A New Model in Forecasting Dynamic Correlations

Ray Y Chou [1] Nathan Liu [2] Chun-Chou Wu [3]

[1] Institute of Economics, Academia Sinica
[2] Department of Management Science, National Chiao-Tung University
[3] Department of International Trade, Chung Yuan Christian University

Abstract

This paper proposes a range-based Dynamic Conditional Correlation (DCC) model, which is an extension of Engle's return-based DCC model. In the DCC model, the conditional correlation coefficients are estimated by a dynamic model for the product of the pair-wise return series with each normalized by their conditional standard deviations. The conditional standard deviation is calculated by using a univariate GARCH for the return series. We use the Conditional Autoregressive Range (CARR) model of Chou (2005a), as an alternative to the univariate GARCH in the DCC first-step estimation. The substantial gain in efficiency in the volatility estimation can induce an efficiency gain in forecasting the time-varying correlation coefficient. Empirical results using in-sample and out-of-sample data are supportive.

Keywords: DCC, CARR, range, dynamic correlation

1. Introduction

It is of primary importance in the practice of portfolio management, asset allocation, and risk management to have an accurate estimate of the covariance matrices for asset prices. When valuing derivatives, forecasts of volatilities and correlations over the whole life of the derivative are usually required. The univariate ARCH/GARCH family of models provides effective tools to estimate the volatilities of individual asset prices. For a survey of this vast literature, see Bollerslev, Chou, and Kroner (1992), and Engle (2004). It is, however, still an active research issue in estimating the covariance or correlation matrices of multiple, especially large sets of asset prices. Early attempts include the VECH model of Bollerslev, Engle, and Wooldridge (1988), the BEKK model of Engle and Kroner (1995), and the constant correlation model of Bollerslev (1990), among others.

In a series of papers, Engle and Sheppard (2001), and Engle (2002a), and Engle, Cappiello, and Sheppard (2003) provide a solution to this problem by using a model entitled the Dynamic Conditional Correlation Multivariate GARCH (henceforth DCC). The conditional covariance estimation problem is simplified by estimating univariate GARCH models for each asset's variance process. Carrying on by using the transformed standardized residuals from the first stage, and estimating a time-varying conditional correlation estimator in the second stage, the DCC model is not linear, but can be estimated simply with the two-stage methods based on the maximum likelihood method.

In this paper, we consider a refinement of the DCC model by utilizing the high/low range data of asset prices. In estimating the volatility of asset prices, there is a growing awareness of the fact that the range data of asset prices can provide sharper estimates and forecasts than the return data based on close-to-close prices. Studies of supporting evidence include Parkinson (1980) and more recently Alizadeh, Brandt, and Diebold (2001), Chou (2005a, 2005b) and Chou, Wu, and Liu (2004). Chou (2005a) proposed the Conditional Autoregressive Range (henceforth CARR) model where can capture the dynamical volatility process and obtained some insightful evidence in real data. In light of the success of the range-based univariate volatility models, it is natural to inquire whether this estimation efficiency can be extended to a multivariate framework, in this case of the DCC model.

2. Correlation Estimation and the DCC Model

The DCC model can be shown as follows.

$$H_t = D_t R_t D_t ,$$

$$R_t = diag\{Q_t\}^{-1/2} Q_t diag\{Q_t\}^{-1/2} .$$

Here, D_t is the $k \times k$ diagonal matrix of time-varying standard deviations from univariate GARCH models with $\sqrt{h_{i,t}}$ on the ith diagonal, where $\sqrt{h_{i,t}}$ is the square root of the estimated variance. and

$$Q_t = S \circ (\iota\iota' - A - B) + A \circ Z_{t-1} Z_{t-1}' + B \circ Q_{t-1}$$

A and B are parameters and o denotes the Hadamard matrix product operator. The symbol ι is a vector of ones and S is the unconditional covariance of the standardized residuals. Finally, $Z_t = D_t^{-1} \times r_t$ are the standardized but correlated residuals. The variable r_t represents the returns of assets. The returns can be either mean zero or the residuals from a filtered time series, i.e.

$r_t \left| I_{t-1} \sim N(0, H_t)\right.$.

The conditional variances of the components of Z_t are, in other words, equal to 1, but the conditional correlation matrix is given by the variable of R_t. It is important to recognize that although the dynamic of the D_t matrix has usually been structured as a standard univariate GARCH model, it can extend to many other types. Later on, we shall propose to use the Conditional Autoregressive Range (CARR) model of Chou (2005a) as an alternative.

As for parameters A and B, it is shown that if A, B, and $(\iota\iota' - A - B)$ are positive semi-definite, then Q_t will be positive semi-definite. If any one of the matrices is positive definite, then Q_t will also be so. For the ijth element of R_t, the conditional correlation matrix is given by $q_{ij,t} / \sqrt{q_{ii,t} q_{jj,t}}$. As to the conditional covariance, it can then be expressed using the product of conditional correlation between these two variables and their individual conditional standard deviations. Engle and Sheppard (2001) show results that simplify finding the necessary conditions for R_t to be positive definite and hence a correlation matrix with a real, symmetric positive semi-definite matrix, with ones on its diagonal line. The log-likelihood of this estimator can be written as:

$$L = -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi) + \log\left|H_t\right| + r_t' H_t^{-1} r_t\right)$$
$$= -\frac{1}{2}\sum_{t=1}^{T}\left(k\log(2\pi) + \log\left|D_t R_t D_t\right| + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t\right)$$

Here, Z_t ~ $N(0,R_t)$ are the univariate GARCH standardized residuals. Based on Engle (2002a)'s argument, we can perform the estimation in two steps. This estimator will no longer be efficient, but still consistent. Let the parameters in D_t be denoted θ and the additional parameters in R_t will be denoted by ϕ. The log-likelihood function can be split into two respective parts:

$L(\theta,\phi) = L_V(\theta) + L_C(\theta,\phi)$.

The former term expresses the volatility part:

$$L_V(\theta) = -\frac{1}{2}\sum_t\left(n\log(2\pi) + \log\left|D_t\right|^2 + r_t' D_t^{-2} r_t\right).$$

The latter term is the correlation component:

$$L_C(\theta,\phi) = -\frac{1}{2}\sum_t\left(\log\left|R_t\right| + Z_t' R_t^{-1} Z_t - Z_t' Z_t\right).$$

The volatility part of the likelihood is the sum of the individual GARCH likelihood if D_t is determined by a GARCH specification.

$$L_V(\theta) = -\frac{1}{2}\sum_t\sum_{i=1}^{k}\left(\log(2\pi) + \log(h_{i,t}) + \frac{r_{i,t}^2}{h_{i,t}}\right).$$

This can be jointly maximized by separately maximizing each term. If D_t is determined by a CARR specification, then the likelihood function of the volatility term is

$$L_V(\theta) = -\frac{1}{2}\sum_t\sum_{i=1}^{k}\left(\log(2\pi) + 2\log(\lambda_{i,t}^*) + \frac{r_{i,t}^2}{\lambda_{i,t}^{*2}}\right),$$

where $\lambda_{i,t}^*$ is the conditional standard deviation as computed from a scaled expected range from the CARR model.

The second part of the likelihood will be used to estimate the correlation parameters. As the squared residuals are not dependent on these parameters, they will not enter the first-order conditions and can be ignored. The two-step approach to maximizing the likelihood is to find

$\hat{\theta} = \arg\max\{L_V(\theta)\}$,

and then take this value as given in the second stage,

$\max_{\phi}\{L_C(\hat{\theta},\phi)\}$.

It is shown in Engle and Sheppard (2001) that under reasonable regularity conditions, consistency of the first step will ensure consistency of the second step. The maximum of the second step will be a function of the first-step parameter estimates, and so if the first step is consistent, then the second step will be too as long as the function is continuous in a neighborhood of the true parameters.

For the CARR volatility structure (return-based conditional volatility model):

$\Re_{k,t} = u_{k,t} \quad u_{k,t} \mid I_{t-1} \sim \exp(1;\cdot)$, k=1,2

$\lambda_{k,t} = \omega_k + \alpha_k \Re_{k,t-1} + \beta_k \lambda_{k,t-1}$,

$z_{k,t}^c = r_{k,t} / \lambda_{k,t}^* \quad$ where $\lambda_{k,t}^* = adj_k \times \lambda_{k,t} \quad adj_k = \dfrac{\overline{\sigma}}{\overline{\hat{\lambda}}_k}$,

where $\Re_{k,t}$ is the high/low range in logarithm, of the k^{th} asset during time interval t, $\overline{\sigma}$ and $\overline{\hat{\lambda}}_k$ are respectively the unconditional variance of the return series and the sample mean of the estimated conditional range of the series k. This is a special case of the multiplicative error model of Engle (2002b). The specification of the exponential distribution of the disturbance term provides a consistent although inefficient estimator for the parameters. For specific discussions also see Chou (2005a).

In the following analysis, we use two alternative versions of DCC. The first one is the standard DCC with mean reversion (henceforth MR_DCC), discussed in Engle (2002). The second one is the integrated DCC (henceforth I_DCC). Both of these two models are simplified versions of the general expression.

The MR_DCC is constructed by the following equation.

$Q_t = S \circ (\iota\iota' - A - B) + A \circ Z_{t-1} Z_{t-1}' + B \circ Q_{t-1}$,

For I_DCC, the dynamic structure simplifies to:

$$Q_t = A \circ Z_{t-1} Z'_{t-1} + (1 - A) \circ Q_{t-1}$$

Like the specific property of volatilities, the correlation also unobservable. We use daily data to construct the proxies for the weekly-realized correlation observations. The purpose of such doing is to extract these so-called “measured correlation”, denoted MCORR, as one kind of benchmark in determining the relative performance of the return-based DCC model and the range-based DCC model for the time being. The $MCORR_t$ is defined as

$$MCORR_t = \frac{1}{\tau} \sum_{i=1}^{\tau} \rho_t^i ,$$

where τ denotes the trading days during the week t and ρ_t^i is the correlation coefficient at the i^{th} trading day of the week *t.* This series is obtained from using the daily returns data and fitting them with a MR_DCC model.

On the other side, we perform the tailor-made regression framework proposed by Mincer and Zarnowitz (1969) for the in-sample comparison.

In constructing the comparison of in-sample data in our subsequent empirical analysis, several related models are included, such as MA(100), EWMA with $\lambda = 0.94$. Here one just uses the estimated correlation regression on the realized correlation, and the correlation is similar in the same manner. For a simple regression, the R-squared can be used as a rough proxy for the model’s performance.

For completeness, we also perform out-of-sample forecast comparisons. It is very straightforward to derive the formulation in computing the out-of-sample conditional correlation for a bivariate MR_DCC specification. Given T as the sample size, the T+1[st] observation is obtained by:

$$\begin{bmatrix} q_{1,T+1} & q_{12,T+1} \\ q_{12,T+1} & q_{2,T+1} \end{bmatrix} = \begin{bmatrix} 1 & \bar{q}_{12} \\ \bar{q}_{12} & 1 \end{bmatrix} \circ \begin{bmatrix} 1-a_1-b_1 & 1-a_3-b_3 \\ 1-a_3-b_3 & 1-a_2-b_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix} \circ \begin{bmatrix} z_{1,T}^2 & z_{1,T} z_{2,T} \\ z_{1,T} z_{2,T} & z_{2,T}^2 \end{bmatrix} + \begin{bmatrix} b_1 & b_3 \\ b_3 & b_2 \end{bmatrix} \circ \begin{bmatrix} q_{1,T} & q_{12,T} \\ q_{12,T} & q_{2,T} \end{bmatrix}$$

where $\rho_{T+1} = q_{12,T+1} / \sqrt{q_{1,T+1} q_{2,T+1}}$.

For the period of t+h, with $h \geq 2$, the correlation is:

$$\begin{bmatrix} q_{1,T+h} & q_{12,T+h} \\ q_{12,T+h} & q_{2,T+h} \end{bmatrix} = \begin{bmatrix} 1 & \bar{q}_{12} \\ \bar{q}_{12} & 1 \end{bmatrix} \circ \begin{bmatrix} 1-a_1-b_1 & 1-a_3-b_3 \\ 1-a_3-b_3 & 1-a_2-b_2 \end{bmatrix} + \begin{bmatrix} a_1+b_1 & a_3+b_3 \\ a_3+b_3 & a_2+b_2 \end{bmatrix} \circ \begin{bmatrix} q_{1,T+h-1} & q_{12,T+h-1} \\ q_{12,T+h-1} & q_{2,T+h-1} \end{bmatrix}.$$

Since the values for the out-of-sample correlation forecast derived from the I_DCC model are constants, we abridge the redundant explanation. In addition to the range-based and return-based DCC models, the MA(100) is introduced for an out-of-sample predictive comparison. Empirically speaking, we still take the value of R-squared as an indication for the comparison of preciseness.

3. Evaluation of Conditional Correlation Forecasts

The data employed in this study comprises 835 weekly observations on the S&P500 Composite (henceforth S&P500), the Nasdaq stock market index, and the yield for 10-year treasury bond (Tbond) spanning the period January 4, 1988 to January 2, 2004. We retrieve the ranges and returns data for the entire period from Yahoo’s database.

Table 1: In-sample Correlations Forecasting

$$MCORR_t = \gamma_0 + \gamma_1 \hat{\rho}_t^{return} + \varepsilon_{1,t}$$
$$MCORR_t = \gamma_0 + \gamma_2 \hat{\rho}_t^{range} + \varepsilon_{2,t}$$
$$MCORR_t = \gamma_0 + \gamma_1 \hat{\rho}_t^{return} + \gamma_2 \hat{\rho}_t^{range} + \varepsilon_{3,t}$$

MCORR		$\hat{\gamma}_0$		$\hat{\gamma}_1$		$\hat{\gamma}_2$		R^2
Panel A: S&P500 and Nasdaq								
MR_DCC	Return	0.251	(10.425)	0.707	(24.691)			0.409
	Range	0.347	(16.937)			0.597	(24.539)	0.389
	BOTH	0.257	(10.524)	0.445	(7.435)	0.258	(5.277)	0.425
I_DCC	Return	0.250	(7.806)	0.705	(18.221)			0.316
	Range	0.280	(10.676)			0.672	(21.554)	0.350
	BOTH	0.171	(5.689)	0.363	(6.545)	0.440	(10.122)	0.392
MA100		0.259	(9.421)	0.699	(21.22)			0.330
Exp. Smoothing		0.375	(20.745)	0.555	(25.732)			0.489
Panel B: S&P500 and Tbond								
MR_DCC	Return	0.030	(4.524)	0.848	(57.328)			0.777
	Range	0.023	(3.789)			0.860	(70.081)	0.800
	BOTH	0.021	(3.621)	-0.399	(-3.513)	1.256	(11.066)	0.803
I_DCC	Return	0.023	(3.539)	0.965	(59.105)			0.794
	Range	0.007	(1.164)			0.945	(72.435)	0.811
	BOTH	0.008	(1.458)	0.128	(1.161)	0.822	(7.730)	0.811
MA100		-0.017	(-1.704)	0.890	(36.625)			0.587
Exp. Smoothing		0.025	(3.947)	0.846	(67.370)			0.792
Panel C: Nasdaq and Tbond								
MR_DCC	Return	0.026	(5.551)	0.851	(57.019)			0.777
	Range	0.029	(6.574)			0.818	(67.500)	0.816
	BOTH	0.031	(7.113)	-0.437	(-5.573)	1.222	(16.445)	0.821
I_DCC	Return	0.032	(6.888)	0.955	(57.549)			0.792
	Range	0.018	(3.987)			0.890	(60.008)	0.802
	BOTH	0.022	(4.970)	0.351	(3.944)	0.572	(6.777)	0.807
MA100		-0.017	(-2.707)	0.871	(42.423)			0.658
Exp. Smoothing		0.030	(6.283)	0.758	(62.599)			0.776

Table 1 describes the in-sample forecasting performance for MCORR. With the exception of the correlation between the S&P500 and Nasdaq indices, all of the estimates of R-squared for the range DCC model are statistically preferred to the return DCC model.

The exponential smoothing model seems to perform well in the sample for correlation forecasting, but poorly for the out-of-sample forecasting. Technically, based on the exponential smoothing model, it obtains a constant value when we predict the out-of-sample correlation between any two variables. Finally, no matter what market data is used, the MA100 is the worst model in our in-sample correlation forecast performance comparison. It is apparently that the range proxy fitting to the in-sample forecasts of correlation is again better on the setting of the DCC model.

We adopt the procedure of a rolling sample to estimate the out-of-sample forecasts using MR_DCC and I_DCC specifications for both the return-based and range-based models. For each individual model, we compute the out-of-sample forecasts for the horizons of 1, 2, 3, and 4 weeks. In all cases, we re-estimated the estimates 100 times. We then use a simple regression to compare the explanatory power of these various forecasts on the realized covariances or correlations.

Table 2: Out-of-Sample Correlations Forecasting

$$MCORR_t = \gamma_0 + \gamma_1 \hat{\rho}_t + \varepsilon_t$$

R-squared	MR DCC		
Forecast horizon	return-based	range-based	MA100
Panel A:	S&P500 and Nasdaq		
1	0.310	0.548	0.044
2	0.279	0.474	0.076
3	0.248	0.407	0.109
4	0.199	0.325	0.153
Panel B:	S&P500 and Tbond		
1	0.491	0.631	0.165
2	0.441	0.547	0.134
3	0.306	0.463	0.098
4	0.229	0.388	0.065
Panel C:	Nasdaq and Tbond		
1	0.593	0.786	0.443
2	0.673	0.756	0.397
3	0.705	0.706	0.348
4	0.719	0.654	0.299

Table 2 reports the value of R^2 from a linear regression of MCORR on each of these out-of-sample forecast series. The overall result is consistent whatever out-of-sample horizon is chosen. We achieve a confirmation that the DCC-range model is more powerful then DCC-return model when the forecasting intervals of correlation are during one-month period. With the exception of the relationship between the Nasdaq index and the Tbond yield for the four weeks out of sample prediction, all of the estimates of R-squared for any other market data witness the inference.

4. Conclsions

In this paper, a new estimator of the time-varying correlation matrices is proposed utilizing the range data by combining the CARR model proposed by Chou (2005a) and the framework of Engle (2002a)'s DCC model. The advantage of this range-based DCC model outperforming the standard return-based DCC model hinges on the relative efficiency of the range over the return data in estimating volatilities. Using weekly returns of S&P500, Nasdaq and 10-year treasury bond rates, we find consistent results that the range-based DCC model outperforms the return-based models in estimating and forecasting correlation matrices, both in-sample and out-of-sample. It also can be applied to larger systems in a manner similar to the application of the return-based DCC model structures that is demonstrated in Engle and Sheppard (2001). Future research will be useful in adopting more diagnostic statistics or tests based on value at risk calculations. Applications to the estimation of optimal portfolio weighting matrices and the calculation of the dynamic hedge ratio in the futures market will also be fruitful.

5. References

[1] Alizadeh, Sassan, Michael Brandt and Francis Diebold. (2002) "Range-based estimation of stochastic volatility models or exchange rate dynamics are more interesting than you think," *Journal of Finance*, 57: 1047-1091.

[2] Bollerslev, Tim, Ray Y. Chou, and Kenneth Kroner. (1992) "ARCH modeling in finance: a review of the theory and empirical evidence." *Journal of Econometrics*, 52: 5-59.

[3] Chou, Ray Y. (2005a) "Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model," *Journal of Money Credit and Banking*, forthcoming

[4] Chou, Ray Y. (2005b) "Modeling the asymmetry of stock movements using price ranges," *Advances in Econometrics*, forthcoming.

[5] Chou, Ray Y., Chun-Chou Wu and Nathan Liu, (2004) "A comparison and empirical study in forecasting abilities of dynamic volatility models," *Journal of Financial Studies.*Vol.12 No.1: 1-25.

[6] Engle, Robert (2002a) "Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, Vol. 20 No. 3:339-350.

[7] Engle, Robert (2002b) "New Frontiers in ARCH Models," *Journal of Applied Econometrics*, Vol. 17 No. 5

[8] Engle, Robert (2004) "Risk and Volatility: Econometric Models and Financial Practice," Nobel Lecture, *American Economic Review* Vol. 94, No. 3,

[9] Engle, Robert and Kevin Sheppard, (2001) "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," *NBER Working Paper* 8554.

[10] Mincer, Jacob and Victor Zarnowitz. (1969) "The evaluation of Economic Forecasts," in J. Mincer, ed, *Economic Forecasts and Expectations (NBER)* 3-46.

[11] Parkinson, Michael (1980) "The extreme value method for estimating the variance of the rate of return," *Journal of Business*, 53: 61-65.