TESTING FORPREDICTIVE ACCURACYUNDERDIFFERENTERROR STRUCTURES

María Planas Lasa

Trabajo de investigación 012/008

Master en Banca y Finanzas Cuantitativas

Directoras :Eva Ferreira y Susan Orbe.

Universidad del País Vasco

Universidad de Castilla-La Mancha

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia

www.finanzascuantitativas.es

Abstract

We propose and discuss tests of null hypothesis of no difference in the accuracy of two competing forecasts. A wide variety of accuracy measures can be used, even if the loss function is not quadratic and even not symmetric. We review Diebold and Mariano (1995) and so we let forecast errors to be non-Gaussian, serially correlated and contemporaneously correlated.

Secondly, we introduce in this discussion regressions of the competing models. All tests are repeated in this new context to see how the results are despite the errors due to the regressions introduced. In this part, we are less restrictive so we permit the errors to be serially correlated and contemporaneously correlated. The serial correlation is due to processes with short-memory and long-memory. Also errors can be in a context of heteroscedasticity.

Finally all tests are evaluated in the context of an empirical example, with real data.

Introduction

In all sciences, prediction is of fundamental importance in order to guide decisions. In economics, comparing the forecast accuracy among competing models is also of a great importance because predictive performance and model adequacy are inextricably linked.

The literature contains thousands of forecasts comparisons. In almost without exception, the forecast accuracy is evaluated with no regressions of the models that are competing and sometimes with no attempt to assess their sampling uncertainty.

Correlation of forecast errors across space and time, as well as other additional complications, make formal comparisons of forecast accuracy difficult. Uncertainty due to the estimation also hinders formal comparisons.

There are some publications offering pessimistic assessments of the possibilities for formal testing. Diebold and Mariano (1995) discussed some formal tests in a context of serial correlation and even with contemporaneous correlation but always in a without any kind of model regressions.

We proceed by detailing our tests procedures in section 1, where we detail the tests as appear in Diebold and Mariano (1995).

In section 2 we evaluate formal tests in a new context: introducing regressions of competing models and even giving the innovation part of the models different structures of serial correlation and contemporaneous correlation. The matching results are detailed in section 3.

Finally, in section 4 we evaluated formal tests with an empirical example based on Audrino and Medeiros (2011).

Section 1

Suppose that Y is an unknown variable we are interested in. Two variables X_1 and X_2 , which can have any dimension, are available to predict the value of Y. The task is to compare the forecast accuracy of these two variables, which means that it is required to study which of these two models is more accurate:

(Model 1) $Y = \alpha_1 + m_1(X_1) + u_1$

(Model 2) $Y = \alpha_2 + m_2(X_2) + u_2$

In particular, we compare these next two linear and individual models:

(Model 1) $Y_t = \alpha_1 + \beta_1 X_{1t} + u_{1t}$

(Model 2) $Y_t = \alpha_2 + \beta_2 X_{2t} + u_{2t}$

There are some publications about this subject. In particular, Diebold and Mariano (1995) study the accuracy of two variables' prediction maintaining always a context in which there is never any kind of estimation. However, they introduce structures to errors permitting them to be serially correlated and even contemporaneously correlated.

In this document, Diebold and Mariano's work will be revised and it is going to be introduced a new context containing estimation in order to see how parametric uncertainty affect the results.

It is wanted the "loss" associated with the individual forecast of variables X_1 and X_2 to be measured. In order to measure it, it is considered the loss function, which is a function depending on the forecast error.

Let's consider two forecasts $\{\hat{Y}_{1t}\}_{t=1}^{T}$ and $\{\hat{Y}_{2t}\}_{t=1}^{T}$ of the time series $\{Y_t\}_{t=1}^{T}$. Let the associated forecast errors be denoted by $\{e_{1t}\}_{t=1}^{T}$ and $\{e_{2t}\}_{t=1}^{T}$, so the loss function is $g(Y_t, \hat{Y}_{it}) = g(e_{it})$ for $i = 1,2$. This means that the loss function will be able to be written depending only on the forecast error, which is supposed to be zero-mean.

We choose quadratic loss function to measure the quality of each fit:

$g(e_{it}) = e_{it}^2 \, i = 1,2.$

The null hypothesis of equal accuracy for two forecasts is

$$H_0: E[g(e_{1t})] = E[g(e_{2t})]$$

while the alternative is

$$H_a: E[g(e_{1t})] \neq E[g(e_{2t})]$$

If we introduce the notation of loss differential, this is $d_t = g(e_{1t}) - g(e_{2t})$, the corresponding null hypothesis is:

$H_0 = E[d] = 0$ (*)

We test the accuracy of the individual prediction of variables X_1 and X_2 by using different tests and statistics. All of them were proposed and/or evaluated by Diebold and Mariano(1995).

They are explained with details in sections 1.1, 1.2 and 1.3.

1.1. Diebold and Mariano test

Let's consider a sample path of loss differential series, $\{d_t\}_{t=1}^{T}$. We denote the sample mean loss differential by $\bar{d}$.

We make the assumption, as the authors did in their publication, that the loss differential series is covariance stationary and short memory, for example, a moving average process.

In large samples, $\bar{d}$is approximately normally distributed with mean μ and variance $\frac{2\pi \hat{f}_d(0)}{T}$ where $\hat{f}_d(0)$ is the spectral density of the loss differential at frequency zero. Thespectral density estimate has relation with autocovariance terms and the serial correlation of the terms of d. Its formula is:

$$\hat{f}_d(0) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma_d(t)$$

where

$$\gamma_d(\tau) = E[(d_t - \mu)(d_{t-\tau} - \mu)]$$

To get a consistent estimate of the spectral density it is necessary to set a truncation lag ,$S(T)$, so a good estimate to $2\pi \hat{f}_d(0)$ can be a weighted sum of sample covariances which weight depends on the truncation lag. The value of the truncation lag keeps relation with the degree of dependence between the errors $\{e_{1t}\}_{t=1}^{T}$and $\{e_{2t}\}_{t=1}^{T}$. In this case, the truncation lag,$S(T)$, will be set to one, as in Diebold and Mariano's publication. The authors did the next approximation:

$$2\pi \hat{f}_d(0) = \sum_{t=-(T-1)}^{(T-1)} l\left(\frac{\tau}{S(T)}\right) \gamma_d(\tau)$$

$$\gamma_d(\tau) = \frac{1}{T} \sum_{t=|\tau|+1}^{T} (d_t - \bar{d})(d_{t-|\tau|} - \bar{d})$$

$$l\left(\frac{\tau}{S(T)}\right) = \begin{cases} 1 & for \left|\frac{\tau}{S(T)}\right| \leq 1 \\ 0 & for \left|\frac{\tau}{S(T)}\right| > 1 \end{cases}$$

Of course, the lag window$l\left(\frac{\tau}{S(T)}\right)$ can be defined by another function. We are maintaining the same function to define the truncation lag as Diebold and Mariano did.

Then, first statistic we are going to take into account is Mariano and Diebold's statistic S_1, which is defined by

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

It´s clear that S_1 is asymptotically distributed by a standard normal distribution. This means that the mean of $\bar{d}$ is supposed to be μ=0.

$S_1$is not thought to be useful in finite small samples. It remains helpful in large samples because of its asymptotic nature.

1.2. Exact finite simple test(The sign test)

Another test we are going to consider and get into practice is the Sign Test. The null hypothesis is a zero median loss differential and this test is thought to be helpful in finite samples.

Assuming that each loss differential seriess are i.i.d we can work the corresponding statistic out, which is (for small samples):

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

where$I_+(d_t)$ takes the value one if $d_t > 0$ and the value zero in the other case. This statistic is distributed by the binomial distribution (under the null hypothesis) with parameter T and ½.This statistic is thought to be helpful with finite samples.

For large samples we work another statistic out which is a light change of S_2. It is defined by

$$S_2^* = \frac{S_2 - 0.5\,T}{\sqrt{0.25T}}$$

which is distributed, under the null hypothesis by a standard normal distribution.

Furthermore, if loss differences are also symmetrically distributed, the median and mean are equal and then, the null hypothesis of Sign test is the same that (*).

Advantages of these two tests

One virtue of both tests explained previously is that loss function does not need to be quadratic and do not need even be symmetric or continuous.

On the one hand, forecast errors do not need to be non-zero mean but they will always be supposed zero-mean. On the other hand forecast errors are allowed to be non-Gaussian and even contemporaneously correlated, as has been assumedinS_2^* test.

Allowance for contemporaneous correlation is important because the forecasts being compared are forecasts of the same endogenous variables. If the endogenous variables is an economic time series, information of forecasters are largely overlapping, so forecasts errors tend to be strongly correlated.

Furthermore, S_1 also allows serial correlation in each one of the error seriess. This is also very important due to forecast errors are serially correlated in general.

1.3. Other tests.

We introduce another three tests which belong to another authors and were also used in Diebold and Mariano (1995).

Previous statistics explained don't impose any kind of loss function g. Now we are going to introduce these three tests which can be used by assuming that $g(.)$ is a quadratic function.

Some other assumptions for each test will be made but it's important to point out that we will do some violations of these assumptions later , in the empirical part of this paper, in order to see how these tests work despite of these violations. We will never violate the assumption corresponding with the zero mean of the errors.

The first one is the Simple F test.

We assume that errors are zero mean, Gaussian, serially uncorrelated and contemporaneously uncorrelated. The test statistic is:

$$F = \frac{e_1{}' e_1}{e_2{}' e_2} \sim F(T, T)$$

where$\{e_{1t}\}_{t=1}^{T}$ and $\{e_{2t}\}_{t=1}^{T}$ are the forecast errors associated to each variable.

It is distributed as $F(T, T)$(under null hypothesis) if the length of both error seriessis T.

As we saidbefore we are going to violate some of the assumptions of this test and it is important to highlight that the presence of contemporaneous correlation, which is one of these violations, has bad effects to the results obtained because numerator and denominator of F are correlated and then, F does not have the F distribution.

The second one is the Morgan-Granger-Newbold test.

We assume that errors are zero mean, Gaussian and serially uncorrelated. With this test we allow the existence of contemporaneous correlation .

The test statistic is:

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

where $\hat{\rho}_{xz} = \frac{x'z}{\sqrt{(x'x)(z'z)}}$ and $x = (e_i + e_j), z = (e_i - e_j)$,

The null hypothesis is equal to zero correlation between $x$and z.

MGN is distributed as Student's t with T-1 degrees of freedom if T is the corresponding length of both forecast errors.

The last one is the Meese-Rogoff test.

We assume that errors are zero mean and Gaussian. The corresponding statistic is:

$$MR = \frac{\hat{\gamma}_{xz}}{\sqrt{\frac{\hat{\Sigma}}{T}}} \sim N(0,1)$$

where $x = \left(e_i + e_j\right), z = \left(e_i - e_j\right)$,$\hat{\gamma}_{xz} = \frac{x'z}{T}$ and $\hat{\Sigma}$is a consistent estimation of the variance of $\hat{\gamma}_{xz}\sqrt{T}$.

In our case, we choose $\hat{\Sigma}$following Diebold and Rudebusch (1991):

$$\hat{\Sigma} = \sum_{\tau=-S(T)}^{S(T)} \hat{\gamma}_{XX}(\tau)\hat{\gamma}_{ZZ}(\tau) + \hat{\gamma}_{XZ}(\tau)\hat{\gamma}_{ZX}(\tau)$$

where:

$S(T) = 1$ (truncation lag)

$$\hat{\gamma}_{XZ}(\tau) = \begin{cases} \frac{1}{T}\sum_{t=\tau+1}^{T} x_t z_{t-\tau}, & \tau \geq 0 \\ \hat{\gamma}_{ZX}(-\tau) & otherwise \end{cases}$$

$$\hat{\gamma}_{ZX}(\tau) = \begin{cases} \frac{1}{T}\sum_{t=\tau+1}^{T} z_t x_{t-\tau}, & \tau \geq 0 \\ \hat{\gamma}_{XZ}(-\tau) & otherwise \end{cases}$$

$$\hat{\gamma}_{XX}(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T} x_t x_{t-\tau},$$

$$\hat{\gamma}_{ZZ}(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T} z_t z_{t-\tau},$$

Of course, we can let some assumptions to be relaxed in some other combinations, but in these cases the distributions become tedious. For more information and details about these, see Diebold and Mariano (1995).

Section 2

In this section, we are going one step further, as we are going to evaluate all tests explained above but we are going to work with validation residuals in spite of simulated errors .This means that we are introducing some regressions.

Our goal is to study two subjects: first one is how the parametric uncertainty due to the regressions affect the tests explained previously and second one is how the structure of the innovation terms affect to the results. Diebold and Mariano (1995) give the innovation term structure with short memory processes. Other structures for the innovation terms are going to be established.

Ferreira and Stute (2009) study this new context with regressions and evaluate how Diebold and Mariano test works in a scenario ofi.i.d innovations. Now this scenario is going to be more complex.

Both variables X_1 and X_2 explain the endogenous variable Y. We want to choose which one has better predictive accuracy.

Our model is:

$$Y_t = a + bX_{1t} + cX_{2t} + u_t \quad (1)$$

The coefficients' values a, b and c are fixed and chosen in order to the systematic part has more information than the innovation part. We have established their values as follows:

$a = 1.8$, $b = 3$ and $c = 2.5$

It is known that each one of these variables X_1 and X_2 are helpful to predict Y but now, in a regression context, we decompose Y into terms depending only on X_1or X_2.

We have to simulate variables X_1 and X_2 and so the variable Y. The number of data simulated is $T + n$.

We will regress next models:

$$Y_t = \alpha_1 + \beta_1 X_{1t} + \varepsilon_{1t} \quad (2)$$

$$Y_t = \alpha_2 + \beta_2 X_{2t} + \varepsilon_{2t} \quad (3)$$

The regressions are done with the first T values of the variables and then, with the validation part of the sample (from T+1 to $T + n$observations) with the corresponding coefficients estimated we calculate the residuals, this means that residuals have length n.

$$\hat{\varepsilon}_{it} = Y_t - \hat{\alpha}_1 - \hat{\beta}_1 X_{it} \qquad t = T+1, \dots, T+n \qquad i = 1,2.$$

Once the residuals have been calculated, all tests can be evaluated changing errors to residuals. Now, we have to take into account that residuals haven't length $T$and some parameter of statistics' distribution under de null hypothesis have changed.

In expression (2), termsε_{1t} correspond with $bX_{2t} + u_t$ (see (1)) and in expression (3) terms ε_{2t} correspond with$aX_{1t} + u_t$. This lead us to some conclusions. First one is that innovations of models (2) and (3) are contemporaneously correlated because of the common random term u_t. They are also serially correlated if terms $u_t$have structure. The second conclusion is that the variances of the exogenous variables will be important in order to choose between models (2) and (3) depending on what we want to measure (the size or the power of these tests).

In our study, our task is not looking for efficient estimations so, for this regressions, Ordinary Least Squares is a good method to apply despite of the possible autocorrelation in the innovation part. Twill take the same values as in Diebold and Mariano's paper and the length of the validation part, n, will take the values $T/2$,T and $2\,T$.

First of all we discuss how these tests work when the innovations $\{u_t\}$ have nostructure , that is, they are with zero-mean, homoscedastic and with no serial correlation. In this case, when we test the statisticS_1, we can work with the version of Ferreira and Stute(1995).This is a good advantage because of the great operational cost of S_1's algorithm.Their version is a little more simple and the corresponding algorithm takes not as longer as does Diebold and Mariano's version.

Secondly, timeseries$\{u_t\}$ will be given the same MA(1) structure as Diebold and Mariano do in their publication, so the parameter of the MA(1) process will take values 0,0.5,0.9.

To build innovations with this structure it is necessary to simulate ani.i.dseries from a standard normal distribution, $\{\varepsilon_t\}_{t=1}^{T+n}$. To get a MA(1) it is only required to do the next transformation, which lead us to errors with variance one:

$$u_t = \frac{1}{\sqrt{1+\theta^2}}(\varepsilon_t + \theta\varepsilon_{t-1})$$

Thirdly, this procedure will be repeated giving$\{u_t\}$ structure following a long memory and covariance stationary process. This model is an AR(1) model without constant and with parameter's value equal to 0.6 and 0.9. The second value of parameter is included to study if the persistence affect to the test. In this case, to get a AR(1) it is only required to do the next transformation:

$$u_t = \phi u_{t-1} + \varepsilon_t \quad , \varepsilon_t \sim i.i.d.N(0,1)$$

Lately, it will be introduced heteroscedasticity of errors but it will be explained with details lately because the method applied is a little different (section 2.1).

We maintain the same significance level ($\alpha = 10\%$), the same number of iterations (10000 in some cases and 5000 in others)as Diebold and Mariano (1995) .

The exogenous variables are both normally distributed, independent with mean zero. The variance of variable $X_1$is always fixed in one and variance of X_2 is 1.5 or 2 if we measure the power of the test. If we want to measure the size of the test, X_2's variance is set to 1.2. This value is calculated imposing ratio of the errors' variance of models to be one. We denote by $\sigma^2$the variance of X_2so we have:

$$Y_{1t} = a + bX_{1t} + u_{1t} \rightarrow u_{1t} = cX_{2t} + u_t \rightarrow Var(u_{1t}) = c^2\,\sigma^2 + Var(u_t)$$

$$Y_{2t} = a + cX_{2t} + u_{2t} \rightarrow u_{2t} = bX_{1t} + u_t \rightarrow Var(u_{2t}) = b^2 + Var(u_t)$$

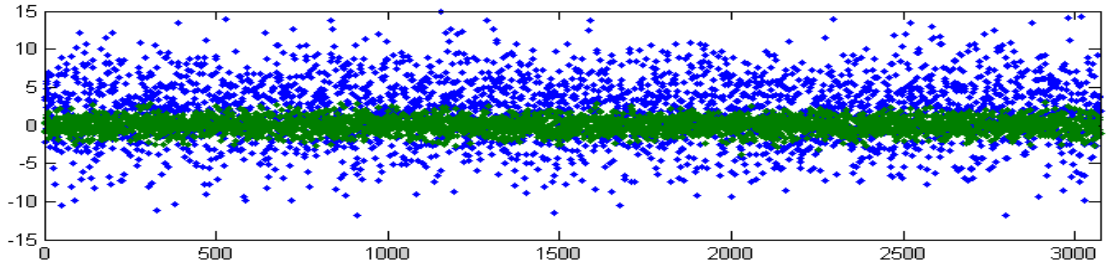
So, to test the size, it has to be$c^2\,\sigma^2 = b^2$, independently of the structure of the innovations $\{u_t\}$. This occurs when $\sigma = 1.2$. When$\sigma$is set to 1.5 or 2, it can be ckecked the corresponding values of ratio$\frac{Var(u_{1t})}{Var(u_{2t})}$ are:

structure	$\sigma = 1.5$	$\sigma = 2$
$\{u_t\}$i.i.d	1.5	2.6
$\{u_t\}$ MA(1) $\forall\theta$	1.5	2.6
$\{u_t\}$ AR(1) $\phi = 0.6$	1.5	2.51
$\{u_t\}$ AR(1) $\phi = 0.9$	1.35	2.12

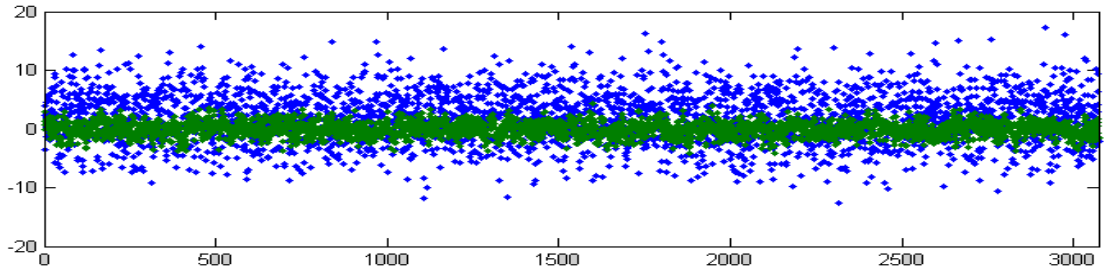
In order to see that the values of coefficients a, b, c are correctly fixed, some figures are represented below. They correspond with the systematic part, $a + bX_{1t} + cX_{2t}$, against the innovation part, u_t. All pictures represented correspond with the case of size of the estimation part T=1024 and size of the validation part n=2048. They also correspond with the case of variance of X_2 equal to 1.2, in which case the size is tested.

Obviously they are only calculated in one iteration, in this case, the last one. There are as many pictures as structures of u_t have been set.

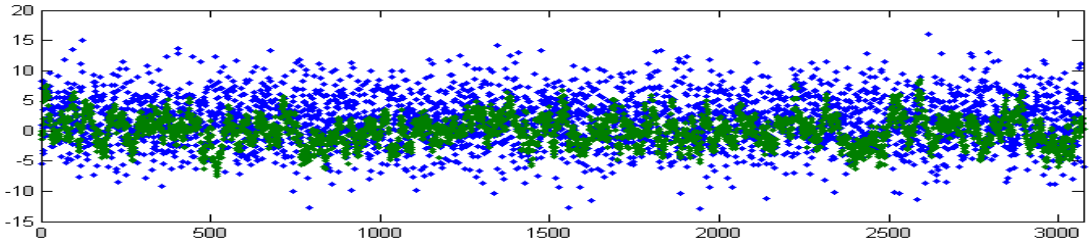
* $u_t$innovations with no structure:



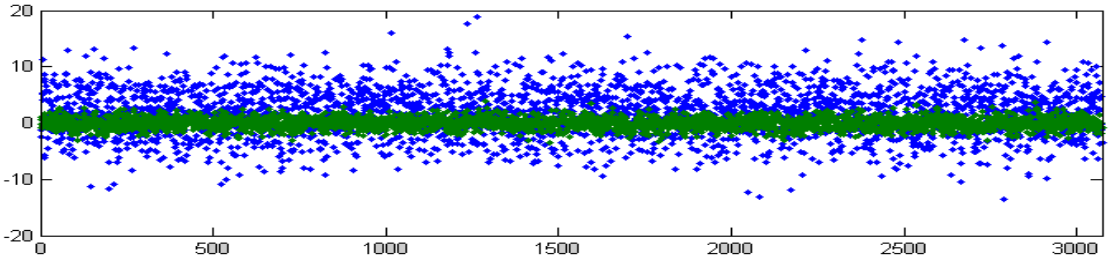
* u_tAR(1) with parameter's value equal to 0.6:



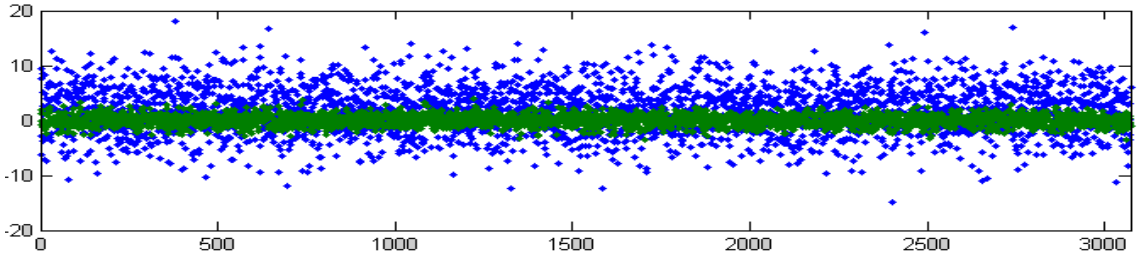
* u_tAR(1) with parameter's value equal to 0.9:



* u_tMA(1) with parameter's value equal to 0.5:



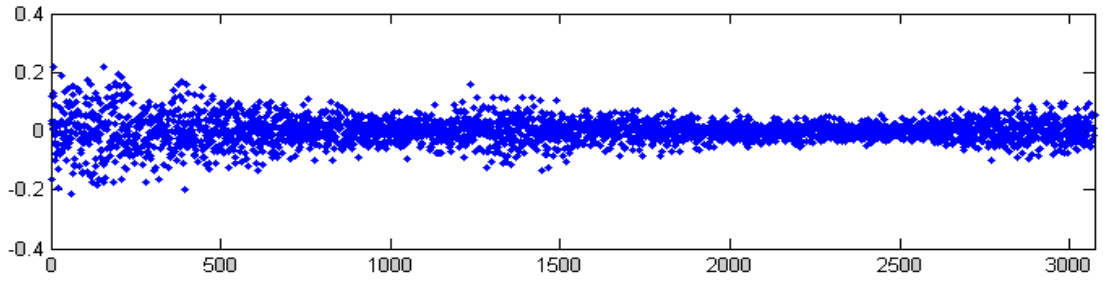
* u_tMA(1) with parameter's value equal to 0.9:



2.1. **Heteroscedasticity case**

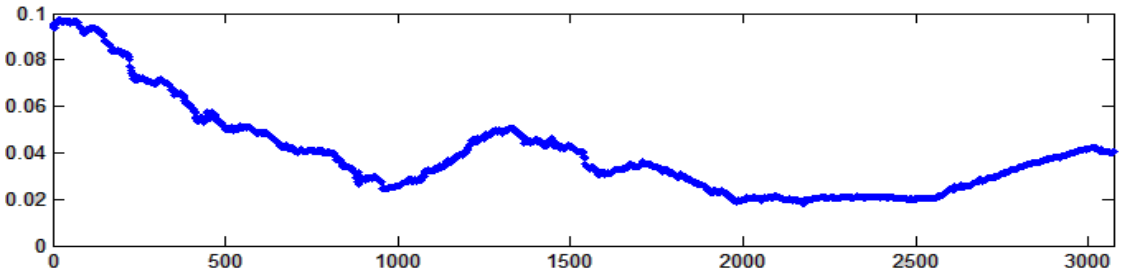
The case of heteroscedasticity is a little more complex. In this case, the innovations have the structure $u_t = w_t Z_t$ where w_t are independent standard normal observations and Z_t is a variable chosen. In this paper, $Z_t$is the time series corresponding to Spanish interest rate with maturity equal to 6 months. They cover to 3/05/1995 from 17/10/2007 (3075 observations). They have been collected from Bloomberg website.

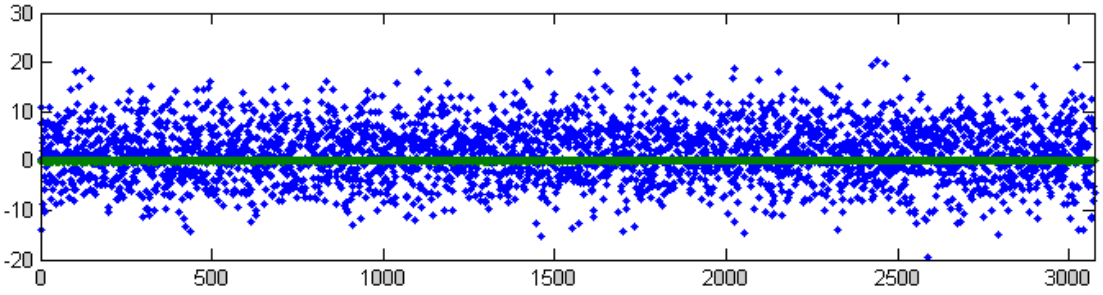
The next figure represents $u_t = w_t Z_t$:



As we can observe, there are some volatility groupages.

Next figures represent the variable Z and the systematic part against the innovation part. Last one corresponds with case of T=1024 and n=2048:





In this case, the method to apply is not Ordinary Least Squares but Generalized Least Squares .

It is clear that we have to transform our model to another onewith constant variance innovations:

$$Y_t = a + bX_{1t} + cX_{2t} + u_t \quad \rightarrow \frac{Y_t}{Z_t} = \frac{a}{Z_t} + b\frac{X_{1t}}{Z_t} + c\frac{X_{2t}}{Z_t} + w_t$$

Once the model has been transformed, Ordinary Least Squares is a good method to apply.

As it has been done in other cases, it is wanted the size of all tests(under heteroscedasticity) to be measured and for this it is necessary to choose the correct variance of variable X_2. The corresponding individual models now are:

$$\frac{Y_t}{Z_t} = \frac{a}{Z_t} + b\frac{X_{1t}}{Z_t} + v_{1t} \rightarrow v_{1t} = c\frac{X_{2t}}{Z_t} + w_t$$

$$\frac{Y_t}{Z_t} = \frac{a}{Z_t} + c\frac{X_{2t}}{Z_t} + v_{2t} \rightarrow v_{2t} = b\frac{X_{1t}}{Z_t} + w_t$$

It is required to calculate the ratio of variances of v_1 and v_2 and for this it's necessary to calculate the variance of $\frac{X_{it}}{Z_t}$. We can do the next approximation as each variable $X = X_i$ is independent from Z:

$$Var\left(\frac{X}{Z}\right) = E\left(\left(\frac{X}{Z}\right)^2\right) - \left(\left(E\left(\frac{X}{Z}\right)\right)^2 = E(X^2)E\left(\frac{1}{Z^2}\right) - \left(E(X)\right)^2\left(E\left(\frac{1}{Z}\right)\right)^2\right.$$

$$= \left(Var(X) + \left(E(X)\right)^2\right)\left(Var\left(\frac{1}{Z}\right) + \left(E\left(\frac{1}{Z}\right)\right)^2\right) - \left(E(X)\right)^2\left(E\left(\frac{1}{Z}\right)\right)^2$$

$$= Var(X)Var\left(\frac{1}{Z}\right) + \left(E(X)\right)^2 Var\left(\frac{1}{Z}\right) + Var(X)\left(E\left(\frac{1}{Z}\right)\right)^2$$

In this case, $E(X) = E(X_i) = 0\ \forall i = 1{,}2$. So the expression useful for us can be simplified and then we have:

$$Var\left(\frac{X}{Z}\right) = Var(X)Var\left(\frac{1}{Z}\right) + Var(X)\left(E\left(\frac{1}{Z}\right)\right)^2$$

To get a good approximation of the moments of the variable $\frac{1}{Z}$ we make the next approximations(seeFerreira and Tusell (1996)):

$$E\left(\frac{1}{Z}\right) = \frac{1}{E(Z)} + \frac{Var(Z)}{(E(Z))^3}$$

$$Var\left(\frac{1}{Z}\right) = \left(-\frac{1}{\left(E(Z)\right)^2}\right)^2 Var(Z) = \frac{Var(Z)}{(E(Z))^4}$$

Following these approximations, we have:

$$Var\left(\frac{X}{Z}\right) = Var(X)Var\left(\frac{1}{Z}\right) + Var(X)\left(E\left(\frac{1}{Z}\right)\right)^2 = Var(X)\left\{\frac{Var(Z)}{\left(E(Z)\right)^4} + \left[\frac{1}{E(Z)} + \frac{Var(Z)}{\left(E(Z)\right)^3}\right]^2\right\}$$

We approximate $Var(Z)$ and $E(Z)$ by the sample variance and sample mean of Z using all data (3072 observations). These respective values are 3.71 and 3.96 and so the ratio of variances of errors of the transformed models is:

$$\frac{Var(v_{1t})}{Var(v_{2t})} = \frac{c^2(Var(X_2))^2\left\{\frac{Var(Z)}{(E(Z))^4} + \left[\frac{1}{E(Z)} + \frac{Var(Z)}{(E(Z))^3}\right]^2\right\} + 1}{b^2\left\{\frac{Var(Z)}{(E(Z))^4} + \left[\frac{1}{E(Z)} + \frac{Var(Z)}{(E(Z))^3}\right]^2\right\} + 1} = \frac{2.5^2(Var(X_2))^2(0.1125) + 1}{9\,(0.1125) + 1}$$

It can be proved that the size will be measured when $Var(X_2) = 1.2$. The corresponding values of theratio when we measure the power of the test are:

$\sqrt{Var(X_2)}$	$\frac{Var(v_{1t})}{Var(v_{2t})}$
1.2	1
1.5	1.28
2	1.89

Section 3:Results.

3.1. Innovations with no structure (i.i.d)

3.1.1. F-test:

F with i.i.d. innovations (σ=1.5)	Size of validation part (10000 iterations)		
T	n=T/2	n=T	n=2*T
64	43.82	63.9	86.11
128	63.89	85.49	97.66
256	84.63	97.7	99.98
512	97.68	99.96	100
1024	99.97	100	100

F with i.i.d. innovations (σ =2)	Size of validation part (10000 iterations)

T	n=T/2	n=T	n=2*T
64	91.74	99.47	100
128	99.41	100	100
256	100	100	100

F with i.i.d.innovations (σ =1.2)	Size of validation part (10000 iterations)		
T	n=T/2	n=T	n=2*T
64	9.13	9.44	9.77
128	10.26	10.18	9.93
256	9.59	9.57	9.90
512	10.18	10.22	10.12
1024	9.73	9.90	9.75

As we can see, F works well despite the estimates. It is correctly sized even when the sample is small. It works well also when we measure the power of the test.As expected, when the variance of X_2increases, power increases too.

3.1.2.Morgan-Granger-Newbold test:

MGN with i.i.d.innovations(σ =1.5)	size of validation part (10000 iterations)		
T	n=T/2	n=T	n=2*T
64	45.80	64.62	86.09
128	64.85	85.56	97.70
256	85.82	97.79	99.99
512	97.85	99.97	100
1024	99.98	100	100

MGNwithi.i.d. innovations (σ =2)	Size of validation part (10000 iterations)		
T	n=T/2	n=T	n=2*T
64	92.68	99.45	100
128	99.54	99.99	100
256	100	100	100

MGN with i.i.d.innovations (σ =1.2)	Size of validation part (10000 iterations)		
T	n=T/2	n=T	n=2*T
64	10.21	9.94	9.99
128	10.20	10.15	10.15
256	10.33	10.82	9.75
512	10.17	9.68	9.72
1024	10.02	10.23	9.60

As we can see, MGNworks as well as F test does.. It is correctly sized even when the sample is small. It works well also when we measure the power of the test. As in F test, when the variance of X_2increases, power increases too.

3.1.3. Meese-Rogoff test:

MR with i.i.d. innovations(σ =1.5)	size of validation part (5000 iterations)		
T	n=T/2	n=T	n=2*T
64	2.90	65.04	95.64
128	13.02	85.92	99.66
256	36.36	97.92	100
512	75.18	100	100
1024	98.01	100	100

MR with i.i.d.innovations (σ =2)	Size of validation part (5000 iterations)		
T	n=T/2	n=T	n=2*T
64	92.68	99.60	100
128	99.46	100	100
256	100	100	100

MR with i.i.d.innovations (σ =1.2)	Size of validation part (5000 iterations)		
T	n=T/2	n=T	n=2*T
64	8.88	9.54	10.06
128	9.34	10.28	9.92
256	10.48	9.98	9.56
512	9.88	9.98	10.36
1024	10.06	10.22	9.54

Meese-Rogoff test is as correctly sized as F test and Morgan-Granger-Newbold tests are under i.i.d.innovations. The power measured with MR is similar to the previous tests ones.

3.1.4. Diebold and Mariano test:

S_1 with i.i.d.innovations(σ =1.5)	size of validation part (5000 iterations)		
T	n=T/2	n=T	n=2*T
128	16.52	86.32	99.64
256	38.12	97.58	100
512	75.18	100	100

S_1 with i.i.d.innovations (σ =2)	Size of validation part (5000 iterations)		
T	n=T/2	n=T	n=2*T
128	99.46	100	100
256	100	100	100

S_1withi.i.d. innovations (σ =1.2)	Size of validation part (5000 iterations)		
T	n=T/2	n=T	n=2*T
128	11.08	10.06	10.04
256	10.98	9.98	9.78
512	9.88	9.74	10.86
1024	10.74	9.50	10.14

Mariano and Diebold's statistic works better than previous tests. It is as correctly sized as others but it reaches greater percentages of power, even in small samples.

3.1.5. The sign test:

S_2^* with i.i.d.innovations(σ =1.5)	size of validation part (5000 iterations)		
T	n=T/2	n=T	n=2*T
64	(S_2) 19.64	38.20	57.72
128	38.42	54.38	79.26
256	56.62	78.96	94.98
512	78.36	95.68	99.82
1024	94.94	99.70	100

S_2^*withi.i.d. innovations (σ =2)	Size of validation part (5000 innovations)		
T	n=T/2	n=T	n=2*T
64	54.72	85.70	98.38
128	84.60	98.30	100
256	98.04	99.98	100
512	100	100	100

S_2^* with i.i.d.innovations (σ =1.2)	Size of validation part (5000 innovations)		
T	n=T/2	n=T	n=2*T
64	5.94	8.52	9.20
128	7.58	8.96	10.18
256	10.10	9.64	9.90
512	9.10	9.70	9.10
1024	10.42	9.64	10.74

As tables show, $S_2^*$is correctly sized, reaches the significance level when T=128 and n=256 but it does not work as well as Diebold and Mariano's test, which reaches the significance level with smaller samples.Table of powers show that percentages are high but not as high as in the case of Diebold and Mariano test.

3.2. Innovations with structure MA(1).

$$u_t = \frac{1}{\sqrt{1+\theta^2}}(\varepsilon_t + \theta\varepsilon_{t-1})$$

3.2.1. F test:

F with MA(1) innovations (σ =1.5)	Size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	43.67	43.92	43.22	64.27	63.97	64.16	87.58	87.67	87.69
128	64.02	64.30	64.17	85.74	85.68	85.57	98.42	98.45	98.48
256	84.69	84.96	85.05	97.69	97.71	97.72	99.99	99.99	99.99
512	97.68	97.81	97.81	99.96	99.94	99.95	100	100	100
1024	99.97	99.98	99.98	100	100	100	100	100	100

F with MA(1) innovations (σ =2)	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	92.21	92.28	91.66	99.43	99.38	99.40	99.99	100	100
128	99.45	99.39	99.45	100	100	100	100	100	100
256	100	99.99	100	100	100	100	100	100	100

F with MA(1) innovations (σ =1.2)	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	9.16	9.23	9.50	9.38	9.45	9.38	9.78	9.47	10.01
128	9.62	9.30	9.29	9.60	9.90	9.83	9.43	9.61	9.63
256	9.61	10.46	9.88	9.59	9.43	10.00	9.71	9.93	9.79
512	9.57	10.16	9.06	9.60	10.09	9.67	10.21	9.86	9.80
1024	10.04	9.55	9.62	10.12	9.86	10.14	10.27	9.93	10.18

When power is measured, F remains unaffected by the autocorrelation. As observed, if the variance of X_2 increases, the power increases too. F is correctly sized. The presence of autocorrelation makes the size of the test be, in general, below the nominal size but always close to 10% .

3.2.2. MGN test

MGN with MA(1) innovations (σ =1.5)	size of the validation part(10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	45.73	45.62	45.78	65.81	65.64	65.75	88.34	88.40	88.32
128	64.83	65.19	65.24	86.16	86.10	86.07	98.51	98.54	98.57
256	84.95	85.20	85.45	97.76	97.75	97.83	99.99	99.99	99.99
512	97.77	97.86	97.86	99.96	99.94	99.95	100	100	100
1024	99.97	99.98	99.98	100	100	100	100	100	100

MGNwith MA(1) innovations (σ =2)	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	92.37	92.48	92.87	99.49	99.58	99.49	100	100	99.99
128	99.48	99.43	99.56	100	100	100	100	100	100
256	100	100	100	100	100	100	100	100	100

MGN with MA(1) innovations (σ =1.2)	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	10.10	10.47	10.05	10.55	10.53	10.46	10.53	10.72	10.56
128	10.50	9.79	9.74	10.03	10.27	10.13	10.09	10.06	9.94
256	9.90	9.66	9.41	9.78	10.03	9.79	10.14	10.26	10.01
512	10.47	9.48	9.82	10.83	10.17	10.42	10.15	10.08	9.64
1024	10.08	9.57	9.86	10.20	10.01	10.52	10.32	9.93	10.32

The results of MGN are similar to F results. MGN remains practically unaffected by the autocorrelation. As observed, if the variance of X_2 increases, the power increases too as in F test. MGN is also correctly sized.

3.2.3. MR test:

MR with MA(1) innovations (σ =1.5)	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	45.54	45.18	45.30	64.38	64.70	65.06	88.38	88.50	88.44
128	64.96	64.62	64.78	86.00	85.92	86.00	98.56	98.54	98.56
256	84.56	84.90	85.20	97.62	97.60	97.72	100	100	100
512	97.70	97.98	98.06	95.94	95.92	95.92	100	100	100
1024	99.96	99.98	99.98	100	100	100	100	100	100

MR with MA(1) innovations (σ =2)	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	92.48	91.96	92.80	99.40	99.42	99.42	99.98	100	100
128	99.38	99.44	99.38	100	100	100	100	100	100
256	100	100	100	100	100	100	100	100	100

MR with MA(1) innovations (σ =1.2)	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	9.68	9.64	8.98	11.00	9.84	9.98	10.18	10.56	10.98
128	10.38	9.92	10.60	10.08	10.04	10.22	10.78	10.32	10.22
256	9.64	9.92	10.26	10.60	10.66	9.98	10.20	10.04	10.22
512	9.66	10.18	10.42	10.08	10.04	10.04	10.52	10.20	10.22
1024	9.74	9.90	9.44	09.56	10.10	9.96	9.36	10.68	9.52

MR results are similar to F and MGN.In this case, we can observe that in general, the presence of serial correlation, makes the size be above the nominal size but extremely closed to the significance level.

3.2.4. Diebold and Mariano test

S_1 with MA(1) innovations (σ =1.5)	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
128	78.36	66.72	67.04	85.68	85.74	86.08	98.56	98.60	98.62
256	85.38	85.96	85.58	97.80	97.54	97.78	99.96	99.98	100
512	97.36	97.88	97.60	99.94	99.98	99.92	100	100	100

S_1 with MA(1) innovations (σ =2)	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
128	99.54	99.56	99.54	100	100	100	100	100	100
256	100	100	100	100	100	100	100	100	100

S_1 with MA(1) innovations (σ =1.2)	size of the validation part(5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
128	20.66	20.78	20.44	10.74	10.94	10.94	10.60	10.68	10.84
256	10.74	10.62	10.84	10.62	10.78	11.14	10.80	10.02	10.30
512	11.06	10.04	9.76	10.92	10.22	11.12	10.46	10.64	10.26
1024	10.16	10.60	10.25	10.81	10.79	10.42	10.48	10.23	10.51

The power of S_1ishigher than previous tests. When power is measured it is clear than greater percentages are reached in small samples. Furthermore, the test is correctly sized but in this case it starts (with the smaller value of T) with results worse than the other tests. Despite this, it is correctly sized and it tends to be slightly above the nominal size.

3.2.5. The sign test

S_2^* MA(1) (σ=1.5)	Size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	(S_2)19.72	20.24	20.32	38.20	37.96	38.20	56.26	58.82	57.40
128	39.48	38.94	38.58	57.32	57.36	57.02	78.82	79.64	78.24
256	57.36	56.76	57.02	77.86	77.68	78.30	95.30	95.64	95.56
512	77.62	77.66	77.56	95.42	95.26	95.22	99.76	99.88	99.92
1024	95.38	95.26	95.62	99.86	99.88	99.92	100	100	100

S_2^* MA(1) (σ =2)	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	55.94	100	100	86.38	100	100	98.70	98.54	98.46
128	85.24	100	100	97.96	100	100	99.98	99.98	100
256	98.12	100	100	99.98	100	100	100	100	100
512	100	100	100	100	100	100	100	100	100
1024	100	100	100	100	100	100	100	100	100

S_2^* MA(1) (σ =1.2)	size of the validation part(5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9	$\theta = 0$	θ=0.5	θ=0.9
64	5.14	10.38	10.68	9.04	8.42	8.16	9.86	9.98	9.30
128	7.74	8.82	9.46	8.84	9.74	9.38	9.10	9.62	9.30
256	9.76	9.66	9.90	9.56	9.82	9.94	9.68	10.44	10.34
512	9.46	9.00	9.22	10.48	10.46	10.66	9.88	8.96	9.70
1024	9.90	10.32	10.00	8.84	10.16	10.10	10.62	10.48	10.38

As we can see, S_2^* works well. When power is measured, it works worse than S_1. It is also correctly sized.

3.3. Innovations with AR(1) structure: $u_t = \phi u_{t-1} + \varepsilon_t \ , \phi = 0.6$

3.3.1. F test

F with AR(1) innovations ϕ =0.6	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	9.00	41.95	91.09	9.71	61.52	99.23	9.89	86.41	100
128	9.96	61.50	99.19	9.61	83.24	100	9.71	98.09	100
256	9.72	82.22	100	9.57	96.80	100	9.77	99.99	100
512	9.52	98.80	100	9.56	99.89	100	10.19	100	100
1024	9.75	99.95	100	9.38	100	100	10.41	100	100

F test is correctly sized and in general below the nominal size but extremely close to the significance level. It reaches appropriate values of power and when the variance of X_2 increases, the power increases too.

3.3.2. MGN test

MGN with AR(1) innovations ϕ =0.6	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	9.81	43.52	91.45	10.52	63.17	99.28	10.54	87.08	100
128	10.66	62.61	99.23	10.07	83.87	100	10.18	98.20	100
256	10.07	82.63	100	9.97	96.92	100	10.03	99.99	100
512	9.79	96.90	100	9.85	99.89	100	10.39	100	100
1024	10.00	99.95	100	9.63	100	100	10.52	100	100

MGN testworks as well as F test does. It is correctly sized .It reaches appropriate values of powers and when the variance of X_2 increases, the power increases too.

3.3.3. MR test

MR with AR(1) innovations ϕ =0.6	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	9.58	42.88	91.10	9.70	62.42	99.14	10.26	87.16	100
128	10.42	62.40	99.28	9.78	83.82	100	10.30	98.08	100
256	10.32	82.72	100	9.80	96.80	100	9.98	100	100
512	9.68	97.08	100	9.78	99.86	100	10.84	100	100
1024	10.62	99.94	100	9.36	100	100	10.14	100	100

MR results are similar to the results of F and MGN.

3.3.4. Diebold and Mariano test

$S_1$with AR(1) innovations ϕ =0.6	size of the validation part(5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
128	12.04	63.98	99.34	10.56	83.54	100	10.56	98.20	100
256	10.96	83.22	100	10.28	96.88	100	10.06	100	100
512	10.34	97.28	100	10.04	99.84	100	10.88	100	100
1024	10.06	100	100	10.36	100	100	10.15	100	100

S_1works well. It is correctly sized. Size of test tends to be slightly above the nominal size but always close to the significance level. It reaches greater values of powers than other tests and when the variance of X_2 increases, the power increases too.

3.3.5. The sign test:

S_2^* with AR(1) innovations ϕ =0.6	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	(S2)5.10	19.12	53.94	8.52	35.86	83.36	9.52	54.58	97.58
128	8.68	36.66	84.22	9.30	54.26	97.66	10.58	75.86	99.94
256	9.96	54.30	97.50	9.70	74.72	99.96	9.84	94.52	100
512	9.46	74.98	99.98	9.78	93.86	100	10.88	99.60	100
1024	10.42	94.26	100	10.04	99.64	100	10.78	100	100

S_2^* is correctly sized. In this case, the values of power are slightly smaller than other cases. In general, powers calculated are as good as in other cases.

3.4. Innovations with AR(1) structure: $u_t = \phi u_{t-1} + \varepsilon_t \ , \phi = 0.9$

3.4.1. F test:

F with AR(1) innovations ϕ =0.9	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	8.62	33.53	81.33	9.08	49.45	95.94	9.96	78.86	99.98
128	8.89	47.10	95.09	9.14	68.69	99.89	9.47	94.66	100
256	8.91	68.23	99.74	8.56	88.65	100	9.67	99.80	100
512	8.49	88.06	100	8.44	98.76	100	9.65	100	100
1024	8.19	98.67	100	8.18	99.98	100	9.95	100	100

With respect the change of the value of the coefficient, results haven´t change practically. Size of the test has been decreased but it is correctly sized. It is maintained always below the nominal size. The power is as good as it was with the value of coefficient equal to 0.6.

3.4.2. MGN test

MGN with AR(1) innovations ϕ =0.9	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	11.24	37.58	80.22	11.46	50.87	94.34	12.74	68.79	99.17
128	10.68	49.67	95.23	11.60	68.53	99.71	12.19	87.22	99.97
256	10.12	70.30	99.75	10.30	88.93	100	11.32	98.50	100
512	10.45	89.51	100	10.52	98.61	100	10.70	99.98	100
1024	10.84	98.77	100	10.10	100	100	10.68	100	100

MGN is correctly sized. It is always slightly above the nominal size. With respect to the power of the test, it works well. The increase of the persistence parameter has not actually change the results drastically.

3.4.3. MR test

MR with AR(1) innovations ϕ =0.9	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	10.04	35.26	79.46	11.40	50.38	94.56	11.96	68.54	99.06
128	10.26	48.94	95.24	11.32	68.20	99.64	12.14	87.40	100
256	11.22	69.40	99.70	10.56	87.84	100	11.10	98.44	100
512	9.76	89.34	100	11.04	98.80	100	11.46	100	100
1024	10.36	99.06	100	9.98	100	100	10.26	100	100

The MR results are similar to MGN ones.

3.4.4. Diebold and Mariano test:

S_1 with AR(1) innovations ϕ =0.9	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
128	Problemas			10.84	72.12	99.86	10.32	95.36	100
256	11.52	71.84	99.74	11.04	89.62	100	10.62	99.88	100
512	10.88	89.26	100	10.12	98.84	100	10.66	100	100
1024	10.48	99.00	100	9.90	99.98	100	10.56	100	100

As it can be observed, S_1works well. It is correctly sized and is the best test if we compare the power reached with other tests.Power reachs higher percentages than other tests if we compare even samples with equal size.

3.4.5. The sign test

S_2^* with AR(1) innovations ϕ =0.9	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	6.30	15.50	42.10	9.14	28.82	70.02	10.50	43.82	89.66
128	9.12	29.80	70.62	9.76	42.54	90.10	10.56	61.00	99.12
256	9.76	42.54	90.64	9.36	61.28	99.42	10.60	83.26	99.96
512	10.30	62.80	99.40	10.04	83.98	100	9.72	97.26	100
1024	10.68	84.42	100	9.76	97.24	100	10.28	99.96	100

$S_2^*$is correctly sized but in this case, as happens with the another value of the parameter of the AR(1) process, the values of power are slightly smaller than other cases. New value of the persistence parameter has not led us to drastic changes in the results.

3.5 .Innovations with heteroscedasticity

Under heteroscedasticity of innovations, when we test the size of F, MGN and MR test, the results are not as good as in previous contexts. All of them show an increasing size as T increases. However, S_1 and S_2^* tests work correctly.

All of five tests work well when we measure the power of tests. S_1works slightly better than others.

3.5.1. F test

F	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	9.52	47.23	94.48	9.25	69.02	99.82	10.19	89.63	100
128	9.41	69.13	99.72	10.82	88.98	99.99	10.78	98.55	100
256	10.23	88.94	100	10.35	98.15	100	11.74	99.96	100
512	10.35	98.24	100	12.40	99.84	100	13.31	100	100
1024	11.96	99.94	100	14.30	100	100	14.29	100	100

3.5.2. MGN test

MGN	size of the validation part (10000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	10.21	49.82	94.61	10.42	71.08	99.76	10.83	89.94	100
128	10.14	69.50	99.70	10.52	89.10	100	11.00	98.36	100
256	10.06	88.54	100	11.18	98.09	100	12.21	99.92	100
512	9.92	98.18	100	12.77	99.82	100	13.51	100	100
1024	12.40	99.99	100	14.26	99.99	100	13.80	100	100

3.5.3. MR test

MR	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	9.38	48.36	94.61	9.62	70.90	99.76	10.30	90.22	100
128	9.68	69.66	99.70	10.58	88.34	100	11.72	98.34	100
256	10.40	88.42	100	11.70	97.92	100	12.32	99.94	100
512	10.12	98.36	100	13.04	99.90	100	13.42	100	100
1024	12.66	99.92	100	14.26	100	100	14.08	100	100

3.5.4. Diebold and Mariano test

S_1	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
128	11.40	70.04	99.64	10.82	89.48	100	11.24	98.40	100
256	9.88	88.04	100	10.26	98.00	100	10.52	99.94	100
512	10.50	98.14	100	9.98	99.92	100	10.72	100	100
1024	9.96	99.94	100	9.64	99.96	100	10.74	100	100

3.5.5. Thesign test

S_2^*	size of the validation part (5000 iterations)								
	n=T/2			n=T			n=2*T		
T	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 1.2$	$\sigma = 1.5$	$\sigma = 2$
64	5.52	21.20	53.26	8.82	40.74	88.80	9.38	62.24	99.20
128	8.74	41.08	88.30	9.24	61.90	99.08	9.74	82.48	99.98
256	9.54	60.54	98.70	9.38	81.62	100	10.16	95.52	100
512	9.12	80.90	99.98	9.98	95.80	100	10.46	99.86	100
1024	10.18	96.96	100	9.78	99.90	100	10.42	100	100

Section 4: An empirical example.

Finally, all tests are going to be applied with real data. In order to follow this section, see Audrino and Medeiros (2011).

In this paper, authors propose a smooth transition tree model for both the conditional mean and variance of the short-interest rate process. All equations have dynamic coefficients. The estimation of such models is addressed and they work the quasi-maximum likelihood estimator. This means that they want the estimates to be efficient.

Our models estimated are not as sophisticated as Audrino and Medeiros (2011) because our task is focused on comparing accuracy predictions.

The data used in this section is provided by this recent publication and all variables have been downloaded from Datastream Results.

The data used in this study are, for all variables, 564 monthly observations which correspond with the time period between January 1960 and December 2006. All data refer to US.

The variables they use are:

$*$ $\boldsymbol{Y_t}$:one-month US Treasury bill rates downloaded from the Fama CRSPTreasury bill files. It is necessary to take first differences. It is the endogenous variable.

Exogenous variables:

$*$ $(\mathbf{60\_}\boldsymbol{month})_t$:annual zero coupon bond yield from the FAMA CRSP bond files.

$*$ $(\boldsymbol{CPI})_t$&$(\boldsymbol{PPI})_t$:inflation and (finished goods) inflation, respectively. It is calculated at time t as$log\,\frac{P_t}{P_{t-12}}$ where P_t is the (seasonally adjusted) inflation index.

$*$ $(\boldsymbol{HELP})_t$:index of help wanted advertising in newspapers.

$*$ $(\boldsymbol{IP})_t$: (seasonally adjusted) growth rate in industrial production.It is calculated at time t as $log\,\frac{I_t}{I_{t-12}}$ where I_t is the (seasonally adjusted) industrial production.
$*$ $(\boldsymbol{UE})_t$:unemployment rate.

$*$ $(\boldsymbol{GDP})_t$:the US gross domestic product.

Authors consider $Y_t$as the endogenous variable .When the author's model is applied to the US short-term interest rate, they find leading indicators for inflation and real activity are the most relevant predictors in characterizing the multiples regime's structure. Furthermore, after comparing the goodness-of-fit of the models they suggest, they find that relevant variables are HELP, GDP and PPI.

First of all we analyze our data and we regress the endogenous variable fitting a linear model with constant term an including all exogenous variables:

$$(M0)\ Y_t = \alpha_0 + \beta_1(60_{month})_t + \beta_2(HELP)_t + \beta_3(CPI)_t + \beta_4(IP)_t + \beta_5(UE)_t + \beta_6(GDP)_t + \beta_7(PPI)_t + \beta_8(D_1)_t + \beta_9(D_2)_t + u_t$$

D_1andD_2 are two categories of two dummies variables. On the one hand, D_1takes value equal to one if the observation corresponds to the period between years 1973-1975 (OPEC oil crises). It takes zero in other case. On the other hand, D_2takes value equal to one if the observation corresponds to the period between years 1979-1982 (Fed experiment). It takes zero in other case. Both have been introduced following the corresponding paper.

The corresponding fit is:

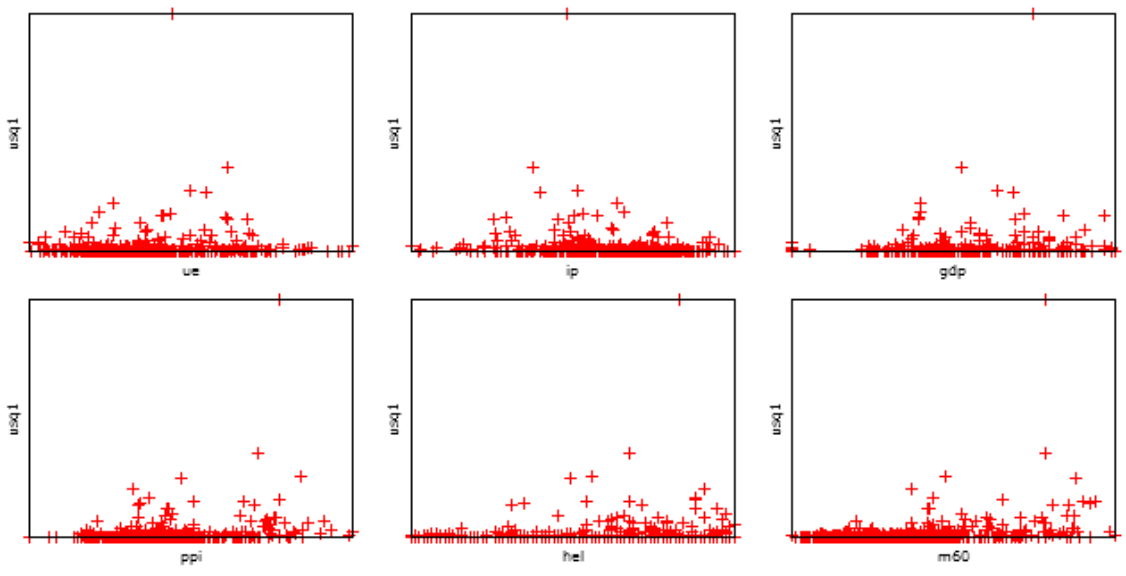
$$\hat{Y}_t = -0.06 - 0.229\,(60_{month})_t + 0.002\,(HELP)_t + (-0.0434)(CPI)_t + (-0.0014)(IP)_t \\ + 0.0004\,(UE)_t + 0.016\,(GDP)_t + 0.02922(PPI)_t - 0.15(D_1)_t + 0.092(D_2)_t$$

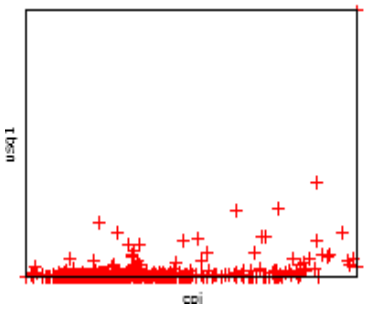
Once the linear model including all exogenous variables has been estimated, we analyze associated residuals and notice indicators of the presence of heteroscedasticity.

Due to this results we proceed by two ways.

Alternative A

We evaluate Breusch and Pagantest and it shows signs of heteroscedasticity. We find relevant variables to explain the heteroscedasticityare 60_{month}, GDP, CPI, D_2.We conclude this after comparing the squares of residuals of regression $(M0)$ with variables and trying several fits to predict the value of the squares of regression $(M0)$. The next figures represent the squares of residuals of model $(M0)$ against each exogenous variable:



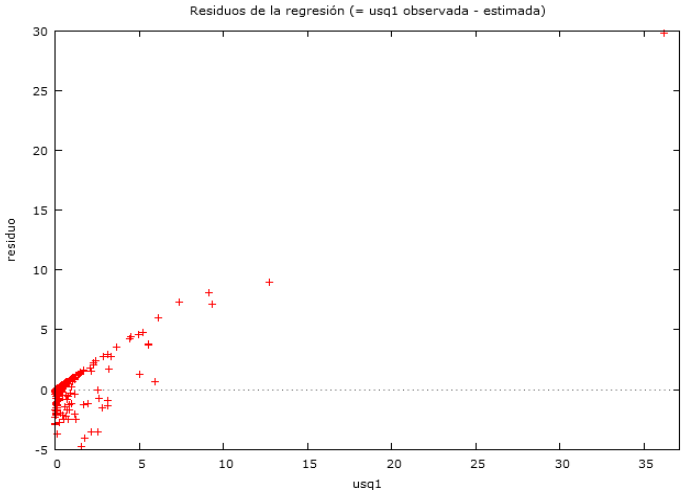


An a priori argument, we cannot choose which variables explain the heteroscedasticity by seeing the figures above.
In order to obtain the estimate values of the time series corresponding to the variance of the innovations in model $(M0)$, we regress the squares of the residuals obtained fitting model $(M0), \hat{u}_t^2$, using variables $60_{month}, GDP, CPI, D_2$.

After severalattempts, we find that the values of the time series corresponding to the variance of the innovations in model $(M0)$ can be explained by next model,

$$\hat{u}_t^2 = 0.00079(GDP)_t{}^2 + 0.00018(CPI)_t{}^4 + \hat{v}_t$$

which fits well with real value of $\hat{u}_t^2$, quality represented in next figure, which shows $\widehat{\hat{u}_t^2}$ against $\hat{u}_t^2$:



Once the previous regression has been done, we consider its corresponding estimated values,$\hat{\sigma}_t^2$,as the variance responsible of the heteroscedasticity.

We transform model $(M0)$ by dividing it by$\sqrt{\hat{\sigma}_t^2}$ to get a new model with constant innovations variance and to be able to do an OLS regression. This is, we applied Feasible Generalized Least Squares Method. After doing this, we notice that themost relevant (transformed) variables in this model are PPI and CPI. Actually, all variables (with the exception of dummies) are relevant but not as relevant as PPI and CPI.

After comparing some models we find that the variables with less relevance are$HELP, D_1, D_2$, opposed to the authors, who conclude that $HELP$ is one of the macroeconomic relevant variables to explain, with their proposed model, changes in short term interest.
We propose next two models for being compared by using all of the tests of accuracy prediction explained in this paper:

$$(M1)\ Y_t = \alpha_0 + \beta_1(60_{month})_t + \beta_2(PPI)_t + \beta_3(IP)_t + \beta_4(UE)_t + \beta_5(GDP)_t + \beta_6(D_1)_t + \beta_7(D_2)_t + \beta_8(CPI)_t + u_{1t}$$

$$(M2)\ Y_t = \mu_0 + \delta_1(60_{month})_t + \delta_2(HELP)_t + \delta_3(CPI)_t + \delta_4(IP)_t + \delta_5(UE)_t + \delta_6(GDP)_t + \delta_7(D_1)_t + \delta_8(D_2)_t + u_{2t}$$

Next table contains the tests results, considering significance level equal to 10%:

	F	MGN	MR	S_1	S_2
T=503, n=60	Does not reject null hypothesis	Does not reject null hypothesis	Does not reject null hypothesis	Does not reject null hypothesis	Does not reject null hypothesis
T=n=281	Does not reject null hypothesis	Rejects null hypothesis	Rejects null hypothesis	Rejects null hypothesis	Rejects null hypothesis

Alternative B

We directly estimate a GARCH(1,1) model in order to get the time series corresponding to the variance of the error. The estimated model obtained is:

$$\hat{Y}_t = -0.0066 - 0.006\,(60_{month})_t + (-0.000749)(HELP)_t + (-0.014)(CPI)_t + (0.0098)(IP)_t + 0.0011\,(UE)_t + 0.019\,(GDP)_t + 0.017(PPI)_t - 0.1382(D_1)_t + (-0.166)(D_2)_t$$

$$\hat{\sigma}_t^2 = 0.0054 + 0.2338\,\hat{u}_{t-1}{}^2 + (0.7662)\hat{\sigma}_{t-1}^2$$

Once we know $\{\hat{\sigma}_t^2\}$, we proceed as in Alternative A, transforming the linear model to another with innovations with constant variance and continue by estimating the transformed model applying OLS (FGLS). After doing this, we found that the unique relevant variable is $HELP$, opposed toAlternativeA results. The second variable in order of relevance is PPI but it has poor explication power.

Following this new results, we propose next two models for being compared by using all of the tests of accuracy predictions explained in this paper:

$$(M1)\ Y_t = \alpha_0 + \beta_1(60_{month})_t + \beta_2(PPI)_t + \beta_3(IP)_t + \beta_4(UE)_t + \beta_5(GDP)_t + \beta_6(D_1)_t + \beta_7(D_2)_t + \beta_8(CPI)_t + u_{1t}$$

$$(M2)\ Y_t = \mu_0 + \delta_1(60_{month})_t + \delta_2(HELP)_t + \delta_3(CPI)_t + \delta_4(IP)_t + \delta_5(UE)_t + \delta_6(GDP)_t + \delta_7(D_1)_t + \delta_8(D_2)_t + u_{2t}$$

Second model is expected to be more accurate but all tests do not lead us to this conclusion.

Next table contains the tests results, considering significance level equal to 10%:

	F	MGN	MR	S_1	S_2
T=503, n=60	Does not reject null hypothesis	Rejects null hypothesis	Rejects null hypothesis	Rejects null hypothesis	Rejects null hypothesis
T=n=281	Does not reject null hypothesis	Does not reject null hypothesis	Does not reject null hypothesis	Rejects null hypothesis	Does not reject null hypothesis

5. Conclusions

We have proposed several tests to compare the accuracy prediction of two competing models. Our task has been to check if these tests conclude good results even if we change the hypothesis of the innovations terms of models. We have permitted innovations to have structure of serial correlation and contemporaneous correlation. The correslation has been due to processes with short memory and with long memory with respective MA or AR processes. We have also analyzed if the increase persistence has a drastic effect on the results, but we conclude that it has not. Even, we have permitted heteroscedasticity structure in the innovation term.

All of the checkings of the tests effectivenesshave been studied in an estimation context. However, in this part of the paper all variables and random innovation terms have been simulated.

In the case of heteroscedastic innovations, it has been necessary to transform the model proposed in order to get another one with innovations with constant variance (GLS), which was not necessary if we only permit correlation in innovation terms.

In spite of permitting correlation in the innovations terms and heteroscedasticity, the corresponding results show the good functioning of all tests, although there were slightly worse in case of heteroscedasticity of innovation terms.

When we set the tests out using real data, we are based on a model recently publicated, Audrino& Medeiros (2011). In this case we notice the presence of heteroscedasticity. It has a more complicated structure than the simulated case.After stablishing two alternatives to capture the heteroscedasticity with proposed models for modeling the variance of the innovation term, we propose two models to be compared expecting one of them to be more

accurate. Results of test show that some of them detect this difference between the accuracy of the models proposed but not all of them.

However we cannot establish a pattern in which these test work correctly based on the length of the validation part , n . Withalternative A, tests which detected the difference in the accuracy prediction work well when n=281 but, in general, with alternativeB, occurs the opposite.

Although heterocedasticity has been taken into account, we are conscious that the systematic part of the model has not been estimated in such a sophisticated way as authors. We include the estimation ofa GARCH model to capture the heterocedasticity but they included a GARCH model with dynamic coefficients. We suspect that the lack of these dynamic coefficients can be a reason of our conclusions.

Also, we have great parametric uncertainty because the number of variables is high and some of them present correlation.

In conclusion, all tests for comparing accuracy prediction of modelshave better functioning when data are simulated and do not work as well when variables are extracted from a real source.

To end, we leave the option to analyze this subject posing a more complex heteroscedasticitycase and posing dynamic coefficients in the systematic part of the model.

6. Bibliography

Francis X. Diebold and Robert S. Mariano (1995), " Comparing Predictive Accuracy", Journal of Business and Economic Statistics, 13, 253-265.

Eva Ferreira and Winfried Stute (2009), "Testing for differences in Predictive Accuracy", Pak. J. Statist. Vol 25(4), 403-417.

Eva Ferreira and Fernando Tusell (1996), "Un modelo aditivo semiparamétrico para estimación de capturas: el caso de las pesquerías de Terranova", Investigaciones económicas. Segunda Época. Volumen XX(1).

Francesco Audrino and Marcelo C. Medeiros (2011), " Modeling and forecasting short-term interest rates: the benefits of smooth regimes macroeconomic variables and bagging", Journal of Applied Econometrics, 26, 999-1022.