Estimate Conditional Correlation with Factor Structure

JUNXIONG GAO*

ABSTRACT

This paper adds a conditional factor structure to correlation dynamics, which presents the covariance matrix by factor loadings and hence shrink the dimension of estimation. Furthermore, the factor structure allows a closed-form solution of the inverse and determinant of the covariance matrix, which simplifies the likelihood function of the dynamic conditional correlation model. Taking the realized correlation from 5-minutes data as the benchmark, the model implies a more precise correlation. In application, the model generates out-of-sample portfolios with higher information ratio and a more precise value at risk measurements.

*Junxiong Gao is part of the Rady School of Management, University of California San Diego.

I. Introduction

Precise estimation of the covariance matrix among assets in both cross-section and times-series dimensions is important for finance practitioners. It is widely employed in portfolio allocation and risk management. However, this exercise in both aspects suffers from the high-dimensional curse: As the opportunity set extends, the estimation of the covariance matrix loses accuracy.

In terms of the cross-sectional or static estimation, an N by N covariance matrix estimated from a T by N sample has a concentration ratio of N/T. That is, the estimator is noisy when N is relatively larger than T. This dilemma is solved through the shrinkage method in Ledoit and Wolf (2004) and Ledoit et al. (2012) because it corrects the bias of the covariance matrix eigenvalues. Alternatively, one can insert a factor structure among the assets to estimate the N by K (number of factors) factor loading instead, which improves the concentration ratio to K/T, then, imply the covariance matrix Q as in Chan et al. (1999) and Ledoit and Wolf (2003):

$$Q = B\Sigma B' + Q_\varepsilon$$

where B is the N by K factor loading, Σ is the covariance among factors, and Q_ε is a diagonal covariance matrix among the residuals.

In this paper, I insert the same cross-section structure to the Dynamic Conditional Correlation (DCC) model to solve high-dimensional dilemma that appears in the time-series estimation, as documented by Engle and Sheppard (2001):

A multivariate generalized autoregressive conditional heteroskedasticity-type (GARCH) model (from Engle (1982) to Engle (2002)) is specified as follows:

$$Q_t = S(1-\alpha-\beta) + \beta Q_{t-1} + \alpha\varepsilon_{t-1}\varepsilon'_{t-1}$$

The model above is known as DCC because it decomposes the covariance to volatility and correlation using the "DeGarch" returns to $\varepsilon_t = D_t^{-1} r_t$, where D_t is a diagonal matrix of each asset's volatility. The covariance likelihood is decomposed to the single-assets GARCH and conditional correlation likelihood. That is,

$$L_c(\theta, \phi) = -\frac{1}{2} \sum_{t=1}^{T} (-\varepsilon_t' \varepsilon_t + \log |R_t| + \varepsilon_t' R_t^{-1} \varepsilon_t)$$

where $R_t = diag\{Q_t\}^{-1} Q_t diag\{Q_t\}^{-1}$ is the normalized correlation matrix. Therefore, dilemma is: As the number of assets increases, the optimization over R_t^{-1} and $|R_t|$ becomes imprecise. Given the "DeGarch" intuition, I treat the terms "Covariance" and "Correlation" as equivalent throughout the context.

As its primary contribution, I solve R_t^{-1} and $|R_t|$ in closed form by inserting a conditional factor structure into the DCC model, which simplifies the likelihood estimation. I use a toy model to sketch the spirit: Consider the most simple case under the capital asset priced model (CAPM) by Sharpe (1964) and Lintner (1975). The set of asset returns is as follows:

$$r_i = \beta_i r_m + \varepsilon_i$$

The covariance between two different assets r_i and r_j is:

$$COV(r_i, r_j) = \beta_i \beta_j \sigma_m^2 = \frac{COV(r_i, r_m) COV(r_j, r_m)}{\sigma_m^2}$$

The correlation between the i,j asset is:

$$\rho_{i,j} = \rho_{i,m} \rho_{j,m}$$

In matrix format, the correlation matrix R is written as:

$$R = XX' + diag\{1 - \rho_{i,m}^2\}$$

where $X = \{\rho_{i,m}\}_{i=1...n}$ is an N by 1 combination of each $\rho_{i,m}$ and the correlation between the ith asset and the market portfolio.

With factor structure, the N by N correlation matrix equals an outer product of an N by 1 vector plus a diagonal matrix. A matrix written in this format has a closed-form solution of inverse and determinant, as demonstrated by Woodbury (1950) and Sherman and Morrison (1950). I extend this simple model to a conditional and arbitrary K factor setting in the main context.

I therefore combine techniques in two strands of literature mentioned above (cross-section and time-series). The points below highlight the advantages of the model proposed herein:

a. Like the static cross-section model, the proposed model shrinks the $\{R\}_{N\times N}$ matrix to a factor loading matrix $\{X\}_{N\times K}$.

b. Though estimating a linear forecasting model for X_t, the correlation matrix takes a quadratic form and, therefore, has more dynamites than the linear model.

c. Like DCC and Dynamical Equivalent Conditional Correlation (DECO), by applying a two-step quasi-maximum likelihood estimate approach, the model establishes a well-known inference property documented by White (1996).

To test the model performance, I compare my model against other variations of the DCC models both in and out-of-sample: The in-sample performance is tested by comparing the model-implied realized measured correlation (from 5-minutes data of stocks in the $S\&P500$ index) under different loss functions. Following the conventional literature, I construct the minimum variance portfolio, that is, tangency portfolio pseudo out-of-sample. Because the model can generate conditional correlation together with the betas, there exist risk management implications, such as computing the marginal value at risk for portfolios.

Like Engle et al. (2017), I can combine my model with the nonlinear shrinkage method in Ledoit et al. (2012) to obtain a better estimation of the unconditional covariance—intercept matrix S in DCC—while retaining the benefit from the conditional factor structure. Following the same intuition, I insert a certain structure to simplify the likelihood estimation.

Here, Engle and Kelly (2012) assumes a pairwise equal correlation to solve R_t^{-1} and $|R_t|$ in closed form, namely, DECO. Engle et al. (1990) and Engle (2009) also present a factor structure, but it is not conditional and uses constant betas. This technique differs from the current research.

The remaining paper is organized as follows: Section II presents the model and estimation method. Section III describes the data and list of benchmark models used, and then reports the in- and out-of-sample performance. Section IV concludes the study.

II. Model and Estimation Method

This section introduces the model specification and estimation steps. Within the same quasi-maximum likelihood estimate approach, the estimators' asymptotic distribution is included in Appendix II.

I. Model Specification

I begin by taking a conditional factor setting. Assume there are K factors for all N assets. The pricing kernel m_t is modeled as:

$$m_t = a_{t-1} - b'_{t-1} f_t, \tag{1}$$

where a_{t-1} is a scalar, b_{t-1} is a K by 1 rotation vector, and f_t is the K by 1 vector of factors. Each stock r_{it} and factor f_{kt} satisfies the pricing equation:

$$E_{t-1}[r_{it} m_t] = 1. \tag{2}$$

Proposition 2.1 Excess Return

The expected excess return under conditions in (1) and (2) is:

$$E_{t-1}[r_{it} - r_t^f] = \beta'_{it} E_{t-1}[f_t]. \tag{3}$$

$$\beta_{it} = COV_{t-1}(r_{it}, f_t) VAR_{t-1}(f_t)^{-1}. \tag{4}$$

$E_{t-1}[f_t]$ is a K by 1 vector of each factor risk premium and β_{it} is the K by 1 dynamic factor loading vector. In this equation, $COV_{t-1}(r_{it}, f_t)$ is the K by 1 vector of the conditional covariance between $r_[it]$ and each factor; $VAR_{t-1}(f_t)$ is the conditional variance matrix among K factors.

Proposition 2.1 can be written in matrix format as follows:

$$E_{t-1}[r_t - r_t^f] = \beta_t E_{t-1}[f_t]. \tag{5}$$

$$\beta_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1}, \tag{6}$$

where $COV_{t-1}(r_t, f_t)$ is the N by K combination of all the assets' factor covariance. Based on proposition 2.1, I model the asset returns as follows:

$$r_t - r_t^f = \beta_t f_t + \varepsilon_t \tag{7}$$

Here, the following assumptions hold:

Factor Model Assumption

2a. f is exogenous: $E_{t-1}(f_t \epsilon_{it}) = 0$ for any i; and

2b. Idiosyncratic returns are independent cross assets: $E_{t-1}(\epsilon_{it}\epsilon_{jt}) = 0, i \neq j$.

The conditional factor setting in this section generalizes the static CAPM model to the extent that it allows for time-varying betas. Further, it is correctly specified if all the risk factors are added in the model. We then derive the conditional distribution of $r_{t|t-1}$:

Proposition 2.2 Conditional Distribution

$$r_{t|t-1} \sim N(\mu_t, H_t) \tag{8}$$

$$\mu_t = \beta_t E_{t-1}[f_t] + r_t^f \tag{9}$$

$$H_t = COV_{t-1}(r_t, f_t)VAR_{t-1}(f_t)^{-1}COV_{t-1}(r_t, f_t)' + VAR_{t-1}(\varepsilon_t) \tag{10}$$

$$\beta_t = COV_{t-1}(r_t, f_t)VAR_{t-1}(f_t)^{-1} \tag{11}$$

Proposition 2.2 indicates that my method is robust for combining with the enormous class of forecasting literature modeling $E_{t-1}[f_t]$. As a guide to practitioners, this method can have both the conditional mean and the variance forecast as its inputs for constructing the optimal portfolio. Because I emphasize conditional correlation, I treat the conditional mean as constants and directly work with demeaned returns.

In terms of the second moment, proposition 2.2 shows that the N by N dynamic covariance matrix can be shrunk to the N by K covariance matrix between asset returns and factors. Under assumption 2a and 2b, by decomposing the conditional covariance $H_t = D_t R_t D_t$, where D_t is the diagonal matrix of each asset's conditional volatility ("DeGarch"), I derive R_t such that I obtain the closed-form solution of its inverse and determinant. This allows me to construct the likelihood function, which, in turn, can be feasibly optimized.

Proposition 2.3 Conditional Correlation

Let $K_t = Corr_{t-1}(f_t)$ be the conditional correlation matrix among K factors. Let X_t be the N by K matrix written in partition form $X_t = [\rho'_{1,t} \dots \rho'_{n,t}]$, where $\rho_{i,t}$ is a K by 1 vector representing the conditional correlation between r_i and each factor.

$$R_t = X_t K_t^{-1} X_t' + diag\{1 - \rho'_{i,t} K_t^{-1} \rho_{i,t}\} \tag{12}$$

$$R_t^{-1} = diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}$$
$$- diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t(K_t + X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t)^{-1}X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\} \quad (13)$$

$$|R_t| = \det(K_t + X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t)\det(K_t^{-1})\prod_i(1-\rho'_{i,t}K_t^{-1}\rho_{i,t}) \quad (14)$$

Proposition 2.3 works as a dynamic multi-factor extension of the toy model in section I. If only a market risk exists, then the correlation among the factors is 1. This makes the construction of the likelihood function in this study trivially simple. The spirit of these tricks is to shrink an infeasible optimization problem, including the dynamic adjustment of an N by N matrix inverse, to a K by K case. Thus, we set a conditional factor model and show important tools for the likelihood simplification. The next section introduces the estimation method based on these tools.

II. Estimation Method

Based on propositions 2.2 and 2.3, the estimation procedure is designed as follows:

Step 1: DCC Normalization

Set a joint DCC among the demeaned assets and factors:

$$\begin{aligned} r^1_{t|t-1} &= [r_{t|t-1} \quad f_{t|t-1}] \\ r^1_{t|t-1} &\sim N(0, D_t R^1_t D_t) \end{aligned} \quad (15)$$

The joint DCC is specified as follows:

GARCH for each r_i^1:

$$D_t^2 = diag\{w_i\} + diag\{\kappa_i\} \circ r_{t-1}r_{t-1}' + diag\{\lambda_i\} \circ D_{t-1}^2$$

DeGARCH:

$$\varepsilon_t^1 = D_t^{-1} r_t^1 \tag{16}$$

$$Q_t = S(1-\alpha-\beta) + \alpha\varepsilon_{t-1}^1 {\varepsilon_{t-1}^1}' + \beta Q_{t-1}$$

$$R_t^1 = diag\{Q_t\}^{-1} Q_t diag\{Q_t\}^{-1}$$

Step 2: Reconstruct the Correlation

Partition the joint correlation R_t^1:

$$R_t^1 = \left\{ \begin{matrix} R_t & X_t \\ X_t' & K_t \end{matrix} \right\} \tag{17}$$

where, R_t and K_t are correlation matrixes among N assets and K factors, respectively, and $X_t = \{\rho_{i,k,t}\}_{N\times K} = [\rho_{1,t}' \ldots \rho_{n,t}']$ contains the correlations between each asset and factor. This partition allows me to construct the likelihood function with the factor model-implied correlation matrix R_t^L in lieu of R_t from the traditional DCC.

$$R_t^L = X_t K_t^{-1} X_t' + diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\} \tag{18}$$

Step 3: Maximize the Likelihood

Construct the likelihood function based on R_t^L. The likelihood is generated by parameters S, α, β and, thus, optimized by the selection of these parameters.

$$R_t^{L^{-1}} = diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}$$
$$- diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t(K_t + X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t)^{-1}X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\} \tag{19}$$

$$\det(R_t^L) = \det(K_t + X'_t diag\{\frac{1}{1-\rho'_{i,t}K_t^{-1}\rho_{i,t}}\}X_t)\det(K_t^{-1})\prod_i(1-\rho'_{i,t}K_t^{-1}\rho_{i,t}). \tag{20}$$

The correlation likelihood is given as follows:

$$L_c(\alpha,\beta,S) = -\frac{1}{2}\sum_{t=1}^{T}(-\varepsilon'_t\varepsilon_t + \log\det(R_t^L) + \varepsilon'_t R_t^{L^{-1}}\varepsilon_t). \tag{21}$$

Equations (18), (19), and (20) build a joint dynamic for R_t^L as a function of X_t. As mentioned in the introduction, the model presents the conditional correlation in a quadratic form as it estimates a simple linear forecast model. The estimator's asymptotic distribution can be derived similar to DCC, as shown in Appendix II.

III. Alternative Estimation Method: MacGyver

A joint likelihood estimate is more efficient, though it requires high computation power. To address this drawback, Engle (2009) proposed a "MacGyver" method to separately estimate $\frac{N(N-1)}{2}$ pairwise DCC, and then take the $\alpha and \beta$ from the average among all the pairwise $\alpha_i and \beta_i$. Though this method requires lower computation power, it not as efficient.

Similarly, one can always implement the method herein in an easier manner: In lieu of building the full matrix dynamic, separately estimate N by K pairs of the bivariate DCC model to obtain each $\rho_{i,k,t}$, and then build R_t by (18):

$$\forall i, \forall k, V_{i,k} = [r_i; F_k], \text{build } n \times k \text{ pairs of DCC for } V_{i,k} \tag{22}$$

$$q_{i,k,t} = s_{i,k} + \alpha_{i,k}\varepsilon_{i,k,t-1}\varepsilon^F_{k,t-1} + \beta_i q_{i,k,t-1} \tag{23}$$

$$\rho_{i,k,t} = \frac{q_{i,k,t}}{\sqrt{\sigma^2_{i,t} h^F_{k,t}}} \tag{24}$$

$$X_t = \{\rho_{i,k,t}\}_{N\times K} = [\rho'_{1,t} \dots \rho'_{n,t}] \tag{25}$$

$$R_t = X_t K_t^{-1} X'_t + diag\{1 - \rho'_{i,t} K_t^{-1} \rho_{i,t}\} \tag{26}$$

$$\forall k, \alpha_k = f(\alpha_{i,k}), \beta_k = f(\beta_{i,k}) \tag{27}$$

Like the MacGyver DCC, the estimates from this method have an unknown inference property. Rather than $\frac{N(N-1)}{2}$ pairs for the DCC, my model only requires N times K pairs of estimation.

III. Empirical Results

In this section, two datasets of the constituents returns from $S\&P500$ index are used to test the model performance for measuring correlation. The first is the 5-minutes price data from 2010 to 2017. After cleaning up the data, 377 stocks returns are available in the full sample. I use this high-frequency dataset to compute the realized correlation as the benchmark, and then compare it with the model-implied under certain loss functions. The second is the daily returns of $S\&P500$ index constituents from 2000 to 2018[1] , used to form out-of-sample portfolios.

I attempt to include some representative models as candidate estimators:

- DCC: original Dynamic Conditional Correlation Model
- DECO: Dynamic Equicorrelation Model, as in Engle and Kelly, 2012

[1] The original data is from 2000 to 2015, with 395 assets that with full history during this period. We then extend our sample by combining new data from 2015 to 2018, with 449 assets.

- DCC–NLS: DCC with nonlinear shrinkage, as in Engle et al. (2017)
- Model: The model proposed in this study; conditional factor structured.
- Model–NLS: The model proposed in this study with nonlinear shrinkage.

Without loss of generality, I use the $S\&P500$ index value-weighted return as the only factor. The trade-off between single and multiple factors is clear: More factors ensure the validity of exogeneity and independence among idiosyncratic returns. The single factor model is more parsimonious for estimation and, thus, less noisy.

I. Loss against Realized Correlation

Based on the high-frequency covariance matrix theory (see Barndorff-Nielsen and Shephard (2004)), I compute the realized correlation $R_t^{realized}$ and the average correlation of a matrix as follows:

$$R_t^{realized} = \frac{1}{n_t} \sum_{i=0}^{n_t} r_{t+i\Delta} r'_{t+i\Delta} \tag{28}$$

$$\bar{\rho}_t = \sum_{i \neq j} \sum_{j} R_t(i,j) \frac{2}{n(n-1)} \tag{29}$$

where n_t is the number of small-time intervals in each unit time.

I further define several loss functions to compare the accuracy of the model-implied correlation:

- Squared Error for the Average Correlation $\bar{\rho}_t$:

$$\overline{SE}_t = (\overline{\rho_t} - \overline{\rho_t^{realized}})^2. \tag{30}$$

- Mean Squared Error:

$$MSE_t = \frac{2}{N(N-1)} \sum_{i \neq j} \sum_{j} (R_t(i,j) - R_t^{realized}(i,j))^2. \tag{31}$$

- Mean Absolute Error:

$$MAE_t = \frac{2}{N(N-1)} \sum_{i \neq j} \sum_{j} \mid (R_t(i,j) - R_t^{realized}(i,j) \mid \tag{32}$$

- Quasi-Normal Likelihood Error:

$$QL_t = \det R_t - \log(\det R_t) - \det R_t^{realized} \tag{33}$$

R_t in these functions denotes the model-implied correlations and $R_t^{realized}$ is measured by 5-minute returns.

The loss of the models against the benchmark over time is reported in Table 1.

[Insert Table 1 here]

Panel A shows the average loss function values of different models and Panel B reports the t-statistics of each model's loss against the DCC model. In general, a nonlinear shrinkage intercept does improve the accuracy of the models. The DCC–NLS and DECO models win over the original DCC under certain loss functions. My model and its combination with the nonlinear shrinkage has the smallest loss and wins over the original DCC model compared with the other models.

In addition, my model naturally yields a dynamic factor loading β_t. Like the partition of conditional correlation matrix, I derive β_t by decomposing the conditional covariance matrix $H_t^1 = VAR_{t-1}([r_t \quad f_t])$:

$$H_t^1 = \begin{Bmatrix} H_t & COV_{t-1}(r_t, f_t) \\ COV_{t-1}(r_t, f_t)' & VAR_{t-1}(f_t) \end{Bmatrix} \tag{34}$$

$$\beta_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} \tag{35}$$

Equation (34) eases the practical applications in the following sections.

II. Model Implication

II.1. Risk Management: Marginal Value at Risk

I now demonstrate how my method applies to risk management for allocating the value at risk of a portfolio. Unlike an allocation problem that determines the optimal weight, a risk management task takes the portfolio weights as given and analyzes the effect of altering asset positions marginally on the portfolio risk.

For any portfolio with N assets:

$$r_{pt} = \sum_{i=1}^{N} w_{it} r_{it} \tag{36}$$

The joint relationship between assets and the portfolio returns is depicted by a conditional factor model:

$$r_{it} = \beta_{it} r_{pt} + \varepsilon_{it} \tag{37}$$

$$\beta_{it} = COV_{t-1}(r_{it}, r_{pt}) VAR_{t-1}(r_{pt})^{-1} \tag{38}$$

This leads to:

$$\sum_{i=1}^{N} w_{it}\beta_{it} = 1, \forall t \tag{39}$$

Equation (39) allows us to link each asset's contribution to the total risk through β_{it}. This argument falls in line with the definition of the marginal value at risk: $\Delta VaR_{it} = \frac{\partial VaR_{pt}}{\partial w_{it} W_t}$ is a partial derivative that measures how a dollar position adjustment shifts the total value at risk. A marginal value at risk (with confidence level θ) is then defined as follows:

$$\Delta VaR_{it} = \frac{\partial VaR_{pt}}{\partial w_{it} W_t} = Z_\theta \frac{\partial \sigma_{pt}}{\partial w_{it}} = Z_\theta \beta_{it} \sigma_{pt} \tag{40}$$

where Z_θ is the critical value for normal distribution, W_t is the total wealth, σ_{pt} is the conditional total volatility, and β_{it} is the dynamic portfolio risk loading. Equation (39) and (40) indicate that:

$$\frac{VaR_{pt}}{W_t} = Z_\theta \sigma_{pt} = \sum_{i=1}^{N} w_{it}\beta_{it} Z_\theta \sigma_{pt} = \sum_{i}^{N} w_{it} \Delta VaR_{it} \tag{41}$$

Equation (41) gives a clear decomposition of each asset's contribution to the portfolio's unite money value at risk. I take the $S\&P500$ 5-minutes data to construct a portfolio weighted by their market capital, and then estimate the joint correlation among all the stocks and the portfolio. The performance of the models is measured comparing it with the value at risk from the realized covariance matrix:

$$\begin{aligned} \Delta VaR_{it}^{realized} &= Z_\theta \beta_{it}^{realized} \sigma_{pt}^{realized} \\ \beta_{it}^{realized} &= COV_{t-1}(r_{it}, r_{pt})^{realized} \{VAR_{t-1}(r_{pt})^{realized}\}^{-1} \end{aligned} \tag{42}$$

I compute the marginal value at risk for each asset with my model, the combination of my model with nonlinear shrinkage and the factor DCC model with the MacGyver method (which also yields dynamic betas by pairwise estimation). Then, the capital weighted square errors, absolute errors, are computed as follows:

$$
\begin{aligned}
MSE_t^{VaR} &= (\sum_i^N w_{it} \Delta VaR_{it} - \sum_i^N w_{it} \Delta VaR_{it}^{realized})^2 \\
MAE_t^{VaR} &= \mid \sum_i^N w_{it} \Delta VaR_{it} - \sum_i^N w_{it} \Delta VaR_{it}^{realized} \mid
\end{aligned}
\tag{43}
$$

The average error over time is reported in Table 1, Panel C. Unlike the loss measured in Panels A and B, the combination of nonlinear shrinkage improves the accuracy substantially. This is because I measure the average accuracy of the N asset dynamics in lieu of all the $\frac{N(N-1)}{2}$ elements in the correlation matrix. The benefit of a precise intercept estimation is often important for reliable practical work.

II.2. Portfolio Allocation

Based on the mean-variance utility developed in Markowitz (1952), the conditional mean and variance is specified as in proposition 2.2. I also assume a constant risk-free rate for convenience.

$$r_{t|t-1} \sim N(\mu_t, H_t) \tag{44}$$

$$\mu_t = \beta_t E_{t-1}[f_t] + r^f \tag{45}$$

$$H_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} COV_{t-1}(r_t, f_t)' + VAR_{t-1}(\varepsilon_t) \tag{46}$$

$$\beta_t = COV_{t-1}(r_t, f_t) VAR_{t-1}(f_t)^{-1} \tag{47}$$

By construction, my model can yield the first moment forecast by estimate $E_{t-1}[f_t]$ in lieu of the entire cross-section of expected returns $E_{t-1}[r_t]$. One can simply combine a risk premium forecast model with my method to derive better portfolios. Because I focus on forecasting the conditional variance, I simply take the historical average risk premium times dynamic factor loading as the expected return, $\hat{\mu}_t = \beta_t \overline{f_t}$. For all my models, because there is no dynamic factor loading, the expected return is given as historical means $\hat{\mu}_t = \overline{r_t}$.

Then, three target portfolios are defined as:

- Global Minimum Variance Portfolio:

$$\widehat{w_{GMV,t}} = \arg\min w' H_t w \quad s.t \quad w'\vec{1} = 1 \tag{48}$$

$$\widehat{w_{GMV,t}} = \frac{H_t^{-1}\vec{1}}{\vec{1}' H_t^{-1}\vec{1}} \tag{49}$$

- Minimum Variance Portfolio (with a required return q, set q = 0.1) (MV):

$$\widehat{w_{MV,t}} = \arg\min w' H_t w \quad s.t \quad w'\vec{1} = 1 \quad w'\mu \geqslant q. \tag{50}$$

$$\widehat{w_{MV,t}} = \frac{C - qB}{AC - B^2} H_t^{-1}\vec{1} + \frac{qA - B}{AC - B^2} H_t^{-1}\mu_t \tag{51}$$

  where $A = \vec{1}' H_t^{-1}\vec{1}, B = \vec{1}' H_t^{-1}\mu_t, C = \mu_t' H_t^{-1}\mu_t$.

- Tangency Portfolio:

$$\widehat{w_{TGC,t}} = \arg\max w'(\mu_t - r_f) - \frac{1}{2}\gamma w' H_t w \quad s.t \quad w'\vec{1} = 1 \tag{52}$$

$$\widehat{w_{TGC,t}} = \frac{H_t^{-1}(\mu_t - r_f)}{\vec{1}' H_t^{-1}(\mu_t - r_f)} \tag{53}$$

I then construct monthly updating portfolios for all the candidates listed before. The procedure is set as follows:

Step 1 : Take the end of t month's last day forecast μ_{t+1}, H_{t+1} to construct portfolios $\widehat{w_{GMV,t+1}}, \widehat{w_{MV,t+1}}, \widehat{w_{TGC,t+1}}$.

Step 2 : Pseudo out-of-sample test the performance and use the realized daily returns in the $t+1$ month (averagely 22 days per month) to compute all the portfolios returns, standard deviations, and Sharpe ratios.

Step 3: Include the $t+1$ month returns in the sample based on a 1,000-day rolling window. This yields the forecast μ_{t+2}, H_{t+2}, and then return to step 1.

Following this recursive pseudo out-of-sample method, I report the average performance over time of the three portfolios generated by the models in Table 2:

[Insert Table 2 here]

From Table 2, I find that, on average, my model and the model with nonlinear shrinkage produce lower volatility portfolios and higher information ratio/Sharpe ratio. Consistent with the literature, both DCC–NLS and DECO produce better results than DCC does.

IV. Conclusions

The study applies the conditional factor model to multivariate covariance models. It combines both cross-section and time-series features of the covariance models to solve the high-dimensional curse.

To improve the methodology on mean-variance allocation, my method incorporates the literature on conditional mean forecasting. Under my framework, the expected return forecasting task is reduced to risk premium forecasting. For professions with strong intuitions on what creates a risk premium, the method is sound for practical use.

As shown in the empirical work, a single factor specification, conditional CAPM setting generates a robust in- and out-of-sample covariance fit. However, in terms of asset pricing research, it is important to further study what factors drive correlation among assets with the proposed method. The model should be extended to a vaster class of risk factor literature as a test of this factor's power to create cross-sectional co-movement. My method thus opens a new channel to test conditional asset pricing models within the second moment.

This study emphasizes the method's performance with a single factor. What factors to be added in the model and to what extend can adding factors increase the prediction power of the model remain open questions for future research.

REFERENCES

Ole E Barndorff-Nielsen and Neil Shephard. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925, 2004.

Louis KC Chan, Jason Karceski, and Josef Lakonishok. On portfolio optimization: Forecasting covariances and choosing the risk model. *The review of Financial studies*, 12(5): 937–974, 1999.

Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

Robert Engle and Bryan Kelly. Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30(2):212–228, 2012.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

Robert F Engle. High dimension dynamic correlations. *The methodology and practice of econometrics: A festschrift in honour of David F. Hendry*, pages 122–148, 2009.

Robert F Engle and Kevin Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, National Bureau of Economic Research, 2001.

Robert F Engle, Victor K Ng, and Michael Rothschild. Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, 45 (1-2):213–237, 1990.

Robert F Engle, Olivier Ledoit, and Michael Wolf. Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, pages 1–13, 2017.

Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5): 603–621, 2003.

Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

Olivier Ledoit, Michael Wolf, et al. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.

John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic Optimization Models in Finance*, pages 131–155. Elsevier, 1975.

Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

Jack Sherman and Winifred J Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1): 124–127, 1950.

Halbert White. *Estimation, inference and specification analysis.* Number 22. Cambridge university press, 1996.

Max A Woodbury. Inverting modified matrices. *Memorandum report*, 42(106):336, 1950.

Appendices

I. Proof of the Conditional Factor Model

Proof of Proposition 2.1

Equation (2) leads to:

$$r_t^f = 1/E_{t-1}[m_t] \tag{54}$$

$$1 = E_{t-1}[r_{it}m_t] = COV_{t-1}(r_{it}, m_t) + E_{t-1}[r_{it}]E_{t-1}[m_t] \tag{55}$$

$$\Rightarrow E_{t-1}[r_{it} - r_t^f] = -COV_{t-1}(r_{it}, m_t)/E_{t-1}[m_t] \tag{56}$$

Combine with the pricing kernel in (1):

$$E_{t-1}[r_{it} - r_t^f] = b'_{t-1}COV_{t-1}(r_{it}, f_t)/E_{t-1}[m_t] \tag{57}$$

By assuming each factor satisfies the pricing equation:

$$E_{t-1}[f_t - r_t^f] = b'_{t-1}VAR_{t-1}(f_t)/E_{t-1}[m_t] \tag{58}$$

As an extension of proof 2.1, let $H_t^K = VAR_{t-1}(f_t)$ be the conditional covariance matrix of factors and $COV_{t-1}(r_{it}, f_t)$ be the k by 1 vector containing all the conditional covariance between r_{it} and each f_{kt}. Then, based on assumptions 2a and 2b:

Proof of Proposition 3.2

$$\forall i \neq j, r_{i,t} = f_t'\beta_{it} + \varepsilon_{i,t}, r_{j,t} = f_t'\beta_{jt} + \varepsilon_{j,t} \tag{59}$$

$$COV_{t-1}(r_{it}, r_{jt}) = \beta_i' VAR_{t-1}(f_t) \ = \beta_i' H_t^K \beta_j \tag{60}$$

$$COV_{t-1}(r_{it}, f_t)' = \beta_i' VAR_{t-1}(f_t) = \beta_i' H_t^K \tag{61}$$

$$COV_{t-1}(r_{jt}, f_t)' = \beta_j' VAR_{t-1}(f_t)) = \beta_j' H_t^K \tag{62}$$

$$\Rightarrow COV_{t-1}(r_{it}, f_t)'(H_t^K)^{-1} COV_{t-1}(r_{jt}, f_t) = COV_{t-1}(r_{it}, r_{jt}) \tag{63}$$

Decompose H_t^K to $D_t^K K_t D_t^K$, and then divide (63) by $\sigma_{i,t}\sigma_{j,t}$:

$$\frac{1}{\sigma_{i,t}} COV_{t-1}(r_{it}, f_t)'(D_t^K)^{-1} K_t^{-1} (D_t^K)^{-1} COV_{t-1}(r_{jt}, f_t) \frac{1}{\sigma_{j,t}} = \rho_{i,j,t} \tag{64}$$

Note that, by definition, $(D_t^K)^{-1} COV_{t-1}(r_{it}, f_t) \frac{1}{\sigma_{i,t}} = \rho_{i,t}$ is the K by 1 vector containing all the correlations between r_{it} and each f_{kt}. Thus,

$$\rho_{i,j,t} = \rho_{i,t}' K_t^{-1} \rho_{j,t}. \tag{65}$$

Equation (65) can be extended to matrix form:

Define X_t as the $N \times K$ matrix containing every $Corr_{t-1}(r_{it}, f_{kt})$. Then, R_t can be presented by $X_t K_t^{-1} X_t'$ plus an adjustment term for diagonal elements:

$$R_t = X_t K_t^{-1} X_t' + diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\}$$

Derive the closed-form solution of R_t^{-1} and $|R_t|$ by applying the following theorem:

Theorem Woodbury Identity

$$(A + UCV')^{-1} = A^{-1} - A^{-1}U(C^{-1} + V'A^{-1}U)^{-1}V'A^{-1} \tag{66}$$

$$\det(A + UCV') = \det(C^{-1} + V'A^{-1}U)\det(C)\det(A) \tag{67}$$

Proof of Proposition 3.3

$$R_t = X_t K_t^{-1} X_t' + diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\} \tag{68}$$

Set $A = diag\{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}\}, U = V = X_t, C = K_t^{-1}$:

$$A^{-1} = diag\{\tfrac{1}{1-\rho_{i,t}' K_t^{-1} \rho_{i,t}}\} \tag{69}$$

$$\det(A) = \textstyle\prod_i (1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}) \tag{70}$$

$$(C^{-1} + V' A^{-1} U) = K_t + X_t' diag\{\tfrac{1}{1-\rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t \tag{71}$$

$$\tag{72}$$

which leads to:

$$R_t^{-1} = diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\}$$
$$- diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t (K_t + X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t)^{-1} X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} \tag{73}$$

$$\det(R_t) = \det(K_t + X_t' diag\{\frac{1}{1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}}\} X_t) \det(K_t^{-1}) \prod_i (1 - \rho_{i,t}' K_t^{-1} \rho_{i,t}) \tag{74}$$

II. Statistical Inference

The likelihood is decomposed into a GARCH part and a correlation part:

$$
\begin{aligned}
L_1(\theta) &= -\frac{1}{2}\sum_t\sum_{i=1}^{n}(\log(2\pi)+\log(h_{i,t})+\frac{r_{i,t}^2}{h_{i,t}}) - \frac{1}{2}\sum_t(\log(2\pi)+\log(h_{m,t})+\frac{r_{m,t}^2}{h_{m,t}}) \\
L_2(\theta,\phi) = L_c(\theta,\phi) &= -\frac{1}{2}\sum_{t=1}^{T}(-\varepsilon_t'\varepsilon_t+\log|R_t|+\varepsilon_t' R_t^{-1}\varepsilon_t) \\
\log f_{1,t} &= -\frac{1}{2}\sum_{i=1}^{n}(\log(2\pi)+\log(h_{i,t})+\frac{r_{i,t}^2}{h_{i,t}}) - \frac{1}{2}(\log(2\pi)+\log(h_{m,t})+\frac{r_{m,t}^2}{h_{m,t}}) \\
\log f_{2,t} &= \frac{1}{2}(\log|R_t|+\varepsilon_t' R_t^{-1}\varepsilon_t)
\end{aligned} \tag{75}
$$

White (1996) Theorem 6.1

Under assumptions C.1–C.6,

$$\sqrt{T}(\hat{\gamma}-\gamma^*) \sim^A N(0, A^{*-1}BA^{*-1})$$

,

where $A^* = \left\{ \begin{matrix} E[\nabla_{\theta\theta}L_1(r_t,\theta^*)] & 0 \\ .E[\nabla_{\theta\phi}L_2(r_t,\theta^*,\phi^*)] & E[\nabla_{\theta\theta}L_2(r_t,\theta^*,\phi^*)] \end{matrix} \right\}.$

and

$B^* = var(\,T^{-\frac{1}{2}}\sum_t(s_{1,t}^{*}{}', s_{2,t}^{*}{}'))\,,$

where $s_{1,t}^* = E[\nabla_\theta L_1(r_t,\theta^*)]$ and $s_{2,t}^* = E[\nabla_\phi L_1(r_t,\theta^*,\phi^*)]$.

Write assumptions in White (1996) and Engle and Kelly (2012):

Assumptions C.1

(a) For all $\theta\in\Theta, \phi\in\Phi$, $E[\log f_{1,t}(r_t,\theta)]$ and $E[\log f_{2,t}(r_t,\theta,\phi)]$ exist and are finite, $\forall t$;

(b) $E[\log f_{1,t}(r_t,\theta)]$ and $E[\log f_{2,t}(r_t,\theta,\phi)]$ are continuous on Θ and Φ, $\forall t$; and

(c) $\{\log f_{1,t}(r_t,\theta)\}$ and $\{\log f_{2,t}(r_t,\theta,\phi)\}$ each obey the strong uniform law of large number.

Assumptions C.2

$f_{1,t}$ and $f_{2,t}$ are each twice continuously differentiable on Θ and Φ, $\forall t$.

Assumptions C.3

For all $\theta \in \Theta, \phi \in \Phi$, $E[\nabla_\theta L_1(r_t, \theta)] < \infty$ and $E[\nabla_\phi L_2(r_t, \theta, \phi)] < \infty$, $\forall t$.

Assumptions C.4

(a) For all $\theta \in \Theta, \phi \in \Phi$, $E[\nabla_{\theta\theta} L_1(r_t, \theta)] < \infty$ and $E[\nabla_{\phi\phi} L_2(r_t, \theta, \phi)] < \infty$;

(b) $E[\nabla_{\theta\theta} L_1(r_t, \theta)]$ and $E[\nabla_{\phi\phi} L_2(r_t, \theta, \phi)]$ are continuous on Θ and Φ;

(c) $\{\nabla'_\theta s_{1,t}(r_t) = \nabla_{\theta\theta} \log f_1(r_t, \theta)\}$ and $\{\nabla'_\phi s_{2,t}(r_t) = \nabla_{\phi\phi} \log f_2(r_t, \theta, \phi)\}$; and

(d) A^* is negative definite.

Assumptions C.5

$E[L_1(r_t, \theta)]$ is uniquely maximized by θ^* interior to Θ, and $E[L_2(r_t, \theta, \phi)]$ is uniquely maximized by ϕ^* interior to Φ.

Assumptions C.6

$\{(T^{-\frac{1}{2}} {s^*_{1,t}}', T^{-\frac{1}{2}} {s^*_{2,t}}')\} \equiv \{(T^{-\frac{1}{2}} \nabla'_\theta L_1(r_t, \theta^*), T^{-\frac{1}{2}} \nabla'_\phi L_2(r_t, \theta^*, \phi^*))\}$ obeys the central limit theorem.

Like DECO, the studied model ensures identification as long as each pairwise DCC is properly identified.

Table 1

Average Loss of each model and T-stats against DCC

Panel A :Loss Under Different Measures					
	DCC	DCC_NLS	DECO	Model	Model_NLS
$\overline{SE}_t$	0.010	0.010	0.010	0.007	0.007
MSE_t	0.078	0.078	0.109	0.065	0.065
MAE_t	0.222	0.222	0.222	0.199	0.199
QL_t	-74.236	-68.768	-71.717	-17.769	-18.377
Panel B: T-stats of Loss(DCC)-Loss(models)					
		DCC_NLS	DECO	Model	Model_NLS
$\overline{SE}_t$		-9.082	1.689	41.048	39.172
MSE_t		-16.399	-55.980	40.390	42.775
MAE_t		1.496	0.916	50.499	53.437
QL_t		22.245	16.434	22.303	22.297
Panel C: Loss of Portfolio VaR					
			DCC(MacGyver)	Model	Model_NLS
MSE_t^{VaR}			0.2888	0.2403	0.8884
MAE_t^{VaR}			0.4958	0.4665	0.1244

Table 2

2001-2018 Daily data. Monthly means, standard deviations IR(information ratio) and sharp ratio are annualized and presented in percent.

Panel A :Global Minimum Variance Portfolio (GMV)					
	Deco	DCC	DCC_NLS	Model	Model_NLS
Mean %	20.059	22.712	21.614	15.590	15.387
Std %	15.213	13.765	12.469	11.792	11.705
IR	0.324	0.595	0.680	0.992	0.981

Panel B :Minimum Variance Portfolio(MV)					
	Deco	DCC	DCC_NLS	Model	Model_NLS
Mean %	24.108	21.519	19.986	14.819	12.169
Std %	14.718	13.432	12.090	12.703	11.340
IR	0.684	0.867	0.958	1.428	1.508

Panel C :Tagency Portfolio(TGC)					
	Deco	DCC	DCC_NLS	Model	Model_NLS
Mean %	-20.163	-36.860	-20.945	11.007	11.156
Std %	36.315	112.290	28.423	15.430	15.555
IR	0.602	0.012	0.875	1.447	1.449
Sharpe Ratio	0.550	-0.015	0.806	1.342	1.345