

PREDICTIVE ACCURACY OF ALTERNATIVE
AUTOREGRESSIVE CONDITIONAL
HETEROSKEDASTICITY MODELS

Carlos Diogo Monteiro Garcia

Master's thesis in Finance

Coordinator:

Prof. Doutor José Dias Curto, Prof. Auxiliar, ISCTE Business School, Departamento de Métodos
Quantitativos

May 2010

Resumo:

Esta tese tem como objectivo comparar alguns dos mais populares modelos de volatilidade, em termos da sua capacidade preditiva. Especificamente, iremos usar três modelos auto-regressivos de heterocedasticidade condicional, GARCH, EGARCH e GJR. Para proceder à comparação entre modelos, iremos servir-nos de alguns dos mais recentes testes de capacidade preditiva: Diebold-Mariano (1995), Diebold-Mariano modificado (1997), Morgan-Granger-Newbold modificado (1997), Harvey-Leybourne-Newbold (1998), Harvey-Newbold (2000) e Hansen (2005).

A nossa análise irá ser feita com base nos índices CAC40, FTSE100, NIKKEI225 e S&P500, para o período de 1 de Janeiro de 1995 até 31 de Dezembro de 2009.

Os resultados obtidos, embora não sendo conclusivos, apontam para uma superior capacidade preditiva dos modelos assimétricos (EGARCH e GJR), face ao GARCH. O facto de não conseguirmos apontar claramente o melhor modelo, de entre os modelos assimétricos, pode ser explicado pelos diversos episódios de volatilidade elevada que tiveram lugar nas últimas duas décadas.

Palavras-chave: Previsão de volatilidade, modelos ARCH, testes de capacidade preditiva, comparação de modelos

Abstract:

The main objective of this thesis is to compare some of the most popular volatility models, in terms of their predictive accuracy. Specifically, we will use three autoregressive conditional heteroskedasticity (ARCH) models, GARCH, EGARCH and GJR. In order to compare these models, we will use some of the most recent predictive accuracy tests: Diebold-Mariano (1995), modified Diebold-Mariano (1997), modified Morgan-Granger-Newbold (1997), Harvey-Leybourne-Newbold (1998), Harvey-Newbold (2000) and Hansen (2005).

We will consider the CAC40, FTSE100, NIKKEI225 and S&p500 indexes in our analysis, from January 1, 1995 through December 31, 2009.

The results obtained, although not being conclusive, point out to a superior predictive accuracy of asymmetric models (EGARCH and GJR), in relation to GARCH. The fact that we can't clearly point out the best model, between the asymmetric ones, may be explained by the several episodes of high volatility that took place over the last two decades.

Keywords: Forecasting volatility, ARCH models, predictive accuracy tests, model comparison

Thank-you note:

I would like to thank my coordinator, Professor José Dias Curto, for accepting tutoring my thesis, helping me to choose the thesis subject and providing me the required bibliography.

Index

1. INTRODUCTION	2
2. VOLATILITY MODELS	3
3. A BRIEF DESCRIPTION OF EXISTING TESTS	5
3.1 Diebold-Mariano (1995).....	5
3.2 A modified Diebold-Mariano test.....	10
3.3 A modified Morgan-Granger-Newbold test	13
3.4 Harvey, Leybourne, Newbold (1998)	18
3.5 Harvey and Newbold (2000)	24
3.6 Peter Hansen (2005).....	30
4. STATISTICAL PROPERTIES OF RETURNS.....	39
5. ECONOMETRIC APPROACH	44
6. ESTIMATION RESULTS	48
In-sample results	48
Out-of sample results	55
7. CONCLUSION	62
8. THESIS LIMITATIONS AND FUTURE RESEARCH	63
9. BIBLIOGRAPHY:.....	64

1. INTRODUCTION

Forecasting can be described as an attempt to foresee the future by examining historical data. In simple terms, a forecast is an estimate for the future value of some variable. Therefore, forecasts may not be confused with guesses or intuition.

In corporate world, managers often predict future events based upon past experience and personal opinion. However, in these days of rapid change, where uncertainty is a reality/constant, forecasting is gaining weight in companies' decisions, being a useful tool for managers. Forecasting is important not only for those who use it, but also for those who create them, since their reputation is directly related.

Given that forecasts play an important role in modern organizations, forecasting accuracy becomes extremely important. In spite the enormous amount of studies found in the literature, forecasting accuracy comparison does not emerge as an easy task, due to several limitations. When the first formal tests of forecasting accuracy came up, the conditions imposed to the loss function and the forecast errors where too restrictive. Specifically, and in accordance to Diebold (1995):

- a) The loss function had to be quadratic

And the forecast errors needed to be:

- b1) Zero mean
- b2) Gaussian
- b3) Serially uncorrelated
- b4) Contemporaneously uncorrelated

Because some of these conditions are difficult to obtain, recent efforts have been made to surpass them and new tests with the relaxation of some conditions were proposed. Indeed, over the last few years, several papers emerged, suggesting different statistical tests to deal with forecasting accuracy.

Thus, and due to the importance of this subject for empirical finance, the main purpose of this thesis is to make use of those tests to compare the forecasting capability of some alternative conditional heteroskedasticity models.

This thesis is organized as follows. Section 2 will present the volatility models to be used in our empirical application. Section 3 will review in a brief way the available literature related with forecasting accuracy tests. Data's statistical properties can be seen in Section 4, while the econometric approach is given in Section 5. Estimation results will be discussed in Section 6 and the conclusions will be made in Section 7. Finally, Section 8 mentions some thesis limitations.

2. VOLATILITY MODELS

Due to the major importance played by risk in financial markets, and modeling and forecasting volatility In fact, modeling and forecasting volatility has become a true focus of attention over the last few years

As early noted by Mandelbrot (1963) and Fama (1965), financial time series vary systemically with time and tend to display periods of unusual large volatility, followed by periods of low volatility. With these findings, Mandelbrot and Fama pointed out the importance of volatility in financial markets. Despite their early findings, however, the efforts to model and forecast volatility only occurred over the last two decades and centered their attention in some stylized facts of asset returns.

Bollerslev et al (1994) pointed out eight empirical regularities of asset returns:

1. Asset returns tend to be leptokurtic;
2. Returns are not i.i.d. (independent and identically distributed) through time. This phenomenon is also known as volatility clustering;
3. Also known as Fisher-Black effect, the leverage effect states that volatility and asset returns are negatively correlated. In other words, price's changes of the same magnitude but different signs will reflect differently in volatility. Specifically, volatility will increase more after a negative change in the asset's price;

4. Information that accumulates when financial markets are closed is reflected in prices after the markets reopen;
5. Forecastable releases of important information are associated with high ex ante volatility. For example, individual firms' stock returns volatility is high around earnings announcements;
6. Volatility and serial correlation are inversely correlated;
7. As observed by Black (1976), a 1% market volatility change typically implies a 1% volatility change for each stock;
8. Measures of macroeconomic uncertainty help to explain changes in stock market volatility.

The first model that seemed to be able to capture some of these stylized facts was proposed by Engle (1982), who launched the first ARCH (autoregressive conditional heteroskedasticity) model. In ARCH (q), the conditional variance is a function of the past q squared innovations. Later, Bollerslev (1986) proposed a generalization of Engle's model, known as GARCH (generalized autoregressive conditional heteroskedasticity), a more parsimonious model than ARCH, as empirical findings suggest. The GARCH (p,q) permits additional dependencies on p lags of the past conditional variance. GARCH (1,1) is the most popular structure for many financial time series.

Although ARCH and GARCH models proved to be able to capture the volatility clusters stylized fact of returns and partially describe the fat tails exhibited by financial data time series, two drawbacks can be pointed out. First, they discharge any influence of the innovations' sign in the conditional variance. Instead, they assume that only the innovations' magnitude is relevant, neglecting the leverage effect. The other ARCHER limitation has to do with the parameters non-negativity restrictions, meaning that they can assume any value, even a negative one.

These limitations led some authors to propose new models. Among them, the EGARCH and the GJR models that will be used in this study.

The EGARCH (Exponential GARCH) model (Nelson, 1991) was constructed in a way that a negative shock leads to a higher conditional variance in the subsequent period than a positive shock would. In other words, EGARCH was constructed to account for the leverage effect. For that reason, the conditional variance is specified in logarithmic form, so that the conditional variance depends on both the size and the sign of lagged residuals. That way, there is no need to impose estimation constraint in order to avoid negative variance.

The GJR model was proposed by Glosten et al. (1993) and is very similar to the Threshold GARCH (TGARCH) model (Zakoian, 1994). “GJR allows a quadratic response of volatility to news with different coefficient for good and bad news, but maintains the assertion that the minimum volatility will result when there is no news (Bollerslev et al, 1994: 2970).” In this model, the leverage effect is modeled with a dummy variable that assumes the value 1 to represent a negative shock and 0 otherwise

In the following section, we will present the accuracy tests that will be used to compare these models in our empirical study.

3. A BRIEF DESCRIPTION OF EXISTING TESTS

3.1 Diebold-Mariano (1995)

Consider two forecasts, $\{\hat{y}_{it}\}_{t=1}^T$ and $\{\hat{y}_{jt}\}_{t=1}^T$, of the time series $\{y_t\}_{t=1}^T$. Let the associated forecast errors be $\{e_{it}\}_{t=1}^T$ and $\{e_{jt}\}_{t=1}^T$, respectively. We wish to assess the expected loss associated with each of the forecasts. Thus, let us consider $g(e_{it})$ and $g(e_{jt})$ as the loss functions.

The null hypothesis of Diebold-Mariano test states that there is no difference between two competing forecasts, in terms of their accuracy skill. Equivalently, the null hypothesis states that the population mean of the differential loss is zero ($E[d_t] = 0$, where $d_t \equiv [g(e_{it}) - g(e_{jt})]$ is the loss differential).

In contrast with the previously developed tests, Diebold-Mariano (1995) test allows the loss function to be non-quadratic and asymmetric. Besides, errors can be non-Gaussian, non-zero mean, serially correlated and contemporaneously correlated.

Diebold-Mariano test statistic is the following:

$$S_1 = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}}, \quad (1)$$

where $\hat{f}_d(0)$ is a consistent estimate of $f_d(0)$, the spectral density of the loss differential at frequency 0 and \bar{d} is the sample mean loss differential which, in large samples, is approximately normally distributed with mean μ and variance $2\pi f_d(0)/T$.

Diebold and Mariano evaluated the finite-sample size of several test statistics, under the null hypothesis. Besides S_1 , two finite-sample reference tests were included in this study: the sign test (S_2) and the Wilcoxon's signed-rank test (S_3). The studentized versions of these two tests (S_{2a} and S_{3a} , respectively) were also studied.

$$S_2 = \sum_{t=1}^T I_+(d_t), \quad (2)$$

where $I_+(d_t) = \begin{cases} 1 & \text{if } d_t > 0 \\ 0 & \text{otherwise} \end{cases}$

$$S_{2a} = \frac{S_2 - 0.5T}{\sqrt{0.25T}} \overset{a}{\sim} N(0,1) \quad (3)$$

$$S_3 = \sum_{t=1}^T I_+(d_t) \text{rank}(|d_t|) \quad (4)$$

$$S_{3a} = \frac{S_3 - \frac{T(T+1)}{4}}{\sqrt{\frac{T(T+1)(2T+1)}{24}}} \overset{a}{\sim} N(0,1) \quad (5)$$

Moreover, the F-test, as well as the MGN (Granger and Newbold, 1977) and MR (Meese and Rogoff, 1988) tests were also included in this analysis.

- **F-test**

The F-test requires all previous referred assumptions to be valid. The null hypothesis of equal forecast accuracy corresponds to equal forecast error variances (by 1 and 2a). By the remaining assumptions, the ratio of simple variances has the usual F distribution under the null hypothesis. Distributed as $F(T, T)$, the test statistic is:

$$F = \frac{e_i' e_i}{e_j' e_j} \quad (6)$$

This test statistics has little practical use due to the conditions imposed to the forecast errors which, as referred before, are very difficult to obtain.

- **Morgan-Granger-Newbold (1977)**

In order to solve the contemporaneous correlation problem, Granger and Newbold employed an orthogonalizing transformation due to Morgan (1939-1940), enabling the relaxation of assumption b4. Thus, the Morgan-Granger-Newbold (MGN) test allows forecast errors to be contemporaneously correlated and maintains all the previously referred assumptions. Under this assumptions, the null hypothesis of equal forecast accuracy is equivalent to zero correlation between x and z ($\rho_{xz} = 0$), where

$$x_t = (e_{it} + e_{jt}) \text{ and } z_t = (e_{it} - e_{jt})$$

$$x = (e_i + e_j) \text{ and } z = (e_i - e_j)$$

Distributed as Student's t with T-1 degrees of freedom, MGN test statistic is the following:

$$MGN = \frac{\hat{\rho}_{xz}}{\sqrt{\frac{1-\hat{\rho}_{xz}^2}{T-1}}} \quad (7)$$

- **Meese-Rogoff (1988)**

Like the MGN test, the Meese-Rogoff (MR) test allows forecast errors to be contemporaneously correlated. In addition, forecast errors can also be serially correlated. Under the remaining assumptions (a, b1 and b2), MR test is asymptotically distributed as standard normal and the test statistic is the following:

$$MR = \frac{\hat{Y}_{xz}}{\sqrt{\frac{\hat{\Sigma}}{T}}} \quad (8)$$

where $\hat{Y}_{xz} = x'z/T$ and $\hat{\Sigma}$ is a consistent estimator of Σ .

It is interesting to note that MR can coincide asymptotically with MGN, when the null hypothesis and assumptions a, b1, b2 and b3 are satisfied.

In these tests' evaluation, Diebold and Mariano presented results for different levels of contemporaneous correlation (ρ), serial correlation (θ) and sample size (T). Gaussian and non-Gaussian forecast errors are also distinguished. Results are shown in the appendix (Tables 26-31).

Let us first discuss the case of Gaussian forecast errors. Here are summarized the main conclusions:

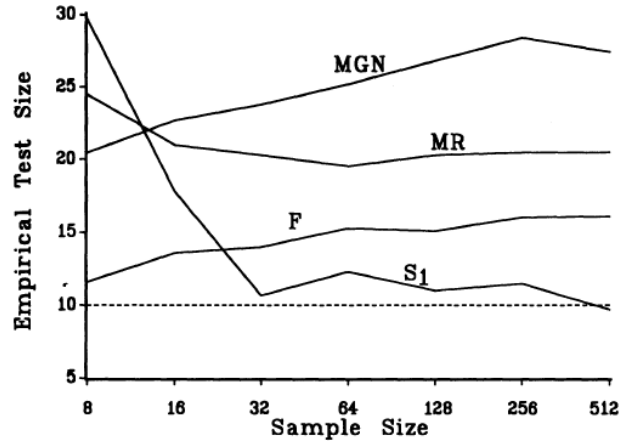
- F is correctly sized when there is no contemporaneous and serial correlation (i.e. $\rho = \theta = 0$) but is missized when any of them is present. Serial correlation pushes empirical size above nominal size, while contemporaneous correlation pushes empirical size severely below nominal size. Clearly, contemporaneous correlation dominates serial correlation. Therefore, in presence of both, F is undersized. This outcome is particularly apparent for large ρ and θ .

- As expected, the MGN test remains correctly sized as long as $\theta = 0$. As we already mentioned, the MGN test allows forecast errors to be contemporaneously correlated. Serial correlation, however, pushes empirical size above nominal size.
- The results obtained for the MR test offer no surprises. As already mentioned, the MR test is robust to both serial and contemporaneous correlation. Indeed, for large samples ($T > 64$), that proved to be truth. Still, in small samples, the MR test showed to be oversized in the presence of serial correlation.
- In large samples, the S_1 test showed to be robust to both contemporaneous and serial correlation. In small samples, however, S_1 is oversized, a behavior particularly similar to the one showed by the MR test, except that the empirical and nominal sizes of the S_1 test converge a bit more slowly, when compared to the MR test.
- Finally, both S_2 and S_3 tests attested to perform well, with nominal and empirical size in close agreement, independently of contemporaneous or serial correlation, as well as sample size. Moreover, this conclusion can be extended to S_{2a} and S_{3a} tests.

In relation to non-Gaussian forecast errors, the most evident result is the drastic missizing of the F, MGN and MR tests, evidence that is common to both small and large samples. S_1 , S_{2a} and S_{3a} , on the other hand, maintain approximately correct size, except for very small sample sizes ($T < 32$). In those cases, S_2 and S_3 continue to perform well.

To summarize, Figure 1 illustrates, in a really perceptible manner, the behavior verified by the F, MGN, MR and S_1 tests, for the non-Gaussian case with $\rho = \theta = 0,5$.

Figure 1: Empirical size, Four Test Statistics; Fat-Tailed case; Theta = Rho = 0,5



3.2 A modified Diebold-Mariano test

As showed before, the original DM test (S_1) performs relatively well for large samples ($T > 32$) but, for small and moderate samples, the test can be quite seriously oversized. Besides, this problem becomes progressively more severe as the forecast horizon grows. So in 1997, Harvey, Leybourne and Newbold proposed some modifications to the original DM test, with the purpose to improve the test's performance for smaller samples. The authors employed an approximately unbiased estimator of the variance of \bar{d} and proved that:

$$E[\hat{V}(\bar{d})] = \left[\frac{n+1-2h+n^{-1}h(h-1)}{-n} \right] V(\bar{d}), \quad (9)$$

where,

$$\bar{d} = n^{-1} \sum_{t=1}^n d_t$$

Therefore, the modified DM test statistic becomes:

$$S_1^* = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n} \right]^{\frac{1}{2}} S_1, \quad (10)$$

where S_1 is the original statistic, n is the number of observations and h is the number of steps-ahead forecasts.

Besides the test statistic, the modified DM test differs from the original one in another way. While the latter uses the critical values of the standard normal distribution, the former uses the critical values of the Student's t distribution with $(n-1)$ degrees of freedom.

In order to conclude about the effects caused by these two modifications, the authors confronted the original Diebold-Mariano test (DM) with the modified Diebold-Mariano test (MDM). The results appear in Table 1. Expected squared errors were taken as the criterion of forecast quality, so that $d_t = e_{1t}^2 - e_{2t}^2$.

Table 1: Percentage of rejections of the true null hypothesis of equal mean squared errors for the original and modified Diebold-Mariano test at nominal 10% level

h	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
1	16.7	13.5	11.6	10.9	10.3	10.6	10.8
	11.0	10.8	10.3	10.4	10.0	10.5	10.7
	13.8	12.0	10.9	10.6	10.1	10.5	10.8
	8.4	9.6	9.7	10.1	9.9	10.4	10.6
2	30.0	20.3	15.1	12.4	11.5	10.9	10.5
	23.9	17.8	13.9	12.0	11.2	10.8	10.5
	21.1	16.4	13.2	11.7	11.1	10.6	10.4
	16.4	14.2	12.2	11.2	10.8	10.5	10.3
3	36.9	26.5	18.3	14.1	11.7	11.3	11.2
	30.9	24.1	17.0	13.7	11.5	11.2	11.1
	22.3	20.4	15.1	12.8	11.0	11.0	10.9
	18.1	18.5	14.3	12.2	10.7	10.8	10.9
4	43.2	30.9	21.3	15.9	13.0	11.6	11.2
	37.4	28.3	20.0	15.4	12.7	11.5	11.1
	20.9	22.1	17.2	13.9	11.8	11.1	11.0
	16.3	19.8	16.1	13.4	11.5	10.9	11.0
5	49.4	34.5	24.5	18.0	13.8	11.9	11.4
	43.5	31.8	23.4	17.4	13.5	11.7	11.3
	16.6	22.1	18.9	15.4	12.4	11.2	11.1
	12.9	19.9	17.8	14.9	12.2	11.1	11.0
6	58.4	37.3	26.7	19.6	14.8	12.2	11.8
	52.7	34.8	25.6	19.1	14.5	12.1	11.7
	13.5	21.8	19.6	16.4	13.2	11.5	11.3
	10.6	19.8	18.8	16.0	12.9	11.4	11.2
7	72.4	39.4	28.8	20.8	15.7	12.7	12.0
	68.3	36.9	27.5	20.4	15.4	12.5	11.9
	12.5	20.4	20.5	17.3	13.8	11.6	11.5
	9.9	18.2	19.5	16.8	13.6	11.6	11.4
8	–	42.6	30.8	22.9	16.3	13.1	12.0
	–	39.8	29.7	22.4	16.0	12.9	11.9
	–	19.2	21.0	18.5	14.2	12.0	11.5
	–	17.4	20.2	18.0	13.8	11.9	11.4
9	–	45.3	32.4	24.5	17.5	13.8	12.2
	–	42.7	31.3	23.9	17.2	13.8	12.2
	–	16.9	21.2	19.5	15.0	12.5	11.6
	–	15.1	20.2	19.0	14.7	12.4	11.6
10	–	49.0	33.4	25.3	17.9	14.2	12.4
	–	46.4	32.3	24.9	17.6	14.1	12.3
	–	15.5	21.1	19.6	15.4	12.7	11.8
	–	14.0	20.2	19.1	15.1	12.6	11.8

Note: The first entry in each cell is for the original test using $N(0, 1)$ critical values, the second for the original test using t_{n-1} critical values, the third for the modified test using $N(0, 1)$ critical values, and the fourth for the modified test using t_{n-1} critical values.

Source: (Harvey et al., 1997:285)

- As pointed out by Diebold and Mariano, their test has showed to be oversized for the case $h=2$. As we can see in Table 1, this problem is extended to longer horizons. Moreover, it becomes increasingly harsh as we move forward into the forecast horizon. The performance achieved by the modified test proved to be significantly better. Nevertheless, the high performance verified for the smallest samples can be seen as casual since that, for longer forecast horizons, the size appears to deteriorate with increasing n before improving again.
- When compared to the original DM test, the empirical and nominal sizes of the MDM test demonstrated to be closer to each other, especially for the smallest samples. This phenomenon also verifies when the errors are autocorrelated and there is contemporaneous correlation.
- The results obtained when using contemporaneous correlation or serial correlation coefficients of 0.5 and 0.9 were a lot similar to those presented in Table 1. Furthermore, the authors discovered that the empirical and nominal sizes tended to move closer with increasing θ .
- The most important conclusion to take is that both improvements showed to improve DM test's performance, although the modification made to the test statistic proved to be the more effective one.

3.3 A modified Morgan-Granger-Newbold test

As pointed out by Diebold and Mariano, in relation to the MGN test, for errors from a Student's t with 6 degrees of freedom (t_6) generating process, the serial correlation pushes the empirical size above the nominal size, meaning the test is oversized. Besides, this problem gets deeper as the sample size increases.

Although Diebold and Mariano have identified this phenomenon, they weren't able to find any explanation. Later, in 1997, Harvey, Leybourne and Newbold concluded that the size distortions verified for the MGN test had to do with the inconsistent estimation of one parameter (D), caused by the non-verification of the assumption $E(\varepsilon_t^2|x_t) = \sigma^2$, in which the usual regression test on $\hat{\beta}$ is based.

$$y_t = \beta x_t + \varepsilon_t, \quad (11)$$

where

$$y_t = e_{1t} + e_{2t} \quad ; \quad x_t = e_{1t} - e_{2t} \quad ; \quad t = 1, 2, \dots, n$$

Under the conditions of theorem 5.3 of White (1984, p.109), that can be applied in this situation,

$$n^{1/2}D^{-\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{d} N(0,1), \quad (12)$$

where

$$D = M^{-2}Q \quad ; \quad Q = V(n^{-1/2} \sum x_t \varepsilon_t) \quad ; \quad M = E(x_t^2)$$

Since $E(\varepsilon_t^2|x_t) = \sigma^2$, that implies $Q = \sigma^2 M$ and, consequently, $D = \sigma^2 M$. Given that $E(\varepsilon_t^2|x_t) = \sigma^2$ does not hold, parameter D is inconsistently estimated. Therefore, in order to solve this problem, the authors employed a consistent estimator of D and proposed a modified MGN test, with the following test statistic:

$$S_2^* = \left[\frac{\sum x_t^2 \hat{\varepsilon}_t^2}{(\sum x_t^2)^2} \right]^{-1/2} \hat{\beta} \quad (13)$$

Although the null distributional result is no longer exact, the authors recommend the comparison this test statistic with critical values from the Student's t distribution with $(n-1)$ degrees of freedom. The comparison between the original and the modified MGN tests is presented in Table 2.

Table 2: Percentage of rejections of the true null hypothesis of equal one-step prediction mean squared errors for the original and modified Morgan-Granger-Newbold tests at nominal 10% level

Normal	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0$	10.2	10.3	10.0	10.1	9.7	10.4	10.7
	19.9	16.0	12.8	11.8	10.6	10.9	10.8
$\rho = 0.5$	10.0	9.9	10.3	10.3	10.2	10.6	10.4
	19.8	15.4	13.0	12.0	11.0	11.2	10.7
$\rho = 0.9$	10.1	9.9	10.2	10.4	10.0	10.5	10.1
	20.4	15.5	12.9	12.0	10.7	11.1	10.6
t_6	$n = 8$	$n = 16$	$n = 32$	$n = 64$	$n = 128$	$n = 256$	$n = 512$
$\rho = 0$	17.9	20.5	22.6	24.8	26.0	26.1	26.8
	26.3	21.2	18.0	15.7	12.9	12.0	10.7
$\rho = 0.5$	16.2	18.5	19.8	22.1	22.4	22.6	23.7
	25.4	20.8	17.5	14.9	12.5	11.4	10.6
$\rho = 0.9$	11.8	12.8	12.6	13.7	13.6	13.7	13.6
	22.0	18.5	15.1	13.3	12.0	10.8	10.3

Note: The first entry in each cell is for the original test, and the second for the modified test.

Source: (Harvey et al., 1997:288)

Let us consider the original test in first place. Up to sampling error, the empirical size matches the nominal size. This result brings no surprises, since the null distribution is known in the normal case. However, for the t_6 error-generating process, the test is seriously over-sized. Moreover, this problem gets deeper as the sample size increases. However, it seems that the excess size problem becomes less severe with increasing contemporaneous correlation between the forecast errors.

Concerning to the modified test, the conclusions are quite contradictory. As predicted by the theory, the modification created a test with the correct size for large samples. Nonetheless, performance in small samples, where the modified test is over-sized, is poor. In fact, in the smallest samples, the modified test is seriously over-sized when the error distribution is normal, and even worse than the original test for the t_6 error-generating process. For that reason, the authors recommend the use of the modified MGN test only for moderately large samples.

Indeed, for the smallest sample sizes, the original MGN test proved to be more powerful than the modified one. Still, as the sample size increases, that advantage vanishes quickly.

Since the test based on the modified statistic performed poorly in small samples, the authors considered also the possibility of a non-parametric approach. Specifically, the authors employed the Spearman's rank correlation test because of their difficulty in handling non-normality. In Table 3, we can see a comparison between the MDM test, the rank correlation variant of the MGN test and the original version of the MGN test. The t_6 generating process was considered for the former two tests.

Table 3: Percentage of rejections of the false null hypothesis of equal one-step prediction mean squared errors for the MDM, the rank correlation variant of MGN and the original MGN tests at 10% level

Normal	$n = 8$ $R = 3$	$n = 16$ $R = 2$	$n = 32$ $R = 1.5$	$n = 64$ $R = 1.375$	$n = 128$ $R = 1.25$	$n = 256$ $R = 1.1875$	$n = 512$ $R = 1.125$
$\rho = 0$	27.1	33.1	28.6	34.4	34.9	40.6	38.0
	28.7	33.2	27.0	32.7	32.6	37.8	35.7
	42.8	38.7	30.7	35.5	34.9	40.9	38.1
$\rho = 0.5$	32.2	39.2	33.9	41.9	42.1	48.0	46.0
	34.3	39.2	32.6	39.3	40.3	45.4	43.2
	50.9	46.4	35.8	43.2	42.7	48.5	46.1
$\rho = 0.9$	59.9	80.1	77.4	86.7	88.3	92.5	92.1
	71.0	81.3	75.0	84.3	86.1	90.3	89.6
	89.6	89.2	81.2	87.9	88.9	92.8	91.9
t_6	$n = 8$ $R = 3$	$n = 16$ $R = 2$	$n = 32$ $R = 1.5$	$n = 64$ $R = 1.375$	$n = 128$ $R = 1.25$	$n = 256$ $R = 1.1875$	$n = 512$ $R = 1.125$
$\rho = 0$	19.1	23.0	20.3	24.2	23.5	25.0	23.1
	27.6	30.9	26.5	30.9	31.0	34.4	32.7
$\rho = 0.5$	25.7	31.1	27.0	31.4	29.7	32.3	30.5
	33.6	37.9	32.4	37.9	37.2	41.5	40.5
$\rho = 0.9$	53.3	77.0	77.2	86.8	86.9	90.9	89.4
	71.6	83.5	79.0	87.2	88.2	92.7	91.6

Note: The first entry in each cell is for the modified Diebold–Mariano test, the second is for the rank correlation variant of the Morgan–Granger–Newbold test, and the third is for the original Morgan–Granger–Newbold test (normal errors only).

Source: (Harvey et al., 1997:290)

Through the analysis of Table 3, Harvey, Leybourne and Newbold came up with the following conclusions:

- For the smallest sample sizes, the MGN test showed to be more powerful than the MDM test. Still, as the sample size increases, that advantage disappears.
- For normal forecast errors, the performance verified by the rank correlation variant of the MGN test is pretty much identical with the performance of the MDM test, as well as the one demonstrated by the original MGN test for large samples. In small samples, however, most of the advantage of the MGN test is lost when ranks are employed.
- In the case of heavy-tailed error distributions, the rank correlation test proved to be rather more powerful than the MDM test.

In conclusion, Harvey, Leybourne and Newbold recommend the use of the modified DM test, in part because of the lack of robustness of the MGN test in the presence of heavily-tailed distributions of the forecast errors. Besides, although the test based on rank correlations performs reasonably well, particularly for heavily-tailed error distributions, it is difficult to see how this test could be extended to deal with forecasts beyond one-step ahead.

3.4 Harvey, Leybourne, Newbold (1998)

When evaluating a forecast, it is often the case that several competing predictors are available. Harvey, Leybourne and Newbold investigated the issue of testing for forecast encompassing when there are two forecasts of the same quantity. The question to ask in these circumstances is whether any of the competing forecasts can add useful information not present in the superior forecast. If an inferior forecast contains no useful information, not present in the superior forecast, we say that the second forecast encompasses the first. In other words, “if a composite predictor formed as a weighted average of two individual forecasts is considered, then one forecast is said to encompass (or be conditionally efficient with respect to) the other if the inferior forecast’s optimal weight in the composite predictor is zero (Harvey and Newbold, 2000: 471)”.

Let (f_{1t}, f_{2t}) be two competing forecasts of the quantity y_t . One-step-ahead predictions were assumed ($h=1$). Let f_{ct} be the combined forecast, formed as a weighted average of the two individual forecasts,

$$f_{ct} = (1 - \lambda)f_{1t} + \lambda f_{2t}, \quad 0 \leq \lambda \leq 1 \quad (14)$$

Then, if $e_{it} = (y_t - f_{it})$, $i = 1, 2$, denote the errors of the individual forecasts and ε_t is the error of the combined forecast, we have:

$$e_{1t} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t \quad (15)$$

In order to evaluate whether f_{2t} contains useful information not present in f_{1t} , Granger and Newbold (1973, 1986) proposed the estimation of regression (15) by ordinary least squares. The null hypothesis is $\lambda = 0$.

When using a record of past forecast errors $(e_{1t}, e_{2t}), t = 1, \dots, n$, it is natural to test for forecast encompassing through a simple least squares regression, which might be expected to perform well when forecast errors are generated by a bivariate normal distribution. However, in accordance to Harvey *et al* (1998), it was reasonable to suspect that forecast error distributions will often be heavily-tailed. In fact, they showed that, when forecast errors are normally distributed, we can stumble on far too many

rejections of a true null hypothesis of forecast encompassing. Consequently, they ran a simulation to evaluate the behavior of the standard regression-based test for forecast encompassing in finite samples. Results can be seen in Table 4.

Table 4: Empirical Sizes of Nominal 5%-level and 10%-level Regression-Based Tests for Forecast Encompassing

<i>n</i>	<i>5%-level</i>		<i>10%-level</i>	
	<i>t₆ errors</i>	<i>t₅ errors</i>	<i>t₆ errors</i>	<i>t₅ errors</i>
8	7.5	8.9	13.1	15.0
16	8.5	10.0	14.0	16.1
32	9.7	10.6	15.3	16.4
64	10.3	11.5	16.1	17.5
128	10.5	12.3	16.5	18.3
256	11.1	12.8	17.0	18.6

Source: (Harvey et al., 1998:256)

As can be noted, the oversizing problem becomes more severe as the number of sample observations increases, with very slow convergence to the asymptotic results found by the authors (0,122 and 0,171 for 5%-level; 0,182 and 0,230 for 10%-level).

Harvey et al. (1998) showed that the standard test for forecast encompassing can be incorrectly sized when the forecast errors are a temporally independent sequence but not normally distributed. “The non-normality problems associated with the standard regression-based test for multiple forecast encompassing can be shown to result from inconsistent estimation of the quantity *Q*, induced by conditional heteroskedasticity in the regression errors” (Harvey and Newbold, 2000: 473). Then, the most obvious modification in this situation is to employ the heteroskedasticity-robust estimator of White (1980), a procedure that can be extended to dependent error sequences.

Given this problem of lack of robustness to non-normality in the standard test,

$$R = \widehat{D}^{-1/2} n^{1/2} \hat{\lambda}, \tag{16}$$

the authors proposed the following test that proved to be robust under these circumstances:

$$R_1 = n^{-1/2} \hat{Q}_1^{-1/2} \sum (e_{1t} - e_{2t})^2 \hat{\lambda} = n^{1/2} \hat{Q}_1^{-1/2} \bar{d} \quad (17)$$

where \bar{d} is the sample mean of the sequence $d_t = (e_{1t} - e_{2t})e_{1t}$

Although \hat{Q}_1 is a consistent estimator for $Q = \text{var}[n^{-1/2} \sum (e_{1t} - e_{2t})\varepsilon_t]$, convergence of the second term to zero is likely to be slow. This being the case, an alternative option is to replace the estimator \hat{Q}_1 by the estimator \hat{Q}_2 . Consequently, the authors proposed a new test statistic:

$$R_2 = n^{-1/2} \hat{Q}_2^{-1/2} \sum (e_{1t} - e_{2t})^2 \hat{\lambda} = n^{1/2} \hat{Q}_2^{-1/2} \bar{d} \quad (18)$$

In both tests, the null hypothesis to be tested is of zero correlation between e_{1t} and $(e_{1t} - e_{2t})$.

Harvey *et al* (1998) made a simulation experiment to evaluate finite sample sizes of R_1 and R_2 tests, along with the original DM test and the modified one proposed by Harvey *et al* (1997). For the standard DM test, normal critical values were used. For the other three tests, however, the authors used t_{n-1} critical values, given the results of Harvey *et al* (1997). Results appear in Table 5.

Table 5: Empirical Sizes of Nominal 5%-level and 10%-level Modified Regression-Based Tests and Diebold-Mariano-type Tests for Forecast Encompassing ($h=1$)

<i>n</i>	<i>Test</i>	<i>5%-level</i>			<i>10%-level</i>		
		<i>N errors</i>	<i>t₆ errors</i>	<i>t₅ errors</i>	<i>N errors</i>	<i>t₆ errors</i>	<i>t₅ errors</i>
8	<i>R</i> ₁	10.1	12.0	13.5	15.8	18.1	19.6
	<i>R</i> ₂	1.6	1.1	1.1	8.5	7.1	7.6
	DM	8.4	7.1	7.4	14.6	13.8	14.5
	MDM	4.4	3.3	3.4	10.2	9.0	9.5
16	<i>R</i> ₁	8.0	9.9	11.4	13.1	15.6	17.1
	<i>R</i> ₂	3.6	3.2	3.0	9.8	9.6	9.6
	DM	6.5	6.2	6.1	12.4	12.3	12.5
	MDM	4.9	4.3	4.3	10.5	10.2	10.4
32	<i>R</i> ₁	6.2	8.9	9.2	11.5	14.5	14.9
	<i>R</i> ₂	4.3	4.3	3.7	9.7	10.6	10.1
	DM	5.4	5.7	5.3	10.8	11.9	11.5
	MDM	4.8	4.8	4.3	9.9	10.9	10.4
64	<i>R</i> ₁	6.0	7.5	7.7	11.3	13.0	13.3
	<i>R</i> ₂	4.9	4.5	4.2	10.5	10.4	10.2
	DM	5.5	5.3	4.9	10.9	11.0	10.7
	MDM	5.1	4.8	4.5	10.7	10.5	10.3
128	<i>R</i> ₁	5.7	6.5	6.8	10.7	12.1	12.1
	<i>R</i> ₂	5.0	4.7	4.4	10.4	10.3	10.2
	DM	5.4	5.0	4.9	10.6	10.6	10.5
	MDM	5.2	4.8	4.6	10.4	10.4	10.3
256	<i>R</i> ₁	5.5	5.9	6.1	10.6	11.3	11.3
	<i>R</i> ₂	5.1	4.9	4.6	10.3	10.4	9.9
	DM	5.3	5.1	4.7	10.4	10.6	10.0
	MDM	5.2	5.0	4.6	10.4	10.4	9.9

Source: (Harvey et al., 1998:257)

As predicted by theory, all four tests have approximately correct sizes in large samples. In small samples, however, the empirical and nominal sizes do not match, even when the forecast-error distribution is bivariate normal. A missizing can be noticed in R_2 and MDM tests, whereas R_1 and DM tests demonstrate to be oversized. In general, the MDM seems to perform better than its competitors and represents a distinct improvement on the standard regression-based test of the previous table when the generating process is bivariate Student's t .

Given the general satisfactory size performance of the MDM test in the case $h=1$, the authors decided to investigate the test's behavior for higher values of h (the steps-ahead forecast). The original DM test was also included in this simulation. Results are shown in Table 6.

Table 6: Empirical Sizes of Nominal 5%-level Diebold-Mariano-type Tests for Forecast Encompassing: Multistep-ahead Prediction (normal errors)

n	Test	$h = 2$	$h = 4$	$h = 6$	$h = 8$
8	DM	12.6	14.3	15.5	—
	MDM	6.4	4.8	2.6	—
16	DM	10.0	11.9	12.5	12.7
	MDM	6.9	7.0	5.8	4.8
32	DM	7.3	10.0	11.0	11.6
	MDM	5.8	7.7	7.5	7.2
64	DM	6.5	8.2	9.7	10.5
	MDM	5.7	6.9	7.8	8.0
128	DM	5.6	6.5	7.4	8.5
	MDM	5.2	6.0	6.6	7.0
256	DM	5.4	5.9	6.2	6.5
	MDM	5.2	5.4	5.8	6.1

Source: (Harvey et al., 1998:258)

As we can see in Table 6, the empirical sizes of the MDM test are really close to the nominal sizes, except for long forecast horizons in small samples. Moreover, the MDM test is generally clearly preferable to the DM test in terms of size properties, suggesting that the Harvey *et al* (1997) modifications are well worth making.

The authors proceeded with further simulations, in order to verify if the power of the previous conclusions remain valid under a more general context. Specifically, for forecast horizons up to four, they also generated forecast errors from bivariate t_5 and t_6 distributions. The results found were relatively close from those for the case of normal errors. Therefore, we can confirm the robustness of the null distribution of the test statistics under heavy-tailed error distributions.

Finally, Harvey, Leybourne and Newbold made a power comparison of four tests for forecast encompassing. Those tests are the standard regression-based test R , the modified regression-based test R_1 , the MDM test and the Spearman's rank correlation test (r_s). Independent sequences of forecast errors (e_{1t}, e_{2t}) were generated from bivariate normal and bivariate t_6 distributions and one-step-ahead forecasts were taken. We can see the estimated size-adjusted powers for the four tests in Table 7

Table 7: Estimated Size-Adjusted Powers of 5%-level Tests for Forecast Encompassing ($h=1$)

n	Test	High power		Moderate power	
		N errors	t_6 errors	N errors	t_6 errors
8 $k_1 = 1$ $k_2 = 2$	R	74.8	67.0	34.2	30.0
	R_1	69.3	59.1	30.9	26.9
	MDM	51.5	51.0	25.2	25.3
	r_s	62.1	54.4	29.3	24.2
16 $k_1 = 1.5$ $k_2 = 3$	R	78.3	66.1	35.4	27.9
	R_1	75.3	61.8	33.0	25.7
	MDM	68.2	59.8	29.6	26.6
	r_s	69.5	62.8	30.4	25.6
32 $k_1 = 2.25$ $k_2 = 4.25$	R	78.5	60.1	36.6	25.8
	R_1	75.5	60.4	35.5	27.5
	MDM	74.1	61.4	35.0	28.2
	r_s	71.6	66.3	33.2	30.0
64 $k_1 = 3.25$ $k_2 = 6.25$	R	76.7	59.8	34.7	25.4
	R_1	76.3	59.2	34.9	26.0
	MDM	76.0	61.5	34.2	26.8
	r_s	72.6	69.0	32.7	29.7
128 $k_1 = 4.75$ $k_2 = 9$	R	75.8	56.0	35.2	24.7
	R_1	75.3	57.4	34.8	26.4
	MDM	75.2	59.2	34.7	26.7
	r_s	71.9	67.9	32.8	31.0
256 $k_1 = 6.75$ $k_2 = 12.75$	R	76.2	54.0	34.1	23.6
	R_1	76.2	56.2	34.5	25.2
	MDM	76.2	57.6	34.5	25.8
	r_s	72.6	66.4	32.6	30.1

Source: (Harvey et al., 1998:258)

As expected, the most considerable differences among the tests occur at the smallest sample sizes, where R is clearly the most powerful of the four tests, for both normal and t_6 errors. For samples inferior to 32, R_1 is somewhat less powerful than R but more powerful than r_s and manifestly more powerful than MDM. For sample sizes larger than 32, however, there is relatively little to choose among R, R_1 and MDM. The poor power performance of MDM test in small samples - particularly for normally distributed errors - is quite regrettable, since its normal sizes are the more reliable ones. To make it worse, the nominal significance levels of R_1 proved to be unreliable in these sample sizes (see Table 5) and the nominal significance levels of R cannot be trusted for any sample size (see Table 4). Therefore, the authors recommend the use of MDM over R and R_1 tests. The same reason can be employed to choose MDM over R_2 , despite the identical size-adjusted power.

In relation to the correlation test, we can verify that it is slightly over performed by the other three tests, when we have a normal error distribution. When the error distribution is bivariate t_6 , however, the test performs relatively well, especially for large samples, where it beats all the other three tests. Hence, in the case in which the forecast errors are an independent sequence, the rank correlation test is certainly a viable alternative to MDM.

3.5 Harvey and Newbold (2000)

In 2000, Harvey and Newbold generalized the forecast encompassing approach to situations where there is more than one competing forecast to compare.

Let (f_{1t}, \dots, f_{Kt}) be K competing forecasts, taken to be unbiased or bias-corrected, of the actual quantity A_t . Assume that the forecasts are made one-step-ahead, with non-autocorrelated errors. f_{ct} is the composite forecast.

$$f_{ct} = (1 - \lambda_1 - \lambda_2 - \dots - \lambda_{K-1})f_{1t} + \lambda_1 f_{2t} + \lambda_2 f_{3t} + \dots + \lambda_{K-1} f_{Kt}, \quad (19)$$

Equivalently,

$$e_{1t} = \lambda_1(e_{1t} - e_{2t}) + \lambda_2(e_{1t} - e_{3t}) + \dots + \lambda_{K-1}(e_{1t} - e_{Kt}) + \varepsilon_t, \quad (20)$$

where $0 \leq \lambda_i \leq 1$, $e_{it} = A_t - f_{it}$ and ε_t is the error of the combined forecast. The null hypothesis that f_{1t} encompasses its competitors (f_2, \dots, f_K) is

$$\lambda_1 = \lambda_2 = \dots = \lambda_{K-1} = 0 \quad (21)$$

The regression-based test for multiple forecast encompassing is an F-test of the joint significance of the parameters in equation (20).

It is convenient to note that regression (20) can be written as

$$y_t = X_t' \beta + \varepsilon_t, \quad (21)$$

where

$$y_t = e_{1t},$$

$$\beta = [\lambda_1 \lambda_2 \dots \lambda_{K-1}]',$$

$$X_t = [(e_{1t} - e_{2t})(e_{1t} - e_{3t}) \dots (e_{1t} - e_{Kt})]'$$

As this is a generalization of the two-forecast regression-based test studied by Harvey *et al* (1998), we expect to deal with the same problems, specifically, the lack of robustness to non-normality in the standard test, caused by conditional heteroskedasticity in the regression errors. Then, having this problem in mind, Harvey and Newbold presented three modified tests that proved to be robust to conditional heteroskedasticity in the regression errors. These tests also allow for forecast error autocorrelation, permitting comparison of forecasts made at horizons greater than one.

- **Modified regression-based tests**

For the first two tests (F_1 and F_2), the authors followed the approach of the two-forecast case applied by Harvey *et al* (1998). Specifically, they employed the heteroskedasticity-robust estimator of White (1980), as well as a robust estimator which is consistent under the null, but not the alternative hypothesis.

In addition, they followed the proposition of Diebold and Mariano (1995) and Harvey et al (1998) that, for h-steps-ahead forecasts, a rectangular kernel with bandwidth (h-1) should be adopted to account for forecast error autocorrelation. The multivariate test statistics proposed by the authors is the following:

$$F_{\alpha} = (K - 1)^{-1}n\hat{\beta}'\hat{D}_{\alpha}^{-1}\hat{\beta} , \quad \alpha = 1,2 \quad (22)$$

where $\hat{D}_{\alpha} = \hat{M}^{-1}\hat{Q}_{\alpha}\hat{M}^{-1}$, $\hat{M} = n^{-1}X'X$ and \hat{Q}_{α} have (i,j) elements

$$\hat{q}_{\alpha,ij} = n^{-1} \left[\sum_{t=1}^n x_{it}x_{jt}u_{\alpha t}^2 + \sum_{m=1}^{h-1} \sum_{t=m+1}^n x_{it}x_{j,t-m}u_{\alpha t}u_{\alpha,t-m} + \sum_{m=1}^{h-1} \sum_{t=m+1}^n x_{i,t-m}x_{jt}u_{\alpha t}u_{\alpha,t-m} \right]$$

with $u_{1t} = \hat{\varepsilon}_t$ and $u_{2t} = y_t$. Harvey and Newbold compare F_{α} with the critical values from $F_{K-1,n-K+1}$ distribution.

- **Modified Diebold-Mariano test**

Proposed by Harvey et al (1997), this test can be seen as an extension of the MDM test, being this the reason why it was named as the modified Diebold-Mariano-type test (MS*).

As showed before, in 1997, Harvey, Leybourne and Newbold proposed some modifications to the Diebold-Mariano test (1995), modifications that proved to be worth making. Later, in 1998, these authors found that the MDM test could be used to test for forecast encompassing, for the two competing forecasts case. Then, Harvey and Newbold (2000) generalized the test for multiple forecast encompassing. The new test statistic takes the form of Hotelling's (1931) generalized T^2 -statistic.

$$MS^* = (K - 1)^{-1}(n - 1)^{-1}(n - K + 1)\bar{d}'\hat{V}^{-1}\bar{d} \quad (23)$$

where

$$\bar{d} = [\bar{d}_1 \bar{d}_2 \dots \bar{d}_{K-1}]',$$

$$\bar{d}_i = n^{-1} \sum d_{it}$$

and \hat{V} is the sample covariance matrix. Although the finite sample result is not exact, Harvey and Newbold maintained the use of $F_{K-1, n-K+1}$ critical values.

Harvey and Newbold ran a simulation study to evaluate the finite sample behavior of the multiple forecast encompassing tests. Their results are displayed in Table 8. Forecast errors were generated from the multivariate normal distribution and the multivariate Student's t-distribution with five and six degrees of freedom. Empirical sizes were calculated for nominal 5%-level and 10%-level tests. $K=3$ was used, meaning that they were testing whether one forecast encompasses two rival predictors.

Table 8: Empirical Sizes of Nominal 5%-level and 10%-level tests for $K=3, h=1$

n	Test	5%-level			10%-level		
		N errors	t_6 errors	t_5 errors	N errors	t_6 errors	t_5 errors
8	F	4.7	9.2	10.8	9.6	16.8	18.2
	F_1	22.1	30.0	31.6	31.8	39.8	41.0
	F_2	0.0	0.0	0.0	0.0	0.0	0.0
	MS^*	2.2	1.7	1.5	6.3	5.0	4.6
16	F	4.8	11.6	14.1	9.4	19.4	22.0
	F_1	14.5	21.6	23.5	22.1	30.1	32.4
	F_2	0.4	0.2	0.3	3.3	2.7	2.6
	MS^*	3.0	2.4	2.3	7.8	6.9	6.5
32	F	4.9	14.7	18.0	9.8	22.5	26.3
	F_1	10.0	15.8	17.9	16.3	23.9	25.3
	F_2	2.5	1.6	1.4	7.0	6.0	5.6
	MS^*	4.2	3.1	2.8	9.1	8.1	7.9
64	F	5.2	16.6	20.0	10.1	24.3	28.9
	F_1	7.8	11.9	12.7	13.3	18.7	20.0
	F_2	4.0	2.6	2.4	8.9	7.5	7.3
	MS^*	5.0	3.5	3.3	9.7	8.5	8.3
128	F	5.0	18.2	22.6	10.1	26.5	31.4
	F_1	6.2	9.0	9.6	11.8	15.5	16.4
	F_2	4.2	3.5	3.2	9.3	8.8	8.1
	MS^*	4.6	4.0	3.5	9.9	9.3	8.5
256	F	4.8	18.7	24.2	10.1	26.9	33.3
	F_1	5.5	7.6	8.1	11.1	13.1	14.4
	F_2	4.5	4.0	3.8	9.9	9.1	8.6
	MS^*	4.7	4.2	4.0	10.1	9.4	8.8
512	F	5.3	20.2	26.2	10.0	29.1	35.2
	F_1	5.5	6.4	6.8	10.4	12.1	12.3
	F_2	5.0	4.5	3.9	9.9	9.5	8.9
	MS^*	5.1	4.6	4.0	10.0	9.7	9.0
10,000	F	4.9	22.5	32.0	9.8	31.8	41.7
	F_1	5.0	5.2	4.9	9.9	10.3	9.9
	F_2	4.9	5.1	4.6	9.9	10.0	9.3
	MS^*	5.0	5.1	4.6	9.9	10.0	9.3

Source: (Harvey and Newbold, 2000:476)

As expected in theoretical grounds, for non-normal forecast errors, the F-test is incorrectly sized. Although it is common to both small and large samples, this over-sizing problem is more severe in the largest samples.

Relating to the other three tests, we can see that they have approximately correct size for the largest samples. For small and moderate samples however, we observe a drastic size distortion of the two modified regression-based tests. Specifically, the F_1 test is over-sized and the F_2 test is under-sized. In its turn, the MS^* test has shown to be robust, although some under-sizing can be noticed in the smallest samples. These evidences are valid to both normal and non-normal errors.

Given the general satisfactory performance of the MS^* test for one-step-ahead prediction, the authors investigated the test's size properties when using forecasts made at horizons greater than one. For $h=2$, forecast errors e_{it} were generated from MA(1) processes $e_{it} = \varepsilon_{it} + \theta\varepsilon_{i,t-1}$, while for larger h these errors were generated as white noise. As we can see in Table 9, the results are not as reliable as for one step-ahead prediction. Specifically, we can note some over-sizing for small samples and larger horizons.

Table 9: Empirical sizes of nominal 5%-level and 10%-level MS^* tests for $K=3$ (normal errors)

n	5%-level			10%-level		
	$h=2, \theta=0$	$h=2, \theta=0.5$	$h=2, \theta=0.9$	$h=2, \theta=0$	$h=2, \theta=0.5$	$h=2, \theta=0.9$
8	9.6	10.0	10.2	14.2	15.3	15.7
16	12.4	10.6	9.6	18.2	17.0	15.8
32	8.9	7.5	6.9	15.2	13.1	12.9
64	6.8	5.7	5.5	12.3	11.2	11.2
128	5.7	5.6	5.6	11.2	10.8	10.8
256	5.1	5.2	5.1	10.2	10.4	10.2
512	5.1	4.9	4.7	10.3	10.2	10.0

n	5%-level			10%-level		
	$h=4$	$h=6$	$h=8$	$h=4$	$h=6$	$h=8$
8	4.9	1.8	–	7.2	2.5	–
16	11.6	7.4	5.2	15.3	10.1	6.9
32	15.1	12.7	10.5	21.0	16.8	14.2
64	12.8	14.1	13.7	18.5	19.4	18.5
128	8.2	10.2	11.8	13.5	15.7	17.5
256	6.5	7.5	8.7	11.5	13.1	14.4
512	5.9	6.5	7.0	11.0	11.6	12.3

Notes: In the upper panel of the table, autocorrelated forecast errors were generated with MA(1) structure and moving average parameter θ . In the lower panel of the table, white noise forecast errors were generated.

Source: (Harvey and Newbold, 2000:477)

Now, we will see the estimated size-adjusted powers of each one of the tests for $K=3$, which are displayed in Table 10. The authors considered one-step-ahead forecasts with errors drawn from the multivariate normal and Student's t_6 -distributions. Since MS^* is a monotonic function of F_2 , these two tests have identical size-adjusted powers (see Harvey and Newbold, 2000).

Table 10: Estimated Size-adjusted Powers of 5%-level tests for $K=3$, $h=1$

n	Population R^2	Test	N errors	t_6 errors
8	0.66	F	71.1	59.9
		F_1	54.9	40.5
		F_2, MS^*	23.4	21.7
16	0.40	F	71.6	55.2
		F_1	61.8	45.4
		F_2, MS^*	44.6	37.3
32	0.22	F	71.1	49.5
		F_1	64.4	47.1
		F_2, MS^*	58.6	46.0
64	0.12	F	71.9	49.1
		F_1	67.6	49.1
		F_2, MS^*	65.4	52.2
128	0.06	F	71.8	44.0
		F_1	70.7	48.4
		F_2, MS^*	69.9	50.9
256	0.031	F	71.7	44.8
		F_1	70.6	49.2
		F_2, MS^*	70.5	51.7
512	0.016	F	72.3	43.4
		F_1	72.5	48.7
		F_2, MS^*	72.3	51.0

Source: (Harvey and Newbold, 2000:479)

Through the analysis of Table 10 we can observe that, for large samples, the size-adjusted powers of the four tests are pretty much identical. For small and moderate samples however, the F -test displays the greater size-adjusted power. Given its superior power properties, the authors feel disappointed with the fact that the F -test is not robust to non-normality. Right after the F -test, we have the F_1 -test, followed by the MS^* test and the F_2 -test. As we can see, the differences between the tests get smaller as the sample size increase. Also, they are more evident for normal errors than t_6 errors. Finally, with no surprises, the tests exhibit higher power for normal errors.

After analyzing all the results, the authors recommended the use of the MS^* test for moderately large samples, due to its good size and reasonable power properties. Still, they warn about this test limitations, when dealing with small samples: under-sizing for one-step-ahead forecasts, over-sizing for multi-step-ahead evaluation and low power.

3.6 Peter Hansen (2005)

In 2005, Hansen proposed a test for Superior Predictive Ability (SPA). When testing for SPA, the question of interest is whether any alternative forecast is better than the benchmark forecast. Testing for SPA is useful for a forecaster who wants to explore whether a better forecasting model is available, compared to the model currently being used to make predictions.

In contrast with the Equal Predictive Ability (EPA) tests, SPA tests were developed to compare more than two competing forecasts. “The distinction is important because the former leads to a simple null hypothesis, whereas the latter leads to a composite hypothesis (Hansen, 2005: 366).”

One of the main complications in composite hypotheses testing is that (asymptotic) distributions typically depend on nuisance parameters. The usual way to handle this problem is to use the *least favorable configuration* (LFC), which is sometimes referred to as “the point least favorable to the alternative”. However, Hansen (2003) proposed a different approach that leads to more powerful tests of composite hypothesis.

Before Hansen, White (2000) proposed a test for Superior Predictive Ability, known as the Reality Check (RC) for data snooping. Also known as data mining, data snooping “occurs when a given set of data is used more than once for purposes of inference or model selection. When such data reuse occurs, there is always the possibility that any satisfactory results obtained may simply be due to chance rather than to any merit inherent in the method yielding the results (White, 2000: 1115)” This is an almost inevitable difficulty when analyzing time-series data, since that there is only one record of information about a given variable of interest.

The test introduced by White provides simple and straightforward procedures for testing the null hypothesis: the best model encountered in a specification search has no predictive superiority over a given benchmark model.

When compared to the RC test, the Hansen test proved to be more powerful and less sensitive to the inclusion of poor and irrelevant alternatives. Hansen test differs from the RC test in two ways. First, a studentized test statistic is employed:

$$T_n^{SPA} = \max \left[\max_{k=1, \dots, m} \frac{n^{1/2} \bar{d}_k}{\hat{\omega}_k}, 0 \right], \quad (24)$$

where $\hat{\omega}_k^2$ is some consistent estimator of $\omega_k^2 \equiv \text{var}(n^{1/2} \bar{d}_k)$ and $\bar{d}_k = n^{-1} \sum_{t=1}^n d_{k,t}$. $d_{k,t}$ denotes the performance of model k relative to benchmark at time t.

Second, a null distribution based on $N_m(\hat{\mu}^c, \hat{\Omega})$ is invoked, where $\hat{\mu}^c$ is a carefully chosen estimator for μ that conforms with the null hypothesis. Specifically, Hansen suggested the estimator:

$$\hat{\mu}^c = \bar{d}_k l_{\{n^{1/2} \bar{d}_k / \hat{\omega}_k \leq -\sqrt{2 \log \log n}\}}, \quad k = 1, \dots, m, \quad (25)$$

where $l_{(\cdot)}$ denotes the indicator function.

Consider the vector of relative performances, $d_t = (d_{1,t}, \dots, d_{m,t})$. Assuming that $\mu = E(d_t)$, the null hypothesis that the benchmark is not inferior to any of the alternatives is:

$$H_0: \mu \leq 0$$

The advantages of the studentized test statistic and the sample dependent null distribution is that they don't rely on stationarity, and are therefore expected to be useful in a more general context. Now, let us discuss each one of these modifications individually.

As shown by Hansen, studentizing the individual statistics will allow a comparison between objects measured in the same "units of standard deviation". Not doing so, will result in a pointless comparison between objects measured in different units. There is one exception where the studentization may reduce the power that occurs when the best performing model has the largest variance. Since poor performing models also tend to have the most erratic performances, the author considered this case to be of little empirical relevance.

The second modification made by Hansen had to do with the possible erosion of RC' power when poor alternatives are included in the analysis. In other words, the p-value associated to the RC test can be increased in an artificial way by adding poor forecasts to the set of alternative forecasts. Naturally, we want to avoid that. Given that the poor alternatives are irrelevant for the asymptotic distribution, a proper test should reduce the influence of these models, while preserving the influence of the models with $\mu_k = 0$. Having this problem in mind, Hansen constructed his test in a way that incorporates all models, while it reduces the influence of alternatives that the data suggest are poor.

Since the test statistics have asymptotic distributions that depend on μ and Ω , these are nuisance parameters. The traditional way to proceed in this case is to replace a consistent estimator for Ω and employ LFC over the values of μ that satisfy the null hypothesis. However, Peter Hansen showed that this approach leads to some rather unfortunate properties when testing for SPA. Therefore, he proposed an alternative way to handle the nuisance dependence of μ , where a data dependent choice for μ is used, rather than $\mu=0$ as dictated by the LFC.

The estimator chosen by Hansen ($\hat{\mu}^C$) was motivated by the law of the iterated logarithm. According to this law:

$$P\left(\lim_{n \rightarrow \infty} \inf \frac{n^{1/2}(\bar{d}_k - \mu_k)}{\omega_k} = -\sqrt{2 \log \log n}\right) = 1, \quad (26)$$

and

$$P\left(\lim_{n \rightarrow \infty} \sup \frac{n^{1/2}(\bar{d}_k - \mu_k)}{\omega_k} = +\sqrt{2 \log \log n}\right) = 1, \quad (27)$$

This way, $\hat{\mu}^C$ meet the necessary asymptotic requirements defined by Hansen. Estimator $\hat{\mu}^C$ proved to account for the fact that poor alternatives should be discarded asymptotically but not in finite samples.

While $\hat{\mu}^c$ separates correctly the good alternatives from the poor ones, there are other threshold rates that also produce valid tests. Because different threshold rates will lead to different p-values in finite samples, it is convenient to determine an upper and lower bound ($\hat{\mu}^u$ and $\hat{\mu}^l$, respectively).

$$\hat{\mu}_k^u = 0 \tag{28}$$

$$\hat{\mu}_k^l \equiv \min(\bar{d}_k, 0) \tag{29}$$

where $k=1, \dots, m$.

In order to obtain the p-values of the three tests for SPA, Hansen followed a bootstrap implementation based on the stationary bootstrap of Politis and Romano (1994). Consequently, we will have six different tests to be estimated, a result of two test statistics (RC and SPA) and three null distributions (one for each estimator, $\hat{\mu}^c$, $\hat{\mu}^u$ and $\hat{\mu}^l$). Hansen studied the size and power properties of these six tests. The rejection frequencies of these tests at levels 5% and 10% can be seen in Tables 11, 12 and 13. Numbers in italic are used when the null hypothesis is true ($\Lambda_1 = 0$). Numbers in standard font represent powers for the various local alternatives ($\Lambda_1 < 0$).

Table 11: Rejection Frequencies under the Null and Alternative ($m=100$ and $n=200$)

Λ_1	Level: $\alpha = 0.05$						Level: $\alpha = 0.10$					
	RC _I	RC _C	RC _M	SPA _I	SPA _C	SPA _M	RC _I	RC _C	RC _M	SPA _I	SPA _C	SPA _M
Panel A: $\Lambda_0 = 0$												
0	<i>0.055</i>	<i>0.053</i>	<i>0.053</i>	<i>0.062</i>	<i>0.060</i>	<i>0.060</i>	<i>0.108</i>	<i>0.101</i>	<i>0.101</i>	<i>0.116</i>	<i>0.110</i>	<i>0.109</i>
-1	0.057	0.054	0.054	0.077	0.074	0.074	0.112	0.105	0.105	0.136	0.129	0.129
-2	0.121	0.111	0.111	0.310	0.280	0.280	0.219	0.197	0.197	0.436	0.389	0.388
-3	0.550	0.471	0.470	0.848	0.764	0.761	0.727	0.620	0.618	0.921	0.845	0.841
-4	0.968	0.888	0.882	0.997	0.979	0.976	0.993	0.947	0.941	1.000	0.990	0.987
-5	1.000	0.996	0.992	1.000	1.000	1.000	1.000	0.999	0.998	1.000	1.000	1.000
Panel B: $\Lambda_0 = 1$												
0	<i>0.013</i>	<i>0.010</i>	<i>0.010</i>	<i>0.026</i>	<i>0.022</i>	<i>0.022</i>	<i>0.035</i>	<i>0.025</i>	<i>0.025</i>	<i>0.055</i>	<i>0.044</i>	<i>0.044</i>
-1	0.013	0.010	0.010	0.047	0.041	0.040	0.036	0.027	0.027	0.087	0.072	0.071
-2	0.036	0.028	0.028	0.312	0.252	0.250	0.084	0.060	0.060	0.436	0.345	0.342
-3	0.301	0.201	0.197	0.862	0.744	0.733	0.516	0.334	0.327	0.928	0.829	0.814
-4	0.896	0.677	0.658	0.998	0.977	0.971	0.971	0.816	0.793	1.000	0.989	0.984
-5	1.000	0.968	0.952	1.000	1.000	0.999	1.000	0.991	0.980	1.000	1.000	1.000
Panel C: $\Lambda_0 = 2$												
0	<i>0.004</i>	<i>0.002</i>	<i>0.002</i>	<i>0.018</i>	<i>0.012</i>	<i>0.012</i>	<i>0.013</i>	<i>0.007</i>	<i>0.006</i>	<i>0.039</i>	<i>0.026</i>	<i>0.026</i>
-1	0.004	0.002	0.002	0.044	0.032	0.032	0.014	0.007	0.006	0.080	0.058	0.056
-2	0.013	0.007	0.006	0.336	0.244	0.238	0.041	0.020	0.019	0.464	0.336	0.324
-3	0.195	0.077	0.073	0.881	0.745	0.721	0.401	0.167	0.152	0.941	0.827	0.799
-4	0.842	0.460	0.414	0.999	0.978	0.968	0.957	0.659	0.598	1.000	0.989	0.982
-5	0.999	0.911	0.855	1.000	1.000	0.999	1.000	0.971	0.934	1.000	1.000	1.000
Panel D: $\Lambda_0 = 5$												
0	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.014</i>	<i>0.007</i>	<i>0.005</i>	<i>0.008</i>	<i>0.001</i>	<i>0.000</i>	<i>0.032</i>	<i>0.013</i>	<i>0.011</i>
-1	0.002	0.000	0.000	0.056	0.031	0.025	0.009	0.001	0.000	0.101	0.054	0.044
-2	0.012	0.001	0.001	0.433	0.273	0.227	0.047	0.005	0.003	0.573	0.370	0.306
-3	0.262	0.032	0.017	0.929	0.787	0.710	0.533	0.088	0.045	0.968	0.860	0.784
-4	0.913	0.336	0.167	1.000	0.986	0.966	0.983	0.581	0.312	1.000	0.995	0.979
-5	1.000	0.894	0.620	1.000	1.000	0.999	1.000	0.974	0.786	1.000	1.000	1.000
Panel E: $\Lambda_0 = 10$												
0	<i>0.003</i>	<i>0.000</i>	<i>0.000</i>	<i>0.016</i>	<i>0.007</i>	<i>0.002</i>	<i>0.011</i>	<i>0.001</i>	<i>0.000</i>	<i>0.036</i>	<i>0.015</i>	<i>0.006</i>
-1	0.004	0.000	0.000	0.080	0.043	0.022	0.014	0.001	0.000	0.149	0.073	0.039
-2	0.037	0.002	0.000	0.532	0.340	0.221	0.128	0.011	0.001	0.675	0.455	0.298
-3	0.487	0.064	0.006	0.953	0.843	0.703	0.768	0.181	0.021	0.980	0.907	0.779
-4	0.973	0.526	0.091	1.000	0.992	0.964	0.997	0.772	0.196	1.000	0.998	0.979
-5	1.000	0.963	0.462	1.000	1.000	0.999	1.000	0.993	0.662	1.000	1.000	1.000

Estimated rejection frequencies for the six tests for SPA under the null hypothesis ($\Lambda_1 = 0$) and local alternatives ($\Lambda_1 < 0$). Thus the rejection frequencies in italic font correspond to Type I errors and those in normal font are correspond to local powers. The reality check of White (2000) is denoted by RC_M and the test advocated by this paper is denoted by SPA_C.

Source: (Hansen, 2005)

Table 12: Rejection Frequencies under the Null and Alternative ($m=100$ and $n=1000$)

Λ_1	Level: $\alpha = 0.05$						Level: $\alpha = 0.10$					
	RC_I	RC_C	RC_M	SPA_I	SPA_C	SPA_M	RC_I	RC_C	RC_M	SPA_I	SPA_C	SPA_M
Panel A: $\Lambda_0 = 0$												
0	<i>0.051</i>	<i>0.048</i>	<i>0.048</i>	<i>0.051</i>	<i>0.048</i>	<i>0.048</i>	<i>0.104</i>	<i>0.098</i>	<i>0.098</i>	<i>0.107</i>	<i>0.100</i>	<i>0.100</i>
-1	0.054	0.051	0.051	0.068	0.064	0.064	0.110	0.103	0.103	0.131	0.122	0.122
-2	0.125	0.116	0.116	0.309	0.282	0.282	0.223	0.202	0.202	0.435	0.391	0.390
-3	0.556	0.480	0.479	0.843	0.762	0.760	0.729	0.624	0.622	0.918	0.842	0.840
-4	0.970	0.889	0.886	0.998	0.980	0.977	0.995	0.945	0.941	1.000	0.992	0.990
-5	1.000	0.996	0.994	1.000	1.000	1.000	1.000	0.999	0.997	1.000	1.000	1.000
Panel B: $\Lambda_0 = 1$												
0	<i>0.011</i>	<i>0.009</i>	<i>0.009</i>	<i>0.020</i>	<i>0.017</i>	<i>0.017</i>	<i>0.031</i>	<i>0.024</i>	<i>0.023</i>	<i>0.050</i>	<i>0.040</i>	<i>0.039</i>
-1	0.011	0.009	0.009	0.043	0.036	0.035	0.033	0.025	0.025	0.086	0.069	0.069
-2	0.034	0.026	0.026	0.312	0.252	0.250	0.084	0.059	0.059	0.436	0.346	0.342
-3	0.316	0.205	0.203	0.859	0.740	0.732	0.520	0.338	0.331	0.927	0.822	0.814
-4	0.900	0.682	0.666	0.999	0.978	0.972	0.973	0.816	0.797	1.000	0.990	0.985
-5	1.000	0.968	0.955	1.000	1.000	0.999	1.000	0.991	0.982	1.000	1.000	1.000
Panel C: $\Lambda_0 = 2$												
0	<i>0.003</i>	<i>0.001</i>	<i>0.001</i>	<i>0.014</i>	<i>0.009</i>	<i>0.009</i>	<i>0.012</i>	<i>0.004</i>	<i>0.004</i>	<i>0.034</i>	<i>0.022</i>	<i>0.021</i>
-1	0.003	0.002	0.002	0.042	0.029	0.028	0.013	0.004	0.004	0.079	0.055	0.054
-2	0.014	0.006	0.006	0.338	0.242	0.236	0.042	0.018	0.017	0.465	0.330	0.322
-3	0.202	0.082	0.077	0.881	0.737	0.720	0.411	0.169	0.159	0.941	0.820	0.798
-4	0.844	0.461	0.428	0.999	0.979	0.969	0.959	0.652	0.602	1.000	0.991	0.983
-5	1.000	0.906	0.861	1.000	1.000	0.999	1.000	0.969	0.936	1.000	1.000	1.000
Panel D: $\Lambda_0 = 5$												
0	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.012</i>	<i>0.005</i>	<i>0.004</i>	<i>0.006</i>	<i>0.000</i>	<i>0.000</i>	<i>0.029</i>	<i>0.011</i>	<i>0.008</i>
-1	0.002	0.000	0.000	0.057	0.028	0.024	0.007	0.001	0.000	0.103	0.051	0.042
-2	0.014	0.001	0.000	0.435	0.267	0.225	0.047	0.004	0.002	0.572	0.364	0.306
-3	0.270	0.029	0.017	0.930	0.777	0.708	0.540	0.084	0.044	0.968	0.851	0.784
-4	0.917	0.328	0.175	0.999	0.987	0.966	0.987	0.554	0.320	1.000	0.995	0.981
-5	1.000	0.877	0.632	1.000	1.000	0.999	1.000	0.966	0.791	1.000	1.000	1.000
Panel E: $\Lambda_0 = 10$												
0	<i>0.003</i>	<i>0.000</i>	<i>0.000</i>	<i>0.013</i>	<i>0.005</i>	<i>0.003</i>	<i>0.010</i>	<i>0.001</i>	<i>0.000</i>	<i>0.033</i>	<i>0.012</i>	<i>0.005</i>
-1	0.003	0.000	0.000	0.083	0.042	0.022	0.013	0.001	0.000	0.145	0.070	0.039
-2	0.039	0.002	0.000	0.534	0.335	0.220	0.128	0.010	0.000	0.672	0.444	0.299
-3	0.498	0.060	0.006	0.954	0.835	0.703	0.762	0.165	0.020	0.980	0.900	0.778
-4	0.974	0.496	0.095	0.999	0.994	0.965	0.997	0.737	0.203	1.000	0.998	0.980
-5	1.000	0.953	0.480	1.000	1.000	0.999	1.000	0.993	0.669	1.000	1.000	1.000

Estimated rejection frequencies for the six tests for SPA under the null hypothesis ($\Lambda_1 = 0$) and local alternatives ($\Lambda_1 < 0$). Thus the rejection frequencies in italic font correspond to Type I errors and those in normal font are correspond to local powers. The reality check of White (2000) is denoted by RC_M and the test advocated by this paper is denoted by SPA_C .

Source: (Hansen, 2005)

Table 11 contains the result for the case where $m=100$ and $n=200$. In the situation where all 100 inequalities are binding ($\Lambda_0 = \Lambda_1 = 0$), we see that the rejection probabilities are close to the nominal levels for all the tests. Through the analysis of Table 11, it is possible to note that the SPA_c-test has an over-rejection by 1%. However, this doesn't seem to be problematic, since this over-rejection disappears when the sample size is increased to $n=1000$, as we can see Table 12. It is interesting to see that the liberal null distribution does not lead to a large over-rejection, which might be explained by a positive correlation across alternatives and a consequent positive correlation between the test statistic and $\hat{\mu}^l$. That way, the critical value will tend to be excessively small when the test statistic is small. Then, this correlation will reduce the over-rejection of the $\hat{\mu}^l$ -based tests, suggesting that Hansen test can be improved. For that, it would be necessary to find a way to incorporate information about the off-diagonal elements of Ω .

Panel A corresponds to the case where $\mu=0$, and is therefore the best possible situation for LFC-based tests. So this is the only situation where the LFC-based tests apply the correct asymptotic distribution. For that reason, it was expectable that the tests that are based on $\hat{\mu}^u = 0$ do well, which indeed happened. Fortunately, SPA_c also performs well in this case. Turning to the configurations where $\Lambda_0 > 0$, it is possible to notice the advantages of using the sample dependent null distribution. For example, in Panel E of Table 11, when $(\Lambda_0, \Lambda_1) = (10, -3)$, while the RC almost never rejects the null hypothesis, SPA_c-test has a power of approximately 84%.

Concerning to Table 13, we only notice a slight over-rejection when all inequalities are binding, ($\Lambda_0 = \Lambda_1 = 0$). The power properties are quite good, despite the fact that 1000 alternative are being compared to the benchmark.

Table 13: Rejection Frequencies under the Null and Alternative ($m=1000$ and $n=200$)

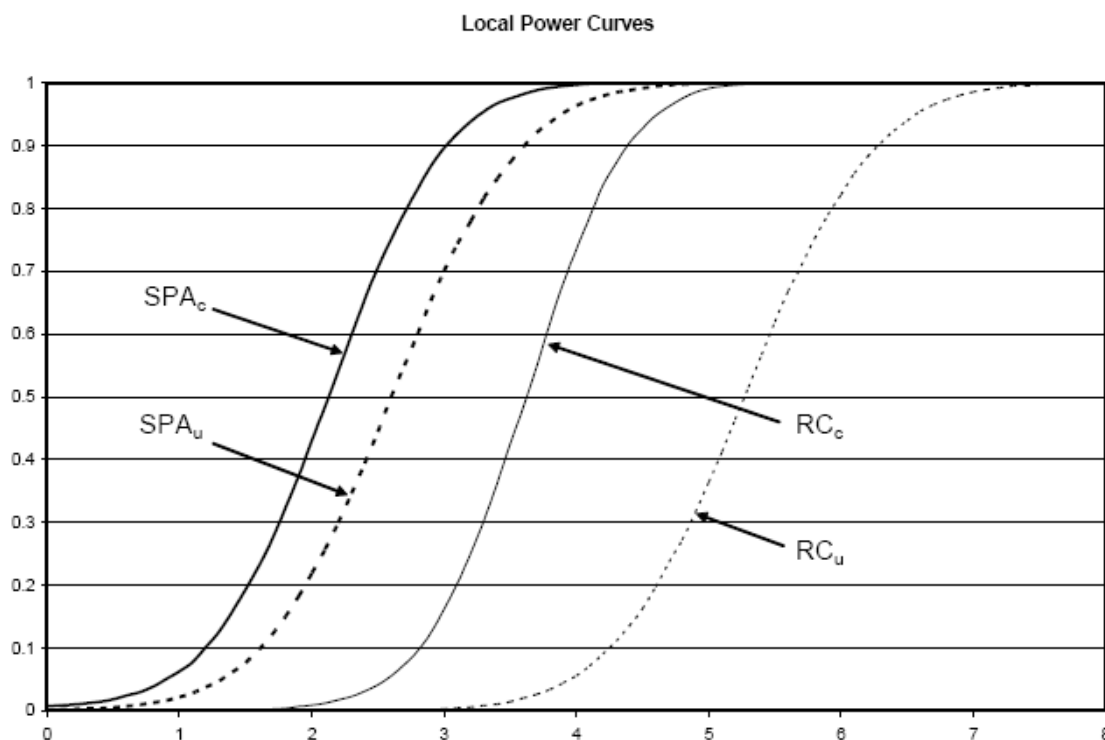
Λ_1	Level: $\alpha = 0.05$						Level: $\alpha = 0.10$					
	RC_I	RC_C	RC_U	SPA_I	SPA_C	SPA_U	RC_I	RC_C	RC_U	SPA_I	SPA_C	SPA_U
Panel A: $\Lambda_0 = 0$												
0	<i>0.049</i>	<i>0.047</i>	<i>0.047</i>	<i>0.064</i>	<i>0.062</i>	<i>0.062</i>	<i>0.106</i>	<i>0.100</i>	<i>0.100</i>	<i>0.125</i>	<i>0.119</i>	<i>0.119</i>
-1	<i>0.049</i>	<i>0.047</i>	<i>0.047</i>	<i>0.066</i>	<i>0.064</i>	<i>0.064</i>	<i>0.106</i>	<i>0.101</i>	<i>0.100</i>	<i>0.128</i>	<i>0.122</i>	<i>0.122</i>
-2	<i>0.061</i>	<i>0.058</i>	<i>0.058</i>	<i>0.173</i>	<i>0.164</i>	<i>0.164</i>	<i>0.128</i>	<i>0.121</i>	<i>0.121</i>	<i>0.269</i>	<i>0.252</i>	<i>0.252</i>
-3	<i>0.288</i>	<i>0.262</i>	<i>0.262</i>	<i>0.658</i>	<i>0.598</i>	<i>0.596</i>	<i>0.434</i>	<i>0.388</i>	<i>0.388</i>	<i>0.770</i>	<i>0.699</i>	<i>0.697</i>
-4	<i>0.815</i>	<i>0.720</i>	<i>0.719</i>	<i>0.980</i>	<i>0.937</i>	<i>0.933</i>	<i>0.917</i>	<i>0.828</i>	<i>0.824</i>	<i>0.994</i>	<i>0.967</i>	<i>0.963</i>
-5	<i>0.998</i>	<i>0.971</i>	<i>0.967</i>	<i>1.000</i>	<i>0.999</i>	<i>0.998</i>	<i>1.000</i>	<i>0.991</i>	<i>0.988</i>	<i>1.000</i>	<i>1.000</i>	<i>1.000</i>
Panel B: $\Lambda_0 = 1$												
0	<i>0.009</i>	<i>0.007</i>	<i>0.007</i>	<i>0.025</i>	<i>0.022</i>	<i>0.022</i>	<i>0.022</i>	<i>0.017</i>	<i>0.017</i>	<i>0.054</i>	<i>0.045</i>	<i>0.045</i>
-1	<i>0.009</i>	<i>0.007</i>	<i>0.007</i>	<i>0.029</i>	<i>0.025</i>	<i>0.025</i>	<i>0.022</i>	<i>0.017</i>	<i>0.017</i>	<i>0.059</i>	<i>0.050</i>	<i>0.050</i>
-2	<i>0.010</i>	<i>0.008</i>	<i>0.008</i>	<i>0.150</i>	<i>0.127</i>	<i>0.127</i>	<i>0.026</i>	<i>0.020</i>	<i>0.020</i>	<i>0.229</i>	<i>0.192</i>	<i>0.191</i>
-3	<i>0.066</i>	<i>0.049</i>	<i>0.049</i>	<i>0.652</i>	<i>0.555</i>	<i>0.548</i>	<i>0.150</i>	<i>0.103</i>	<i>0.102</i>	<i>0.759</i>	<i>0.652</i>	<i>0.643</i>
-4	<i>0.502</i>	<i>0.345</i>	<i>0.339</i>	<i>0.980</i>	<i>0.924</i>	<i>0.916</i>	<i>0.701</i>	<i>0.500</i>	<i>0.488</i>	<i>0.993</i>	<i>0.956</i>	<i>0.947</i>
-5	<i>0.965</i>	<i>0.813</i>	<i>0.794</i>	<i>1.000</i>	<i>0.998</i>	<i>0.997</i>	<i>0.994</i>	<i>0.907</i>	<i>0.886</i>	<i>1.000</i>	<i>1.000</i>	<i>0.999</i>
Panel C: $\Lambda_0 = 2$												
0	<i>0.001</i>	<i>0.000</i>	<i>0.000</i>	<i>0.015</i>	<i>0.011</i>	<i>0.011</i>	<i>0.005</i>	<i>0.002</i>	<i>0.002</i>	<i>0.035</i>	<i>0.026</i>	<i>0.025</i>
-1	<i>0.001</i>	<i>0.000</i>	<i>0.000</i>	<i>0.020</i>	<i>0.015</i>	<i>0.015</i>	<i>0.005</i>	<i>0.002</i>	<i>0.002</i>	<i>0.043</i>	<i>0.032</i>	<i>0.032</i>
-2	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.155</i>	<i>0.115</i>	<i>0.113</i>	<i>0.006</i>	<i>0.003</i>	<i>0.003</i>	<i>0.233</i>	<i>0.172</i>	<i>0.167</i>
-3	<i>0.016</i>	<i>0.007</i>	<i>0.007</i>	<i>0.669</i>	<i>0.544</i>	<i>0.525</i>	<i>0.054</i>	<i>0.022</i>	<i>0.022</i>	<i>0.779</i>	<i>0.636</i>	<i>0.616</i>
-4	<i>0.291</i>	<i>0.125</i>	<i>0.117</i>	<i>0.985</i>	<i>0.923</i>	<i>0.906</i>	<i>0.516</i>	<i>0.243</i>	<i>0.224</i>	<i>0.994</i>	<i>0.954</i>	<i>0.940</i>
-5	<i>0.901</i>	<i>0.576</i>	<i>0.529</i>	<i>1.000</i>	<i>0.999</i>	<i>0.996</i>	<i>0.980</i>	<i>0.744</i>	<i>0.683</i>	<i>1.000</i>	<i>1.000</i>	<i>0.998</i>
Panel D: $\Lambda_0 = 5$												
0	<i>0.000</i>	<i>0.000</i>	<i>0.000</i>	<i>0.011</i>	<i>0.005</i>	<i>0.004</i>	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.029</i>	<i>0.012</i>	<i>0.009</i>
-1	<i>0.000</i>	<i>0.000</i>	<i>0.000</i>	<i>0.019</i>	<i>0.010</i>	<i>0.008</i>	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.044</i>	<i>0.020</i>	<i>0.016</i>
-2	<i>0.000</i>	<i>0.000</i>	<i>0.000</i>	<i>0.199</i>	<i>0.122</i>	<i>0.101</i>	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.291</i>	<i>0.180</i>	<i>0.148</i>
-3	<i>0.011</i>	<i>0.000</i>	<i>0.000</i>	<i>0.748</i>	<i>0.570</i>	<i>0.505</i>	<i>0.045</i>	<i>0.004</i>	<i>0.002</i>	<i>0.843</i>	<i>0.664</i>	<i>0.589</i>
-4	<i>0.303</i>	<i>0.036</i>	<i>0.017</i>	<i>0.993</i>	<i>0.939</i>	<i>0.897</i>	<i>0.575</i>	<i>0.098</i>	<i>0.050</i>	<i>0.998</i>	<i>0.967</i>	<i>0.930</i>
-5	<i>0.936</i>	<i>0.387</i>	<i>0.207</i>	<i>1.000</i>	<i>0.999</i>	<i>0.996</i>	<i>0.992</i>	<i>0.605</i>	<i>0.356</i>	<i>1.000</i>	<i>1.000</i>	<i>0.998</i>
Panel E: $\Lambda_0 = 10$												
0	<i>0.001</i>	<i>0.000</i>	<i>0.000</i>	<i>0.012</i>	<i>0.004</i>	<i>0.003</i>	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.029</i>	<i>0.011</i>	<i>0.004</i>
-1	<i>0.001</i>	<i>0.000</i>	<i>0.000</i>	<i>0.025</i>	<i>0.012</i>	<i>0.007</i>	<i>0.002</i>	<i>0.000</i>	<i>0.000</i>	<i>0.054</i>	<i>0.024</i>	<i>0.011</i>
-2	<i>0.001</i>	<i>0.000</i>	<i>0.000</i>	<i>0.259</i>	<i>0.156</i>	<i>0.097</i>	<i>0.004</i>	<i>0.000</i>	<i>0.000</i>	<i>0.366</i>	<i>0.226</i>	<i>0.141</i>
-3	<i>0.031</i>	<i>0.001</i>	<i>0.000</i>	<i>0.815</i>	<i>0.633</i>	<i>0.495</i>	<i>0.109</i>	<i>0.006</i>	<i>0.000</i>	<i>0.891</i>	<i>0.726</i>	<i>0.579</i>
-4	<i>0.508</i>	<i>0.064</i>	<i>0.005</i>	<i>0.996</i>	<i>0.958</i>	<i>0.892</i>	<i>0.765</i>	<i>0.175</i>	<i>0.018</i>	<i>0.999</i>	<i>0.981</i>	<i>0.926</i>
-5	<i>0.983</i>	<i>0.531</i>	<i>0.099</i>	<i>1.000</i>	<i>1.000</i>	<i>0.995</i>	<i>0.998</i>	<i>0.753</i>	<i>0.210</i>	<i>1.000</i>	<i>1.000</i>	<i>0.998</i>

Estimated rejection frequencies for the six tests for SPA under the null hypothesis ($\Lambda_1 = 0$) and local alternatives ($\Lambda_1 < 0$). Thus the rejection frequencies in italic font correspond to Type I errors and those in normal font are correspond to local powers. The reality check of White (2000) is denoted by RC_U and the test advocated by this paper is denoted by SPA_C .

Source: (Hansen, 2005)

The power curves for the tests that employ $\hat{\mu}^c$ and $\hat{\mu}^u$ are shown in Figure 2, for the case where $m=100$, $n=200$ and $\Lambda_0=20$. The power curves are based on tests that aim at a 5% significance level, and their rejection frequencies are plotted against a range of local alternatives. For the power curves in Figure 2, we conclude that the RC test is dominated by the three other tests. There is a significant increase in power when using the consistent distribution. Moreover, a quite similar improvement is achieved when we use the standardized test statistic, T_n^{SPA} . In fact, according to Hansen's calculations, to regain the power that is lost by using LFC instead of the sample dependent null distribution, it would be necessary a sample size 1.49 times larger, meaning that we are tossing 33% of the data when using the LFC. Besides, when we drop the studentization, a 65% of the data is being discarded. Dropping both modifications is equivalent to tossing away 84% of the data.

Figure 2: Local power curves of the four tests, SPA_c , SPA_u , RC_c and RC_u , for the simulation experiment where $m=100$, $\Lambda_0=20$ and μ_1/\sqrt{n} ($= -\Lambda_1$) ranges from 0 to 8 (the x-axis). The power curves quantify the power improvements from the two modifications of the Reality Check.



Source: (Hansen, 2005)

4. STATISTICAL PROPERTIES OF RETURNS

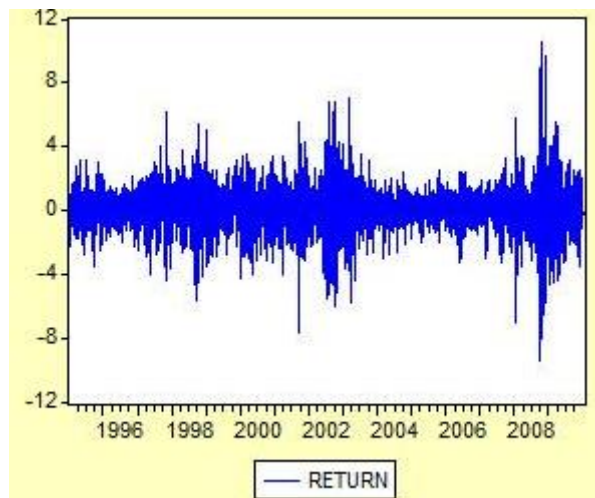
Our data base is formed by the daily closing prices of the CAC40, FTSE100, NIKKEI 225 and S&P500 indexes from January 1, 1995 through December 31, 2009.¹

In order to obtain the daily stock returns (r_t), we followed the conventional procedure:

$$r_t = 100 \times \log\left(\frac{P_t}{P_{t-1}}\right), \quad (29)$$

where P_t is the daily closing price of period t and P_{t-1} is the daily closing price of period $t-1$. Figures 3, 4, 5 and 6 show us the behavior of our four indexes over time.

Figure 3: CAC40 returns



¹ All data series were drawn from the following website: <http://finance.yahoo.com/>

Figure 4: FTSE100 returns

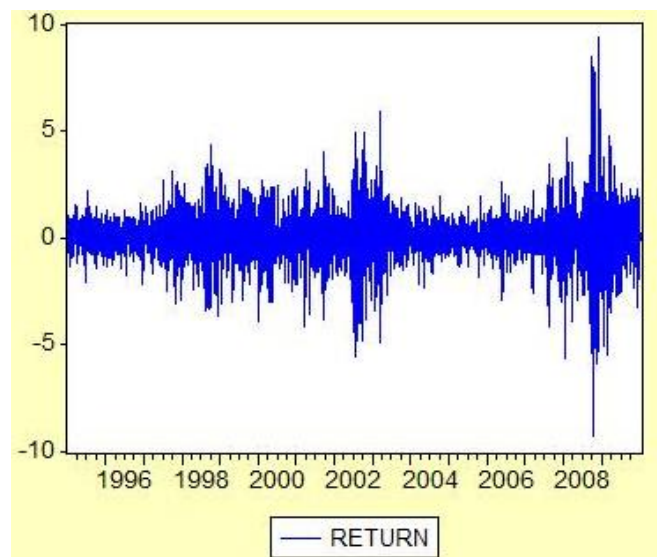


Figure 5: NIKKEI225 returns

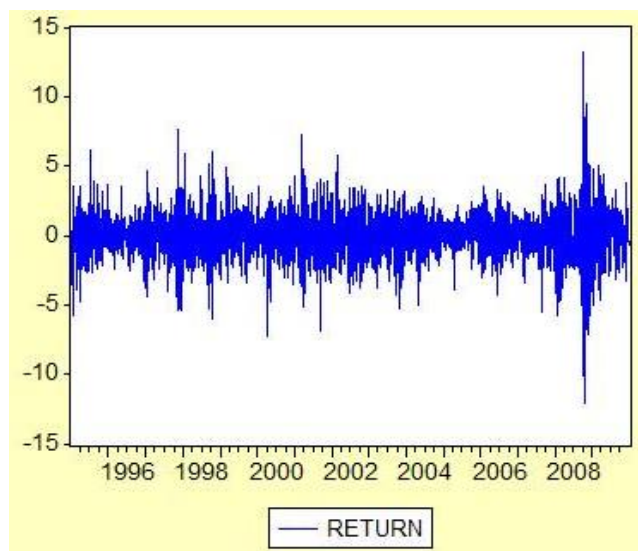
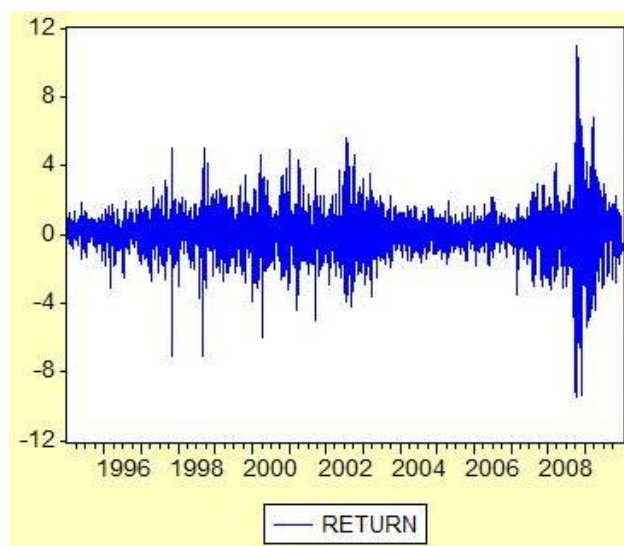


Figure 6: S&P500 returns



It is evident that the ending of 2008 represent the most volatile period over the last 15 years, which is as a result of the sub-prime crisis that affected the whole world. Although it started in the USA, this crisis quickly spread all around the globe, as it can be confirmed by our four geographically scattered indexes (two from Europe, one from Asia and another one from Northern America). We can also identify another high volatility periods as for example: the October 1997 Asia mini-crash, the 1998 Russian financial crisis, the March 2000 dot-com bubble crash or the post-9/11 incident in 2001.

Table 14 provides a general overview of the data used.

Table 14: Statistical properties of returns

	CAC40	FTSE100	NIKKEI225	S&P500
Observations	3803	3788	3686	3777
Mean	0,0193	0,015	-0,0169	0,0235
Median	0,0466	0,0504	0,0018	0,0683
Maximum	10,5946	9,3842	13,2346	10,9572
Minimum	-9,4715	-9,2646	-12,111	-9,4695
Standard Deviation	1,483	1,2313	1,5822	1,2823
Skewness	-0,0278	-0,1359	-0,1912	-0,2001
Kurtosis	7,6934	9,175	8,3519	11,1562
Jarque-Bera	3491,055 [0,0000]	6029,919 [0,0000]	4421,575 [0,0000]	10494,45 [0,0000]

As we can see, excepting NIKKEI225, the mean returns all are positive. NIKKEI225 also seems to be the most volatile index, since it has the lower and the high value, as well as the higher standard deviation. Without surprises, the Jarque-Bera test (Jarque and Bera, 1987) rejects the normality assumption for each of the series.

The Ljung-Box statistics on the returns, computed at a tenth-order lag, are shown in Figures 7, 8, 9 and 10.

Figure 7: Ljung-Box test for the CAC40 returns

Date: 05/01/10 Time: 19:32
 Sample: 1/03/1995 12/31/2009
 Included observations: 3803

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.026	-0.026	2.5825	0.108
		2	-0.030	-0.031	6.0377	0.049
		3	-0.070	-0.071	24.459	0.000
		4	0.047	0.042	32.779	0.000
		5	-0.059	-0.061	45.830	0.000
		6	-0.020	-0.026	47.427	0.000
		7	-0.007	-0.006	47.639	0.000
		8	0.043	0.031	54.561	0.000
		9	-0.027	-0.024	57.330	0.000
		10	-0.001	-0.002	57.334	0.000

Figure 8: Ljung-Box test for the FTSE100 returns

Date: 05/01/10 Time: 19:38
 Sample: 1/03/1995 12/31/2009
 Included observations: 3788

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.035	-0.035	4.6695	0.031
		2	-0.048	-0.050	13.525	0.001
		3	-0.092	-0.096	45.527	0.000
		4	0.076	0.068	67.680	0.000
		5	-0.056	-0.061	79.431	0.000
		6	-0.046	-0.052	87.404	0.000
		7	0.009	0.014	87.733	0.000
		8	0.062	0.043	102.33	0.000
		9	0.001	0.004	102.33	0.000
		10	-0.013	-0.002	103.01	0.000

Figure 9: Ljung-Box test for the NIKKEI225 returns

Date: 05/01/10 Time: 19:39
 Sample: 1/04/1995 12/30/2009
 Included observations: 3686

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.042	-0.042	6.4260	0.011
		2 -0.035	-0.037	10.890	0.004
		3 -0.002	-0.005	10.905	0.012
		4 -0.017	-0.018	11.933	0.018
		5 -0.017	-0.019	12.977	0.024
		6 -0.013	-0.016	13.594	0.035
		7 0.011	0.009	14.078	0.050
		8 0.002	0.002	14.100	0.079
		9 -0.026	-0.026	16.541	0.056
		10 0.039	0.036	22.024	0.015

Figure 10: Ljung-Box test for the S&P500 returns

Date: 05/01/10 Time: 19:40
 Sample: 1/03/1995 12/31/2009
 Included observations: 3777

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.068	-0.068	17.401	0.000
		2 -0.065	-0.070	33.313	0.000
		3 0.027	0.018	36.139	0.000
		4 -0.012	-0.013	36.690	0.000
		5 -0.035	-0.034	41.369	0.000
		6 0.007	0.000	41.581	0.000
		7 -0.040	-0.044	47.757	0.000
		8 0.023	0.019	49.766	0.000
		9 -0.001	-0.005	49.768	0.000
		10 0.029	0.033	53.028	0.000

For the CAC40 index, the Ljung-Box test tells us that period t and period $t-1$ are not correlated, since we do not reject the null hypothesis at a 5% significance level. Still, we reject all the other null hypothesis, meaning that all the other autocorrelation coefficients are statistically significant at a 5% significance level.

Concerning the FTSE100 and the S&P500 indexes, it is possible to conclude that the returns are temporally correlated, since we reject (at a 5% level) the null hypothesis that the autocorrelation coefficients are equal to zero. Then, we can say that these series were not generated through a white noise process.

In its turn, for the NIKKEI225 index we can say that period t is correlated with period $t-7$, $t-8$ and $t-9$, if we assume a 10% significance level. For the remaining periods, we reject the null at a 5% level. Therefore, we conclude that every autocorrelation coefficient is statistically significant. That way, The NIKKEI225 returns were not generated through a white noise process.

Since we identified relevant autocorrelation in every index, we will consider an order five autoregressive model to remove the linear dependency in the series.

5. ECONOMETRIC APPROACH

The empirical distribution of a financial asset can be described as the sum of a predictable part with an unpredictable part:

$$r_t = E(r_t|\Phi_{t-1}) + u_t, \quad (30)$$

where Φ_{t-1} is the relevant information set until, and including, $t-1$.

Based on our previous findings, we computed the conditional mean equation, $E(r_t|\Phi_{t-1})$, as a fifth-order autoregressive process, AR(5), in order to remove the observed linear dependency in the returns:

$$r_t = m + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \phi_4 r_{t-4} + \phi_5 r_{t-5} + u_t, \quad (31)$$

where $u_t = z_t \sigma_t$ and the standardized innovations (z_t) are assumed to be i.i.d. with Student's t distribution.

Figures 11, 12, 13 and 14 show us the Ljung-Box test applied to the residuals after fitting the AR(5) to the series.

Figure 11: Ljung-Box test for the CAC40 residuals

Sample: 1/03/1995 12/31/2009
Included observations: 3798

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.002	-0.002	0.0099	0.921
		2	0.001	0.001	0.0117	0.994
		3	0.000	0.000	0.0120	1.000
		4	-0.004	-0.004	0.0776	0.999
		5	0.002	0.002	0.0960	1.000
		6	-0.026	-0.026	2.7566	0.839
		7	-0.006	-0.006	2.8915	0.895
		8	0.031	0.031	6.5641	0.584
		9	-0.022	-0.022	8.3941	0.495
		10	-0.003	-0.003	8.4273	0.587

Figure 12: Ljung-Box test for the FTSE100 residuals

Sample: 1/03/1995 12/31/2009
Included observations: 3783

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.003	-0.003	0.0376	0.846
		2	0.004	0.004	0.1125	0.945
		3	-0.003	-0.003	0.1546	0.985
		4	-0.004	-0.004	0.2117	0.995
		5	0.003	0.003	0.2397	0.999
		6	-0.049	-0.049	9.2889	0.158
		7	0.017	0.017	10.442	0.165
		8	0.044	0.045	17.752	0.023
		9	0.004	0.004	17.818	0.037
		10	-0.007	-0.008	18.001	0.055

Figure 13: Ljung-Box test for the NIKKEI225 residuals

Sample: 1/04/1995 12/30/2009
Included observations: 3681

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.000	-0.000	0.0003	0.987
		2	0.000	0.000	0.0003	1.000
		3	-0.000	-0.000	0.0004	1.000
		4	-0.001	-0.001	0.0041	1.000
		5	-0.000	-0.000	0.0042	1.000
		6	-0.014	-0.014	0.7611	0.993
		7	0.010	0.010	1.1141	0.993
		8	0.003	0.003	1.1393	0.997
		9	-0.024	-0.024	3.2718	0.953
		10	0.039	0.039	8.8091	0.550

Figure 14: Ljung-Box test for the S&P500 residuals

Sample: 1/03/1995 12/31/2009
Included observations: 3772

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.000	0.000	3.E-07	1.000
		2	-0.002	-0.002	0.0090	0.996
		3	0.000	0.000	0.0090	1.000
		4	0.001	0.001	0.0125	1.000
		5	-0.002	-0.002	0.0313	1.000
		6	-0.002	-0.002	0.0475	1.000
		7	-0.042	-0.042	6.8485	0.445
		8	0.026	0.026	9.4426	0.306
		9	-0.004	-0.005	9.5181	0.391
		10	0.030	0.030	12.842	0.233

As we can see, the AR(5) successfully removed the linear dependency in CAC40, NIKKEI225 and S&P500 series up to the tenth-lag. For the FTSE100, however, the linear dependency was removed to the seventh-lag only, indicating that we could have used an autoregressive process of superior order than five. Still, the results obtained with the AR(5) are quite satisfactory and allow us to consider that the residues are now white noise.

For the conditional variance of $u_t: E(u_t^2 | \Phi_{t-1}) = \sigma_t^2$, we have considered our three conditional heteroskedasticity models, GARCH, EGARCH and GJR.

As stated by Bollerslev et al. (1992) as well as Hansen and Lunde (2005), when modeling financial assets returns volatility, the (1,1) specification in ARCH (p,q) models is rather satisfactory. For that reason, we will use p=1 and q=1.

$$GARCH: \sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (32)$$

$$EGARCH: \ln \sigma_t^2 = \omega + \alpha_1 \frac{|u_{t-1}|}{\sigma_{t-1}} + \gamma_1 \frac{u_{t-1}}{\sigma_{t-1}} + \beta_1 \ln \sigma_{t-1}^2 \quad (33)$$

$$GJR: \sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \gamma_1 I_{t-1} u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (34)$$

where $\omega, \alpha_1, \gamma_1, \beta_1$ are unknown parameters, $I_{t-1} = 1$ if $u_{t-1} < 0$ and $I_{t-1} = 0$ if $u_{t-1} \geq 0$. All models were estimated through maximum likelihood (MLE).

Since volatility itself is not directly observable, establishing the effectiveness of the volatility forecast involves the use of a “volatility proxy” that may constitute an imperfect estimate of the true volatility, as mentioned by Andersen and Bollerslev (1998), for example. Following the conventional approach, squared returns are used as a proxy for the latent volatility process. According to Patton (2006) the squared return on an asset over the period t (assuming a zero mean return) is a conditionally unbiased estimator of the true unobserved conditional variance of the asset over the period t.

Our original sample was divided in two parts: the in-sample (January 3, 1995 to December 31, 2004) and the out-of-sample (January 3, 2005 to December 31, 2009). That way, the parameters for the conditional variance equation are estimated for the first ten years. The remaining five years were considered as the forecast period.

In the following section, we will present and discuss both in-sample and out-of-sample results.

6. ESTIMATION RESULTS²

In-sample results

Figures 15 to 26 report the in-sample results for CAC40, FTSE100, NIKKEI225 and S&P500. In order to compare these results, we will make use of three likelihood based goodness-of-fit criteria. Those criteria are the maximum log-likelihood, the Akaike Information Criteria (Akaike, 1978) and the Schwarz Bayesian Criteria (Scharwz, 1978). The first one is obtained from the maximum likelihood estimation and the bigger it is, the better it is. On the contrary, we want Akaike Information Criteria (AIC) and Schwarz Bayesian Criteria (SBC) to be as lower as possible.

Figure 15: Log-likelihood for GARCH (CAC40)

LogL: LL1
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 17:02
 Sample: 1/13/1995 12/31/2004
 Included observations: 2518
 Evaluation order: By observation
 Convergence achieved after 12 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.070245	0.023399	3.002116	0.0027
ARE(1)	-0.001729	0.021416	-0.080737	0.9357
ARE(2)	-0.009103	0.020529	-0.443443	0.6574
ARE(3)	-0.059459	0.019898	-2.988234	0.0028
ARE(4)	-0.000668	0.020728	-0.032242	0.9743
ARE(5)	-0.042420	0.020351	-2.084470	0.0371
OMEGA(1)	0.015571	0.006225	2.501569	0.0124
ALPHA(1)	0.064730	0.009062	7.143035	0.0000
BETA(1)	0.928201	0.009841	94.32250	0.0000
TDF(1)	15.76205	3.947159	3.993264	0.0001
Log likelihood	-4210.339	Akaike info criterion	3.352136	
Avg. log likelihood	-1.672096	Schwarz criterion	3.375294	
Number of Coefs.	10	Hannan-Quinn criter.	3.360540	

² All calculations were programmed in Eviews 5.0. The program work files are displayed in the Appendix.

Figure 16: Log-likelihood for GJR (CAC40)

LogL: LL2
 Method: Maximum Likelihood (BHHH)
 Date: 05/02/10 Time: 00:57
 Sample: 1/13/1995 12/31/2004
 Included observations: 2518
 Evaluation order: By observation
 Convergence achieved after 10 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.043813	0.023642	1.853169	0.0639
ARE(1)	0.002556	0.021455	0.119157	0.9052
ARE(2)	-0.004572	0.020441	-0.223664	0.8230
ARE(3)	-0.057036	0.020132	-2.833066	0.0046
ARE(4)	0.004313	0.020646	0.208884	0.8345
ARE(5)	-0.038750	0.020223	-1.916138	0.0553
OMEGA(1)	0.021377	0.006084	3.513727	0.0004
ALPHA(1)	0.017966	0.009169	1.959452	0.0501
BETA(1)	0.930785	0.009415	98.86679	0.0000
GAMMA(1)	0.077534	0.013277	5.839522	0.0000
TDF(1)	18.00287	5.200847	3.461527	0.0005
Log likelihood	-4193.716	Akaike info criterion	3.339727	
Avg. log likelihood	-1.665495	Schwarz criterion	3.365201	
Number of Coefs.	11	Hannan-Quinn criter.	3.348972	

Figure 17: Log-likelihood for EGARCH (CAC40)

LogL: LL3
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 17:03
 Sample: 1/13/1995 12/31/2004
 Included observations: 2518
 Evaluation order: By observation
 Failure to improve Likelihood after 34 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	-0.095101	0.215474	-0.441360	0.6590
ARE(1)	-0.103082	0.160317	-0.642988	0.5202
ARE(2)	-0.413312	0.144787	-2.854622	0.0043
ARE(3)	-0.408168	0.177242	-2.302889	0.0213
ARE(4)	-0.723035	0.206900	-3.494613	0.0005
ARE(5)	-0.390459	0.161822	-2.412892	0.0158
OMEGA(1)	-0.078033	0.014516	-5.375658	0.0000
BETA(1)	0.986868	0.003698	266.8873	0.0000
ALPHA(1)	0.081243	0.014183	5.728031	0.0000
GAMMA(1)	-0.034232	0.009778	-3.500919	0.0005
TDF(1)	18.81993	6.356323	2.960821	0.0031
Log likelihood	-4183.592	Akaike info criterion	3.331685	
Avg. log likelihood	-1.661474	Schwarz criterion	3.357159	
Number of Coefs.	11	Hannan-Quinn criter.	3.340930	

Figure 18: Log-likelihood for GARCH (FTSE100)

LogL: LL1
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:33
 Sample: 1/13/1995 12/31/2004
 Included observations: 2518
 Evaluation order: By observation
 Convergence achieved after 12 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.055388	0.017084	3.242153	0.0012
ARE(1)	0.003415	0.021369	0.159829	0.8730
ARE(2)	-0.028088	0.020268	-1.385826	0.1658
ARE(3)	-0.058029	0.020574	-2.820427	0.0048
ARE(4)	-0.006758	0.020651	-0.327223	0.7435
ARE(5)	-0.038531	0.020170	-1.910340	0.0561
OMEGA(1)	0.008267	0.003087	2.678247	0.0074
ALPHA(1)	0.073359	0.010438	7.028017	0.0000
BETA(1)	0.920583	0.010808	85.17889	0.0000
TDF(1)	17.41807	5.055271	3.445527	0.0006
Log likelihood	-3480.699	Akaike info criterion	2.772596	
Avg. log likelihood	-1.382327	Schwarz criterion	2.795754	
Number of Coefs.	10	Hannan-Quinn criter.	2.781001	

Figure 19: Log-likelihood for GJR (FTSE100)

LogL: LL2
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:33
 Sample: 1/13/1995 12/31/2004
 Included observations: 2518
 Evaluation order: By observation
 Failure to improve Likelihood after 3 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.045215	0.017033	2.654608	0.0079
ARE(1)	0.004872	0.020922	0.232849	0.8159
ARE(2)	-0.028147	0.019768	-1.423828	0.1545
ARE(3)	-0.059451	0.020071	-2.962025	0.0031
ARE(4)	-0.006434	0.020491	-0.313975	0.7535
ARE(5)	-0.041238	0.019784	-2.084458	0.0371
OMEGA(1)	0.007096	0.002600	2.728939	0.0064
ALPHA(1)	0.015134	0.006171	2.452379	0.0142
BETA(1)	0.942399	0.009284	101.5112	0.0000
GAMMA(1)	0.066336	0.012061	5.500145	0.0000
TDF(1)	15.37516	3.762267	4.086676	0.0000
Log likelihood	-3460.483	Akaike info criterion	2.757334	
Avg. log likelihood	-1.374298	Schwarz criterion	2.782808	
Number of Coefs.	11	Hannan-Quinn criter.	2.766579	

Figure 20: Log-likelihood for EGARCH (FTSE100)

LogL: LL3
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:33
 Sample: 1/13/1995 12/31/2004
 Included observations: 2518
 Evaluation order: By observation
 Failure to improve Likelihood after 29 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.336443	0.140418	2.396013	0.0166
ARE(1)	-0.504176	0.193697	-2.602917	0.0092
ARE(2)	-0.146429	0.187363	-0.781526	0.4345
ARE(3)	0.269354	0.150483	1.789926	0.0735
ARE(4)	0.209251	0.155966	1.341649	0.1797
ARE(5)	-0.004411	0.127696	-0.034544	0.9724
OMEGA(1)	-0.089659	0.015842	-5.659653	0.0000
BETA(1)	1.002548	0.005649	177.4645	0.0000
ALPHA(1)	0.065492	0.015489	4.228417	0.0000
GAMMA(1)	-0.067345	0.018462	-3.647864	0.0003
TDF(1)	18.69202	4.890611	3.822022	0.0001
Log likelihood	-3440.617	Akaike info criterion	2.741555	
Avg. log likelihood	-1.366409	Schwarz criterion	2.767029	
Number of Coefs.	11	Hannan-Quinn criter.	2.750800	

Figure 21: Log-likelihood for GARCH (NIKKEI225)

LogL: LL1
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:33
 Sample: 1/17/1995 12/30/2004
 Included observations: 2454
 Evaluation order: By observation
 Convergence achieved after 13 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	-0.001661	0.026887	-0.061765	0.9508
ARE(1)	-0.025907	0.021736	-1.191891	0.2333
ARE(2)	-0.026320	0.020644	-1.274937	0.2023
ARE(3)	0.001858	0.020224	0.091849	0.9268
ARE(4)	-0.016204	0.020673	-0.783840	0.4331
ARE(5)	-0.008437	0.020194	-0.417792	0.6761
OMEGA(1)	0.066470	0.018577	3.578043	0.0003
ALPHA(1)	0.064747	0.011681	5.543070	0.0000
BETA(1)	0.906313	0.016774	54.03203	0.0000
TDF(1)	8.518345	1.180024	7.218791	0.0000
Log likelihood	-4335.214	Akaike info criterion	3.541332	
Avg. log likelihood	-1.766591	Schwarz criterion	3.564989	
Number of Coefs.	10	Hannan-Quinn criter.	3.549929	

Figure 22: Log-likelihood for GJR (NIKKEI225)

LogL: LL2
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:33
 Sample: 1/17/1995 12/30/2004
 Included observations: 2454
 Evaluation order: By observation
 Convergence achieved after 9 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	-0.024091	0.026425	-0.911651	0.3620
ARE(1)	-0.023770	0.021544	-1.103361	0.2699
ARE(2)	-0.022978	0.020634	-1.113610	0.2654
ARE(3)	0.003800	0.019942	0.190557	0.8489
ARE(4)	-0.010671	0.020403	-0.523017	0.6010
ARE(5)	-0.004068	0.020146	-0.201913	0.8400
OMEGA(1)	0.058471	0.015264	3.830769	0.0001
ALPHA(1)	0.014666	0.009476	1.547805	0.1217
BETA(1)	0.914799	0.014809	61.77524	0.0000
GAMMA(1)	0.091298	0.017289	5.280694	0.0000
TDF(1)	8.849325	1.225894	7.218672	0.0000
Log likelihood	-4318.824	Akaike info criterion	3.528789	
Avg. log likelihood	-1.759912	Schwarz criterion	3.554811	
Number of Coefs.	11	Hannan-Quinn criter.	3.538245	

Figure 23: Log-likelihood for EGARCH (NIKKEI225)

LogL: LL3
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:34
 Sample: 1/17/1995 12/30/2004
 Included observations: 2454
 Evaluation order: By observation
 Failure to improve Likelihood after 28 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	1.630947	0.301084	5.416913	0.0000
ARE(1)	-0.348918	0.120325	-2.899809	0.0037
ARE(2)	-0.756098	0.202649	-3.731079	0.0002
ARE(3)	0.322908	0.135553	2.382152	0.0172
ARE(4)	0.069349	0.145969	0.475091	0.6347
ARE(5)	-0.453273	0.129613	-3.497121	0.0005
OMEGA(1)	-0.167452	0.025060	-6.682038	0.0000
BETA(1)	0.994430	0.009560	104.0152	0.0000
ALPHA(1)	0.142442	0.027577	5.165269	0.0000
GAMMA(1)	0.035189	0.022525	1.562266	0.1182
TDF(1)	9.314539	1.346267	6.918790	0.0000
Log likelihood	-4306.680	Akaike info criterion	3.518892	
Avg. log likelihood	-1.754963	Schwarz criterion	3.544914	
Number of Coefs.	11	Hannan-Quinn criter.	3.528348	

According to Tables 15-23, EGARCH is the best model to sculpt the conditional variance of CAC40, FTSE100 and NIKKEI225 returns. GJR comes in second place, leaving GARCH as the worst model. As expected, both asymmetric models (EGARCH and GJR) beat GARCH in terms of goodness-of-fit measures, proving that they successfully capture the leverage effect.

For the S&P500, however, the results are quite surprisingly, given that EGARCH is beaten by GARCH. Besides, GJR appears to be the most accurate predictor of S&P500 return's volatility.

Figure 24: Log-likelihood for GARCH (S&P500)

LogL: LL1
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:33
 Sample: 1/13/1995 12/31/2004
 Included observations: 2511
 Evaluation order: By observation
 Convergence achieved after 12 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.095789	0.017118	5.595619	0.0000
ARE(1)	-0.005668	0.021150	-0.268006	0.7887
ARE(2)	-0.043434	0.019971	-2.174794	0.0296
ARE(3)	-0.040139	0.020357	-1.971747	0.0486
ARE(4)	-0.005623	0.019917	-0.282317	0.7777
ARE(5)	-0.047659	0.019890	-2.396167	0.0166
OMEGA(1)	0.005425	0.002203	2.462696	0.0138
ALPHA(1)	0.058842	0.008711	6.754815	0.0000
BETA(1)	0.938982	0.008484	110.6733	0.0000
TDF(1)	8.338986	1.152180	7.237570	0.0000
Log likelihood	-3570.416	Akaike info criterion	2.851785	
Avg. log likelihood	-1.421910	Schwarz criterion	2.874996	
Number of Coefs.	10	Hannan-Quinn criter.	2.860210	

Figure 25: Log-likelihood for GJR (NIKKEI225)

LogL: LL2
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:33
 Sample: 1/13/1995 12/31/2004
 Included observations: 2511
 Evaluation order: By observation
 Failure to improve Likelihood after 3 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	0.085359	0.016961	5.032668	0.0000
ARE(1)	-0.004505	0.020729	-0.217327	0.8280
ARE(2)	-0.043413	0.019462	-2.230658	0.0257
ARE(3)	-0.036655	0.019689	-1.861688	0.0626
ARE(4)	-0.007334	0.019329	-0.379407	0.7044
ARE(5)	-0.047255	0.019166	-2.465610	0.0137
OMEGA(1)	0.006445	0.001734	3.716666	0.0002
ALPHA(1)	0.002853	0.004367	0.653246	0.5136
BETA(1)	0.953653	0.006880	138.6054	0.0000
GAMMA(1)	0.067760	0.010090	6.715294	0.0000
TDF(1)	9.307245	1.321933	7.040632	0.0000
Log likelihood	-3543.240	Akaike info criterion	2.830936	
Avg. log likelihood	-1.411087	Schwarz criterion	2.856469	
Number of Coefs.	11	Hannan-Quinn criter.	2.840204	

Figure 26: Log-likelihood for EGARCH (NIKKEI225)

LogL: LL3
 Method: Maximum Likelihood (BHHH)
 Date: 04/25/10 Time: 16:35
 Sample: 1/13/1995 12/31/2004
 Included observations: 2511
 Evaluation order: By observation
 Failure to improve Likelihood after 127 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
M(1)	47.17346	693.2812	0.068044	0.9458
ARE(1)	92.43422	1355.957	0.068169	0.9457
ARE(2)	-32.95817	493.5092	-0.066783	0.9468
ARE(3)	-8.482755	122.2264	-0.069402	0.9447
ARE(4)	1.519235	22.72387	0.066856	0.9467
ARE(5)	3.417582	51.47944	0.066387	0.9471
OMEGA(1)	-0.012623	0.014512	-0.869800	0.3844
BETA(1)	0.973283	0.009012	107.9998	0.0000
ALPHA(1)	0.000530	0.007792	0.068055	0.9457
GAMMA(1)	0.000732	0.010816	0.067721	0.9460
TDF(1)	11.44423	1.782029	6.422023	0.0000
Log likelihood	-3573.197	Akaike info criterion	2.854797	
Avg. log likelihood	-1.423018	Schwarz criterion	2.880330	
Number of Coefs.	11	Hannan-Quinn criter.	2.864064	

Out-of sample results

In the out-of-sample analysis, we will use the tests presented in Section II to identify the best conditional heteroskedasticity model (GARCH, GJR or EGARCH). Specifically, we will use the Diebold-Mariano (1995), the modified Diebold-Mariano (1997), the modified Morgan-Granger-Newbold (1997), the Harvey-Leybourne-Newbold (1998), the Harvey-Newbold (2000) and the Hansen (2005) tests.

It is important to note that these tests are measured in terms of their forecasting errors. Therefore, in order to choose the best model, we will choose the lowest test value, as we want to minimize forecasting errors.

According to the original Diebold-Mariano test (Tables 15 and 16), the conclusions are rather mixed.

Table 15: Diebold-Mariano test (MSE loss functions)

	MSE		
	<i>GJR-EGARCH</i>	<i>GJR-GARCH</i>	<i>EGARCH-GARCH</i>
CAC40	-0,6331 [0,5266]	-2,2020 [0,0267]	-1,5028 [0,1329]
FTSE100	-0,6058 [0,5447]	-2,2895 [0,0221]	-0,4768 [0,6335]
NIKKEI225	-0,8069 [0,4197]	-1,5563 [0,1196]	-1,5865 [0,1126]
SP500	-2,0376 [0,0416]	-1,7054 [0,0881]	1,6137 [0,1066]

Table 16: Diebold-Mariano test (MAE loss functions)

	MAE		
	<i>GJR-EGARCH</i>	<i>GJR-GARCH</i>	<i>EGARCH-GARCH</i>
CAC40	2,2929 [0,0219]	-3,0082 [0,0026]	-3,7035 [0,0002]
FTSE100	3,3451 [0,0008]	-4,9535 [0,0000]	-4,7296 [0,0000]
NIKKEI225	2,5920 [0,0095]	-0,3429 [0,7317]	-3,9617 [0,0001]
SP500	2,8888 [0,0039]	-4,0828 [0,0000]	-3,6106 [0,0003]

In terms of Mean Squared Errors (MSE) loss function, we point out two notes. First, GJR appears to beat GARCH, as we reject the null hypothesis that there is no difference in the accuracy of two competing forecasts and the test value is negative. Still, this is only true for CAC40 and FTSE100 (at 5% significance level) and SP500 (at 10% significance level). For NIKKEI225 the difference in the predictive accuracy is not statistically significant.

Second, GJR only beats EGARCH for the SP500 stock index. Thus, GJR seems to be the most appropriate volatility forecasting model, although this conclusion cannot be generalized to all the stock indexes under analysis. Also important is the fact that, in spite of the loss function being always lower for EGARCH (when compared to the symmetric GARCH), the differences are not statistically significant.

In terms of Absolute Squared Errors (MAE) loss function, the conclusions are somewhat clearer. As we can see in Table 16, with one exception, the null is always rejected, indicating accuracy differences among the competing models. The results point out to EGARCH as the best model to predict volatility, followed by GJR and lately, GARCH.

Without surprises, the values obtained for the modified Diebold-Mariano test (Tables 17 and 18) are practically the same of those we've found for the original test. Thus, the conclusions are exactly the same.

Table 17: Modified Diebold-Mariano test (MSE loss functions)

	MSE		
	<i>GJR-EGARCH</i>	<i>GJR-GARCH</i>	<i>EGARCH-GARCH</i>
CAC40	-0,6329 [0,5268]	-2,2011 [0,0277]	-1,5022 [0,1330]
FTSE100	-0,6055 [0,5448]	-2,2886 [0,0221]	-0,4766 [0,6336]
NIKKEI225	-0,8066 [0,4199]	-1,5556 [0,1198]	-1,5859 [0,1128]
SP500	-2,0368 [0,0417]	-1,7047 [0,0883]	1,6130 [0,1067]

Table 18: Modified Diebold-Mariano test (MAE loss functions)

	MAE		
	<i>GJR-EGARCH</i>	<i>GJR-GARCH</i>	<i>EGARCH-GARCH</i>
CAC40	2,2929 [0,0219]	-3,0082 [0,0026]	-3,7035 [0,0002]
FTSE100	3,3451 [0,0008]	-4,9535 [0,0000]	-4,7296 [0,0000]
NIKKEI225	2,5920 [0,0095]	-0,3429 [0,7317]	-3,9617 [0,0001]
SP500	2,8888 [0,0039]	-4,0828 [0,0000]	-3,6106 [0,0003]

The values' resemblance can be easily understood. Looking at equation (10) we conclude that the difference between these two tests depends on the number of h-step-ahead forecasts assumed. Since we use h=1, we can rewrite equation (10):

$$S_1^* = \left[\frac{n-1}{n} \right]^{\frac{1}{2}} S_1, \tag{35}$$

Given that, $\lim_{n \rightarrow \infty} \left(\frac{n-1}{n} \right)^{1/2} = 1$ then, $\lim_{n \rightarrow \infty} S_1^* = S_1$.

As in the original and modified Diebold-Mariano tests, the null hypothesis of the modified MGN test also states that there is no difference in the accuracy of two competing forecasts.

Table 19: Modified MGN test

	<i>GARCH-EGARCH</i>	<i>GARCH-GJR</i>	<i>EGARCH-GJR</i>
CAC40	0,0117 [0,9906]	0,0006 [0,9995]	0,0022 [0,9982]
FTSE100	0,0018 [0,9986]	0,0000 [1,0000]	0,0030 [0,9976]
NIKKEI225	0,0061 [0,9951]	0,1072 [0,9146]	0,0033 [0,9974]
SP500	-0,0008 [0,9993]	0,0000 [1,0000]	0,0002 [0,9999]

Table 19 allow us to observe that, regardless the index, the null is never rejected. Then, we can conclude that GJR, GARCH and EGARCH are equally accurate, in terms of volatility prediction.

Harvey et al. (1998) proposed two tests to compare predictive ability of competing models. Let us discuss the results of those two tests (R_1 and R_2) individually.

Table 20: R_1 test

	R1					
	<i>GARCH-EGARCH</i>	<i>GARCH-GJR</i>	<i>EGARCH-GJR</i>	<i>EGARCH-GARCH</i>	<i>GJR-GARCH</i>	<i>GJR-EGARCH</i>
CAC40	0,2689 [0,7880]	0,3234 [0,7464]	0,1201 [0,9044]	-0,0401 [0,9680]	-0,2493 [0,8031]	-0,0133 [0,9894]
FTSE100	0,1611 [0,8720]	0,0313 [0,9750]	0,1626 [0,8708]	0,0615 [0,9510]	-0,0228 [0,9818]	0,0275 [0,9781]
NIKKEI225	17,8633 [0,0000]	4,4203 [0,0000]	0,3680 [0,7129]	-5,6295 [0,0000]	-2,8792 [0,0040]	-0,0120 [0,9904]
SP500	0,0509 [0,9594]	0,7682 [0,4424]	0,4368 [0,6622]	0,5020 [0,6157]	-0,4170 [0,6767]	-0,0476 [0,9620]

Table 21: R_2 test

	R2					
	<i>GARCH-EGARCH</i>	<i>GARCH-GJR</i>	<i>EGARCH-GJR</i>	<i>EGARCH-GARCH</i>	<i>GJR-GARCH</i>	<i>GJR-EGARCH</i>
CAC40	0,2119 [0,8322]	0,2444 [0,8069]	0,1365 [0,8914]	-0,0386 [0,9692]	-0,1996 [0,8418]	0,0132 [0,9895]
FTSE100	0,1387 [0,8897]	0,0304 [0,9758]	0,1941 [0,8461]	0,0655 [0,9477]	-0,0223 [0,9822]	0,0267 [0,9787]
NIKKEI225	0,9470 [0,3436]	1,2924 [0,1962]	0,5823 [0,5604]	-0,8492 [0,3958]	-1,5321 [0,1255]	-0,0121 [0,9903]
SP500	0,0484 [0,9614]	0,4345 [0,6640]	0,7757 [0,4379]	1,0080 [0,3135]	-0,2943 [0,7685]	-0,0500 [0,9601]

A quick view over Tables 20 and 21 show a quite few contradictory results. Let us use the comparison of GARCH and EGARCH to illustrate that. Concerning to F_1 test for the FTSE100 index, the value of the first row tell us that EGARCH encompasses GARCH, due to the positive and statistically significant test value. Logically, if EGARCH encompasses GARCH, then GARCH cannot encompass EGARCH. However, that's the exact conclusion to be taken of the fourth row, where we have a positive and statistically significant value. Instead, we should have a negative value. Having this in mind, let us proceed to a more detailed analysis of the results.

Regarding R_1 , the conclusions are quite contradictory. For the CAC40 index, the positive and statistically significant values obtained in the three first rows indicate GJR as the best model, since it encompasses both GARCH and EGARCH. Also, EGARCH encompasses GARCH. The negative and statistically significant values obtained in the last three rows confirm this idea. For the other three indexes, the conclusions are not that clear. Specifically, for the FTSE100 index, we can only conclude that GJR encompasses GARCH, while for the S&P500 index, GJR seems to beat both GARCH and EGARCH. For the NIKKEI225 index, the most apparent result is GJR encompassing EGARCH.

Regarding R_2 test, the most obvious outcome is the statistical significance of every single test value, indicating noteworthy differences among our three models. For the CAC40 index, the only conclusion to take is the fact that GARCH is the weakest model, since it is encompassed by the other two. For the FTSE100 and S&P500 indexes, the conclusions made for the R_1 test can be applied for the R_2 test. Specifically, for the FTSE100 index, we conclude that GJR encompasses GARCH, while for the S&P500 index, GJR beats both GARCH and EGARCH. Lastly, the results obtained for the NIKKEI225 lead us to conclude that GJR is the best model, since it encompasses both GARCH and EGARCH.

In relation to the Harvey-Newbold tests, Tables 22-24 show us that all three tests perform quite similarly.

Table 22: F_1 test

	F1		
	GARCH	EGARCH	GJR
CAC40	12,8675 [0,0003]	5,7174 [0,0169]	8,2504 [0,0041]
FTSE100	7,7096 [0,0056]	5,5814 [0,0183]	4,0676 [0,0439]
NIKKEI225	8,9479 [0,0028]	5,0701 [0,0245]	7,2534 [0,0072]
SP500	5,5736 [0,0184]	1,4778 [0,0001]	1,9524 [0,1626]

Table 23: F_2 test

	F2		
	GARCH	EGARCH	GJR
CAC40	9,5509 [0,0020]	4,1623 [0,0415]	6,6718 [0,0099]
FTSE100	6,8349 [0,0090]	4,5132 [0,0338]	3,8269 [0,0507]
NIKKEI225	7,0985 [0,0078]	3,7428 [0,0533]	6,2320 [0,0127]
SP500	4,8139 [0,0284]	1,1965 [0,0006]	1,7975 [0,1803]

Table 24: MS^* test

	MS*		
	GARCH	EGARCH	GJR
CAC40	9,6153 [0,0020]	4,1726 [0,0413]	6,7015 [0,0097]
FTSE100	6,8666 [0,0089]	4,5258 [0,0336]	3,8355 [0,0504]
NIKKEI225	7,1340 [0,0077]	3,7512 [0,0530]	6,2588 [0,0125]
SP500	4,8285 [0,0282]	1,2070 [0,0005]	1,7986 [0,1801]

Concerning F_1 test, we reject the null hypothesis for the three models (at a 5% significance level), meaning that neither of them encompasses another. Consequently, any of them can be improved when combined with the two other models. This is true for CAC40, FTSE100 and NIKKEI225. For S&P500, however, we fail to reject the null hypothesis for GJR, meaning that this model encompasses the other two. Then, GARCH and EGARCH contain no useful information not present in GJR.

For F_2 and MS^* tests, the conclusions are exactly the same. Concerning CAC40 and FTSE100 indexes, we always reject the null hypothesis that one model encompasses another (at a 5% significance level). Therefore, as we already seen for F_1 test, GARCH, EGARCH and GJR models can be improved when combined among themselves. In relation to NIKKEI225, it is possible to see that both F_2 and MS^* tests exhibit a p-value slightly superior to the considered 5% significance level. Then, failing to reject the null hypothesis, we conclude that EGARCH encompasses GARCH and GJR. However, a more flexible approach can lead us to another conclusion.

In fact, it is important to see that we could easily reject the null hypothesis, if using a slightly superior significance level (6% for example). Therefore, although we can say that EGARCH encompasses the other two models, we don't expect those differences to be that substantial. Finally, like it happened for F_1 , so too F_2 and MS^* fail to reject the null hypothesis for GJR when using S&P500. As a result, we can say that all three Harvey-Newbold tests point out the GJR as the best model to predict S&P500 volatility.

At last, we have Hansen (2005) test, which results are shown in Table 25.

Table 25: Hansen test

	<i>GARCH-EGARCH</i>	<i>GARCH-GJR</i>	<i>EGARCH-GJR</i>
CAC40	0,0000 [0,9072]	0,9895 [0,0446]	5,03639 [0,0000]
FTSE100	0,0000 [0,7391]	0,0000 [0,9222]	1,2577 [0,0000]
NIKKEI225	0,0000 [0,9141]	6,5472 [0,0000]	7,0252 [0,0000]
SP500	0,0000 [0,7352]	0,0000 [0,9001]	1,2291 [0,0000]

As we fail to reject the null hypothesis that the benchmark (GARCH) is not inferior to EGARCH, we conclude that these two models are equally able to predict CAC40, FTSE100, NIKKEI225 and S&P500 volatility. When the comparison involves EGARCH (benchmark) and GJR, we conclude that the latter outperforms the first, since we reject the null hypothesis and we have a positive value for the test statistic. This conclusion is valid for all four indexes. Lastly, we stumble on two opposite remarks, when comparing GARCH and GJR models. In one hand, the results achieved for CAC40 and NIKKEI225 lead us to conclude that GJR beats GARCH (benchmark), as we reject the null hypothesis and the test value is positive. On the other hand, as we don't reject the null hypothesis, we can say that GJR and GARCH are equally competent to predict FTSE100 and S&P500 volatility.

7. CONCLUSION

The main purpose of this thesis is to compare the forecasting capability of some alternative conditional heteroskedasticity models (GARCH, EGARCH and GJR). Our data base is formed by the daily closing prices of the CAC40, FTSE100, NIKKEI 225 and S&P500 indexes from January 1, 1995 through December 31, 2009. This sample was then divided in two: the in-sample (January 3, 1995 to December 31, 2004) and the out-of-sample (January 3, 2005 to December 31, 2009). Next, we will present the main conclusions for both in-sample and out-of-sample analyses.

In relation to the in-sample analyses, we used three likelihood based goodness-of-fit criteria to compare our models. Those criteria are the maximum log-likelihood, the AIC and the SBC. As expected, due to the leverage effect, all these criteria pointed out the asymmetric models (EGARCH and GJR) as the best predictors of volatility. Besides, EGARCH seems to be the most accurate volatility predictor, as it beats GJR in every index, excepting S&P500.

In the out-of-sample analysis, the results are not that clear, since we reach different results for different indexes and different tests. Still, there are some general conclusions that can be made. First, with exception of the modified Diebold-Mariano (1997) test and the F_1 test proposed by Harvey and Newbold (2000), all these tests point out the existence of accuracy differences among models, indicating that there are some better than others. Then, as already seen in the in-sample analyses, the asymmetric models seem to outperform GARCH. The major problem is then to choose between GJR and EGARCH.

Regarding the Diebold-Mariano (1995) and the modified Diebold-Mariano (1997) tests, EGARCH beats GJR, who beats GARCH, in terms of MAE loss function. In terms of MSE loss function, however, this fact does not apply. Moreover, it does not apply to the NIKKEI225 index too. The Harvey and Newbold (2000) F_2 and MS^* tests, also seem to favor EGARCH, although the accuracy differences are not categorical. On the other hand, both Harvey-Leybourne-Newbold (1998) appear to choose GJR as the best model.

Unfortunately then, we were not able to point out any model as the absolute best accurate predictor of volatility. Still, it is fair to say that the dataset analyzed contains several episodes of regional and global “market stress”, who provoked high volatilities.

Although we want a volatility model to be able to forecast volatility in these extreme situations, we are aware of the difficulties. Still, nowadays' models already incorporate a lack of several stylized facts, being useful tools, although not perfect, to predict volatility.

8. THESIS LIMITATIONS AND FUTURE RESEARCH

In our empirical application, we used the daily squared returns as a proxy of daily volatility. However, as found by Andersen and Bollerslev (1998) squared returns are a noisy estimator. Thus, in the future we will have to consider an alternative estimator such as the realized variance based on intraday volatility.

Another point for future research is to analyze which test is statistically more appropriate for a particular models comparison. For example, in accordance to Hansen and Lunde (2005), when the comparison involves nested models it is more appropriate to apply a test for equal predictive accuracy (EPA), such as that of Harvey and Newbold (2000) when compared to the SPA, for example. Thus, future research should involve the characteristics of the models in order to select the most appropriate test to compare the models forecasting accuracy.

9. BIBLIOGRAPHY:

Akaike, H. (1978), *Time series analysis and control through parametric models*, Applied Time Series, D.F Findley (ed.), Academic Press, New York.

Andersen, T.G. and Bollerslev, T. (1998), Answering the skeptics: yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885-905.

Black, F. (1976), Studies of stock price volatility changes, Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, *American Statistical Association*, 177-181.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307-327.

Bollerslev, T., Chou, R. and Kroner, K. (1992), Arch modeling in finance, *Journal of Econometrics* 52, 5-59.

Bollerslev, T., R.F. Engle and D.B. Nelson (1994), "ARCH Models," in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics, Volume IV*, 2959-3038. Amsterdam: North-Holland.

Diebold, F. X. and Mariano, R.S. (1995), Comparing Predictive Accuracy, *Journal of Business & Economic Statistics* 13, 253-263.

Engle, R.F. (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987-1007.

Fama, E.F. (1965), The behavior of stock prices, *Journal of Business* 38, 34-105.

Glosten, L., Jagannathan, R. and Runkle, R. (1993), On the relationship between the expected value and the volatility of the nominal excess returns on stocks, *Journal of Finance* 48, 1779-1801.

Granger, C.W.J., and Newbold, P. (1973), Some comments on the evaluation of economic forecasts, *Applied Economics* 5, 35-47.

Granger, C.W.J., and Newbold, P. (1977), *Forecasting economic time series*, Orlando, FL: Academic Press.

Granger, C.W.J., and Newbold, P. (1986), *Forecasting economic time series* (2nd ed.), Orlando, FL: Academic Press.

Hansen, P.R. and Lunde, A. (2003), Consistent preordering with an estimated criterion function, with an application to the evaluation and comparison of volatility models, *Brown University Working Paper*.

Hansen, P.R. and Lunde, A. (2005), A forecast comparison of volatility models: does anything beat a GARCH (1,1)?, *Journal of Applied Econometrics* 20, 873-889.

Hansen, P.R. (2005), A test for superior predictive ability, *Journal of Business and Economic Statistics* 23, 365-380.

Harvey, D.I., Leybourne, S.J. and Newbold P. (1997), Testing the equality of prediction mean squared errors, *International Journal of Forecasting* 13, 281-291.

Harvey, D.I., Leybourne, S.J. and Newbold P. (1998), Tests for forecast encompassing, *Journal of Business and Economic Statistics* 16, 254-259.

Harvey, D.I and Newbold P. (2000), Tests for multiple forecast encompassing, *Journal of Applied Econometrics* 15, 471-482.

Hotelling H. (1931), The generalization of Student's ratio, *Annals of Mathematical Statistics* 2, 360-378.

Jarque, C.M. and Bera, A.K (1987), A test for normality of observations and regression residuals, *International Statistics Review* 55 (2), 163-172.

Mandelbrot, B. (1963), The variation of certain speculative prices, *Journal of Business* 36, 394-419.

Meese, R.A., Rogoff, K. (1988), Was it real? The exchange rate-interest differential relation over the modern floating-rate period, *Journal of Finance* 43, 933-948.

Morgan, W.A. (1939-1940), A test for significance of the difference between the two variances in a sample from a normal bivariate population, *Biometrika* 31, 13-19.

Nelsen, D.B. (1991), Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347-370.

Patton, A.J. (2006), Volatility forecast comparison using imperfect volatility proxies, Research Paper 175, *Quantitative Finance Research Centre*, University of Technology, Sydney.

Scharwz, G. (1978), Estimating the dimension of a model, *Annals of Statistics* 6, 461-464.

White, H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817-838.

White, H. (1984), *Asymptotic theory for econometricians*, Orlando, FL: Academic Press.

White, H. (2000), A reality check for data snooping, *Econometrica* 68, 1097-1126

Zakoian, J-M. (1994) Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* 18, 931-955.

Appendix:

Table 26: Empirical size under quadratic loss, Test Statistic F

<i>T</i>	ρ	<i>Gaussian</i>			<i>Fat-tailed</i>		
		$\theta = .0$	$\theta = .5$	$\theta = .9$	$\theta = .0$	$\theta = .5$	$\theta = .9$
8	.0	9.85	12.14	14.10	14.28	15.76	17.21
8	.5	7.02	9.49	11.42	9.61	11.64	13.02
8	.9	.58	1.26	1.86	.57	1.13	1.79
16	.0	9.83	12.97	14.85	16.47	18.59	19.78
16	.5	7.30	10.11	11.89	11.14	13.55	14.94
16	.9	.47	.99	1.55	.34	.70	1.13
32	.0	9.88	12.68	14.34	18.06	19.55	20.35
32	.5	6.98	9.50	11.22	21.30	21.00	21.37
32	.9	.23	.55	1.00	.01	.07	.23
64	.0	9.71	13.05	14.62	29.84	29.72	29.96
64	.5	6.48	9.25	10.62	23.48	23.93	24.15
64	.9	.16	.47	.79	.02	.12	.29
128	.0	10.30	13.41	14.99	30.34	30.95	31.26
128	.5	7.01	10.13	11.64	24.89	25.01	25.16
128	.9	.16	.50	.74	.11	.44	.73
256	.0	10.01	13.05	14.65	31.07	31.12	31.24
256	.5	7.37	10.31	11.78	25.48	25.45	25.70
256	.9	.19	.51	.80	.51	1.13	1.44
512	.0	10.22	13.51	15.25	31.45	32.38	32.60
512	.5	7.53	10.16	11.49	26.35	26.92	16.95
512	.9	.18	.50	.85	.81	1.58	2.06

NOTE: *T* is sample size, ρ is the contemporaneous correlation between the innovations underlying the forecast errors, and θ is the coefficient of the MA(1) forecast error. All tests are at the 10% level. 10,000 Monte Carlo replications are performed.

Source: (Diebold-Mariano, 1995:258)

Table 27: Empirical size under quadratic loss, Test Statistic MGN

<i>T</i>	ρ	<i>Gaussian</i>			<i>Fat-tailed</i>		
		$\theta = .0$	$\theta = .5$	$\theta = .9$	$\theta = .0$	$\theta = .5$	$\theta = .9$
8	.0	10.19	14.14	17.94	18.10	21.89	25.65
8	.5	9.96	14.66	18.61	16.00	20.51	24.19
8	.9	9.75	14.53	18.67	11.76	16.31	20.00
16	.0	10.07	14.34	17.54	20.33	24.54	27.08
16	.5	9.56	14.37	17.95	37.15	36.18	25.66
16	.9	10.02	14.70	18.20	12.01	16.76	19.81
32	.0	9.89	15.04	18.00	22.94	26.32	28.72
32	.5	10.08	15.11	17.95	20.23	23.76	26.20
32	.9	9.59	15.32	18.25	12.75	17.78	20.54
64	.0	10.09	15.37	17.99	24.56	28.15	30.00
64	.5	9.95	15.18	18.15	21.10	25.18	27.28
64	.9	10.26	15.67	18.49	12.98	18.09	20.53
128	.0	9.96	15.09	17.59	26.47	29.50	30.94
128	.5	10.23	15.07	17.48	23.62	26.82	28.51
128	.9	10.11	15.05	18.05	14.34	18.89	21.56
256	.0	10.28	15.62	18.37	27.39	30.74	32.46
256	.5	10.60	16.02	18.44	23.81	28.38	30.31
256	.9	10.11	15.48	17.91	14.15	19.43	22.03
512	.0	10.12	15.34	17.68	27.64	30.55	32.14
512	.5	10.05	14.96	17.66	24.10	27.40	29.28
512	.9	9.90	15.09	17.53	14.78	19.16	21.49

NOTE: *T* is sample size, ρ is the contemporaneous correlation between the innovations underlying the forecast errors, and θ is the coefficient of the MA(1) forecast error. All tests are at the 10% level. 10,000 Monte Carlo replications are performed.

Source: (Diebold-Mariano, 1995:258)

Table 28: Empirical size under quadratic loss, Test Statistic MR

<i>T</i>	ρ	<i>Gaussian</i>			<i>Fat-tailed</i>		
		$\theta = .0$	$\theta = .5$	$\theta = .9$	$\theta = .0$	$\theta = .5$	$\theta = .9$
8	.0	9.67	19.33	22.45	16.16	25.26	27.62
8	.5	9.50	19.00	22.07	14.81	24.50	26.99
8	.9	9.66	19.51	22.85	11.23	21.28	24.14
16	.0	9.62	13.92	14.72	19.94	22.56	23.06
16	.5	10.02	13.88	14.96	17.70	21.04	21.26
16	.9	10.04	13.82	14.94	11.76	15.68	16.70
32	.0	9.96	10.98	11.12	22.78	22.86	21.72
32	.5	9.68	11.46	11.66	19.78	20.32	20.14
32	.9	9.86	11.62	11.96	12.42	13.54	13.46
64	.0	10.32	11.02	11.04	24.50	22.60	21.58
64	.5	9.84	10.56	10.64	21.44	19.48	18.84
64	.9	9.58	10.58	10.34	13.38	13.38	13.20
128	.0	9.78	10.54	10.44	25.86	22.90	21.54
128	.5	10.02	11.04	11.18	22.76	20.26	19.44
128	.9	10.76	11.28	11.38	13.44	13.52	12.92
256	.0	10.04	9.90	9.58	27.16	23.74	22.70
256	.5	10.32	9.92	9.82	24.00	20.50	19.18
256	.9	9.92	10.16	10.34	13.38	12.70	12.24
512	.0	9.94	10.48	10.56	26.92	23.40	21.78
512	.5	9.52	10.56	10.48	23.56	20.52	19.36
512	.9	9.80	9.82	9.88	13.96	12.98	12.74

NOTE: *T* is sample size, ρ is the contemporaneous correlation between the innovations underlying the forecast errors, and θ is the coefficient of the MA(1) forecast error. All tests are at the 10% level. At least 5,000 Monte Carlo replications are performed.

Source: (Diebold-Mariano, 1995:259)

Table 29: Empirical size under quadratic loss, Test Statistic S_1

T	ρ	<i>Gaussian</i>			<i>Fat-tailed</i>		
		$\theta = .0$	$\theta = .5$	$\theta = .9$	$\theta = .0$	$\theta = .5$	$\theta = .9$
8	.0	31.39	31.10	31.03	31.62	29.51	29.07
8	.5	31.37	30.39	29.93	31.21	29.71	29.36
8	.9	31.08	30.19	30.18	31.18	30.12	29.75
16	.0	20.39	19.11	18.94	19.26	18.50	18.32
16	.5	20.43	19.52	18.86	19.57	17.67	17.63
16	.9	20.90	19.55	19.59	20.15	18.38	18.16
32	.0	12.42	12.28	12.18	11.30	11.64	11.56
32	.5	13.32	13.22	12.94	11.54	10.66	10.84
32	.9	12.60	13.38	13.22	11.16	11.22	11.50
64	.0	12.47	12.11	11.94	12.44	11.62	11.36
64	.5	12.76	12.49	12.35	12.10	12.26	12.10
64	.9	12.21	12.23	12.03	13.00	12.36	12.16
128	.0	11.72	11.94	12.04	11.48	10.72	10.28
128	.5	11.44	11.72	11.60	10.84	10.96	10.96
128	.9	11.76	11.26	11.34	11.50	10.66	10.86
256	.0	11.11	10.65	10.66	12.06	11.67	11.79
256	.5	10.90	10.39	10.48	12.16	11.46	11.60
256	.9	10.69	10.79	10.75	11.51	11.59	11.16
512	.0	11.15	10.67	10.63	10.06	9.46	9.62
512	.5	10.90	10.39	10.49	9.94	9.66	9.76
512	.9	10.31	10.09	10.05	10.12	10.12	10.06

NOTE: T is sample size, ρ is the contemporaneous correlation between the innovations underlying the forecast errors, and θ is the coefficient of the MA(1) forecast error. All tests are at the 10% level. At least 5,000 Monte Carlo replications are performed.

Source: (Diebold-Mariano, 1995:259)

Table 30: Empirical size under quadratic loss, Test Statistics S_2 and S_{2a}

T	ρ	<i>Gaussian</i>			<i>Fat-tailed</i>		
		$\theta = .0$	$\theta = .5$	$\theta = .9$	$\theta = .0$	$\theta = .5$	$\theta = .9$
<i>S₂, nominal size = 25%</i>							
8	.0	22.24	22.48	22.38	23.94	23.46	23.34
8	.5	22.14	23.46	22.16	23.08	24.80	23.06
8	.9	22.24	23.02	22.66	22.92	23.26	22.86
<i>S₂, nominal size = 14.08%</i>							
16	.0	13.46	13.26	13.14	13.62	13.06	13.76
16	.5	14.22	13.46	12.92	13.70	13.24	13.62
16	.9	13.08	13.84	13.28	12.86	13.06	13.20
<i>S₂, nominal size = 15.36%</i>							
32	.0	14.36	14.52	14.28	14.54	14.32	14.30
32	.5	14.36	14.06	13.94	15.08	14.36	15.02
32	.9	14.68	14.62	13.46	14.94	14.76	14.52
<i>S_{2a}, nominal size = 10%</i>							
64	.0	9.72	9.92	9.42	9.68	10.36	10.44
64	.5	9.66	10.34	9.68	9.52	10.06	10.00
64	.9	10.84	9.46	10.34	9.40	8.98	10.02
<i>S_{2a}, nominal size = 10%</i>							
128	.0	11.62	11.62	11.84	12.22	12.20	11.42
128	.5	11.66	11.62	11.90	12.06	11.94	11.44
128	.9	11.22	11.72	11.28	12.06	10.76	11.40

NOTE: T is sample size, ρ is the contemporaneous correlation between the innovations underlying the forecast errors, and θ is the coefficient of the MA(1) forecast error. At least 5,000 Monte Carlo replications are performed.

Source: (Diebold-Mariano, 1995:260)

Table 31: Empirical size under quadratic loss, Test Statistics S_3 and S_{3a}

T	ρ	<i>Gaussian</i>			<i>Fat-tailed</i>		
		$\theta = .0$	$\theta = .5$	$\theta = .9$	$\theta = .0$	$\theta = .5$	$\theta = .9$
<i>S₃, nominal size = 25%</i>							
8	.0	22.50	22.92	22.90	23.26	23.34	21.96
8	.5	22.98	22.26	23.06	23.42	23.86	22.88
8	.9	23.16	22.36	24.24	24.26	23.32	23.34
<i>S₃, nominal size = 10.92%</i>							
16	.0	10.62	10.06	10.40	10.16	10.42	9.84
16	.5	10.38	10.92	10.32	10.54	10.94	10.34
16	.9	10.64	10.18	9.62	10.58	10.96	10.64
<i>S₃, nominal size = 10.12%</i>							
32	.0	10.72	10.28	9.30	9.90	10.00	9.98
32	.5	10.56	10.00	10.02	10.40	10.64	10.30
32	.9	10.92	10.44	10.30	10.46	9.96	10.70
<i>S_{3a}, nominal size = 10%</i>							
64	.0	9.38	9.54	9.16	9.64	9.24	8.84
64	.5	9.80	10.02	9.66	9.58	8.82	8.78
64	.9	9.90	9.24	9.68	9.92	9.78	10.00
<i>S_{3a}, nominal size = 10%</i>							
128	.0	9.94	9.70	9.12	9.82	9.04	8.46
128	.5	9.52	10.00	9.32	10.08	9.24	9.20
128	.9	9.46	9.64	9.42	9.28	9.22	9.26

NOTE: T is sample size, ρ is the contemporaneous correlation between the innovations underlying the forecast errors, and θ is the coefficient of the MA(1) forecast error. At least 5,000 Monte Carlo replications are performed.

Source: (Diebold-Mariano, 1995:260)

Program workfile:

For the CAC40 index:

(the exact same calculations were made to the other three indexes)

```
load "cac40"
```

```
series rt=return
```

```
series rt2=rt^2
```

```
sample s0 1995/01/03 1995/01/16
```

```
sample s1 1995/01/13 2004/12/31
```

```
sample s2 2005/01/03 2009/12/31
```

```
smpl s1
```

```
'GARCH GARCH GARCH
```

```
equation eq1
```

```
eq1.arch rt c
```

```
'show eq1.output
```

```
'declare coef vectors to use in the likelihood
```

```
coef(1) m = eq1.c(1)
```

```
coef(1) omega = eq1.c(2)
```

```
coef(1) alpha = eq1.c(3)
```

```
coef(1) beta = eq1.c(4)
```

```
coef(4) gamma = 0
```

```
coef(3) delta = 0
```

```
coef(1) tdf = 3
```

```
coef(5) are = 0
```

```
smpl s0
```

```
series res = rt-m(1)
```

```
series siggarch = @var(rt)
```

```
!pi = @acos(-1)
```

```
'set up the likelihood GARCH
```

```
logl ll1
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
ll1.append @logl logl
```

```
ll1.append res = rt-m(1)-are(1)*rt(-1)-are(2)*rt(-2)-are(3)*rt(-3)-are(4)*rt(-4)-are(5)*rt(-5)
```

```
'set up the likelihood GARCH
```

```
ll1.append siggarch = omega(1)+alpha(1)*res(-1)^2+beta(1)*siggarch(-1)
```

```
'likelihood equation
```

```
ll1.append z =res^2/siggarch/(tdf(1)-2)+1
```

```
ll1.append logl = @gammalog((tdf(1)+1)/2) - @gammalog(tdf(1)/2) - @log(!pi)/2 - @log(tdf(1)-2)/2 -  
@log(siggarch)/2-(tdf(1)+1)*@log(z)/2
```

```
'estimate and display results
```

```
smpl s1
```

```
ll1.ml(b)
```

```
'show ll1.output
```

```
'conditional variance out-of-sample GARCH
```

```
smpl s2
```

```
res=rt-m(1)-are(1)*rt(-1)-are(2)*rt(-2)-are(3)*rt(-3)-are(4)*rt(-4)-are(5)*rt(-5)
```

```
siggarch=omega(1)+alpha(1)*res(-1)^2+beta(1)*siggarch(-1)
```

```
'MSE GJR
```

```
series dgarch1
```

```
dgarch1=(siggarch-rt2)^2
```

```
'MAE GJR
```

```
series dgarch2
```

```
dgarch2=@abs(siggarch-rt2)
```

```
'GJR GJR GJR GJR
```

```
equation eq1
```

```
eq1.arch rt c
```

```
'show eq1.output
```

```
'declare coef vectors to use in the likelihood
```

```
'coef(1) m=0.1
```

```
'coef(1) omega=0.1
```

```
'coef(1) alpha=0.1
```

```
'coef(1) beta=0.1
```

```
'coef(1) gamma=0.1
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
'coef(1) tdf=3
'coef(5) are=0.1

smpl s0
series res=rt-m(1)
series siggjr=@var(rt)
!pi=@acos(-1)

'set up the likelihood GJR
logl ll2
ll2.append @logl logl
ll2.append res=rt-m(1)-are(1)*rt(-1)-are(2)*rt(-2)-are(3)*rt(-3)-are(4)*rt(-4)-are(5)*rt(-5)

'set up the likelihood GJR
ll2.append n1=res(-1)<0
ll2.append siggjr=omega(1)+alpha(1)*res(-1)^2+beta(1)*siggjr(-1)+gamma(1)*res(-1)^2*n1

'likelihood equation
ll2.append z=res^2/siggjr/(tdf(1)-2)+1
ll2.append logl=@gammalog((tdf(1)+1)/2)-@gammalog(tdf(1)/2)-@log(!pi)/2-@log(tdf(1)-2)/2-
@log(siggjr)/2-(tdf(1)+1)*@log(z)/2

'estimate and display results
smpl s1
ll2.ml(b)
'show ll2.output

'conditional variance out-of-sample GJR
smpl s2
res=rt-m(1)-are(1)*rt(-1)-are(2)*rt(-2)-are(3)*rt(-3)-are(4)*rt(-4)-are(5)*rt(-5)
n1=res(-1)<0
siggjr=omega(1)+alpha(1)*res(-1)^2+beta(1)*siggjr(-1)+gamma(1)*n1*res(-1)^2
'MSE GJR
series dgjr1
dgjr1=(siggjr-rt2)^2
'MAE GJR
series dgjr2
dgjr2=@abs(siggjr-rt2)
```


Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

EGARCH EGARCH EGARCH

equation eq1

eq1.arch rt c

'show eq1.output

'declare coef vectors to use in the likelihood

'coef(1) m=eq1.c(1)

'coef(1) omega=eq1.c(2)

'coef(1) alpha=eq1.c(3)

'coef(1) beta=eq1.c(4)

'coef(4) gamma=0

'coef(3) delta=0

'coef(1) tdf=3

'coef(5) are=0

smpl s0

series sigegarch=@var(rt)

series reseg=rt-m(1)

!pi=@acos(-1)

'set up the likelihood EGARCH

logl ll3

ll3.append @logl logl

ll3.append reseg=rt-m(1)-are(1)*rt(-1)-are(2)*rt(-2)-are(3)*rt(-3)-are(4)*rt(-4)-are(5)*rt(-5)

'set up the likelihood EGARCH

ll3.append $\log(\text{sigegarch})=\omega(1)+\beta(1)*\log(\text{sigegarch}(-1))+\alpha(1)*\text{abs}(\text{reseg}(-1))/\sqrt{\text{sigegarch}(-1)}+\gamma(1)*(\text{reseg}(-1)/\sqrt{\text{sigegarch}(-1)})$

'likelihood equation

ll3.append $z=\text{res}^2/\text{sigegarch}/(\text{tdf}(1)-2)+1$

ll3.append $\text{logl}=\text{@gammalog}((\text{tdf}(1)+1)/2)-\text{@gammalog}(\text{tdf}(1)/2)-\log(!\pi)/2-\log(\text{tdf}(1)-2)/2-\log(\text{sigegarch})/2-(\text{tdf}(1)+1)*\log(z)/2$

'estimate and display results

smpl s1

ll3.ml(b)

'show ll3.output

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
'conditional variance out-of.sample EGARCH
smpl s2
reseg=rt-m(1)-are(1)*rt(-1)-are(2)*rt(-2)-are(3)*rt(-3)-are(4)*rt(-4)-are(5)*rt(-5)
sigegarch=@exp(omega(1)+beta(1)*log(sigegarch(-1))+alpha(1)*@abs(reseg(-1))/@sqrt(sigegarch(-
1))+gamma(1)*(reseg(-1)/@sqrt(sigegarch(-1))))
series degarch1
degarch1=(sigegarch-rt2)^2
'MAE EGARCH
series degarch2
degarch2=@abs(sigegarch-rt2)
```

'COMPUTING THE HARVEY-NEWBOLD (2000) TEST

```
series e1t=rt2-sigegarch
vector(1000) yt
stomna(e1t,yt)
series e2t=rt2-sigegarch
series e3t=rt2-siggjr
series de12t=e1t-e2t
series de13t=e1t-e3t
```

```
equation eq3
eq3.ls e1t de12t de13t
'show eq3.output
!obs=@obssmpl
!s2=@se^2
!K=@ncoef
eq3.makesresids res1
vector(1000) resf1
stomna(res1,resf1)
vector(!K) LAMB
for !i=1 to !K
LAMB(!i)=@coefs(!i)
next
```

```
matrix(1000,2) X
group HNg1 de12t de13t
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
stomna(HNg1,X)
matrix(1000,2) HatM
matrix(1000,2) HatQ
matrix(1000,2) HatD
HatM=(!obs^(-1))*@transpose(X)*X
HatQ=!s2*HatM
HatD=@inverse(HatM)*HatQ*@inverse(HatM)
```

'F STANDARD TEST

```
matrix(1,1) F11
F11=@transpose(lamb)*@inverse(HatD)*lamb
scalar F=!obs*(!K-1)^(-1)*@trace(F11)
if F>0 then !ProbF=1-@cdfist(F,!K-1,!obs-!K+1)
else
!ProbF=@cdfist(F,!K-1,!obs-!K+1)
endif
```

'COMPUTING F1

```
!q=0
!cm=@columns(HatM)
!rm=@rows(HatM)
matrix(!cm, !rm) HatQ1
for !i=1 to !cm
for !j=1 to !rm
for !l=1 to !obs
!q=!q+X(!l,!i)*X(!l,!j)*resf1(!l)^2
next
HatQ1(!i,!j)=!q*!obs^(-1)
!q=0
next
next
matrix(!rm, !cm) HatD1
HatD1=@inverse(HatM)*HatQ1*@inverse(HatM)

matrix(1,1) F12
F12=@transpose(lamb)*@inverse(HatD1)*lamb
scalar F1=!obs*(!K-1)^(-1)*@trace(F12)
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
if F1>0 then !ProbF1=1-@cdfdist(F1,!K-1,!obs-!K+1)
else
!ProbF1=@cdfdist(F1,!K-1,!obs-!K+1)
endif
```

```
'COMPUTING F2
```

```
!q=0
!cm=@columns(HatM)
!rm=@rows(HatM)
matrix(!cm, !rm) HatQ2
for !i=1 to !cm
for !j=1 to !rm
for !l=1 to !obs
!q=!q+X(!l,!i)*X(!l,!j)*yt(!l)^2
next
HatQ2(!i,!j)=!q*!obs^(-1)
!q=0
next
next
```

```
matrix(!rm, !cm) HatD2
```

```
HatD2=@inverse(HatM)*HatQ2*@inverse(HatM)
```

```
matrix(1,1) F13
```

```
F13=@transpose(lamb)*@inverse(HatD2)*lamb
```

```
scalar F2=!obs*(!K-1)^(-1)*@trace(F13)
```

```
if F2>0 then !ProbF2=1-@cdfdist(F2,!K-1,!obs-!K+1)
```

```
else
```

```
!ProbF2=@cdfdist(F2,!K-1,!obs-!K+1)
```

```
endif
```

```
'COMPUTING MS*
```

```
scalar MS=(!obs-(!K-1)*F2)^(-1)*(!obs-!K+1)*F2
```

```
if MS>0 then !ProbMS=1-@cdfdist(MS,!K-1,!obs-!K+1)
```

```
else
```

```
!ProbMS=@cdfdist(MS,!K-1,!obs-!K+1)
```

```
endif
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
table(7,2) resulta
setcolwidth(resulta,1,30)
resulta(1,1)="Statistics of the test"
resulta(1,2)="Test value"
resulta(1,3)="P-value"
resulta(2,1)="F-standard"
resulta(3,1)="F1"
resulta(4,1)="F2"
resulta(5,1)="MS*"
resulta(2,2)=F
resulta(2,3)=!ProbF
resulta(3,2)=F1
resulta(3,3)=!ProbF1
resulta(4,2)=F2
resulta(4,3)=!ProbF2
resulta(5,2)=MS
resulta(5,3)=!ProbMS
```

```
show resulta
```

'COMPUTING THE DIEBOLD-MARIANO TEST

```
series dmsv1=(siggarch-rt2)^2
'MAE MSV-EGARCH
series dmsv2
dmsv2=@abs(siggarch-rt2)
```

'MSE

```
series d11=dgjr1-degarch1
series d12=dgjr1-dmsv1
series d13=degarch1-dmsv1
```

```
equation eq11
eq11.ls d11 c
scalar DM11=@tstats(1)
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
if DM11<0 then !ProbDM11=2*@cnorm(DM11)
else
!ProbDM11=2*(1-@cnorm(DM11))
endif
```

```
equation eq12
eq12.ls d12 c
scalar DM12=@tstats(1)
if DM12<0 then !ProbDM12=2*@cnorm(DM12)
else
!ProbDM12=2*(1-@cnorm(DM12))
endif
```

```
equation eq13
eq13.ls d13 c
scalar DM13=@tstats(1)
if DM13<0 then !ProbDM13=2*@cnorm(DM13)
else
!ProbDM13=2*(1-@cnorm(DM13))
endif
```

'MAE

```
series d21=dgjr2-degarch2
series d22=dgjr2-dmsv2
series d23=degarch2-dmsv2
```

```
equation eq21
eq21.ls d21 c
scalar DM21=@tstats(1)
if DM21<0 then !ProbDM21=2*@cnorm(DM21)
else
!ProbDM21=2*(1-@cnorm(DM21))
endif
```

```
equation eq22
eq22.ls d22 c
scalar DM22=@tstats(1)
if DM22<0 then !ProbDM22=2*@cnorm(DM22)
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
else
!ProbDM22=2*(1-@cnorm(DM22))
endif

equation eq23
eq23.ls d23 c
scalar DM23=@tstats(1)
if DM23<0 then !ProbDM23=2*@cnorm(DM23)
else
!ProbDM23=2*(1-@cnorm(DM23))
endif
```

```
table(7,3) DM
setcolwidth(DM,1,30)
```

```
DM(1,1)="Test Statistics"
DM(2,1)="DM11"
DM(3,1)="DM12"
DM(4,1)="DM13"
DM(5,1)="DM21"
DM(6,1)="DM22"
DM(7,1)="DM23"
DM(1,2)="Test Value"
DM(2,2)=DM11
DM(3,2)=DM12
DM(4,2)=DM13
DM(5,2)=DM21
DM(6,2)=DM22
DM(7,2)=DM23
DM(1,3)="P-value"
DM(2,3)=!ProbDM11
DM(3,3)=!ProbDM12
DM(4,3)=!ProbDM13
DM(5,3)=!ProbDM21
DM(6,3)=!ProbDM22
DM(7,3)=!ProbDM23
```

```
show DM
```

'COMPUTING THE MODIFIED DM (1997)

```

scalar mDM11=((!obs+1-2*1+!obs^(-1)*1*(1-1))/(!obs))^(1/2)*DM11
if mDM11<0 then !Prob_mDM11=2*@cnorm(mDM11)
else
!Prob_mDM11=2*(1-@cnorm(mDM11))
endif

```

```

scalar mDM12=((!obs+1-2*1+!obs^(-1)*2*(1-1))/(!obs))^(1/2)*DM12
if mDM12<0 then !Prob_mDM12=2*@cnorm(mDM12)
else
!Prob_mDM12=2*(1-@cnorm(mDM12))
endif

```

```

scalar mDM13=((!obs+1-2*1+!obs^(-1)*2*(1-1))/(!obs))^(1/2)*DM13
if mDM13<0 then !Prob_mDM13=2*@cnorm(mDM13)
else
!Prob_mDM13=2*(1-@cnorm(mDM13))
endif

```

```

scalar mDM21=((!obs+1-2*1+!obs^(-1)*2*(1-1))/(!obs))^(1/2)*DM21
if mDM21<0 then !Prob_mDM21=2*(1-@cnorm(mDM21))
else
!Prob_mDM21=2*(1-@cnorm(mDM21))
endif

```

```

scalar mDM22=((!obs+1-2*1+!obs^(-1)*2*(1-1))/(!obs))^(1/2)*DM22
if mDM22<0 then !Prob_mDM22=2*@cnorm(mDM22)
else
!Prob_mDM22=2*(1-@cnorm(mDM22))
endif

```

```

scalar mDM23=((!obs+1-2*1+!obs^(-1)*2*(1-1))/(!obs))^(1/2)*DM23
if mDM23<0 then !Prob_mDM23=2*@cnorm(mDM23)
else
!Prob_mDM23=2*(1-@cnorm(mDM23))
endif

```


Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
table(7,3) mDM
setcolwidth(mDM,1,30)

mDM(1,1)="Test Statistics"
mDM(2,1)="mDM11"
mDM(3,1)="mDM12"
mDM(4,1)="mDM13"
mDM(5,1)="mDM21"
mDM(6,1)="mDM22"
mDM(7,1)="mDM23"
mDM(1,2)="Test Value"
mDM(2,2)=mDM11
mDM(3,2)=mDM12
mDM(4,2)=mDM13
mDM(5,2)=mDM21
mDM(6,2)=mDM22
mDM(7,2)=mDM23
mDM(1,3)="P-value"
mDM(2,3)=!Prob_mDM11
mDM(3,3)=!Prob_mDM12
mDM(4,3)=!Prob_mDM13
mDM(5,3)=!Prob_mDM21
mDM(6,3)=!Prob_mDM22
mDM(7,3)=!Prob_mDM23
```

```
show mDM
```

```
'COMPUTING THE MODIFIED MGN (1997)
```

```
series y12=e1t+e2t
series y13=e1t+e3t
series y23=e2t+e3t
series x12=e1t-e2t
series x13=e1t-e3t
series x23=e2t-e3t
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
vector(2000) x_12  
stomna(x12, x_12)  
vector(2000) x_13  
stomna(x13, x_13)  
vector(2000) x_23  
stomna(x23, x_23)
```

```
equation mMGN12  
mMGN12.ls y12 x12  
matrix(1,1) beta_mMGN12  
beta_mMGN12=@coefs
```

```
equation mMGN13  
mMGN13.ls y13 x13  
matrix(1,1) beta_mMGN13  
beta_mMGN13=@coefs
```

```
equation mMGN23  
mMGN23.ls y23 x23  
matrix(1,1) beta_mMGN23  
beta_mMGN23=@coefs
```

```
mMGN12.makesresids residmMGN12  
mMGN13.makesresids residmMGN13  
mMGN23.makesresids residmMGN23
```

```
vector(2000) resid_mMGN12  
vector(2000) resid_mMGN13  
vector(2000) resid_mMGN23  
stomna(residmMGN12, resid_mMGN12)  
stomna(residmMGN13, resid_mMGN13)  
stomna(residmMGN23, resid_mMGN23)
```

```
for !t=1 to !obs
```

```
matrix(1,1) S2_mMGN12
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
S2_mMGN12=(x_12(!t)^2*resid_mMGN12(!t)^2)/((x_12(!t)^2)^2)^(-1/2)*beta_mMGN12
scalar S2_mMGN12_testvalue=S2_mMGN12(1,1)
if S2_mMGN12_testvalue<0 then !Prob_mMGN12=2*@ctdist(S2_mMGN12_testvalue,!obs-1)
else
!Prob_mMGN12=2*(1-@ctdist(S2_mMGN12_testvalue,!obs-1))
endif

matrix(1,1) S2_mMGN13
S2_mMGN13=(x_13(!t)^2*resid_mMGN13(!t)^2)/((x_13(!t)^2)^2)^(-1/2)*beta_mMGN13
scalar S2_mMGN13_testvalue=S2_mMGN13(1,1)
if S2_mMGN13_testvalue<0 then !Prob_mMGN13=2*@ctdist(S2_mMGN13_testvalue,!obs-1)
else
!Prob_mMGN13=2*(1-@ctdist(S2_mMGN13_testvalue,!obs-1))
endif

matrix (1,1) S2_mMGN23
S2_mMGN23=(x_23(!t)^2*resid_mMGN23(!t)^2)/((x_23(!t)^2)^2)^(-1/2)*beta_mMGN23
scalar S2_mMGN23_testvalue=S2_mMGN23(1,1)
if S2_mMGN23_testvalue<0 then !Prob_mMGN23=2*@ctdist(S2_mMGN23_testvalue,!obs-1)
else
!Prob_mMGN23=2*(1-@ctdist(S2_mMGN23_testvalue,!obs-1))
endif

next

table (4,3) mMGN
setcolwidth(mMGN,1,30)

mMGN(1,1)="Test Statistics"
mMGN(1,2)="Test Value"
mMGN(1,3)="P-value"
mMGN(2,1)="mMGN12"
mMGN(3,1)="mMGN13"
mMGN(4,1)="mMGN23"
mMGN(2,2)=S2_mMGN12_testvalue
mMGN(3,2)=S2_mMGN13_testvalue
mMGN(4,2)=S2_mMGN23_testvalue
mMGN(2,3)=!Prob_mMGN12
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
mMGN(3,3)=!Prob_mMGN13
```

```
mMGN(4,3)=!Prob_mMGN23
```

```
show mMGN
```

```
'COMPUTING THE HARVEY-LEYBOURNE-NEWBOLD (1998)
```

```
series x21=e2t-e1t
```

```
series x31=e3t-e1t
```

```
series x32=e3t-e2t
```

```
vector(2000) x_21
```

```
stomna(x21, x_21)
```

```
vector(2000) x_31
```

```
stomna(x31, x_31)
```

```
vector(2000) x_32
```

```
stomna(x32, x_32)
```

```
equation HLN12
```

```
HLN12.ls e1t x12
```

```
matrix(1,1) lambda_HLN12
```

```
lambda_HLN12=@coefs
```

```
equation HLN13
```

```
HLN13.ls e1t x13
```

```
matrix(1,1) lambda_HLN13
```

```
lambda_HLN13=@coefs
```

```
equation HLN23
```

```
HLN23.ls e2t x23
```

```
matrix(1,1) lambda_HLN23
```

```
lambda_HLN23=@coefs
```

```
equation HLN21
```

```
HLN21.ls e2t x21
```

```
matrix(1,1) lambda_HLN21
```

```
lambda_HLN21=@coefs
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

equation HLN31

HLN31.ls e3t x31

matrix(1,1) lambda_HLN31

lambda_HLN31=@coefs

equation HLN32

HLN32.ls e3t x32

matrix(1,1) lambda_HLN32

lambda_HLN32=@coefs

HLN12.makesresids residHLN12

HLN13.makesresids residHLN13

HLN23.makesresids residHLN23

vector(2000) resid_HLN12

vector(2000) resid_HLN13

vector(2000) resid_HLN23

stomna(residHLN12, resid_HLN12)

stomna(residHLN13, resid_HLN13)

stomna(residHLN23, resid_HLN23)

HLN21.makesresids residHLN21

HLN31.makesresids residHLN31

HLN32.makesresids residHLN32

vector(2000) resid_HLN21

vector(2000) resid_HLN31

vector(2000) resid_HLN32

stomna(residHLN21, resid_HLN21)

stomna(residHLN31, resid_HLN31)

stomna(residHLN32, resid_HLN32)

'ASSUMING H=1

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

for !t=1 to !obs

matrix(2000,2000) HatQ1_HLN12

HatQ1_HLN12=(!obs^(-1))*(x_12(!t)^2)*(resid_HLN12(!t)^2)

matrix(2000,2000) HatQ1_HLN13

HatQ1_HLN13=(!obs^(-1))*(x_13(!t)^2)*(resid_HLN13(!t)^2)

matrix(2000,2000) HatQ1_HLN23

HatQ1_HLN23=(!obs^(-1))*(x_23(!t)^2)*(resid_HLN23(!t)^2)

matrix(1,1) R1_HLN12

R1_HLN12=(!obs)^(-1/2)*HatQ1_HLN12(!t)^(-1/2)*x_12(!t)^2*lambda_HLN12

scalar R1_HLN12_testvalue= R1_HLN12(1,1)

if R1_HLN12_testvalue<0 then !Prob_R1_HLN12=2*@cnorm(R1_HLN12_testvalue)

else

!Prob_R1_HLN12=2*(1-@cnorm(R1_HLN12_testvalue))

endif

matrix(1,1) R1_HLN13

R1_HLN13=(!obs)^(-1/2)*HatQ1_HLN13(!t)^(-1/2)*x_13(!t)^2*lambda_HLN13

scalar R1_HLN13_testvalue= R1_HLN13(1,1)

if R1_HLN13_testvalue<0 then !Prob_R1_HLN13=2*@cnorm(R1_HLN13_testvalue)

else

!Prob_R1_HLN13=2*(1-@cnorm(R1_HLN13_testvalue))

endif

matrix(1,1) R1_HLN23

R1_HLN23=(!obs)^(-1/2)*HatQ1_HLN23(!t)^(-1/2)*x_23(!t)^2*lambda_HLN23

scalar R1_HLN23_testvalue= R1_HLN23(1,1)

if R1_HLN23_testvalue<0 then !Prob_R1_HLN23=2*@cnorm(R1_HLN23_testvalue)

else

!Prob_R1_HLN23=2*(1-@cnorm(R1_HLN23_testvalue))

endif

matrix(2000,2000) HatQ1_HLN21

HatQ1_HLN21=(!obs^(-1))*(x_21(!t)^2)*(resid_HLN21(!t)^2)

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
matrix(2000,2000) HatQ1_HLN31
```

```
HatQ1_HLN31=(!obs^(-1))*(x_31(!t)^2)*(resid_HLN31(!t)^2)
```

```
matrix(2000,2000) HatQ1_HLN32
```

```
HatQ1_HLN32=(!obs^(-1))*(x_32(!t)^2)*(resid_HLN32(!t)^2)
```

```
matrix(1,1) R1_HLN21
```

```
R1_HLN21=(!obs)^(-1/2)*HatQ1_HLN21(!t)^(-1/2)*x_21(!t)^2*lambda_HLN21
```

```
scalar R1_HLN21_testvalue= R1_HLN21(1,1)
```

```
if R1_HLN21_testvalue<0 then !Prob_R1_HLN21=2*@cnorm(R1_HLN21_testvalue)
```

```
else
```

```
!Prob_R1_HLN21=2*(1-@cnorm(R1_HLN21_testvalue))
```

```
endif
```

```
matrix(1,1) R1_HLN31
```

```
R1_HLN31=(!obs)^(-1/2)*HatQ1_HLN31(!t)^(-1/2)*x_31(!t)^2*lambda_HLN31
```

```
scalar R1_HLN31_testvalue= R1_HLN31(1,1)
```

```
if R1_HLN31_testvalue<0 then !Prob_R1_HLN31=2*@cnorm(R1_HLN31_testvalue)
```

```
else
```

```
!Prob_R1_HLN31=2*(1-@cnorm(R1_HLN31_testvalue))
```

```
endif
```

```
matrix(1,1) R1_HLN32
```

```
R1_HLN32=(!obs)^(-1/2)*HatQ1_HLN32(!t)^(-1/2)*x_32(!t)^2*lambda_HLN32
```

```
scalar R1_HLN32_testvalue= R1_HLN32(1,1)
```

```
if R1_HLN32_testvalue<0 then !Prob_R1_HLN32=2*@cnorm(R1_HLN32_testvalue)
```

```
else
```

```
!Prob_R1_HLN32=2*(1-@cnorm(R1_HLN32_testvalue))
```

```
endif
```

```
series dtHLN12=x12*e1t
```

```
series dtHLN13=x13*e1t
```

```
series dtHLN23=x23*e2t
```

```
series dtHLN21=x21*e2t
```

```
series dtHLN31=x31*e3t
```

```
series dtHLN32=x32*e3t
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
vector(2000) dt_HLN12  
stomna(dtHLN12, dt_HLN12)
```

```
vector(2000) dt_HLN13  
stomna(dtHLN13, dt_HLN13)
```

```
vector(2000) dt_HLN23  
stomna(dtHLN23, dt_HLN23)
```

```
matrix(2000,2000) HatQ2_HLN12  
HatQ2_HLN12=(!obs)^(-1)*dt_HLN12(!t)^2
```

```
matrix(2000,2000) HatQ2_HLN13  
HatQ2_HLN13=(!obs)^(-1)*dt_HLN13(!t)^2
```

```
matrix(2000,2000) HatQ2_HLN23  
HatQ2_HLN23=(!obs)^(-1)*dt_HLN23(!t)^2
```

```
vector(2000) dt_HLN21  
stomna(dtHLN21, dt_HLN21)
```

```
vector(2000) dt_HLN31  
stomna(dtHLN31, dt_HLN31)
```

```
vector(2000) dt_HLN32  
stomna(dtHLN32, dt_HLN32)
```

```
matrix(2000,2000) HatQ2_HLN21  
HatQ2_HLN21=(!obs)^(-1)*dt_HLN21(!t)^2
```

```
matrix(2000,2000) HatQ2_HLN31  
HatQ2_HLN31=(!obs)^(-1)*dt_HLN31(!t)^2
```

```
matrix(2000,2000) HatQ2_HLN32  
HatQ2_HLN32=(!obs)^(-1)*dt_HLN32(!t)^2
```


Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
matrix(1,1) R2_HLN12
```

```
R2_HLN12=(!obs)^(-1/2)*HatQ2_HLN12(!t)^(-1/2)*x_12(!t)^(2)*lambda_HLN12
```

```
scalar R2_HLN12_testvalue= R2_HLN12(1,1)
```

```
if R2_HLN12_testvalue<0 then !Prob_R2_HLN12=2*@cnorm(R2_HLN12_testvalue)
```

```
else
```

```
!Prob_R2_HLN12=2*(1-@cnorm(R2_HLN12_testvalue))
```

```
endif
```

```
matrix(1,1) R2_HLN13
```

```
R2_HLN13=(!obs)^(-1/2)*HatQ2_HLN13(!t)^(-1/2)*x_13(!t)^(2)*lambda_HLN13
```

```
scalar R2_HLN13_testvalue= R2_HLN13(1,1)
```

```
if R2_HLN13_testvalue<0 then !Prob_R2_HLN13=2*@cnorm(R2_HLN13_testvalue)
```

```
else
```

```
!Prob_R2_HLN13=2*(1-@cnorm(R2_HLN13_testvalue))
```

```
endif
```

```
matrix(1,1) R2_HLN23
```

```
R2_HLN23=(!obs)^(-1/2)*HatQ2_HLN23(!t)^(-1/2)*x_23(!t)^(2)*lambda_HLN23
```

```
scalar R2_HLN23_testvalue= R2_HLN23(1,1)
```

```
if R2_HLN23_testvalue<0 then !Prob_R2_HLN23=2*@cnorm(R2_HLN23_testvalue)
```

```
else
```

```
!Prob_R2_HLN23=2*(1-@cnorm(R2_HLN23_testvalue))
```

```
endif
```

```
matrix(1,1) R2_HLN21
```

```
R2_HLN21=(!obs)^(-1/2)*HatQ2_HLN21(!t)^(-1/2)*x_21(!t)^(2)*lambda_HLN21
```

```
scalar R2_HLN21_testvalue= R2_HLN21(1,1)
```

```
if R2_HLN21_testvalue<0 then !Prob_R2_HLN21=2*@cnorm(R2_HLN21_testvalue)
```

```
else
```

```
!Prob_R2_HLN21=2*(1-@cnorm(R2_HLN21_testvalue))
```

```
endif
```

```
matrix(1,1) R2_HLN31
```

```
R2_HLN31=(!obs)^(-1/2)*HatQ2_HLN31(!t)^(-1/2)*x_31(!t)^(2)*lambda_HLN31
```

```
scalar R2_HLN31_testvalue= R2_HLN31(1,1)
```

```
if R2_HLN31_testvalue<0 then !Prob_R2_HLN31=2*@cnorm(R2_HLN31_testvalue)
```

```
else
```

```
!Prob_R2_HLN31=2*(1-@cnorm(R2_HLN31_testvalue))
```

```
endif
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
matrix(1,1) R2_HLN32
R2_HLN32=(lobs)^(-1/2)*HatQ2_HLN32(!t)^(-1/2)*x_32(!t)^(2)*lambda_HLN32
scalar R2_HLN32_testvalue= R2_HLN32(1,1)
if R2_HLN32_testvalue<0 then !Prob_R2_HLN32=2*@cnorm(R2_HLN32_testvalue)
else
!Prob_R2_HLN32=2*(1-@cnorm(R2_HLN32_testvalue))
endif
```

```
next
```

```
table(10,3) HLN
setcolwidth(HLN,1,30)
```

```
HLN(1,1)="Test Statistics"
HLN(2,1)="R1_HLN12"
HLN(3,1)="R1_HLN13"
HLN(4,1)="R1_HLN23"
HLN(5,1)="R2_HLN12"
HLN(6,1)="R2_HLN13"
HLN(7,1)="R2_HLN23"
HLN(8,1)="R1_HLN21"
HLN(9,1)="R1_HLN31"
HLN(10,1)="R1_HLN32"
HLN(11,1)="R2_HLN21"
HLN(12,1)="R2_HLN31"
HLN(13,1)="R2_HLN32"
HLN(1,2)="Test Value"
HLN(2,2)=R1_HLN12_testvalue
HLN(3,2)=R1_HLN13_testvalue
HLN(4,2)=R1_HLN23_testvalue
HLN(5,2)=R2_HLN12_testvalue
HLN(6,2)=R2_HLN13_testvalue
HLN(7,2)=R2_HLN23_testvalue
HLN(8,2)=R1_HLN21_testvalue
HLN(9,2)=R1_HLN31_testvalue
HLN(10,2)=R1_HLN32_testvalue
HLN(11,2)=R2_HLN21_testvalue
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
HLN(12,2)=R2_HLN31_testvalue
HLN(13,2)=R2_HLN32_testvalue
HLN(1,3)="P-value"
HLN(2,3)=!Prob_R1_HLN12
HLN(3,3)=!Prob_R1_HLN13
HLN(4,3)=!Prob_R1_HLN23
HLN(5,3)=!Prob_R2_HLN12
HLN(6,3)=!Prob_R2_HLN13
HLN(7,3)=!Prob_R2_HLN23
HLN(8,3)=!Prob_R1_HLN21
HLN(9,3)=!Prob_R1_HLN31
HLN(10,3)=!Prob_R1_HLN32
HLN(11,3)=!Prob_R2_HLN21
HLN(12,3)=!Prob_R2_HLN31
HLN(13,3)=!Prob_R2_HLN32
```

```
show HLN
```

```
'COMPUTING THE PETER HANSEN TEST (2005)
```

```
scalar d_bar12_hansen
d_bar12_hansen=!obs^(-1)*@sum(x_12)
scalar var_x12=@var(x_12)

scalar Hat_omega12_hansen
Hat_omega12_hansen=@sqrt(var_x12)

scalar spa_hansen12
spa_hansen12=(!obs^(1/2)*d_bar12_hansen)/Hat_omega12_hansen

scalar spa_hansen12_testvalue
if spa_hansen12>0 then spa_hansen12_testvalue=spa_hansen12
else spa_hansen12_testvalue=0
endif

scalar sqrt_hansen=-@sqrt(@log(@log(!obs)))
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
scalar Hat_miu12_hansen
if spa_hansen12<sqrt_hansen then Hat_miu12_hansen=d_bar12_hansen
else
Hat_miu12_hansen=0
endif

!Prob_spa_hansen12=2*(1-@cnorm((spa_hansen12_testvalue-
Hat_miu12_hansen)/Hat_omega12_hansen))
```

```
scalar d_bar13_hansen
d_bar13_hansen=!obs^(-1)*@sum(x_13)
scalar var_x13=@var(x_13)
```

```
scalar Hat_omega13_hansen
Hat_omega13_hansen=@sqrt(var_x13)
```

```
scalar spa_hansen13
spa_hansen13=(!obs^(1/2)*d_bar13_hansen)/Hat_omega13_hansen
```

```
scalar spa_hansen13_testvalue
if spa_hansen13>0 then spa_hansen13_testvalue=spa_hansen13
else spa_hansen13_testvalue=0
endif
```

```
scalar sqrt_hansen=-@sqrt(@log(@log(!obs)))
```

```
scalar Hat_miu13_hansen
if spa_hansen13<sqrt_hansen then Hat_miu13_hansen=d_bar13_hansen
else
Hat_miu13_hansen=0
endif
```

```
!Prob_spa_hansen13=2*(1-@cnorm((spa_hansen13_testvalue-
Hat_miu13_hansen)/Hat_omega13_hansen))
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

```
scalar d_bar23_hansen
d_bar23_hansen=!obs^(-1)*@sum(x_23)
scalar var_x23=@var(x_23)

scalar Hat_omega23_hansen
Hat_omega23_hansen=@sqrt(var_x23)

scalar spa_hansen23
spa_hansen23=(!obs^(1/2)*d_bar23_hansen)/Hat_omega23_hansen

scalar spa_hansen23_testvalue
if spa_hansen23>0 then spa_hansen23_testvalue=spa_hansen23
else spa_hansen23_testvalue=0
endif

scalar sqrt_hansen=-@sqrt(@log(@log(!obs)))

scalar Hat_miu23_hansen
if spa_hansen23<sqrt_hansen then Hat_miu23_hansen=d_bar23_hansen
else
Hat_miu23_hansen=0
endif

!Prob_spa_hansen23=2*(1-@cnorm((spa_hansen23_testvalue-
Hat_miu23_hansen)/Hat_omega23_hansen))

table (4,3) Hansen
setcolwidth(Hansen,1,30)

Hansen(1,1)="Test Statistics"
Hansen(1,2)="Test Value"
Hansen(1,3)="P-value"
Hansen(2,1)="Hansen12"
Hansen(3,1)="Hansen13"
Hansen(4,1)="Hansen23"
Hansen(2,2)=spa_hansen12_testvalue
Hansen(3,2)=spa_hansen13_testvalue
Hansen(4,2)=spa_hansen23_testvalue
Hansen(2,3)=!Prob_spa_hansen12
```

Predictive Accuracy of Alternative Autoregressive Conditional Heteroskedasticity Models

Hansen(3,3)=!Prob_spa_hansen13

Hansen(4,3)=!Prob_spa_hansen23

show Hansen