Dynamic vine copula models for high-frequency data

Murilo A. P. Pereira
HEC Montréal

Abstract: This study investigates intraday patterns in the comovements of financial stock returns, focusing on the importance of flexible dependence structures on density forecasting accuracy. We propose a dynamic canonical vine copula method, which models complex dependence patterns, including both time-varying and asymmetric dependencies in the upper and lower tails among financial assets. Utilizing a pair copula decomposition approach, this research analyzes 1-minute frequency returns of 10 U.S. financial stocks in March 2020, a period marked by Covid-19 market turmoil. Our findings highlight the critical role of tail dependencies and time-varying parameters in accurately modeling and forecasting intraday returns.

Keywords: forecasting; regular vine; pair-copula constructions; time-varying copulas;

Acknowledgments: I am indebted to Marcelo Medeiros for his helpful comments during my pre-doc research and for kindly sharing his dataset.

1 Introduction

The main goal of this work is to explore commonalities between financial stock returns within the day. Interactions among financial assets are crucial, serving diverse roles in finance. Such comovements have implications for risk management, asset allocation, and policy-making. Consequently, investors, risk managers, and regulators must closely monitor and respond to evolving dynamics. Specifically, understanding the dependence structures between stocks becomes critical for assessing the risk of a portfolio throughout the day, rather than focusing on the risks of individual stocks. Intraday dependencies in financial asset returns pose significant challenges, and several open questions persist in the current literature.

In high-frequency settings, previous empirical research has suggested that the variability of information arrival throughout the day is one potential reason for expecting time-varying dependencies in intraday financial asset movements. Several studies focus on volatility modeling, they have indicated that the volatility of high-frequency returns tend to be greater at market opening compared to other times of the day (Andersen and Bollerslev (1997) and Andersen, Thyrsgaard and Todorov (2021)). Research on the patterns of intraday dependence of stock price changes remains limited. The work by Koopman, Lit, Lucas and Opschoor (2018) advances our understanding in this field. They have developed a dynamic model for the intraday dependence between discrete stock price changes using an equidependence structure.

In this paper I explore the patterns in intraday dependencies in financial stock returns, and evaluate whether accounting for flexible dependencies increase the forecasting accuracy of comovements. This research specifically focuses on high-frequency movements during the Covid-19 market turmoil in March 2020. Utilizing a dynamic canonical vine copula approach, I construct a hierarchical framework that accommodates time-varying tail dependencies and asymmetric effects. Previous studies have highlighted the importance of accounting for key features such as asymmetric dependence and heavy tails in modeling financial asset comovements. In this study, I tackle these aspects and further explore the pair copula decomposition proposed by Bedford and Cooke (2002) to model complex dependencies using simple elements known as pair-copulas. I consider time variation in the dependence structure of pair-copulas endowing copula parameters with an autoregressive dynamic and a forcing variable, which is used as the innovation. This time-varying structures are introduced in Tse and Tsui (2002) and Patton (2006).

This research employs 1-minute frequency returns data for 10 U.S. financial stocks throughout March 2020. Following an initial exploratory analysis, I conducted an in-sample investigation to evaluate the goodness of fit for various bivariate copulas, including Elliptical, Archimedean, and Mixture copulas. The findings indicate that the Student-t distribution offers the best fit for most bivariate copulas within the in-sample assessment. Consequently, the in-sample performance of the model that does not impose any restriction on the distribution family, which is called the benchmark model, and the one using exclusively Student-t pairs is nearly the same. Furthermore, this research presents evidence of intraday time variations in the dependence structure.

In addition, I also include an out-of-sample forecasting application, which entails 1-minute density forecasts of the dependence structure, employing a C-vine copula structure with varied bivariate links. The performance of the benchmark model, the one that select bivariate family out of several distributions, is statistically superior over all other specification, except for the C-vine model with Student-t distribution, which performs equally well. Notably, the incorporation of asymmetrical tail dependencies in the Mixture copula does not enhance the out-of-sample forecasting accuracy using log-likelihood based measures.

Extensive research has addressed modeling conditional dependence in multivariate financial time series, notably through multivariate GARCH models such as CCC-GARCH (Bollerslev, 1990) and DCC-GARCH (Engle, 2002). These approaches primarily utilize correlation or covariance matrices to estimate conditional dependence. However, the assumptions on the distributions for each return series are often limited to distribution that need to be explicitly defined. The introduction of the copula function by Sklar (1959), and its detailed examination by Joe (1997) and Nelson (2006), paved the way for copula-based GARCH models. These models, further developed by Aas and Berg (2009), Ausin and Lopes (2010), and Patton (2006), isolate the modeling of the joint density function for marginal time series and the multidimensional copula density, allowing for more flexible dependence structures that better capture the empirical features of the financial data.

Time-varying copulas have emerged as a key tool for describing dependence dynamics in economics and finance (Manner and Reznikova, 2012). Several methods exist for investigating dependencies in limited cross-sectional dimensions, as detailed in Patton (2013). However, techniques for large asset collections are scarce, primarily due to the curse of dimensionality. Recently, there have been efforts to address this issue by utilizing factor copulas (Oh and Patton (2017), Oh and Patton (2023), and Opschoor, Lucas, Barra and van Dijk (2021)). Specifically, Opschoor et al. (2021) presents a multi-factor copula with dynamic loadings for daily data. Their study offers a computationally simple way to work with copulas when the cross-sectional number of variables is large. However, the proposed model is not flexible in the tails and it relies on pre-specified cluster assignments. The approach that I follow in this work has similarities with Koopman et al. (2018), their framework uses GAS copulas with time-varying parameters and Skellam marginals to analyse intraday dependence among stock price changes. They show that in the US stock market, the dependence starts low but gradually increases throughout the day. However, their approach does not incorporate tail dependence, aspect I address in this research. Furthermore, their analysis employs a simple equidependence structure.

Vine copula GARCH models addresses some limitations of traditional copula-GARCH models. Introduced by Joe (1997) and further developed by Bedford and Cooke (2001) as a type of graphical model, vines allow for the detailed analysis of conditional dependence between pairs of random variables. This approach is key to understanding how each pair of return series depends on each other through bivariate conditional copulas. Kurowicka and Cooke (2006) introduced the concept of Gaussian vines, and Aas, Czado, Frigessi and Bakken (2009) expanded on this by outlining how to construct a vine copula GARCH model, including simulation algorithms, model selection, and the study of tail dependence. Further developments and interesting empirical application of vine-copulas, including dynamic vines, were explored in Nikoloulopoulos, Joe and Li (2012) and Tófoli, Ziegelmann, Silva Filho and Pereira (2019).

The remainder of this paper proceeds as follows. Section 2 presents the canonical vine copula garch model proposed in this study and Section 3 presents the empirical application, describing the data and discussing the main results. Section 4 offers some concluding remarks.

2 Vine copula GARCH model

In this section, I present the dynamic C-vine copula model and offer a primer on canonical vine copula theory. Then, I describe how to introduce dynamics in the dependence structure. According to Sklar (1959), a multivariate cumulative distribution function F with marginals $F_1, ..., F_n$ may be written as

$$F(x_1, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)), \tag{1}$$

for a n-dimensional copula C. The joint probability density function, for an absolutely continuous distribution function with strictly increasing, continuous marginals densities $F_1, ..., F_n$ is given by

$$f(x_1, \ldots, x_n) = c_{1\ldots n}(F_1(x_1), \ldots, F_n(x_n)) \cdot f_1(x_1) \ldots f_n(x_n). \tag{2}$$

for a uniquely identified n-variate copula density $c_{1\ldots n}(\cdot)$.

Copulas offer a methodology for isolating the marginal structure from the dependency structure. The primary objective of this research is to thoroughly investigate this dependency structure. I explore the pair-copula decomposition presented by Bedford and Cooke (2002), which allows the construction of complex multivariate dependency models through a series of simple, interconnected blocks, using trees.

2.1 Pair-copula construction of c-vine copula

Vine copulas in d-dimensions are constructed through a process of progressively combining bivariate linking copulas in a hierarchical structure. Diverse types of vine copulas can be constructed, Bedford and Cooke (2002) presented a graphical model called regular vine. Two boundary cases are the canonical vine (C-vine) and the drawable vine (D-vines), as discussed by Aas et al. (2009). In this paper we focus on C-vine copulas.

In a d-dimensional C-vine, the pairs at level 1 are 1, i, for $i = 2, ..., d$, and for level ℓ $(2 \leq \ell < d)$, the conditional pairs are $\ell, i|1, \ldots, \ell - 1$ for $i = \ell + 1, \ldots, d$. In this structure, conditional copulas are specified for variables ℓ and i given those indexed as 1 to $\ell - 1$. The decomposition of a multivariate density in pair-copulas requires marginal conditional distribution, which can be computed using Joe (1996) equation, given by

$$F(x|\boldsymbol{v}) = \frac{\partial C_{x,v_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j}))}{\partial F(v_j|\boldsymbol{v}_{-j})}. \tag{3}$$

Following Aas et al. (2009), the density of C-vines is

$$f(\mathbf{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}\big(F_{i|i+1\ldots i+j-1}(x_i|\mathbf{x}_{i+1;i+j-1}), F_{i+j|i+1,\ldots,i+j-1}(x_{i+j}|\mathbf{x}_{i+1:i+j-1})\big) \tag{4}$$

where index j represents the trees and index i the edges in each tree. The decomposition is given by pair-copulas and marginals.

Following the approach outlined in Aas et al. (2009), this study will utilize the maximum pseudo-likelihood method for estimating the parameters of the pair-copula decomposition. The computational feasibility of multivariate vine copulas arises from their densities being expressible in terms of bivariate linking copulas. The structure of the log-likehood of equation 4 allows to adopt an efficient sequential estimation procedure. First, I estimate the parameters of the marginal distributions, then I compute the parameters of the bivariate copulas associated with different levels of the vine. Estimation is carried conditionally on the parameters of previous steps. Initially, the marginal distribution parameters are determined using maximum likelihood estimation, and the log-returns are converted into uniform values. These uniform values are then used as inputs for the pair-copulas at the first level of the C-vine. Then, the procedure involves sequentially transforming data and estimating pair-copula parameters, based on prior levels. This iterative process continues for all trees.

2.2 Building blocks

This research considers copula families with flexible tail behaviors in the estimation process of C-vine copula. This families present different degrees of asymmetries and upper/lower tail dependence.

- Elliptical copulas: Gaussian and Student-t
- Archimedean copulas: Gumbel, Survival Gumbel and Clayton
- Mixture copulas: BB1, BB7, BB8

The Gaussian and the Student-t copulas are reflection symmetric, but only the Student-t copula has upper and lower tail dependence. In the familly of Archimedean copulas, the Gumbel copulas has upper tail dependence, while the Survival Gumbel and Clayton have only lower tail dependence. Finally, the Mixture copulas have different upper and lower tail dependence. For every pair of transformed data I use the Akaike Information Criterion (AIC) to identify the superior model.

2.3 Dynamic c-vine copulas

The last subsection demonstrated that the C-vine copula models effectively breaks down a high-dimensional density into products of pair-copulas. Building on this concept, a dynamic C-vine copula model can be developed by integrating dynamic pair-copulas into this structure.

While most of the existing studies on vine copulas in finance consider static parameters, previous research indicates that the correlation among returns varies (Ang and Chen, 2002). Consequently, as in Tófoli et al. (2019), the approach of this study enhance the C-vine copula model with dynamic features, allowing the pair-copulas' dependence parameters to adjust over time.

The dynamics of the dependence parameter is based on Tse and Tsui (2002) and Patton (2006). The time variation of the dependence parameter $\theta_{i,i+j|v_{ij}}$ of the pair-copula $c_{i,i+j|v_{ij}}$, is defined as

$$\theta_{i,i+j|v_{ij},t} = \Lambda((1-a-b)\omega + a\theta_{i,i+j|v_{ij},t-1} + b\psi_{t-1}) \tag{5}$$

where Λ is a logistic transformation used to keep the parameter in its correct interval in the estimation procedure. Equation 5 contains an autoregressive term with coefficient a and a forcing variable, ψ_t, with coefficient b. The identification of the forcing variable can be complicated in certain cases, for the Elliptical copulas the forcing variable is the sample conditional correlation given past m-period data. Copulas from the Archimedean family and other similar copulas, which lack correlation parameters, face challenges in being adapted for time-varying scenarios. In this work, I utilize the sample conditional Kendall's tau given past m-period data for Archimedian copulas, and the mean absolute difference between the transformed data over the past m-observations for Mixture copulas. This transformed data is defined as $u_{i|v_{ij},t} = F_{i|v_{ij}}(x_{i,t}|\mathbf{x}_{v_{ij},t})$ and $u_{i+j|v_{ij},t} = F_{i+j|v_{ij}}(x_{i+j,t}|\mathbf{x}_{v_{ij},t})$.

2.4 Marginal model

This paper works with high-frequency financial returns of stocks traded at the New York Stock Exchange (NYSE). A key issue when modelling intraday data is to account for temporal dependence and intraday seasonality. In this work, I employ the parsimonious multiplicative componenet GARCH model from Engle and Sokalska (2012) to first filter the data and then work with standardized residuals to analyse time-varying dependence across assets.

Define the continuously compounded return as $r_{t,i}$, where t denotes the day and i the regularly spaced time interval at which the return was calculated. Under this model, the conditional variance is a multiplicative product of daily, diurnal and stochastic (intraday) components, so that the return process may be represented as

$$r_{t,i} = \mu_{t,i} + \varepsilon_{t,i}\varepsilon_{t,i} = (q_{t,i}\sigma_t s_i)\, z_{t,i} \tag{6}$$

where $q_{t,i}$ is the stochastic intraday volatility, σ_t a daily endogenously determined forecast volatility, s_i the diurnal volatility in each regularly spaced interval $i, z_{t,i}$ the i.i.d $(0,1)$ standardized innovation. The seasonal (diurnal) part of the process is defined as

$$s_i = \frac{1}{T}\sum_{t=1}^{T}\left(\varepsilon_{t,i}^2/\sigma_t^2\right). \tag{7}$$

Dividing the residuals by the diurnal and daily volatility gives the normalized residuals, defined as

$$\bar{\varepsilon}_{t,i} = \varepsilon_{t,i}/\left(\sigma_t s_i\right)$$

and the the stochastic component of volatility is

$$q_{t,i}^2 = \left(\omega + \sum_{j=1}^{m}\zeta_j v_{jt}\right) + \sum_{j=1}^{p}\alpha_j\bar{\varepsilon}_{t-j}^2 + \sum_{j=1}^{q}\beta_j q_{t-j}^2 \tag{8}$$

In the dependence analysis, for each stock, the standardized residuals are transformed to uniform scores $u_{i,t}$ using the empirical distribution of the data. If the marginal distribution is accurately defined, transforming the standardized residuals using the probability integral transform (PIT) will result in a uniform distribution in [0, 1]. This outcome is essential to identify the copulas during the dependence analysis.

3 Empirical study

3.1 Data

The dataset consist of intraday returns of stocks traded on the NYSE. This research works with intraday returns for 10 U.S. financial stocks with 1-minute frequency obtained from TAQ database for 22 trading days from March 1, 2020 to March 31, 2020. The stocks are JPMorgan (JPM), Capital One Financial Corporation (COF), Citigroup (C), American International Group (AIG), Morgan Stanley (MS), American Express (AXP), Wells Fargo (WFC), Bank of America (BAC), U.S. Bancorp (USB), Goldman Sachs (GS). The group of financial stocks of this research is the same as in Koopman et al. (2018). Figure 1 presents the intraday 1-minute returns for four stocks of the dataset in March 2, 2020. The figure shows that the volatility of returns varies considerably within the day, especially for Morgan Stanley (MS), which exhibit a decreasing volatility behavior. Table 1 presents descriptive statistics for the 10 selected financial companies for all trading days in this analysis. The table reports ticker symbol (Code), mean return (Mean), standard deviation (SD), maximum return (Max), and minimum return (Min). In a typical trading, there are 390 1-minute returns each day.

3.2 Marginal analysis

In order to handle intraday seasonality and temporal dependence, I employ the multiplicative component GARCH$(1, 1)$ model, as outlined in Subsection 2.4. The model assumes that innovations follow a standard Student-t distribution and the mean process for each marginal is described using ARMA(1,1) model. The daily volatility in equation 6 follows an GJR-GARCH$(1, 1)$ model, estimated with a rolling window of 756 days re-fitted every 10 days. Figure 2 panel (a) shows the estimated daily volatility for one selected stock (Citigroup) in 2020. Daily volatility spikes in March and then abruptly decreases in May. The figure shows another spike in July, volatility returns to its initial level only at the end of the year.

The multiplicative component GARCH model is designed to capture intraday volatility patterns, Figure 2 displays key illustrations regarding this intraday dynamics for Citigroup. Panels (b), (c) and (d) consider data from March 1, 2020 until March 31, 2020. This period is notably as one of the most turbulent in US stock market history, affected by the Covid-19 pandemic. On March 16, 2020 the S&P 500 index experiencec a drop of approximately 12% from its lowest point of the day to the previous day's closing price. Figure 2 panel (b) illustrates the estimated intraday volatility seasonality, as in equation 7. Volatility peaks shortly after the market opens, then it gradually decreases. The estimated intraday stochastic component, denoted as $q_{t,i}$ in Equation 8, is depicted in panel (c). Finally, panel (d) displays the total volatility, the plot shows that volatility peaks in the middle of the month.

3.3 Dependence modeling

After the marginal estimations, it's necessary to select the structure of the C-Vine copula model for the 10-dimensional dataset. Trees are selected using maximum spanning trees method with respect to the absolute value of the empirical Kendall's tau, based of the probability integral transformations of the marginal standardized residuals. The tree structure is in table 2. For easy of presentation, I enumerate the financial stocks from 1 to 10 in the following order: JPM (1), COP (2), C (3), AIG (4), MS (5), AXP (6), WFC (7), BAC (8), GS (9), USB (10).

Table 3 presents in-sample model comparison for the entire period of analysis based on model complexity, as denoted by the number of parameters, the goodness of fit (log-likelihood), and two criteria for model selection: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Additionally, an aggregated model labeled "All" and also called the benchmark model, is considered. This model does not adhere to any specific bivariate copula family restrictions, selecting the copula family with the AIC criteria. This inclusion aims to leverage one of the vine copula methods' primary advantages: the flexibility to choose the most suitable copula family for each pair copula independently.

The Student-t and the benchmark model demonstrate the highest log-likelihood values table 3, indicating a superior fit to the data over the other models. Correspondingly, these models also exhibit the lowest AIC and BIC scores. Despite having a high number of parameters, the Mixture copulas BB1, BB7, and BB8 do not show the best performance based on the log-likelihood, and the information criteria, AIC and BIC, which penalize for increased complexity. Additionally, the Clayton and Gumbel models, with the least parameters, present a relatively lower fit to the data.

The inclusion of a model without family restrictions underscores the flexible nature of vine copula methods in model specification, demonstrating the possibility to choose the best fitting copula family for each pair copula. This analysis highlights the critical balance between model complexity and data fit in the selection of statistical models, with the Student-t and the flexible specification "All" emerging as superior across the metrics under evaluation.

Tables 4 and 5 present the in-sample parameter estimates for the most flexible C-vine copula model. In addition to choosing the bivariate copula distribution, I also estimate both a static and a dynamic version of such copulas and we use the AIC information criteria to choose the best of all specifications. The tables show a nuanced dependence structure across different asset pairs. The analysis primarily selects the Student-t distribution, family 2, for both static and dynamic copulas, indicating its superior fit in capturing the tails and asymmetries present in this financial dataset. In the Student-t distribution, parameters θ_1 and θ_2 denote correlation and degrees of freedom, respectively. For dynamic models, additional parameters (ω, a, and b) are provided, reflecting the temporal evolution in dependencies. Notably, only a few pairs are modeled using the BB8 (Family 10) and Survival BB8 (Family 20) families, suggesting only specific cases where these families offer a better fit according to the AIC. For BB8 copulas, parameters θ_1 and

θ_2 denote the estimated parameters of the BB8 copula (Joe, 1997).

Tables 4 shows the that static Student-t bivariate copulas dominate in the first tree, where nearly all selections are static with only one dynamic exception. This indicates that static relationships are initially considered adequate for capturing the interactions between assets. As we progress in the tree structure, it exhibits a notable shift towards incorporating more dynamic copulas, reflecting an adjustment to better capture the evolving dependencies among assets. Despite this shift, in the later trees, depicted in table 5, static versions present a better fit, highlighting the complexity of asset relationships and the necessity to tailor the copula choice to specific asset pairs.

3.4 Out-of-sample forecasting

In this subsection I conduct an out-of-sample forecasting analysis, predicting the copula density of returns over the next 1-minute. This analysis considers all models presented in this study, the Elliptical, Archimedian and Mixture copulas, including the rotated specifications for Archimedian and Mixture copulas. All models in this comparison utilize the same marginal distributions and, consequently, the same probability integral transforms. Therefore, any observed differences in performance are directly attributable to how each model approaches and captures the dynamics of dependence.

For the estimation strategy I adopt a dynamic approach by estimating the static parameters over a fixed period of 10 trading days. These parameters are then utilized to predict the return density for each subsequent 1-minute interval over the following five days. After this period, we update the parameter estimates to reflect the most recent data, and predict the following five days.

As in Koopman et al. (2018), I evaluate the copula probability density function forecasts of the different models through the sum of the log score of the whole day,

$$S_t(M_j) = \sum_{i=1}^{390} \log c_t(u_i; R_i \,|\, \mathcal{F}_{i-1}), \tag{9}$$

where i is the 1-minute interval and R_i is the one-step-ahead forecast of each model j.

The evaluation of the forecasting performance through the sum of log scores reveals considerable variability in the effectiveness of different copula models in predicting the 1-minute ahead returns density. The analysis of Diebold-Mariano (DM) test statistics further corroborate these findings, offering a statistical comparison of forecast accuracy between each copula model and a benchmark.

In table 6, the "Mean" row represents the average sum of log scores over the forecasting period for each model, indicating the overall performance in capturing the returns density. The Δ_{LL} row shows the difference in performance relative to the most flexible model that combines several families, denoted "All". A positive Δ_{LL} indicates a model underperformance compared to the benchmark model, whereas a negative Δ_{LL} suggests superior performance. The "DM" row reports the Diebold-Mariano test statistic for each model against the benchmark, where a statistically significant positive value would indicate that the model in question performs significantly worse than the "All" model, and a negative value suggests significantly better performance.

The Normal copula does not perform as well as the most flexible model, as evidenced by its higher mean score and a positive DM statistic, indicating its inferior forecasting ability in this forecasting period. The average score value for the C-vine copula with normal bivariate links is 1213.0, while the value for the benchmark model is 1296.1. In contrast, the Student-t copula demonstrates a near-equivalent performance to the benchmark model, with a low Δ_{LL} and DM statistic, illustrating that the difference in their forecasting accuracy is statistically insignificant. This suggests the Student-t copula's robustness in capturing the dynamics of financial returns within the 1-minute forecasting horizon.

Contrarily, the Clayton copula significantly underperforms, highlighted by its substantially lower mean score and the highest positive DM value. Meanwhile, the Gumbel copula and the Mixture copulas (BB1, BB7, BB8) exhibit intermediate performance. Specifically, the BB8 specification is the model with the best between the three mixture copulas. However, the performance still statistically inferior when compared to the benchmark model.

4 Conclusion

This paper investigates intraday patterns in the dependencies of financial stock returns. It evaluates whether accounting for flexible dependencies improves the forecasting accuracy of comovements. By employing a dynamic canonical vine copula method, this work has developed a model to capture flexible dependence patterns, including time-varying upper/lower tail dependence and asymmetric dependence among financial assets. Building on previous research that underscores the importance of modeling key features of financial data, this study applies a pair copula decomposition approach to analyze complex dependencies with simple pair-copulas, incorporating time-varying dependence structures through autoregressive dynamics into copula parameters.

This study has analyzed 1-minute frequency returns of 10 U.S. financial stocks throughout March 2020. After an initial exploratory analysis, it evaluates the fit of various bivariate copulas, Elliptical, Archimedean, and Mixture copulas, finding that the Student-t distribution provides the best fit in-sample. I have documented intraday variations in dependence measures. In particular, time-varying pair-copulas are selected most of the time in the second level and third level of the C-vine copula, while static copulas provided a better fit in deeper trees. The performance of models without distributional restrictions is comparable to those using exclusively Student-t pairs. As noted by Nikoloulopoulos et al. (2012), pair-copulas with Student-t distribution tend to be best based on a likelihood or AIC comparison. However, for inferences involving tails, the strategy to choose the pair-copula family should not just be likelihood-based but also depend on tail dependence measure and extreme quantiles. In future research, I aim to address this point in a backtesting *Value at Risk* analysis.

In this work, I have also explored out-of-sample forecasting with the C-vine copula model employing diverse bivariate links. The most flexible model, selecting bivariate families from multiple distributions, shows statistically superior performance according to Diebold-Mariano test, except when compared to the C-vine model with Student-t distributions, which performs equally well. Finally, incorporating asymmetric tail dependencies with the Mixture copula have not improved forecasting accuracy based on log-likelihood measures.

References

Aas, K., Berg, D., 2009, Models for construction of multivariate dependence – a comparison study, The European Journal of Finance 15(7-8), 639–659.

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009, Pair-copula constructions of multiple dependence, Insurance: Mathematics and Economics 44(2), 182–198.

Andersen, T. G., Bollerslev, T., 1997, Intraday periodicity and volatility persistence in financial markets, Journal of Empirical Finance 4, 115–158.

Andersen, T. G., Thyrsgaard, M., Todorov, V., 2021, Recalcitrant betas: Intraday variation in the cross-sectional dispersion of systematic risk, Quantitative Economics 12, 647–682.

Ang, A., Chen, J., 2002, Asymmetric correlations of equity portfolios, Journal of Financial Economics 63, 443–494.

Ausin, M. C., Lopes, H. F., 2010, Time-varying joint distribution through copulas, Computational Statistics Data Analysis 54(11), 2383–2399.

Bedford, T., Cooke, R., 2001, Probability density decomposition for conditionally dependent random variables modeled by vines, Ann. Math. Artif. Intell. 32, 245–268.

Bedford, T., Cooke, R., 2002, Vines - a new graphical model for dependent random variables, Annals of Statistics.

Bollerslev, T., 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model, The Review of Economics and Statistics 72(3), 498–505.

Engle, R., 2002, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, Journal of Business Economic Statistics 20, 339–50.

Engle, R. F., Sokalska, M. E., 2012, Forecasting intraday volatility in the US equity market. Multiplicative component GARCH, The Journal of Financial Econometrics 10, 54–83.

Joe, H., 1996, Families of \$m\$-variate distributions with given margins and \$m(m-1)/2\$ bivariate dependence parameters.

Joe, H., 1997, Multivariate Models and Multivariate Dependence Concepts (1st ed.), Chapman and Hall/CRC.

Koopman, S. J., Lit, R., Lucas, A., Opschoor, A., 2018, Dynamic discrete copula models for high-frequency stock price changes, Journal of Applied Econometrics 33, 966–985.

Kurowicka, D., Cooke, R., 2006, Uncertainty analysis with high dimensional dependence modelling., Wiley, New York.

Manner, H., Reznikova, O., 2012, A survey on time-varying copulas: Specification, simulations, and application, Econometric Reviews 31, 654–687.

Nelson, R., 2006, An introduction to copulas, Springer-Verlag, New York.

Nikoloulopoulos, A. K., Joe, H., Li, H., 2012, Vine copulas with asymmetric tail dependence and applications to financial return data, Computational Statistics Data Analysis 56, 3659–3673.

Oh, D. H., Patton, A. J., 2017, Modeling dependence in high dimensions with factor copulas, Journal of Business & Economic Statistics 35, 139–154.

Oh, D. H., Patton, A. J., 2023, Dynamic factor copula models with estimated cluster assignments, Journal of Econometrics 237, 105374.

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 & Economic Statistics 39, 1066–1079.

Patton, A., 2006, Modelling asymmetric exchange rate dependence, International Economic Review 47(2), 527–556.

Patton, A., 2013, Copula Methods for Forecasting Multivariate Time Series, Handbook of Economic Forecasting, Elsevier.

Sklar, A., 1959, Fonctions de répartition à n dimensions et leurs marges, Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231.

Tófoli, P., Ziegelmann, F. A., Silva Filho, O., Pereira, P., 2019, Dynamic d-vine copula model with applications to value-at-risk (var), Journal of Time Series Econometrics 11, 20170016.

Tse, Y. K., Tsui, A. K. C., 2002, A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, Journal of Business & Economic Statistics 20(3), 351–362.

Code	Mean	SD	Max	Min
JPM	1.59	0.0028	0.0319	-0.0350
COF	-2.78	0.0033	0.0377	-0.0440
C	-0.34	0.0035	0.0400	-0.0410
AIG	-1.88	0.0037	0.0479	-0.0342
MS	1.57	0.0030	0.0284	-0.0272
AXP	0.65	0.0029	0.0285	-0.0290
WFC	-0.93	0.0029	0.0389	-0.0385
BAC	1.53	0.0029	0.0330	-0.0314
USB	1.59	0.0030	0.0323	-0.0290
GS	1.16	0.0026	0.02361	-0.0274

Table 1: Descriptive statistics of 10 ten selected stocks for all trading days in March, 2020. The table reports ticker symbol (Code), mean return (Mean), standard deviation (SD), maximum return (Max), and minimum return (Min).

Tree	**Edge**	**Tree**	**Edge**	**Tree**	**Edge**	**Tree**	**Edge**
1	10,1	3	10,9;5,1	5	9,8;2,10,5,1	7	7,4;9,8,2,10,5,1
1	1,5	3	10,2;5,1	5	8,7;2,10,5,1	7	4,6;9,8,2,10,5,1
1	1,9	3	10,8;5,1	5	8,4;2,10,5,1	7	4,3;9,8,2,10,5,1
1	1,2	3	10,7;5,1	5	8,6;2,10,5,1	8	7,6;4,9,8,2,10,5,1
1	1,8	3	10,4;5,1	5	8,3;2,10,5,1	8	7,3;4,9,8,2,10,5,1
1	1,7	3	10,6;5,1	6	9,7;8,2,10,5,1	9	6,3;7,4,9,8,2,10,5,1
1	1,4	3	10,3;5,1	6	9,4;8,2,10,5,1		
1	1,6	4	9,2;10,5,1	6	9,6;8,2,10,5,1		
1	1,3	4	2,8;10,5,1	6	9,3;8,2,10,5,1		
2	10,5;1	4	2,7;10,5,1				
2	5,9;1	4	2,4;10,5,1				
2	5,2;1	4	2,6;10,5,1				
2	5,8;1	4	2,3;10,5,1				
2	5,7;1						
2	5,4;1						
2	5,6;1						
2	5,3;1						

Table 2: Tree structure of C-vine copula

	Number of parameters	Log-likelihood	AIC	BIC
Normal	57	716.88	-1431.78	-1424.72
Student-t	58	762.71	-1520.85	-1504.69
Clayton	45	530.50	-1059.02	-1051.96
Gumbel	45	684.20	-1366.42	-1359.36
BB1	90	734.95	-1465.92	-1451.8
BB7	90	700.20	-1396.4	-1382.29
BB8	90	742.36	-1480.73	-1466.62
All	58	762.77	-1520.98	-1504.82

Table 3: This table presents a comparison of various statistical models based on their number of parameters, log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) scores. Models include Normal, Student-t, Clayton, Gumbel, BB1, BB7, BB8, and a combined model labeled "All". Each row details the model's performance metrics.

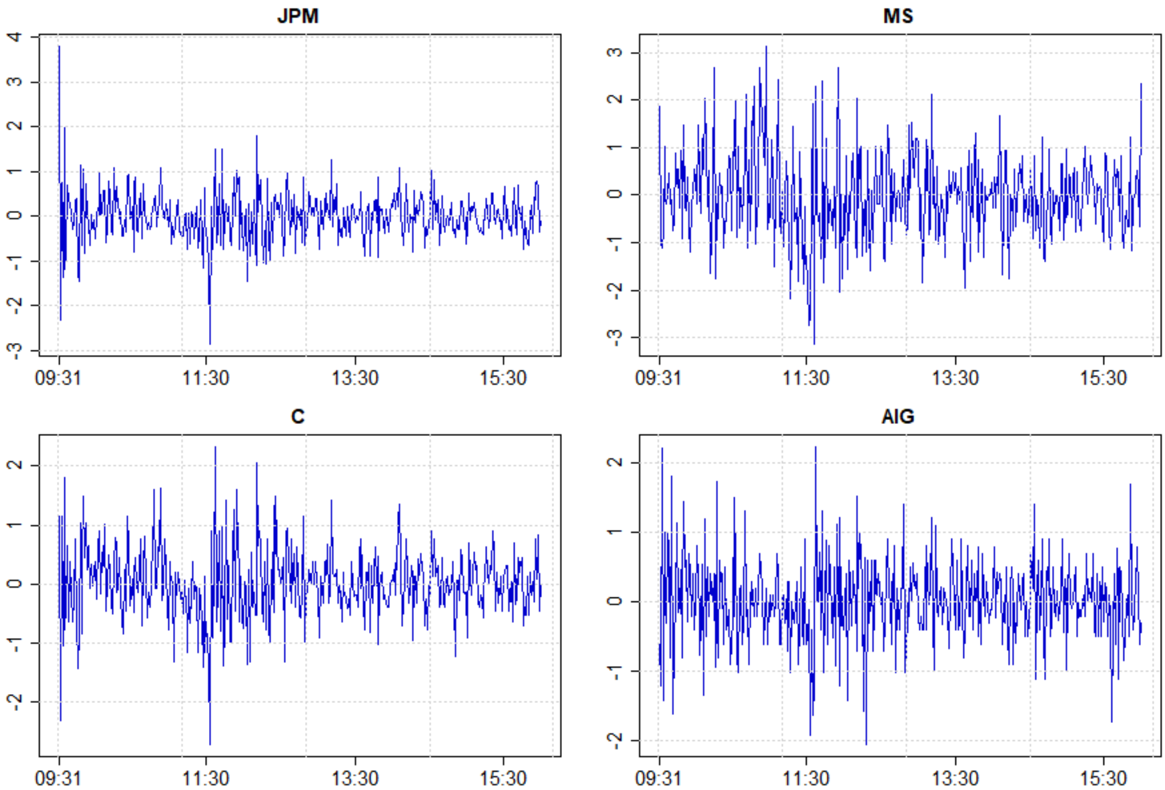

Figure 1: Intraday 1-minute returns for four stocks of the dataset in March 2, 2020. The stocks are JPM, MS, C, and AIG.

Pair	Type	Family	θ_1	θ_2	ω	a	b
1_5	static	2	0.715	7.541			
1_10	static	2	0.726	6.566			
1_2	static	2	0.62	9.981			
1_8	static	2	0.776	8			
1_9	static	2	0.716	8.735			
1_4	dynamic	2		11.951	0.539	0.961	0.029
1_7	static	2	0.709	8.641			
1_6	static	2	0.622	9.135			
1_3	static	2	0.743	7.181			
5_10	dynamic	2		9.571	0.351	0.683	0.039
5_2	static	10	3.874	0.498			
5_8	dynamic	2		11.026	0.379	0.976	0.012
5_9	static	2	0.464	10.326			
5_4	static	2	0.271	13.878			
5_7	dynamic	2		10.71	0.318	0.986	0.01
5_6	dynamic	2		11.88	0.287	0.995	0.005
5_3	dynamic	2		10.728	0.407	0.977	0.015
10_2	static	20	4.243	0.322			
10_8	dynamic	2		12.521	0.262	0.985	0.005
10_9	dynamic	2		17.284	0.187	0.902	0.009
10_4	static	20	3.978	0.357			
10_7	dynamic	2		9.855	0.334	0.999	0.001
10_6	dynamic	2		24.986	0.186	0.998	0.001
10_3	dynamic	2		12.115	0.239	0.991	0.004

Table 4: In-sample parameter estimates for the most flexible C-vine copula model for all trading days in March, 2020.

Pair	Type	Family	θ_1	θ_2	ω	a	b
2_8	static	2	0.128	18.439			
2_9	dynamic	2		18.954	0.15	0.998	0.001
2_4	dynamic	2		28.555	0.161	0.995	0.004
2_7	static	2	0.094	23.731			
2_6	static	2	0.253	17.658			
2_3	static	2	0.149	18.298			
8_9	static	2	0.133	14.367			
8_4	static	10	1.178	0.823			
8_7	static	20	6	0.246			
8_6	static	2	0.06	21.576			
8_3	static	2	0.224	12.261			
9_4	static	2	0.096	20.914			
9_7	static	2	0.099	14.343			
9_6	static	2	0.137	18.27			
9_3	static	2	0.111	22.229			
4_7	static	2	0.064	30			
4_6	static	20	1.727	0.496			
4_3	static	2	0.082	30			
7_6	static	2	0.034	30			
7_3	static	2	0.085	18.943			
6_3	static	2	0.039	25.611			

Table 5: In-sample parameter estimates for the most flexible C-vine copula model for all trading days in March, 2020.

	All	Normal	Student-t	Clayton	Gumbel	BB1	BB7	BB8
Mean	1296.1	1213.0	1298.6	903.7	1130.7	1240.0	1206.8	1265.4
Δ_{LL}		82	-2.69	392	165	56	90	31
DM		1.67	-0.31	8.34	0.4	0.8	3.25	1.65

Table 6: Out-of-sample comparison of various statistical model based on different copula families. The "Mean" row represents the average sum of log scores. The benchmark model is the one denoted as "All". Δ_{LL} represent the deviations from the benchmark in term of average sum of log scores. The "DM" row reports the Diebold-Mariano test statistic for each model against the benchmark model.

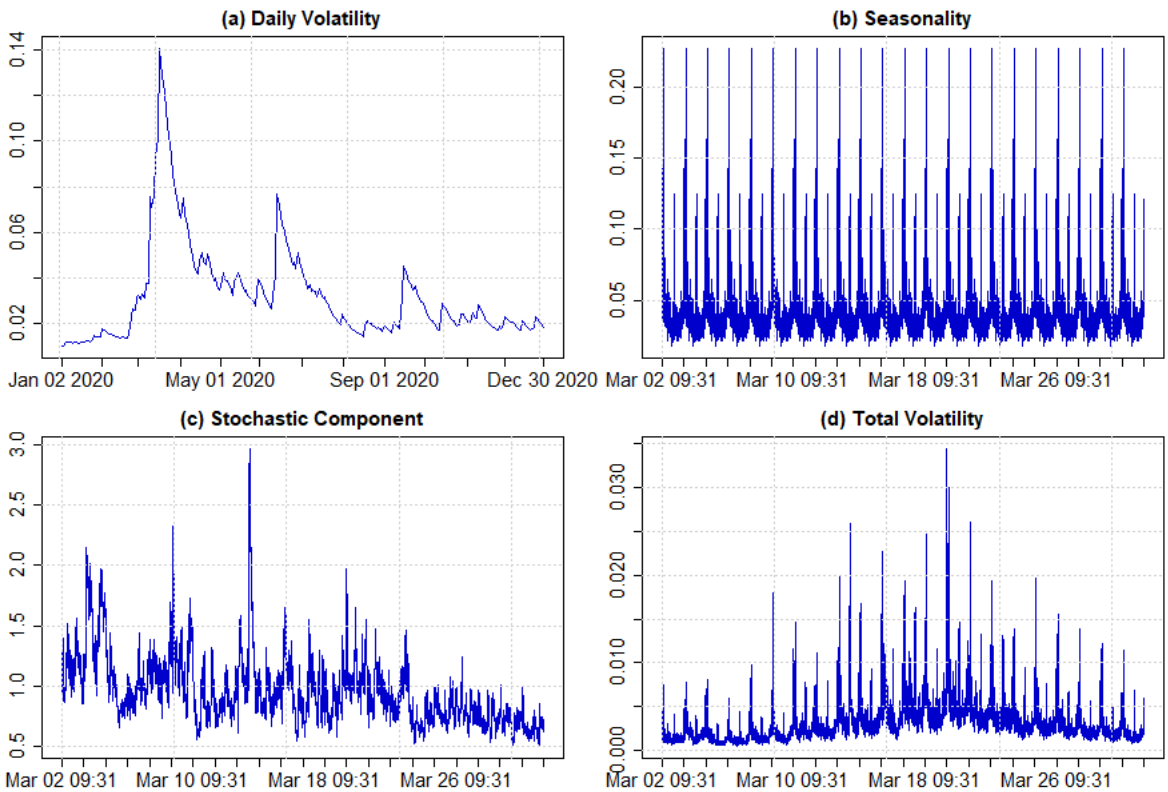

Figure 2: Panel (a) shows the estimated daily volatility for Citygroup in 2020. Panels (b), (c) and (d) present the seasonal component, stochastic component, and total volatility, respectively. The financial stock is Citigroup, data ranges from March 1, 2020 until March 31, 2020.