Stock Price Forecasting Using A Dependence Structure

Paul Mucci, Eun-Joo Lee, and Seung-Hwan Lee

***Abstract* — It is important to incorporate diverse dependence structures between stocks when managing a stock portfolio. Copulas are a useful statistical tool to capture dependence structure, dealing with both the linear and non-linear association that may occur in the tails of data. Financial time series datasets often exhibit volatility clustering that affects price forecasting accuracy. This work proposes the initial use of the principal component analysis followed by a copula and GARCH model that filters the effect of the volatility clustering in the series. For illustration, we consider ten banks from which Bank of America and PNC Financial Services Group are chosen, and then we project their future price movements through simulations. Since they are selected in terms of the principal component analysis, the procedures could help the proposed model to become a more widely used tool in forecasting financial stock performance.**

***Keywords* — Banking industry, copula, GARCH, PCA, price forecasting, tail dependence.**

I. Introduction

Accurate stock research is crucial to successful active investment management in a dynamic marketplace. Analysts rely on historical pricing to model future returns for companies, and consistently attempt to outperform the market on a risk adjusted basis. Traditional equity research entails forecasting future returns on company financials (cash flows, balance sheets and income statements), then deriving fair value pricing on a per year end basis. Newer research methods have begun leveraging statistical models, focused solely on past pricing and technical analysis rather than company financials. Statistical methods generate ranges of future pricing rooted in probability and variance.

Banking provides economic insight on many levels. The banking industry falls under the financial sector, providing liquidity, risk transfer and currency access for consumers. In the markets, banking stocks bridge the gap between equity and fixed income vehicles due to heavy influence from the interest rate environment. Consistent banking returns indicate economic strength, and likewise banks tend to be hit the hardest during recessionary periods. The impact of banking on an economy combined with historical tail dependences in pricing provides researchers with reasonable insight into future sector movements.

Many challenges arise when dealing with large data sets with many variables (or features). Examples include the difficulty in handling, modeling, and interpreting data. Principal component analysis (PCA hereafter) is one way to resolve such issues, making possible to identify trends and patterns of data sets in higher dimensions. It is a statistical technique for reducing the dimensionality of a data set while preserving as much information as possible. By reducing the dimensionality, some important characteristics of the data become distinguishable. This results in fewer new (uncorrelated) variables called principal components that are easier to analyze. Readers can refer to [1] for details. Since PCA minimizes the loss of information of an original data set, interpretations with the principal components, whose dimensions are lower than that of the original data set, will reflect the patterns and trends of the data set well. In our work, Bank of America (BAC hereafter) and PNC Financial Services Group (PNC hereafter) are chosen by PCA out of ten banks that play major roles in U.S. markets. The initial ten selected banks are intentionally diverse: Market caps range between \$83.58 - \$440.29 billion, and assets under management (AUM) span \$521 billion to \$3.758 trillion-Bloomberg. Per PCA, banks with the highest historical correlation to the sector are selected as the best predictors of future movement of the sector. Forecasting performance of the two most correlated banks (PNC & BAC) will implicitly forecast returns for the entire sector. Thus, forecasting performances of the two banks, BAC, and PNC, will implicitly forecast returns for the entire sector.

Financial time series datasets often exhibit volatility clustering varying over time (see [2]), a concept referred to as "heteroscedasticity." This may lead to under or over-forecasting if only a mean model is utilized. Such time-variant variance appears in residuals of BAC and PNC as well. The generalized

Published on May 20, 2022.
P. Mucci, Illinois Wesleyan University, USA.
(e-mail: pmucci@iwu.edu)
E-J. Lee, Millikin University, USA.
(email: elee@millikin.edu)
S-H. Lee, Illinois Wesleyan University, USA.
(corresponding email: slee@iwu.edu)

autoregressive conditional heteroskedastic (GARCH hereafter) model (see [3] and [4]) is a way to overcome the heteroscedastic issue. In this work, to deal with the conditional mean and variance of the data, we employ a combination of the two models, ARMA (autoregressive moving average) and GARCH. In particular, the ARMA(1,1)-GARCH(1,1) model is used for this data.

Linear correlation coefficients may not be appropriate to measure association between two or more variables, when data contains extreme values at the tails. It is known that financial data shows a tendency of such tail dependence, which is consequently misleading when estimating total required economic capital. For stock price data, a large gain or loss will have a large impact on future prices. Thus, tail events need to be considered in data analysis when modeling. Copulas are useful tools to model multivariate distributions, capturing various possible tail dependence structures. In this work, we utilize Student's t copula in [5] for BAC and PNC filtered by the ARMA(1,1)-GARCH(1,1).

This paper is organized as follows: Section II provides an overview of methods employed for this work, including PCA, time series modeling and dependence modeling. The procedures of applying the methods to the data and its numerical results are presented in Section III. Section IV gives concluding remarks.

II. Method

A. Principal Component Analysis

Configuration issues are prevalent in statistical analysis with high dimensional datasets, in which there are many features through variables. An example includes high dimensional sparse data that may lead to complexity in modeling, or erroneous conclusions. Principal Component Analysis (PCA) is one way to deal with data of high dimensionality, explaining most of the variability using a fewer number of the variables. This useful statistical technique simplifies large datasets by reducing its dimensionality in terms of deleting highly correlated features, preserving as much variance in the data as possible. That is, it characterizes the new feature space by the eigenvectors and eigenvalues of a covariance, making it possible to transform data from a highly dimensional space into a lower dimensional space. Some features with relatively low information in the eigenvalues are eliminated to obtain fewer principal components in a lower dimensional space into which data are projected.

Let $X_j, j = 1, \ldots, n$, be a vector of data, where n is the number of input variables. An eigenvector that becomes a principal component that corresponds to the maximum eigenvalue accounts for the greatest portion of the variance in the data, creating a new feature reflecting a characteristic of data in a high dimensional space. For example, the first principal component of the data is obtained by the linear combination of X_j's, $Y_1 = \sum_{j=1}^{n} e_{1j} X_j$, that has the maximum variance, or equivalently the maximum eigenvalue, among all linear combinations through e_{1j} such that $\sum_{j=1}^{n} e_{1j}^2 = 1$. The second component, Y_2, is obtained in the same way from all linear combinations that are uncorrelated to Y_1. Let $\lambda_i, i = 1, \ldots, n$, denote the eigenvalues. We then repeat this process until all principal components are constructed, and then select a few of the largest $\lambda_i's$ from the eigenvalues sorted from the largest to the smallest, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, in order to constitute a low dimensional space through a few principal components chosen by the chosen $\lambda_i's$.

For our data, the first principal component has the greatest percentage (77%) of the variance retained followed by the second principal component (15%). This implies the first and second principal components account for the proportion of 92% of the variation in the data. See Table I for details. Note that in the table, the first principal component is denoted by PC1, the second principal component is denoted by PC2, and other principal components are denoted in the same manner. Figure 1 shows the corresponding scree plot drawn across all principal components 1 - 10, displaying cumulative percentage of explained variance. Ten banks described in Table II are considered, from which two banks are identified as the most influential by the correlation on the new features, PC1 and PC2. The two banks are BAC and PNC.

It is worth noting that the initial PCA on a selected group of stocks will determine how useful the model will be; Highly correlated stocks will yield better results than selected stocks with lower correlations. Additionally, the PCA implicitly delivers insight regarding which stocks in a sector are most closely related to a sectors aggregate index.

TABLE I: PERCENTAGE OF VARIANCE EXPLAINED BY PRINCIPAL COMPONENTS

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
*SD	2.787	1.242	0.615	0.326	0.317	0.218	0.154	0.109	0.102	0.090
PV	0.776	0.154	0.037	0.011	0.011	0.005	0.002	0.001	0.001	0.001
CP	0.777	0.931	0.969	0.979	0.989	0.994	0.997	0.998	0.999	1.000

*SD = standard deviation, PV = percentage of variance (×100%), CP = cumulative percentage (×100%).

TABLE II: CORRELATION BETWEEN BANK AND PRINCIPAL COMPONENTS

	*BAC	C	GS	JPM	MS	PNC	SCHW	TF	USB	WFC
PC1	0.347	0.305	0.324	0.332	0.316	0.355	0.339	0.343	0.309	0.121
PC2	-0.106	0.342	-0.181	-0.224	-0.331	-0.071	-0.088	0.078	0.356	0.730

*Banks are listed by ticker as follows: Bank of America (BAC), Citigroup (C), Goldman Sachs (GS), J.P. Morgan (JPM), Morgan Stanley (MS), PNC Financial Services Group (PNC), Charles Schwab (SCHW), Truist Financial (TF), U.S. Bancorp (USB) & Wells Fargo (WFC).

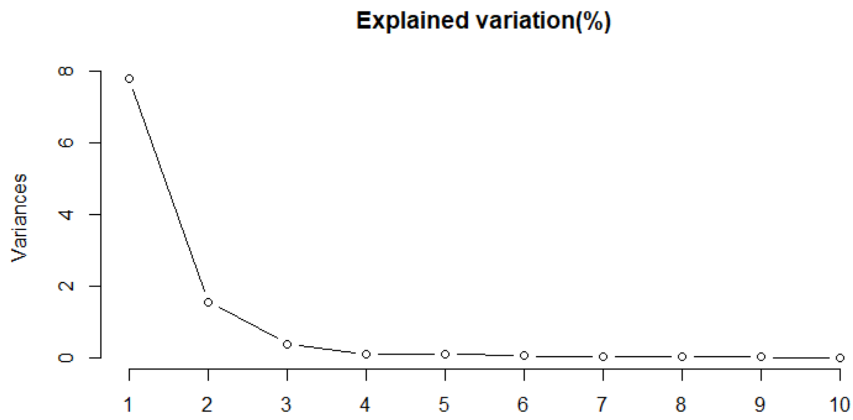

Fig. 1. PCA scree plot of the data.

B. Time Series Modeling

Let x_t denote the values of a time series at time t. Then, the value of the series one time before time t can be represented by x_{t-1}. If the present value depends on the proceeding value, the present value of the series is represented by the autoregressive (AR) model. The AR model of order p, denoted by AR(p), states that given the past, the conditional mean of x_t is $\varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p}$. For example, the AR(1) model is used for forecasting when a value is determined by the immediately preceding value of the series. Another model, along with the AR model, that accounts for the autocorrelation is the moving average, or MA, model. The MA model with the mean zero case of order q, denoted by MA(q), is $x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$ which implies past errors. With the combination of the AR(p) and MA(q) models, the general autoregressive moving average model of p and q, denoted by ARMA(p,q), postulates

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t. \tag{1}$$

For financial data, Student's t distribution is typically assumed for the errors, $\boldsymbol{\varepsilon_t}$**,** referred to as innovations or shocks. Although the ARMA model has many advantages as a conditional mean model in the short-term time series forecasting (see [6]), its usage is restricted to cases of homoscedasticity. This assumption is often violated since financial time series data generally exhibits non-constant variance (see [2]). Specifically, financial time series datasets often exhibit volatility clustering varying over time, so the ARMA model may lead to under or over-forecasting. One way to overcome this issue is to use the GARCH model (see [3], [4]) which filters the effect of the volatility clustering in the series. The GARCH(p,q)

$$\varepsilon_t = \sigma_t \eta_t,$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{2}$$

model is

where $\boldsymbol{\alpha_0 > 0}$, $\boldsymbol{\alpha_i \geq 0, \beta_j \geq 0}$ such that $\boldsymbol{\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1}$, and η_t is a random sample with mean 0 and variance 1. See [4] and [11] for details.

In this work we utilize a GARCH model, together with an ARMA model to specify conditional mean and conditional variance (heteroscedasticity). Particularly, the ARMA(1,1)-GARCH(1,1) model with Student's t-distribution innovations was chosen based on the AIC (Akaike Information Criteria), whose procedures are described in *1)* of *B* in section III.

It is known that sets of financial data have significant tail events that occur with low probability at the ends of tails of a distribution. To model such tail dependence, copula can be employed, which will be described in the following section. We use copula to model dependence structures of the financial time series filtered by the ARMA-GARCH model in the development of forecasting procedures.

C. Dependence Modeling

1) Student's t copula

Copula is a distribution function with p -dimensions on an interval $[0,1]^p$, with standard uniform marginal distributions. Let C be a copula and suppose that $F_1,\ldots,F_p$ are univariate distribution functions. Then, $C(F_1(x_1),\ldots,F_p(x_p))$ represents a multivariate distribution function with marginal distributions $F_1,\ldots,F_p$. Then, every multivariate distribution function F can be written as

$$F(x_1,\ldots,x_p) = C(F_1(x_1),\ldots,F_p(x_p)) \tag{3}$$

for some copula C according to Sklar's theorem (see [7] and [8]), which is uniquely determined on the interval $[0,1]^p$ for the multivariable distribution. Put another way, the distribution function F is a multivariate distribution function with $F_1,\ldots,F_p$, if C is a copula and $F_1,\ldots,F_p$ are marginal distribution functions.

Let $X = (X_1,\ldots,X_p)$ be a random vector that has marginal distribution functions. Assume that the marginal distributions are continuous and strictly increasing. Then, a copula function of their multivariate distribution can be uniquely determined by a multivariate distribution function F from (3) by obtaining

$$C(u_1,\ldots,u_p) = F(F_1^{-1}(u_1),\ldots,F_p^{-1}(u_p)) \tag{4}$$

where $F_1^{-1},\ldots,F_p^{-1}$ are respective inverse functions of $F_1,\ldots,F_p$. Thus, the copula of the multivariate distribution function F can be regarded as the distribution function of componentwise transformations of the random vector $X = (X_1,\ldots,X_p)$. Note that the copula remains invariant under the assumption that transformations of the components of the vector X are strictly increasing, which is a salient feature of copula compared to the correlation coefficient.

In this work, we consider a Student's t copula (see [5] and [9]). Capturing extreme events is important in studying the overall financial market pattern, since such events, like unexpected political tension for example, could result in stress on the market (see [10]). Student's t copula specifies different levels of correlations in the tails, making it useful in analyzing the effect of the dependence structure on the extreme events that have heavy-tailed distributions. Student's t copula of X given by (4) is derived from the multivariate t distribution (see [5]),

$$C_{\nu,R}^t(u_1,\ldots,u_p) = t_{\nu,R}(t_\nu^{-1}(u_1),\ldots,t_\nu^{-1}(u_p)) \tag{5}$$

where t_ν is the marginal distribution of a multivariate distribution function $t_{\nu,R}$ with degrees of freedom ν and correlation coefficient matrix R, and t_ν^{-1} denotes the function of a univariate t distribution with the degrees of freedom ν which controls the heaviness of tails of distributions.

Simulation from the copula, $C_{\nu,R}^t$, is simple to perform. To generate a multivariate t-distributed random vector X, an algorithm is based on the equation (3) that leads to $U = (t_\nu(X_1),\ldots,t_\nu(X_p))$. We can impose any univariate marginal distributions on Student's t copula. In the next section, scatter plots of simulated data generated by the algorithm are used for visual exploration of Student's t copula, particularly in the tails of the distributions.

2) Tail Dependence

Characterization of tail dependence of two or more random variables is the most useful feature of a copula. To put it simply, let X_1 and X_2 be continuous random variables that have marginal distributions F_1 and F_2, respectively. Then, the dependence coefficients in the tails of the bivariate distribution of X_1 and X_2 are defined as

$$\begin{aligned}\lambda_L(X_1,X_2) &= \lim_{u\downarrow 0} P(X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u)),\\ \lambda_U(X_1,X_2) &= \lim_{u\uparrow 1} P(X_2 > F_2^{-1}(u) | X_1 > F_1^{-1}(u))\end{aligned} \tag{6}$$

respectively, where λ_L and λ_U represent the lower and upper tail dependence coefficients. See [5] and [11] for details. The tail dependence expression says the probabilities of having a high (low) extreme value of X_2 given that a high (low) extreme value of X_1 occurs. For example, large values of λ_U imply that the occurrence of joint extreme values is more likely, when compared to low values of λ_U. If $\lambda_U = 0$, then X_1 and X_2 are said to be asymptotically independent in the upper tail, i.e., no tail dependence. If λ_U lies in (0,1], then X_1 and X_2 are asymptotically dependent in the upper tail. Interpretation for the lower tail dependence coefficient λ_L is analogous. The two coefficients, the measures of λ_L and λ_U, are equivalent for Student's t copula, because of the symmetric property of the distribution. Let $\lambda_L = \lambda_U = \lambda$. Then, the tail dependence coefficient is easily calculated by using the following formula,

$$\lambda = 2t_{\nu+1}(-\sqrt{\nu+1}\sqrt{1-R_{12}}/\sqrt{1+R_{12}}) \tag{7}$$

with R_{12} being the off-diagonal element of R .

Table III presents coefficients of tail dependence of Student's t copula, showing associations between two variables in the tails that vary with different levels of degrees of freedom and correlation coefficients as described in [12]. The coefficient of tail dependence increases as the value of R_{12} and ν increases and decreases, respectively. This in turn results in the increased strength of tail dependence, implying suitability of Student's t copula for extreme events.

TABLE III: COEFFICIENTS OF TAIL DEPENDENCE FOR STUDENT'S t COPULA

	*-1	-0.9	-0.8	-0.5	-0.2	0	0.2	0.5	0.8	0.9	1
**1	0	0.1024	0.1474	0.2468	0.3333	0.3968	0.4544	0.5641	0.7196	0.8003	1
2	0	0.0171	0.0351	0.0955	0.1679	0.2254	0.2929	0.4226	0.6220	0.7295	1
3	0	0.0032	0.0093	0.0405	0.0917	0.1393	0.2010	0.3318	0.5527	0.6776	1
10	0	0	0	0.0002	0.0023	0.0078	0.0220	0.0845	0.2948	0.4643	1
30	0	0	0	0	0	0	0.0001	0.0031	0.0733	0.2113	1
∞	0	0	0	0	0	0	0	0	0	0	1

*Linear correlation coefficient ($R_{12} = R$), ** Degrees of freedom (ν)

In order to visually explore behaviors of Student's t copula, scatterplots are shown in Fig. 2. The plots are obtained based on simulated data generated by the algorithm to be described in section 3.2.1. For various degrees of freedom, ν=2, 10, 25, 50, with $R_{12} = 0.75$, the scatterplots demonstrate the associations of variables, graphically evaluating the tail dependence according to the thickness of tails. Note that the plots tend to extend to the top right corner and the bottom left corner as the degrees of freedom gets increased. For example, the form of tail dependence with ν=2 is less noticeable in the plot than ν=50. This implies that the heavier the tails, the stronger the dependence is in the tails. This kind of dependence structure of the distributions is not well defined by the correlation alone.

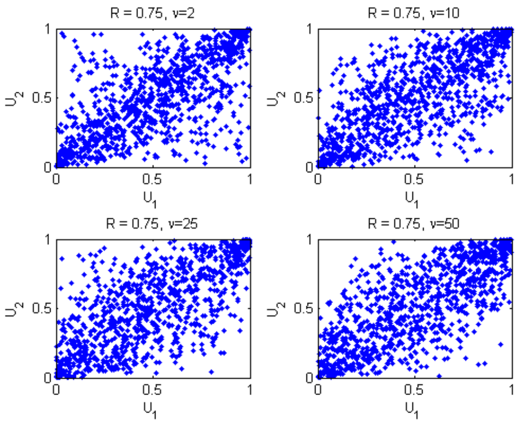

Fig. 2. Tail dependence structure of Student's t -copula.

III. Application

A. Data Description

1) Banking and the Economy

The financial sector provides value exchange, intermediation, risk transfer and liquidity to the economy. Banks, insurance firms, real estate brokers, consumer finance companies, mortgage lenders and REITs are all components of this sector. The duty of the financial sector in an economy is to ensure sustainable growth through facilitating efficient allocation of resources.

Banks are responsible for providing liquidity and risk transfer to consumers. Liquidity offers borrowers the ability to finance investments through raising debt, resulting in the development of new business and technological innovation. Banks also transfer risks to their balance sheets, shielding consumers from heightened risks of decentralized loaning. By accepting the consequences of defaulting loans, banks stimulate economic growth by shouldering economic risk. It has been found that the higher the [developed a] banking sector is, the higher [economic] output production exists. See [13] for details.

2) Banking Trends

Online banking, AI and Fintech are all drivers of the digitalization of banking. In 2006, an estimated 80% of banks offered online banking services (see [14]), and that number has only increased. Today, 89% of U.S. households that bank utilize mobile banking in some aspect, and 70% of those individuals use mobile banking as the primary means of handling their accounts (see [15]).

Hybrid work models and contract employee utilization have grown significantly post pandemic. It is estimated that nearly 50% of the U.S Labor force currently works from home, and 54% plan to continue that trend (see [16]). The gig economy represents 60 million workers in the U.S., and by 2027 it is estimated that the majority of U.S. employees will be contract based. Banks are using contract work to outsource non-core jobs such as marketing, HR and procurement (see [17]).

Due to inflation running at 30-year highs and fiscal policy heavily focused on spending, the FED (Federal Reserve Bank) has signaled interest rate hikes as soon as March 2022. The FED of Kansas City President Esther George has stated that the FED's current stance is “out of sync” with the economy. Likewise, the FED of Philadelphia has supported a quarter percent rate hike soon (see [18] and [19]). Additionally, FED officials have projected at least 3 interest rate hikes in 2022 with some reports indicating as many as 6 or 7. Raises are most likely to occur quarterly with 25-50 basis point increases at each time (see [20]). Historically banks experience slower but more consistent growth during periods of rate increases contrasted to the volatility they experience in low or negative interest rate environments (see [21]).

B. Numerical Studies

1) Procedures

Let S_t be a stock’s closing price quoted at the end of trading day**.** The rates of the log return changes are used as data, as follows,

$$x_{t+1} = \log\frac{S_{t+1}}{S_t} = \log S_{t+1} - \log S_t, \tag{8}$$

where S_t is a stock’s closing price at time t. Note that taking the log difference makes non-stationary data stationary, leading to symmetry pattern. Figure 3 displays graphs of logarithmic closing prices of BAC and PNC and their differences.

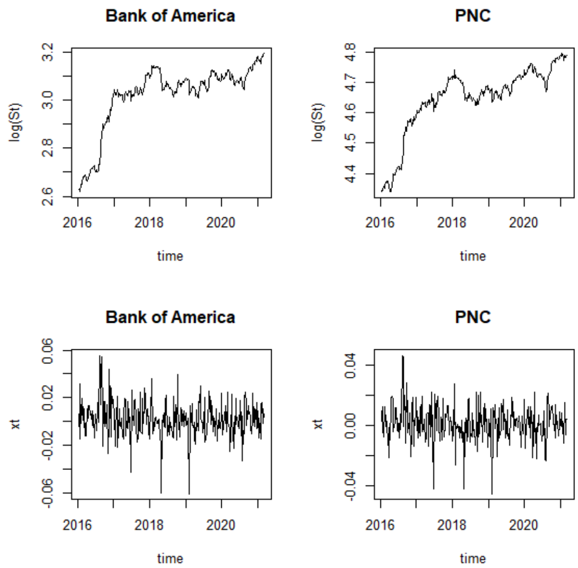

Fig. 3. Log and log-difference of prices of Bank of America and PNC.

It is observed in Fig. 3 that there exists volatility clustering changes over time in both series of BAC and PNC. This indicates the variance of the error term is not constant across time series data and suggests a use of the GARCH model to minimize potential errors in forecasting that may happen in the future. To make deeper insights and more informed decision, we utilize the ACF (auto-correlation function) plots shown in Figure 4. It exhibits such conditional heteroscedasticity of the return series. To analytically assist the graphical results, the Ljung-Box tests (see [22]) are conducted in a quantitative way. It formally assesses autocorrelation of the series, enhancing the objectivity of the graphical results which may result in the emergence of researcher bias. The Ljung-Box tests for both BAC and PNC give p-values < .0001. Both graphical and quantitative methods validate the appropriateness of the GARCH model, confirming time varying conditional volatility of both time series data.

Table IV shows descriptive statistics that describes basic features of the data, such as skewness and kurtosis. It is observed from the table that BAC has a skewness near 0, while the kurtosis is higher than 3. Similar patterns are observed in PNC. This implies they are symmetrically distributed with heavier tails than the normal distribution with mean 0 and variance 1. P-values obtained by the Jarque-Bera test also indicate the non-normality of the data. For data not normally distributed, Student's t distribution would be employed as a legitimate model for the data, since it has a greater chance of producing the values that deviate away from the mean, dealing with a realistic calculation of, for example, value at risk. Thus, the GARCH model with Student's t innovations is appropriate to use for heavy-tailed time series (see [3]).

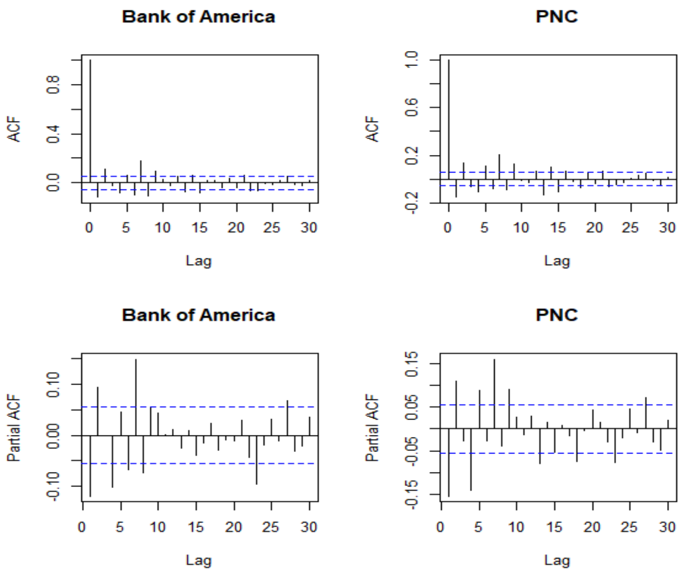

Fig. 4. ACF of BAC and PNC returns and squared returns.

TABLE IV: DESCRIPTIVE STATISTICS OF THE LOG DIFFERERENCE

	Min	Max	Mean	Median	SD	Skewness	Kurtosis	Jarque-Bera
BAC	-0.17	0.16	0	0	0.02	-0.11	12.55	< 0.001
PNC	-0.17	0.16	0	0	0.02	-0.46	12.38	< 0.001

In order to choose one among many specifications of the GARCH and ARMA models, we utilize AIC, and it turns out that for the data, GARCH(1,1) appears to be the most appropriate to account for the time-varying volatility of the series. It is known that the most used heteroscedastic model is GARCH(1,1) when working with financial time series (see [23]). Similarly, the most suitable model dealing with conditional mean seems to be ARMA(1,1), among many other specifications of p and q. Thus, in this work, the time series model that is associated with copula is ARMA(1,1)-GARCH(1,1). Table V summarizes the parameter estimates of the model that are obtained by maximum likelihood, along with the AIC results.

TABLE V: PARAMETER ESTIMATES FOR ARMA(1,1)-GARCH(1,1) WITH STUDENT'S T-INOVATIONS

Variables	μ	ϕ	θ	α	β	AIC
BAC	0.0012	-0.6267	0.6045	0.1157	0.8541	-5.3583
PNC	0.0013	-0.5619	0.5172	0.1523	0.8262	-5.5693

With the copula-based ARMA(1,1)-GARCH(1,1) model associated with the parameter estimates, we generate a simulated return series with a size 2,000 for BAC and PNC for an investigation into their future performance of them. The following steps specify the simulation process:

S1. Convert standardized residuals of ARMA(1,1)-GARCH(1,1) to U(0,1) samples.
S2. Fit the model (i.e., copula-based ARMA(1,1)-GARCH(1,1)) to the samples.
S3. Simulate bivariate data of U_1 and U_2 that represent BAC and PNC, respectively.
S4. Get $F_1^{-1}(U_1)$ and $F_2^{-1}(U_2)$, and then convert back into the original scales.

S5. Forecast n-steps-ahead.

2) Numerical Results

Based on the distribution that is composed of 2,000 simulated data points, Table VI presents return estimates using Student's t copula associated with ARMA(1,1)-GARCH(1,1), where the estimates are in percentages and forecasts of the return rate for the future performance over the next day through the tenth day are labeled D1, D2,..., D10. The first, second and third quartiles of the distribution are indicated by Q1, Q2 and Q3, respectively. Table VII compares corresponding predicted and actual prices in dollars.

TABLE VI: 10-STEP AHEAD FORECAST (%) OF BAC AND PNC

Variabe	Quartile	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
BAC	Q1	-1.006	-1.039	-1.019	-1.032	-1.025	-1.029	-1.027	-1.029	-1.028	-1.029
	Q2	0.144	0.112	0.132	0.193	0.127	0.122	0.125	0.124	0.125	0.124
	Q3	1.300	1.268	1.289	1.276	1.285	1.280	1.283	1.282	1.283	1.282
PNC	Q1	-0.819	-0.882	-0.855	-0.878	-0.873	-0.883	-0.885	-0.891	-0.895	-0.900
	Q2	0.168	0.111	0.143	0.125	0.135	0.129	0.133	0.131	0.132	0.131
	Q3	1.162	1.110	1.148	1.134	1.150	1.149	1.157	1.116	1.165	1.117

TABLE VII: COMPARISON OF PREDICTED AND ACTUAL PRICE ($) OF BAC AND PNC

Variabe	Quartile/ Actual	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
BAC	Q1	41.52	41.50	41.51	41.51	41.51	41.51	41.51	41.51	41.51	41.51
	Q2	42,00	41.99	41.99	42.02	41.99	41.99	41.99	41.99	41.99	41.99
	Q3	42.48	42.47	42.48	42.47	42.48	42.47	42.48	42.48	42.48	42.48
	Actual	43.05	42.95	42.86	42.25	42.87	43.13	44.01	44.03	43.91	44.`3
PNC	Q1	190.65	190.53	190.58	190.54	190.55	190.53	190.53	190.52	190.51	190.50
	Q2	192.55	192.44	192.50	192.47	192.49	192.48	192.48	192.48	192.48	192.48
	Q3	194.46	194.36	194.43	194.41	194.44	194.44	194.45	194.37	194.47	194.38
	Actual	197.86	195.41	195.13	193.35	195.99	195.78	198.15	198.47	194.44	200.39

Aside from a large jump on D1 (9/27/21), the variation of stock prices was mostly captured by the projections. On D1, PNC and BAC closed at $43.05 and $197.86, up 2.93% and 2.66% respectively from the previous day's close. News of forecasted rate hikes and the 10-year bond yield tracking above 1.5% intraday for the first time since June 2021 drew investors to banking stocks. Projections for PNC were semi-accurate over the 10-day period neglecting the initial D1 volatility; pricing grew 1.28% ($195.41 to $200.39) over the period vs. D2 Q2 to D10 Q3 growth forecasting of 1.01% ($192.44 to $194.38). BAC experienced faster growth; D2 Q2 to D10 Q3 ($41.99 to $42.48) growth projections assumed 1.16% appreciation, however the stock jumped 2.50% during the same period, placing it's results at the upper echelon of aggressive growth (D1 Q2 to D10 Q3) projections.

It is worth noting that a stock closing 2.5% higher than the previous day occurs about 2.3% of the time, correctly placing the growth on D1 higher than the projected Q3 growth of 1.3%, signaling the rarity of the event. Closing >2.5% DoD (day over day) occurs ~6.3 trading days per year.

C. Applications and Improviements

Algorithmic (or "robo") trading can consider information derived from this model as additional indicators to make trades. Day traders may also find the model helpful; by reducing the input data to just days or even intraday prices, an actively updating model can provide moving indicators which can trigger trading as the investor deems necessary.

Potential adjustments to the model include decreasing input data or increasing projection duration. Regarding the former, deriving the model from a smaller initial data set will yield projections based on recent movement, and potentially lower the occurrence/size of tail dependencies, albeit at the expense of accuracy. The latter option will also have diminished accuracy as prediction time lengthens, however farther out projections are useful in long strategies.

IV. Concluding Remarks

The primary objective of a stock portfolio is to maximize profit and minimize losses through management's decision making. Through research, we can identify market trends and forecast future movements to aid in trading decisions. Analyzing diverse relationships between banks in the market will provide insight into financial sector trends and patterns, revealing hidden information regarding its impacts. A copula is a useful tool to reveal such dependent associations through its work with tail dependence. Heteroscedasticity is a factor leading to erroneous conclusions in forecasting, and this is often observed when working with financial time series datasets. To incorporate dependence and overcome the heteroscedasticity issue, the copula-based ARMA(1,1)-GARCH(1,1) model was developed and utilized to forecast the movements of Bank of America and PNC. The banks were chosen by principal component analysis, so their movements aim to reflect the behavior of the entire financial sector. Likewise, the results obtained demonstrate that the forecast is consistent with actual values, with the ability for adjustment to suit an investor's needs.

CONFLICT OF INTEREST

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Hotelling H. Analysis of Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology.* 1993; 24: 417-441.
[2] Tsay RS. Analysis of Financial Time Series, Wiley; 2002.
[3] Bollerslev T. Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*. 1986;31: 307-327.
[4] Engle R. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*. 1982; 50: 987-1007.
[5] Embrechts P, McNeil A, Straumann D. Correlation and Dependency in Risk Management: Properties and Pitfall, *In Risk Management: Value at Risk and Beyond*, ed. M.A.H. Dempster, Cambridge University Press, Cambridge. 2002; 176-223.
[6] Box G, Jenkins G. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day; 1970.
[7] Nelsen R. An Introduction to Copulas, New York, Springer-Verlag; 2006.
[8] Sklar A. Functions de Repartition a n Dimensions et leurs Merges, Publication of the lnstitute of Statistics, University of Paris. 1959;8: 229-231. French.
[9] Demarta S, McNeil A. The t Copula and related Copulas, *International Statistical Review*. 2005; 73: 111-129.
[10] Aktas H, Oncu S. The Stock Market Reaction to Extreme Events: The Evidence from Turkey. *International Research Journal of Finance and Economics. 2006*; 6: 78-85.
[11] Joe H. Families of m-variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Paremeters. In L. Ruschendorf and B. Schweizer and M.D. Taylor (Ed.), *Distributions with Fixed Marginals and Related Topics*; 1996.
[12] Lee S, Lee E, Vlk J. Forecasting Stock Returns using a Copula-GARCH model, Lambert; 2017.
[13] Bakar HO, Sulong Z. The Role of Financial Sector on Economic Growth: Theoretical and Empirical Literature Reviews Analysis, *Journal of Global Economics 6*. 2018.
[14] Sarreal R. History of Online Banking: How Internet Banking Went Mainstream, *Go Banking Rates*; 2019.
[15] Phaneuf A. State of mobile banking in 2022: Top apps, features, statistics and mobile trends, *Business Insider*; 2022.
[16] Kolmar C. 39 MUST-KNOW WORK FROM HOME STATISTICS [2022], *Zippia*. 2021.
[17] Caminiti S. 4 gig economy trends that are radically transforming the US job market, *CNBC*; 2019.
[18] Derby M. Fed's George: Current Fed Stance 'Out Of Sync' With Economic Outlook, *Wall Street Journal*; 2022.
[19] Derby, M. Fed's Harker Leans Against 50 Basis-Point Rate Hike, *Wall Street Journal*; 2022.
[20] Timiraos N. Fed Officials Project Three Interest Rate Rises in 2022 and Accelerate Stimulus Wind-Down, *Wall Street Journal*; 2021.
[21] Lofton O, Ulate M. How Do Low and Negative Interest Rates Affect Banks?, *Federal Reserve Bank of San Francisco Economic Letter*; 2021.
[22] Ljung GM, Box GE. On a Measure of a Lack of Fit in Time Series Models, *Biometrika*. 1978; 65: 297-303.
[23] Bera A, Higgins M. ARCH models: Properties, Estimation and Testing, *Journal of Economic Surveys*. 1993; 7: 305-366.