# Density Forecast of Financial Returns Using Decomposition and Maximum Entropy*

Tae-Hwy Lee,† He Wang,‡ Zhou Xi,§ and Ru Zhang¶

September 14, 2021

**Abstract**

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 due to the multiplicative decomposition in estimation of the density forecast model, we impose a moment constraint that the conditional mean forecast is set to match with the sample mean. Imposing such a moment constraint operates a shrinkage and tilts the density forecast of the decomposition model to produce the improved maximum entropy density forecast. An empirical application to forecasting density of the daily stock returns demonstrates the benefits of using the decomposition and imposing the moment constraint to obtain the improved density forecast. We evaluate the density forecast by comparing the logarithmic score, the quantile score, and the continuous ranked probability score. We contribute to the literature on the density forecast and the decomposition models by showing that the density forecast of the decomposition model can be improved by imposing a sensible constraint in the maximum entropy framework.

*Key Words*: Decomposition, Copula, Moment constraint, Maximum entropy, Density forecast, Logarithmic score, Quantile score, VaR, Continuous ranked probability score.

*JEL Classification*: C1, C3, C5

*The authors would like to thank Raffaella Giacomini (editor) and the two anonymous referees, Amos Golan, Yongmiao Hong, Ivan Jeliazkov, Roger H. Moon, Hashem Pesaran, Aman Ullah, and seminar participants at Korea Institute of Finance, UC Irvine, UC Riverside, University of Southern California, Xiamen University, and Yonsei University, for many valuable comments.

†Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu.

‡Corresponding author. Department of Insurance, University of International Business and Economics, Beijing, China. E-mail: hewang72@gmail.com.

§Citigroup. E-mail: tedshady520zx@gmail.com.

¶JPMorgan Chase. E-mail: zhangruany@gmail.com.

# 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 (as discussed below in this section). 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 and imposing a simple moment constraint (for shrinkage via constrained maximization of an entropy). The improved density forecast is useful for financial risk management as it can produce a better risk forecast in the conditional high moments (variance, skewness) and conditional quantiles in terms of Value-at-Risk (VaR). While the decomposition allows us to model the two components in much richer specifications, the two predicted components can amplify the estimation errors as it 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). Controlling the latter would improve the density forecast from the decomposition model.

To do that, 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 historical mean (HM) forecast will give less estimation error than the more complex model, we use the maximum entropy principle for the out-of-sample density forecast subject to the HM moment constraint that is to match the mean forecasts from the decomposition model with the HM. When the mean forecast from the decomposition model deviates from the HM model, imposing the conditional mean constraint tilts the density forecast of the decomposition model and produce a more stable maximum entropy density forecast. The tilted density forecast would improve over the original (without tilting) density forecast of the decomposition model if the correct constraint is imposed. The underlying reason for the benefit of imposing the constraint is the "shrinkage" principle, that 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. Many 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 $\text{sign}(r_t) = \mathbf{1}(r_t > 0) - \mathbf{1}(r_t < 0)$. 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-day ahead return density forecast $f_t(r_{t+1})$ conditional on the information at time $t$, which can be obtained from the joint density $f_t(|r_{t+1}|, \text{sign}(r_{t+1}))$. 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 $\mathbf{1}(r_{t+1} < 0)$ displays conditional mean serial dependence, namely, if $E[\mathbf{1}(r_{t+1} < 0) | I_t]$ 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 $\sigma_{t+1}^2 = E\left[(r_{t+1} - \mu_{t+1})^2 | I_t\right]$ be the conditional variance. Then Christoffersen and Diebold (2006) note that

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

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 $\mathbf{1}(r_{t+1}<0)$ 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 referenced earlier 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 that the sign and the absolute return are not independent. We find DP1 is clearly rejected by Hong and Li (2005) test. If the dependence is weak, shrinkage toward the independence can benefit the density forecasts. We consider different copula function for the dependence of the sign of return and absolute value of the return, including the independent copula, and compare different density forecast models by a test in line with Diebold and Mariano (1995), Amisano and Giacomini (2007), Bao, Lee, and Saltoğlu (2007), Gneiting and Ranjan (2011), and Granziera, Hubrich, and Moon (2014).

The decomposition model may not generate satisfactory moment predictions even for forecasting 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 daily stock returns, the predicted conditional mean fluctuates rather excessively, often deviating unreasonably too far from the historical mean (HM) predictions. The plot of HMs of the daily stock returns over rolling windows (shifting one day forward with dropping the oldest day 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 functions will include 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 (2014) 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 improves the density forecast of the financial returns $\{r_{t+1}\}$.

The contribution of our paper is to show how to stabilize the decomposition model for the density forecast by imposing a moment condition in the maximum entropy framework. Anatolyev and Gospodinov (2010) develop the copula model for the decomposition model to model the dependence between $|r_t|$ and $\text{sign}(r_t)$, and examine the decomposition model for the mean forecasts of stock returns in terms of mean squared forecast errors or mean absolute forecast errors. We examine the decomposition model for the density forecast in terms of scoring rules (logarithmic score, quantile scores, and continuous ranked probability score). We show that the density forecast of the decomposition model can be improved in terms of these scoring rules by imposing a simple moment constraint by maximum entropy.

The rest of the paper is structured as follows. Section 2 introduces the normal density forecast model. Section 3 considers 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 function. In Section 4, we consider the maximum entropy decomposition density forecast with the mean moment constraint imposed. In Section 5 the scoring rules are discussed to evaluate the density forecast models and to compare their predictive abilities. Section 6 includes empirical results where the decomposition density forecast model and the maximum entropy density forecast with imposing a moment constraint are compared. Section 7 offers concluding remarks.

## 2 Normal Density Forecast Model

A simple density forecast of $r_{t+1}$ at time $t$ is the normal density forecast model

$$f_{t+1}(r) = \frac{1}{\sqrt{2\pi\sigma_{t+1}^2}} \exp\left\{-\frac{(r-\mu_{t+1})^2}{2\sigma_{t+1}^2}\right\}, \qquad r \in (-\infty \ \ +\infty), \tag{2}$$

where $\mu_{t+1} := E(r_{t+1}|I_t)$ and $\sigma_{t+1}^2 := E\left[(r_{t+1}-\mu_{t+1})^2 |I_t\right] = \gamma_0 + \gamma_1 (r_t - \mu_t)^2 + \gamma_2\sigma_t^2$. For the conditional mean specification, it is common to assume that $\mu_{t+1}$ is equal to the historical mean (HM). According to the weak form of efficient market hypothesis, the best forecast of the conditional mean is HM. Consider a linear predictive regression for the conditional mean

$$r_{t+1} = \alpha + x_t'\beta + \varepsilon_{t+1}. \tag{3}$$

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 Welch and Goyal (2008). Welch and Goyal (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. See also Campbell and Thompson (2008). We confirmed that the conditional mean forecast using a covariate performs worse than the 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 $\hat{\sigma}_{t+1}^2 = \hat{\gamma}_{0,t} + \hat{\gamma}_{1,t}(r_t - \bar{r}_t)^2 + \hat{\gamma}_{2,t}\hat{\sigma}_t^2$ is the predicted according to the GARCH(1,1) model using the estimated parameter values at time $t$. This normal density forecast model with $\hat{\mu}_{t+1} = \bar{r}_t$ (HM) will be referred to as Model 1 (or M1 in short).

## 3 Decomposition Density Forecast 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:

$$r_{t+1} = |r_{t+1}| \times \mathrm{sign}(r_{t+1}) =: U_{t+1}V_{t+1}, \tag{4}$$

where $U_{t+1} = |r_{t+1}|$ and $V_{t+1} = \text{sign}(r_{t+1})$. To get the density forecast of $r_{t+1}$, several assumptions about the joint and marginal distribution of $U_{t+1}$ and $V_{t+1}$ are needed.

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. We test for the independence between the sign and absolute value of the return with applying Hong and Li (2005). For our empirical applications, the independence is clearly rejected, which indicates the need to incorporate the dependence between them. The test procedure and results are available but not reported for space. One way to incorporate dependence between $U_{t+1}$ and $V_{t+1}$ is to include copula in the joint density function $f_{t+1}^{UV}(u,v)$, as

$$f_{t+1}^{UV}(u,v) = f_{t+1}^{U}(u) \times f_{t+1}^{V}(v) \times c\left(F_{t+1}^{U}(u), F_{t+1}^{V}(v)\right), \tag{5}$$

where $F_{t+1}^{U}(u) = \Pr\left(U \leq u | I_t\right)$ and $F_{t+1}^{V}(v) = \Pr\left(V \leq v | I_t\right)$. The third term $c\left(F_{t+1}^{U}(u), F_{t+1}^{V}(v)\right)$ is the copula density function such that $c(w_1, w_2) = \frac{\partial^2 C(w_1,w_2)}{\partial w_1 \partial w_2}$, where $w_1 = F_{t+1}^{U}(u)$ and $w_2 = F_{t+1}^{V}(v)$, and $C(\cdot,\cdot)$ is defined such that $F_{t+1}^{UV}(u,v) = C(F_{t+1}^{U}(u), F_{t+1}^{V}(v))$ is the joint cumulative distribution function (CDF). Moreover, the conditional copula (distribution) 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 $u = (F_{t+1}^{U})^{-1}(w_1)$ and $v = (F_{t+1}^{V})^{-1}(w_2)$ denote the inverse functions of the marginal CDFs of $U_{t+1}$ and $V_{t+1}$.

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,

$$U_{t+1} = \psi_{t+1} e_{t+1}, \tag{6}$$

where $\psi_{t+1} = E\left(U_{t+1} | I_t\right)$ is the conditional mean of the absolute return and $\{e_{t+1}\}$ is an i.i.d. positive random variable with $E\left(e_{t+1} | I_t\right) = 1$. 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 the Burr and gamma distributions 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:

$$\psi_{t+1} = \delta_0 + \delta_1 U_t + \delta_2 \psi_t, \tag{7}$$

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,

$$f_{t+1}^U(u) = \frac{1}{\psi_{t+1}} \exp\left(-\frac{1}{\psi_{t+1}} u\right), \qquad u > 0. \tag{8}$$

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

Next, the marginal density of the sign of the return $V_{t+1} = \mathbf{1}(r_{t+1} \geq 0) - \mathbf{1}(r_{t+1} < 0)$ 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:

$$\begin{aligned} f_{t+1}^V(v) &= \begin{cases} p_{t+1} & \text{if} \quad v = 1 \\ 1 - p_{t+1} & \text{if} \quad v = -1 \end{cases} \\ &= p_{t+1}^{\frac{v+1}{2}} (1 - p_{t+1})^{\frac{1-v}{2}}, \end{aligned} \tag{9}$$

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, $\hat{p}_{t+1} = \frac{1}{R} \sum_{s=t-R+1}^{t} \mathbf{1}(r_s > 0)$. 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 and $b = 0$. More complicated models can be used to estimate $p_{t+1}$. For instance, $G(\cdot)$ can be the standard normal CDF or the logistic function for probit or logit models. However, these more complicated models may not work better in out-of-sample forecast 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 $f_{t+1}^V(v) = \hat{p}_{t+1}^{\frac{v+1}{2}} (1 - \hat{p}_{t+1})^{\frac{1-v}{2}}$.

Due to the Bernoulli-type distribution (discrete with the bounded support) 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

$$f_{t+1}^{UV}(u,v) = f_{t+1}^{U}(u)\rho_{t+1}^{\frac{v+1}{2}}(1-\rho_{t+1})^{\frac{1-v}{2}}, \tag{10}$$

where

$$\rho_{t+1} = \rho_{t+1}(F_{t+1}^{U}(u)) = 1 - \partial C(F_{t+1}^{U}(u), 1-p_{t+1})/\partial w_1. \tag{11}$$

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, the joint density of $U$ and $V$ can be obtained from

$$\begin{aligned} f_{t+1}^{UV}(u,v) &= \frac{\partial F_{t+1}^{UV}(u,v)}{\partial u} - \frac{\partial F_{t+1}^{UV}(u,v-2)}{\partial u} \\ &= \frac{\partial C(F_{t+1}^{U}(u), F_{t+1}^{V}(v))}{\partial u} - \frac{\partial C(F_{t+1}^{U}(u), F_{t+1}^{V}(v-2))}{\partial u} \\ &= f_{t+1}^{U}(u)\left[\frac{\partial C(F_{t+1}^{U}(u), F_{t+1}^{V}(v))}{\partial w_1} - \frac{\partial C(F_{t+1}^{U}(u), F_{t+1}^{V}(v-2))}{\partial w_1}\right]. \end{aligned} \tag{12}$$

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

$$f_{t+1}^{UV}(u,v) = f_{t+1}^{U}(u)\left[\frac{\partial C(F_{t+1}^{U}(u), 1-p_{t+1})}{\partial w_1} - 0\right] = f_{t+1}^{U}(u)(1-\rho_{t+1}). \tag{13}$$

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

$$f_{t+1}^{UV}(u,v) = f_{t+1}^{U}(u)\left[1 - \frac{\partial C(F_{t+1}^{U}(u), 1-p_{t+1})}{\partial w_1}\right] = f_{t+1}^{U}(u)\rho_{t+1}. \tag{14}$$

Putting these together yields the expression in (10).

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 in Eq (11) 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

$$
\begin{aligned}
C^{\text{Frank}}(w_1, w_2, \theta) &= -\frac{1}{\theta} \log \left( 1 + \frac{(e^{-\theta w_1} - 1)(e^{-\theta w_2} - 1)}{(e^{-\theta} - 1)} \right), \ \theta \in (-\infty, +\infty) \backslash \{0\}, \\
c^{\text{Frank}}(w_1, w_2, \theta) &= \frac{\theta(1 - e^{-\theta}) e^{-\theta(w_1 + w_2)}}{[(1 - e^{-\theta}) - (1 - e^{-\theta w_1})(1 - e^{-\theta w_2})]^2}, \\
\rho_{t+1}^{\text{Frank}}(w_1, p_{t+1}, \theta) &= \left( 1 - \frac{1 - e^{-\theta(1 - p_{t+1})}}{1 - e^{\theta p_{t+1}}} e^{\theta(1 - w_1)} \right)^{-1}.
\end{aligned}
$$

(3) Clayton copula

$$
\begin{aligned}
C^{\text{Clayton}}(w_1, w_2, \theta) &= \left( w_1^{-\theta} + w_2^{-\theta} - 1 \right)^{-1/\theta}, \ \theta \in [-1, +\infty) \backslash \{0\}, \\
c^{\text{Clayton}}(w_1, w_2, \theta) &= \frac{(1 + \theta)(w_1^{-\theta} + w_2^{-\theta} - 1)^{-\frac{1}{\theta} - 2}}{(w_1 w_2)^{\theta + 1}}, \\
\rho_{t+1}^{\text{Clayton}}(w_1, p_{t+1}, \theta) &= 1 - \left( 1 + \frac{(1 - p_{t+1})^{-\theta} - 1}{w_1^{-\theta}} \right)^{-1/\theta - 1}.
\end{aligned}
$$

(4) Farlie-Gumbel-Morgenstern copula (FGM)

$$
\begin{aligned}
C^{\text{FGM}}(w_1, w_2, \theta) &= w_1 w_2 \left( 1 + \theta(1 - w_1)(1 - w_2) \right), \ \theta \in [-1, 1], \\
c^{\text{FGM}}(w_1, w_2, \theta) &= 1 + \theta - 2\theta w_1 - 2\theta w_2 + 4\theta w_1 w_2, \\
\rho_{t+1}^{\text{FGM}}(w_1, p_{t+1}, \theta) &= 1 - (1 - p_{t+1})(1 + \theta p_{t+1}(1 - 2w_1)).
\end{aligned}
$$

Notice that Frank copula and FGM copula are symmetric copula, while Clayton copula is asymmetric copula capturing the lower tail dependence. For Frank and FGM copula, $\theta < 0$ implies negative dependence and $\theta > 0$ implies positive dependence. For Clayton copula and Frank 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 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:

$$
\begin{aligned}
\hat{\theta} &= \arg\max_{\theta} \sum_{t=1}^{n} \log \left( f_{t+1}^U(u) \ \left[ \rho_{t+1}(F_{t+1}^U(u), \theta) \right]^{\frac{v+1}{2}} \ \left[ 1 - \rho_{t+1}(F_{t+1}^U(u), \theta) \right]^{\frac{1-v}{2}} \right) \qquad (15) \\
&= \arg\max_{\theta} \sum_{t=1}^{n} \log f_{t+1}^U(u) + \frac{v+1}{2} \log \rho_{t+1}(F_{t+1}^U(u), \theta) + \frac{1-v}{2} \log \left[ 1 - \rho_{t+1}(F_{t+1}^U(u), \theta) \right].
\end{aligned}
$$

Since the marginal density $f^U_{t+1}(u)$ 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 $\hat{F}^U_{t+1}(u)$, and then in the second step we get the MLE of $\theta$ by

$$\hat{\theta} = \arg\max_{\theta} \sum_{t=1}^{n} \frac{v+1}{2} \log(\rho_{t+1}(\hat{F}^U_{t+1}(u), \theta)) + \frac{1-v}{2} \log\left[1 - \rho_{t+1}(\hat{F}^U_{t+1}(u), \theta)\right]. \tag{16}$$

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).

The joint density forecast $f^{UV}_{t+1}(u,v)$ of $U_{t+1}$ and $V_{t+1}$ will be called the decomposition model. The decomposition model using a different copula function is referred to as Model 2 (or M2 in short). In particular, Model 2 with Independent copula is referred to as Model 2-I, Model 2 with Frank copula as Model 2-F, Model 2 with Clayton copula as Model 2-C, and Model 2 with Farlie-Gumbel-Morgenstern copula as Model 2-FGM.

# 4 Maximum Entropy Decomposition Density Forecast Model

Previous studies, such as Anatolyev and Gospodinov (2010), did not consider the possible problem that the decomposition model may not generate satisfactory moment predictions even for the first moment (mean). The possible problem with the joint density forecast model using the decomposition is that the mean prediction $E\left(U_{t+1}V_{t+1}|I_t\right)$ from the joint density forecast function $f^{UV}_{t+1}\left(u,v\right)$ of the decomposition model (Model 2) may not be equal to $E\left(r_{t+1}|I_t\right)$ from Model 1, even if $r_{t+1} \equiv U_{t+1}V_{t+1}$ is an identity. In fact, 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. Ignoring this problem could amplify the estimation errors as it makes the conditional moment forecasts be subjected to multiplicative estimation errors which are of higher order of magnitude. The mean prediction $E\left(U_{t+1}V_{t+1}|I_t\right)$ can deviate from the mean forecast of Model 1 for which HM is used for $E\left(r_{t+1}|I_t\right)$. In other words, the estimated decomposition model may not satisfy the mean moment condition $E\left(U_{t+1}V_{t+1} - r_{t+1}|I_t\right) = 0$. To solve this problem, imposing the conditional mean constraint tilts the density forecast of the decomposition model and improves over the original (without tilting) density forecast of the decomposition model. Imposing such a moment constraint operates a shrinkage and tilts the density forecast of the decomposition model to produce the improved maximum entropy density forecast. With this in mind, we want to impose the (conditional) moment constraint that

$$E\left(U_{t+1}V_{t+1}|I_t\right) = \mu_{t+1}. \tag{17}$$

We consider the HM for $\mu_{t+1}$.

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 $f_{t+1}^{UV}(u,v)$:

$$\max_{h_{t+1}^{UV}(u,v)} = -\int\int\left(\log\frac{h_{t+1}^{UV}(u,v)}{f_{t+1}^{UV}(u,v)}\right)h_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v \tag{18}$$

$$\text{subject to} \quad \int\int m_t(u,v)h_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v = 0, \tag{19}$$

$$\text{and} \quad \int\int h_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v = 1, \tag{20}$$

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 maximum entropy (ME) density forecast satisfying the moment constraint of (19). The moment constraint in (17) is rewritten as (19) with

$$m_t(u,v) = uv - \mu_{t+1}. \tag{21}$$

Note that the expectation in (17) 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 (19) with (21) will make the new joint density forecast $h_{t+1}^{UV}(u,v)$ have the same mean forecast $\mu_{t+1}$ as Model 1.

The maximization of (18) subject to the moment constraint (19) 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

$$h_{t+1}^{UV}(u,v) = f_{t+1}^{UV}(u,v)\exp\left[\eta_t^* + \lambda_t^* m_t(u,v)\right], \tag{22}$$

where

$$\lambda_t^* = \arg\min_{\lambda_t} I_t(\lambda_t), \tag{23}$$

$$\eta_t^* = -\log I_t(\lambda_t^*), \tag{24}$$

$$I_t(\lambda_t) = \int\int \exp\left[\lambda_t m_t(u,v)\right] f_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v. \tag{25}$$

That is, to get the solution $h_{t+1}^{UV}(u,v)$, we construct 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 in particular, Giacomini and Ragusa (2014).

Note that the objective function of (18) is the (negative) *conditional* Kullbeck-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,

$$\begin{aligned} KLIC(h_{t+1}^{UV}, f_{t+1}^{UV}) &= \int\int \left(\log \frac{h_{t+1}^{UV}(u,v)}{f_{t+1}^{UV}(u,v)}\right) h_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v \\ &= \int\int \log\exp\left[\eta_t + \lambda_t m_t(u,v)\right] h_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v \\ &= \eta_t \int\int h_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v + \lambda_t \int\int m_t(u,v) h_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v \\ &= \eta_t. \end{aligned} \tag{26}$$

Note that $\eta_t = KLIC(h_{t+1}^{UV}, f_{t+1}^{UV})$ measures the gain from $h_{t+1}^{UV}$ (Model 3) over $f_{t+1}^{UV}$ (Model 2) in terms of the logarithmic score (33). The gain is non-negative since $\eta_t \geq 0$ as shown in (34). 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 $I_t(\lambda_t)$, 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 (25)

$$I_t(\lambda_t) = E_t \exp[\lambda_t m_t(u,v)], \tag{27}$$

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 (10), the conditional density of $V_{t+1}$ conditioning on $U_{t+1}$ is given by:

$$f_{t+1}^{V|U}(v|u) = \frac{f_{t+1}^{UV}(u,v)}{f_{t+1}^{U}(u)} = \rho_{t+1}^{\frac{1+v}{2}}(1-\rho_{t+1})^{\frac{1-v}{2}}, \qquad v \in \{-1, 1\}. \tag{28}$$

The decomposition model (Model 2) with the HM moment constraint $\hat{\mu}_{t+1} = \bar{r}_t$ imposed will be called Model 3 (or M3 in short). For each copula function, we will denote the model using the name of copula, e.g., Model 3-I (or M3-I) with Independent copula. The labels for other copula functions are made similarly. Model 2 is the original decomposition density forecast model $f_{t+1}^{UV}(u,v)$, while Model 3 is the tilted maximum entropy density forecast model $h_{t+1}^{UV}(u,v)$. We will compare Model 2 and Model 3 to examine if imposing the HM moment constraint can improve the density forecast of the decomposition model.

**Remark 1.** If the relative entropy between $h^{UV}$ (Model 3) and $f^{UV}$ (Model 2) is big, that means the moment constraint is not satisfied by the decomposition model density $f^{UV}$. This can arise from misspecifications in the marginal models for the components $(U,V)$ or their copula.

We consider the ACD-like model for $f^U$ in Eq (7) with the exponential density in Eq (8), and the marginal model for $V$ is $f^V$ in Eq (9) with $\Pr(V_{t+1} = 1|I_t) = \Pr(V_{t+1} = 1)$. Hence, it is possible that the mean $E(r_{t+1}|I_t)$ from Model 1 and the mean $E(U_{t+1}V_{t+1}|I_t)$ from Model 2 may not be the same even though $r_{t+1} \equiv U_{t+1}V_{t+1}$ is an identity. The moment condition is simply $E(r_{t+1} - U_{t+1}V_{t+1}|I_t) = 0$. Most empirical finance literature on the return prediction shows that HM is a hard-to-beat benchmark for the mean forecast, e.g., Welch and Goyal (2008), Campbell and Thompson (2008). Hence, we believe $E(r_{t+1}|I_t) = HM$ is a good model for the mean and thus we impose the constraint that $E(U_{t+1}V_{t+1}|I_t) = HM$. The benefit of the moment condition is when the mean of $f^{UV}$ does not satisfy it. In such case, $f^{\ UV}$ is tilted to satisfy the moment condition. The resulting ME density $h^{UV}$ can generate not only the more stable mean forecast (HM) but also the better density forecast in the logarithmic score (since KLIC is non-negative).

**Remark 2.** If the HM mean constraint is believed to be true and it is not satisfied by the mean of the density forecast from the decomposition model, the marginal models and copula in the decomposition model may not be fully correctly specified. Solutions may be to modify the ACD-like model in Eq (7) to include the leverage effect or the exogenous predictors, to modify Eq (8) for the marginal density forecast of $U$ using the more flexible density instead of the exponential density, to modify Eq (9) and the link function in $\Pr(V_{t+1} = 1|I_t) = G(a + x_t'b)$ as mentioned earlier, to use nonparametric marginal densities of $U, V$, or to use a nonparametric copula function in Eq (5). However, in this paper, our interest is to see how imposing a simple moment condition can improve the density forecast of the decomposition model.

**Remark 3.** If we apply more moment constraints, the maximum entropy density can improve the density of the decomposition model even more. We have not considered different moment constraints other than the mean constraint. What we have done is to match the first moment. If we go for matching higher order moment conditions such as variance, skewness, kurtosis, and etc, it would improve the density forecast even more. If we go for imposing conditions that are implied from economic theory, as considered by Giacomini and Ragusa (2014), it could also improve the density forecasts.

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 $\hat{\mu}_{t+1} = 0$. In this case, the analytical expression of $I_t(\lambda_t)$ is obtained as follows:

$$
\begin{aligned}
I_t(\lambda_t) &= \int\int \exp[\lambda_t m(u,v)] f_{t+1}^{UV}(u,v)\mathrm{d}u\mathrm{d}v \\
&= \int\int_{v=-1} f_{t+1}^{UV}(u,v)\exp[\lambda_t m_t(u,v)]\mathrm{d}v\mathrm{d}u + \int\int_{v=1} f_{t+1}^{UV}(u,v)\exp[\lambda_t m_t(u,v)]\mathrm{d}v\mathrm{d}u \\
&= \int f(u,-1)\exp[\lambda_t m(u,-1)]\mathrm{d}u + \int f(u,1)\exp[\lambda_t m(u,1)]\mathrm{d}u \\
&= \int_0^\infty \frac{1}{\psi_{t+1}}\exp(-\frac{1}{\psi_{t+1}}u)(1-p_{t+1})\exp(-\lambda_t u)\mathrm{d}u + \int_0^\infty \frac{1}{\psi_{t+1}}\exp(-\frac{1}{\psi_{t+1}}u)p_{t+1}\exp(\lambda_t u)\mathrm{d}u \\
&= \frac{\frac{1}{\psi_{t+1}}(1-p_{t+1})}{\frac{1}{\psi_{t+1}}+\lambda_t} + \frac{\frac{1}{\psi_{t+1}}p_{t+1}}{\frac{1}{\psi_{t+1}}-\lambda_t} \qquad (29)
\end{aligned}
$$

A plot of $I_t(\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_t = 0.55$ is given in Figure 1(a). A plot of $I_t(\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_t = 0.65$ is given in Figure 1(c). 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 $I_t(\lambda_t) = \frac{1}{S}\sum_{s=1}^S \exp[\lambda_t m_t(u_t^s, v_t^s)]$ 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

$$
\frac{\mathrm{d}I_t(\lambda_t)}{\mathrm{d}\lambda_t} = -\frac{1-p_{t+1}}{(\frac{1}{\psi_{t+1}}+\lambda_t)^2} + \frac{p_{t+1}}{(\frac{1}{\psi_{t+1}}-\lambda_t)^2} = 0. \qquad (30)
$$

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

$$
\lambda_t^* = \frac{\frac{1}{\psi_{t+1}}\left(-1+2\sqrt{p_{t+1}(1-p_{t+1})}\right)}{2p_{t+1}-1}, \qquad (31)
$$

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 1(a) and Figure 1(c), these figures are magnified into Figure 1(b) and Figure 1(d) 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

$$
I_t(\lambda_t^*) = \frac{(1-p_{t+1})(2p_{t+1}-1)}{2p_{t+1}-2+2\sqrt{p_{t+1}(1-p_{t+1})}} + \frac{p_{t+1}(2p_{t+1}-1)}{2p_{t+1}-2\sqrt{p_{t+1}(1-p_{t+1})}}. \qquad (32)
$$

It is interesting to see that $I_t(\lambda_t^*)$ does not depend on $\psi_{t+1}$ but depends only on $p_{t+1}$.[1] Figure 1(e) 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 (31) is not defined for $p_{t+1} = 0.5$, the limiting value $\lim_{p_{t+1} \to 0.5} I_t(\lambda_t^*) = 1$ as shown in Figure 1(e), 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). As tabulated in Figure 1(f), the gain $\eta_t^*$ in the new ME density $h^{UV}$ increases as $p_{t+1}$ deviates from 0.5.

# 5 Evaluation of Density Forecast Models

To evaluate different density forecast models (Model 1, Model 2, Model 3), a *proper* scoring rule can be compared. A scoring rule $S(f, y)$ of the density forecast $f$ is a real value evaluated at the realized value $y$ of a random variable $Y$. A score rule is positive-oriented (negative-oriented) if it is to be maximized (minimized). A larger expected positive-oriented (negative-oriented) scoring rule means that the associated density forecast is better (worse). For a proper positive-oriented scoring rule $S(f, y)$, the expected score of the true density is always greater than the expected score of any other density. A positive-oriented scoring rule is proper if $E_f S(f, Y) \geq E_f S(g, Y)$ for all densities $f$ and $g$, where $E_f S(f, Y) = \int S(f, y) f(y) \mathrm{d}y$ is the expected score value of $S(f, y)$ under the density function $f$, and $E_f S(g, Y) = \int S(g, y) f(y) \mathrm{d}y$ is the expected score value of $S(g, y)$ under the density function $f$. If the equality holds only if $f = g$, then the score function is strictly proper. If the true conditional density of $Y$ is $f$, then the density forecast $f$ is ideal. A negative-oriented scoring rule is proper if $E_f S(f, Y) \leq E_f S(g, Y)$. See Gneiting and Raftery (2007) and Gneiting and Ranjan (2011). In this section, we consider three proper scoring rules: logarithmic score (LS), quantile score (QS), and continuous ranked probability score (CRPS). LS is positive-oriented, while QS and CRPS are negative-oriented.

[1] This is due to the particular moment condition $m_t(u, v) = uv - \mu_{t+1}$ and $\mu_{t+1} = 0$ (21) 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 $p_{t+1}$.

### 5.1 Logarithmic Score

The logarithmic score (LS) is defined as

$$LS(f, y) \equiv \log f(y). \tag{33}$$

The difference of the expected logarithmic scores is the KLIC divergence measure. For example, when comparing the performance of $h^{UV}$ (Model 3) and $f^{UV}$ (Model 2), the difference of the expected log scores of $h$ and $f$ is

$$KLIC(h, f) = E_h \left[LS(h, Y) - LS(f, Y)\right] = E_h \left[\log h\left(Y\right) - \log f\left(Y\right)\right], \tag{34}$$

where $Y = (U, V)$. $KLIC(h, f) \geq 0$ due to the Jensen's inequality applied to the logarithmic function which is concave. See Rao (1965), White (1994), and Ullah (1996). To estimate the logarithmic score, we use the sample of the total number of observations $T$, divided into $R$ in-sample observations and $P$ out-of-sample observations ($T = R + P$). For in-sample (IS) estimation, the expected logarithmic score $E\left[LS(f, Y)\right]$, for $f$ being the density from Model 1, Model 2, or Model 3, is estimated by

$$LS_{IS} = \frac{1}{R}\sum_{t=1}^{R} LS_t, \tag{35}$$

where $LS_t = LS\left(f_t, y_t\right)$ is the logarithmic score of the density estimated at time $t$, and evaluated at the realized value $y_t$ at time $t$ (where $y_t = r_t$ for Model 1 and $y_t = (u_t, v_t)$ for Models 2, 3). For out-of-sample (OOS) forecasting, the expected logarithmic score is estimated by the out-of-sample average of predicted logarithmic scores

$$LS_{OOS} = \frac{1}{P}\sum_{t=R}^{T-1} LS_{t+1}, \tag{36}$$

where $LS_{t+1} = LS\left(f_{t+1}, y_{t+1}\right)$ is the predicted logarithmic score of the one-period ahead density forecast $f_{t+1}$ made at time $t$ and evaluated at the realized value $y_{t+1}$ at time $t+1$. The out-of-sample forecast evaluation period is for $t = R, \ldots, T-1$, with the total $P$ observations in the OOS. The density forecast with a higher value of $LS$ is the better density forecast.

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

$$LS(f_{t+1}, y_{t+1}) = LS\left(f_{t+1}^{UV}, (u_{t+1}, v_{t+1})\right) = \log f_{t+1}^{UV}\left(u_{t+1}, v_{t+1}\right). \tag{37}$$

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

$$LS\left(f_{t+1}^{UV}, (u_{t+1}, v_{t+1})\right) = \log f_{t+1}^{U}\left(u_{t+1}\right) + \frac{1+v_{t+1}}{2}\log\rho_{t+1} + \frac{1-v_{t+1}}{2}\log(1-\rho_{t+1}). \tag{38}$$

We compare density forecasts by the average out-of-sample logarithmic scores in (36) where $LS_t$ is the logarithmic score $LS\left(f_{t+1}^{UV}, (u_{t+1}, v_{t+1})\right)$ 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 $t+1$.

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

$$f_{t+1}(r) = \sum_{v\in\{-1,1\}} f_{t+1}^{UV}\left(u, v\right) = f_{t+1}^{UV}\left(-r, -1\right) + f_{t+1}^{UV}\left(r, 1\right) \tag{39}$$

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 $f_{t+1}^{UV}\left(u_{t+1}, v_{t+1}\right)$

$$\frac{1}{P}\sum_{t=R}^{T-1}\log f_{t+1}\left(r_{t+1}\right) = \frac{1}{P}\sum_{t=R}^{T-1}\log f_{t+1}^{UV}\left(u_{t+1}, v_{t+1}\right). \tag{40}$$

So we can actually compare Model 1 and Model 2 via the scoring rule of the joint density of $U$ and $V$. In Section 6 for empirical applications, the results using the logarithmic scores $LS_{IS}, LS_{OOS}$ are presented in Table 1.

## 5.2 Value-at-Risk Forecasts and Quantile Score

An important application of the density forecast models is to make risk forecasts. Since a "better" density forecast is 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 necessarily be a better 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. As reported later, we find the density forecasts from Model 2 generates higher average logarithmic score 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 the support away from the mean, possibly from tails.

Therefore we examine if the decomposition model's density and the maximum entropy density can also generate superior density forecast in tails. We focus on Value-at-Risk (VaR), the quantiles of the density forecasts. To be specific, we will invert the conditional distribution functions of the density forecasts to obtain the conditional quantile forecasts, namely, $\mathrm{VaR}_{t+1}(\alpha) = F_{t+1}^{-1}(\alpha)$, for a given tail probability $\alpha \in (0,1)$. 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 $\mathrm{VaR}_{t+1}(\alpha)$ with $\alpha = 0.01$.

The normal density forecast (M1) 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 $\hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\Phi^{-1}(\alpha)$ where $\Phi(\cdot)$ is the CDF of the standard normal distribution. For the decomposition models M2 and M3, it is difficult to derive the VaR analytically as there is no explicit solution, 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 $h_{t+1}^{U}(u)$, and then generate $V_{t+1}$ based on the conditional density of $h^{V|U}(v|u)$, and since $V_{t+1}$ is binary, the conditional density should also be Bernoulli-type. Substitute (10) and (21) to (22), the joint density of $U_{t+1}$ and $V_{t+1}$ under moment constraint can be written as:

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

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:

$$\begin{aligned}
h^U_{t+1}(u) &= h^{UV}_{t+1}(u,-1) + h^{UV}_{t+1}(u,1) \\
&= f^U_{t+1}(u)(1-\rho_{t+1})\exp(\eta^*_t - \lambda^*_t\hat{\mu}_{t+1} - \lambda^*_t u) + f^U_{t+1}(u)\rho_{t+1}\exp(\eta^*_t - \lambda^*_t\hat{\mu}_{t+1} + \lambda^*_t u) \\
&= \exp(\eta^*_t - \lambda^*_t\hat{\mu}_{t+1}) f^U_{t+1}(u)[(1-\rho_{t+1})\exp(-\lambda^*_t u) + \rho_{t+1}\exp(\lambda^*_t u)],
\end{aligned}$$

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

$$\begin{aligned}
h^{V|U}_{t+1}(v|u) &= \frac{h^{UV}_{t+1}(u,v)}{h^U_{t+1}(u)} \\
&= \frac{\rho_{t+1}^{\frac{1+v}{2}}(1-\rho_{t+1})^{\frac{1-v}{2}}\exp(\lambda^*_t uv)}{(1-\rho_{t+1})\exp(-\lambda^*_t u) + \rho_{t+1}\exp(\lambda^*_t u)}
\end{aligned}$$

However, it is not always easy to generate $u$'s from the $h^U_{t+1}(u)$ 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 $h^U_{t+1}(u)$ is to use the probability integral transformation (PIT). Since we assume that $u$ follows an exponential distribution with mean equal to $\psi_{t+1}$, we have $f^U_{t+1}(u) = \frac{1}{\psi_{t+1}}\exp\left(-\frac{1}{\psi_{t+1}}u\right)$. Substituting this expression into the marginal distribution $h^U_{t+1}(u)$, we get

$$h^U_{t+1}(u) = \exp(\eta^*_t - \lambda^*_t\hat{\mu}_{t+1})\frac{1}{\psi_{t+1}}\exp\left(-\frac{1}{\psi_{t+1}}u\right)[(1-\rho_{t+1})\exp(-\lambda^*_t u) + \rho_{t+1}\exp(\lambda^*_t u)],$$

and the CDF

$$\begin{aligned}
H^U_{t+1}(u) &= \int h^U_{t+1}(u)du \\
&= \exp(\eta^*_t - \lambda^*_t\mu_{t+1})\frac{1}{\psi_{t+1}}\frac{1-\rho_{t+1}}{\frac{1}{\psi_{t+1}}+\lambda^*_t}\left[1-\exp\left(-\left(\frac{1}{\psi_{t+1}}+\lambda^*_t\right)u\right)\right] \\
&+ \exp(\eta^*_t - \lambda^*_t\mu_{t+1})\frac{1}{\psi_{t+1}}\frac{\rho_{t+1}}{\frac{1}{\psi_{t+1}}-\lambda^*_t}\left[1-\exp\left(-\left(\frac{1}{\psi_{t+1}}-\lambda^*_t\right)u\right)\right].
\end{aligned}$$

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, $H^U_{t+1}(u)$ 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 $i = 1, 2, \ldots$ generate $U_i \sim \text{Uniform}(0,1)$ and $X_i \sim f_X(x)$. Calculate

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

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

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 $u \sim h_{t+1}^U(u)$, $h_{t+1}^U$ is the target density and $f_{t+1}^U$ is the candidate-generating density. Since it is easy to generate $u \sim f_{t+1}^U(u)$, we will set $u \sim f_{t+1}^U(u)$ as the $X$ variable above, and $u \sim h_{t+1}^U(u)$ to be the $Y$ variable above, and then $u \sim h_{t+1}^U(u)$ 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.

To compare the VaR forecasts from different density forecast models, we define the quantile score (QS) from the "check function" of Koenker and Bassett (1978). The QS of a quantile $F^{-1}(\alpha)$ for a given probability level $\alpha \in (0\ 1)$ is defined as

$$QS(f, y; \alpha) \equiv \left[\alpha - \mathbf{1}\left(y \leq F^{-1}(\alpha)\right)\right]\left(y - F^{-1}(\alpha)\right). \tag{41}$$

The QS is a negative-oriented scoring rule, i.e., a lower QS means that the associated density forecast is better for the VaR forecasts. Saerens (2000), Bertail, Haefke, Politis and White (2004), Komunjer (2005), Bao, Lee and Saltoğlu (2006), and Gneiting (2011) show that the QS or the check loss function can be regarded as a (negative) quasi-likelihood, therefore QS can provide a measure of the lack-of-fit of a quantile model.

Once we obtained a VaR, $F^{-1}(\alpha)$, by the numerical method discussed above, we plug it to the above expression to compute QS. For in-sample (IS) estimation, the expected quantile score

$E\left[QS(f,Y;\alpha)\right]$, for $f$ being the density from Model 1, Model 2, or Model 3, is estimated by

$$QS_{IS}\left(\alpha\right)=\frac{1}{R}\sum_{t=1}^{R}QS_{t}\left(\alpha\right), \tag{42}$$

where $QS_t\left(\alpha\right)=QS\left(f_t,y_t;\alpha\right)$ is the quantile score of the density estimated at time $t$ and evaluated at the realized value $y_t$ at time $t$ (where $y_t=r_t$ for Model 1 and $y_t=(u_t,v_t)$ for Models 2, 3). For out-of-sample (OOS) forecast, the predictive QS is estimated by computing the average out-of-sample check loss

$$QS_{OOS}\left(\alpha\right)=\frac{1}{P}\sum_{t=R}^{T-1}QS_{t+1}\left(\alpha\right), \tag{43}$$

where $QS_{t+1}\left(\alpha\right)=QS\left(f_{t+1},y_{t+1};\alpha\right)$ is the quantile score of the one-period ahead density forecast $f_{t+1}$ made at time $t$ and evaluated at the realized value $y_{t+1}$ at time $t+1$. For example, for Model 2, $QS_t\left(\alpha\right)=QS\left(f_{t+1}^{UV},(u_{t+1},v_{t+1});\alpha\right)$ with the density forecast $f_{t+1}^{UV},(u_{t+1},v_{t+1})$ made at time $t$ and evaluated at the realized absolute return $u_{t+1}$ and the realized sign $v_{t+1}$ at time $t+1$. In Section 6 for empirical applications, the results using the quantile scores $QS_{IS}\left(\alpha\right),QS_{OOS}\left(\alpha\right)$ are presented in Table 2.

Besides the quantile score, the VaR's from different density models can be evaluated by the coverage probability. See Corradi, Fosten, and Gutknecht (2020). Given the quantile $F^{-1}\left(\alpha\right)$, the in-sample and out-of-sample coverage probabilities are estimated by

$$\begin{aligned}\hat{\alpha}_{IS}&=\frac{1}{R}\sum_{t=1}^{R}\left[\mathbf{1}(y_t<F_t^{-1}\left(\alpha\right))\right], \\ \hat{\alpha}_{OOS}&=\frac{1}{P}\sum_{t=R}^{T-1}\left[\mathbf{1}(y_{t+1}<F_{t+1}^{-1}\left(\alpha\right))\right].\end{aligned} \tag{44}$$

A density model is more desirable if these empirical coverage probabilities of its quantiles are closer to the true coverage probabilities. In Section 6 for empirical applications, the results for $\hat{\alpha}_{IS},\hat{\alpha}_{OOS}$ are presented in Table 3.

## 5.3 Continuous Ranked Probability Score

The continuous ranked predictive score (CRPS; Matheson and Winkler 1976; Gneiting and Raftery 2007; Gneiting and Ranjan 2011) is a negative-oriented scoring rule to evaluate density forecast $f$. Let the probability score, $PS\left(f,z\right)=\left[\int_{-\infty}^{z}f(y)\mathrm{d}y-\mathbf{1}\left(y\leq z\right)\right]^2$, be the quadratic distance between a binary event $\mathbf{1}\left(y\leq z\right)$ and its expectation $E\left[\mathbf{1}\left(y\leq z\right)\right]=\Pr\left(y\leq z\right)=\int_{-\infty}^{z}f(y)\mathrm{d}y=$

$F(z)$ which is the corresponding CDF of the density $f(y)$. CRPS is defined as the integration of $PS(f,z)$ over $z \in \mathbb{R}$ :

$$CRPS(f,y) \equiv \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{z} f(y)\mathrm{d}y - \mathbf{1}(y \leq z) \right]^2 \mathrm{d}z. \tag{45}$$

Gneiting and Ranjan (2011) note that CRPS is equivalently expressed in terms of the quantile score in (41):

$$CRPS(f,y) = 2\int_0^1 QS(f,y;\alpha)\,\mathrm{d}\alpha, \tag{46}$$

so that CRPS reflects the quantile scores for all values of $\alpha$ from 0 to 1. Thus we first compute $QS(f_t, y_t; \alpha)$ using VaR($\alpha$) from $\int_{-\infty}^{VaR(\alpha)} f(y)\mathrm{d}y = \alpha$ and approximate CRPS by

$$CRPS(f_t, y_t) \approx \frac{2}{J-1}\sum_{j=1}^{J-1} QS(f_t, y_t; \alpha_j), \tag{47}$$

as in Gneiting and Ranjan (2011), with $\alpha_j = \frac{j}{J}$ and $J = 100$. For in-sample (IS) estimation, the expected CRPS, $E[CRPS(f,Y)]$, for $f$ being the density from Model 1, Model 2, or Model 3, is estimated by

$$CRPS_{IS} = \frac{1}{R}\sum_{t=1}^{R} CRPS_t, \tag{48}$$

where $CRPS_t \equiv CRPS(f_t, y_t)$ is the CRPS of the density estimated at time $t$ and evaluated at the realized value $y_t$ at time $t$. For out-of-sample (OOS) forecast, the expected CRPS is estimated by computing the average out-of-sample CRPS

$$CRPS_{OOS} = \frac{1}{P}\sum_{t=R}^{T-1} CRPS_{t+1}, \tag{49}$$

where $CRPS_{t+1} \equiv CRPS(f_{t+1}, y_{t+1})$ is the CRPS of the one-period ahead density forecast $f_{t+1}$ made at time $t$ and evaluated at the realized value $y_{t+1}$ at time $t+1$. With letting $y = (u,v)$, we compare density forecasts by the average out-of-sample CRPS in (49) where $CRPS_{t+1} = CRPS\left(f_{t+1}^{UV}, (u_{t+1}, v_{t+1})\right)$ for Model 2 and $CRPS_{t+1} = CRPS\left(h_{t+1}^{UV}, (u_{t+1}, v_{t+1})\right)$ for Model 3, evaluated at the realized absolute return $u_{t+1}$ and the realized sign $v_{t+1}$ at time $t+1$. In Section 6 for empirical applications, the results for $CRPS_{IS}, CRPS_{OOS}$ are presented in Table 4.

### 5.4 Test of Comparing Predictive Ability

To compare density forecast models, we test for the null hypothesis that two density forecasts $f, g$ are equal in a scoring rule, $H_0 : E[S(f,Y)] - E[S(g,Y)] = 0$. The scoring rule $S(f,y)$

is $LS(f,y)$, $QS(f,y)$, or $CRPS(f,y)$. The alternative hypothesis that $g$ is better than $f$ is $H_1 : E[S(f,Y)] - E[S(g,Y)] < 0$ for a positive-oriented scoring rule such as LS, and $H_1 : E[S(f,Y)] - E[S(g,Y)] > 0$ for a negative-oriented scoring rule such as QS and CRPS.

The test statistic for comparing the in-sample (IS) estimated densities $f_t$ and $g_t$ is $t_{IS} = \sqrt{R}\left(\bar{S}_R^f - \bar{S}_R^g\right)/\hat{\sigma}_R$, where $\bar{S}_R^f = \frac{1}{R}\sum_{t=1}^R S(f_t, y_t)$, $\bar{S}_R^g = \frac{1}{R}\sum_{t=1}^R S(g_t, y_t)$, and $\hat{\sigma}_R^2 = \frac{1}{R}\sum_{t=1}^R [S(f_t, y_t) - S(g_t, y_t)]^2$. The test statistic for comparing the out-of-sample (OOS) density forecasts is $t_{OOS} = \sqrt{P}\left(\bar{S}_P^f - \bar{S}_P^g\right)/\hat{\sigma}_P$, where $\bar{S}_P^f = \frac{1}{P}\sum_{t=R}^{T-1} S(f_{t+1}, y_{t+1})$, $\bar{S}_P^g = \frac{1}{P}\sum_{t=R}^{T-1} S(g_{t+1}, y_{t+1})$, and $\hat{\sigma}_P^2 = \frac{1}{P}\sum_{t=R}^{T-1} [S(f_{t+1}, y_{t+1}) - S(g_{t+1}, y_{t+1})]^2$. The test statistics are asymptotically standard normal under $H_0$. In Section 6 for empirical applications, we report the t-statistics and their asymptotic p-values.

# 6 Empirical Applications

## 6.1 Data

In our empirical studies, we use the daily S&P500 index. Denote by $P_t$ the S&P500 index at day $t$. We consider the density forecast of stock return. The one-day return from day $t$ to day $t+1$ is defined as $r_{t+1} \equiv P_{t+1}/P_t - 1$. We consider two periods of daily data. Data Period 1 is from 1/4/2007 to 2/12/2009 with $T = 532$ observations. We divide the whole sample into $R = 117$ in-sample observations and $P = 355$ pseudo out-of-sample observations. Data Period 2 is from 1/2/2019 to 12/31/2020 with $T = 505, R = 168$, and $P = 337$. Both periods contain obvious downturns, while the lengths of the downturns are different. Data Period 1 includes the 2007-08 Great Recession downturn for about 17 months, from October 2007 to February 2009. Data Period 2 includes the Covid-19 downturn for about 1 month, from February 20, 2020 to March 16, 2020. The plots of the S&P500 index and daily returns for both data periods are given in Figure 2.

We consider both in-sample estimation and out-of-sample forecasting. The in-sample estimation uses the $R$ in-sample observations to estimate parameters of a model. The out-of-sample forecasting is using rolling windows of the fixed size $R$ to estimate models. That is, at each time $t \in \{R, \ldots, T-1\}$ we use the data starting from $t - R + 1$ and ending at time $t$ to estimate parameters of a model and then make one-day ahead density forecast for the next day $t+1$.

### 6.2 Results

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 strong dependence between $U_{t+1}$ and $V_{t+1}$ is evidenced in the in-sample and out-of-sample estimated copula function parameter $\theta$, $\rho$-function, logarithmic score, quantile score, and empirical coverage probability, and to lesser degree in CRPS.

Figure 3 plots the out-of-sample estimated copula parameter $\theta$ for Frank (blue solid line), Clayton (red dashed line) and FGM (black dashed line) copula over time. Note that the in-sample fitted copula parameter $\theta$ for Data Period 1 are $\theta = 0.2929, 0.1548, 0.1371$, for Data Period 2 are $\theta = 0.5982, 0.0904, 0.2909$, for Frank, Clayton, and FGM copula respectively. We can see that from the figures all the three copula parameters are away from 0, indicating that the absolute return and the sign of the return are not independent. The figure shows the dependence is time-varying over the OOS, sometimes switching between positive and negative dependence between the absolute return and the sign of the return.

Figure 4 plots the estimated in-sample and out-of-sample $\rho$-functions for Frank (blue solid line), Clayton (red dashed line), FGM (black dashed line), and Independent (green solid line) copula functions over time. For Independent copula, $\rho_{t+1} = p_{t+1}$. The magnitude of the improvement from Model 2-I to Model 2 with dependent copula depends on how far away $\rho_{t+1}$ deviates from $p_{t+1}$.

Figure 5 plots the pattern of $I_t(\lambda_t)$ with $\lambda_t$ for the in-sample estimation for the two data periods using the four copula functions. In Model 3, we compute the optimal $\lambda_t^*$ by minimizing $I_t(\lambda_t)$ with respect to $\lambda_t$. From the graphs we can see that the optimal $\lambda_t^* \leq 1$, which indicates that the optimal $\eta_t^* = -\log I(\lambda_t^*) \geq 0$ and therefore $I(\lambda_t^*) \leq 1$. The pattern of $I(\lambda_t)$ for Frank copula and for FGM copula are similar and lines are overlapped, which can be explained by their overlapped $\rho$-functions in Figure 4(a)(b). In Figure 5(a), among the four copulas, the Clayton copula performs the best as it is associated the lowest $\lambda_t^*$. This can be confirmed from Table 1 that among the four in-sample logarithmic scores for Data Period 1, logarithmic score for M3-C, 3.3582, is the highest among M3 density forecasts.

Table 1 shows the in-sample and out-of-sample average values of the logarithmic scores for different density forecast models. We compare density forecasts from Models 1, 2, 3 to see whether the decomposition model improves from the normal distribution model and especially whether imposing moment constraint can improve the decomposition model. The t-statistics and p-values are shown. We observe the following empirical findings.

1. Comparing the logarithmic scores for Model 1 and Model 2, we observe that the decomposition model improves substantially upon the normal distribution model. The t-statistics reported in the rows of M2 are to compare the logarithmic score of M1 with that of each of M2's. The negative t-statistics in most cases indicate M2's are better than M1. For example, for Data Period 1, the in-sample logarithmic score jumps from 3.2513 for Model 1 to $3.3431, 3.3440, 3.3481, 3.3440$ for Model 2 using Independent, Frank, Clayton, and FGM copula functions, respectively, with the p-values ranging $0.102 \sim 0.114$.

2. Model 3 improves upon Model 2 when the HM constraint imposed, for all the in-sample results and most cases of the out-of-sample results. For example, for Data Period 1, the average out-of-sample logarithmic score goes up from $2.6567, 2.6498, 2.6566$ for Model 2 to $2.6583, 2.6558, 2.6579$ for Model 3 using Frank, Clayton, and FGM copula functions, respectively. This also holds for the independent copula with the logarithmic score 2.5528 for Model 2-I and the logarithmic score 2.6485 for Model 3-I. The estimated logarithmic scores of Model 3 are higher than those of corresponding Model 2, i.e., imposing the HM moment constraint does improve Model 2 and produces the improved density forecast Model 3, for all copula functions.

3. A t-statistic is negative for improvement in the logarithmic score. For example, for Data Period 1, the in-sample test result in the row of M2-I is $-1.205$, which is negative, meaning that M2-I has better predictive ability than M1. The in-sample test results in the row of M2-F, $-1.217$, indicating that M2-F has better predictive ability than M1. The in-sample test result in the row of M3-F, $-1.320$, showing that M3-F has better predictive ability than M2-F. In many cases, the statistics are statistically significant. For example, for Data Period 1, the out-of-sample statistic is $-1.876$ in the row of M3-C with p-value 0.030, indicating that the improvement in the performance of M3-C upon M2-C is statistically significant.

Next, let us turn to evaluating the density forecasts for their tail risk forecasts. Table 2 reports the in-sample quantile scores and the average out-of-sample predictive quantile scores, as well as the test results. In order to see more decimal numbers, we have multiplied 100 to the quantile scores. The results show that the decomposition model density forecasts using different copula functions (Model 2) produce better VaR forecasts than the normal density forecast model (Model 1), and the maximum entropy density forecasts with the HM moment constraint imposed on the decomposition model (Model 3) produce even better VaR forecasts for $\alpha = 0.01$ quantile. We observe the following empirical findings.

1. Comparing the quantile scores, we see that Model 2 improves upon Model 1 in many cases but the results are mixed in some cases. A t-statistic is positive for the improvement in the quantile score. In many cases Model 2 is better than Model 1 in the in-sample estimation but the results are mixed for the OOS forecasting. The performance of Model 2 is not stable for the out-of-sample VaR forecasts.

2. However, it is clear that Model 3 improves Model 2 for Frank, Clayton, and FGM copula functions, in both in-sample estimation of VaR and out-of-sample VaR forecasts. This shows that imposing the moment condition by the maximum entropy stabilizes the performance of Model 2.

Table 3 reports the in-sample and the average out-of-sample empirical coverage probabilities of VaR(0.01). In order to show more decimal numbers, we have multiplied 100 in Table 3. Model 2 gives more accurate empirical coverage probabilities (closer to $\alpha = 0.01$) than Model 1, and Model 3 gives even more accurate empirical coverage probabilities and better than Model 2. In general, VaR forecasts from the decomposition model with the HM moment constraint (Model 3) are most accurate in the empirical coverage probabilities ($\hat{\alpha}$) as they are closest to the true coverage probability $\alpha$.

Finally, Table 4 reports the in-sample CRPS scores and the average out-of-sample predictive CRPS scores of density models as well as the test results. In order to show more decimal numbers, we have multiplied 100 to the CRPS values. We use $J = 100$ in (47) to approximate the integral in (46).

1. Comparing the CRPS results, we find that M3 has smaller CRPS than M2, and M2 has smaller CRPS than M1, for the in-sample results for Data 1 and for the OOS results for Data 2. Note that the smaller the CRPS is, the better the density model is. However, for the other cases, the results are mixed.

2. A t-statistic is positive for improvement in CRPS. For the in-sample results for Data 1, Model 3 has better predictive ability than Model 2 with positive t-statistics and significant p-values ranging $0.035 \sim 0.108$. M2 has better predictive ability than M1 with significant p-values around $0.020 \sim 0.029$. Other cases show mixed results with insignificant p-values.

# 7 Conclusions

We consider the multiplicative decomposition of the financial returns into the sign and modulus. Anatolyev and Gospodinov (2010) study the decomposition model for the return predictability in the conditional mean. We show the decomposition model produces better density forecasts of stock returns than the models without using the decomposition. However, the decomposition model involves rich specifications of the component densities, it may be subject to misspecification and amplified estimation errors from the multiplicative structure, and may not generate stable moment predictions. Motivated by this, we show how to impose a simple moment constraint by maximum entropy to improve the density forecast of the decomposition model in terms of the logarithmic score. We also find the improvement of the maximum entropy density forecasts in terms of the quantile score and continuous ranked probability score.

Several extensions can be considered. First, if we apply more moment constraints, the maximum entropy density can improve the density of the decomposition model even more. We have not considered different moment constraints other than the mean constraint to match the first moment. If we go for matching higher order moments such as variance, skewness, kurtosis, and etc., it would improve the density forecast even more. If we go for imposing conditions that are implied from economic theory, as considered by Giacomini and Ragusa (2014), it can also improve the density forecasts further. Second, as considered by Ferreira and Santa-Clara (2011) an additive decomposition of stock returns may be considered for the convolution of the component densities for the density forecast of the whole stock returns. Third, instead of the moment equality constraint considered in this paper, some inequality constraints may be considered as in Campbell and Thompson (2008), Moon and Schorfheide (2009), and Lee, Tu, and Ullah (2014, 2015).

# References


Amisano, G. and Giacomini, R. (2007). "Comparing density forecasts via weighted likelihood ratio tests". *Journal of Business and Economic Statistics* 25: 177-190.

Anatolyev, S. and Gospodinov, N. (2010). "Modeling Financial Return Dynamics via Decomposition". *Journal of Business and Economic Statistics* 28(2): 232-245.

Bao, Y., Lee, T.H., and Saltoğlu, B. (2007). "Comparing Density Forecast Models". *Journal of Forecasting* 26(3): 203-225.

Bauwens, L. and Giot, P. (2000). "The Logarithmic ACD model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks". *Annales d'Économie et de Statistique* 60: 117-149.

Bera, A.K. and Bilias, Y. (2002). "The MM, ME, ML, EL, EF and GMM approaches to estimation: a synthesis". *Journal of Econometrics* 107: 51-86.

Bera, A.K. and Park, S.Y. (2008), "Optimal Portfolio Diversification Using the Maximum Entropy Principle". *Econometric Reviews* 27(4-6): 484-512.

Campbell, J.Y. and Thompson, S. (2008), "Predicting the equity premium out of sample: Can anything beat the historical average?" *Review of Financial Studies* 21(4): 1511-1531.

Chen, X., Fan, Y. and Tsyrennikov, V. (2006). "Efficient Estimation of Semiparametric Multivariate Copula Models". *Journal of American Statistical Association* 101: 1228-1240.

Christoffersen, P.F. and Diebold, F.X. (2006). "Financial Asset Returns, Market Timing, and Volatility Dynamics". *Management Science* 52: 1273-1287.

Corradi, V., Fosten, J., and Gutknecht, D. (2020). "Conditional Quantile Coverage: an Application to Growth-at-Risk". SSRN: https://ssrn.com/abstract=3670575.

Csiszár, I. (1975). "I-Divergence Geometry of Probability Distributions and Minimization Problems". *The Annals of Probability* 3(1): 146-158.

Ding, Z., Granger, C.W.J., Engle, R.F. (1993). "A Long Memory Property of Stock Market Returns and a New Model". *Journal of Empirical Finance* 1: 83-106.

Engle, R.F. and Russell, J.R. (1998). "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data". *Econometrica* 66(5): 1127-1162.

Ferreira, M.A. and Santa-Clara, P. (2011). "Forecasting stock market returns: The sum of the parts is more than the whole". *Journal of Financial Economics* 100: 514–537.

Giacomini, R. and Ragusa, G. (2014). "Theory-coherent Forecasting". *Journal of Econometrics* 182(1): 145-155.

Golan, A., Judge, G., and Miller, D. (1996). *Maximum Entropy Econometrics Robust Estimation with Limited Data*. Wiley, New York.

Gneiting, T. (2011). "Making and Evaluating Point Forecasts". *Journal of the American Statistical Association* 106(494): 746-762.

Gneiting, T. and Raftery, A.E. (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation". *Journal of the American Statistical Association* 102(477): 359-378.

Gneiting, T. and Ranjan, R. (2011). "Comparing Density Forecasts Using Threshold- and Quantile-Weighted Scoring Rules". *Journal of Business and Economic Statistics* 29(3): 411-422.

Grammig, J. and Maurer, K.-O. (2000). "Non-monotonic Hazard Functions and the Autoregressive Conditional Duration Model". *Econometrics Journal* 3: 16-38.

Granger, C.W.J. and Ding, Z. (1995a). "Some Properties of Absolute Return. An Alternative Measure of Risk". *Annales d'economie et de statistique* 40: 67-91.

Granger, C.W.J. and Ding, Z. (1995b). "Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets". Department of Economics, Univeristy of California, San Diego, unpublished paper.

Granziera, E., Hubrich, K., and Moon, H.R. (2014). "A Predictability Test for a Small Number of Nested Models". *Journal of Econometrics* 182(1): 174-185.

Hong, Y., and Li, H. (2005). "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates". *Review of Financial Studies* 18: 37-84.

Imbens, G.W., Spady, R.H., Johnson, P. (1998). "Information Theoretic Approaches to Inference in Moment Condition Models". *Econometrica* 66: 333-357.

Jaynes, E. T. (1957). "Information Theory and Statistical Mechanics". *Physical Review* 106: 620-630.

Jaynes, E. T. (1968). "Prior Probabilities". *IEEE Transactions on Systems Science and Cybernetics* 4: 227-241.

J.P. Morgan. (1995). "Riskmetrics Technical Manual". 3 ed.

Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation". *Econometrica* 65(4): 861-874.

Kitamura, Y., Tripathi, G., Ahn, H. (2004). "Empirical likelihood-based inference in conditional moment restriction models". *Econometrica* 72(6): 1667-1714.

Koenker, R. and G. Bassett (1978). "Regression Quantiles". *Econometrica* 46: 33-50.

Korkie, B., Sivakumar, R., and Turtle, H. (2002). "The Dual Contributions of Information Instruments in Return Models: Magnitude and Direction Predictability". *Journal of Empirical Finance* 9: 511-523.

Kullback, S. and Leibler, R. A. (1951). "On Information and Sufficiency". *Annals of Mathematical Statistics* 22: 79-86.

Lee, T.H., Tu, Y. and Ullah, A. (2014). "Nonparametric and Semiparametric Regressions Subject to Monotonicity Constraints: Estimation and Forecasting". *Journal of Econometrics* 182(1): 196-210.

Lee, T.H., Tu, Y. and Ullah, A. (2015). "Forecasting Equity Premium: Global Historical Average versus Local Historical Average and Constraints", *Journal of Business and Economic Statistics* 33(3): 393-402.

Lunde, A. (1999). "A Generalized Gamma Autoregressive Conditional Duration Model". Aarhus University, Unpublished Working Paper.

Maasoumi, E. (1993). "A Compendium to Information Theory in Economics and Econometrics". *Econometric Reviews* 12: 137-181.

Matheson, J.E., and Winkler, R.L. (1976), "Scoring Rules for Continuous Probability Distributions". *Management Science* 22: 1087-1096.

Moon, H.R. and Schorfheide, F. (2009). "Estimation with overidentifying inequality moment conditions". *Journal of Econometrics* 153: 136-154.

Park, S.Y. and Bera, A.K. (2006). "Maximum Entropy Autoregressive Conditional Heteroskedasticity Model". *Journal of Econometrics* 150: 219-230.

Rao, C.R. (1965). *Linear Statistical Inference and Its Applications.* John Wiley and Sons, Inc., New York.

Robertson, J.C., Tallman, E.W. and Whiteman, C.H. (2005). "Forecasting Using Relative Entropy". *Journal of Money, Credit, and Banking* 37(3): 383-401.

Rydberg, T.H. and Shephard, N. (2003). "Dynamics of Trade-by-Trade Price Movements: Decomposition and Models". *Journal of Financial Econometrics* 1: 2-25.

Rydén, T., Teräsvirta, T. and Åsbrink, S. (1998). "Stylized Facts of Daily Return Series and the Hidden Markov Model". *Journal of Applied Econometrics* 13: 217-244.

Shih, J.H. and Louis, T.A. (1995). "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data". *Biometrics* 51: 1384-1399.

Song, P. X.-K. Fan, Y., and Kalbfleisch, J.D. (2005). "Maximization by Parts in Likelihood Inference". *Journal of the American Statistical Association* 100(472): 1145-1158.

Stengos, T. and Wu, X. (2010). "Information-Theoretic Distribution Test with Applications to Normality". *Econometric Reviews* 29(3): 307-329.

Ullah, A. (1996). "Entropy, Divergence and Distance Measures With Econometric Applications" *Journal of Statistical Planning and Inference* 49: 137-162.

Welch, I. and Goyal, A. (2008). "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction". *The Review of Financial Studies* 21(4): 1455-1508.

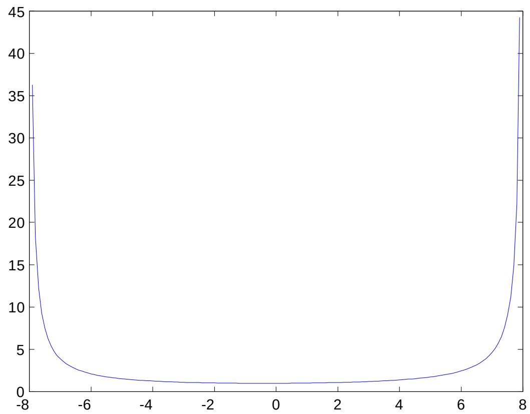
White, H. (1994). *Estimation, Inference, and Specification Analysis.* Cambridge: Cambridge University Press.

Zellner, A. (1994). "Bayesian method of moments/instrumental variable (BMOM/IV) analysis of mean and regression models" *Proceedings of Bayesian Statistical Science* of the American Statistical Association. Also appeared in 1996: J. C. Lee, W. C. Johnson, and A. Zellner (eds.): *Modeling and Prediction: Honoring Seymour Geisser.* Springer-Verlag, pp. 61-75.

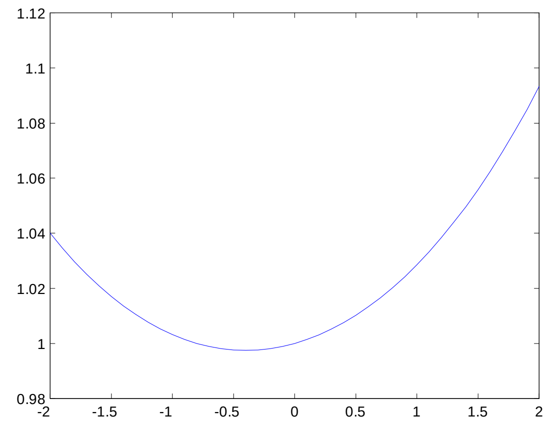
Zhang M. Y., Russell, J., and Tsay, R.S. (2001). "A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data". *Journal of Econometrics* 104: 179-207.

**Figure 1. Plots of $I(\lambda_t)$ and $I(\lambda_t^*)$**

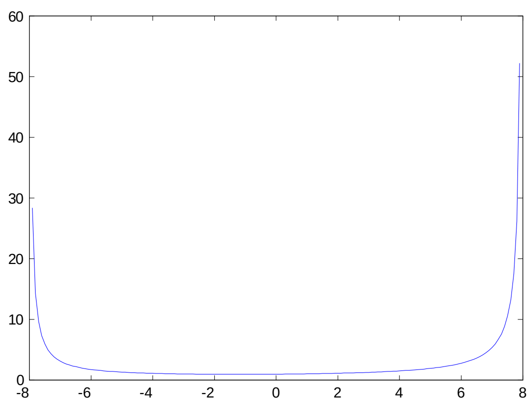
(a) $I(\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.55$



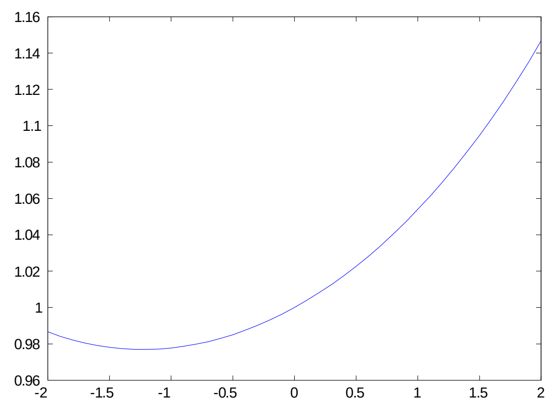

(b) $I(\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.55$



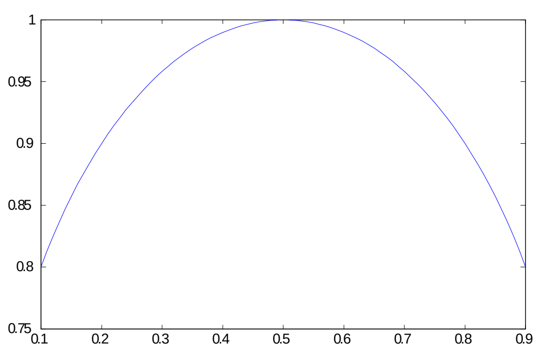

(c) $I(\lambda_t)$ with $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.65$




(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 $p_{t+1}$

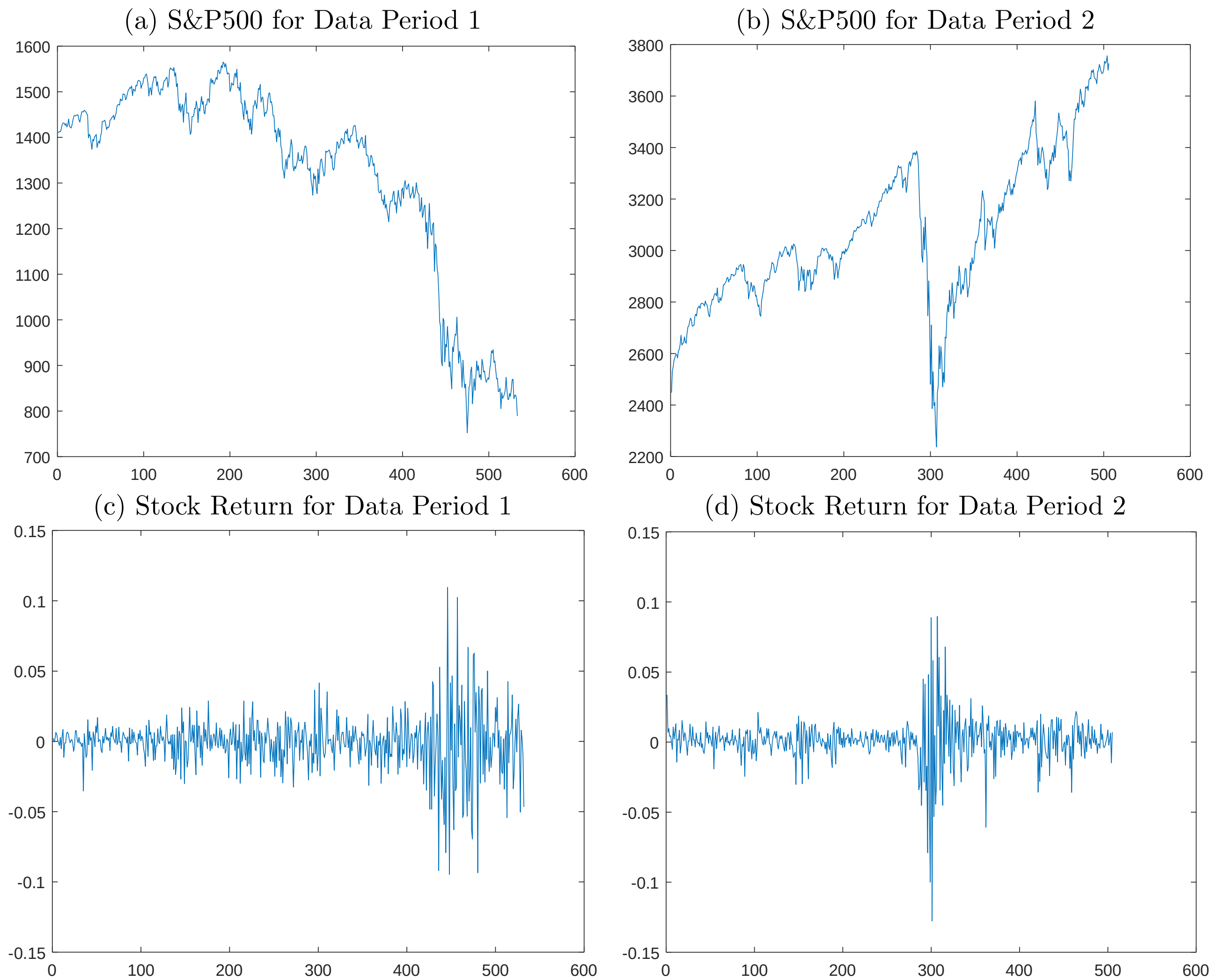



(f) Optimal $I(\lambda_t^*)$ and $\eta_t^*$ for some values of $p_{t+1}$

| $p_{t+1}$ | $I_t(\lambda_t^*)$ | $\eta_t^*$ |
|---|---|---|
| 0.51 | 0.9999 | 0.0004 |
| 0.55 | 0.9975 | 0.0025 |
| 0.65 | 0.9770 | 0.0233 |
| 0.70 | 0.9583 | 0.0426 |
| 0.80 | 0.9000 | 0.1054 |
| 0.90 | 0.8000 | 0.2231 |

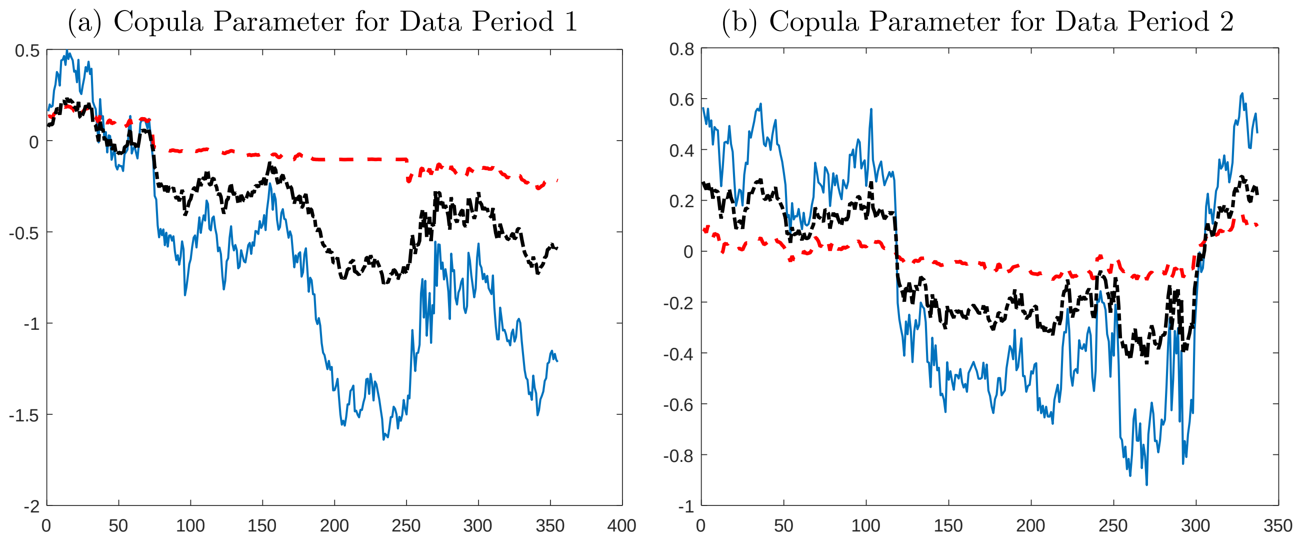
Notes: Panels (a,b,c,d) are plots of of $I(\lambda_t)$ against $\lambda_t$ for fixed values of $\frac{1}{\psi_{t+1}}$ and $p_{t+1}$. Panels (a,b) are plots of $I(\lambda_t)$ against $\lambda_t$ for fixed values of $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.55$. Panel (b) magnifies Panel (a) for $-2 < \lambda_t < 2$. Panels (c,d) are plots of $I(\lambda_t)$ against $\lambda_t$ for fixed values of $\frac{1}{\psi_{t+1}} = 8$ and $p_{t+1} = 0.65$. 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. S&P500 Index and Stock Return**



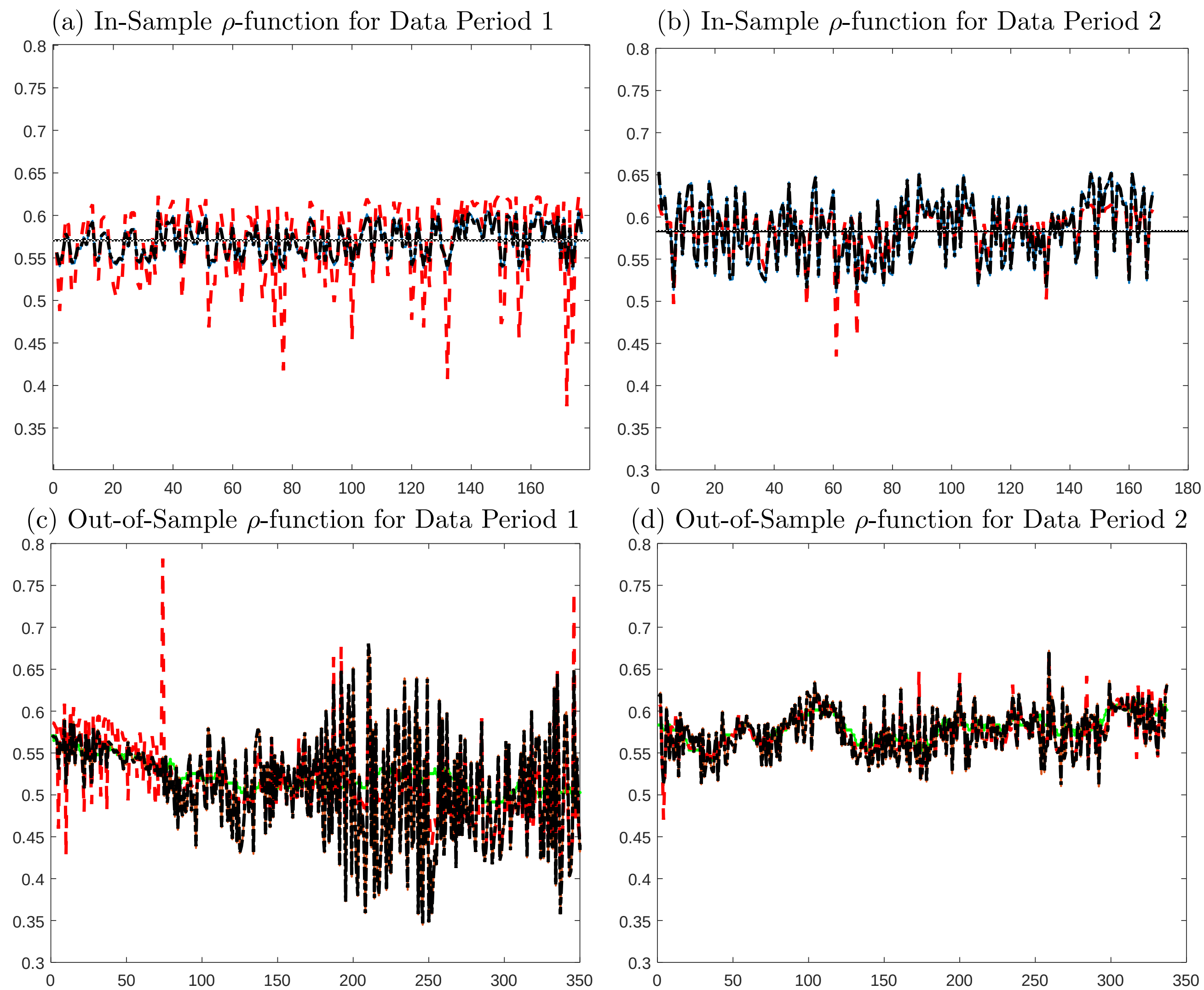

Notes: Figure 2 plots the S&P500 index and stock return $r_{t+1}$, for both data periods. Data Period 1 is from 1/4/2007 to 2/12/2009 with 532 observations. Data Period 2 is from 1/2/2019 to 12/31/2020 with 505 observations. Both periods contain obvious downturns, while the lengths of the downturns are different. Data Period 1 includes the 2007-08 Great Recession for about 17 months, from October 2007 to February 2009. Data Period 2 includes the Covid-19 downturn for about 1 month, from February 20, 2020 to March 16, 2020.

**Figure 3. Out-of-Sample Estimated Copula Parameter $\theta$**




Notes: Figure 3 plots the out-of-sample estimated copula parameter $\theta$ for Frank (blue solid line), Clayton (red dashed line) and FGM (black dashed line) copula over time. Note that the in-sample fitted copula parameter $\theta$ for Data Period 1 are $\theta = 0.2929, 0.1548, 0.1371$, for Data Period 2 are $\theta = 0.5982, 0.0904, 0.2909$, for Frank, Clayton, and FGM copula, respectively. All three copula parameter estimates $\hat{\theta}$ are changing over time and mostly away from 0, indicating that the absolute return and the sign of the return are not independent.

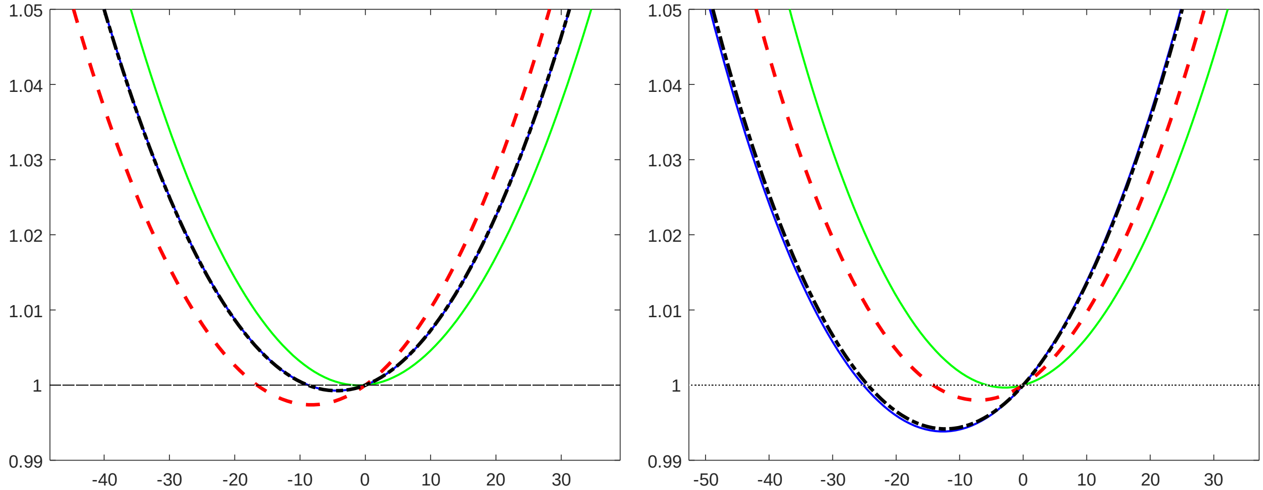
**Figure 4. In-Sample and Out-of-Sample Estimated $\rho$-function**




Notes: Figure 4 plots the estimated in-sample and out-of-sample $\rho$-function for Frank (blue solid line), Clayton (red dashed line), FGM (black dashed line), and Independent (green solid line) copula functions over time. For Independent copula, $\rho_{t+1} = p_{t+1}$. The magnitude of the improvement from Model 2-I to Model 2 with dependent copula depends on how far away $\rho_{t+1}$ deviates from $p_{t+1}$.

**Figure 5. In-Sample Pattern of $I_t\left(\lambda_t\right)$ with $\lambda_t$**

(a) Pattern of $I_t\left(\lambda_t\right)$ with $\lambda_t$ for Data Period 1 (b) Pattern of $I_t\left(\lambda_t\right)$ with $\lambda_t$ for Data Period 2




Notes: Figure 5 plots the pattern of $I_t\left(\lambda_t\right)$ with $\lambda_t$ for the in-sample period for Frank (blue solid line), Clayton (red dashed line), FGM (black dashed line), and Independent (green solid line) copula. In Model 3, we compute the optimal $\lambda_t^*$ by minimizing $I_t\left(\lambda_t\right)$ with respect to $\lambda_t$. It shows that the optimal $\lambda_t^* \leq 1$, which indicates that the optimal $\eta_t^* = -\log I\left(\lambda_t^*\right) \geq 0$ and $I\left(\lambda_t^*\right) \leq 1$. For Model 3 with the independent copula, $I\left(\lambda_t^*\right) = 1$ for Data Period 1, and $I\left(\lambda_t^*\right) = 0.9997$ for Data Period 2. The in-sample pattern of $I\left(\lambda_t\right)$ for Frank copula and for FGM copula are similar (lines are overlapped). This can be explained by the overlapped $\rho$-function of Frank and FGM copula in Figure 4(a)(b). The overlapped $\rho$-function indicates the overlapped joint density, and then indicates the overlapped $I_t\left(\lambda_t\right)$.

**Table 1. In-Sample (IS) and Out-of-Sample (OOS) Logarithmic Scores (LS) for Density Forecasts**

| | Data 1 IS | | Data 1 OOS | | Data 2 IS | | Data 2 OOS | |
|---|---|---|---|---|---|---|---|---|
| | $LS_{IS}$ | t-stat (p-val) | $LS_{OOS}$ | t-stat (p-val) | $LS_{IS}$ | t-stat (p-val) | $LS_{OOS}$ | t-stat (p-val) |
| M1 | 3.2513 | | 2.5695 | | 3.3468 | | 3.0241 | |
| M2-I | 3.3431 | −1.205 (0.114) | 2.5528 | 0.284 (0.612) | 3.3983 | −1.467 (0.071) | 3.0380 | −0.186 (0.426) |
| M2-F | 3.3440 | −1.217 (0.102) | 2.6567 | −1.485 (0.069) | 3.4017 | −1.623 (0.052) | 3.1634 | −1.865 (0.031) |
| M2-C | 3.3481 | −1.270 (0.102) | 2.6498 | −1.366 (0.086) | 3.4000 | −1.545 (0.061) | 3.1637 | −1.869 (0.031) |
| M2-FGM | 3.3440 | −1.216 (0.112) | 2.6566 | −1.482 (0.069) | 3.4016 | −1.621 (0.053) | 3.1633 | −1.863 (0.031) |
| M3-I | 3.3500 | −1.449 (0.073) | 2.6485 | −1.319 (0.094) | 3.3987 | −0.174 (0.431) | 3.1595 | −2.187 (0.014) |
| M3-F | 3.3525 | −1.320 (0.093) | 2.6583 | −0.697 (0.243) | 3.4079 | −0.744 (0.229) | 3.1621 | 0.510 (0.695) |
| M3-C | 3.3582 | −1.216 (0.112) | 2.6558 | −1.876 (0.030) | 3.4020 | −0.429 (0.334) | 3.1628 | 0.330 (0.629) |
| M3-FGM | 3.3526 | −1.362 (0.087) | 2.6579 | −0.602 (0.273) | 3.4075 | −0.723 (0.235) | 3.1620 | 0.508 (0.694) |

Notes: (a) Model 1 (M1) is the normal density forecast model with the conditional mean $\hat{\mu}_{t+1} = \bar{r}_t$ (HM) and the conditional variance $\hat{\sigma}^2_{t+1} = \gamma_{0,t} + \gamma_{1,t}\varepsilon^2_t + \gamma_{2,t}\sigma^2_t$ (GARCH). See Section 2. (b) Model 2 (M2) is the decomposition model using different copula functions. Model 2 with Independent, Frank, Clayton, FGM copula are labeled as M2-I, M2-F, M2-C, M2-FGM. See Section 3. (c) Model 3 (M3) is the decomposition model with the HM moment constraint $\hat{\mu}_{t+1} = \bar{r}_t$ imposed. For each copula function, we will denote the model using the name of copula such as M3-I, M3-F, M3-C, and M3-FGM. See Section 4. (d) The t-statistics reported in the rows of M2 are computed to compare M1 v. M2. Since $LS$ is to be maximized, t-stat will be negative if $LS(M1) < LS(M2)$ when M2 improves over M1. Similarly, the t-statistics reported in the rows of M3 are to compare M2 v. M3 for each copula function. The comparison is for the same copula for M2 v. M3. The asymptotic p-values are computed from the standard normal distribution for the one-sided test, e.g., $H_0 : E(LS(\text{M2}) - LS(\text{M3})) = 0$ against $H_1 : E(LS(\text{M2}) - LS(\text{M3})) < 0$.

**Table 2. In-Sample (IS) and Out-of-Sample (OOS) Quantile Scores (QS) for VaR(0.01) Forecasts**

| | Data 1 IS | | Data 1 OOS | | Data 2 IS | | Data 2 OOS | |
|---|---|---|---|---|---|---|---|---|
| | $QS_{IS}$ | t-stat (p-val) | $QS_{OOS}$ | t-stat (p-val) | $QS_{IS}$ | t-stat (p-val) | $QS_{OOS}$ | t-stat (p-val) |
| M1 | 0.0397 | | 0.0726 | | 0.0413 | | 0.0597 | |
| M2-I | 0.0360 | 1.056 (0.145) | 0.0772 | −0.254 (0.600) | 0.0336 | 1.227 (0.110) | 0.0598 | 0.101 (0.460) |
| M2-F | 0.0389 | 0.346 (0.145) | 0.0798 | −0.697 (0.757) | 0.0390 | 1.287 (0.110) | 0.0576 | 0.398 (0.345) |
| M2-C | 0.0401 | −0.251 (0.599) | 0.0786 | −0.501 (0.691) | 0.0362 | 1.250 (0.106) | 0.0575 | 0.408 (0.342) |
| M2-FGM | 0.0389 | 0.342 (0.366) | 0.0798 | −0.701 (0.758) | 0.0390 | 1.287 (0.099) | 0.0580 | 0.336 (0.369) |
| M3-I | 0.0401 | −1.810 (0.965) | 0.0780 | −0.541 (0.706) | 0.0364 | −1.720 (0.957) | 0.0572 | 1.271 (0.102) |
| M3-F | 0.0304 | 1.435 (0.076) | 0.0747 | 0.879 (0.190) | 0.0315 | 1.099 (0.136) | 0.0569 | 0.105 (0.458) |
| M3-C | 0.0308 | 1.440 (0.075) | 0.0759 | 0.641 (0.261) | 0.0313 | 1.195 (0.116) | 0.0575 | 0.001 (0.500) |
| M3-FGM | 0.0303 | 1.441 (0.075) | 0.0747 | 0.875 (0.191) | 0.0321 | 1.063 (0.144) | 0.0569 | 0.161 (0.436) |

Notes: (a) Reported are the QS values multiplied by 100. (b) The t-statistics reported in the rows of M2 are to compare M1 v. M2. Since QS is to be minimized, t-stat will be positive if M2 improves in VaR over M1. Similarly, the t-statistics reported in the rows of M3 are to compare M2 v. M3 for each copula function. The comparison for M2 v. M3 is for the same copula. The asymptotic p-values of the one-sided test are reported for, e.g., $H_0 : E(QS(\text{M2}) - QS(\text{M3})) = 0$ against $H_1 : E(QS(\text{M2}) - QS(\text{M3})) > 0$.

**Table 3. In-Sample (IS) and Out-of-Sample (OOS) Coverage Probability for VaR(0.01) Forecasts**

| | Data 1 IS | Data 1 OOS | Data 1 IS | Data 2 OOS |
|---|---|---|---|---|
| | $\hat{\alpha}_{IS}$ | $\hat{\alpha}_{OOS}$ | $\hat{\alpha}_{IS}$ | $\hat{\alpha}_{OOS}$ |
| M1 | 2.8249 | 3.6620 | 2.9762 | 2.6706 |
| M2-I | 2.8249 | 1.4085 | 2.3810 | 1.4837 |
| M2-F | 2.8249 | 1.4085 | 2.3810 | 1.4837 |
| M2-C | 2.8249 | 1.4085 | 2.3810 | 1.4837 |
| M2-FGM | 2.8249 | 1.4085 | 2.3810 | 1.4837 |
| M3-I | 2.8249 | 1.4085 | 2.3810 | 1.4837 |
| M3-F | 1.1299 | 1.1268 | 1.7857 | 0.8902 |
| M3-C | 1.1299 | 1.1268 | 1.7857 | 0.8902 |
| M3-FGM | 1.1299 | 1.1268 | 1.7857 | 0.8902 |

Notes: Reported are the values of $\hat{\alpha}$ multiplied 100 so that all the numbers are in percentage. The true $\alpha = 1\%$. Most of the empirical coverage probabilities of Model 2 are closer to 1% than those of Model 1, and most of the empirical coverage probabilities of Model 3 are closer to 1% than those of Model 2.

**Table 4. In-Sample (IS) and Out-of-Sample (OOS) CRPS for Density Forecasts**

| | Data 1 IS | | Data 1 OOS | | Data 2 IS | | Data 2 OOS | |
|---|---|---|---|---|---|---|---|---|
| | $CRPS$ | t-stat (p-val) | $CRPS$ | t-stat (p-val) | $CRPS$ | t-stat (p-val) | $CRPS$ | t-stat (p-val) |
| M1 | 0.5103 | | 1.1576 | | 0.4592 | | 0.8077 | |
| M2-I | 0.5011 | 2.057 (0.020) | 1.1787 | −2.187 (0.986) | 0.4567 | 0.858 (0.196) | 0.8158 | −0.800 (0.788) |
| M2-F | 0.5011 | 1.975 (0.024) | 1.1757 | −2.426 (0.992) | 0.4581 | 0.262 (0.397) | 0.8152 | −0.886 (0.812) |
| M2-C | 0.5007 | 1.902 (0.029) | 1.1754 | −2.281 (0.989) | 0.4570 | 0.643 (0.260) | 0.8153 | −0.824 (0.795) |
| M2-FGM | 0.5011 | 1.975 (0.024) | 1.1754 | −2.433 (0.993) | 0.4580 | 0.293 (0.385) | 0.8152 | −0.882 (0.811) |
| M3-I | 0.4988 | 1.248 (0.106) | 1.1719 | 1.597 (0.055) | 0.4569 | −0.103 (0.541) | 0.8141 | 0.313 (0.377) |
| M3-F | 0.4970 | 1.746 (0.040) | 1.1852 | −0.991 (0.839) | 0.4590 | −0.187 (0.574) | 0.8133 | 0.334 (0.369) |
| M3-C | 0.4975 | 1.240 (0.108) | 1.1758 | −0.062 (0.525) | 0.4577 | −0.213 (0.584) | 0.8149 | 0.069 (0.472) |
| M3-FGM | 0.4968 | 1.812 (0.035) | 1.1850 | −0.965 (0.833) | 0.4589 | −0.191 (0.576) | 0.8133 | 0.333 (0.370) |

Notes: (a) Reported are the continuous ranked probability scores multiplied by 100. (b) The t-statistic reported in an M2 row is to compare M1 v. M2 in that row. Since CRPS is to be minimized, t-stat will be positive if M2 improves over M1. Similarly, the t statistic in a row of M3 is to compare M2 v. M3 for each copula function. The comparison is for the same copula for M2 v. M3. The values below the t-statistics are p-values of the one-sided test, e.g., $H_0 : E(CRPS(\text{M2}) - CRPS(\text{M3})) = 0$ against $H_1 : E(CRPS(\text{M2}) - CRPS(\text{M3})) > 0$.