Dynamic mixed frequency density pooling

Audronė Virbickaitė [*†] Hedibert F. Lopes [‡] Martina Danielova Zaharieva [§]

Abstract

This paper proposes an alternative to combining information available from daily and intraday data to model and predict the dependence structure of equity returns. The two data sources are combined via density pooling approach, where the individual joint densities are represented as a copula function and the pooling weights are potentially time-varying. The dependence structure in the daily frequency case is extracted from a standard multivariate volatility model while the high-frequency counterpart is based on additive Inverse Wishart model (AIW). It is evident that while the AIW model preforms individually best, incorporating low frequency information via density pooling provides significant gains in predictive model performance. Finally, a portfolio allocation exercise quantifies the gains in terms of risk measures, such as Value-at-Risk and Expected Shortfall.

Keywords: Density combination; High Frequency; Mixture; Realized.

JEL: C58, C11.

[*]Corresponding author

[†]Department of Quantitative Methods, CUNEF, Calle Pirineos 55, Madrid 28040, audrone.virbickaite@cunef.edu, Phone: 0034 914 480 892

[‡]School of Mathematical and Statistical Sciences at Arizona State University, USA, and Insper Institute of Education and Research, Sao Paulo, Brazil, Hedibert.Lopes@asu.edu / hedibertfl@insper.edu.br

[§]Department of Quantitative Methods, CUNEF, Calle Pirineos 55, Madrid 28040, martina.zaharieva@cunef.edu

1 Introduction

Since the advent of the availability of the high frequency financial data, there has been a surge of research on how to use, model and predict measures extracted from such data (McAleer & Medeiros 2008). As a result, high frequency data based models have proved to be powerful competitors to the standard modeling approaches, which are based on daily data. This is especially the case for financial time series volatility and co-volatility modeling.

Alternatively, instead of picking either one or another modeling approach, some authors combine the best of the both worlds by augmenting the low frequency models with the high frequency information, see, for example Engle 2002, Ghysels et al. 2004, 2005, Shephard & Sheppard 2010, Noureldin et al. 2012, Hansen et al. 2012, 2014. Such combinations rely on a convoluted approach, where the high frequency measure enters the equation of interest as an exogenous covariate. Continuing with this line of research, we propose an approach to combine information arising from high and low frequency data to model and predict daily co-dependence structures between financial assets. Differently from the previous research, we combine low and high frequency information not through parameters, but through the combination of densities. In particular, we model the dependence structure of multiple financial returns as a weighted sum of two predictive densities, the first arising from low frequency data and the second - from high frequency data. Such combinations are also known as opinion pools (the name was first proposed by Stone 1961).

Combination of predictive densities is a recent topic and has had a lot of popularity in financial and macroeconomic literature (Hall & Mitchell 2007, Jore et al. 2010, Geweke & Amisano 2011, Billio et al. 2013, Aastveit et al. 2014, Del Negro et al. 2016). Hall & Mitchell (2007) and Geweke & Amisano (2011) rely on log predictive scores to calculate recursive combination weights, which, in long run, reach some stable equilibrium. Del Negro et al. (2016) also use the log scoring rule for modeling dynamic combination weights. Alternatively, Billio et al. (2013) consider dynamic weights that are based on the model residuals and not on the log scores. In order to control for the effect of the particular weighting scheme, we consider four options for optimal density pooling: equally weighted, static (Geweke & Amisano 2011), naïve dynamic (Jore et al. 2010) and dynamic (Del Negro et al. 2016). For a general introduction to aggregating probability distributions refer to Clemen & Winkler

(2007), among others.

We construct each one of the two multivariate distributions using copula functions. This is a convenient solution when the focus of the modeling is explicitly on the dependence structure rather then on the individual series dynamics. Modeling the dependence via copula has also some practical advantages. It allows to simplify the assessment of the marginal distributions since the copula parameter and its functional form is of major interest, and avoids dealing with highly parametrized and possibly nonstandard multivariate density functions. Models where copula parameters are obtained from daily data are considered a standard approach in the financial times series literature (Dias & Embrechts 2004, Patton 2006, Rodriguez 2007, Ausín & Lopes 2010). On the other hand, models where dependence structures are obtained from high frequency data are sparse. Fengler & Okhrin (2016) propose a realized copula where the parameter of the assumed copula family is estimated in a method-of-moments fashion from the intraday data. Okhrin & Tetereva (2017) extend the work of Fengler & Okhrin (2016) by introducing the realized hierarchical Archimedean copula, which is more flexible and available in larger dimensions. In addition, Salvatierra & Patton (2015) exploit this idea by introducing a realized correlation measure in the generalized autoregressive score (GAS) model of Creal et al. (2013) and using the resulting dynamic correlation as a copula parameter.

In this article, we rely on a Bayesian estimation approach in two stages. In the first stage, conditional on the marginals the joint predictive density is estimated given each individual model. In particular, the daily data is transformed to the unit interval by standardizing the de-meaned equity returns using the corresponding realized volatility measure and then applying the probability integral transform. This approach has two advantages. First, by using the readily available realized volatility the number of parameters to be estimated is significantly reduced. And second, it produces approximately standard normal marginals which are very easy to handle. Then, we fit a copula model to the resulting uniformly distributed data to obtain the joint predictive density for the returns. In the second stage, the density pooling weights assigned to a low and to a high frequency model are obtained.

We use daily and intraday equity return data from 2001-2009 to produce multiple variants of competing density pools differing in (i) the respective underlying low and high frequency modeling strategy and (ii) the density pooling approach. Empirical results show that pooled models outperform the best individual model in terms of the entire density forecasts as well in the left tail. We also find that the

preference for the model arising from the high frequency data is negatively correlated with the market volatility. Finally, we perform a Global Minimum Variance (GMV) portfolio allocation exercise in order to quantify the economic gains in using the proposed approach. The results confirm the benefits of pooling when focusing on the tails.

The paper is organized as follows. Section 2 presents the pooled copula model, estimation approach and model evaluation. Section 3 contains real data application for a 10-variate dataset. Finally, Section 4 concludes.

2 Methodology

The diagram in Figure 1 summarizes the main contribution of the paper: combining information arising from high and low frequency data in copula modeling framework. The choice to model the joint distribution via copulas is of pure convenience from the computational as well as from the methodological perspective. As noted in Opschoor et al. (2021), when the cross-section dimension d is large, specifying and estimating the marginals separately might considerably ease the computational burden. And, as mentioned before, such approach allows to focus on modeling the dependence structure explicitly, independently from the marginals. In particular, we are interested in estimating a copula density for some uniformly distributed data $c(u_t|\mathcal{M}_{HF}, \mathcal{M}_{LF})$. As seen from the diagram, such approach consists of three mayor tasks:

- Modeling the dynamics of the covariance matrices arising from low frequency data via model called $\mathcal{M}_{LF}$. Such model includes standard specifications for multivariate co-volatilities.

- Modeling the dynamics of the covariance matrices arising from high frequency data via model called $\mathcal{M}_{HF}$. We concentrate on additive Inverse Wishart approach (Jin & Maheu 2013, 2016).

- Modeling the dynamics of the combination weights ω_t. Here we consider four options, covering large part of the variety of linear combination strategies.

Even though both models, high- and low- frequency data based, essentially try to capture the

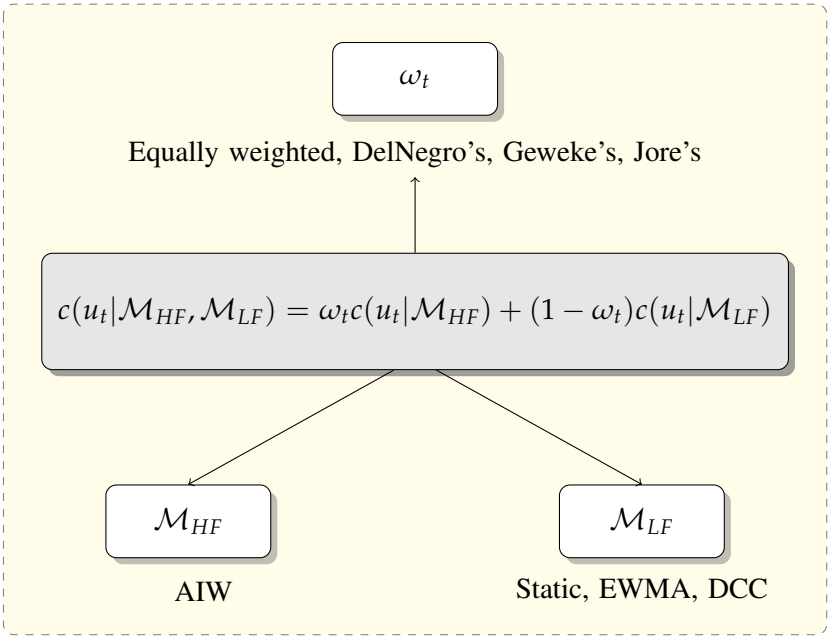

Figure 1: Model ingredients: combining high ($\mathcal{M}_{HF}$) and low ($\mathcal{M}_{LF}$) frequency models via density pooling with time varying weights ω_t.

dependence structure between the standardized returns, they exhibit very different properties. The low-frequency data based models consider the entire series of historical daily data and the estimated co-volatility processes are usually smooth. On the other hand, the high frequency data based models can capture instantaneous changes in co-variation and predict accordingly, however, are more likely to be "contaminated" by the market micro-structure noise. As noted in Kapetanios et al. (2015), some models might be useful while the markets are in decline, meanwhile other models might be more informative when the markets are booming. Therefore, the time-varying pooling weights might also indicate if the preference for one model or another is correlated to the overall market conditions.

We start by defining $r_{i,t}$ as the de-meaned log returns (in %) for day t and asset i such that $t = 1,\ldots,T$ and $i = 1,\ldots,d$:

$$r_{i,t} = 100 \times \left(\log \frac{P_{i,t}}{P_{i,t-1}} - \mathrm{E}\left[\log \frac{P_{i,t}}{P_{i,t-1}} \right] \right),$$

where $P_{i,t-1}$ and $P_{i,t}$ are the prices at the beginning and at the end of the period, respectively.

Next, we present an approach of how to combine information arising from high and low frequency

data for dependence modeling between daily financial returns. This can be done by relying on density combination approach. As noted in Clemen & Winkler (2007), there are two major approaches: linear and logarithmic pools. Linear opinion pool is a weighted linear combination of predictive probabilities, meanwhile multiplicative averaging results into a logarithmic opinion pool. It has been shown that the logarithmic combinations result into unimodal, less dispersed (Rufo et al. 2012) and symmetric (Kascha & Ravazzolo 2010) densities as compared to the ones obtained through a linear combination - quite the opposite of empirically observed features of the financial returns. Therefore, in order to accommodate the possible asymmetries and fat tails in the co-dependence, in this paper we rely on linear pools only.

Linear combination of individual densities given models $\mathcal{M}$ is given by:

$$p(r_t) = \sum_{j=1}^{N} \omega_j p(r_t|\mathcal{M}_j), \quad t = 1, \ldots, T,$$

where N is a number of alternative models, ω_j are the combination weights and $p(\cdot)$ are the candidate densities, originating from different models. The dependence structure between low frequency returns $r_{i,t}$ can be modeled by combining (i) a model estimated from daily returns with (ii) a model estimated from high frequency returns. In other words, there are several alternative models that provide predictive distributions for d-variate return vector $r_t = (r_{1,t}, \ldots, r_{d,t})'$.

One convenient way to model the potentially high dimensional joint density $p(\cdot)$ is by separating the dependence structure from the dynamics of the marginals. Furthermore, the treatment of the marginal density can be substantially simplified by taking advantage of the available *ex post* realized volatility measure defined by $RV_{i,t} = \sum_{j=1}^{J} \tilde{r}_{i,t,j}^2$. Here, $\tilde{r}_{i,t,j}$ is a l-minute log-return for day t and J is the number of l-minute intervals in a trading day (Barndorff-Nielsen & Shephard 2002, Andersen et al. 2003, Barndorff-Nielsen & Shephard 2004). For an excellent review of realized volatility refer to McAleer & Medeiros (2008). The de-meaned log returns are standardized by the realized volatility measure and some unconditional standard deviation[1] $z_{i,t} = r_{i,t}/(\sqrt{RV_{i,t}} \cdot \sigma_i)$. As seen in Andersen et al. (2000, 2001), it is safe assume that $z_{i,t} \sim N(0,1)$[2]. Finally, call $u_{i,t} = \Phi_1(z_{i,t})$ a probability inte-

[1] σ_i is a scaling factor which allows the standard deviation of the returns to deviate from the RV measure, see Jin & Maheu (2013, 2016).

[2] Andersen et al. (2000, 2001) find that the distributions of the returns scaled by realized standard deviations are approximately Gaussian.

gral transform of the $z_{i,t}$, where $\Phi_1(\cdot)$ is a cumulative distribution function for the univariate standard Normal distribution and the resulting variables are uniformly distributed $u_{i,t} \overset{iid}{\sim} \mathcal{U}(0,1) \forall i = 1, \ldots, d$ (serially uncorrelated). This maneuver helps us reduce the number of parameters and computational burden of the estimation procedure. Moreover, we circumvent the discussion of the numerous possible modeling approaches for the marginals and focus on the dependence structure instead.

The dependence structure of the resulting probability integral transforms can be easily modeled using copulas. To define a copula we consider a collection of random variables $Y_1, \ldots, Y_d$ with corresponding distribution functions $F_i(y_i) = P[Y_i \leq y_i]$ for $i = 1 \ldots, d$ and a joint distribution function $H(y_1, \ldots, y_d) = P[Y_1 \leq y_1, \ldots, Y_d \leq y_d]$. Then, according to a theorem by Sklar (1959), there exists a copula C such that

$$H(y_1, \ldots, y_d) = C(F_1(y_1), \cdots, F_d(y_d)).$$

In other words, it is possible to separate the dependence structure from the marginals. Copulas are defined in the unit hypercube $[0,1]^d$, where d is the dimension of the data, and all univariate marginals are uniformly distributed. For excellent treatment of copulas and areas of applications refer to McNeil et al. (2005), Nelsen (2006), Joe (2015), Patton (2012).

In this paper we use Gaussian and t copulas, since they are available in high dimensions ($d > 2$) and their implementation is straight-forward. Gaussian copula, even though widely used, does not allow for fat-tailed co-dependence, an assumption that can be relaxed by using the t copula. Nonetheless, one could also consider even more flexible vine copulas (Brechmann & Czado 2015, Loaiza-maya & Smith 2018) or inversion copulas (Demarta & McNeil 2005, Smith et al. 2012, Loaiza-Maya & Smith 2020), for example, which are also available in higher dimensions and can capture stylized features, observed in financial time series, such as heteroscedasticity, pair-wise fat tails and asymmetry.

Call $u_t = (u_{1,t}, \ldots, u_{d,t})'$ the collection of uniformly distributed data at time t. The d-variate Gaussian copula has the following distribution and density functions (Joe 2015):

$$C(u_t; \boldsymbol{R}) = \Phi_d(\Phi_1^{-1}(u_{1,t}), \ldots, \Phi_1^{-1}(u_{d,t}); \boldsymbol{R}),$$
$$c(u_t; \boldsymbol{R}) = \frac{\phi_d(\Phi_1^{-1}(u_{1,t}), \ldots, \Phi_1^{-1}(u_{d,t}); \boldsymbol{R})}{\prod_{i=1}^{d} \phi_1(\Phi_1^{-1}(u_{i,t}))}.$$

Here $\Phi_d(\cdot; \boldsymbol{R})$ and $\phi_d(\cdot; \boldsymbol{R})$ are a d-variate standard Normal distribution and density functions with a correlation matrix $\boldsymbol{R}$.

The d-variate t copula has the following distribution and density functions (Joe 2015):

$$C(u_t; \boldsymbol{R}, \eta) = T_{d,\eta}(T_{1,\eta}^{-1}(u_{1,t}), \dots, T_{1,\eta}^{-1}(u_{d,t}); \boldsymbol{R}),$$
$$c(u_t; \boldsymbol{R}, \eta) = \frac{t_{d,\eta}(T_{1,\eta}^{-1}(u_{1,t}), \dots, T_{1,\eta}^{-1}(u_{d,t}); \boldsymbol{R})}{\prod_{i=1}^{d} t_{1,\eta}(T_{1,\eta}^{-1}(u_{i,t}))}.$$

Here $T_{1,\eta}$, $T_{d,\eta}(\cdot; \boldsymbol{R})$, $t_{1,\eta}$ and $t_{d,\eta}(\cdot; \boldsymbol{R})$ are the univariate and d-variate t distribution and density functions with parameter $\eta > 0$ and correlation matrix $\boldsymbol{R}$. When $\eta \to \infty$ the t copula becomes a Gaussian copula.

Note that this model specification for the time series data implies that the dependence structure captured by the copula function and the dependence parameter $\boldsymbol{R}$ remain the same across all t. By using the density combination approach, we are able to relax both of these assumptions by allowing $\boldsymbol{R}$ to vary in time and a copula function $c(\cdot)$ to be a weighted average of two copulas where the weights change over time.

There is another important reason to focus on Gaussian copula, at least for the high-frequency model. We make use of a fact that the variance-covariance (or correlation) matrix, estimated from the de-meaned and standardized log returns $z_{i,t}$ (given they are approximately Normally distributed) is equivalent to the copula parameter $\boldsymbol{R}$. This result is valid for the Gaussian copula with standard Normal marginals only and is a result of Hoeffding's lemma and Sklar's theorem, for details see Fengler & Okhrin (2016). This can be easily seen in the bi-variate case. If we consider two random variables Y_1 and Y_2 with standard normal marginal distributions $F_1(y_1)$ and $F_2(y_2)$ and a joint distribution function $H(y_1, y_2)$, we can find the covariance (in this case also correlation for standardized data)

$Cov(y_1, y_2) = \sigma_{12}$ by inserting the copula function in Hoeffding's lemma:

$$\begin{aligned}\sigma_{12} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H(y_1, y_2) - F_1(y_1)F_2(y_2)]\, dy_1 dy_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [C(F_1(y_1), F_2(y_2); \boldsymbol{R}) - F_1(y_1)F_2(y_2)]\, dy_1 dy_2 \\ &= \kappa.\end{aligned}$$

When $C(F_1(y_1), F_2(y_2); \boldsymbol{R})$ corresponds to the above defined Gaussian copula, the resulting scalar κ is just the off-diagonal element of the matrix $\boldsymbol{R}$. Therefore, those two copula functions allow the one-to-one mapping between the copula dependence parameter and the linear dependence measure. This result is relevant for our work since it permits to use the realized correlation, obtained from high frequency data, as a copula parameter $\boldsymbol{R}$.

Finally, given our proposed framework, the resulting density to describe daily dependence structure at time t can be written in terms of a copula density pool:

$$c(u_t|\mathcal{M}_{LF}, \mathcal{M}_{HF}) = \omega_t c(u_t|\mathcal{M}_{HF}) + (1 - \omega_t)c(u_t|\mathcal{M}_{LF}), \tag{1}$$

where $\mathcal{M}_{HF}$ and $\mathcal{M}_{LF}$ present the models estimated using high (HF) and low frequency (LF) data.

For example, consider JP Morgan (JPM) and Bank of America (BAC) log return series, used later in real data application. Left panel of Figure 2 draws the realized correlations (in grey), sample correlation (thick black horizontal line) and rolling-window correlations (a window of 50, black line), meanwhile middle and left panels draw different dependence structures for calm and more volatile periods between standardized BAC/JPM returns and pairwise estimated t-copula degrees of freedom. As we can observe, the dependence structure can be described by using multiple alternative measures, static and dynamic. Moreover, the dependence structure not only changes in strength, but also in tail thickness. The proposed mixed frequency pooled copula would allow for all these features observed in real data.

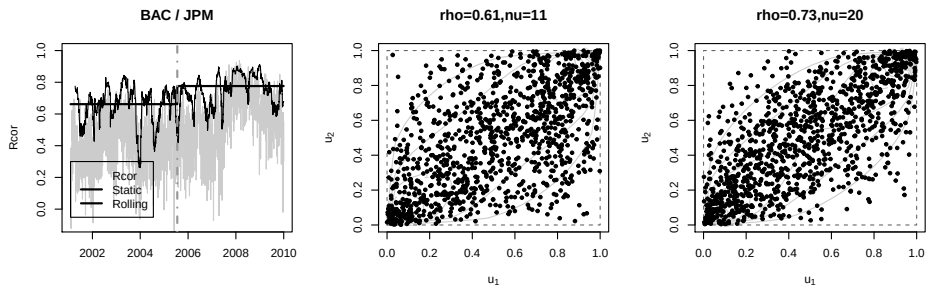

Figure 2: Left: BAC and JPM realized correlations, sample correlation (for two sub-samples) and rolling-window correlation, middle and right: different dependence structures for the first half of the sample (2001/02/01- 2005/07/20) and the second half of the sample (2005/07/21-2009/12/31) between the standardized BAC/JPM returns and pairwise estimated t-copula degrees of freedom.

2.1 Choosing the weights

In this paper we attempt to cover a large part of the types of linear pooling schemes by focusing on four different approaches: equally weighted, static (Geweke & Amisano 2011), naïve dynamic (Jore et al. 2010) and dynamic (Del Negro et al. 2016).

Geweke & Amisano (2011) propose to maximize the log predictive score function at each point in time:

$$\omega^{Gew}_{T+k+1} = \arg\max_{\omega} f(\omega), \quad \text{such that} \tag{2}$$
$$f(\omega) = \sum_{t=1}^{T+k} \log[\omega c(u_t|\mathcal{M}_{HF}) + (1-\omega)c(u_t|\mathcal{M}_{LF})],$$

where $c(u_t|\mathcal{M}_{HF})$ and $c(u_t|\mathcal{M}_{LF})$ are predictive copula densities for u_t and $k = 1, \ldots, K$ is the out of sample evaluation period. Even though the weights are recalculated at each time point, this weighting scheme is considered static because for a large K the weights will reach a stable equilibrium (Del Negro et al. 2016). Another approach is to use the log-score 'rolling' weights that are calculated at each time t using $\tilde{m}$ lags, as defined in Jore et al. (2010):

$$\omega^{Jore}_{T+k+1,\tilde{m}} = \frac{\exp[\sum_{\tau=T+k+1-\tilde{m}}^{T+k} \log c(u_\tau|\mathcal{M}_{HF})]}{\sum_{r=\{HF,LF\}} \exp[\sum_{\tau=T+k+1-\tilde{m}}^{T+k} \log c(u_\tau|\mathcal{M}_r)]}. \tag{3}$$

We call this a naïve time-varying weighting approach. The main difference between the weights in Eqs. (2) and (3) is that Geweke's approach considers the predictive densities from the entire sample,

meanwhile Jore's weighting scheme gives importance only to the last $\tilde{m}$ observations.

Finally, similarly as in Del Negro et al. (2016) we allow for persistence in weights by introducing a latent variable s_t. This gives rise to a dynamic weighting scheme:

$$\begin{aligned} s_t &= \beta s_{t-1} + \sqrt{1-\beta^2}\xi_t, \quad \xi_t \sim \mathcal{N}(0,1) \\ \omega_t^{DN} &= \Phi(s_t). \end{aligned} \tag{4}$$

The unconditional mean of s_t is 0 and the unconditional variance is 1. Parameter β controls the persistence of the weight dynamics and when $\beta = 1$ the process reduces to a random walk.

Important to note is that for the sake of simplicity in this paper we consider only a few of the available linear pools since the main goal is to investigate the potential benefits of high and low frequency data combinations. By no means we wish to present a horse-race between the pooling methods, and undoubtedly considering a more flexible/advanced density combination method would most likely result into better performing models. For example, a similar approach to ours was proposed by McAlinn (2021), where macroeconomic data from different frequencies were synthesized using Bayesian predictive synthesis (McAlinn & West 2019). Such approach is particularly powerful and beneficial when models are dependent, as shown by Takanashi & McAlinn (2021). Finally, since the data is multivariate, one can also consider weighing each series separately (McAlinn et al. 2020), a potentially beneficial route if the marginal distributions form a part of the overall model. Therefore, the use of Bayesian predictive synthesis in our proposed modelling framework should definitely be pursued in a follow-up research agenda.

2.2 Low frequency covariance modeling

First, we consider several standard approaches to model the dynamics of the correlation matrix that arises from the low frequency (daily) data.

Call Ω a variance-covariance matrix of the observed standard Normally-distributed standardized returns $z_t = (z_{1,t}, \ldots, z_{d,t})'$. Then the corresponding correlation matrix is $\boldsymbol{R} = (\text{diag}\Omega)^{-1/2}\Omega(\text{diag}\Omega)^{-1/2}$. We start with the most straightforward way to measure dependence is by using a sample correlation matrix. The dependence between u_t is modeled either by fitting a Gaussian or t copula with a static

correlation matrix $\boldsymbol{R}$, estimated given the daily data up to time t. Another possible model, a dynamic one, is the Exponentially Weighted Moving Average (EWMA) specification, popularized by the RiskMetrics© (RM) model of J.P. Morgan. The persistence parameter for the RM model can be either fixed ($\lambda = 0.94$, as recommended in the RiskMetrics model for daily data) or estimated:

$$\Omega_t = (1-\lambda) z_{t-1} z_{t-1}' + \lambda \Omega_{t-1}.$$

This model is very simple and easy to justify: the covariance at time t depends on the previous period's covariance adjusted by the most recent shock. Finally, one can generalize the EWMA model to include an intercept term, resulting into is a Dynamic Conditional Correlation (DCC) model (Tse & Tsui 2002, Engle 2002):

$$\Omega_t = \overline{\Omega} \odot (\iota\iota' - A - B) + A \odot z_{t-1} z_{t-1}' + B' \odot \Omega_{t-1},$$

where $\odot$ is the Hadamard product of two equally-sized matrices (element-by-element multiplication), ι is a vector of ones, parameter matrices A, B can be replaced with scalars a, b and $\overline{\Omega}$ is a sample variance covariance matrix. Naturally, the model choice for daily variance-covariance matrix is not limited to the models outlined above. For extensive reviews of existing multivariate volatility models refer to Asai et al. (2006), Bauwens et al. (2006), Silvennoinen & Teräsvirta (2009), among others.

2.3 High frequency covariance modeling

As mentioned before, Fengler & Okhrin (2016) show that Gaussian copula's parameter $\boldsymbol{R}_t$ can be estimated using the correlation matrix of the original data (log returns in our case). In the high frequency data setting, the correlation matrix of the returns can be estimated via Rcor_t - realized correlation measure, obtained from intraday data (Noureldin et al. 2012):

$$\text{Rcor}_t = (\text{diag}\,\text{Rcov}_t)^{-1/2} \text{Rcov}_t (\text{diag}\,\text{Rcov}_t)^{-1/2},$$

where Rcov_t is a realized covariance measure. Modeling the dynamics of the realized covariance matrices is a notoriously difficult task due to high dimensions and positive-definite restrictions on the

matrices. One way to do it is to decompose/transform the variance-covariance matrix and use some standard time-series techniques to model the transformed series. Such approach is employed by Bauer & Vorkink (2011), Chiriac & Voev (2011), among others. Another way is to model the dynamics of the realized variance covariance matrices directly by using Wishart distributions (Gourieroux et al. 2009, Jin & Maheu 2013, 2016). In Jin & Maheu (2013) the scale matrix in the Wishart distribution follows either additive or multiplicative component structure, and the authors find that the additive structure performs better. They compare their proposed model to multiple other models, such as Cholesky-VARFIMA of Chiriac & Voev (2011), Wishart auto-regressive of Gourieroux et al. (2009), vec-MGARCH of Ding & Engle (2001), DCC of Engle (2002). The authors find that the additive Wishart model produces superior density forecasts for all forecast horizons. The work of Jin & Maheu (2016) extends their previous work by considering infinite mixtures of Inverse Wishart distributions for covariance modeling. Such additive models capture strong persistence in the covariances and fat-tailed distributions of the returns.

Next, we present the additive component model, introduced in Jin & Maheu (2013, 2016). Consider a sequence of realized covariance matrices Rcov_t of dimension $d \times d$, $t = 1, \ldots, T$. The additive component Inverse Wishart AIW(L) model is given by:

$$\begin{aligned} \text{Rcov}_t &\sim \mathcal{IW}((\nu - d - 1)V_t, \nu), \\ V_t &= B_0 + \sum_{j=1}^{L} B_j \odot \Gamma_{t-1,l_j}, \\ B_j &= b_j b_j', \quad j = 1, \ldots, L, \\ \Gamma_{t-1,l_j} &= 1/l_j \sum_{i=1}^{l_j} \text{Rcov}_{t-i}. \end{aligned} \tag{5}$$

Here, $\mathcal{IW}(A, b)$ is the Inverse-Wishart distribution with scale matrix A and degrees of freedom b. We set $l_1 = 1$ and further l_js indicate how many past observations are used to form a component Γ_{t-1,l_j} and L is the number of autoregressive components. B_0 is a symmetric positive definite matrix and is set to $B_0 = (\iota\iota' - B_1 - \ldots - B_K) \odot \overline{\text{Rcov}}$ so that the long-run mean of the covariances is equal to the sample mean. This model is similar to the well-known Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) of Corsi (2009), where the log realized volatility is modeled as an

AR(22) process.

3 Posterior Inference and Model Selection

3.1 Posterior inference

For posterior inference and prediction we rely on Bayesian computation, in particular, Markov Chain Monte Carlo (MCMC) methods. In order to estimate the model in Eq. (1), we first sample from the posterior of the individual models $\mathcal{M}_{HF}$ and $\mathcal{M}_{LF}$. Conditional on those samples the density pooling weights in Eqs. (2)-(4) can be obtained. Next, we describe in short the posterior sampling details for each of the models presented in Sections 2.2 and 2.3.

Static model. Consider an Inverse-Wishart prior on the unconditional variance covariance matrix $\pi(\Omega) = \mathcal{IW}(\Omega; I_d(\nu_0 - d - 1), \nu_0)$, $\nu_0 \geq d + 1$, so that $E[\Omega] = I_d$ and I_d id the d-dimensional unit matrix. Given the observed standardized approximately Normally distributed data $z_t = (z_{1,t}, \ldots, z_{d,t})'$, where $z_{1:T} = (z_1', \ldots, z_T')'$, the parameter Ω can be sampled directly from the posterior $p(\Omega|z_{1:T}) = \mathcal{IW}(z_{1:T}z_{1:T}' + I_d(\nu_0 - d - 1), \nu_0 + T)$, see Appendix A in the Online Supplementary Material[3] for derivations. The correlation matrix that is used as a copula parameter is obtained as $\boldsymbol{R} = (\text{diag}\Omega)^{-1/2}$ $\Omega(\text{diag}\Omega)^{-1/2}$.

RiskMetrics. The estimated RiskMetrics (RMe) model contains only one parameter $\lambda \in (0, 1)$. We assume a Beta prior $\pi(\lambda) = \mathcal{B}(\lambda; a_\lambda, b_\lambda)$ so that $0 < \lambda < 1$. Given that the data is Normally distributed, the likelihood can be easily written as a function of Ω_t (or $\boldsymbol{R}_t$). Our target density is the posterior $p(\lambda|u_t) \propto \prod_{t=1}^{T} c(u_t|\boldsymbol{R}_t)\pi(\lambda)$, which is of a non-standard form. Therefore, we can sample from $p(\lambda|u_t)$ via Random Walk Metropolis Hastings (RWMH) step, for $m = 1, \ldots, M$, where M is the length of the MCMC chain, and given some starting value $\lambda^{(0)}$:

1. At iteration m draw a new value of $\tilde{\lambda}$ from a Normal proposal distribution $\mathcal{N}(\lambda^{(m-1)}, V_\lambda)$.
2. Accept the new draw with probability $\alpha = \max\left\{1, p(\tilde{\lambda})/p(\lambda)\right\}$.
3. If the draw is accepted, set $\lambda^{(m)} = \tilde{\lambda}$, if not, set $\lambda^{(m)} = \lambda^{(m-1)}$.

[3] Available at `https://sites.google.com/view/audravirbickaitephd`.

Tuning the parameter V_λ allows to control the acceptance ratio.

DCC and DCC-t. Similarly as in the RMe model above, the parameters for the scalar DCC and DCC-t models (a, b, η) can be sampled via RWMH. The priors for the parameters (a, b) are assumed Beta so that $0 < a, b < 1$, and the prior for the degrees of freedom of the Student-t distribution, η, is exponential. We sample (a, b, η) jointly in one step from a trivariate Normal proposal distribution given some starting values $(a, b, \eta)^{(0)}$. The algorithm iterates via M-H steps and we always reject the draws for which $a + b > 1$ in order to ensure for the process to be mean-reverting. For a DCC model we have only parameters (a, b).

AIW. For estimation of the AIW model, same as in Jin & Maheu (2013, 2016), we use MH within Gibbs. We assume $L = 2$ and call $\boldsymbol{b} = (b_1', b_2')$. The priors for the model parameters are

$$\pi(\nu, b_1, b_2, l_2) = \mathcal{E}_{\nu>d+1}(\nu; \xi_\nu) \cdot \mathcal{N}_{2d}(\boldsymbol{b}; 0, V_b) \cdot \mathcal{U}_Z(l_2; a_l, b_l).$$

Here $\mathcal{E}(\cdot)$ is Exponential distribution, $\mathcal{N}_{2d}(\cdot)$ is a $2d$-variate Normal distribution and $\mathcal{U}_Z(\cdot)$ is a discrete Uniform distribution. Given some starting values $(l_2, \nu, \boldsymbol{b})^{(0)}$ the algorithm iterates through the following for $m = 1, \ldots, M$:

1. Sample ν via RWMH from the conditional posterior:
   $p(\nu|l_2, \boldsymbol{b}, \text{Rcov}_{1:T}) \propto \pi(\nu) \prod_t g_{IW}(\text{Rcov}_t|l_2, \nu, \boldsymbol{b})$., where g_{IW} is the density function of the Inverse-Wishart distribution.

2. Sample $\boldsymbol{b} = (b_1', b_2')$ via RWMH jointly from the $2d$-variate Normal proposal, where the first elements of each vector are truncated to be positive for identification purposes. Same as in Jin & Maheu (2016) we reject such draws of $\boldsymbol{b}$ where B_0 is not positive definite or the absolute value of any element of $\sum_{i=1}^{2} B_i$ is not less than 1.

3. Sample l_2 via RWMH using Poisson increments that can be either positive or negative with equal probability.

Pooling weights. Estimation of the static and naïve time-varying weights is straightforward and can be done by applying the formulas in Eqs. (2) and (3) on the log predictive scores at each MCMC iteration after the estimation is carried out for all models individually. As for the time-varying persistent weights ω_t^{DN}, we employ a variant of Particle MCMC called Particle Marginal Metropolis-Hastings sampler (Andrieu et al. 2010). In particular, we use a bootstrap filter of Gordon et al. (1993) for the latent state s_t filtering and a standard MH step with Normal prior truncated at (-1,1) $\pi(\beta) = \mathcal{TN}_{(-1,1)}(\beta; m_\beta, V_\beta)$ with a random walk proposal for the persistence parameter β.

3.2 Model selection

In order to compare model performance we consider 1-step-ahead density prediction. 1-step-ahead horizon was also considered by Billio et al. (2013), for example [4]. For that purpose, we calculate the correlation matrices for $t+1$. For the static and fixed-parameter RiskMetrics (RMf) model, the marginal predictive is available analytically:

$$
\begin{aligned}
p^{\text{static}}(u_{t+1}|z_{1:t}) &= x^{-1} t_d\left(z_{t+1}|0, \frac{I_d(\nu_0 - d - 1) + z'_{1:t} z_{1:t}}{\nu_0 + t - d + 1}, \nu_0 + t - d + 1\right), \\
p^{\text{RMf}}(u_{t+1}|z_{1:t}) &= x^{-1} \phi_d(z_{t+1}|\boldsymbol{R}_{t+1}(\lambda)),
\end{aligned}
$$

where $z_{t+1} = (\Phi^{-1}(u_{1,t+1}), \ldots, \Phi^{-1}(u_{d,t+1}))'$, $x = \prod_{i=1}^{d} \phi_1(z_{i,t+1})$ and $\boldsymbol{R}_{t+1}(\lambda)$ is a correlation matrix from the fixed-parameter RiskMetrics model with known parameter λ. Note, that the marginal predictive for the static model is the Student-t density, which is a result of the Normal- Inverse Wishart conjugacy.

For DCC and estimated parameter RiskMetrics (RMe) models, described in Section 2.2, the posterior predictive distributions are given by

$$
\begin{aligned}
p^{\text{DCC}}(u_{t+1}|z_{1:t}, \theta_{DCC}) &= x^{-1} \phi_d(z_{t+1}|\boldsymbol{R}_{t+1}(\theta_{DCC})), \\
p^{\text{DCCt}}(u_{t+1}|z_{1:t}, \theta_{DCCt}) &= \left(\prod_{i=1}^{d} t_{1,\eta}(T_{1,\eta}^{-1}(u_{i,t+1}))\right)^{-1} t_{d,\eta}(u_{t+1}|\boldsymbol{R}_{t+1}(\theta_{DCCt})), \\
p^{\text{RMe}}(u_{t+1}|z_{1:t}, \theta_{RMe}) &= x^{-1} \phi_d(z_{t+1}|\boldsymbol{R}_{t+1}(\theta_{RMe})).
\end{aligned}
$$

[4] As a robustness check we also consider 5-step-ahead prediction, see Online Supplementary Material for results.

Here $\theta_{DCC} = (\overline{\Omega}, a, b)$, $\theta_{DCCt} = (\overline{\Omega}, a, b, \eta)$ and $\theta_{RMe} = \lambda$ are the estimated parameters for DCC, DCC-t and RMe models.

Finally, the posterior predictive density for the AIW model is:

$$p^{\text{AIW}}(u_{t+1}|z_{1:t}, \theta_{AIW}) = x^{-1} t_d \left(z_{t+1}|0, \frac{v-d-1}{v-d+1} V_{t+1}, v-d+1 \right),$$

where θ_{AIW} is a vector of the parameters in the AIW model.

The marginal predictive densities that account for parameter uncertainty for DCC, DCC-t, RMe and AIW models $p(u_{t+1}|z_{1:t})$ can be obtained using the MCMC output:

$$p(u_{t+1}|z_{1:t}) = \int p(u_{t+1}|z_{1:t}, \theta) p(\theta|z_{1:t}) d\theta \approx \frac{1}{M} \sum_{m=1}^{M} p(u_{t+1}|z_{1:t}, \theta^{(m)}),$$

where $(\theta^{(1)}, \ldots, \theta^{(M)})$ are the M posterior samples obtained from the MCMC.

The model comparison is carried out via predictive Bayes Factors (BF) given K out of sample observations. The BF between model 0 ($\mathcal{M}_0$) and model 1 ($\mathcal{M}_1$) is defined as (West 1986, Kass & Raftery 1995):

$$BF_{T:T+K} = \frac{p(u_{T:T+K}|z_{1:T}, \mathcal{M}_0)}{p(u_{T:T+K}|z_{1:T}, \mathcal{M}_1)},$$

where $p(u_{T:T+K}|z_{1:T}, \mathcal{M}_r) = \prod_{j=1}^{K} p(u_{T+j}|z_{1:T+j-1}, \mathcal{M}_r)$. The exact calculation of $p(u_{T:T+k}|z_{1:T}, \mathcal{M}_r)$ might be time consuming due to expanding time horizon, i.e., the model has to be re-estimated k times. For notational convenience we drop the conditioning on the model $\mathcal{M}_r$ and instead of conditioning on

$z_{1:T}$ we can condition on $u_{1:T}$ because $u_{i,t} = \Phi(z_{i,t})$. Then we can write:

$$\begin{aligned} p(u_{T:T+k}|u_{1:T}) &= \prod_{j=1}^{k} p(u_{T+j}|u_{1:T+j-1}) \\ &= \prod_{j=1}^{k} \int p(u_{T+j}|\theta) p(\theta|u_{1:T+j-1}) d\theta \\ &\overset{\text{T large}}{\approx} \prod_{j=1}^{k} \int p(u_{T+j}|\theta) \hat{p}(\theta) d\theta, \quad \text{where} \quad \theta^{(1)}, \ldots, \theta^{(M)} \sim \hat{p}(\theta), \\ &\approx \prod_{j=1}^{k} \frac{1}{M} \sum_{m=1}^{M} p(u_{T+j}|\theta^{(m)}). \end{aligned}$$

The marginal predictive distribution of $u_{T:T+k}$ can be approximated using a posterior sample of estimated model parameters $\hat{p}(\theta)$ up till time T (instead of re-estimating the model k times).

Another necessary measure, that is used for calculating the pooling weights, is the log predictive score (LPS):

$$LPS = \sum_{t=T}^{T+k-1} \log p(u_{t+1}|z_{1:t}). \tag{6}$$

Finally, we also compare the predictive model performance for the lower q^* percentile. Similar metrics were also considered by Delatola & Griffin (2011) and Opschoor et al. (2021), among others. Define the log predictive tail score ($LPTS$) measure as follows:

$$LPTS_{q^*} = \sum_{t=T}^{T+k-1} I[u_{t+1} < q] \times \log p(u_{t+1}|z_{1:t}),$$

where q is a $d \times 1$ vector and $I[u_{t+1} < q] = \prod_{i=1}^{d} I[u_{i,t+1} < q_i]$ with $q_i \in [0,1]$. Here $I[a]$ denotes the indicator function which equals 1 if condition a is fulfilled and 0 otherwise. We select $q = [q_1, \ldots, q_d]$ such that $K^{-1} \sum_{t=T}^{T+k-1} I[u_{t+1} < q] = q^*$, for $q^* = 0.5, 0.25, 0.10$ (Opschoor et al. 2021). In other words, we look at the LPS from Eq.(6) only when the d-variate data is jointly in the lower region $[0, q_1] \times \ldots \times [0, q_d]$.

4 Real Data Application

4.1 Data description

The daily and intraday equity return data as well as the realized variance and covariance data is from Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009) "Oxford-Man Institute's realized library", Oxford-Man Institute, University of Oxford[5], from the Multivariate HEAVY Paper of Noureldin et al. (2012). The data is from 2001/02/01 till 2009/12/31, 2242 data points in total. High frequency returns and the realized covariance measures are extracted as described in Noureldin et al. (2012), using 5-minute returns with subsampling. The dataset contains some of the most liquid stocks in the Dow Jones Industrial Average (DJIA) index. These are: Alcoa (AA), American Express (AXP), Bank of America (BAC), Coca Cola (KO), Du Pont (DD), General Electric (GE), International Business Machines (IBM), JP Morgan (JPM), Microsoft (MSFT), and Exxon Mobil (XOM). Online Supplementary Material contains the descriptive statistics for all assets. For the ease of exposition, the first set of descriptive plots is for BAC and JPM returns. Figure 3 draws the log returns together with the realized standard deviations, QQ-plots for the standardized returns $(z_{1,t}, z_{2,t})$ against the Normal distribution and histograms for the probability integral transforms (PITs) against the Uniform distribution. The corresponding plots for all ten assets can be found in the Online Supplementary Material. As seen from the plots, the time series data includes calm and volatile episodes. The QQ-plot indicates that the data, standardized by the RV measure, is approximately Normally distributed, as shown by Andersen et al. (2000, 2001). This is confirmed by the probability integral transforms of the standardized returns $u_{i,t} = \Phi(z_{i,t})$ which are uniformly distributed.

4.2 Prior specification and estimation

The prior hyperparameters for the variance-covariance matrix in the static model are set to $\mathcal{IW}(\Omega; I_d, 10)$. For the RiskMetrics model the prior is $\pi(\lambda) = \mathcal{B}(\lambda; 10, 3)$, for the DCC is $\pi(a, b) = \mathcal{B}(a; 3, 10)\mathcal{B}(b; 10, 3)$, and for the DCC-t is $\pi(a, b, \eta) = \mathcal{B}(a; 3, 10)\mathcal{B}(b; 10, 3)\mathcal{E}(\eta; 0.1)$. The priors for the AIW model are $\pi(\nu) = \mathcal{E}_{\nu>d+1}(\nu; 0.1)$ and $\pi(\boldsymbol{b}) = \prod_{j=1}^{10} \mathcal{N}(b_{1,j}; 0, 10)\mathcal{N}(b_{2,j}; 0, 10)$. In general, all priors are rather uninformative but proper. The size of the MCMC chain is $M = 50k$ for all models, retaining the first

[5] Available at Oxford-Man Institute of Quantitative Finance Realized Library

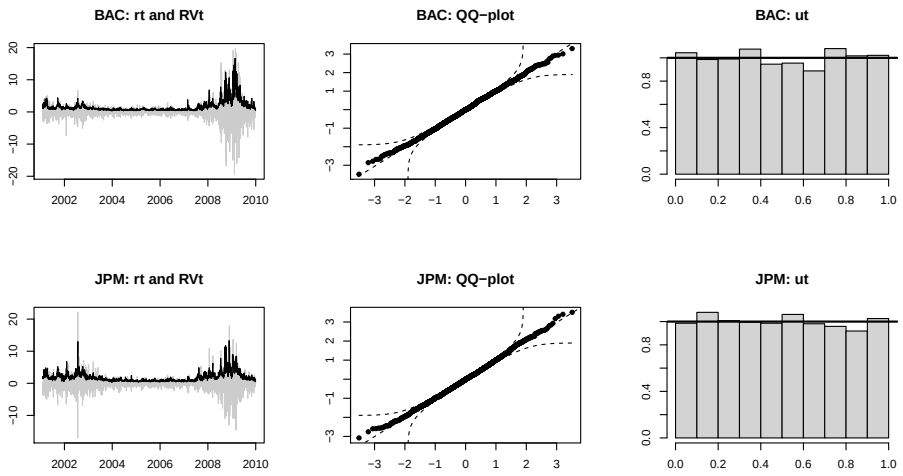

Figure 3: First column: log returns (open-to-close, in gray) and square root of realized volatilities (in black); Second column: QQ-plots of the standardized returns against the Normal distribution; Third column: histograms of the probability integral transforms against the Uniform density for BAC (top row) and JPM (bottom row) assets.

half as burn-in and thinning every 25th observation from the second half, resulting into posterior samples of 1000 observations. For the RWMH steps the proposal variances were adjusted such that the acceptance rate is around 0.5 for univariate parameter vectors and around 0.10 - 0.30 for multivariate parameter vectors. For sampling l_2 Poisson increments have the rate parameter equal to either 1.5 or 2, depending on the acceptance probability. All MCMC chains have converged after $50k$ iterations. Appendix B in the Online Supplementary Material contains parameter estimations results and trace plots for the parameters for all models. Appendix B also contains the robustness check study, where all models are re-estimated using different hyper-parameter values resulting into more vague priors. The results for all the models using different hyperparameter values have remained virtually identical.

4.3 Full model results

For estimation we have used almost all available data, retaining the last year for the out-of-sample performance evaluation. In particular, the data used for estimation is from 2001/02/01 till 2008/12/31 (1990 data points) and out-of-sample evaluation period is from 2009/01/02 till 2009/12/31 (252 data points). Table 1 presents the average LPS for the $K = 252$ out of sample observations for five low frequency data based models and a high frequency model. Based on the LPS we can see that the DCC-t

model performs the best among the low frequency data based models and AIW is the best overall. Figure 4 draws expanding-window log predictive Bayes Factors for each of the models, where the static model is the benchmark. Bayes Factors that are positive means that the model outperforms the static specification. AIW, DCC and DCC-t provide superior out of sample density forecasts, meanwhile the more restrictive RiskMetrics models are comparable to the static model. At the beginning of the evaluation period, when the markets are in turmoil, the HF data based model is clearly superior, however, when the markets are not so volatile, the DCC-t model can be a comparable alternative (e.g. around 2009/05).

Table 1: 1-step-ahead log predictive scores (*LPS*) for all individual models: Static, RiskMetrics fixed (RMf), RiskMetrics estimated (RMe), Dynamic conditional correlation with Gaussian and t copulas (DCC and DCC-t) and Additive Inverse Wishart (AIW) for 2009/01/02-2009/12/31 out-of-sample period ($K = 252$ observations).

Static	RMf	RMe	DCC	DCC-t	AIW
-3134.07	-3135.31	-3134.72	-3125.25	-3119.98	-3111.74

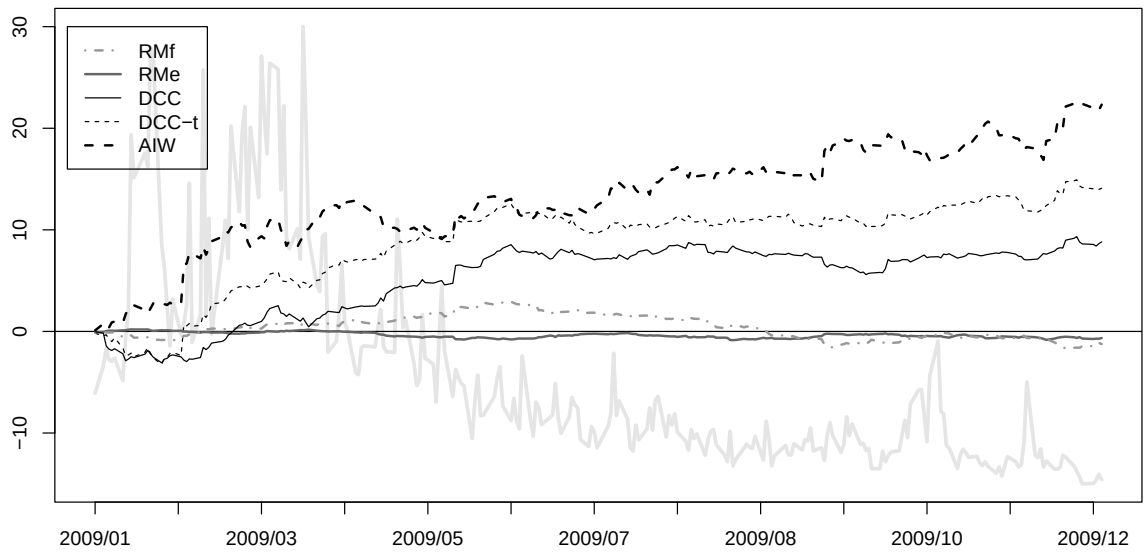

Figure 4: Expanding-window predictive log Bayes Factors with a Static model as a benchmark for all individual models: RiskMetrics fixed (RMf), RiskMetrics estimated (RMe), Dynamic conditional correlation with Gaussian and t copulas (DCC and DCC-t) and Additive Inverse Wishart (AIW) for 2009/01/02-2009/12/31 out-of-sample period ($K = 252$ observations). Average standardized realized volatility (in gray) in the background.

Next, we perform the predictive density combination exercise, described in Section 2.2. As seen

from the Figure 4, model preference is non-constant and some models, that might appear "universally" the best, in certain periods are outperformed by others (e.g. AIW vs DCC-t). So, instead of choosing a single model for density prediction, we combine predictive densities using several alternative weighting schemes. For illustration purposes, we combine DCC-t and AIW models only, them being the best models in LF and HF model classes. However, this can be easily extended in calculating the weights for all possible models.

Figure 5 draws the posterior average of the weights for the HF component (AIW model) for the four weighting schemes: equally weighted, Geweke's as in Eq. (2), Jore's with $\tilde{m} = \{1, 5, 10\}$ as in Eq. (3) and Del Negro (DN) as in Eq. (4). Jore's weights are more volatile because they take into consideration only the last $\tilde{m}$ observations, meanwhile Geweke's weight takes into consideration the entire out of sample period up till the time when the weights are calculated and reach a seemingly stable level of around 0.6. Drawing the 95% credible intervals around Geweke's weight indicate that the HF component weight is almost always different from 0.5. DN weights are not as volatile as Jore's, however, after careful inspection, one can notice that they both follow a similar pattern. The only difference is that Jore's weight is a few lags behind, meaning that DN weight capture the changes in external environment instantaneously. Overall, DN weights fluctuate around 0.5, which is not surprising given that both models, DCC-t and AIW, perform rather similarly, especially during the calm period.

The bottom plot of the Figure 5 draws expanding-window predictive log Bayes Factors for density combinations and individual models, with AIW model as the benchmark. All four combination schemes outperform the best individual AIW model. Geweke's and equally-weighted performs the worst, mainly because it does not re-balance the weights to adjust to rapidly changing environment. Surprisingly, Del Negro's scheme performs as good as Geweke's and equal weights. Overall, all four weighting schemes produce significant improvement in 1-step-ahead density prediction as compared to individual LF and HF models. As a robustness check, we have performed 5-step-ahead density forecasts, see Appendix C in the Online Supplementary Material. We found that generally the results hold.

As a robustness check, we have also considered three other pools: two best low-frequency models, some other two low-frequency models and the AIW plus some other low frequency model. We found that both low frequency combinations perform rather poorly, meanwhile the AIW plus some other

low frequency model presents substantial gains. Therefore, we argue that the observed gains are from pooling different frequencies rather than from the pooling itself. The details of this exercise can be found in the Appendix C in the Online Supplementary Material.

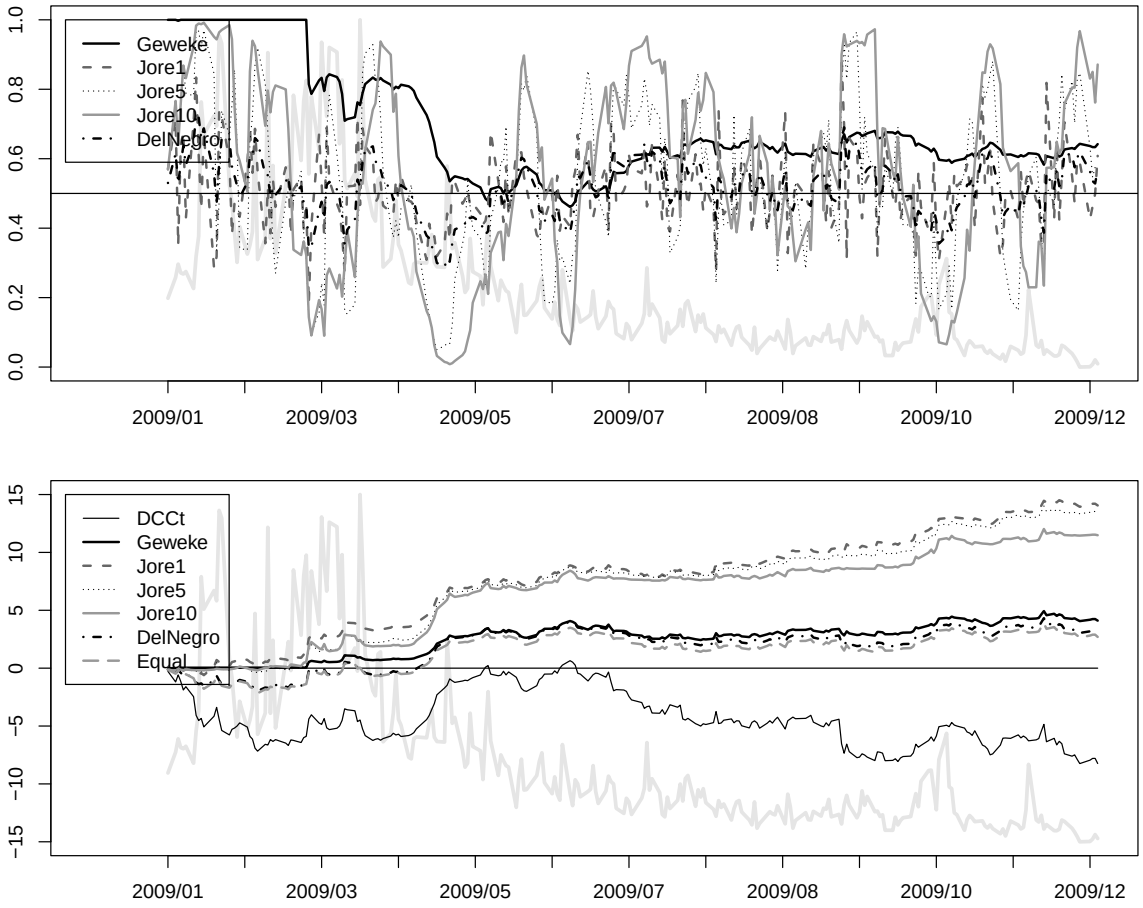

Figure 5: Results for pooling the high-frequency AIW-Gaussian and low-frequency DCC-t copulas. Top plot: posterior mean of the high-frequency component weight for the five different weighting schemes. Bottom plot: expanding-window predictive log Bayes Factor for density combinations and individual models, with Additive Inverse Wishart (AIW) as benchmark, for 2009/01/02-2009/12/31 out-of-sample period ($K = 252$ observations). Average standardized realized volatility (in gray) in the background.

Next, Figure 6 draws the posterior densities for the *average per observation* out-of-sample LPS and $LPTS_{q^*}$ for lower 10, 25 and 50% quantiles only for some individual and some of the pooled models (to ease the readability of the graph). Note that ideally one would look at 5% or even 1% quantiles, however, with the out-of-sample period being rather short ($K = 252$), the resulting sample size would

be very small[6]. The posterior densities also tell us if the differences in these average LPS and $LPTS$ are statistically significant or not. Top left plot indicates that the differences between average overall LPS is statistically significant, with pooled models providing the best predictive out of sample performance, and DCC-t model being the worst. The results change somewhat if we look only at the 50% lower quantile (top right plot). Here we see that the preference for the high-frequency based model is not so clear, because the DCC-t and AIW intervals overlap. Pooled models remain the best, with Jore's 1-period mixing weights on the top. If we look at the first quartile (bottom left) instead of the lower half, we see that the model ordering stays the same but the differences between the models decrease even further. Finally, in the 10% lower quantile (bottom right plot) both single component models and two of the pooled models perform virtually the same, where only Jore's pooling scheme is significantly better than the rest. These results show that different models perform differently depending on the metric being used (whole distribution vs the tail of the distribution), therefore, conceptually it is impossible to find such model that is universally "the best". In this paper the best model is characterized as the one which provides the highest log predictive score, because we are interested in the entire predictive distribution of the returns. Nonetheless, if one is interested exclusively in the tails for example, the log predictive tail score would be a more appropriate metric to be used in calculating the pooling weights. For example, Kapetanios et al. (2015) propose to model weights being dependent on some variable of interest, which could be some metric related to the lower region of a predictive density.

Finally, we are interested to see if the preference for the high-frequency model is correlated to overall market conditions. The preference for the high-frequency model is measured as the pooling weight of the high-frequency component in various pooling schemes. We also consider the difference between the predictive log likelihoods between the two best models: AIW and DCC-t. Positive values would mean that the AIW model is preferred, meanwhile negative values mean that the AIW model is outperformed by the DCC-t. As a proxy for the market volatility we take the average standardized realized volatility (that was obtained using the 5-minute returns with subsampling, see Noureldin et al. 2012) over the 10 assets. As alternative proxies, we also consider the equally weighted market portfolio realized volatility, the MCap[7] weighted market portfolio volatility and daily VIX index[8] for

[6]There are 12.6 observations in the 5% quantile and only 2.52 in the 1% quantile.

[7]Market capitalization (MCap) data is from June of 2021.

[8]VIX is the Chicago Board Options Exchange's Volatility Index, based on S&P500 index options.

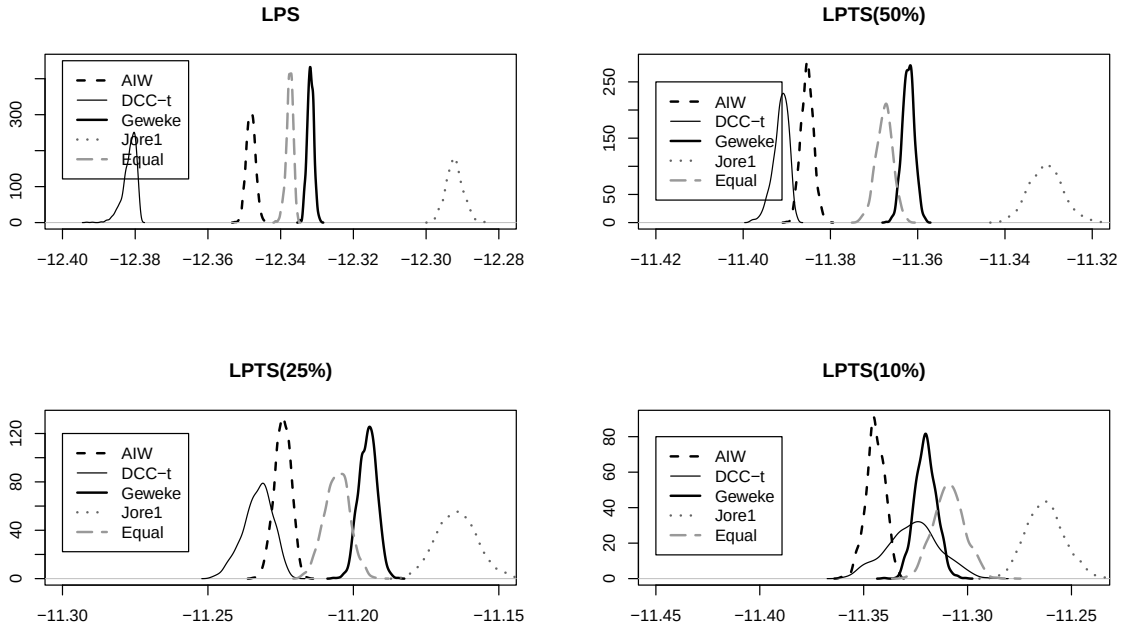

Figure 6: Posterior densities for the average per observation 1-step-ahead log predictive score and log predictive tail scores for the lower 50%, 25% and 10% quantiles for 2009/01/02-2009/12/31 out-of-sample period ($K = 252$ observations). Additive Inverse Wishart (AIW) and Dynamic conditional correlation with t copula (DCC-t) are the high and low frequency models, meanwhile the pooled models are according to Geweke's, Jore's and equally weighted schemes.

the corresponding period. Table 2 reports the posterior medians of the sample correlation coefficients between the preference for the high-frequency weight and the market volatility proxies. Except for Geweke's weights, the rest are negatively correlated, meaning that the preference for the HF model is negatively correlated with the market volatility. The correlations are relatively small, however, their 95% posterior credible intervals almost never include zero (except for Del Negro's weights). We argue that when the market volatility increases the realized variance-covariance, used in the AIW model, becomes more contaminated by the market microstructure noise. Hansen & Lunde (2006), for example, have found that the noise is negatively correlated with the returns. Therefore, the preference for the high-frequency component might decrease because of less reliable estimators during the volatile times.

Table 2: Posterior medians of sample correlations between the preference for high-frequency model and four proxies for the market volatility for 2009/01/02-2009/12/31 out-of-sample period ($K = 252$ observations). The preference for the high-frequency model is measured as a high-frequency component weight in various pooling schemes as well as the difference between the daily log likelihood (diff:logLik) between the AIW and DCC-t models. The proxies for the market volatility are: average standardized realized volatility (avrg RV), equally weighted market portfolio realized volatility (Mkt:eql), MCap weighted market portfolio realized volaltity (Mkt:MCap) and VIX index.

	Geweke	Jore1	Jore5	Jore10	DelNegro	diff:logLik
avrg RV	0.624	-0.028	-0.076	-0.108	-0.053	-0.031
Mkt:eql	0.604	-0.024	-0.069	-0.099	-0.046	-0.027
Mkt:Mcap	0.624	-0.016	-0.067	-0.097	-0.042	-0.020
VIX	0.722	-0.011	-0.026	-0.035	-0.005	-0.015

4.4 Portfolio allocation exercise

We wish to quantify how the use of one model versus another translates into a better performing portfolio in terms of economic gains. Note that even though we have the explicit form of the K 1-step-ahead predictive densities for the 10-variate return series at each MCMC iteration, the closed-form expressions for the variance-covariance matrix for the pooled models are analytically unavailable. Therefore, we employ a similar approach as in Ausín & Lopes (2010) and Opschoor et al. (2021), where at each MCMC iteration and for each out-of-sample point we draw N replications from the 10-variate predictive distribution, where N is some large number[9]. Given this simulated data, we then can calculate the 1-step-ahead variance covariance matrix and perform the portfolio weight calculation. The procedure can be summarized as follows. For each $m = 1, \ldots, M$ and for each $k = 1, \ldots, K$:

1. Simulate N replications of $u_{T+k}^{(m)}$ from $p(u_{T+k}|z_{1:T+k-1})$ and transform the uniformly distributed data to predictive returns $r_{T+k}^{*(m)}$ via the corresponding quantile function. Since we wish to be able to discriminate the models based on their predictive performance of the dependence structure only, all predictive returns have the same marginals with the *ex-post* realized variance as the volatility.

2. Calculate the empirical 1-step-ahead variance-covariance matrix $\Sigma_{T+k}^{(m)}$ of the predictive returns and obtain the solution to the quadratic programming problem ana-

[9] $N = 10,000$ in our case.

lytically:

$$w_{T+k}^{(m)} = \min \quad w_{T+k}^{\prime(m)} \Sigma_{T+k}^{(m)} w_{T+k}^{(m)} \quad \text{s.t.} w_{T+k}^{\prime(m)} \iota = 1, \quad \text{with} \quad w_{T+k}^{(m)} = \frac{\Sigma_{T+k}^{-1(m)} \iota}{\iota' \Sigma_{T+k}^{-1(m)} \iota}.$$

3. Given the estimated predictive optimal portfolio weights $w_{T+k}^{(m)}$ and the actual *ex-post* observed returns r_{T+k} and realized variance-covariance matrices Rcov_{T+k}, we can calculate various *ex-post* portfolio metrics of interest. Since we have M of such weight vectors for each time period, we can also have the entire posterior distributions of these quantities.

In particular, we calculate the $K = 252$ sequence of realized portfolio returns, portfolio volatilities and Sharpe ratios. Given the $K = 252$ realized portfolio returns, we can also calculate the empirical 5 or 10% quantiles or the expected value in these quantiles, which is the GMV portfolio Value-at-Risk (VaR) and the Expected Shortfall (ES). Finally, same as in Opschoor et al. (2021), we also calculate portfolio turnover (TO), concentration (CO) and short position (SP):

$$\begin{aligned}
TO_{T+k}^{(m)} &= \sum_{i=1}^{d} \left| w_{i,T+k+1}^{(m)} - w_{i,T+k}^{(m)} \frac{1 + r_{i,T+k}}{1 + w_{T+k}^{\prime(m)} r_{T+k}} \right|, \\
CO_{T+k}^{(m)} &= \sum_{i=1}^{d} \left(w_{i,T+k}^{(m)2} \right)^{1/2}, \\
SP_{T+k}^{(m)} &= \sum_{i=1}^{d} w_{i,T+k}^{(m)} \times I[w_{i,T+k}^{(m)} < 0].
\end{aligned}$$

Here $w_{i,T+k}^{(m)}$ is the ith element of the GMV portfolio weight vector at iteration $m = 1, \ldots, M$ for out-of-sample period $k = 1, \ldots, K$. The portfolio turnover measures the value of portfolio that is bought/sold when passing from time $T + k$ to $T + k + 1$. An investor would prefer smaller values of the TO_t because it implies less transaction costs. Portfolio concentration and portfolio short position measures how extreme the portfolio weights are. An investor would prefer such model that provides the smallest concentration and the largest short position measures.

Table 3 reports the posterior medians of various GMV portfolio metrics based on the 1-step-ahead predictions for the competing models. The high-frequency data based AIW model provides the smallest

variance producing 1-step-ahead portfolios. A similar result was observed in Opschoor et al. (2021), where the authors note that "(...) for economic criteria (GMV portfolio), simpler models prevail". The authors attribute this result to the nature of the global minimum variance criterion: the increased estimation uncertainty of more complex models, such as pooled models in our case, might overpower the differences in minimum variance, which are typically very small. Therefore, a more comprehensive criteria, based on the full density forecasts, reflects better the gains in using pooled models.

Additionally, Jore's pooled model provides the best results in terms of 5 and 10% Value-at-Risk and Expected Shortfall of the 1-step-ahead portfolio returns. In terms of portfolio turnover, DCC-t model performs the best, with Jore's pool being the second best. Meanwhile, in terms of portfolio concentration and short position, AIW performs the best with Geweke's pool being the second best.

Table 3: GMV portfolio results based on 1-step-ahead predicitons for 2009/01/02-2009/12/31 out-of-sample period ($K = 252$ observations). The table reports the posterior medians of various Global Minimum Variance portfolio metrics for the pooled models: Geweke's, Jore's and equally weighted, as well as two best individual models, Additive Inverse Wishart (AIW) and Dynamic Conditional Correlation with t copula (DCC-t).

	Geweke	Jore1	Equal	AIW	DCC-t
VaR5%	-1.285	-1.267	-1.285	-1.289	-1.281
VaR10%	-0.924	-0.910	-0.926	-0.926	-0.922
ES5%	-1.969	-1.965	-1.991	-1.967	-2.018
ES10%	-1.525	-1.521	-1.539	-1.522	-1.557
var	0.742	0.747	0.748	0.737	0.758
mean	0.032	0.036	0.035	0.033	0.037
Sharpe	0.087	0.090	0.089	0.090	0.088
TO	3.598	3.140	3.522	3.969	3.066
CO	0.671	0.672	0.672	0.664	0.680
SP	-0.178	-0.180	-0.180	-0.171	-0.189

5 Discussion and Conclusion

In this work we propose a mixed frequency copula-based approach that allows to model dependence between equity returns by using information arising from data sampled at different frequencies. We rely on density pooling approach to combine alternative copula models to describe the daily dependence structure.

In particular, we pool two copula densities, where parameters are obtained from low and high

frequency data. For high frequency copula parameter we use a realized correlation measure. We model the dynamics of the realized variance-covariance matrices via additive Inverse Wishart model with Gaussian copula, meanwhile for the low-frequency dependence structure we have considered five standard models: static, RiskMetrics fixed, RiskMetrics estimated and DCC, all with Gaussian copula, and DCC with t copula. As expected, DCC-t model always performed the best among the low frequency data based models. In the real data application, even though the overall log predictive scores favor the AIW model, incorporating information arising from the low frequency data improves the predictive model performance. We show that the gains arise not from density pooling itself, but from pooling different frequencies, also, that the results also hold for longer prediction horizons.

For future research one could consider an infinite component mixture for high-frequency data based models (Jin & Maheu 2016). Also, a more flexible pooling scheme, such as Bayesian predictive synthesis would result in overall better models. Finally, the use of more flexible copulas, such as inversion copula for example, should also be considered.

Acknowledgments

Partially supported by the grant PID2020-113192GB-I00 (Mathematical Visualization: Foundations, Algorithms and Applications) from the Spanish MICINN.

References

Aastveit, K. A., Gerdrup, K. R., Jore, A. S. & Thorsrud, L. A. (2014), 'Nowcasting GDP in real time: A density combination approach', *Journal of Business and Economic Statistics* **32**(1), 48–68.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**(1), 43–67.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2000), 'Exchange rate returns standardized by Realized Volatility are (Nearly) Gaussian', *NBER Working Paper Series* **7488**, 1–21.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Andrieu, C., Doucet, A. & Holenstein, R. (2010), 'Particle Markov chain Monte Carlo methods', *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **72**(3), 269–342.

Asai, M., McAleer, M. & Yu, J. (2006), 'Multivariate Stochastic Volatility: A Review', *Econometric Reviews* **25**(2-3), 145–175.

Ausín, M. C. & Lopes, H. F. (2010), 'Time-Varying Joint Distribution Through Copulas', *Computational Statistics and Data Analysis* **54**(11), 2383–2399.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(2), 253–280.

Barndorff-Nielsen, O. E. & Shephard, N. (2004), 'Econometric analysis of realized covariation: high frequency covariance, regression and correlation in financial economics', *Econometrica* **72**(3), 885–925.

Bauer, G. H. & Vorkink, K. (2011), 'Forecasting multivariate realized stock market volatility', *Journal of Econometrics* **160**(1), 93–101.

Bauwens, L., Laurent, S. & Rombouts, J. V. K. (2006), 'Multivariate GARCH Models: a Survey', *Journal of Applied Econometrics* **21**(1), 79–109.

Billio, M., Casarin, R., Ravazzolo, F. & Van Dijk, H. K. (2013), 'Time-varying combinations of predictive densities using nonlinear filtering', *Journal of Econometrics* **177**(2), 213–232.

Brechmann, E. C. & Czado, C. (2015), 'COPAR - Multivariate time series modeling using the copula autoregressive model', *Applied Stochastic Models in Business and Industry* **31**(4), 495–514.

Chiriac, R. & Voev, V. (2011), 'Modelling and Forecasting Multivariate Realized Volatility', *Journal of Applied Econometrics* **26**(6), 922–947.

Clemen, R. T. & Winkler, R. L. (2007), Aggregating probability distributions, *in* E. Ward, R. Miles & D. von Winterfeldt, eds, 'Advances in Decision Analysis: From Foundations to Applications', Cambridge University Press, pp. 154–176.

Corsi, F. (2009), 'A Simple Approximate Long-Memory Model of Realized Volatility', *Journal of Financial Econometrics* **7**(2), 1–23.

Creal, D., Koopman, S. J. & Lucas, A. (2013), 'Generalized Autoregressive Score Models with Applications', *Journal of Applied Econometrics* **28**(5), 777–795.

Del Negro, M., Hasegawa, R. B. & Schorfheide, F. (2016), 'Dynamic prediction pools: An investigation of financial frictions and forecasting performance', *Journal of Econometrics* **192**(2), 391–405.

Delatola, E.-I. & Griffin, J. E. (2011), 'Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility', *Bayesian Analysis* **6**(4), 901–926.

Demarta, S. & McNeil, A. J. (2005), 'The t copula and related copulas', *International Statistical Review* **73**(1), 111–129.

Dias, A. & Embrechts, P. (2004), 'Dynamic copula models for multivariate high-frequency data in finance', *Manuscript, ETH Zurich* **81**(November), 1–42.

Ding, Z. & Engle, R. F. (2001), 'Large Scale Conditional Covariance Matrix Modeling , Estimation and Testing', *NYU Working Paper No. FIN-01-029* .

Engle, R. F. (2002), 'New Frontiers for ARCH Models', *Journal of Applied Econometrics* **17**(5), 425–446.

Fengler, M. R. & Okhrin, O. (2016), 'Managing risk with a realized copula parameter', *Computational Statistics and Data Analysis* **100**, 131–152.

Geweke, J. & Amisano, G. (2011), 'Optimal prediction pools', *Journal of Econometrics* **164**(1), 130–141.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004), 'The MIDAS Touch: Mixed Data Sampling Regression Models', *UCLA: Finance* pp. 1–33.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2005), 'There is a risk-return trade-off after all', *Journal of Financial Economics* **76**(3), 509–548.

Gordon, N., Salmond, D. & Smith, A. (1993), 'Novel approach to nonlinear/non-Gaussian Bayesian state estimation', *IEE Proceedings F (Radar and Signal Processing)* **140**(2), 107–113.

Gourieroux, C., Jasiak, J. & Sufana, R. (2009), 'The Wishart Autoregressive process of multivariate stochastic volatility', *Journal of Econometrics* **150**(2), 167–181.

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

Hansen, P. R., Huang, Z. & Shek, H. H. (2012), 'Realized GARCH: a Joint Model for Returns and Realized Measures of Volatility', *Journal of Applied Econometrics* **27**(6), 877–906.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**(2), 127–161.

Hansen, P. R., Lunde, A. & Voev, V. (2014), 'Realized beta GARCH: a Multivariate GARCH Model with Realized Measures of Volatility', *Journal of Applied Econometrics* **29**(5), 774–799.

Jin, X. & Maheu, J. M. (2013), 'Modeling realized covariances and returns', *Journal of Financial Econometrics* **11**(2), 335–369.

Jin, X. & Maheu, J. M. (2016), 'Bayesian semiparametric modeling of realized covariance matrices', *Journal of Econometrics* **192**(1), 19–39.

Joe, H. (2015), *Dependence Modeling with Copulas*, CRC Press, Boca Raton, FL.

Jore, A. S., Mitchell, J. & Vahey, S. P. (2010), 'Combining forecast densities from VARs with uncertain instabilities', *Journal of Applied Econometrics* **25**(4), 621–634.

Kapetanios, G., Mitchell, J., Price, S. & Fawcett, N. (2015), 'Generalised density forecast combinations', *Journal of Econometrics* **188**(1), 150–165.

Kascha, C. & Ravazzolo, F. (2010), 'Combining Inflation Density Forecasts', *Journal of Forecasting* **29**(1-2), 231–250.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes Factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Loaiza-maya, R. & Smith, M. S. (2018), 'Time series copulas for heteroskedastic data', *Journal of Applied Econometrics* **33**(3), 332–354.

Loaiza-Maya, R. & Smith, M. S. (2020), 'Real-Time Macroeconomic Forecasting With a Heteroscedastic Inversion Copula', *Journal of Business and Economic Statistics* **28**(2), 470–486.

McAleer, M. & Medeiros, M. C. (2008), 'Realized volatility: A review', *Econometric Reviews* **27**(1-3), 10–45.

McAlinn, K. (2021), 'Mixed-frequency Bayesian predictive synthesis for economic nowcasting', *Journal of the Royal Statistical Society. Series C: Applied Statistics* **00**(March), 1–21.
URL: *doi.org/10.1111/rssc.12500*

McAlinn, K., Aastveit, K. A., Nakajima, J. & West, M. (2020), 'Multivariate Bayesian Predictive Synthesis in Macroeconomic Forecasting', *Journal of the American Statistical Association* **115**(531), 1092–1110.

McAlinn, K. & West, M. (2019), 'Dynamic Bayesian predictive synthesis in time series forecasting', *Journal of Econometrics* **210**(1), 155–169.

McNeil, A. J., Frey, R. & Embrechts, P. (2005), *Quantitative risk management: Concepts, techniques, and tools*, Princeton University Press.

Nelsen, R. B. (2006), *An Introduction to Copulas, 2nd Edition*, Springer.

Noureldin, D., Shephard, N. & Sheppard, K. (2012), 'Multivariate High-Frequency-Based Volatility (HEAVY) Models', *Journal of Applied Econometrics* **27**(6), 907–933.

Okhrin, O. & Tetereva, A. (2017), 'The Realized Hierarchical Archimedean Copula in Risk Modelling', *Econometrics* **5**(2), 26.

Opschoor, A., Lucas, A., Barra, I. & van Dijk, D. (2021), 'Closed-Form Multi-Factor Copula Models With Observation-Driven Dynamic Factor Loadings', *Journal of Business and Economic Statistics* **39**(4), 1066–1079.

Patton, A. J. (2006), 'Modelling Asymmetric Exchange Rate Dependence', *International Economic Review* **47**(2), 527–556.

Patton, A. J. (2012), 'A review of copula models for economic time series', *Journal of Multivariate Analysis* **110**, 4–18.

Rodriguez, J. C. (2007), 'Measuring financial contagion: a Copula approach', *Journal of Empirical Finance* **14**(3), 401–423.

Rufo, M. J., Martín, J. & Pérez, C. J. (2012), 'Log-linear pool to combine prior distributions: A suggestion for a calibration-based approach', *Bayesian Analysis* **7**(2), 411–438.

Salvatierra, I. D. L. & Patton, A. J. (2015), 'Dynamic copula models and high frequency data', *Journal of Empirical Finance* **30**, 120–135.

Shephard, N. & Sheppard, K. (2010), 'Realising the Future: Forecasting with High-Frequency-Based Voality (HEAVY) Models', *Journal of Applied Econometrics* **25**(2), 197–231.

Silvennoinen, A. & Teräsvirta, T. (2009), Multivariate GARCH models, *in* T. G. Andersen, R. A. Davis, J. Kreiss & T. Mikosch, eds, 'Handbook of Financial Time Series', Springer-Verlag, pp. 201–226.

Sklar, A. (1959), 'Fonctions de répartition à n dimensions et leurs marges', *Publications Inst. Statis., Univ. Paris-VIII* pp. 229–231.

Smith, M. S., Gan, Q. & Kohn, R. J. (2012), 'Modelling dependence using skew t copulas: Bayesian inference and applications', *Journal of Applied Econometrics* **27**(3), 500–522.

Stone, M. (1961), 'The opion pool.', *Annals of Mathematical Statistics* **32**, 1339–1342.

Takanashi, K. & McAlinn, K. (2021), 'Predictions with dynamic Bayesian predictive synthesis are exact minimax', *arXiv preprint arXiv:1911.08662* pp. 1–27.

Tse, Y. K. & Tsui, A. K. C. (2002), 'A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations', *Journal of Business and Economic Statistics Economic Statistics* **20**(3), 351–362.

West, M. (1986), 'Bayesian Model Monitoring', *Journal of the Royal Statistical Society. Series B (Methodological)* **48**(1), 70–78.