



**EKONOMIHÖGSKOLAN**  
Lunds Universitet

## **The model confidence set – choosing between models**

Kandidatuppsats i nationalekonomi  
Av: Jeanette Johansson  
Handledare: Hossein Asgharian  
Datum: 28 Oktober, 2005

## *Abstract*

My data consists of closing prices between January 2000 and December 2004. My interest is, by looking at the return, to apply the Model Confidence Set. I work with ten volatility models. The Model Confidence set is analogous to a confidence interval of a specific parameter. This means that the Model Confidence Set is used to choose model/models that is/are considered to be “best”. The empirical exercise is based on ten models, and by applying the method I came to the conclusion that the Asymmetric Garch (1,1) was inferior compared to the other models.

Keywords: forecasting error, hypothesis testing, model selection and forecasting

## 1) Introduction

Forecasting is today of great importance in economics. It is important to be precise when one is forecasting because forecasts are used to guide decisions. The precision of forecasting is also important to those who produce forecasts because the producers' reputations (and fortunes) rise and fall with forecast accuracy (Diebold and Mariano 1995).

Traditional research within financial economics focused on the mean of stock returns. Today researchers are more interested in the volatility of these returns. Volatility forecasting is used within several areas. Analysis of market timing decisions and portfolio selection are two examples (Brailsford and Faff, 1996). Another area is comparison of forecast accuracy. This area is of importance to economists who are interested in comparing competing economic models. The focus of this paper is on the forecasting precision of 20-day index market volatility from 10 different statistical models. The basic methodology involves the estimation of parameters of each model using an initial set of data (daily returns).

There is a large number of conditional volatility models to dispose of today. The first two were the ARCH model of Engle (1982) and the GARCH model of Bollerslev (1986). A volatility function can be composed into two parts, one predictable component and one unpredictable. Researchers have focused on the predictable part.

The conditional variance of a time series is important for pricing derivatives, calculating risk and hedging. The conditional variance cannot be observed. That is why it is estimated by the realized variance. Initially the conditional variance was estimated by the squared return. This estimate typically resulted in a poor performance. By using the realized variance as an estimate, instead of the squared return, Andersen & Bollerslev (1998) showed that volatility models perform well. Another argument for using the realized variance instead of the squared return was introduced by Hansen & Lunde (2003). They show that if one is using the squared return as the estimate for the conditional variance the empirical ranking may be inconsistent for the true ranking. Several papers have tried to answer the following question: Which model is the best? There are a couple of reasons to why it is difficult to establish if there exists a "best" model:

- 1) Asset returns often do not contain enough information to identify a single volatility model as “best”. (Hansen, Lunde & Nason, 2003).
- 2) The set of competing models is large. (Hansen, Lunde & Nason, 2005)

Volatility models attempts to explain the changing behaviour of the volatility in the daily return of my chosen index. As I mentioned before there exist a number of papers discussing the evaluation and comparison of volatility models. One example is Hansen & Lunde (2001) who report that there is no evidence that the GARCH (1,1) model is outperformed by other models in an analysis of exchange rates. But, on the contrary, when the analysis evaluates IBM stock returns the conclusion is that the GARCH (1,1) is inferior to the other volatility models in  $M_0$ .

Another paper examining model selection techniques is the one of Brailsford and Faff (1993). They conducted an extensive model exercise based on Australian data. They examined asymmetric volatility models and found support for the Glosten, Jagannathan & Runkle (1993) modified GARCH model. Specifically, they came to the conclusion that the GJR-GARCH (3,1) was the superior model. The Poon and Granger study examines 93 papers. Seventeen of these compare alternative versions of GARCH. They concluded that GARCH-models dominated the ARCH model.

The purpose of my paper is to determine which forecasting model/models that is/are significantly inferior from a set of competing models. The starting point is to reduce the set of competing models to a smaller set of models, the Model Confidence Set (MCS). The Model Confidence Set is similar to the confidence interval of a parameter. A confidence interval is a sequence of numbers in which the true number is believed to lie with, in my case, a 95% probability. The interval can be interpreted in different ways. One way is by looking at the midpoint of the interval. The value in the middle can be seen as the “best” guess for the true answer. Thus, the MCS can be seen as the approach of choosing the “best” forecasting model/models.

The Model Confidence Set method has a couple of advantages compared to methods involving a single model. It doesn't require that the “true” model is among the set of competing models,  $M_0$ . A MCS doesn't rule out a model unless it is found to be significantly inferior compared to the other models. Also, in reality it is often so that *two or more* models are considered to be equally likely. That is why it is more appealing to work with a set of forecasting models. This implies that MCS dominates methods that

require a single model being selected as “best”. A test that is quite similar to the MCS is the test for superior predictive ability (SPA). SPA can test if other models significantly outperform a particular “benchmark” model. A simple random walk model is the benchmark model in Hansen (2001). The benchmark is compared to a large number of regression-based forecasts and the empirical result showed that the random walk model was significantly outperformed. However, it is worth to point out that the Model Confidence Set can consist of a single model (Hansen, Lunde and Nason 2003).

My set of competing models consists of the following models: symmetric and asymmetric Garch (1,1), symmetric and asymmetric Garch (1,2), symmetric and asymmetric Garch (2,1), Random walk model, Historical mean model, Moving Average model and an Exponential Smoothing model. To construct a Model Confidence set I use a test for equal predictive ability (EPA). The Pantula (1989) testing method is used. This test erases the models that are found to be statistically significant, that is rejecting the null hypothesis (Hill, Griffiths & Judge). The null hypothesis is explained in section 3.2. The set of surviving models,  $M^*$ , is the MCS. This set contains, with a 95 % probability, the ‘best’ model.

The volatility models examined in this paper are mentioned above. I implement the Model Confidence set with this set of models. There is one finding of my empirical analysis: When testing the null hypothesis I could delete the asymmetric GARCH (1,1) model. That is, I find no evidence that the asymmetric GARCH (1,1) is superior to the other models.

My data are daily closing prices from the Stockholm Stock exchange general index between January 1 2000 and December 31 2004. I have chosen to base my empirical part only on this index.

This paper is organized as follows. Section 2 describes the theory of some basic calculations I am going to make. Section 3 contains Introduction to the Model Confidence Set and hypothesis testing. In section 4, I describe my ten chosen volatility models. Section 5 describes my data and the empirical results that were obtained. Section 6 is the discussion. Section 7 concludes.

## 2 Some basic calculations

*In this part I am going to describe some basic calculations I have made to be able to perform the Model Confidence Set approach. The compounded return are based on closing prices.*

### 2.1 Compounded return

Let  $\{p_t\}$  be the closing price on day  $t$ . My analysis involves 20-day volatility forecasts. The continuously compounded daily returns are calculated as:

$$R_{m,t} = \ln(p_{m,t} / p_{m,t-1}) \quad (2.1.1)$$

$p_{m,t}$  is the closing price from Stockholm stock exchange general index and  $R_{m,t}$  is the continuously compounded return on trading day  $t$  in month  $m$ .

### 2.2 How to determine the forecasting error.

My sample is divided into an estimation period and an evaluation period where the estimation period and evaluation period is one year respectively (Hansen, Lunde & Nason 2003). The parameters of the volatility models are estimated using the calculated daily returns above. These estimates (based on one year) are used to make 20-day-ahead forecasts for the following year (Hansen and Lunde 2001). My estimates for year 2000 are used to make 20-day forecasts for 2001. The same idea applies for 2001, 2002, 2003 and 2004. Thus model  $m$ ,  $m=1, \dots, 10$  gives a sequence of forecasts denoted by  $h_{t,m}^2$ .

Estimation period
Forecasting period

The purpose is to determine the forecasting error. The forecasting error is the difference between the forecast for model  $m$  at time  $t$  and the conditional variance. I have calculated the variance from the daily returns. The conditional variance is this variance at time  $t$ .

The conditional variance,  $\sigma_\tau^2$ , is not known so an estimate is used. This estimate is given by

$$\hat{\sigma}_\tau^2 = \frac{1}{N-1} \sum_{\tau=1}^N (R_\tau - \bar{R})^2 \quad (2.2.1)$$

where  $\bar{R}$  is the mean return in year a. a=2000...2004

$\sigma_\tau^2$  is the realized variance, which is a precise measure of  $\sigma_\tau^2$ . I would like to point out that it is important to use accurate measures of the conditional variance to avoid an inconsistent ranking of the alternatives (Hansen & Lunde 2003).

Each of the forecasts was compared to  $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_{10}^2$ , using a loss function L. Let the first model, m=1, be a fix model that is compared to the rest of the models. Each model leads to a sequence of losses

$$L_{m,\tau} \equiv L(h_{m,\tau}^2, \sigma_\tau^2) = (h_{m,\tau}^2 - \hat{\sigma}_\tau^2)^2. \quad (2.2.2)$$

$\tau=1,2,\dots,t$   $\tau=1$  is the first 20-day forecasting period.  $\tau=t$  is the last 20-day forecasting period.

Equation (2.2.2) is known as the mean square error, MSE. (Hansen, Lunde & Nason 2003) rank models according to their expected loss using the MSE.  $h_{m,\tau}^2$  is model  $i$ 's forecast of  $\sigma_\tau^2$ .

### 3 Theory behind the Model Confidence Set

*The approach of the MCS involves hypothesis testing. I will explain how this test works and how to decide whether to accept or reject a null hypothesis. I will also mention the main idea with the Model Confidence Set. Hansen, Lunde and Nason (2003) implement the Model Confidence Set with a study of volatility models of return.*

#### 3.1 Introduction to the model confidence set

The empirical problem is to determine which models are in  $M^*$  and which are not. It is known that a confidence interval is an interval that contains a specific parameter with probability  $1-\alpha$ . The same interpretation applies for the MCS. The MCS contains the best model/models with probability 95 %, where  $\alpha=0,05$  is the level of significance. Thus, the MCS is a random data-dependent set of models that includes the best forecasting model(s), as a standard confidence interval covers the population parameter with a certain probability.

#### 3.2 Hypothesis testing

My null hypothesis is  $H_0: E(d_{i,j,\tau}) = 0$ . The interpretation of this is that one is testing the null hypothesis that the expected value of the difference between the two loss functions is 0 (Diebold and Mariano). That is, I want to test the null hypothesis that the benchmark model is as good as any other model in terms of the MSE.

The alternative hypothesis is  $H_1: E(d_{i,j,\tau}) \neq 0$ . I want to test the null for my set of candidate models  $M_0$ . When testing  $H_0$  I used the t-test. My test value can be expressed as

$$t_{i,j} = \frac{d_{i,j-0}}{se(d_{i,j,\tau/50})} \quad (3.2.1)$$

where se denotes the standard error.

Thus, the t-test is a way to test if a model is significant.

How does one know whether to reject or accept a null hypothesis? The null hypothesis,  $H_0$  is the same as an statement.  $H_0$  is tested against an alternative hypothesis denoted  $H_1$ .



The theory of hypothesis testing is concerned with developing rules for deciding whether to reject or not reject the null hypothesis. The rule applied in this paper is the confidence interval. Since I use the t-test in this paper the confidence interval becomes:

$\Pr [-t_{\alpha/2} \leq \text{test value} \leq + t_{\alpha/2}] = 1-\alpha$ .  $t_{\alpha/2}$  is called the critical value.

If my test value lies within the critical value I will *not reject* the null hypothesis (Gujarati). Not rejecting the null hypothesis implies that t is a “statistically significant” statistic (Hill, Griffiths and Judge 1997).

When I performed this test I could eliminate the Asymmetric Garch (1,1)-model. Thus I have decided which models that belong to the model confidence set.

### 3.3 Approach to determine the worst performing model.

Section 2.2 described the forecasting error and how it was determined. The definition of the forecasting error was given by equation (2.2.2). The coming equation is extended to compare the existing models. That is, I want to study the forecasting error but this time between models.

The loss differential between models  $i$  and  $j$  is given by

$$d_{i,j,\tau} = L(h_{i,\tau}^2, \hat{\sigma}_{\tau}^2) - L(h_{j,\tau}^2, \hat{\sigma}_{\tau}^2) \quad (3.3.1)$$

$$\forall i, j \neq i ; i, j = 1, \dots, m \quad , \tau = 1, \dots, t$$

The first term is the deviation for model  $i$  from the conditional variance. The second term is the deviation for model  $j$  from the conditional variance. Then the difference between these two terms was calculated.

To be able to determine the worst performing model it was necessary to define the following two equations and equation (3.3.1).

I define the variables

$$1) \bar{d}_{i,j} = (\sum_{\tau=1}^{50} d_{i,j,\tau}) / 50 \quad (3.3.2)$$

$$2) d_i = \left[ \sum_{\tau=1}^{50} \sum_{j=1}^{m-1} d_{i,j,\tau} \right] / [50*(m-1)] \quad (3.3.3)$$

Equation (3.3.2) measures how model  $i$  perform in proportion to model  $j$ . Model  $i$  is fix and model  $j$ , where  $j=1..9$  are every possible combination together with model  $i$ .

Equation (3.3.3) can be seen as how model  $i$  perform in proportion to the average of all the models in  $M$ .

$M=M_0$  and is the full set of competing models.

$$\text{var}(\bar{d}_i) = \frac{1}{50*(m-1)} \text{var}(d_{i,j,\tau}) \quad (3.3.4)$$

where

$$\text{var}(d_{i,j,\tau}) = \text{var}(\text{all observations in } d_{i,j,\tau} \text{ for } \forall j). \quad (3.3.5)$$

Equation (3.3.5) was interpreted as the variance of all the models observations. The measure Hansen, Lunde and Nason (2003) use to rank the models in  $M$  is

$$v_{i,M} \equiv \frac{\bar{d}_i}{se(\bar{d}_i)} \quad (3.3.6)$$

Thus, the larger the statistic  $v_{i,M}$  is the worse is this model performing and its rank decreases.

The worst performing model has the highest  $v_{i,M}$

#### 4)Volatility models

*In this part I will talk about my ten volatility models. I will also try to make an interpretation of what the models mean. There is a considerable number of models used for forecasting volatility. I have chosen to examine both simpler models as the random walk and the historical mean model as well as more complex models.*

##### 4.1 Random walk model

Under a random walk model, the best forecast of this period volatility is the last period observed volatility (Brailsford and Faff 1996).

$$\hat{\sigma}_{\tau}^2 (\text{RW}) = \sigma_{\tau-1}^2 \quad (4.1.1)$$

$$\tau=1,2,3,\dots,t$$

The random walk model says that the forecasted volatility should be equal to the last observed volatility, which is a volatility that is already known. In practice this mean that anyone can speculate in for example stocks and make money on it (Gujarati 2003).

The forecasts was calculated for one year respectively. Each year consists of fifty 20-day forecasts. Thus  $\tau=1,2,3,\dots,50$ .  $\tau=1$  is the first 20-day forecast and  $\tau=50$  is the last 20-day forecast.

##### 4.2 Historical mean model

Under the assumption of a stationary mean, the best forecast of this period volatility is a long-term average of past-observed volatilities. The assumption of a stationary mean is equivalent to a process with a constant variance (Brailsford and Faff 1996).

$$\hat{\sigma}_{\tau}^2 (\text{LTM}) = \frac{1}{\tau-1} \sum_{j=1}^{\tau-1} \sigma_j^2 \quad (4.1.2)$$

The random walk model made a forecast based on the last observed volatility. Notice the difference from the historical mean model, which forecast with an average of past-observed volatilities.

A 20-day forecast of the volatility is equal to the average of all past observed 20-day forecasts.

#### 4.3 Moving average (MA- $\alpha$ ) model

Market analysts often use a moving average as a predictor of mean returns. The moving average technique is often used in traditional time-series analysis. The choice of the moving average estimation period was chosen arbitrary ( Brailsford and Faff 1996).

The forecast of the moving average model is expressed as

$$\hat{\sigma}_{\tau}^2 (\text{MA}) = \frac{1}{\alpha} \sum_{j=m-\alpha}^{m-1} \sigma_j^2 \quad (4.1.3)$$

$$\tau = 1, 2, 3, \dots, t$$

The moving average period was set to six. The term on the right side of equation (4.1.3) imply that the moving average model is forecasted with an average of the last six 20-day forecasts.

The moving average model focuses on the most six recent observed 20-day volatilities. The historical mean model focuses on all the past-observed volatilities and calculates the mean of these.

Imagine that the volatility for the first 20-day period is extremely high. Because the historical mean model will take this high volatility in consideration it could result in a poor forecast. This could be a reason to why the moving average model is a better volatility model compared with the historical mean model.

#### 4.4 Exponential smoothing model

According to Dimson and March (1990), an exponential smoothing model is used to forecast volatility. This model is also designed to track changes in the volatility (Gujarati 2003).

$$\hat{\sigma}_{\tau}^2 (\text{ES}) = \Theta \hat{\sigma}_{\tau-1}^2 (\text{ES}) + (1 - \Theta) \sigma_{\tau-1}^2 \quad (4.1.4)$$

The first term on the right side is the last periods forecast and the second term is the last periods observed volatility (Brailsford and Faff 1996).

The smoothing parameter,  $\theta$ , is constrained to lie between zero and one. The optimal value of  $\theta$  must be determined empirically. I let Eviews estimate this parameter. If it is zero, the right hand side of the equation will only consist of the realized variance, that is the exponential smoothing model collapses to the random walk model.

The 20-day forecast is a function of the last 20-day forecast and the last 20-day observed volatility.

The four models described so far were used to model the levels of the previous data. This means that the forecasts calculated the future variance from previous variance. Thus, the variance didn't change over time. Instead it is interesting to study the fact that the variance of a time series does vary over time. How can one model such "varying variance"? This was the reason that the symmetric GARCH model and asymmetric GARCH model was chosen as volatility models.

#### 4.5 Symmetric Generalized autoregressive conditional heteroskedasticity (GARCH)(p,q)

The GARCH model is a generalization of the ARCH model. Heteroskedasticity implies that a time series doesn't have a constant variance (Gujarati 2003) that is it doesn't behave like the Random walk model or any of the other models described earlier in this section.

The GARCH process introduced by Bollerslev (1986) recognizes the difference between the unconditional and the conditional variance allowing the latter to change over time as a function of past errors. The assumption underlying a GARCH model is thus that the volatility (variance) change as time goes on. During some periods volatility is relatively high; during other periods it is relatively low.

The conditional variance can be stated as

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{k=1}^q \alpha_k u_{t-k}^2 \quad (4.1.5)$$

If the lag length parameters are set to 1 this implies that the conditional variance is calculated with the most recent observation of the past error,  $u_{t-k}^2$ , and the most recent 20-day forecasting volatility,  $\sigma_{t-j}^2$ . If the sum of my estimated parameters,  $\alpha$  and  $\beta$  are larger than 1 then the Garch process diverges.  $\alpha$  is the weight connected to the past forecasting error.  $\beta$  is the weight assigned to the conditional variance and  $\omega$  is the weight connected to the unconditional variance.

#### 4.6 Asymmetric GARCH (p,q)

The conditional variance is expressed as

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{m=1}^q \alpha_m u_{t-m}^2 d_{t-m} + \sum_{k=1}^q \alpha_k u_{t-k}^2 \quad (4.1.6)$$

$p$  and  $q$  are called lag length parameters. I estimated the symmetric and asymmetric Garch-models over  $p,q=1,2$ . The dummy variable is just a constant with the value 1.  $u_{t-k}^2$  is the (squared) forecasting error connected to the Arch term in the model.

## 5) Empirical part using the Model Confidence Set

*This section accounts for my data and empirical results. I will explain how the data is used to make forecasts for my chosen volatility models. In the section for the empirical result I will mention which model/models that is/are considered to be inferior. I will also try to explain why a model is inferior.*

### 5.1 Data and method

My data are daily closing prices from the SAX-index (Stockholm stock exchange). My empirical part is based on the realized variance. The realized variance was based on closing prices with a sample period that runs from January 1 2000 through December 31 2004. The data was collected from Six Trust database.

When the data was absolute I decided the estimation period and the forecasting period. The estimation period has a time horizon of one year that is I have estimates for 2000, 2001, 2002, 2003 and 2004 respectively. Based on my estimates for 2000 I make a forecast for 2001. The same applies for the other years. At the end of each forecasting period I calculated the forecasts of the realized variance. The forecasts is denoted by  $h^2_{t,m}$ . The realized variance is a function of the return  $r$  and the mean  $\mu$ . The mean is calculated for each year.

The first step was to estimate the parameters of my chosen volatility models. Based on the daily return data from 2000 I estimated the parameters. Note that the daily return was calculated from equation 2.1.1.

These parameters are then used to make 20-day-ahead forecasts for 2001. Thus, my first 20-day-ahead forecast is from January 1 2001 to January 20 2001. By this method I get *fifty* 20-day forecasts for 2001.

Then, I estimate the same parameters, but this time based on my daily return data from 2001. In the same way I make 20-day-ahead forecasts for 2002. To get the forecasts for each model I use the equations for my ten volatility models described earlier.

There are several other methods for constructing the realized variance and several of these are discussed in Andersen, Bollerslev & Diebold (2003). A disadvantage using different measures of the realized variance is that it can lead to different results.

## 5.2 Empirical Results

The conditional variance,  $\sigma^2_t$ , is thus the main object of interest and my analysis includes 10 specifications that make up my set of candidate models,  $M_0$ .

The first empirical finding was established when I tested the null hypothesis. I earlier described how one should do to be able to reject or not reject the null. I used the t-test when I performed this method. I use a two-tail test. My critical values are  $-2.02$  and  $+2.02$ . These values are based on 48 degrees of freedom and the significance level  $\alpha$  is set to 0.05.

Table 5.2.1

Test value	Asymmetric Garch (1,1)
Symmetric Garch (1,1)	-4,63997
Symmetric Garch (1,2)	-4,63989
Asymmetric Garch (1,2)	-4,616
Symmetric Garch (2,1)	-4,63998
Asymmetric Garch (2,1)	-3,91873
Random walk	-4,63988
Historical mean	-4,6398
Moving average	-4,63991
Exponential Smoothing	-4,64022

By noticing from table 5.2.1 that the test values for the asymmetric Garch (1,1) lies in the critical region I reject the null hypothesis. This mean that the Asymmetric Garch (1,1) don't belong to the Model Confidence Set. If my MCS would have been just one model, the result becomes stronger. Section 6 accounts for the relation between a significance test and a potential poor model, the Asymmetric Garch (1,2). In table 5.2.2 I have calculated  $v_{i,M}$  which is the measure Hansen, Lunde and Nason (2003) use to rank the models. If one is looking at the number for the asymmetric Garch (1,1) one can see that this variance is higher than for the other models.



It is naturally also of interest to study the models *in* the Model Confidence Set. How will the remaining nine models be ranked according to the information I have? The statistic  $v_{i,m}$  was used as a measure to rank models. The model with the maximum  $v_{i,m}$  was the worst performing model.

Table 5.2.2

<b>Model</b>	<b><math>V_{i,m}</math></b>	<b>Rank</b>
Exponential smoothing	-1,7062	1
Symmetric Garch (1,1)	-1,7054	2
Symmetric Garch (2,1)	-1,705421	3
Moving average	-1,70513	4
Symmetric Garch (1,2)	-1,7051	5
Random walk	-1,705079	6
Historical mean	-1,70485	7
Asymmetric Garch (1,2)	-1,52746	8
Asymmetric Garch (2,1)	-0,285323	9
Asymmetric Garch (1,1)	4,729705	10

That an observation has a higher/large variance means that it is located further away from the regression line, that is, it is more possible that the observation is an outlier. Based on the empirical part it wasn't possible to establish if the Asymmetric Garch (1,1) model really was inferior or if the model was simply 'unlucky' in this sample. An alternative way to determine if a model was inferior was to study the estimated parameters. A presentation and analyse of these will be introduced in section six.

## 6) Discussion

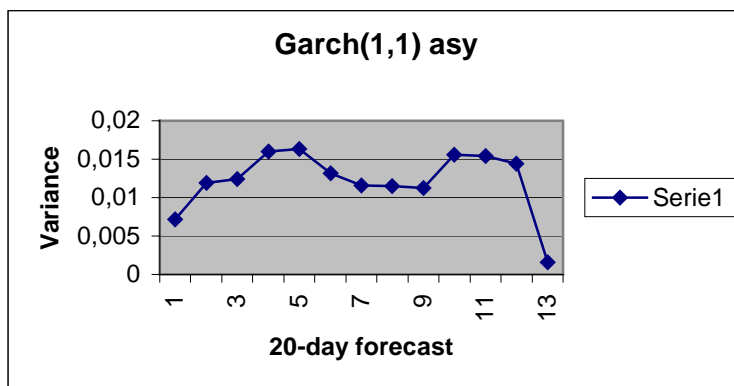
In this paper I have studied the model confidence set. The MCS method was introduced by Hansen, Lunde and Nason (2003). The theory and empirical part in this paper is therefore based on the paper by Hansen et. Al.

I want to, based on the empirical part, determine which forecasting models are “better” from my set of 10 competing models  $M_0$ . That is, I want to determine which models are in  $M^*$ .

I apply the MCS to 10 different volatility models to be able to judge their ability to forecast the volatility of returns from the Stockholm stock exchange index. There is a wide sphere of volatility models. The motive for this large number of models is that the possibility to make empirical findings and economic interpretations increases (Hansen and Lunde). When I had applied the model confidence set to my ten models I came to the conclusion that the Asymmetric Garch(1,1) didn't belong to the MCS. I have estimated parameters for 2000, 2001, 2002, 2003 and 2004. Below is a table containing the estimates for the Asymmetric Garch (1,1) in year 2001.

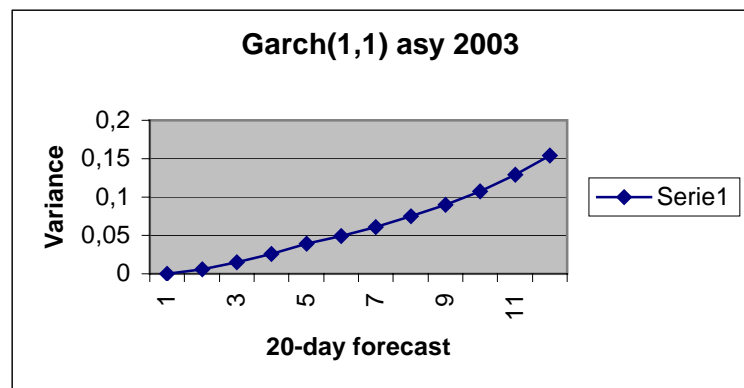
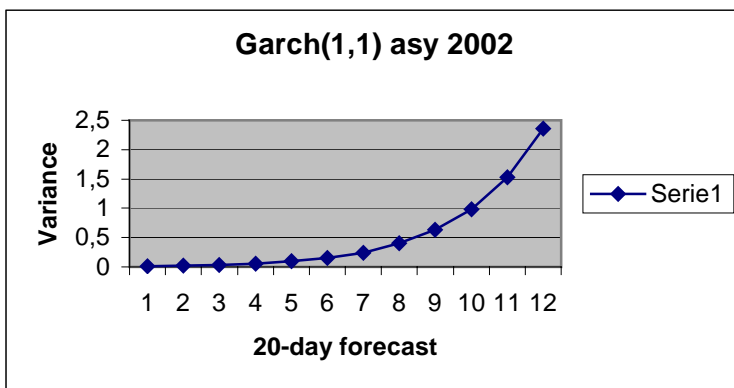
Omega	7.18E-06
Alfa	-0.052614
(RESID<0)*Alfa	0.158819
Beta	0.953199
Unconditional variance	-0,00012

If I look at the estimated parameters in year 2001 and sum these estimates I get a sum that is greater than 1. That the sum of my estimates is larger than 1 is another way to say that this model has a volatility that either strongly decreases or strongly increases.



This conclusion agrees with the graph above. In the beginning the graph has a calmer look, which imply that it doesn't contain so many jumps. But at the end of the year the forecasted conditional is strongly downward sloping.

In the same way, if I consider the estimates for 2002 and 2003 I can see that the sum is above 1, thus the forecasted conditional variance doesn't converge. This is also in line with the increasing slope for the Asymmetric Garch (1,1). Also, the Garch-term, the weight connected to the past conditional variance is greater then 1 in 2002 and 2003 which could explain that the forecasted variances for these two years are upward sloping.



I mentioned earlier that the main goal with the model confidence set method is to determine, from a set of competing models, which model/models are the 'best'. Using the word 'best' is related to the meaning of the word significance (Hansen, Lunde and Nason 2005). That a variable is significant with the significance level 5% is just another way to say that there is a 5 % probability that I will reject the true hypothesis (Gujarati). An interpretation of this is that the model confidence set may contain several poor models in the finite sample,  $M^*$ . Could there be a model that seems to be inferior (significant)? It is easy to see from appendix 1 that the estimates for the Asymmetric Garch (1,2) for 2001 and 2002 sum to more then one. Based on the analyse for the Asymmetric Garch (1,1) the Asymmetric Garch (1,2) shouldn't belong to the Model Confidence set. Could there be any other explanation?

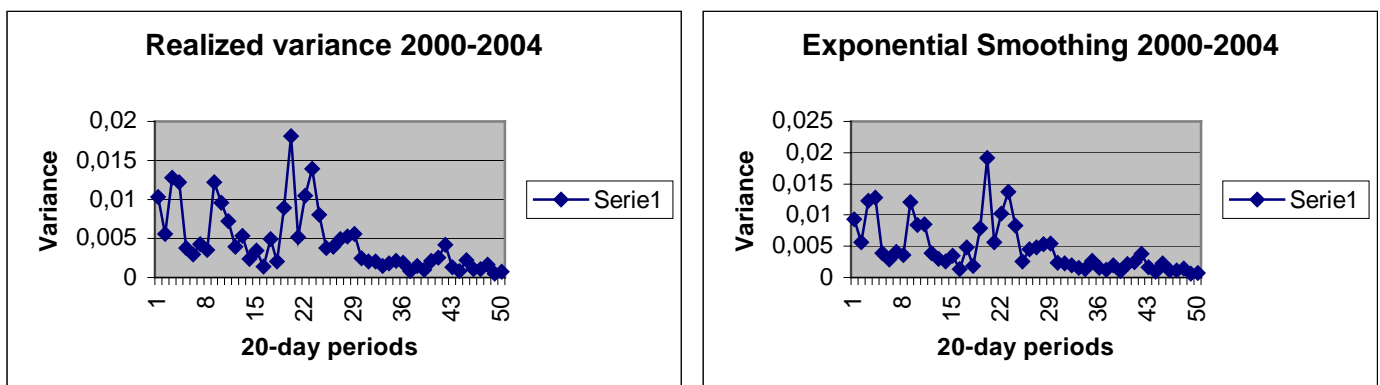
I observe the MSE for the Asymmetric Garch (1,2) below. I ranked models to their expected loss using the mean square error.

MSE	G (1,1) sy	G (1,1) asy	G (1,2) sy	G (1,2) asy	G (2,1) sy	G (2,1) asy	Rand.w	Hist.mean	Movave.	Exp.smoot.
	1,1E-05	0,19051	1,4E-05	0,00233	1,04E-05	0,01723	1,46E-05	1,7E-05	1,4E-05	3,3E-07

That is, I looked at how much the forecast for my model deviates from the realized variance. The MSE for the Asymmetric Garch (1,2) is 0.0023. That is a quite low number. This could be a reason that the model belongs to the MCS, although the model doesn't describe the variation in the realized variance very well. The MSE criterion is more sensitive to outliers. A result is thereby that the model confidence set contains a larger number of models when one is using the MSE as a loss function compared to the MAD loss function.

Although, the MCS will contain the best model(s) asymptotically. That is, the MCS procedure delete a model only if it is found to be significantly inferior to another model.

In this paper I have set the level of confidence,  $1-\alpha$ , to 95%. Hansen et al. define the 'best' model as the model whose forecasts produce the minimum expected loss. The Exponential Smoothing model has the lowest deviation in the loss function. This is a interesting observation. If I compare the appearance for the Exponential Smoothing model and the realized variance I see that they are quite similar.



In appendix 1 I have presented the estimates for the Exponential Smoothing model in 2001 and 2002. I can see that they are quite close to 0, which mean that the model *almost* behaves like the random walk model. The random walk model says that the forecasted volatility is equal to the last observed volatility.

In theory it is often so that the number of models in the MCS will increase as the level of confidence increases. This is reasonable because the region of acceptance become larger (Gujarati). When the MCS contain a single model, this model is very likely to be the

truly superior model. Under the condition of the loss function that was applied when one is constructing the model confidence set. Based on this theory I can say that the MCS contains nine out of the ten models. Thus, the MCS is able to separate superior from inferior volatility models, even if it is just one single model, the asymmetric Garch (1,1), which is considered to be inferior.

There are several advantages with using the MCS method. First, it require less assumptions/information compared to other methods for model selection. One doesn't need to have knowledge of the optimal forecasting model when one is applying the MCS.

In most empirical investigations it's not possible to settle on a single model because of limitations in the data. In appendix 2 I present descriptive statistics for my data (return). The values of the kurtosis for respective year are less then three. This mean that larger movements in the return is less likely. Thus, the closing prices for Stockholm stock exchange index didn't make any larger jumps between 2000 and 2004. This implies that the sample variance should be low for the same period, which is consistent with appendix 2.

The statistics for the skewness in 2002 and 2003 are positive. This means that the probability is larger for an increase in the return (Gujarati 2003). The model confidence set method recognizes this problem. Instead, the MCS usually chooses the models that perform significantly better than the rest.

The MCS method doesn't contain many problems when it has to classify inferior models as truly inferior. Thus, the method is a powerful tool when it comes to choosing the best set of forecasting models. Maybe this is why the model confidence set method is useful for central banks, other parts of government and in financial markets when problems need to be studied (Hansen, Lunde and Nason 2005).

## 7) Conclusion

I have evaluated ten volatility models by using the Model Confidence Set. The issue of my paper is to determine which model/models that belong in the Model Confidence Set. The MCS is analogous to a confidence interval in the sense that the MCS contains a specific model or models. A confidence interval is known to contain a specific parameter (number).

When applying the MCS I came to the conclusion that the Asymmetric Garch(1,1) was inferior and thereby didn't belong in the Model Confidence Set. I observed that the test value, using the t-test, was in the critical region for this model. This is one of the reasons to why the model is considered to be inferior. Also, the  $v_{i,M}$  value showed that the Asymmetric Garch(1,1) is the worst performing model between the ten competing. I also mentioned that it is possible that the MCS can contain one or more poor models. I mentioned that the Asymmetric Garch(1,2) was a possible model for this dilemma. Although, the method is a powerful tool when it comes to choosing the best set of forecasting models. Finally, the ranking between the models in the MCS is as follows

- 1) Exponential Smoothing model
- 2) Symmetric Garch (1,1) model
- 3) Symmetric Garch (2,1) model
- 4) Moving average model
- 5) Symmetric Garch (1,2) model
- 6) Random walk model
- 7) Historical mean model
- 8) Asymmetric Garch (1,2) model
- 9) Asymmetric Garch (2,1) model

The empirical results that were obtained were based on data from the Stockholm stock exchange market. If a different set of data was chosen the empirical findings could be different.

## Literature

Akgiray, V.(1989), 'Conditional Heteroscedasticity in Time Series of Stock Returns : Evidence and Forecasts', *Journal of Business* 62, 55-80

Andersen, T.G & Bollerslev, T.(1998), 'Answering the sceptics: Yes standard volatility models do provide accurate forecasts', *International Economic Review* 39(4), 885-905

Andersen, T.G, Bollerslev, T. & Diebold, F.X(2003), 'Parametric and nonparametric volatility measurement, in Y.Ait-Sahalia & L.P Hansen, eds, 'forthcoming in *Handbook of Financial Econometrics*', Vol.1, Elsevier-North Holland, Amsterdam

Bollerslev, t.(1986), 'Generalized Autoregressive conditional heteroskedasticity', *Journal of Econometrics* 31, 307-327

Brailsford, Timothy J & Faff Robert W. (1993),'Modelling Australian stock market volatility', *Australian Journal of Management* 18, 109-132

Brailsford, Timothy J & Faff, Robert W.(1996), 'An evaluation of volatility forecasting techniques', *Journal of Banking and Finance* 20, 419-438

Brooks, Chris *Journal of Forecasting* 17:1 59-80

Diebold and Mariano (1995), 'Comparing Predictive Accuracy', *Journal of Business and Economic statistics* 13, 253-263

Dimson, E. & Marsh P. (1990),'Volatility forecasting without data-snooping', *Journal of Banking and Finance* 14, 399-421

Engle R.F (1982) , Autoregressive conditional heteroskedasticity with estimates of the variance of U.K inflation, *Econometrica* 50, 987-1008.

Engle, Robert F and Victor K. Ng., 'Measuring and testing the impact of news on volatility' *Journal of Finance* 48, 1749-1778

Glosten, L.R , Jagannathan and D. E. Runkle (1993) ' On the relation between the expected value and the volatility of the Nominal excess Return on stocks', *Journal of Finance* 48 1779-1801

Gujarati Damodar N., (2003) *Basic Econometrics fourth edition*

Hansen P.R (2001) 'A test for superior predictive ability'.

[www.econ.brown.edu/fac/Peter\\_Hansen](http://www.econ.brown.edu/fac/Peter_Hansen)

Hansen, P.R & Lunde, A (2001), 'A forecast comparison of volatility models : Does anything beat a GARCH(1,1)?'. [www.econ.brown.edu/fac/Peter\\_Hansen](http://www.econ.brown.edu/fac/Peter_Hansen)

Hansen, P.R & Lunde A. (2003), 'Consistent preordering with an estimated criterion function, with an application to the evaluation and comparison of volatility models'.

[http://www.econ.brown.edu/fac/Peter\\_Hansen](http://www.econ.brown.edu/fac/Peter_Hansen)

Hansen, P.R & Lunde, A. & Nason, James M. (2003), 'Choosing the Best Volatility Models:

The Model Confidence Set Approach', [www.econ.brown.edu/fac/Peter\\_Hansen](http://www.econ.brown.edu/fac/Peter_Hansen)

Hansen, P.R , Lunde A & Nason James M. (2005), 'Model Confidence Sets for Forecasting Models', [www.econ.brown.edu/fac/Peter\\_Hansen](http://www.econ.brown.edu/fac/Peter_Hansen)

Hentschel, L. (1995), 'All in the family Nesting symmetric and asymmetric GARCH models', *Journal of Financial Economics* 39, 71-104

Hill R. Carter, Griffiths William E., Judge George G. (1997) *Undergraduate Econometrics second edition*

Pagan Adrian R. & Schwert William G.(1990), ' Alternative models for conditional stock volatility', *Journal of Econometrics* 45, 267-290.



Pantula S.G (1989), 'Testing for unit roots in time series data', *Econometric Theory* 5, 256-71

Poon and Granger 'Forecasting volatility in financial markets' *Journal of Economic Literature* 41, 478-539

Poterba, James M and Summers (1986), *American Economic Review* 76, 1142-1151

Schwert, G William, 'Why does stock volatility change over time?' *Journal of Finance* 44,1115-1153

## Appendix 1

	2001	2002
<i>Symmetric Garch (1,1)</i>		
Omega	3.36E-05	1.28E-05
Alfa	0.062178	0.135356
Beta	0.831317	0.828525
Unconditional variance	0,000315	0,000354
<i>Symmetric Garch (1,2)</i>		
Omega	5.81E-06	1.34E-05
Alfa (1)	0.012989	0.119716
Beta (1)	<b>1.782808</b>	<b>1.274924</b>
Beta (2)	-0.813633	-0.430982
Unconditional variance	0,000326	0,000369
<i>Asymmetric Garch (1,1)</i>		
Omega	7.18E-06	9.76E-07
Alfa(1)	-0.052614	-0.058167
(RESID<0)*Alfa(1)	0.158819	0.092384
Beta (1)	0.953199	<b>1.021626</b>
Unconditional variance	-0,00012	-1,7E-05
<i>Asymmetric Garch (1,2)</i>		
Omega	7.35E-06	1,99E-07
Alfa(1)	-0.054981	-0,05159
(RESID<0)*Alfa (1)	0.160993	0,09376
Beta (1)	0.948481	0,997744
Beta (2)	0.005233	0,005961
Unconditional variance	-0,00012	-4,3E-06
<i>Asymmetric Garch (2,1)</i>		
Omega	9,53E-06	7.14E-06
Alfa (1)	-0,09604	-0.085355
Alfa (2)	0,041819	0.109538
(RESID<0)*Alfa (1)	0,154816	0.136002
Beta (1)	0,941431	0.890768
Unconditional variance	-0,00023	-0,00014
<i>Exponential Smoothing</i>	0,001	0,001

Above is a table consisting of the estimated parameters for some chosen symmetric and asymmetric Garch models and the Exponential Smoothing model. Year 2001 and year 2002 was chosen. Omega, alfa and beta were my estimated parameters. The unconditional variance, denoted  $\sigma^2$ , was calculated with the help of the three parameters.

## Appendix 2

### Return year 2000

Mean	-0,0006
Standard Error	0,001136
Median	0,000458
Standard Deviation	0,018005
Sample Variance	0,000324
Kurtosis	-0,20792
Skewness	-0,0761

### Return year 2001

Mean	-0,00072
Standard Error	0,001204
Median	-0,00219
Standard Deviation	0,019044
Sample Variance	0,000363
Kurtosis	1,112719
Skewness	-0,00758

### Return year 2002

Mean	-0,00166
Standard Error	0,001176
Median	-0,00268
Sample Variance	0,000346
Kurtosis	1,42339
Skewness	0,481454

### Return year 2003

Mean	0,000941
Standard Error	0,000757
Median	0,00113
Sample Variance	0,000143
Kurtosis	0,615521
Skewness	0,053967

### Return year 2004

Mean	0,000598
Standard Error	0,000559
Median	0,001212
Sample Variance	7,88E-05
Kurtosis	1,377604
Skewness	-0,58536