected scores $[E_h S(h, y) - E_h S(f, y)]$ is the KLIC divergence measure. The logarithmic score is strictly proper because

$$KLIC(h, f) = E_h [S(h, y) - S(f, y)] = E_h [\log h(y) - \log f(y)] \geq 0$$

(38)
due to the Jensen’s inequality applied to the logarithmic function which is concave. See Rao (1965), White (1994), and Ullah (1996). Hence, we wish to find the density forecast model that gives the highest expected logarithmic score. For out-of-sample forecast, the expected score is estimated by computing the average out-of-sample scores

$$\hat{S}_P = \frac{1}{P} \sum_{t=T-P}^{T-1} S_t,$$

(39)

where $S_t = S(f_{t+1}, y_{t+1})$ is the score of the density forecast made at time $t$ and evaluated at the realized value $y_{t+1}$ at time $t + 1$. The density forecast with a higher value of $\hat{S}_P$ is the better density forecast.

To apply the logarithmic score to the decomposition model, let $y = (u, v)$ and

$$S(f_{t+1}, y_{t+1}) = S(f_{t+1}^{UV}, (u_{t+1}, v_{t+1})) = \log f_{t+1}^{UV} (u_{t+1}, v_{t+1}).$$

(40)
The joint density forecast can be improved by improving any of the two marginal density forecasts and a copula density forecast, i.e., any of the three terms in the right hand side of

\[
S \left( f_{t+1}^{UV}(u_{t+1}, v_{t+1}) \right) = \log f_{t+1}^{U}(u_{t+1}) + \frac{1 + v_{t+1}}{2} \log \rho_{t+1} + \frac{1 - v_{t+1}}{2} \log(1 - \rho_{t+1}).
\]  

(41)

We compare density forecasts by the average out-of-sample logarithmic scores

\[
\hat{S}_P = \frac{1}{P} \sum_{t=T-P}^{T-1} S_t,
\]

where \( S_t \) is the logarithmic score \( S \left( f_{t+1}^{UV}(u_{t+1}, v_{t+1}) \right) \) of the joint density forecast made at time \( t \), and evaluated at the realized absolute return \( u_{t+1} \) and the realized sign \( v_{t+1} \) at time \( t+1 \).

Finally it should be noted that the logarithmic score of this joint density of \( U \) and \( V \) can be compared with that of the normal density forecast model (Model 1) as well. Since \( r_{t+1} \equiv U_{t+1}V_{t+1} \), the normal density forecast model conditional on the information set \( I_t \) in (2) can be seen as

\[
f_{t+1}(r) = \sum_{v \in \{-1,1\}} f_{t+1}^{UV}(u,v) = f_{t+1}^{UV}(-r,-1) + f_{t+1}^{UV}(r,1)
\]

because the Jacobian of the transformation from \( (r = uv, w = v) \) to \( (u = r/w, v = w) \) is 1. Further, since the logarithmic score is evaluated at one realized value \( r_{t+1} = u_{t+1}v_{t+1} \) at each time \( t \), the average out-of-sample logarithmic score value of the density forecast of Model 1 is expressed as that of the joint density forecast \( f_{t+1}^{UV}(u_{t+1}, v_{t+1}) \)

\[
\frac{1}{P} \sum_{t=T-P}^{T-1} \log f_{t+1}(r_{t+1}) = \frac{1}{P} \sum_{t=T-P}^{T-1} \log f_{t+1}^{UV}(u_{t+1}, v_{t+1}).
\]

So we can actually compare Model 1 and Model 2 via the scoring rule of the joint density of \( U \) and \( V \).

5 Risk Forecast

The most important application of the density forecast models is to make risk forecasts. Since a “better” density forecast are in terms of the expected logarithmic score which is estimated by the average out-of-sample logarithmic score evaluated at \( \{(u_{t+1}, v_{t+1})\} \), the evaluation of the density forecast models is to compare the overall performance over the entire support \( u \in \mathbb{R}^+ \) and \( v \in \{-1,1\} \). A better density forecast over the entire support may not be the
best density forecast for a certain subset of the support. It would be interesting to compare
the density forecast models over the tail part of the support. Furthermore, as the mean
forecast was constrained to be fixed at ZM or HM, the difference of the different models may
be from the support away from the mean. The density forecasts from Model 2 generates
higher average log score $\hat{S}_P$ than Model 1, while the mean forecasts from Model 2 are not
necessarily better than those from Model 1. That means the improvement should come from
tails.

We wish to examine if the decomposition and the maximum entropy can also generate
superior density forecast in tails. There are several different ways to evaluate the tail density
forecasts, e.g., Gneiting and Ranjan (2011), Diks, Panchenko and van Dijk (2011). In this
paper, however, we focus on VaR, the quantiles of the density forecasts. To be specific,
we will invert the conditional density forecasts to obtain the conditional quantile forecasts,
namely, $\text{VaR}(\alpha)$ for a given tail probability $\alpha$. While we could do more by computing other
tail/risk measures such as the risk spectrum (the weighted average of return quantiles with
weights reflecting different risk aversion), the expected short fall, the loss given default, or
unexpected loss, we will focus on the forecasts of $\text{VaR}(\alpha)$ with $\alpha = 0.01$.

In this section, we set the Riskmetrics model of J.P. Morgan (1995) as a benchmark. The
Riskmetrics model forecasts quantile, denoted by $\hat{q}_{t+1}$, as follows:

$$\hat{q}_{t+1}(\alpha) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \Phi^{-1}(\alpha),$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal CDF so that $\Phi(0.01) = -2.33$, $\hat{\mu}_{t+1} = \bar{r}_t = \frac{1}{t} \sum_{j=1}^{t} r_j$ is the HM estimated from the recursive expanding window, and $\hat{\sigma}_{t+1}$ is estimated
by the exponentially weighted moving average (EWMA)

$$\hat{\sigma}_{t+1}^2 = 0.94\hat{\sigma}_{t}^2 + 0.06(r_t - \hat{\mu}_{t+1})^2.$$

Another benchmark is the VaR forecast from normal density forecast (M1). Since the
normal density forecast is determined by its mean forecast $\hat{\mu}_{t+1}$ and standard deviation
forecast $\hat{\sigma}_{t+1}$ from GARCH(1,1), the VaR forecast is given by $\hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \Phi^{-1}(\alpha)$.

For the decomposition models M2 and M3, it is difficult to derive the VaR analytically
as there is no explicit solutions, so we use numerical methods to obtain VaR forecasts from
the density forecast. Once the joint density forecast \( f_{t+1}^{UV}(u, v) \) in M2 or the joint density forecast under the moment constraint \( h_{t+1}^{UV}(u, v) \) in M3 are computed, we can again generate \( S \) random draws of \( \{u_t^s, v_t^s\}_{s=1}^S \) from \( f_{t+1}^{UV}(u, v) \) or \( h_{t+1}^{UV}(u, v) \), and calculate the return by \( r_t^s = u_t^s v_t^s \) and find the \( \alpha \)-percentile by taking the \([\alpha S]\) lowest return from the random draws as the forecasted VaR(\( \alpha \)) for a given \( \alpha \).

To generate the random draws of \( \{u_t^s, v_t^s\}_{s=1}^S \) from \( f_{t+1}^{UV}(u, v) \), we can just follow the same procedure in the last section. However, to generate the random draws of \( \{u_t^s, v_t^s\}_{s=1}^S \) from \( h_{t+1}^{UV}(u, v) \) is less straightforward. Still we follow similar steps, that we first generate \( U_{t+1} \) based on its marginal density \( h_t^U(u) \), and then generate \( V_{t+1} \) based on the conditional density of \( h_{t+1}^{V|U}(v|u) \), and since \( V_{t+1} \) is binary, the conditional density should also be Bernoulli-type. Substitute (22) and (27) to (28), the joint density of \( U_{t+1} \) and \( V_{t+1} \) under moment constraint can be written as:

\[
h_{t+1}^{UV}(u, v) = f_{t+1}^{U}(u) \rho_{t+1} \frac{1-u}{2} \exp \left[ \eta_{t}^* + \lambda_{t}^*(uv - \mu_{t+1}) \right].
\]

Since the sign part is binary where \( V_{t+1} \) can only take value of 1 or \(-1\), the marginal distribution of \( U_{t+1} \) can then be written as:

\[
h_t^U(u) = h_{t+1}^{UV}(u, -1) + h_{t+1}^{UV}(u, 1)
= f_{t+1}^{U}(u)(1 - \rho_{t+1}) \exp(\eta_t^* - \lambda_t^* \hat{\mu}_{t+1} - \lambda_t^* u) + f_{t+1}^{U}(u)\rho_{t+1} \exp(\eta_t^* - \lambda_t^* \hat{\mu}_{t+1} + \lambda_t^* u)
= \exp(\eta_t^* - \lambda_t^* \hat{\mu}_{t+1}) f_{t+1}^{U}(u)[(1 - \rho_{t+1}) \exp(-\lambda_t^* u) + \rho_{t+1} \exp(\lambda_t^* u)],
\]

then the conditional density of \( V_{t+1} \) given realizations of \( U_{t+1} \) is given by:

\[
h_{t+1}^{V|U}(v|u) = \frac{h_{t+1}^{UV}(u, v)}{h_{t+1}^{U}(u)}
= \frac{\rho_{t+1} \frac{1-u}{2} \exp(\lambda_t^* uv)}{(1 - \rho_{t+1}) \exp(-\lambda_t^* u) + \rho_{t+1} \exp(\lambda_t^* u)}.
\]

However, it is not always easy to generate \( u \)'s from the \( h_t^U(u) \) function since it is not a common density function especially when the \( \rho_{t+1} \) function is complicated for some copulas. To solve this problem, we consider two possible methods. The straight method to generate \( u \) from \( h_t^U(u) \) is to use the probability integral transformation (PIT). Since we assume that

\[
...
$u$ follows an exponential distribution with mean equal to $\frac{1}{\psi_{t+1}}$, and let $\theta_{t+1} = \frac{1}{\psi_{t+1}}$, we have

$$f_{t+1}^U(u) = \frac{1}{\psi_{t+1}} \exp\left(-\frac{1}{\psi_{t+1}} u\right).$$

Substituting this expression into the marginal distribution $h_{t+1}^U(u)$, we get

$$h_{t+1}^U(u) = \exp(\eta_t^* - \lambda_t^* \theta_{t+1}) \frac{1}{\psi_{t+1}} \exp\left(-\frac{1}{\psi_{t+1}} u\right) \left[(1 - \theta_{t+1}) \exp(-\lambda_t^* u) + \theta_{t+1} \exp(\lambda_t^* u)\right],$$

and the CDF

$$H_{t+1}^U(u) = \int h_{t+1}^U(u) du$$

$$= \exp(\eta_t^* - \lambda_t^* \theta_{t+1}) \frac{1}{\psi_{t+1}} \frac{1 - \theta_{t+1}}{1 + \lambda_t^*} \left[1 - \exp\left(-\frac{1}{\psi_{t+1}} \left(\frac{1}{\psi_{t+1}} + \lambda_t^* \right) u\right)\right]$$

$$+ \exp(\eta_t^* - \lambda_t^* \theta_{t+1}) \frac{1}{\psi_{t+1}} \frac{\theta_{t+1}}{1 + \lambda_t^*} \left[1 - \exp\left(-\frac{1}{\psi_{t+1}} \left(\frac{1}{\psi_{t+1}} - \lambda_t^* \right) u\right)\right].$$

Once we get this analytical expression of the CDF of $U_{t+1}$, we can first generate random numbers from uniform distribution, and solve for $u$ from the inverse of its CDF function evaluated at the realizations of the uniform distribution. However, $H_{t+1}^U(u)$ is a highly nonlinear function in $u$, and the only way to solve the inverse function of the CDF is through numerical methods, which could be very inefficient and inaccurate and thus affect the random draws of $u$, so we do not consider using PIT in the empirical application.

An alternative method is the Metropolis-Hastings algorithm to generate $u$’s from $h_{t+1}^U(u)$. The sketch of this algorithm is as follows. Let $Y \sim f_Y(y)$ (target density) and $X \sim f_X(x)$ (candidate-generating density), and $f_Y$ and $f_X$ have common support. If it is easy to generate $X$ from $f_X(x)$, then the following algorithm can generate $Y$ from $f_Y(y)$:

1. Generate $X_0 \sim f_X(x)$. Set $Z_0 = X_0$.

2. For $i = 1, 2, \ldots, S$, generate $U_i \sim \text{Uniform}(0, 1)$ and $X_i \sim f_X(x)$. Calculate

$$\alpha_i = \min\left\{\frac{f_Y(X_i)}{f_X(X_i)} \cdot \frac{f_X(Z_{i-1})}{f_Y(Z_{i-1})}, 1\right\}.\$$

$$Z_i = \begin{cases} X_i & \text{if } U_i \leq \alpha_i \\
Z_{i-1} & \text{if } U_i > \alpha_i \end{cases}.\$$

Then, as $i \to \infty$, $Z_i$ will converge to $Y$ in distribution.
See Casella and Berger (2002) and Chib and Greenberg (1995) for an introduction to the Metropolis-Hastings algorithm. In terms of our notation to generate \( u \sim h^U_{t+1}(u) \), \( h^U_{t+1} \) is the target density and \( f^U_{t+1} \) is the candidate-generating density. Since it is easy to generate \( u \sim f^U_{t+1}(u) \), we will set \( u \sim f^U_{t+1}(u) \) as the \( X \) variable above, and \( u \sim h^U_{t+1}(u) \) to be the \( Y \) variable, and then \( u \sim h^U_{t+1}(u) \) can be generated applying the Metropolis-Hastings algorithm. Using these \( u's \) we can get the VaR forecast according to the numerical method discussed above.

We evaluate the VaR forecasts from different density forecast models in terms of the predictive quantile loss and the empirical coverage probability. To compare the productivity of VaR from different density forecasts, we use the “check function” of Koenker and Bassett (1978). The expected check loss of quantile \( q_t(\alpha) \) for a given left tail probability level \( \alpha \) has the form

\[
L(\alpha) = E[\alpha - 1(r_t < q_t(\alpha))][r_t - q_t(\alpha)].
\]

Saerens (2000), Bertail, Haefke, Politis and White (2004), Komunjer (2005), Bao, Lee and Saltoğlu (2006), and Gneiting (2011) show that the check function can be regarded as a quasi-likelihood, therefore the expected check loss \( L(\alpha) \) can provide a measure of the lack-of-fit of a quantile model. Once we obtained the out-of-sample VaR forecasts \( \hat{q}_t(\alpha)'s \), we can plug them into the above expression and evaluate the out-of-sample expected check function as:

\[
\hat{L}_P(\alpha) = \frac{1}{P} \sum_{t=R+1}^{T} [\alpha - 1(r_t < \hat{q}_t(\alpha))][r_t - \hat{q}_t(\alpha)].
\]

Then a model which gives the VaR forecast \( \hat{q}_t(\alpha)'s \) with the minimum loss value of \( \hat{L}_P(\alpha) \) is considered as the best model.

As an alternative evaluation of risk forecast, when the CDF of \( r_t \) is continuous in a neighborhood of \( q_t(\alpha) \), \( q_t(\alpha) \) minimizes \( L(\alpha) \) and makes a condition for the correct conditional coverage probability

\[
\alpha = E[1(r_t < q_t(\alpha))|I_{t-1}],
\]

so \( \{\alpha - 1(r_t < q_t(\alpha))\} \) is a martingale difference sequence, which can be used to form a conditional moment test to evaluate VaR forecasts. Without deriving such a formal test.
statistic, we report only the empirical out-of-sample coverage probability defined as 

\[
\hat{\alpha}_P = \frac{1}{P} \sum_{t=R+1}^{T} 1 (r_t < \hat{q}_t(\alpha)),
\]

where \(\hat{q}_t(\alpha)\) is a forecast of \(q_t(\alpha)\). The density forecast model which gives \(\hat{\alpha}_P\) closest to the nominal value \(\alpha\) is the preferred model.

6 Empirical Results

6.1 Data

In our empirical studies, we use the data set of Goyal and Welch (2008). In addition to their original monthly return, we calculate the annualized monthly stock return as in Campbell and Thompson (2008). We consider the density forecast of both stock return and equity premium. Since the difference between equity premium and stock return is the risk free rate which is relatively small and smooth compared to the equity premium, the equity premium should have similar distribution properties as those of the stock return discussed above. Therefore we can apply the decomposition model incorporating dependence as well as imposing constraints to equity premium as well. Denote by \(P_t\) the S&P500 index at month \(t\). The monthly one-month return from month \(t\) to month \(t+1\) is defined as \(R_t(1) \equiv P_{t+1}/P_t - 1\), and one-month excess return is denoted as \(Q_t(1) \equiv R_t(1) - r^f_t\) with \(r^f_t\) being the risk-free interest rate. Following Campbell, Lo and MacKinlay (1997, page 10), we define the \(k\)-period return from month \(t\) to month \(t+k\) as

\[
R_t(k) \equiv \frac{P_{t+k}}{P_t} - 1 \\
= \left( \frac{P_{t+k}}{P_{t+k-1}} \right) \times \cdots \times \left( \frac{P_{t+1}}{P_t} \right) - 1 \\
= (1 + R_{t+k-1}(1)) \times \cdots \times (1 + R_t(1)) - 1 \\
= \left[ \prod_{j=1}^{k} (R_{t+k-j}(1) + 1) \right] - 1.
\]
and following Campbell and Thompson (2008) we define the $k$-period excess return as

$$Q_t(k) \equiv (1 + R_{t+k-1}(1) - r_{t+k-1}^f) \times \cdots \times (1 + R_t(1) - r_t^f) - 1$$

$$= (Q_{t+k-1}(1) + 1) \times \cdots \times (Q_t(1) + 1) - 1$$

$$= \left[ \prod_{j=1}^{k} (Q_{t+k-j}(1) + 1) \right] - 1. \quad (43)$$

We allow $r_{t+1} = R_t(k)$ or $Q_t(k)$ when making the density and risk forecast and consider $k = 1, 3, 12$ as denoted in Campbell and Thompson (2008).

We consider the data from May 1937 to December 2002 with the total of 788 observations, since although we did not report the result of using the covariates, we did used them and compared with the results without covariates, and this time period includes the most updated data for all 13 predictors. We divide the whole sample equally into $R$ in-sample observations and $P$ pseudo out-of-sample observations, with $R = P = 394$. The models are estimated using rolling windows of the fixed size $R$. That is, at each time $t$ we use the data starting from $t - R + 1$ and ending at time $t$ to estimate parameters of a model and then make one-period ahead forecast for the next period $t + 1$. For annualized or quarterly aggregated monthly data, to avoid using future information, we only use data up to month $t - 11$ or $t - 2$ for estimation.

6.2 Results

Table 1 shows the test statistic and its asymptotic $p$ value for test of independence between $U_{t+1}$ and $V_{t+1}$. The null hypothesis of independence is clearly rejected for both stock return and equity premium and for all $k$. The test statistics are huge numbers from standard normal distribution and the $p$-values are zero. Many papers have already cited that the stock returns exhibit negative skewness, which is an evidence of dependence between the absolute return and the sign of the return. The test results are formal confirmation of this dependence and are in accordance with the literature.

Figure 2 plots the estimated copula parameter $\theta$ for Frank, Clayton and FGM copula over time, and we can see that from the figures all the three copula parameters are away from 0, which again indicates that the absolute return and the sign of the return are not
independent, since when the parameter converges to zero, all the three copula converges to independent copula. Moreover, from these figures we can also tell whether there is positive or negative dependence between the absolute return and the sign of the return. For stock returns, the dependence keeps to be positive since \( \theta \) remains positive for all time period. Yet for equity premium, it seems there is a sign change around the 1980s, which shows there may be a structural break for the dependence between the absolute return and the sign of the return.

Figure 3 plots the estimated \( \rho_{t+1} \) for Frank, Clayton, FGM, and Independent copula functions over time. For Independent copula, \( \rho_{t+1} = p_{t+1} \). The magnitude of the improvement from M2-I to M2 with dependent copula depends on how far away \( \rho_{t+1} \) deviates from \( p_{t+1} \). Using the Independent copula can be regarded as a shrinkage in \( \theta \) towards zero or equivalently from \( \rho_{t+1} \) to \( p_{t+1} \). Depending on the strength in the dependence, the shrinkage of ignoring the dependence may benefit the out-of-sample density forecast.

Table 2 shows the out-of-sample average values of the log scores for different density forecast models for the stock return and equity premium. We evaluate and compare density forecasts from M1, M2 and M3 to see whether decomposition model improves from normal distribution models, whether relaxing DP1 using a copula can helps improving M2, and whether imposing moment constraint can improve the density forecast. First, comparing the log scores for M1 and M2 independent copula, we observe that the decomposition model with independent copula improve substantially upon the normal distribution model, especially for annually and quarterly aggregated data, for both the stock return and equity premium. Moreover, using the three dependent copula functions further improves substantially for both stock return and equity premium and all the aggregation levels. For example, the average log score jump from \(-0.0399\) for M1 to \(0.1763\) for M2 using independent copula, and then to \(0.6434\) for M2 using Frank copula.

In addition, to see the effect of imposing moment constraint on the density forecast, compare M3 with M2. For stock return, M3 will improve upon M2 when the constraint imposed is correct. For stock return, M3 always improves M2, yet for equity premium, only for quarterly aggregated data, M3 can improve upon M2. Adding moment constraints in
general improves the density forecast if the ZM or HM constraint is correct. The average log scores all go up for either ZM or HM or both moment constraints except for $R(3)$ the three month return. The ZM constraint is preferred for equity premium for monthly and quarterly aggregated data and while the HM constraint is better for equity premium and stock return for monthly and annually aggregated data. Since the annually aggregated return is the accumulated return over 12 months, the aggregate return would deviate more from ZM, and thus for 12 month aggregated returns, HM would be a better constraint than ZM. For monthly or quarterly aggregated returns, the mean does deviate far from zero so that ZM may be a better constraint. For equity premium which is equal to the stock return minus the risk free rate, the means of $Q(k)$ is closer to 0 than those of $R(k)$. That is why the ZM constraint works better for equity premium for shorter return period while the HM constraint is preferred for longer return period.

The magnitude of the improvement from M2 to M3 depends on how far away the mean forecast from M2 deviates from the mean constraint we impose in M3. To see this look at Figure 4, which plots the mean forecasts from different decomposition models and the historical mean forecasts. Figure 4(d) shows that for annually aggregated stock return the decomposition models make mean forecasts much deviating from the historical mean forecast, but in Figure 4(a) the annually aggregated equity premium have the mean forecasts from decomposition models that are close to the historical mean forecasts. That explains Table 3 that the average log scores increase more for annually aggregated stock return than those for annually aggregated equity premium when we impose the historical mean constraint.

Now let us turn to out-of-sample risk forecast using VaR. Tables 3 and 4 report the predictive quantile loss and the empirical coverage probability for different aggregation levels $(k)$ of stock return $R(k)$ and equity premium $Q(k)$. In order to see more decimal numbers, we have multiplied 100 to both the predictive quantile check loss values as well as the empirical coverage probability value, so that all the numbers for $\hat{\alpha}_p$ in these tables are in percentage. From these tables we can see the decomposition model of density forecast produces better VaR forecasts than Riskmetrics and normal distribution model for 1% quantile since we can always find at least one, and most cases more than one decomposition models among M2 and
M3 using different copula functions which will produce a smaller loss. However, if we choose the 5% and 10% quantile VaR to forecast, the Riskmetrics model is hard to beat in terms of expected loss, but it may not always give a more accurate empirical coverage probability. The intuition behind why decomposition model works for 1% quantile is due to the marginal density assumption for the absolute return $U_{t+1}$. Since we assume an exponential distribution for $U_{t+1}$ which has heavier tails than the normal distribution, it will make a more extreme and thus conservative VaR forecast, and thus is consistent with the fact that stock returns does have large drops in cases of financial crisis. To sum up, the density forecast using the decomposition model generates higher (larger in absolute value) VaR forecasts, yet it will generate less loss compared to Riskmetrics when extreme events happen since reserve according to a higher VaR is able to absorb the loss during crisis. In general, VaR forecast from the decomposition model is more efficient in that the predictive quantile loss is smaller and the empirical coverage probability is better for a given $\alpha$.

This argument is further supported by Figure 5. The VaR forecasts for $\alpha = 1\%$ are shown from four models: Riskmetrics, M1-ZM, M3-ZM-I, M3-ZM-C. The VaR forecasts from M2 were similar to those of M3 do not vary much in terms of figure, so we just select the two from M3-ZM as an representative, and other decomposition models will follow similarity. From the figure we can see decomposition models do generate more “conservative” VaR forecast in that their VaR forecasts are larger in absolute value than those from M1 and the Riskmetrics model for all confidence levels. The Riskmetrics model produces a small absolute value of VaR forecast. Therefore, in terms of 1% level VaR, where it captures the ability to cover loss from extreme negative return events from possible crisis, the decomposition model always works better than Riskmetrics since its conservative VaR forecast can efficiently account for those extreme events which is likely to happen in the financial market. However, for the larger confidence levels $\alpha = 5\%, 10\%$ for less extreme events, a conservative VaR forecast may not be appropriate since it costly to require large reserve fund for such less extreme events. For $\alpha = 5\%, 10\%$ (not reported for space), it is found that the decomposition model is not as good as Riskmetrics.

Tables 3, 4 also show that decomposition models improve more for lower $k$. For $k = 1$ and
3 months aggregated returns, decomposition models can always beat the benchmark model in terms of both predictive quantile loss and the empirical coverage probability. But for \( k = 12 \) month return, the difference is not that large. This is because when we aggregate the return, extreme negative returns become less rare since it is averaged out by other normal or positive returns and there will be no significant loss if one looks at the overall return. Thus the decomposition model loses its advantage of making a larger and more conservative VaR forecast. However, since the most common use of VaR is to give instructions for the reserve fund for possible extreme events like financial crisis on the daily or monthly return basis, the decomposition model is more useful compared to Riskmetrics in terms of its higher coverage probability and lower predictive quantile loss in the event of large loss happens.

7 Conclusion

The density and risk forecasts are based on the decomposition of the (excess) financial returns into its sign and modulus, and a copula function is used to model the dependence between the two components of the decomposition. This paper explores three empirical questions on the density forecast and the risk forecast: (a) Is the decomposition useful? (Is M2 better than M1?) (b) Is the shrinkage of imposing the independence constraint useful? (Is M2 using a copula better than M2-I?) (c) Is the shrinkage of imposing the ZM/HM moment constraints useful? (Is M3 better than M2?) The first question is to examine whether the density forecast from the decomposition model improves upon the traditional normal density forecast model and whether the risk forecast from the decomposition model can produce improved VaR forecasts relative to traditional risk forecasts from Riskmetrics and normal density forecast model. The second and the third questions are, noting that a simple model often beats a more sophisticated one for the out-of-sample forecasting, we examine possible benefits of shrinkage from imposing constraints. We consider two types of constraints: one is to impose independence between the two components of the decomposition and the other constraint is to impose the ZM or HM moment constraint.

We answer these questions from the empirical analysis using the monthly, quarterly, and annual stock returns and equity premium (in monthly frequency). The decomposition model
M2 (that incorporates DP2 and DP3) is better than normal model M1. As the sign and absolute return are found to be dependent, the decomposition model which incorporates the dependence works better. Imposing DP1 makes worse. Imposing moment constraints can improve density forecasts as long as such constraint is proper. Furthermore, the decomposition density forecast model produces better risk forecasts than Riskmetrics at lower tail at $\alpha = 1\%$ in terms of lower predictive check loss in the event of large loss happens. Thus the decomposition model has higher potential to absorb loss during crisis because the common use of VaR is to provide the required capital reserve against possible extreme events. The decomposition density forecast model seems to provide viable risk forecasts, alternative to Riskmetrics and normal density forecast models.

This paper is about how we improve density and risk forecasting with using a multiplicative decomposition and with imposing a conditional moment equality constraint. Several extensions can be considered. First, motivated by the recent paper by Ferreira and Santa-Clara (2011) who consider forecasting stock market returns using the additive decomposition of stock returns the convolution (density of the sum of the parts) of the component density of the parts of the return may be shown to be better than the density forecast of the whole, namely “the sum of the parts is more than the whole” in density forecasts. Second, unlike the equality constraint considered in this paper, we can consider various inequality constraints. For example, Campbell and Thompson (2008) and Lee, Tu, and Ullah (2013, 2014) consider imposing the inequality constraint on the parameter space. Also it will be interesting to see how the current paper may be extended to imposing the inequality constraint on the conditional moments such as $E(r_{t+1}|I_t) > 0$, $E(U_{t+1}V_{t+1}|I_t) > 0$, or the constraint that the conditional skewness $< 0$, as considered by Moon and Schorfheide (2009) in a different context. We plan to report these results in a separate paper.
References


Table 1. Test for Independence

<table>
<thead>
<tr>
<th></th>
<th>$Q_t(12)$</th>
<th>$Q_t(3)$</th>
<th>$Q_t(1)$</th>
<th>$R_t(12)$</th>
<th>$R_t(3)$</th>
<th>$R_t(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_P$</td>
<td>144.7009</td>
<td>158.6323</td>
<td>168.6566</td>
<td>210.8172</td>
<td>213.3336</td>
<td>168.9395</td>
</tr>
<tr>
<td>p-value</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
</tbody>
</table>

Note: Reported are the test statistic $\hat{T}_P$ in (19) for the null hypothesis of independence of $|r_t|$ and sign($r_t$). The asymptotic p-values are in brackets.
Table 2. Average Out-of-Sample Logarithmic Scores $\hat{S}_P$ for Density Forecasts

<table>
<thead>
<tr>
<th></th>
<th>$Q_t(12)$</th>
<th>$Q_t(3)$</th>
<th>$Q_t(1)$</th>
<th>$R_t(12)$</th>
<th>$R_t(3)$</th>
<th>$R_t(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1-ZM</td>
<td>-0.0399</td>
<td>0.8976</td>
<td>1.6613</td>
<td>-0.3840</td>
<td>0.8541</td>
<td>1.6544</td>
</tr>
<tr>
<td>M1-HM</td>
<td>-0.1333</td>
<td>0.8507</td>
<td>1.6560</td>
<td>-0.1309</td>
<td>0.8550</td>
<td>1.6595</td>
</tr>
<tr>
<td>M2-I</td>
<td>0.1763</td>
<td>1.0764</td>
<td>1.6611</td>
<td>0.0267</td>
<td>1.0757</td>
<td>1.6610</td>
</tr>
<tr>
<td>M2-F</td>
<td>0.6434</td>
<td>1.3557</td>
<td>1.7189</td>
<td>0.5209</td>
<td>1.3515</td>
<td>1.7093</td>
</tr>
<tr>
<td>M2-C</td>
<td>0.6493</td>
<td>1.3543</td>
<td>1.7178</td>
<td>0.5200</td>
<td>1.3529</td>
<td>1.7061</td>
</tr>
<tr>
<td>M2-FGM</td>
<td>0.6436</td>
<td>1.3558</td>
<td>1.7189</td>
<td>0.5232</td>
<td>1.3514</td>
<td>1.7093</td>
</tr>
<tr>
<td>M3-ZM-I</td>
<td>0.1582</td>
<td>1.3573</td>
<td>1.6609</td>
<td>0.1661</td>
<td>1.3258</td>
<td>1.6791</td>
</tr>
<tr>
<td>M3-ZM-F</td>
<td>0.6479</td>
<td>1.3591</td>
<td>1.7203</td>
<td>0.4918</td>
<td>1.3322</td>
<td>1.7031</td>
</tr>
<tr>
<td>M3-ZM-C</td>
<td>0.6535</td>
<td>1.3584</td>
<td>1.7196</td>
<td>0.4880</td>
<td>1.3351</td>
<td>1.6973</td>
</tr>
<tr>
<td>M3-ZM-FGM</td>
<td>0.6481</td>
<td>1.3591</td>
<td>1.7203</td>
<td>0.4949</td>
<td>1.3321</td>
<td>1.7031</td>
</tr>
<tr>
<td>M3-HM-I</td>
<td>0.0620</td>
<td>1.3516</td>
<td>1.6566</td>
<td>0.2679</td>
<td>1.3471</td>
<td>1.6844</td>
</tr>
<tr>
<td>M3-HM-F</td>
<td>0.6580</td>
<td>1.3513</td>
<td>1.7155</td>
<td>0.5998</td>
<td>1.3409</td>
<td>1.7095</td>
</tr>
<tr>
<td>M3-HM-C</td>
<td>0.6635</td>
<td>1.3504</td>
<td>1.7148</td>
<td>0.5975</td>
<td>1.3464</td>
<td>1.7037</td>
</tr>
<tr>
<td>M3-HM-FGM</td>
<td>0.6583</td>
<td>1.3513</td>
<td>1.7155</td>
<td>0.6019</td>
<td>1.3409</td>
<td>1.7094</td>
</tr>
</tbody>
</table>
Table 3. Average Out-of-Sample Check Loss $\hat{L}_P(\alpha)$ for VaR(0.01) Forecasts

<table>
<thead>
<tr>
<th></th>
<th>$Q_t(12)$</th>
<th>$Q_t(3)$</th>
<th>$Q_t(1)$</th>
<th>$R_t(12)$</th>
<th>$R_t(3)$</th>
<th>$R_t(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskMetrics</td>
<td>1.6993</td>
<td>0.9429</td>
<td>0.5946</td>
<td>1.9249</td>
<td>0.9783</td>
<td>0.5997</td>
</tr>
<tr>
<td>M1-ZM</td>
<td>0.9800</td>
<td>0.4382</td>
<td>0.1980</td>
<td>0.7569</td>
<td>0.4083</td>
<td>0.1858</td>
</tr>
<tr>
<td>M1-HM</td>
<td>1.1243</td>
<td>0.4902</td>
<td>0.2038</td>
<td>1.0250</td>
<td>0.4783</td>
<td>0.1973</td>
</tr>
<tr>
<td>M2-I</td>
<td>0.8533</td>
<td>0.2687</td>
<td>0.1816</td>
<td>0.8224</td>
<td>0.2671</td>
<td>0.1770</td>
</tr>
<tr>
<td>M2-F</td>
<td>0.8600</td>
<td>0.2692</td>
<td>0.1820</td>
<td>0.7662</td>
<td>0.2753</td>
<td>0.1764</td>
</tr>
<tr>
<td>M2-C</td>
<td>0.8538</td>
<td>0.2687</td>
<td>0.1818</td>
<td>0.8034</td>
<td>0.2645</td>
<td>0.1773</td>
</tr>
<tr>
<td>M2-FGM</td>
<td>0.8677</td>
<td>0.2693</td>
<td>0.1820</td>
<td>0.7561</td>
<td>0.2763</td>
<td>0.1764</td>
</tr>
<tr>
<td>M3-ZM-I</td>
<td>0.8341</td>
<td>0.2798</td>
<td>0.1805</td>
<td>0.8097</td>
<td>0.2735</td>
<td>0.1778</td>
</tr>
<tr>
<td>M3-ZM-F</td>
<td>0.7738</td>
<td>0.2716</td>
<td>0.1738</td>
<td>0.8567</td>
<td>0.2899</td>
<td>0.1705</td>
</tr>
<tr>
<td>M3-ZM-C</td>
<td>0.7737</td>
<td>0.2709</td>
<td>0.1745</td>
<td>0.8522</td>
<td>0.2917</td>
<td>0.1704</td>
</tr>
<tr>
<td>M3-ZM-FGM</td>
<td>0.7741</td>
<td>0.2716</td>
<td>0.1738</td>
<td>0.8601</td>
<td>0.2878</td>
<td>0.1705</td>
</tr>
<tr>
<td>M3-HM-I</td>
<td>0.9191</td>
<td>0.2780</td>
<td>0.1804</td>
<td>0.8738</td>
<td>0.2708</td>
<td>0.1765</td>
</tr>
<tr>
<td>M3-HM-F</td>
<td>1.0845</td>
<td>0.2754</td>
<td>0.1735</td>
<td>1.1466</td>
<td>0.2797</td>
<td>0.1758</td>
</tr>
<tr>
<td>M3-HM-C</td>
<td>1.0920</td>
<td>0.2761</td>
<td>0.1747</td>
<td>1.1298</td>
<td>0.2779</td>
<td>0.1775</td>
</tr>
<tr>
<td>M3-HM-FGM</td>
<td>1.0802</td>
<td>0.2754</td>
<td>0.1735</td>
<td>1.1520</td>
<td>0.2799</td>
<td>0.1758</td>
</tr>
</tbody>
</table>
Table 4. Average Out-of-Sample Coverage Probability $\hat{P}$ for VaR(1%) Forecasts

<table>
<thead>
<tr>
<th></th>
<th>$Q_t(12)$</th>
<th>$Q_t(3)$</th>
<th>$Q_t(1)$</th>
<th>$R_t(12)$</th>
<th>$R_t(3)$</th>
<th>$R_t(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RiskMetrics</td>
<td>0.7614</td>
<td>1.5228</td>
<td>1.7766</td>
<td>0.7614</td>
<td>1.5228</td>
<td>1.5228</td>
</tr>
<tr>
<td>M1-ZM</td>
<td>7.1066</td>
<td>4.0609</td>
<td>2.7919</td>
<td>3.5533</td>
<td>3.0457</td>
<td>2.5381</td>
</tr>
<tr>
<td>M1-HM</td>
<td>8.6294</td>
<td>5.3299</td>
<td>2.7919</td>
<td>6.5990</td>
<td>4.5685</td>
<td>3.0457</td>
</tr>
<tr>
<td>M2-I</td>
<td>4.8223</td>
<td>1.2690</td>
<td>1.2690</td>
<td>2.5381</td>
<td>1.2690</td>
<td>1.0152</td>
</tr>
<tr>
<td>M2-F</td>
<td>4.8223</td>
<td>1.5228</td>
<td>1.2690</td>
<td>2.7919</td>
<td>1.7766</td>
<td>1.2690</td>
</tr>
<tr>
<td>M2-C</td>
<td>4.8223</td>
<td>1.5228</td>
<td>1.2690</td>
<td>2.5381</td>
<td>1.5228</td>
<td>1.0152</td>
</tr>
<tr>
<td>M2-FGM</td>
<td>4.8223</td>
<td>1.5228</td>
<td>1.2690</td>
<td>2.7919</td>
<td>1.7766</td>
<td>1.2690</td>
</tr>
<tr>
<td>M3-ZM-I</td>
<td>4.5685</td>
<td>1.2690</td>
<td>1.2690</td>
<td>2.2843</td>
<td>1.0152</td>
<td>0.7614</td>
</tr>
<tr>
<td>M3-ZM-F</td>
<td>3.2995</td>
<td>0.5076</td>
<td>0.7614</td>
<td>1.0152</td>
<td>0.2538</td>
<td>0.5076</td>
</tr>
<tr>
<td>M3-ZM-C</td>
<td>3.2995</td>
<td>0.5076</td>
<td>0.7614</td>
<td>1.0152</td>
<td>0.2538</td>
<td>0.5076</td>
</tr>
<tr>
<td>M3-ZM-FGM</td>
<td>3.2995</td>
<td>0.5076</td>
<td>0.7614</td>
<td>1.0152</td>
<td>0.2538</td>
<td>0.5076</td>
</tr>
<tr>
<td>M3-HM-I</td>
<td>5.3299</td>
<td>1.5228</td>
<td>1.2690</td>
<td>2.7919</td>
<td>1.0152</td>
<td>0.7614</td>
</tr>
<tr>
<td>M3-HM-F</td>
<td>5.3299</td>
<td>1.2690</td>
<td>0.7614</td>
<td>4.3147</td>
<td>1.2690</td>
<td>1.0152</td>
</tr>
<tr>
<td>M3-HM-C</td>
<td>5.3299</td>
<td>1.2690</td>
<td>0.7614</td>
<td>4.5685</td>
<td>1.2690</td>
<td>1.0152</td>
</tr>
<tr>
<td>M3-HM-FGM</td>
<td>5.3299</td>
<td>1.2690</td>
<td>0.7614</td>
<td>4.3147</td>
<td>1.2690</td>
<td>1.0152</td>
</tr>
</tbody>
</table>
Figure 1. Plots of $I (\lambda_t)$ and $I (\lambda_t^*)$

(a) $I (\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.55$

(b) $I (\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.55$

(c) $I (\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.65$

(d) $I (\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.65$

(e) Optimal $I (\lambda_t^*)$ as a function of $p_{t+1}$

(f) Optimal $I (\lambda_t^*)$ and $\eta_t^*$ for some values of $p_{t+1}$

<table>
<thead>
<tr>
<th>$p_{t+1}$</th>
<th>$I_t (\lambda_t^*)$</th>
<th>$\eta_t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.51</td>
<td>0.9999</td>
<td>0.0004</td>
</tr>
<tr>
<td>0.55</td>
<td>0.9975</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.65</td>
<td>0.9770</td>
<td>0.0233</td>
</tr>
<tr>
<td>0.70</td>
<td>0.9583</td>
<td>0.0426</td>
</tr>
<tr>
<td>0.80</td>
<td>0.9000</td>
<td>0.1054</td>
</tr>
<tr>
<td>0.90</td>
<td>0.8000</td>
<td>0.2231</td>
</tr>
</tbody>
</table>

Notes: Panels (a,b,c,d) are plots of $I(\lambda_t)$ against $\lambda_t$ for fixed values of $\frac{1}{\psi_{t+1}}$ and $p_{t+1}$. Panels (a,b) are plots of $I(\lambda_t)$ against $\lambda_t$ for fixed values of $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.55$ which are the
estimated values for the annualized monthly equity premium $Q_t(12)$. Panel (b) magnifies Panel (a) for $-2 < \lambda_t < 2$. Panels (c,d) are plots of $I(\lambda_t)$ against $\lambda_t$ for fixed values of $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.65$ which are the estimated values for the annualized monthly stock returns $R_t(12)$. Panel (d) magnifies Panel (c) for $-2 < \lambda_t < 2$. Panel (e) is a plot of the optimal $I(\lambda^*)$ against $p_{t+1}$.
Figure 2. Estimated copula parameter $\theta$ over time

(a) $Q_t(12)$

(b) $Q_t(3)$

(c) $Q_t(1)$

(d) $R_t(12)$

(e) $R_t(3)$

(f) $R_t(1)$

Note: $\hat{\theta}$ for Frank copula is denoted in blue solid line, for Clayton copula in red dashed line, and for FGM copula in black dashed line.
Note: $\hat{\theta}$ for Frank copula is denoted in blue solid line, for Clayton copula in red dashed line, and for FGM copula in black dashed line.
Figure 4. Mean Forecasts from the Decomposition Models (Model 2) and HM Forecasts

(a) $Q_t(12)$

(b) $Q_t(3)$

(c) $Q_t(1)$

(d) $R_t(12)$

(e) $R_t(3)$

(f) $R_t(1)$

Note: The mean forecasts from M2-F using Frank copula (blue solid), from M2-C using Clayton copula (red dashed), and from M2-FGM using FGM copula (black dashed). Green line is the historical mean of real data.
Figure 5. Risk Forecasts: VaR(0.01) for $R_t (3)$

Note: Plotted are the four VaR($\alpha$) forecasts for $R_t (1)$: RiskMetrics (black dashed), M1-ZM (red dashed), M3-ZM using Independent copula (green), and M3-ZM using Clayton copula (blue).