STOCK SELECTION AND RISK MEASUREMENT IN THE HONG KONG STOCK EXCHANGE-A MULTIVARIATE COPULA-GARCH APPROACH

[1,*]BIREN HU, [2]JIANXU LIU, [3]JIRAKOMSIRISRISAKULCHAI, [4]PATHAIRATPASTPIPATKUL, [5]SONGSAK SRIBOONCHITTA

[1,2,3,4,5]Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
Email: [1]birenhu@hotmail.com

Abstract - In this paper, we conduct the stock selection with the traditional and modified Sharpe ratio(SR) methods to obtain three groups of target portfolios on the Hong Kong Stock Exchange(HKSE). We measure the Value at Risk (VaR) of the log returns of the portfolios by using seven Copula-GARCH models with the dependent structure. The empirical results show three findings: (1) Copula function explicitly captured the dependence structure of the selected stocks;(2) The t-Copula-GARCH with unstructured covariance matrix is the best adequate model to all three portfolios; (3) Conventional Sharpe ratio method has the best performance in stock selection, and the copula-GARCH model captures VaR quite well.

Keywords - Stock Selection, Risk Measurement, Copula-GARCH, HKSE

I. INTRODUCTION

Under non-linear strictly increasing property, the linear correlation was not invariant, and financial data were more highly correlated during volatile markets and during market downturns. (Hotta et al., 2008) There were some disadvantages by using common multivariate distribution to measure risk. It was precisely because copula approach owns the properties of capturing tail dependence and nonlinear correlation. It implied copula functions were conductive to estimating VaR because the VaR was concerned with the tail of the distribution. To describe volatility of financial data, we combined GARCH and multivariate copulas instead of a classical joint normal one. We used the conditional copula-GARCH model forbeing more reasonable and adequate. In financial market, the stocks selection to form the portfolios and the risk measurement of the selected portfolios were extremely important assignments for financial institutes, investment managers, and financial analysts. This paper used the Sharpe ratio methods to conduct stocks selection on HKSE to create portfolios. Then the portfolios were imposed to carry out the multivariate VaR estimation by using Copula-GARCH models to measure the risk and capture the dependence structure of the portfolios. Sharpe ratio was usually an instrument of stocks selection to build the portfolio and the stocks in the portfolio as target equity pool was normally assumed were identical independent distributed. Recently years, there were plenty of research studies on using Copula-GARCH to analyze the VaR with the dependent structure of the portfolio of financial asset. Nevertheless, the target portfolios of financial assets were given with lacking of the stocks selection process. In order to overcome this drawback, we combined both of the two aspects of the stock selection and the VaR with a dependence structure by using Copula-GARCH model which covered through whole process of financial activities. This paper aimed at selecting stocks from HKSE as the target portfolios, modeling the dependence structure with the daily returns of the chosen portfolios, forecasting the VaR by rolling windows with the adequate dependence structured models. William Sharpe(1994) introduced that the Sharpe ratio indicated the expected differential return in terms of one-unit risk associated with the differential return. Once the portfolios have been selected, we used seven Copula-Garch methods to fit the models. Copula function described the joint relations that link the marginal distribution into a joint one.Kole et al. (2007) considered important the elliptical and asymmetric copulas for risk management. Huang et al. (2009) described the method to estimate the VaR of the portfolio via using copula-GARCH based model, and indicate that this approach was robust on the estimation of VaR. In the volatility and dependence analysis (Liu et al, 2015) found that the multidimensional elliptical and asymmetric copula with component expected shortfall model had better performance for systemic risk measurement of the international financial system. Guerard Jr. et al (2015) studied the level of testing for stock selection and portfolio construction, and found that using multi-factor risk-controlled portfolio returns rejects the data mining corrections test null hypothesis. After fitting the Copula-GARCH model, it turns out the t-Copula-GARCH with unstructured covariance matrix was the best fitted model for all three target portfolios. Finally, we forecast the VaR by using the rolling windows by the adequate model of the Copula-GARCH model. The VaR is expressed as a maximum value of the expected loss of a given asset value (or a portfolio) over a given period of time with a given probability. Therefore, under 3 different levels of significant of 1%, 5% and 10%, we estimated the VaR based on the t-Copula-GARCH and find out the amount of the violated observation in the different significant level.

The section follows introduction was the methodology of this paper. Then we provided the empirical results on section 3. At the last, we made the conclusion as section 4 of the paper to summarize the final ideas.

II. METHODOLOGY

This paper systematically analyzed the topic combining stock selection by using Sharpe ratio methods, risk measurement by using a Multivariate Copula-GARCH model on the HKSE. By comparing seven copula models,the result indicated that the t-Copula-GARCH with unstructured covariance matrix was the best fitted model for 3 portfolios which selected from HKSE market during 2010-2016; Subsequently,forecasting the VaR were all acceptable for all three groups of portfolio and the portfolio 1, the traditional Sharpe ratio wide standard deviation, had got the best performance among the three Sharpe ratio methods which based on the copula models (see Table 2).

2.1. Sharpe Ratio of traditional stock selection

The Sharpe ratio is simply the return per unit of risk (represented by variability). The unit of risk can be measured with the standard deviation, the value at risk, and the expected shortfall of the returns. Based on the three sigma, we could figure out the three Sharpe ratios.

1) Sharpe ratio based on the standard deviation:

$$\mathrm{SharpeRatio}_{\mathrm{STD}} = \frac{\mu - R_f}{\sigma},$$

Where μ is portfolio expected return, R_f is risk free rate, and σ is standard deviation of portfolio.

2) Sharpe ratio based on the modified Value at Risk:

$$\mathrm{SharpeRatio}_{\mathrm{MVaR}} = -\frac{\mu - R_f}{R_f - \mathrm{MVaR}_p},$$

Where μ is portfolio expected return, R_f is risk free rate, MVaR_p is the modified Value at Risk of the portfolio.

3) Sharpe ratio based on the modified expected shortfall:

$$\mathrm{SharpeRatio}_{\mathrm{MES}} = -\frac{\mu - R_f}{R_f - \mathrm{MES}_p},$$

Where μ is portfolio expected return, R_f is risk free rate, MES_p is the modified expected shortfall of the portfolio. The original Sharpe Ratio from William Sharpe explained Sharpe ratio with higher value represent a better performance in terms of the combination of "risk" and return. Therefore, the traditional Sharpe Ratio is a method of the risk-adjusted measurement of return, and it uses standard deviation to represent financial risk. Noticeably, modified VaR and Expected Shortfall (ES) are recently introduced for the asymmetric distribution of assets return which based on the Cornish-Fisher expansion.

2.2.The GARCH (1,1) model

The return of financial data is non-linear, asymmetric and time-varying. The application for the ARCH model is to capture the volatility pattern of the financial time series which allows for a long lags of the conditional variance while The GARCH model allows for a longer memory and much more flexible lag structures. The basic GARCH (1,1) model can be expressed as

$$r_t = \mu + \sigma_t z_t,$$
$$\varepsilon_t = \sigma_t z_t,$$
$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

Where r_t is the return at time t,z_t is the standardized residual, σ_t^2 is the conditional variance of r_t. The GARCH model is accepted as one of the best models to describe volatility. However, it is important to note that there are three disadvantages to this model. Firstly, in both positive and negative cases, it can only model symmetric returns as the conditional variance depending on the square of the error term, which does not take the sign into account. Secondly, the variation in volatility is not explained. Finally, the GARCH model is still too short regarding the tail model. It is simply too short even with the innovation of standardized student t distribution. After modeling marginal distribution, we standardize the residuals obtained in the GARCH process and square them then we obtain the standardized residuals matrix of each preferable conditional distribution. After that we estimate the copula functions, which will be expressed in the next section where the estimation methods of copula functions are made through the maximum likelihood estimation method. AIC are then again used to select the best fitted copula family for each assets pair.

2.3. Multivariate copula

After obtaining the marginal distribution of the residuals from the best fitted GARCH model, we use the multivariate copula to capture the dependence structure of the portfolios.

1) Multivariate Gaussian copula

The Gaussian copula is a distribution over the unit cube $[0,1]^d$. It is constructed from a multivariate normal distribution over R^d by using a probability integral transform. With R being a d×d correlation matrix and u is the marginal distribution, the multivariate copula cdf is:

$$C(u;R) = \phi_d\left(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d);R\right),$$

$u \in [0,1]^d$and the copula density is

$$C(u;R) = \frac{\phi_d(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_1);R)}{\prod_{j=1}^{d}\phi(\Phi^{-1}(u_j))},\ u \in [0,1]^d,$$

Where ϕ_d is the joint multivariate normal distribution of u. The correlation structures of exchangeable (ex) and unstructured (un), and the corresponding correlation matrix is as follows:

$$R_{ex} = \begin{pmatrix} 1 & \rho_1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_1 \\ \rho_1 & \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_1 & \rho_1 & 1 \end{pmatrix},$$

$$R_{un} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_4 & \rho_5 \\ \rho_2 & \rho_4 & 1 & \rho_6 \\ \rho_3 & \rho_5 & \rho_6 & 1 \end{pmatrix}.$$

In this study, the Gauss copula with the 2 type of structures is used to handle the time series behavior of longitudinal data.

2) Multivariate t-copula

The t-copula function is based on the multivariate t distribution. Since the degree of tail dependency could be changed by altering the degrees of freedom, therefore, it's application is still growing.

$$C(u;R,v) = \frac{t_{d,v}(T_{1,v}^{-1}(u_1),\ldots,T_{1,v}^{-1}(u_d);R)}{\prod_{j=1}^{d} t_{1,v}[T_{1,v}^{-1}(u_j)]}, u \in [0,1]^d,$$

where the student-t copula with 2 structures (ex and un) of the correlation matrix R is used, same as in the Gauss copula as above. v is the degrees of freedom, u is the marginal distributions and d is the dimensions of u. Then $t_{d,v}$ is the standardized multivariate Student's distribution and $T_{1,v}^{-1}$ is the inverse of the univariate Student's distribution.

3) Multivariate Clayton copula

Clayton (1978), for the first time, proposed the multivariate Clayton family as below

$$C(u;\gamma) = \left(1 + \sum_{i=1}^{d}(u_i^{-\gamma} - 1)\right)^{-1/\gamma},$$

where $C(u;\gamma)$ is a d dimensional Clayton copula function which is the so-called generator function. u is the marginal distribution function and γ is a parameter within some parameter space.

4) Multivariate Frank copula

The multivariate Frank copula first appeared in Frank (1979). It is discussed at length in Genest(1987) and can be expressed as follows

$$C(u;\delta) = \psi\left(\sum_{j=1}^{d}\psi^{-1}(u_j)\right) = -\delta^{-1}\log\left\{1 - \frac{\prod_j(1-e^{-\delta u_j})}{(1-e^{-\delta})^{d-1}}\right\},\ u \in [0,1]^d,$$

where ψ is the generator and the ψ^{-1} is the inverse function, u is the marginal distribution, $\delta \neq 1$ is some parameter, d is the dimension of the variables.

5) Multivariate Gumbel copula

Gumbel (1960) introduced the multivariate Gumbel family. Since it has been, It has been known as the Gumbel-Hougaard family because of being discussed with Hougaard(1986).

$$C(u;\delta) = \exp\left\{-\left[\sum_{j=1}^{d}[-\log u_j]^{\delta}\right]^{1/\delta}\right\} = \psi\left(\sum_{j=1}^{d}\psi^{-1}(u_j;\delta);\delta\right) = \exp\{-A(x;\delta)\},$$

where ψ is the generator and the ψ^{-1} is as an inverse form, u is the marginal distribution of the variable, $\delta > 1$ is some parameter.The VaR estimation imposes a copula based GARCH model including three different distributions on error terms: the normal distribution, the student-t distribution and the standardized skew Student-t distribution. The steps of estimation are as follows: First of all, the estimated results of the preferable copula function were used for generating the random numbers; secondly, applying the inverse of the function of a corresponding preferred conditional distribution of the marginal to obtain the standardized residuals; finally, we forecast the VaR by using the rolling windows at one period ahead (t+1) by the adequate model of the Copula-GARCH model with the preferred conditional distribution.

III. EMPIRICAL RESULTS

3.1. The data

In 2010, the HKSE had approximately 1,300 stocks in main board. The data was obtained from Yahoo finance using the function 'getSymbols' of the package 'quantmod' in R. The data was collected daily from the 1st of January 2010 to the 15st of December 2016. The collected data was cleaned of missing values and other variables, and was run all the models to select the stock portfolio. Companies with incomplete data, because some stocks were suspended, were deleted in this time period, thus, 700 stocks are used. The dataset of daily log returns of the closing price has been separated to two parts as an estimation period from the 1st of January 2010 to the 31st of December 2014 with totally 1246 observations and a forecasting period from the 1st of January 2015 to the 15st of December 2016 with totally 499 observations.

3.2 Results for stock selection

Our dataset contains daily log returns of three portfolios with 5 stocks each over the period. An available stock tickers data of Hong Kong stock market provides approximately 700 stocks which are used in my paper to choose target stocks to form 3 portfolios with 5 best stocks for each by applying the SR ranking. Nine shares are chosen by all three

portfolios such as Entertainment Group Limited(0027.HK, GE), Naga Corp Ltd(3918.HK, NC), New Ocean Energy Holdings Limited(0342.HK, NOE), Sands China Ltd.(1928.HK, SC),SJM Holdings Limited (0080.HK, SJM), 0525.HK(China Ruifeng Renewable Energy Holdings Limited, CRRE), Yuxing InfoTech(8005.HK, YIT), Pak Fah Yeow International Ltd.(0239.HK, PFYI) and Trad Chi Med(0570.HK, TCM).

3 stocks are selected in these 3 groups at the same time, respectively GE, NC and NOE.In portfolio 1,the top 5 chosen shares are GE, NC, NOE, SCL and SJM; In portfolio 2, the top 5 are CRRE, GE NC, NOE and YIT; In portfolio 3,the top 5 are NC, GE, NOE, PFYI and TCM.

3.3 Results for copula model selection

In this part, we firstly estimate all the alternative models and then select the preferable models in terms of AIC. There are seven Copula-GARCH methods with three distributions for each group. According with Akaike 1974 the smaller AIC values are better when comparing more models, the best model is t-copula-UN with sstd in portfolio1, portfolio 2 and portfolio 3.

3.4 Estimation of Parameters

We then estimate the correlation coefficients of the t-Copula model to get the covariance matrices for the 3 groups of portfolio which shows in Table 1. The estimation results provide the evidence that portfolio 1 has the highest correlation more than other two groups which coincide s with the realistic concept of the high reward getting along with the high risks. On the contrary, the values of the correlation matrix from portfolio 2 and portfolio 3 are relatively lower which indicates the diversification of risk. Specifically, SCL and GE in portfolio 1 has the highest correlation 0.6796 and both companies belong to the Hotel & Entertainment industry; Both belonging to the Hotel & Entertainment industry, NC and GE from both portfolio 2 and portfolio 3 has the highest correlation which are respectively 0.3135and 0.3129. Surprisingly, the correlation of the three pairs of stocks are all from the Hotel & Entertainment industry with the highest correlation.

Portfolio 1					
	GE	NC	NOE	SCL	SJM
GE	1	0.3074*** (0.0257)	0.1376*** (0.0285)	0.6796*** (0.0135)	0.6773*** (0.0136)
NC		1	0.1536*** (0.0284)	0.2688*** (0.0263)	0.2865*** (0.0263)
NOE			1	0.1661*** (0.028)	0.1737*** (0.0283)
SCL				1	0.6234*** (0.0157)
SJM					1
Portfolio 2					
	CRRE	GE	NC	NOE	YIT
CRRE	1	0.1206*** (0.0259)	0.0804*** (0.0262)	0.0547** (0.0262)	0.0339* (0.0261)
GE		1	0.3135*** (0.0249)	0.1335*** (0.0281)	0.1009*** (0.0281)
NC			1	0.1537*** (0.0279)	0.0155 (0.0285)
NOE				1	0.0913*** (0.0282)
YIT					1
Portfolio 3					
	NC	GE	NOE	PFYI	TCM
NC	1	0.3129*** (0.025)	0.1545*** (0.0279)	0.0684** (0.0276)	0.1176*** (0.0283)
GE		1	0.1341*** (0.0282)	0.118*** (0.0272)	0.1639*** (0.0278)
NOE			1	0.0523* (0.0277)	0.0667** (0.0288)
PFYI				1	0.0211 (0.0278)
TCM					1

Table 1.Estimation of parameters for selected copula

Conversely, GE from the Hotel & Entertainment industry and NOE from the Oil & Gas industry in portfolio 1 has the lowest correlation 0.1376; CRRE from the Utilities industry and YIT from the IT Hardware industry in portfolio 2 has the lowest correlation 0.0339.NOE rom the Oil & Gas industry and PFYI from the Health & Personal care industry in portfolio 3 has the lowest correlation 0.0523. NC and YIT in portfolio 2 has no correlation. PFYI and TCM both from the Health& Personal care industry in portfolio 3 has no correlation. The marginal distributions for each stock in the portfolios are not iid. but a dependent structure. Also the results of the covariance matrix are significant. We can conclude that the chosen Copula model adequately captured the dependent characteristic of the portfolios.

3.5 Value at Risk measurement

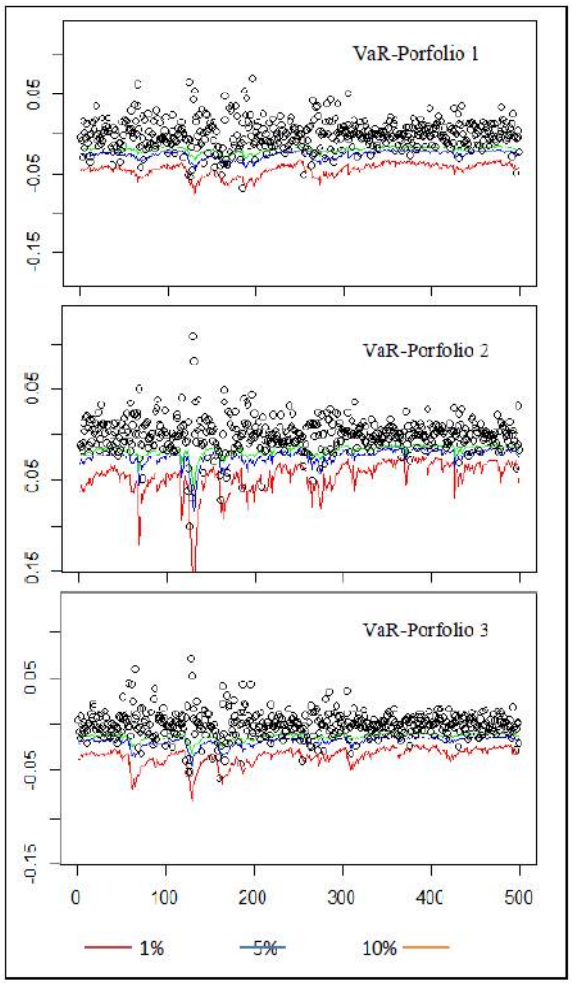


Figure1.VaR for portfolios at α = 1%, 5% and 10%

The performance of the VaR forecasting can be measured by using the number of violations of the VaR that the forecasting is based on the adequate model the copula-GARCH model in Table 2. We set α=0.01, α=0.05 and α=0.1 for the VaR estimation for 5 portfolios of 3 groups of the return of the stock price in HKSE. Figure 1. shows the VaR plot we estimate using the best fitted copula (t-copula-UN) for different groups of portfolios.

The paper imposes 1745 observations of the return for each individual stock which the first 1246 observations as the in-sample observations and the last 499 as the out-sample one. Therefore, we start to estimate the VaR_{1247} at a time t=1247. By using the rolling window method, we estimate 499 VaRs one by one with 1246 previous observation each. Since there are 499 sample-out observations, it has a total of 499 tests for VaR. The number of violations of the VaR were estimated by applying copula functions which shown in Table 2. The result shows the numbers of sample observations that stayed out of the critical value.

No. observations=499	α =0.01	α =0.05	α =0.1
Expected no. of violations	5	25	50
Portfolio 1	4	27	62
Portfolio 2	3	30	63
Portfolio 3	3	36	68

Table 2. Number of violations of the VaR estimation

From the results in Table 2., the estimation of the number of violations calculated in portfolio 1 is closer to the expected number of violations with a 99%, 95% and 90% level of confidence (α = 0.01,α =0.05 and α = 0.1). Besides, Figure 1. shows the VaR plot we estimate using the T-copula with sstd model at α = 0.01, α =0.05 and α = 0.1. VaR is an estimate of investment loss in the worst case scenario with a relatively high level of confidence. In this Figure the VaR of portfolio is located almost below the portfolio returns, and describes the expectation of investment loss well. The portfolio return of VaR with a 99% confidence is surely lower than that with a 95% confidence. By using the best fitted t-Copula-GARCH-UN-sstd model for three portfolios, the results of the VaR shows the portfolio 1 which selected by the traditional Sharpe ratio with standard deviation has the best performance in terms of the closest risk measure of the VaR value.

CONCLUSIONS

In financial market, it is important to precisely estimate the financial portfolios in order to allocate the investment capitals correctly. However, traditional portfolio method's restriction on the normality of the joint distribution limit the effectiveness of the estimation of the portfolio. We relax the assumption of the normal joint distribution to improve the performance of the estimation and forecasting.

First, using the traditional Sharpe ratio method to select stocks is workable, and has good performance in HKSE market. There is relatively low dependence

among most of selected stocks, while SCL and SJM, GE and SJM have relatively high relationship because of the similar business scope. These enterprises also have good performance in the HKSE, which supports the stock selection with the Sharpe ratio method.

Second, according to AIC, the multivariate Student-t copula in each group of portfolio returns describes the dependence structure quite well, and produces the best results of the reliable VaR limit.

Finally, multivariate copula-GARCH model successfully calculate VaR at a level 0.01 for the three group portfolios. Therefore, we conclude that the multivariate copula method has the feature of flexibility in distribution which is more appropriate in studying highly volatile financial markets comparing the traditional methods.

REFERENCES

[1] T. Berger, and M. Missong, 'Copulas and Portfolio Strategies: An Applied Risk Management Perspective', Journal of Risk, 17 (2014), 51-91.

[2] John B Guerard, Harry Markowitz, and GanLin Xu, 'Earnings Forecasting in a Global Stock Selection Model and Efficient Portfolio Construction and Management', International Journal of Forecasting, 31 (2015), 550-60.

[3] Luiz Koodi Hotta, Edimilson C Lucas, and Helder P Palaro, 'Estimation of Var Using Copula and Extreme Value Theory',(2006).

[4] J. J. Huang, K. J. Lee, H. M. Liang, and W. F. Lin, 'Estimating Value at Risk of Portfolio by Conditional Copula-Garch Method', Insurance Mathematics & Economics, 45 (2009), 315-24.

[5] E. Jondeau, and M. Rockinger, 'The Copula-Garch Model of Conditional Dependencies: An International Stock Market Application', Journal of International Money and Finance, 25 (2006), 827-53.

[6] E. Kole, K. Koedijk, and M. Verbeek, 'Selecting Copulas for Risk Management', Journal of Banking & Finance, 31 (2007), 2405-23.

[7] Jianxu Liu, Songsak Sriboonchitta, Panisara Phochanachan, and Jiechen Tang, 'Volatility and Dependence for Systemic Risk Measurement of the International Financial System', in Integrated Uncertainty in Knowledge Modelling and Decision Making: 4th International Symposium, Iukm 2015, Nha Trang, Vietnam, October 15-17, 2015, Proceedings, ed. by Van-Nam Huynh,Masahiro Inuiguchi and Thierry Demoeux (Cham: Springer International Publishing, 2015), pp. 403-

[8] J. Patton, 'A Review of Copula Models for Economic Time Series', Journal of Multivariate Analysis, 110 (2012), 4-18.

[9] William F Sharpe, 'The Sharpe Ratio', The journal of portfolio management, 21 (1994), 49-58.

[10] M. S. Smith, 'Copula Modelling of Dependence in Multivariate Time Series', International Journal of Forecasting, 31 (2015), 815-33.

[11] J. F. Sun, E. W. Frees, and M. A. Rosenberg, 'Heavy-Tailed Longitudinal Data Modeling Using Copulas', Insurance Mathematics & Economics, 42 (2008), 817-30.

★★★