

'Carbon trading thickness and market efficiency' by Alberto Montagnoli and Frans P. de Vries. Energy Economics, Volume 32, Issue 6, November 2010, pp. 1331 – 1336.

Published in Energy Economics by Elsevier. Energy Economics, Volume 32, Issue 6, November 2010, pp. 1331 - 1336.

This is the peer reviewed version of this article.

*NOTICE: this is the author's version of a work that was accepted for publication in Energy Economics. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Energy Economics, VOL 32, ISSUE 6, (November 2010). DOI 10.1016/j.eneco.2010.04.001.*

# Carbon Trading Thickness and Market Efficiency\*

Alberto Montagnoli      Frans P. de Vries<sup>†</sup>

Stirling Management School, Division of Economics, University of Stirling, UK

April 2, 2010

## Abstract

This note tests for the efficient market hypothesis (EMH) in the market for CO<sub>2</sub> emission allowances in Phase I and Phase II of the European Union Emissions Trading Scheme (EU ETS). As usually is the case in emerging and non-competitive markets such as the EU ETS, trading often not occurs on a frequent basis. This has adverse implications for both the gains from permit trade as well as biases the EMH tests. Variance ratio tests are employed to adjust for the thin trading effect. The results indicate that Phase I—the trial and learning period— was inefficient, whereas the first period under Phase II shows signs of restoring market efficiency.

**JEL classification:** C14, G14, Q50

**Keywords:** carbon trading, efficient market hypothesis, thin trading, variance ratio tests, EU ETS

---

\*The authors thank two anonymous referees for helpful comments.

<sup>†</sup>Corresponding author. Stirling Management School, Division of Economics, University of Stirling, Stirling FK9 4LA, UK, Telephone: +44-1786-467485, Fax: +44-1786-467469, e-mail: [f.p.devries@stir.ac.uk](mailto:f.p.devries@stir.ac.uk).

# 1 Introduction

The European Union Emissions Trading Scheme (EU ETS) for trading carbon dioxide ( $\text{CO}_2$ ) emissions has generated a great interest among academics and practitioners alike to try to assess the functioning and actual behavior of this relatively young market. In this market, regulated firms as well as other investors can buy or sell emission allowances. From an investment point of view, an assessment of the corresponding market behavior is a necessary step for the correct implementation of (carbon) management strategies and as such relevant for investors, risk managers and environmental policymakers. At the heart of this is the efficient market hypothesis (EMH) asserting, in its weak form, that a market is efficient if its current price reflects all available information. This implies that investors cannot earn abnormal profit by exploiting past information (e.g., Fama, 1970). In this paper, we use the idea that the weak form efficiency can be tested using the random walk hypothesis and we utilize variance ratio tests to investigate whether the returns on the  $\text{CO}_2$  market follow a martingale difference sequence.

However, in non-competitive and emerging markets, such as the EU ETS, there is often too little trade (e.g., Wirl, 2009). It is well known that market frictions characterized through infrequent or “thin” trading adversely affects the gains from permit trade (Liski, 2001) as well as seriously biases the result of the EMH tests and introduces the problem of serial correlation (Miller et al., 1994). As a consequence, thin trading has direct implications for effective risk management in  $\text{CO}_2$  or other type of pollution markets. This note aims at examining to what extent adjusting for the possibility of thin trading affects the inferences drawn from testing the EMH of the EU ETS, hence assessing the role of expectations with respect to the  $\text{CO}_2$  returns in this market.

Whilst the literature on the price dynamics of  $\text{CO}_2$  allowances as part of the EU ETS is steadily increasing, the issue of thin trading in relation to the EMH has not been addressed so far. The closest to our contribution is the study by Daskalakis and Markellos (2008), who empirically test for the weak form efficiency in the European carbon market. They find no econometric support that the market is behaving efficiently. Among other things, Seifert et al. (2008) present a stochastic equilibrium model which incorporates the main features of the EU carbon market. Using an autocorrelation analysis they show that  $\text{CO}_2$  prices exhibit non-stationary behavior and that its evolution is not different from the U.S.  $\text{SO}_2$  market, i.e., the EU ETS is informational efficient. Paoletta and Taschini (2008) undertake a pure econometric analysis addressing the heteroskedasticity and the unconditional tail distribution behavior of the  $\text{SO}_2$  and  $\text{CO}_2$  spot market returns. They propose the use of a mixed-normal GARCH model to describe and forecast the returns on the  $\text{CO}_2$  allowances. Benz and Trück (2009) look at the  $\text{CO}_2$  spot price dynamics and at the volatility of the returns and advocate the implementation of Markov switching and AR-GARCH models. Finally, Daskalakis et al. (2009) show that the EU ETS spot prices exhibit jumps and non-stationary behavior.

Our contribution extends the discussion and aforementioned literature by evaluating the EMH for the EU ETS with the explicit adjustment for the possibility of thin trading. We particularly use a series of

variance ratio (VR) tests. These tests have been widely used in finance research, but have, to the best of our knowledge, not been applied to analyze the functioning of tradable permit markets or other “commodity” markets.<sup>1</sup> The VR tests resemble the class of non-parametric tests which have the advantage of preserving flexibility in the functional specifications, in particular in the context of tradable permits (or quota) (e.g., Oude Lansink and van der Vlist, 2008, p.488).

The paper proceeds with a description of the empirical framework. Section 3 discusses the data and provides some basic statistics. Results are presented in Section 4, followed by conclusions in Section 5.

## 2 Empirical framework

Our empirical methodology uses a series of variance ratio tests to investigate whether the EU ETS is efficient. Lo and MacKinlay (1988, 1989) first exploited the idea that variance of a random walk process is linear in all sampling intervals. This means that if the series under investigation follows a random walk, then its variance increases linearly with time, i.e., the variance of a  $k$ -period change must be  $k$  times the variance of the 1-period change. The VR of a  $k$ -period series can formally be defined as:

$$VR(k) = \frac{var(x_t + x_{t-1} + \dots + x_{t-k+1})/k}{var(x_t)} = 1 + 2 \sum_{i=1}^{k-1} \frac{(k-i)}{k} \rho_i, \quad (1)$$

where  $\rho_i$  is the  $i$ th lag autocorrelation coefficient of  $x_i$ . Eq. (1) shows that  $VR(k)$  is a linear combination of the 1st  $(k-1)$  autocorrelation coefficients with linearly decreasing weights. This implies that  $VR(k) = 1$  under the EMH.

Lo and MacKinlay (1988, 1989) present a number of statistics for testing the EMH hypothesis based on the estimated variance ratio:

$$VR(k) = \frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(1)}, \quad (2)$$

where  $\hat{\sigma}^2(1)$  is the unbiased estimator of the one-period variance. The form of the statistic depends on the particular random walk model assumed under the null hypothesis (see Campbell et al., 1997). There are, however, various problems with the original VR test. First, the test is based on asymptotic approximation and it is subject to severe size distortions and low power. To correct for this problem non-parametric tests have been suggested (e.g., Wright, 2000). Second, the null hypothesis is tested on selected values of  $k$ . In order to accept the hypothesis that a time series has mean reverting properties, one should be able not rejecting the hull for all values of  $k$ ; however, conducting separate tests for each value of  $k$  leads to size distortions. It is therefore necessary to conduct a joint test to investigate whether the null hypothesis holds for all values of  $k$ .<sup>2</sup> One problem still remains: even with these tests the researcher needs to choose the  $k$  holding periods. This, however, is an arbitrary choice since there is not an optimal value of  $k$ . To overcome

---

<sup>1</sup>An exception is Charles and Darné (2009), who apply VR tests to the crude oil market.

<sup>2</sup>Examples of multiple variable ratio tests are presented in Chow and Denning (1993) and Kim and Shamsuddin (2008).

this problem, Choi (1999) has suggested an entirely data-dependent procedure to determine the value of  $k$ . As proven in Kim (2009), this methodology, combined with a bootstrap technique, shows no size distortions and has substantially higher power than its competitors.

### 3 Data and summary statistics

We analyze and test the CO<sub>2</sub> return data for Phase I and the first time period of Phase II. The price data on both phases come from *BlueNext Spot*, which is the major spot market for EU ETS allowances covering about 75% of the market. The sample for Phase I covers the period 27 June 2005 until 28 December 2007; the sample for Phase II covers the period 26 February 2008 until 30 December 2009. This yields 627 and 471 observations respectively. Fig. 1 shows the evolution of the spot price for CO<sub>2</sub> allowances for both these phases respectively. At a first glance, the graph for Phase I suggests that the CO<sub>2</sub> prices were relatively stable, varying between 20-30 Euro per ton. Fig. 1 also clearly shows the sharp downward fall of the carbon price immediately after information became public by the end of April 2006 that there was a *de facto* overallocation of allowances. After this sharp fall the spot price recovers slightly and increases from 10 to about 18 Euro and subsequently varies around 15 Euro for a while. Starting September 2006, there is a relatively fast downward adjustment to a price close to zero, and the price remains low until the end of Phase I. The sample period for Phase II shows a less irregular CO<sub>2</sub> price pattern. Starting off with a price of about 20 Euro in February 2008, there is a steady increase approaching 29 Euro per ton by the end of June 2008. Since then the prices reveal a downward trend until February 2009.

[FIGURE 1 ABOUT HERE]

Figs. 2 and 3 show for both Phase I and II the daily CO<sub>2</sub> returns<sup>3</sup> (upper panel), the corresponding distribution (middle panel) and the QQ-plot<sup>4</sup> (lower panel). Fig 2 shows that the returns are essentially zero during the first period of Phase I; however, deviation from zero starts to occur during the second half of Phase I. Comparing this with the returns during the first period of Phase II (see upper panel Fig. 3), it appears that the CO<sub>2</sub> returns are showing slightly more variation. This is confirmed by the density graphs in Figs. 2 and 3 (see middle panels). The density graph and QQ-plot against the normal distribution shows that the returns distribution also exhibits fat tails, confirming the kurtosis statistics as shown in Table 1.

[TABLE 1 ABOUT HERE]

The series exhibit significant level of skewness and kurtosis. The positive skewness implies that the returns are flatter to the right compared to the normal distribution. The kurtosis reported indicates that the return distributions have sharp peaks compared to a normal distribution. The Jarque-Bera statistics

<sup>3</sup>We calculated the log returns as:  $return_t = \ln(p_t/p_{t-1})$ , where  $p_t$  denotes the allowance price at time  $t$ .

<sup>4</sup>A QQ-plot is a graphical method for comparing two probability distributions with each other. In our case we plotted the quantiles of a normal distribution (straight line) against the quantiles of the return data. If the return data follow a normal distribution then the two lines should (roughly) coincide.

for testing normality confirm the significant non-normality of returns. Notice, however, that the level of skewness and kurtosis is much smaller in Phase II than in Phase I. This indicates that the distribution of the returns under Phase I shows a longer right tail and tend to be more concentrated on the tails of the distributions. Although Fig. 1 indicates some volatility of the CO<sub>2</sub> prices, the number of trades in CO<sub>2</sub> emissions were very limited. In fact, for the covered periods, the average number of trades per day in 2005, 2006, 2007 and 2008 was no more than 4, 8, 3 and 1 respectively.<sup>5</sup> We believe this is rather small and as such we contend the EU ETS can be classified as a thin market. Therefore, we test the EMH for the EU ETS while adjusting specifically for thin trading. To this we turn next.

[FIGURES 2 AND 3 ABOUT HERE]

## 4 Results

In this section, we test for the EMH hypothesis for the EU ETS allowance prices of CO<sub>2</sub> using a series of VR tests.<sup>6</sup> Under the null hypothesis the series follows a random walk and the variance-ratios are expected to be equal to one. The test is implemented for different holding periods,  $k$ . In particular, we test given values  $k = 2$ ,  $k = 5$ ,  $k = 10$ ,  $k = 20$  and  $k = 40$  days respectively. As suggested by Deo and Richardson (2003) the choice of  $k$  is relatively short. Table 2 reports the test statistics for the non-parametric rank-based tests ( $R1$  and  $R2$ ) and signed-based test ( $S1$ ) as detailed in Wright (2000) for the observed raw market data over  $k$  number of lags. Following Hoque et al. (2007) we reject the EMH in the case of two or more rejections at the usual level of statistic significance. The null hypothesis is rejected in Phase I for all holding periods while the EMH is accepted in Phase II when  $k = 5$  and  $k = 10$ .

[TABLE 2 ABOUT HERE]

Although Wright's (2000) VR test has the ability to perform well when the daily returns are non-normal and non-stationary, it is quite susceptible to size distortions due to sequential trading because it assumes that the test statistics computed at different time intervals —or holding periods  $k$ — are uncorrelated. As mentioned in the previous section, this is a problem associated with all the individual VR test. To overcome this possible problem we apply a multiple test as in Kim and Shamsuddin (2008).

The first three rows in Table 3 reports the values of their non-parametric VR-based test on ranks and signs. The results reinforce our previous finding, i.e. reject the null hypothesis that the CO<sub>2</sub> returns during Phase I followed an *i.i.d.* process. All overwhelming number of rejections is in the right tail of the distributions, suggesting that serial correlation is positive. However, results for Phase II show that the market appears to be efficient; the VR test indicates that we cannot reject the EMH at conventional statistical significance levels. In addition, we implemented the automatic VR test as in Choi (1999) and Kim (2009). Table 3 reports

---

<sup>5</sup>These represent the average number of trades as recorded by the EEX Emissionsmarkt/Emission Market.

<sup>6</sup>See the Appendix for a formal description of the tests implemented.

the  $p$ -value of the obtained with the wild bootstrap technique.<sup>7</sup> The results contradict with the previous finding; the test rejects the EMH of the CO<sub>2</sub> market. This implies some cautions in the interpretation of the results.

In testing the EMH in thin markets it is necessary to take into account the magnitude of trading that typically characterizes these markets. This market characteristic is associated with the asset (the tradable permit) not being exchanged at every consecutive time interval. Two strategies are followed to overcome this problem. The first strategy that should correct for some market distortion involves the use of low/medium frequency data. The second strategy involves removing the effect of thin trading by a moving average process, which reflects the periods of non-trading. In particular, Miller et al. (1994) show that the adjusted data can be obtained by estimating an AR(1) process. For the adjusted data we then repeat Wright's (2000) VR test. These results are contained in Table 3. The results confirm the rejection of the EMH for Phase I but not for Phase II.

The tests with the adjusted data are presented in the last two columns of Table 3. Again the results seems inconclusive; the automatic variance ratio test (AVR\*) test for the daily adjusted data and the weekly data are contradicting. Although we could speculate that the weekly data are better placed to correct for the illiquidity of the market, there might be another reason worth to be investigated. As we observed earlier, the CO<sub>2</sub> market exhibited a higher degree of volatility after the sharp adjustment in late April 2006 at which information was publicly disclosed signalling that emissions were considerably below the allocation of emission allowances. This information disclosure had a direct impact on the carbon price and accordingly reshaped the expectations of the market (e.g., Ellerman and Joskow, 2008). To correct for this trend break, we divided the data of Phase I into two periodic sub-samples. The first period goes from 24/06/2005 to 26/04/2006 and the second from 27/04/2006 to 28/12/2007. This latter period clearly shows a downward adjustment of the carbon price and reflects an adjustment of the expectations in the sense that the initial cap on aggregate emissions appeared to be too lenient. We conducted a structural break test on the price time series and confirms a break around that period.<sup>8</sup>

We repeated all the previous exercises for the two sub-samples of Phase I and Table 4 reports the results. We used only daily data, since the use of weekly data would make the sample too small. The empirical outcome from the automatic VR test suggests a rejection of the EMH for the first part of the Phase I, while we cannot reject the hull hypothesis for the second sub-sample. As a robustness check we also conducted the Chow and Denning (1993) and Kim (2006) tests. They reject the hypothesis that the market for the first period of Phase I was efficient.

[TABLE 3 ABOUT HERE]

[TABLE 4 ABOUT HERE]

---

<sup>7</sup>We experimented with both Mammen and normal distribution and the results appear to be qualitatively identical.

<sup>8</sup>The results are available upon request.

It should be noted that the VR test presented above has the ability to reject *linear* dependency only. Therefore, before drawing a final conclusion about the data-generating process characterizing the returns in the carbon market, we should be able to reconfirm the hypothesis after testing for the absence of *non-linear* dependence. Following Hsieh (1991) we applied the BDSL test (see Brock et al., 1996) for this independence on the residuals of the ARMA model. If we reject the null hypothesis then the series has a high probability to be non-linear, or exhibits chaotic characteristics. The ability to detect an *i.i.d.* process is subject to the choice of the embedding dimension  $m$  and the bound  $\varepsilon$ .<sup>9</sup> If we select a value for  $\varepsilon$  that is too small, the null hypothesis of a random i.i.d process will be accepted too often irrespective of it being true or false. As well, it is not safe to choose too large a value for  $m$ . To deal with this problem Brock et al. (1991) suggest that for a large sample size (i.e.,  $T > 500$ )  $\varepsilon$  should equal 0.5, 1.0, 1.5 and 2 times the standard deviations of the data. Given these concerns we present both the  $p$ -value based on asymptotic theory and on a bootstrap, where the latter was based on 1000 replications. Our results are presented in Table 5 and reconfirms support for the hypothesis that Phase II of the CO<sub>2</sub> market follows a weak EMH, while we can reject the i.i.d. hypothesis for Phase I.

## 5 Concluding remarks

This paper tests the efficiency of the EU ETS carbon market. Although market efficiency is generally important, it is particularly relevant when the market is relatively immature and in an emerging state. Assessing the behavior of market participants in the EU ETS, as reflected through the behavior of CO<sub>2</sub> spot price dynamics, is especially relevant for effective risk management and (carbon) investment strategies. However, in the relatively young European carbon market the number of trades were very limited for the period in our sample. It is this infrequent or thin trading that has negative implications for both the gains from permit trade as well as the statistical inferences for testing the EMH.

For the sample periods 27 June 2005 to 28 December 2007 (Phase I) and 26 February 2008 to 4 April 2009 (Phase II) the EMH is tested through variance-ratio tests while adjusting for thin trading. The results show that the EU ETS was inefficient during Phase I but efficient during the first period of Phase II. This suggests that the carbon market shows the first signs of maturation after the learning and trial period in Phase I.

## References

- [1] Benz, E., Trück, S., 2009, Modeling the price dynamics of CO<sub>2</sub> emission allowances. *Energy Economics* 31, 4-15.

---

<sup>9</sup>In the BDSL test  $m$  is defined as the number of consecutive points used in the set. While  $\varepsilon$  is part of the correlation integral  $C_{m,N}(\varepsilon)$ , it essentially “counts up the number of  $m$  – *histories* that lie within a hypercube of size  $\varepsilon$  of each other” (Patterson and Ashley, 2000, p.41).



- [2] Brock, W.A., Hsieh, D.A., LeBaron, B., 1991, Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence. MIT Press, Massachusetts.
- [3] Brock, W.A., Dechert, W., Scheinkman, J., LeBaron, B., 1996, A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197-235.
- [4] Campbell, J.Y., Lo, A.W, MacKinlay, A.C., 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.
- [5] Charles, A., Darné, O., 2009, The efficiency of the crude oil markets: evidence from variance ratio tests. *Energy Policy* 37, 4267-4272.
- [6] Choi, I., 1999, Testing the random walk hypothesis for real exchange rates. *Journal of Applied Econometrics* 14, 293-308.
- [7] Chow, K.V., Denning, K.C., 1993, A simple multiple variance ratio test. *Journal of Econometrics* 58, 385-401.
- [8] Daskalakis, G., Markellos, R., 2008, Are the European carbon markets efficient? *Review of Futures Markets* 17, 103-128.
- [9] Daskalakis, G., Psychoyios, D., Markellos, R., 2009, Modelling CO<sub>2</sub> emission allowance prices and derivatives: evidence from the European trading scheme. *Journal of Banking and Finance* 33, 1230-1241.
- [10] Deo, R.S., Richardson, M., 2003, On the asymptotic power of the variance ratio test. *Econometric Theory* 19, 231-239.
- [11] Ellerman, A.D., Joskow, P.J., 2008, *The European Union's Emissions Trading System in Perspective*. Report prepared for the Pew Center on Global Climate Change.
- [12] Fama, E., 1970, Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25, 383-417.
- [13] Hoque, H.A.A.B., Kim, J.H., Pyun, C.S., 2007, A comparison of variance ratio tests of random walk: a case of Asian emerging stock markets. *International Review of Economics and Finance* 16, 488-502.
- [14] Hsieh, D.A., 1991, Chaos and nonlinear dynamics: application to financial markets. *Journal of Finance* 46, 837-877.
- [15] Kim, J.H., 2006, Wild bootstrapping variance ratio tests. *Economics Letters* 92, 38-43.
- [16] Kim, J.H., 2009, Automatic variance ratio test under conditional heteroskedasticity. *Finance Research Letters* 6, 179-185.

- [17] Kim, J.H., Shamsuddin, A., 2008, Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *Journal of Empirical Finance* 15, 518-53.
- [18] Liski, M., 2001, Thin versus thick CO<sub>2</sub> market. *Journal of Environmental Economics and Management* 41, 295-311.
- [19] Lo, A.W., MacKinlay, A.C., 1988, Stock market prices do not follow random walk: evidence from a simple specification test. *Review of Financial Studies* 1, 41-66.
- [20] Lo, A.W., MacKinlay, A.C., 1989, The size and power of the variance ratio test in finite samples: a Monte Carlo investigation. *Journal of Econometrics* 40, 41-66.
- [21] Miller, M.H., Muthuswamy, J., Whaley, R.E., 1994, Mean reversion of standard and poor's 500 index basis changes: arbitrage-induced or statistical illusion? *Journal of Finance* 49, 479-513.
- [22] Oude Lansink, A., van der Vlist, A., 2008, Non-parametric modelling of CO<sub>2</sub> emission quota. *Journal of Agricultural Economics* 59, 487-497.
- [23] Paoletta, M.S., Taschini, L., 2008, An econometric analysis of emission allowance prices. *Journal of Banking and Finance* 32, 2022-2032.
- [24] Patterson, D.M., Ashley, R.A., 2000, *A Non-Linear Time Series Workshop*. Kluwer Academic Publishers, London.
- [25] Seifert, J., Uhrig-Homburg, M., Wagner, M., 2008, Dynamic behavior of CO<sub>2</sub> spot prices. *Journal of Environmental Economics and Management* 56, 180-194.
- [26] Wirl, F., 2009, Oligopoly meets oligopsony: the case of permits. *Journal of Environmental Economics and Management* 58, 329-337.
- [27] Wright, J.H., 2000, Alternative variance-ratio tests using ranks and signs. *Journal of Business and Economic Statistics* 18, 1-9.

## Appendix: VR tests

In this appendix we present the formal definition of the various VR tests implemented in this paper. The central hypothesis of the VR methodology used was first exploited by Lo and MacKinlay (1988) the idea that the variance of random walk process is linear in all sampling intervals, i.e., the variance of a  $y_t - y_{t-k}$  will be  $k$  times the variance of the first difference.

### A.1 Rank and sign test

The variance ratio test proposed by Wright (2000) is a non-parametric test based on ranks ( $R_1$  and  $R_2$ ) and signs ( $S_1$  and  $S_2$ ). These types of tests are exact under the independence and identical distribution assumption, whereas the tests based on signs are exact even under conditional heteroskedasticity. Wright's suggested  $R_1$  and  $R_2$  can be defined as:

$$\begin{aligned} R_1 &= \left( \frac{\frac{1}{Tk} \sum_{t=k}^T (r_{1,t} + \dots + r_{1,t-k+1})^2}{\frac{1}{T} \sum_{t=k}^T r_{1,t}^2} - 1 \right) \times \phi(k)^{-1/2}, \\ R_2 &= \left( \frac{\frac{1}{Tk} \sum_{t=k}^T (r_{2,t} + \dots + r_{1,t-k+1})^2}{\frac{1}{T} \sum_{t=k}^T r_{2,t}^2} - 1 \right) \times \phi(k)^{-1/2}, \end{aligned}$$

where

$$\begin{aligned} r_{1,t} &= \frac{(r(y_t - \frac{T+1}{2}))}{\sqrt{\frac{(T-1)(T+1)}{12}}}, \\ r_{2,t} &= \Phi^{-1} \frac{r(y_t)}{(T+1)}, \end{aligned}$$

and where  $\Phi^{-1}$  is the inverse of the cumulative standard normal cumulative distribution and  $\phi(k)$  is defined as:

$$\phi(k) = \frac{2(2k-1)(k-1)}{3kT}.$$

The tests based on signs of returns is given by:

$$\begin{aligned} S_1 &= \left( \frac{\frac{1}{Tk} \sum_{t=k}^T (s_t + \dots + s_{t-k+1})^2}{\frac{1}{T} \sum_{t=k}^T s_t^2} - 1 \right) \times \phi(k)^{-1/2}, \\ S_1 &= \left( \frac{\frac{1}{Tk} \sum_{t=k}^T (s_t(\bar{\mu}) + \dots + s_{t-k+1}(\bar{\mu}))^2}{\frac{1}{T} \sum_{t=k}^T s_t^2(\bar{\mu})} - 1 \right) \times \phi(k)^{-1/2}, \end{aligned}$$

where  $s_t = 2u(y_t, 0)$ ,  $s_t(\bar{\mu}) = 2u(y_t, \bar{\mu})$  and

$$u(y_t, q) = \begin{cases} 0.5 & \text{if } y_t > q \\ -0.5 & \text{if otherwise} \end{cases}.$$

The critical values of the  $R_1$ ,  $R_2$ ,  $S_1$  and  $S_2$  test can be computed by simulating their exact sampling distribution. Kim and Shamsuddin (2008) propose a test to overcome the problem that VR tests at various aggregation intervals lead to rejection rates larger than the nominal size. Their multiple test takes the form  $JS = \max_{1 \leq i \leq l} |S_l(k_i)|$ . The  $JS$  statistic has the same sampling distribution and can be obtained in a similar way for  $S_1(k)$ . The null hypothesis that the series of a random walk is accepted is when the  $JS$  statistic is smaller than the critical value.

## A.2 Automatic variance ratio test

This test was first implemented by Choi (1999) to overcome the problem associated with the choice of the holding period. He presented a procedure which is entirely data-dependent. The test can be written as:

$$VR(k) = 1 + 2 \sum_{i=1}^{T-1} z(i/k) \hat{\rho}(i),$$

where  $\hat{\rho}(i)$  is the autocorrelation function and  $z(x)$  is the quadratic spectral *kernel* defined as:

$$z(x) = \frac{25}{12\pi^2 x^2} \left[ \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right].$$

Choi (1999) showed that the automatic variance ratio test follows a standard normal distribution as  $k \rightarrow \infty, T \rightarrow \infty, T/k \rightarrow \infty$ , and the process is *i.i.d.*

$$AVR(k) = \sqrt{T}/k(VR(k) - 1)/\sqrt{2} \rightarrow N(0, 1).$$

Recently, Kim (2009) has shown that the AVR test has a serious size distortion and suggests to use a wild bootstrap procedure to correct this problem.<sup>10</sup>

---

<sup>10</sup>We thank an anonymous referee for pointing out this reference.

Table 1: Descriptive statistics

	Returns Phase I	Returns Phase II
Mean	-0.787	-0.110
Median	0.000	0.000
Maximum	66.666	10.547
Minimum	-40.000	-10.285
Std. Deviation	8.178	2.773
Skewness	0.9142	-0.181
Kurtosis	14.840	4.425
Jarque-Bera	3744.2	42.373
Probability	0.000	0.000
N.Observations	627	470

Table 2: VR tests daily and weekly returns

		$K = 2$	$K = 5$	$K = 10$	$K = 20$	$K = 40$
Phase I	$R1$	2.861**	2.995**	2.866**	2.179**	2.582**
unadj. returns	$R2$	1.453	1.622*	1.561*	1.004	1.343*
	$S1$	5.675**	6.698**	7.689**	9.016**	11.414**
Phase I	$R1$	5.298**	4.919**	4.226**	3.244**	3.424**
adj. returns	$R2$	3.919**	3.468**	2.786**	1.902*	2.063*
	$S1$	6.600**	7.887**	8.162**	8.428**	10.148**
Phase II	$R1$	1.821	0.608	0.518	0.419	0.269
unadj. returns	$R2$	2.050*	0.789	0.782	0.624	0.561
	$S1$	1.383	0.353	0.349	0.204	-0.22
Phase I	$R1$	0.032	-0.943	-0.686	-0.451	-0.351
adj. returns	$R2$	0.157	-0.871	-0.532	-0.335	-0.121
	$S1$	-0.877	-1.517	-0.752	-0.635	-0.833

**Notes:** \*\*\* and \*\* denote the significance at the 5% and 10% levels, respectively.

Table 3: VR tests daily and weekly returns

		Daily undaj. returns	Daily adj. returns	Weekly returns
Phase I	<i>R1J</i>	2.995**	5.298***	1.273
	<i>R2J</i>	1.622	3.919***	1.241
	<i>S1J</i>	11.414***	10.148***	2.271**
	<i>AVR*</i>	0.186	0.828	0.014**
Phase II	<i>R1J</i>	1.821	0.943	0.881
	<i>R2J</i>	2.053*	0.871	0.586
	<i>S1J</i>	1.383	1.517	0.816
	<i>AVR*</i>	0.594	0.508	0.802

**Notes:** ‘\*\*\*’ and ‘\*\*’ denote the significance at the 5% and 10% levels, respectively. *R1J*, *R2J*, *S1J* are the joint Wright tests statistics; *AVR\** is the *p*-value for the wild bootstrap for the Automatic Variance Ratio Test.

Table 4: VR tests daily and weekly returns for Phase I sub-samples

		Undaj. returns	Adj. returns
Sub-sample 1	<i>R1J</i>	2.257**	1.059
	<i>R2J</i>	2.401**	1.134
	<i>S1J</i>	1.876	0.795
	<i>CD(R1)</i>	2.331*	3.521**
	<i>CD(R2)</i>	1.728	2.549*
	<i>MV*</i>	0.330	0.092*
	<i>AVR*</i>	0.000**	0.000**
Sub-sample 2	<i>R1J</i>	1.251	3.013**
	<i>R2J</i>	0.271	1.837
	<i>S1J</i>	13.759***	4.197**
	<i>CD(R1)</i>	2.880**	1.709
	<i>CD(R2)</i>	1.713	1.093
	<i>MV*</i>	0.218	0.698
	<i>AVR*</i>	0.126	0.698

**Notes:** ‘\*\*\*’ and ‘\*\*’ denote the significance at the 5% and 10% levels, respectively. *R1J*, *R2J*, *S1J* are the joint Wright tests statistics, CD1 and CD2 are the statistics for the Chow Denning test, *MV\** is the *p*-value for the the wild bootstrap Chow-Denning (1993) test; *AVR\** is the *p*-value for the wild bootstrap for the Automatic Variance Ratio Test.

Table 5: BDSL test results for phase I and II

Phase I						
Dimension	$\epsilon = 0.50$	$\epsilon = 1.00$	$\epsilon = 2.00$	$\epsilon = 0.50$	$\epsilon = 1.00$	$\epsilon = 2.00$
	Unadjusted returns			Adjusted returns		
Bootstrap						
2	0.000	0.011	0.292	0.000	0.003	0.066
3	0.000	0.000	0.274	0.000	0.002	0.021
4	0.000	0.000	0.039	0.000	0.000	0.007
Asymptotic theory						
2	0.000	0.005	0.300	0.000	0.002	0.076
3	0.000	0.001	0.289	0.000	0.000	0.014
4	0.000	0.000	0.025	0.000	0.000	0.007
Phase II						
Dimension	$\epsilon = 0.50$	$\epsilon = 1.00$	$\epsilon = 2.00$	$\epsilon = 0.50$	$\epsilon = 1.00$	$\epsilon = 2.00$
	Unadjusted returns			Adjusted returns		
Bootstrap						
2	0.083	0.299	0.329	0.395	0.498	0.256
3	0.243	0.387	0.294	0.454	0.383	0.150
4	0.321	0.384	0.234	0.364	0.272	0.105
Asymptotic theory						
2	0.075	0.328	0.343	0.452	0.548	0.263
3	0.250	0.434	0.317	0.512	0.438	0.156
4	0.348	0.453	0.263	0.435	0.313	0.123

**Notes:** Only  $p$ -values are reported under the null hypothesis that the time series is a serial i.i.d. process. All calculations are done using the non-linear toolkit Patterson and Ashley (2000)

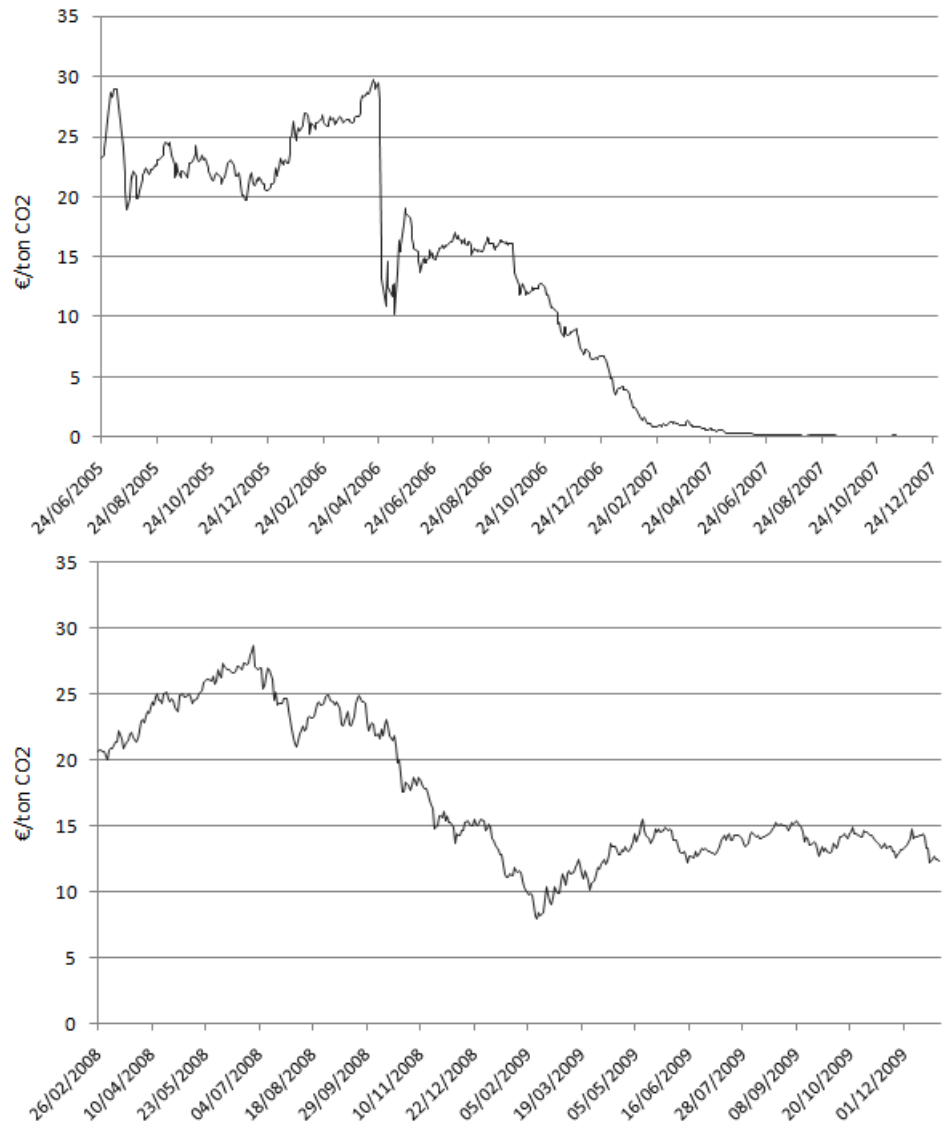


Figure 1: Daily CO<sub>2</sub> spot prices in Phase I (upper panel) and Phase II (lower panel).



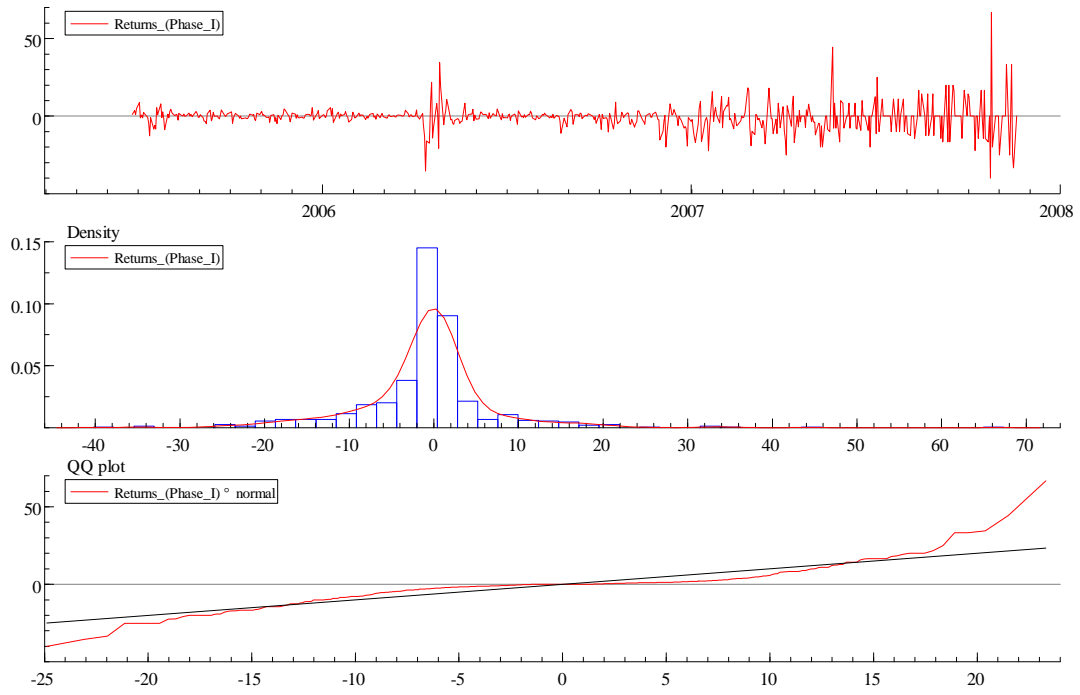


Figure 2: Daily CO<sub>2</sub> returns, density and QQ-plot Phase I.

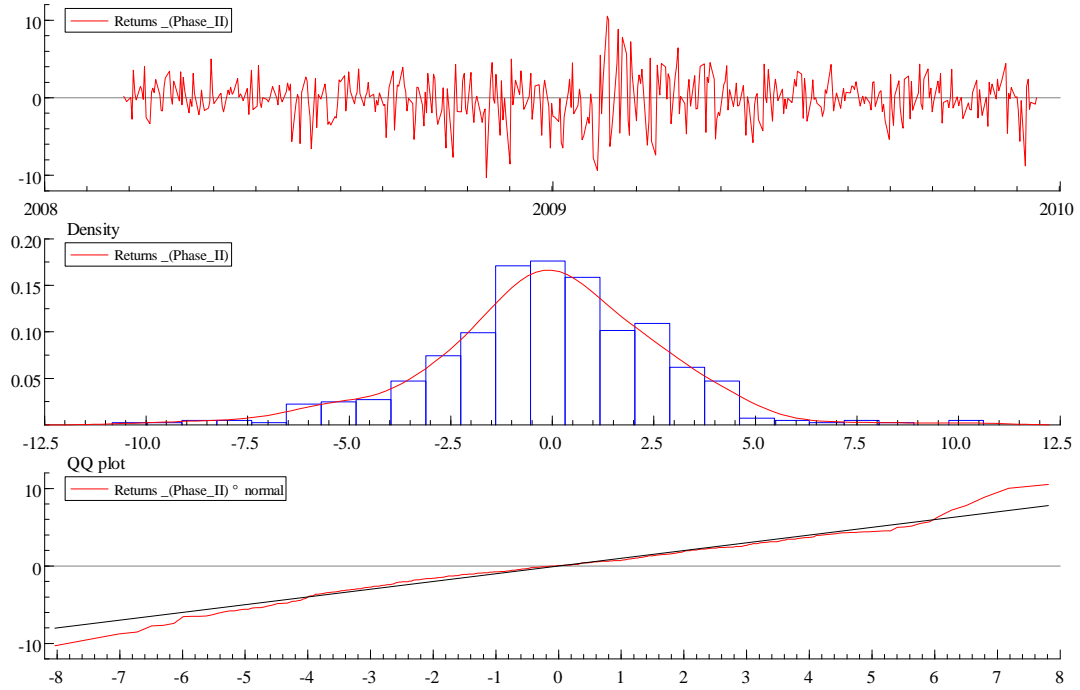


Figure 3: Daily CO<sub>2</sub> returns, density and QQ-plot Phase II.