



Munich Personal RePEc Archive

## **RIN Market: price behavior and its forecast**

Kakorina, Ekaterina

European University at Saint-Petersburg

16 February 2014

Online at <https://mpra.ub.uni-muenchen.de/54062/>  
MPRA Paper No. 54062, posted 09 Mar 2014 23:43 UTC

## **RIN Market: price behavior and its forecast**

### **Abstract**

In the 90th the Kyoto Protocol was signed and a market for emissions emerged. This market has one problem: it is too difficult to measure how much the company is polluting. The USA solved this problem by creating a similar market, namely the RIN (Renewable Identification Number) market. Unlike emissions, presently RINs are traded without the exchange. The importance of the RIN trading is likely to increase in the future and the goal of this paper is to research the RIN price behavior and to forecast the prices using ARMA-t-GARCH models. This paper shows that it is not important how to estimate these series (separately or together), because the estimation of parameters are very similar and the forecasted gaps are similar too. Also the common estimation using DCC-GARCH model made it possible to ascertain that these series have positive correlation in each pair.

### **Introduction**

For the two recent decades one of the world's most important questions has been about ecology. More and more countries think about an ecological situation, most of them are concerned that the road transport emissions are higher than 50% of all CO<sub>2</sub> emissions and they are continuing to grow.

In 90th the Kyoto Protocol was signed and a number of countries created a market for emissions. The main idea of this market is that every plant has emission quota which says how much CO<sub>2</sub> emissions it can have. If she one doesn't use up all the quota, it can sell a part of it to another manufacturer. In the opposite situation, if the company has used up all the quote, it can buy some from someone else. This method is good for stimulating the producers to reduce pollution. The main problem of this market is that it is very difficult to measure how much the company is polluting.

The USA solved this problem by creating a similar market in 2005. It is called the RIN market. RIN stands for Renewable Identification Number, but it is not only number, it is also a financial asset which a producer can trade like emissions. Because of similarity with emissions, many Americans call RIN a tax. It means that the government stimulates blenders to add ethanol to gasoline before selling it to the service stations. Every gallon of ethanol which was produced in the USA or was imported to this country must have one RIN, and every blender has own RFS (Renewable Fuel Standards), which specifies the minimum amount of RIN it must possess. A blender can trade RINs in a way similar to trading emissions, and it is important that they can trade RINs separately from ethanol.

As it was said before, both the emission market and the RIN market have many common characteristics, but there is one big difference. The emissions are traded on an exchange. Many

countries trade emissions with others. For example, Russia doesn't have the local emission market, but it has the opportunity to sell its quota. Unlike emissions, RIN are traded without the exchange, but it is reasonable to suppose that in some years it will be possible to buy or sell RINs on the exchange.

As the importance of RIN trading is likely to increase, the goal of this paper is to research the RIN price behavior and to forecast the price. More information about RINs and RIN prices including the data we analyzed are placed in Section 1. A detailed description of analysis method (GARCH model), the rationale for choosing this method as well as our results regarding the RIN market are placed in Section 2. The critique of this method and a proposal of a possible way to improve it, linear interpolation, are placed in Section 3. The common model for the estimation of prices behavior and correlations (multivariate GARCH) is placed in Section 4. All results of this paper are repeated in the Conclusion.

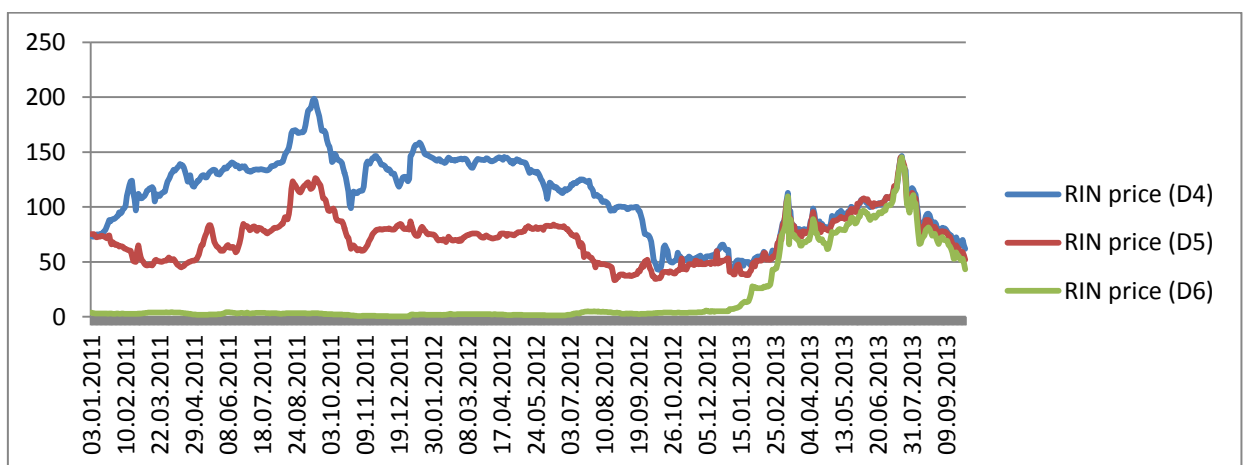
## 1. Data description

There are two groups of RINs: advanced and conventional. Advanced RIN is for ethanol which is produced from cellulose, bio-mass or other biofuels. Conventional RIN is only for corn ethanol.

As it is difficult to find a free database, this paper analyzes only prices of three advanced RINs: biomass-based diesel (D4), advanced biofuels (D5) and cellulose (D6). The daily prices were taken from the site of a young American company EcoEngineers<sup>1</sup>, which offers a wide range of RIN services like consulting, reporting and certification. This company presents prices only since the beginning of 2011, so the prices were analyzed from January, 2011 to September 2013 (691 observations for each series).

RIN prices are very unstable (Picture 1). They very often change their values. And approximately in December, 2013 the price of D6 RIN went up and became closer to the prices of other RIN types.

Picture 1. RIN prices (in cents)



<sup>1</sup> <http://www.ecoengineers.us/>

For the following analysis it is important to attest that the series is stationary, so one checks the order integrability using the Dickey-Fuller test. The null hypothesis is there is a unit root at some level of confidence. If series has one unit root, it means that the first difference is stationary (the characteristics of the series remain constant over time). For the D4 RIN prices the Dickey-Fuller test says that the series has minimum one unit root with the probability 0.4 (it is more than 10%), in other words, the series is not stationary; for the first difference of the series the test says that the probability to be wrong if one rejects the null hypothesis equals 0, so this new series is stationary. In the table 1, the reader can find the results of the Dickey-Fuller test for all series. All of the original series are not stationary, but all of the first differences are stationary.

Table 1. Probabilities for the Dickey-Fuller test

	D4 RIN prices	D5 RIN prices	D6 RIN prices
The original series	0,4216	0,1191	0,7091
The first difference	0	0	0

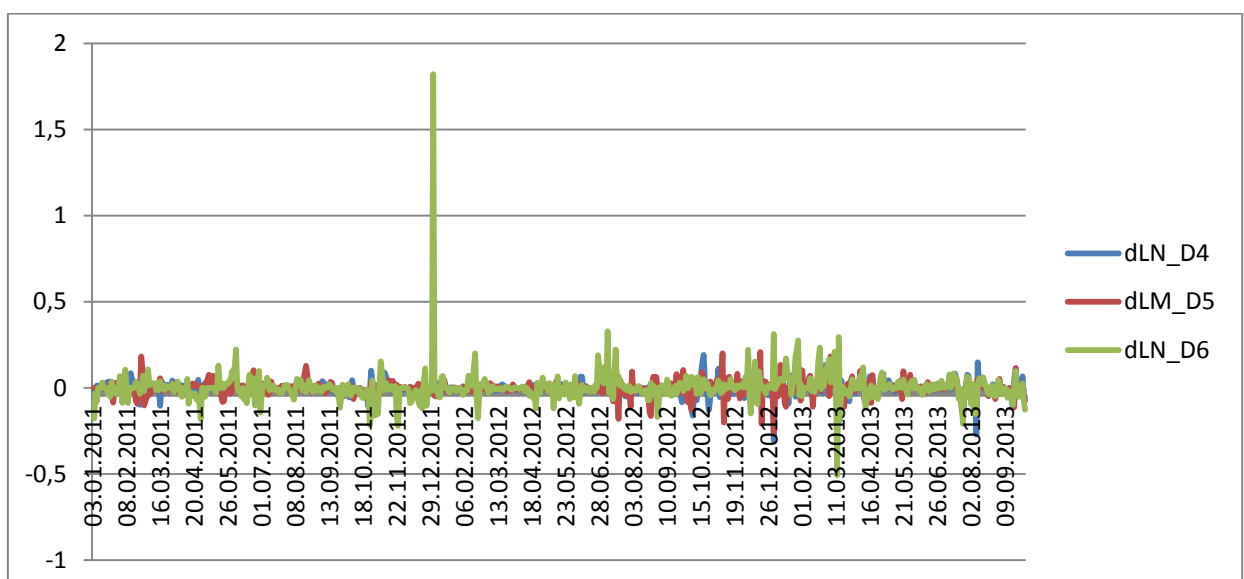
According to the receiving information, the series should be reconstructed to the first difference or returns. Researchers often use the first difference of logarithmic prices, but since the equation,

$$r_t = d\ln(P_t) = \ln(P_t) - \ln(P_{t-1}) = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln\left(1 + \frac{P_t - P_{t-1}}{P_{t-1}}\right) \approx \frac{P_t - P_{t-1}}{P_{t-1}}$$

where  $r_t$  is return and  $P_t$  is price of current period, is right for prices which are not too different from each other, so it is possible to use the ratio of the first difference of prices and the previous price.

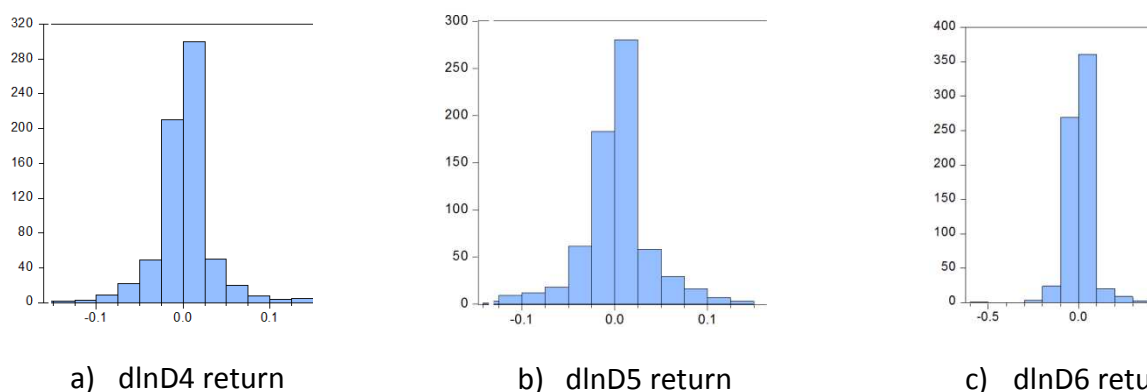
On the picture 2 the reader can see the return of all series. Almost all of observations have relatively little variation (volatility), except one observation. More information about a modeling of this outlier will be in Section 3.

Picture 2. The returns of D4 RIN price, D5 RIN price and D6 RIN price



The next important point is a distribution. Three histogram of all returns are on the picture 3 (a,b,c).

Picture 3. Histograms of returns



These histograms look like histograms for the normal distribution, so in order to check this supposition we use the Jarque-Bera test. The null hypothesis of the test is a distribution is normal. For all returns the probability of a mistake if the null hypothesis is rejected equals 0, so it means these distributions are not normal, so it is probably Student.

## 2. Univariate GARCH

After the short description of the data it is reasonable to give an explanation what model will be used for the estimation. Because of the similarity of the emission market and the RIN market, it is possible to use a paper about the investigation of the Carbon Financial Instrument (CFI) (Sabbaghi, Sabbaghi, 2011).

The authors estimated GARCH model. In their case, the CFI was non-stationary, but the CFI return was stationary. Using daily data the authors estimated t-GARCH model which used Student distribution. Referring to Tsay's paper (Tsay, 2010) they argued that the t-GARCH (1,1) model is the most effective specification for applied econometric modeling. Despite that, they checked all models from t-GARCH(1,1) to t-GARCH(4,3) and discovered that t-GARCH(1,1) had really the least Schwarz statistic among the estimated models.

The main idea of Sabbaghis' paper was a removal of zero returns from the series. The zero return means that maybe at that day it was no trade, so this observation can be excluded as soon as holiday observations are not recorded. According to this method, 71, 122 and 104 zero observations were removed from D4, D5 and D6 RIN returns, respectively.

Below the reader can find the way that was used to estimate t-GARCH models for RIN returns.

Firstly, it is necessary to test the series on the ARCH affect. A time series has autoregressive conditional heteroscedastic (ARCH) effects if it exhibits conditional heteroscedasticity or autocorrelation in the squared series. If model doesn't have the ARCH affect, a model doesn't have a GARCH part, only an ARMA one.

There are two tests, ARCH test and Ljung-Box test. Engle's ARCH test is a Lagrange multiplier test to assess the significance of ARCH effects. The null hypothesis is

$$H_0: a_0 = a_1 = a_2 = \dots = a_m = 0$$

in an equation

$$e_t^2 = a_0 + a_1 e_{t-1}^2 + \dots + a_m e_{t-m}^2 + u_t$$

where  $e_t$  is the residual series and  $u_t$  is a white noise error process. The test statistic for Engle's ARCH test is the usual F statistic for the regression on the squared residuals. Under the null hypothesis, the F statistic follows a  $\chi^2$  distribution with  $m$  degrees of freedom. A large critical value indicates rejection of the null hypothesis in favor of the alternative.

The second test is Ljung-Box test checked that the first  $m$  lags of the sample autocorrelation function of the  $e_t$  series are zeros, so the null hypothesis is

$$H_0: p_0 = p_1 = p_2 = \dots = p_m = 0$$

Under the null hypothesis,

$$Q(m) = N(N + 2) \sum_{h=1}^m \frac{p_h^2}{N - h}$$

follows to the  $\chi^2$  distribution with  $m-g$  degrees of freedom, where  $N$  is the length of the observed time series,  $g$  is a number of parameters in a model. Also Ljung-Box test is possible to use for checking autocorrelation of the residuals, so the researcher can understand if a model has ARMA part.

All results of these two tests for D4 RIN returns are presented in the table 2, and the reader can notice that in all case tests chose  $H_1$ , it means the model which will be estimated is ARMA-t-GARCH.





















Table 2. Results of ARCH test and Ljung-Box test.

m		ARCH-test	Ljung-Box(e)	Ljung-Box(e <sup>2</sup> )
5	H	1	1	1
	pValue	3.1969e-04	7.1106e-10	2.6498e-05
	ARCHstat/Qstat	23.1236	51.4150	28.7044
	CriticalValue	11.0705	11.0705	11.0705
10	H	1	1	1
	pValue	1.5027e-04	9.7649e-10	6.9474e-08
	ARCHstat/Qstat	34.5279	62.9997	53.1621
	CriticalValue	18.3070	18.3070	18.3070
15	H	1	1	1
	pValue	0.0024	6.5791e-10	9.4227e-07
	ARCHstat/Qstat	35.0978	74.6389	56.6460
	CriticalValue	24.9958	24.9958	24.9958

The similar results were received for the other series, so ARMA-t-GARCH model will be estimated for all three series.

Secondly, the order of ARMA model and the order of GARCH model should be identified. The order of ARMA model is showed by a correlogram of the residual series. If the autocorrelation function cuts short on n lag, it means the model is MA(n), if the partial correlation function cuts short on m lag, the model is AR(m). On the picture 4, the reader can see the correlogram of residuals for D4 RIN returns. It said that the maximum logical order of the model is ARMA(3,2).

Picture 4. The correlogram of residuals for D4 RIN returns

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.241	0.241	36.191	0.000
		2	0.132	0.079	47.107	0.000
		3	-0.039	-0.093	48.046	0.000
		4	-0.056	-0.041	50.033	0.000
		5	-0.047	-0.010	51.415	0.000
		6	-0.094	-0.079	56.919	0.000
		7	-0.056	-0.020	58.909	0.000
		8	-0.000	0.033	58.909	0.000
		9	0.068	0.062	61.810	0.000
		10	0.043	-0.002	63.000	0.000

Similarly to Omid and Navid Sabbaghi (Sabbaghi, Sabbaghi, 2011), different models ranging from AR(1)-t-GARCH(1,1) or MA(1)-t-GARCH(1,1) to ARMA(3,2)-t-GARCH(4,4) were estimated. The AR(1)-t-GARCH(1,1) was chosen as the model which had the least BIC statistic. The same result was received for D5 RIN returns, so the model has the following structure:

$$r_t = \varphi_0 + \varphi_1 r_{t-1} + V_t$$

$$V_t = \sqrt{h_t} u_t \quad u_t \sim t(n)$$

$$h_t = \gamma + \alpha V_{t-1}^2 + \beta h_{t-1}$$

where  $r_t$  is return,  $h_t$  is volatility and  $n$  is the number of degree of freedom.

The model for the D6 RIN returns has another structure, because as it was said above this series has one shock, so because of it the first equation has the additional parameter, and the best model, according the BIC statistic, is AR(1)-t-GARCH(1,1).

$$r_t = \varphi_0 + \varphi_1 r_{t-1} + \varphi_2 D + V_t$$

$$D = \begin{cases} 1 & \text{if } t = 190 \\ 0 & \text{otherwise} \end{cases}$$

$$V_t = \sqrt{h_t} u_t \quad u_t \sim t(n)$$

$$h_t = \gamma + \alpha V_{t-1}^2 + \beta h_{t-1}$$

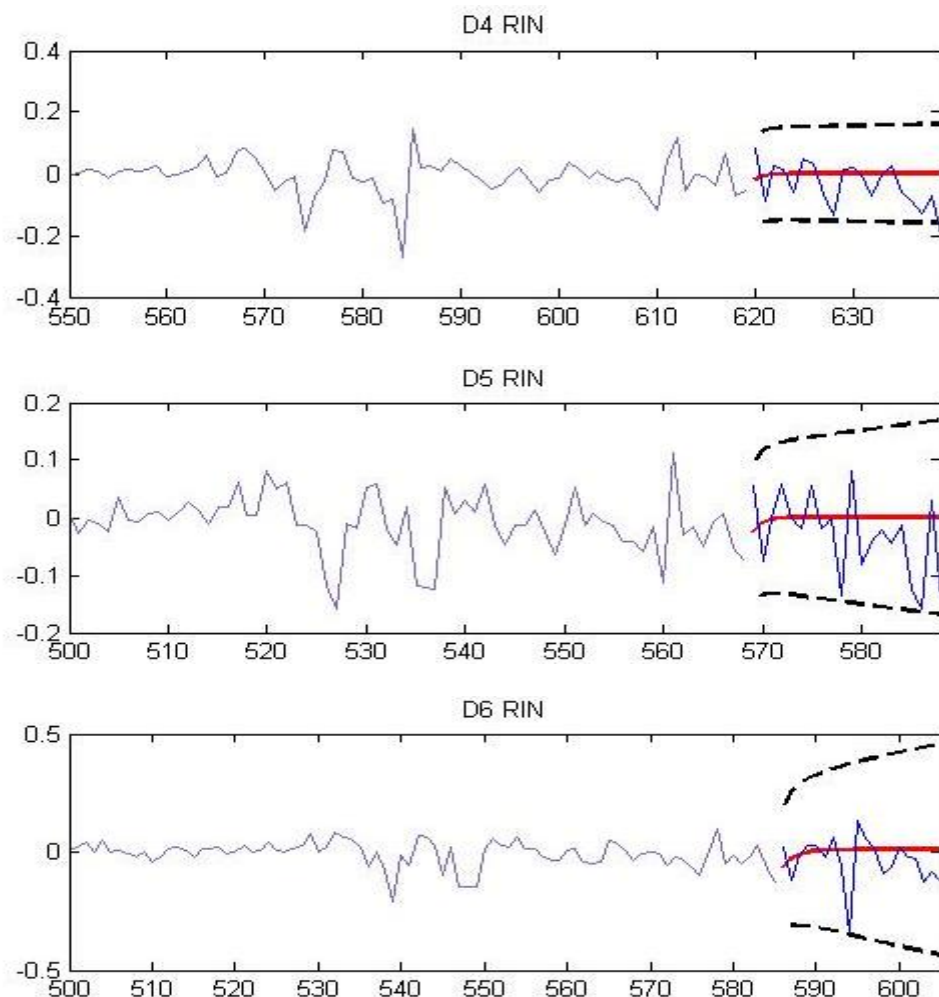
Finally, the residuals in all models were checked on ARCH effect using both ARCH test and the Ljung-Box test and there was no that effect in each model. All results of estimation are presented in Table 3. It is interesting to notice that all columns have the similar parameter values.

Table 3. Estimated parameters

parameter	D4 RIN	D5 RIN	D6 RIN
$\varphi_0$	0,0004	0,0004	0,005
$\varphi_1$	0,367	0,345	0,543
$\varphi_2$	–	–	0,854
$\gamma$	0,00004	0,0001	0,0009
$\alpha$	0,341	0,405	0,885
$\beta$	0,658	0,594	0,114
$n$	3,931	3,787	3,212

On the picture 5 the forecasted returns are showed. The red line is forecasted returns, the blue line is real returns and two black lines are the bounds of the 5% gap. Almost all real values for 20 forecasted periods are found in the gap, except several last observations of D4 RIN series.

Picture 5. Forecasted returns





In this section the price behavior and the price forecast were presented. All series has the similar structure, it is AR(1)-t-GARCH(1,1) model. All of them forecast relatively well, almost all real returns are in the predicted gap. So it is one of the possible ways to forecast future returns and hedge risks.

### 3. Linear Interpolation

In the previous section, it was described how it is possible to estimate three RIN returns series separately. The zero observations were excluded from the series. This method can be criticized, because if the company fixed the price value, it means that some trade operations happened and the new price was similar by chance. So suppose that it is true. Using the linear interpolation, it is possible to find the “real” value of zero returns and after that to estimate a multivariate GARCH model in the next section that shows correlations of different series.

The linear interpolation is one of the simplest methods of polynomial interpolation which is the interpolation of a given data set by a polynomial: given some points, find a polynomial which goes exactly through these points. The linear interpolation takes two data points, for instance  $(x_a, y_a)$  and  $(x_b, y_b)$  and the interpolant is given by:

$$y = y_a + (y_b - y_a) \frac{x - x_a}{x_b - x_a}$$

at the point  $(x, y)$ .

If  $y$  and  $x$  is changed to time period and returns, this formula will look like:

$$r_t = r_{t-1} + (r_{t+1} - r_{t-1}) \frac{t - (t - 1)}{(t + 1) - (t - 1)} = \frac{1}{2} (r_{t+1} + r_{t-1})$$

In Table 4 the reader can see the number of zeros and their percentage in the whole sample before and after the linear interpolation. It is easy to notice that the percentage of zeros decrease considerably, maybe, it will make possible to receive more accurate estimations of parameters and the exact forecast of returns.

Table 4. Quantity of zeros and their percentage

series of returns	before		after	
	number	percentage	number	percentage
D4 RIN	71	10	7	1
D5 RIN	122	18	18	3
D6 RIN	104	15	26	4

### 4. Multivariate GARCH

In this section the reader can find the detailed description of common estimated model using the data after the linear interpolation, which was presented above.

The multivariate GARCH (MGARCH) models model multiple time series vectors together accounting for the interactions between these series such as correlations, spillovers.

There are several different types of MGARCH models:

- VEC-GARCH  
The model imposes a GARCH specification not only on the conditional volatilities, but also on the conditional covariances. This model is very flexible, but because it has so many parameters it is not very practical to estimate. The problem with the VEC-GARCH model is that it is hard to guarantee that a variance matrix is positive definite.
- BEKK-GARCH  
This model guarantees the positive definiteness of the variance matrix. One problem with this model is that for some sets of parameters there might be identification problems.
- CCC-GARCH  
This model assumes that the correlations are constant over time. Two weaknesses of the CCC-GARCH model are that it assumes that correlations are constant and there are no interactions between correlations.
- ECCG-GARCH  
It is model where volatility spillovers are allowed. The advantage of ECCG-GARCH model over CCC-GARCH model is that in the latter model the autocorrelations of the squared observations decline monotonically, in the former model these autocorrelations need have a monotonic decline starting from the first lag.
- DCC-GARCH  
It is model where the correlation matrix is allowed to change over time.

According to all this information, it was chosen the last type of GARCH models (DCC-GARCH), which has the following structure:

$$\begin{aligned}
 r_t &= \varphi_0 + \varphi_1 r_{t-1} + \varphi_2 D + V_t \\
 D &= \{d_{it}\} \\
 d_{3t} &= \begin{cases} 1 & \text{if } t = 252 \\ 0 & \text{otherwise} \end{cases} \\
 u_t &= D_t^{-1/2} V_t \quad u_t \sim t(n) \\
 R_t &= P_t Q_t P_t \\
 P_t &= \text{diag} \left( \text{diag}(Q_t)^{-\frac{1}{2}} \right) = \text{diag} \left( \left[ q_{11,t}^{-\frac{1}{2}}, q_{22,t}^{-\frac{1}{2}}, q_{33,t}^{-\frac{1}{2}} \right] \right) \\
 Q_t &= (1 - \delta_1 - \delta_2) S + \delta_1 u_{t-1} u_{t-1}' + \delta_2 Q_{t-1} \\
 D_t &= \text{diag} \left( h_t^2 \right) \\
 h_t &= [h_{1t}, h_{2t}, h_{3t}]' \\
 h_{i,t} &= \gamma + \alpha_i V_{i,t-1}^2 + \beta_i h_{i,t-1}
 \end{aligned}$$

where  $r_t$  is a  $k \times 3$  matrix of returns,  $V_t$  is a  $k \times 3$  matrix of errors,  $R_t$  is the correlation matrix and  $Q_t$  is an unstandardized correlation matrix. The first and the last equations are made under previous estimated models.

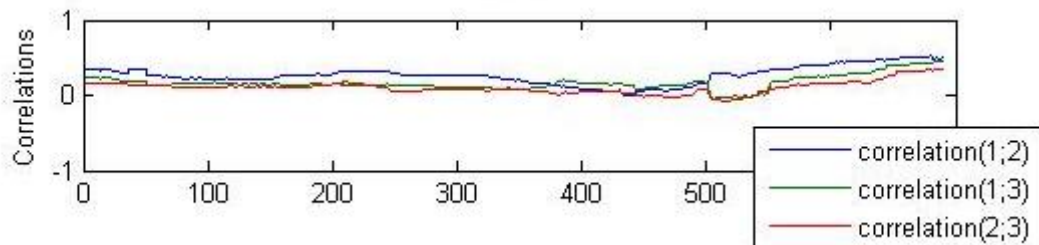
In table 5 the reader can find the results of parameters estimation using DCC-GARCH model (AR(1)-DCC-t-GARCH(1,1)). It is interesting to notice that these values are very similar to values which were received estimating separately.

Table 5. The results of the estimation of DCC-GARCH model with Student distribution

	D4 RIN	D5 RIN	D6 RIN
$\varphi_0$	-0,0003	0,000002	-0,0007
$\varphi_1$	0,4909	0,4095	0,4838
$\varphi_2$	–	–	0,912
$\gamma$	0,0001	0,0004	0,0013
$\alpha$	0,3179	0,4721	0,7088
$\beta$	0,672	0,503	0,282
$\delta_1$	0,0061		
$\delta_2$	0,9938		
$n$	4,2856		

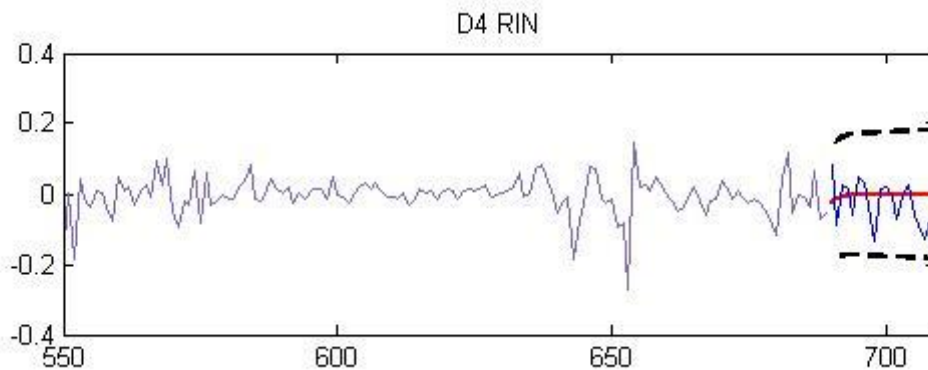
The main difference between the estimation separately and together is estimation of pair correlations. On the picture 6 it is showed that for almost the whole observed period correlations were positive and varied from 0 to 0,5 (1 - D4 RIN, 2 – D5 RIN, 3 – D6 RIN). It means that no one can hedge others security, but this information can help to forecast, because if one series has shock, other series will change too. And it is interesting that the all correlations have increased for last 200 days.

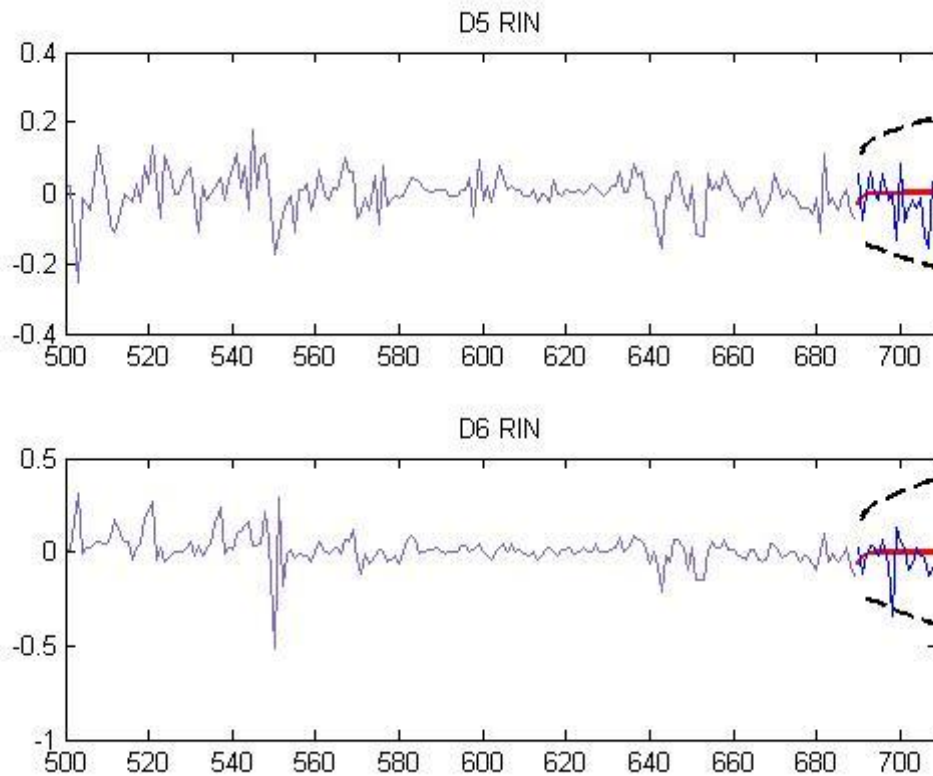
Picture 6. Correlations of RIN pairs



As it was previously, the main target of the DCC-GARCH estimation is the forecast of returns. The gaps of forecasted returns are similar to the gaps that were presented above, maybe, because of the alike parameter values (Picture 7).

Picture 7. The forecasted returns





### Conclusion

This paper shows that it is not important how to estimate these series (separately or together), because the estimations of parameters are very similar and the forecasted gaps are similar too. Also the common estimation using DCC-GARCH model made it possible to ascertain that these series have positive correlation in each pair of series. It means that no one can hedge others security, but this information can help to forecast, because if one series has shock, other series will change too. This information will help traders when RINs are traded in an exchange.

In the future it will be interesting to find security that will hedge these assets. The idea is to test the oil prices that can be negatively correlated with RINs, because for blenders oil and RIN are substitutes.

### Bibliography

Sabbaghi, O, Sabbaghi, N, Carbon Financial Instruments, thin trading, and volatility: Evidence from the Chicago Climate Exchange // *The Quarterly Review of Economics and Finance* 51, 2011, pp. 399– 407

Tsay, R, *Analysis of financial time series* (3rd ed.). Hoboken: Wiley, 2010