Filtering with Wavelets may be worse than you think

Eugenio Fernández
Universidad Rey Juan Carlos

Ignacio Olmeda
Universidad de Alcalá

June 1, 2000

Abstract

One of most promising applications of wavelets is in the field of nonparametric statistical estimation, in which one wants to estimate an unknown signal from some noisy data. Donoho and Johnstone (1994, 1995) have developed a simple and yet powerful methodology for nonparametric regression and smoothing based on the principle of wavelet shrinkage (removing noise by shrinking wavelets coefficients towards zero) referred as the Waveshrink algorithm. In order to select the best values for the parameters of the waveshrink algorithm several approaches have been proposed. Each of the methods suggested have their pros and cons depending on the particular domain of application. Nevertheless a basic question is which of the methods dominates under a forecasting criterion. To put in simple words: which method is best?. The purpose of this paper is twofold, first we analyze the potential advantages of wavelet shrinkage methods for financial time series prediction, then we identify the best combination of srinkage parameters. Our results show that, in general, filtering may result harmful and that the choice the parameters is more critical for some parameters than for others.

1 Introduction

Wavelets theory has been developed with ideas taken from different fields such as applied mathematics, signal processing or physics. They have been applied in many statistical areas (Donoho and Johnstone, 1994, 1995; Fan et al. 1993; Johnstone et al., 1992) such as regression, pattern analysis, density estimation or forecasting of time series. In the dynamic case, wavelets have been demonstrated to be a very powerful tool when dealing with phenomena that evolve rapidly in time.

One of great successful stories of wavelets is in the field of nonparametric statistical estimation, in which one wants to estimate an unknown signal $f(x_t)$ from some noisy data y_t. Donoho and Johnstone (1994, 1995) have developed a simple and yet powerful methodology for nonparametric regression and smoothing based on the principle of wavelet shrinkage (removing noising by shrinking wavelets coefficients towards zero) referred as the Waveshrink algorithm. Shrinkage essentially rests on three simple principles: signal features can be represented by just a few wavelet coefficients, noise affects all wavelets coefficients, and by shrinking wavelet coefficients towards zero, the noise can be removed while preserving features. The algorithm can be described as a three step procedure: 1) Data are transformed into a set of wavelets coefficients applying the discrete wavelet function; 2) a shrinkage of the coefficients is performed; 3) the shrunken wavelet coefficients are transformed back in the domain of the original data.

1.1 Wavelet function estimation

Suppose the data are given by $y_t = f(x_t) + e_t$, $t = 1, 2, ..., n$ where $f(t)$ is a $L^2(\Re)$ discrete signal, and e_t are independent and identically distributed normal errors. The wavelet estimators are a special class of orthogonal series estimators of the form:

$$f(x) = \sum_{j \in Z} \sum_{k \in Z} d_{jk} \psi_{jk}(x)$$

where $d_{jk} = \int_{\Re} f(x) \psi_{jk}(x) dx$ are the wavelets coefficients of f. The wavelet coefficients are localized in time and frequency, this time-frequency localization is the main reason why wavelets are useful for function approximation. The basis function wavelets are usually of the form:

$$\psi_{jk}(x) = 2^{\frac{j}{2}} \psi(2^j x - k)$$

where $\psi(x)$ is a particular kind of function so that $\psi_{jk}(x)$ forms an orthogonal basis, $\psi(x)$ and all its derivatives up to order m exist and decrease rapidly and $\psi(x)$ is orthogonal to all polynomials of degree $m-1$.

1.2 Wavelet shrinkage

The discrete wavelet transform can be represented by an orthogonal matrix W:

$$w = Wy$$

where w are the wavelet coefficients. The idea of wavelet shrinkage is to modify the coefficients by some procedure obtaining a new set of coefficients $\widetilde{w}$ and the perform an inverse transformation to obtain:

$$\widetilde{f} = W^T \widetilde{w}$$

where $\widetilde{f}$ is an estimate of f at x.
The Waveshrink algorithm can be described as:

1. Apply the wavelet transform with J levels to the signal y, obtaining wavelets detail and smooth coefficients $w = (d_1, ..., d_J, s_J)$
2. Shrink the detail coefficients at the j finest scales to obtain new detail coefficients $\widetilde{d_1} = \delta_c(d_1), \widetilde{d_2} = \delta_c(d_2), ..., \widetilde{d_j} = \delta_c(d_j)$, applying the function $\delta(d)$ to shrink d towards zero.
3. Apply the inverse discrete wavelet transform using the detail coefficients $\widetilde{w} = (\widetilde{d_1}, \widetilde{d_2}, ..., \widetilde{d_j}, d_{j+1,...,}d_J, s_J)$ to obtain the waveshrink estimate $\widetilde{f}$.

Generally, the shrinkage function δ may take two basic forms, "soft shrinkage" :

$$\delta_c(d) = \begin{Bmatrix} 0 & if \quad |d| \leq c \\ sign(d)(|d| - c) & if \quad |d| > c \end{Bmatrix}$$

and "hard shrinkage":

$$\delta_c(d) = \begin{Bmatrix} 0 & if \quad |d| \leq c \\ d & if \quad |d| > c \end{Bmatrix}$$

where $c = \lambda * \sigma$ (other shrinkage forms have been also proposed such as *semisoft shrinkage*, Gao and Bruce, 1996;and *non-negative garrote*, Gao, 1997). In order to select the best values for the parameters of the waveshrink algorithm, the thresholding level, λ, and an estimate of the noise variance, σ, several approaches have been proposed. To select the threshold λ, four different rules have been proposed: "rigsure" (selection using the principle of Stein's Unbiased Risk Estimate), "sqtwolog" (fixed form threshold equal to $\sqrt{2 * log(n)}$, "heursure" (selection using a mixture of the first two options), and finally, "minimaxi" (selection using minimax principles). With respect to the scale of the noise, there are three basic rules: *"one"* is the basic model with $\sigma = 1$, *"sln"* use the finest scale detail coefficients to estimate a single factor for all levels and *"mln"* employs a separate scale factor for each level.

Each of the methods proposed have, obviously, their pros and cons depending on the particular domain of application. Nevertheless a basic question is which (if any) of the methods dominates under a forecasting criterion. To put in simple words: which method is best?.

2 Procedure employed and Database

As we have previously mentioned, we are interested in wavelets as a technique for noise reduction. Our approch is to apply wavelet shrinkage methods to obtain a "clean" signal which is then used to estimate several linear models. If too few noise is removed from the data, the linear models estimated over the clean signal should perform approximately as well as the linear model on the corresponding noisy time series. Also, if the function is "oversmoothed", the patterns present in the data are removed and forecasts should be also poor. This suggest that there may be a balance between too much filtering and too few filtering so that forecasts may be improved.

The approach adopted is quite simple but extremally computationally intensive: we consider all the posible combinations of wavelet parameters and apply each of the combinations to filter the noise of a set of financial time series. Then, we construct a variety of linear models for each of the filtered series and perform a dynamic forecasting exercise. Finally we conduct an analysis on the relative superiority of each of the filtering methods by comparing the results with the ones obtained with the non-filtered series.

The choice of the type of Wavelet is based on previous work. Daubechies' wavelet is probably the most widely used applied studies while the *Coiflet* wavelet has been also applied in a financial context (e.g. Thomason, 1996; Abecasis, 1997). Finally, we have employed the *Symlet* wavelet to complete our study, since it has been employed less frequently in the literature.

Since we consider three types of wavelets (*Coiflet 4, Daubechies 4, Symlet 4*), two levels of decomposition ($J = 4, 6$), two levels of shrinkage ($j = 2, 4$ for $J = 4$ and $j = 4, 6$ for $J = 6$), three scales of noise (*one, sln, mln),* two shrinkage functions *(soft* and *hard)* and four threshold rules *(rigsure, sqtwolog, heursure, minimaxi*), we have $3x2x2x3x2x4 = 288$ diferent filterings. In what follows, each one of the filters will be represented as an 6-upla *(threshold, wavelet, noise, function, level, resolution),* so that, for example, *(Rigsure, Coiflet, sln, hard, 0, 6)* means that we have employed a filter using *rigsure* as the threshold rule, the *Coiflet* wavelet, *sln* as the scale of noise, *hard* as the srinkage function, *j=4* and *J=6.*

The database is composed of daily closing prices (Reuters) of 14 indexes of the main european markets as well as the Nikkei and S&P500, the period of study spans from 10/11/1995 to 10/15/1997. The indexes employed are CAC40 (France), IBEX35 (Spain), DAX30 (Germany), AEX (Netherlands), FTSE100 (United Kingdom), MIB30 (Italy), BEL20 (Belgium), ATX (Austria), OMX (Sweden), PSI (Portugal), KFX (Denmark), FOX (Finland), OBX (Norway), SMI (Switzerland), NIKKEI (Japan) and S&P500 (United States).

Firstly, and to justify the use of the wavelet transform, we verified non-stationarity in the log prices by means of the augmented Dickey-Fuller test (Dickey and Fuller, 1979), which did not allow to reject the null of the existence of a unit root. All series showed leptokurtic and left-skewed, so that, the Jarque-Bera statistic permited to reject null of normality of the unconditional distribution of the returns. Also, we verified the evidence of some linear structure in the returns with the Ljung-Box test (we rejected no autocorrelation of order ten at the 5% level in all series) and comprobated that the correlation is removed with a simple autoregressive model of order five. Similarly to Swanson and White (1997) our study employs log prices instead of returns (a main diference with most of the empirical studies). The employment of log prices allows us to simplify assumptions about the "true" data generating process, specifically, the existence or not of cointegrating relations.

2.1 Univariate and bivariate specifications

In our first set of experiments we will compare the predictions obtained from linear models when they are estimated over filtered as well as nonfiltered series, we proceeded as follows. First we set an observational window of size 256 (approximately one year of daily observations), begining in the first observation of the series, and filter this subseries through each one of the 288 combination of wavelets parameters to obtain a set of filtered series. Then, we fit, for each ones of the filtered series, univariate autoregressive models of the form:

$$\log P_t^1 = \alpha_0 + \sum_{i=1}^{3} I^1 \alpha_i^1 \log P_{t-i}^1 + \varepsilon_t \tag{1}$$

where P_t^j is the index j ($j = 1, 2, ..., 16$) at time t, and use them to forecast next day log price. Finally, the diference between the forecast and the real observation is computed. We roll the window one observation ahead and continue the process until the end of the sample is reached. We also proceed similarly with the nonfiltered series, computing the error, and use this as the benchmark to compare each of the filterings.

Throughout our paper, the loss function employed to test for accuracy is the mean squared error (mse):

$$mse = \frac{1}{n} \sum (p_i - r_i)^2 \tag{2}$$

where p_i is the forecast, r_i the real data and n the number of forecats (which, in our case, is equal to 266). To normalize, the mse of each of the models and filters is divided by the mse of a random walk (no drift) and expressed in percentage terms, so that a ratio less than 100 means that the particular model is better than the random walk.

The procedure adopted with the bivariate specification is identical, except that in this case our models also include past information of the S&P500 index (we also use 3 lags). The models are then of the form:

$$\log P_t^k = \alpha_0 + \sum_{i=1}^{3} \alpha_i^k \log P_{t-it}^k + \sum_{j=1}^{3} \beta_j^{SP500} \log P_{t-j}^{SP500} + \varepsilon_t \tag{3}$$

2.2 Multivariate specifications

We have also analyze the possible improvements of filtering through wavelets in a multivariate context. To alleviate the computational burden we will focus just in five of the indexes, DAX30, CAC40, FTSE100, PSI20 and AEX. These indexes are chosen because they are sufficiently representative of the European stock markets and they represent diferent degrees of predictability (being the PSI20 index the most predictable, in terms of magnitude, during the period of study and the FTSE100 the least predictable). For each of the indexes, we employ five stockmarket indexes as regressors, the indexes are selected to maximize the correlation with the "target" index. Te regressors employed in each one of the cases are, DAX30, AEX, ATX, BEL20 and KFX, for the DAX; CAC40, IBEX35, AEX, FTSE100 and OMX, for the CAC40; FTSE100, AEX, CAC40, OMX and BEL20 for the FTSE100; PSI20, DAX30, AEX, OMX and CAC40, for the PSI20 and AEX, OMX, DAX30, FTSE100 and BEL20 for the AEX.

The multivariate models employed are parameterized as:

$$\log P_t^1 = \alpha_0 + \sum_{i=1}^{5} I^1 \alpha_i^1 \log P_{t-i}^1 + ... + \sum_{i=1}^{5} I^k \alpha_i^k \log P_{t-i}^k + \varepsilon_t \tag{4}$$

where P_t^k is the index k $(k = 1, 2, ..., 5)$ at time t, and $I^k, k = 1, ..., 5$ is equal to one in the index k is included in the model and zero otherwise (so that when $I^k = 0, \forall k \neq 1$ the model is just an $AR(5)$ in the logs). Since we have 5 possible regressors, the maximum number of models is 32 $(= 2^5)$. The simplest model includes the past values of the index as the information set while the most complicate employs all the corresponding five indexes. The models are estimated, againg, using a rolling window of a fixed size of 256 filtered observations and, in this case, and to check for the robustness of our results, always using five lags (a trading week, approximately).

As we have 522 daily observations (which imply 266 forecasts) and 32 diferent models, the number of independent forecasts for each filtering is $8544 (= 267x32)$. Also, since there are 288 diferent filters and five series, the whole number of forecasts is 12.303.135, which gives an idea of the computational burden of the problem.

3 Results

3.1 Univariate and bivariate forecasts

To present our results, and due to the magnitude of data, we have ellaborated figures 1 and 2 which provide a visual intuition. Since the errors of some of the forecasts obtained with many of the filterings were orders of magnitude higher than the errors obtained with the nonfiltered series, we will just focus on the best 50 filterings. In all the figures we represent in the the y-axis the porcentual *mse* ratio of the corresponding filter against the one obtained by a random walk, so that a number higher than 100 denotes a worse behavior than the latter. In the x-axis we represent the sorted order of each ones of the the best 50 filterings, so that, for example, *x=1* may correspond to a filter of the form *(Rigsure, Coiflet, sln, hard, 0, 6).* We have also plotted the ratio against the random walk of the model estimated on the nonfiltered series, which we represent by an horizontal line. To the extent that the filters are significantly below this line, wavelet filtering will be beneficious, independently if models are able or not to beat the random walk, to the extent that the filters are significantly below this line and also significantly below 100, models could be useful in predicting future stockmarket prices.

As we can see, the univariate models are generally useless for predicting stockmarket prices in virtually all the cases. For the nonfiltered series, with the exception of the KFX, OBX , PSI20 and BEL20 we find ratios slightly over 100. Of the mentioned four cases, only the PSI20 shows a significative improvement while the others are clearly marginal. The second thing to mention is that the filters errors are generally smaller than the nonfilterd ones (the exceptions are the DAX30, AEX, S&P500 and KFX). Maybe unfortunately, we find that the improvements are insignificant, since the most important is for the SMI and it is around 2%.

For the bivariate models the results change in some sense. First, in all the cases the improvements against the random walk for the nonfiltered series is obvious in all the cases. The forecasts are in some cases a 30% lower than the ones obtained with the random walk (DAX30) and in the worst case (NIKKEI) they are around 5% better. Focussing on the results with the filterings, we find though that the conclusions are similar to the ones obtained with the univariate models. In this case, the filters are generally worse for CAC40, DAX30, EAEX, OMX and KFX and the improvements are absolutely marginal (around 1%).

As we see, there are no significative differences when we compare the filters in “good” (the models beat the random walk) or “bad” (the model does not beat the random walk) situations. Also, note that we are exclussively focussing on the performance of the *best* filterings that are selected *ex-post*. This means that, with no previous knowledge, the experimenter would obtain much worse results by applying a naive filter than without using it.

3.2 Multivariate forecasts

Now we will analyze the effect of filtering in a multivariate context. Again, and to give an intuitive vision of the results, we present all of them as points in a cartesian plane. Since, in this case, we have 32 diferent models for each ones of the filters, our representation will be slightly diferent to the preceeding ones. In this case we plot, for each one of the 32 multivariate models, the results obtained with each ones of the 288 filterings, so that for each one of the models we obtain a family of points that represent the correponding *mse* ratios. Again we will focus only on some of the models (the best 25) since the worst ones produced errors orders of magnitude higher and their representation would rest clarity to our presentation. In the fiures 4 to 6, the *x-axis* denote each one of the models (sortered in increasing *mse* order, as before) and the *y-axis* denote the respective ratio to the random walk. For example, the point (23,0.96) means that model 23 improves the random walk by a 4% while the point (17,1.07) means that model 17 is a 7% worse than the random walk. We plot the errors obtained by the models with the nonfiltered series in red and the errors obtained using the filtered series in yellow. To give a clearer view, the filterings are also sorted in increasing *mse* ratio for each ones of the models so that, for example, model 1 for filtering number 166 may not be the same as model 1 for filtering number 38. Finally, we also plot the ratios obtained with the nonfiltered series.

In figure 4 we show the results obtained with all the filterings. Again, a ratio smaller than one should be interpreted as a better performance than a random walk. As it can be seen, most of the filterings make the estimated models exhibit worse performance than the ones built on the nonfiltered series, since the yellow lines lie above the red line line. Note that filtering may serverely damage the forecasts since, especially for the worst models, the ratio is significantly higher than one.

To see more clearly the results, we propose two additional figures. In the first one (figure 5) we present only the filterings for which the best model is at least as good as the random walk. It can be seen that there are much less filters which have this property. Note that for the first series (CAC40), some of the filters permit to improve the forecasts than the ones obtained with the nonfiltered series but in all the other four cases, most of the filters are worse.

In the next figure (figure 6) we plot only the filters for which the best model improves the best nonfiltered model. As it can be seen, again for the CAC40, many filters have this property, while for the rest of the models they are very few (e.g. four for the DAX30) and the improvements are negligible. In the case of CAC40, some of the improvements seem to be significative, since they are around a 6%.

As a conclusion, we have confirmed the results obtained with univariate and bivariate models. The filters generally provided worse forecasts than the ones obtained using "raw" series.

Now we will turn our attention to the effects of each one of the filtering parameters in the quality of the forecasts obtained. This aspect is important since, as we have seen, there is a considerable variation in the quality of the predictions depending on the selected filter, so that it is relevant to investigate if some configurations dominate.

To see the effects of each one of the six parameters, we will compute the number of models that are better than the random walk keeping a particular parameter fixed. For example, we compute the number of models that are better than the random walk *and* employ the *Rigsure* thresholding, then we repeat for each on the thresholding methods. If, say, the number of models that employ *Rigsure* is considerably higher than the number of models that use other thresholdings we can conclude *ceteris paribus* that *Rigsure* compares favourably to the other methods. Tables 1 to 6 give the details for each one of the parameters, we show the percentage of the models that beat the random walk and include a particular factor (the number of models are 122, 166, 107, 180 and 131 for CAC, DAX, FTSE, PSI and AEX, respectively).

With respect to thresholding, there are no significative diferences among *Rigsure*, *Heursure* and *Minimax*, while *Sqtwolog* seems clearly inferior. The wavelet type does not seem to be of clear importance but the *Coiflet* dominates slightly. Respecting the estimation of noise, *one* is clearly superior, while *mln* is the worst. *Hard* shrinkage clearly dominates over *soft* and, finally, changes in the resolution and level do not seem to be very important. Note that these results are consistent along the series employed, which showed diferent degrees of predictability.

Until now, we have focused on analyzing qualitatively the questions of wheter filtering or not and the choice of the optimal filter. Now we will employ a statistical test to verify the statistical significance of the improvements, for which we will employ the nonparametric test of Diebold and Mariano (1995). Let us suppose we have two predictions p_i^1, p_i^2 (say that p_i^1 is the i prediction of a model built on the nonfiltered series while p_i^2 is the i prediction of the same model built on the filtered series) and let $g(p_i^{1,2}, r_i)$ be a loss function, (for example the *mean squerd error*: $g(p_i^{1,2}, r_i) = \frac{1}{n}\sum_{i=1}^{n}(p_i^{1,2} - r_i)^2$). The diference of the loss functions for prediction i will be: $d_i = g(p_i^1, r_i) - g(p_i^2, r_i)$. Then, it can be shown that, under the null of equal predictive power:

$$DM = \frac{\overline{d}}{\sqrt{\frac{2\pi \widehat{f}_d(0)}{n}}} \sim N(0,1) \tag{5}$$

where

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{6}$$

and $\widehat{f}_d(0)$ is a consistent estimator of the spectral density of the difference of predictive errors at frequency zero.

Our idea consists on comparing the predictive accuracy of two models: the first one, which employs non-filtered data and a second one estimated over the filtered time series. If DM is bigger than 1,96, we can reject the null hypothesis (at a 5% level) of equal predictive power in favor of the alternative that model 2 provides better predictions (that is, filtering is beneficious). On the contrary, when the DM value is smaller than -1,96, we can conclude that filtering damages the forecasts.

We will use, again, a graphical presentation of the results. We restrict our attention to CAC40 and DAX30, which represent examples of the relative "goodness" and "badness" of filtering. In Figures 7 (for CAC40) and 8 (for DAX30) we present the value of the DM statistic for each of the filters which provide at least one model better than the nonfiltered series. If a filter is always useful, it will exhibit DM statistics (represented in the y-axis) consistently above 1,96, while if is is perjudicial it will provide DM statistics below -1,96. In other case, we can not say that filtering is bebeficious nor perjudicial in terms of forecasting accuracy. As it can be seen for the CAC40 only in four cases we find a consistent statistical improvement by filtering while in the rest of the cases (excepting the model ranked in the twelfth position) the filtered series permit to obtain equivalen forecasts (the values of the DM statistic are in the interval [-2,2]). In the case of the DAX30, the filters appear to produce more pronounced results, while very few filters permit to obtain better forecasts the vast majority of them are irrelevant or perjudicial.

4 Conclusions

In this paper we have explored the sensitivity of the forecasts obtained by employing a denoising procedure based on wavelet shrinkage. We have shown that, in the particular application chosen, filtering with wavelets does not permit to obtain more accurate forecasts and, in fact, it can damage them. Of course, our results are just a first approximation to the problem but we think they are illustrative on the potential damage of a naive use of wavelets in financial time series prediction.

Another finding is that the shrinkage do not seem to be very sensitive to certain parameters while, for others, the differences are clear: we find that the wavelet type is not crucial but the *one* rule of noise estimation dominates, as it does *hard* shrinkage over *soft* shrinkage. Again, more work to evaluate the potential superiority of some filterings is needed.

References

[1] Abecasis, A. M.and E. S. Lapenta (1997): "Modelling the Merval Index with Neural Networks and the Discrete Wavelet Transform". *Journal of Computational Intelligence in Finance*, 5, 15-19.

[2] Daubechies, I. (1988): *Ten Lectures on Wavelets*. Philadelphia: SIAM.

[3] Dickey, D.A. and W.A. Fuller (1979): "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, 74, 427-431.

[4] Donoho, D.L., and I.M. Johnstone (1994): "Ideal spatial adaptation by wavelet shrinkage". *Biometrika*, 81, 425–455.

[5] Donoho, D.L., and I.M. Johnstone (1995): "Adapting to unknown smoothness via wavelet shrinkage", *Journal of the American Statististical Association*, 90, 1200–1224.

[6] Fan, J., P., M. Martin and P. Patil (1993): "Adaption to high spatial inhomogeneity based on wavelets and on local linear smoothing", Tech. Rept. CMA-SR18-93. Australian National University.

[7] Johnstone, I.M., G. Kerkyacharian and D. Picard (1992): "Estimation d'une densité de probabilité par méthode d'ondelettes", *Comptes Rendus Acad. Sciences Paris,* 315, 211-216.

[8] Smith, M., and T. Barnwell (1986): "Exact reconstruction techniques for tree-structurated subband coders", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34, 434-441.

[9] Swanson, N. R. and H. White (1997):" Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models", *International Journal of Forecasting,* 13, 439-461.

[10] Thomason, M. R. (1997): "Financial Forecasting with Wavelets Filters and Neural Networks". *Journal of Computational Intelligence in Finance*, 5, . 27-32.

[11] Gao, H.-Y. (1997): "Wavelet shrinkage denoising using non-negative garrote", Statistical Sciences Division, MathSoft Inc.

Table 1: Effect of thresholding

	CAC40	*DAX30*	*FTSE100*	*PSI20*	*AEX*
Rigsure	30%	30%	35%	29%	32%
Sqtwolog	10%	14%	15%	18%	9%
Heursure	32%	30%	35%	29%	32%
Minimax	29%	27%	16%	23%	27%

Table 2: Effect of Wavelet type

	CAC40	*DAX30*	*FTSE100*	*PSI20*	*AEX*
Coiflet	41%	34%	38%	38%	37%
Daubechies	25%	34%	30%	36%	33%
Symlet	34%	32%	32%	26%	30%

Table 3: Effect of noise

	CAC40	*DAX30*	*FTSE100*	*PSI20*	*AEX*
mln	10%	16%	9%	13%	11%
one	69%	58%	76%	53%	61%
sln	21%	27%	15%	33%	27%

Table 4: Effect of function

	CAC40	*DAX30*	*FTSE100*	*PSI20*	*AEX*
Soft	31%	35%	31%	37%	24%
Hard	69%	65%	69%	63%	76%

Table 5: Effect of level

	CAC40	*DAX30*	*FTSE100*	*PSI20*	*AEX*
0	49%	48%	50%	48%	47%
1	51%	42%	50%	52%	53%

Table 6: Effect of resolution

	CAC40	*DAX30*	*FTSE100*	*PSI20*	*AEX*
4	52%	52%	51%	52%	53%
6	48%	48%	49%	48%	47%

Figure 1: MSE ratio against a random walk, univariate models

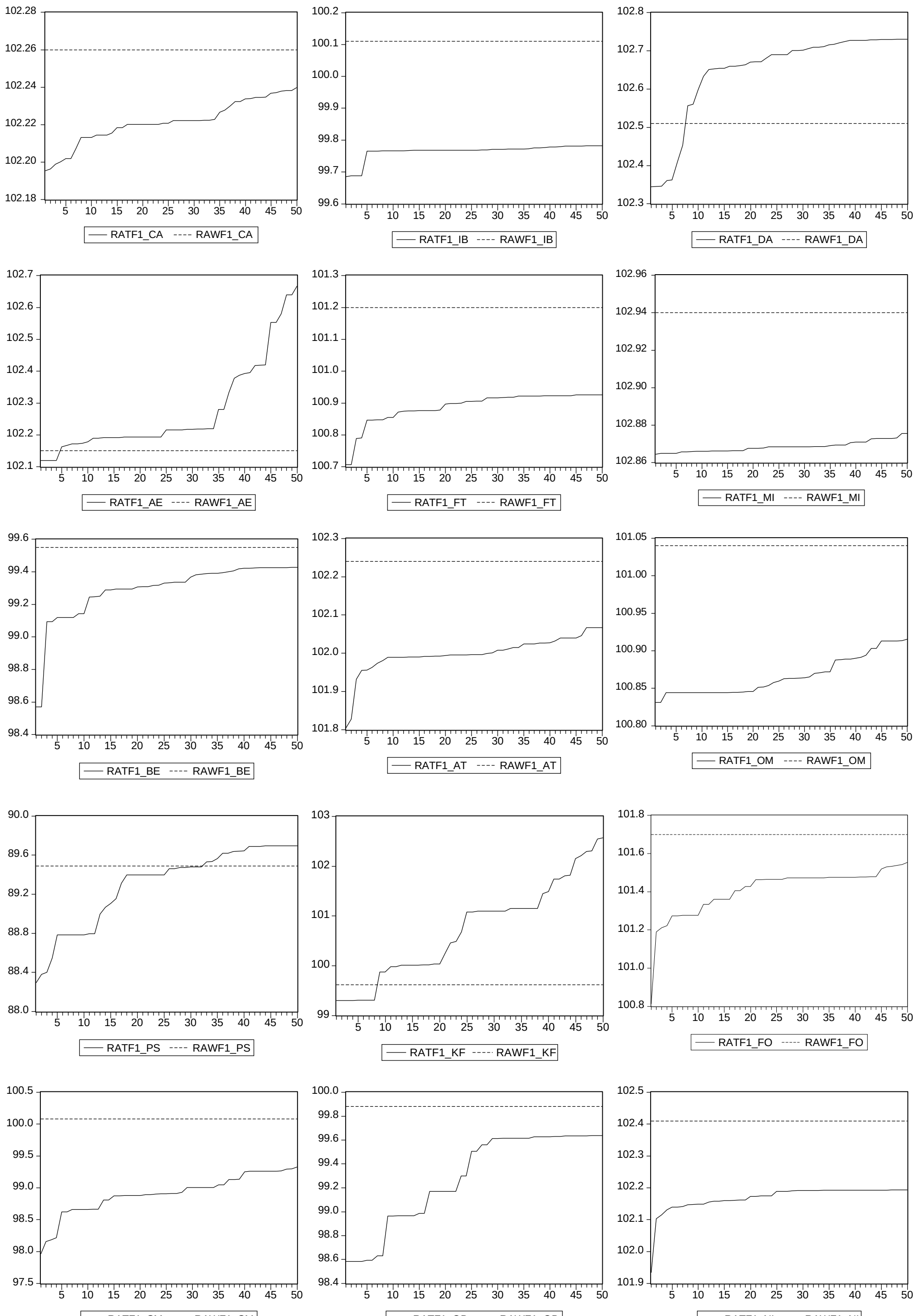


Figure 2: MSE ratio against random walk, univariate models (cont.)

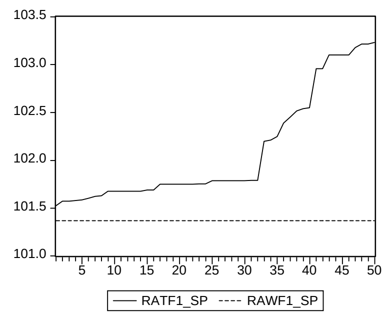


Figure 3: MSE against a random walk, bivariate models (S&P500 included in the model)

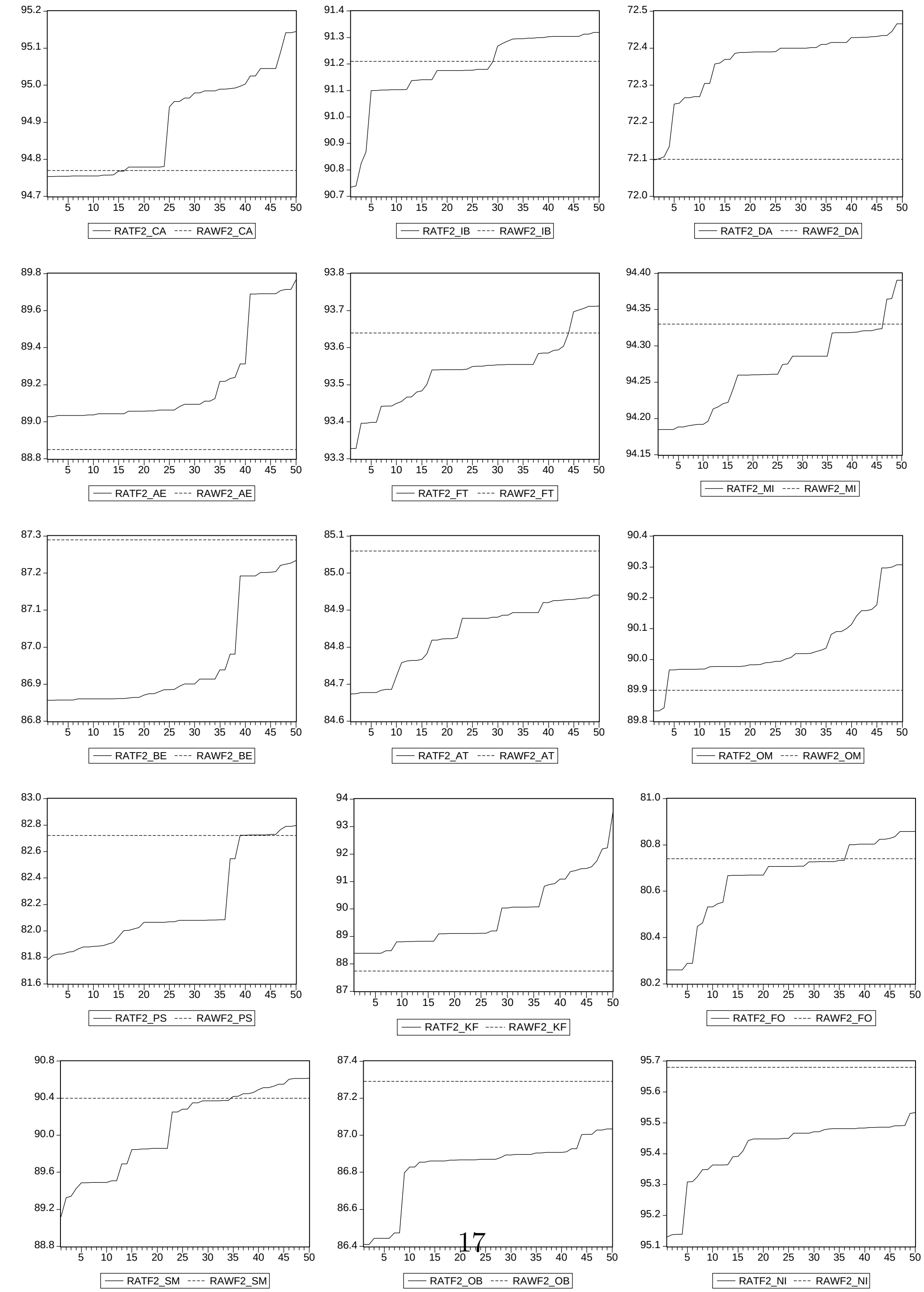


Figure 4: All filterings: CAC, FTSE, DAX, PSI, AEX

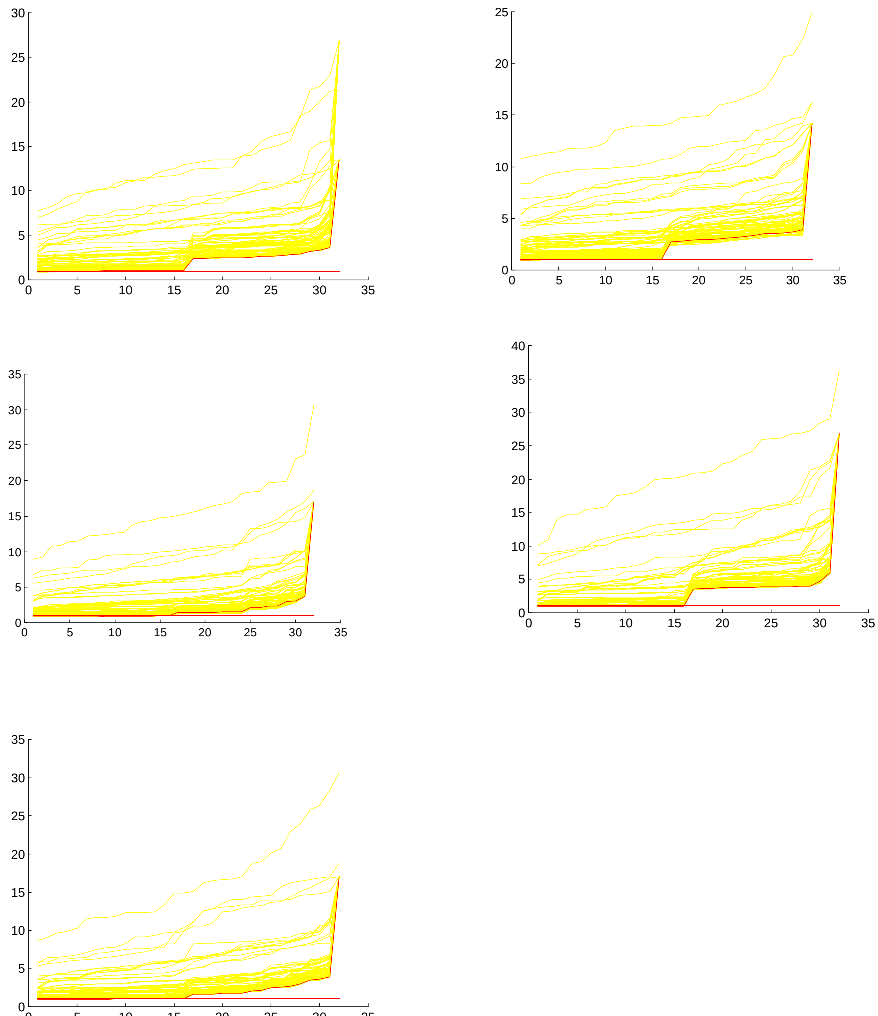


Figure 5: Filterings better than the random walk: CAC, FTSE, DAX, PSI, AEX

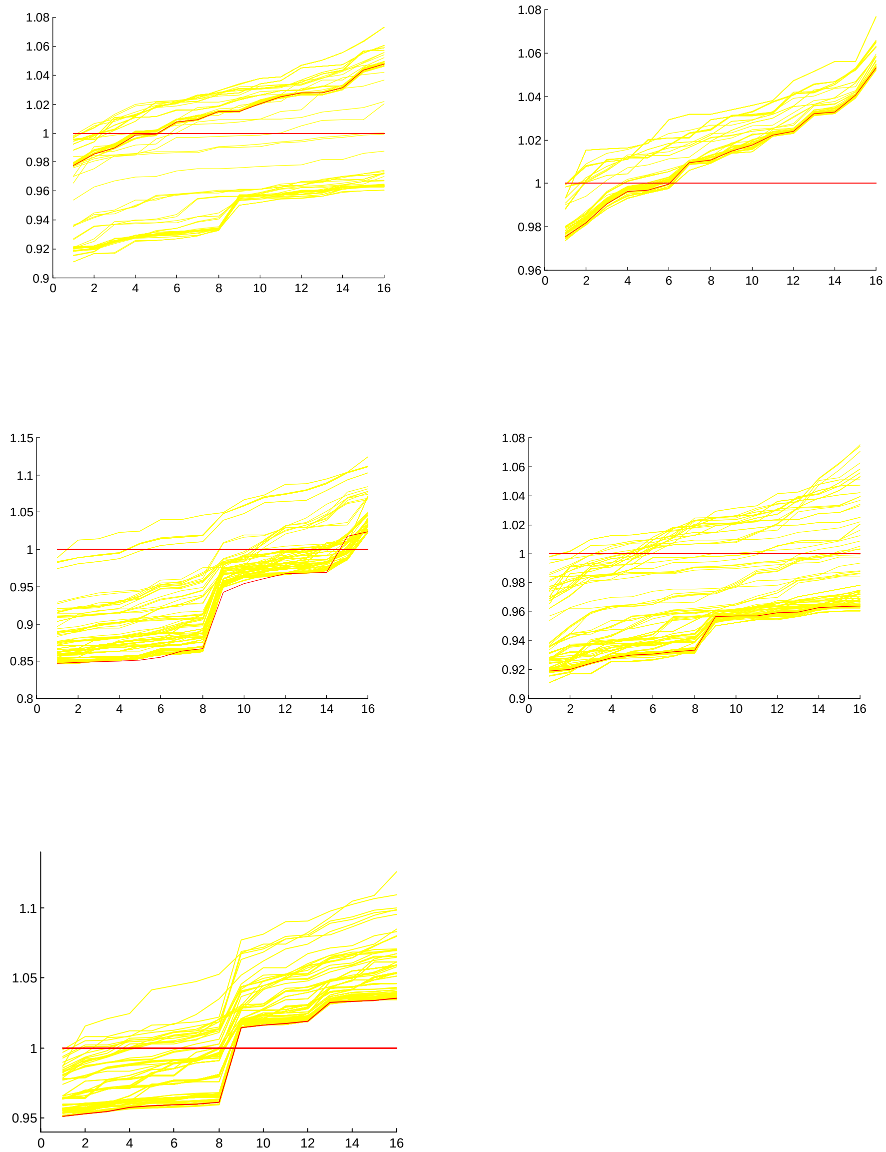
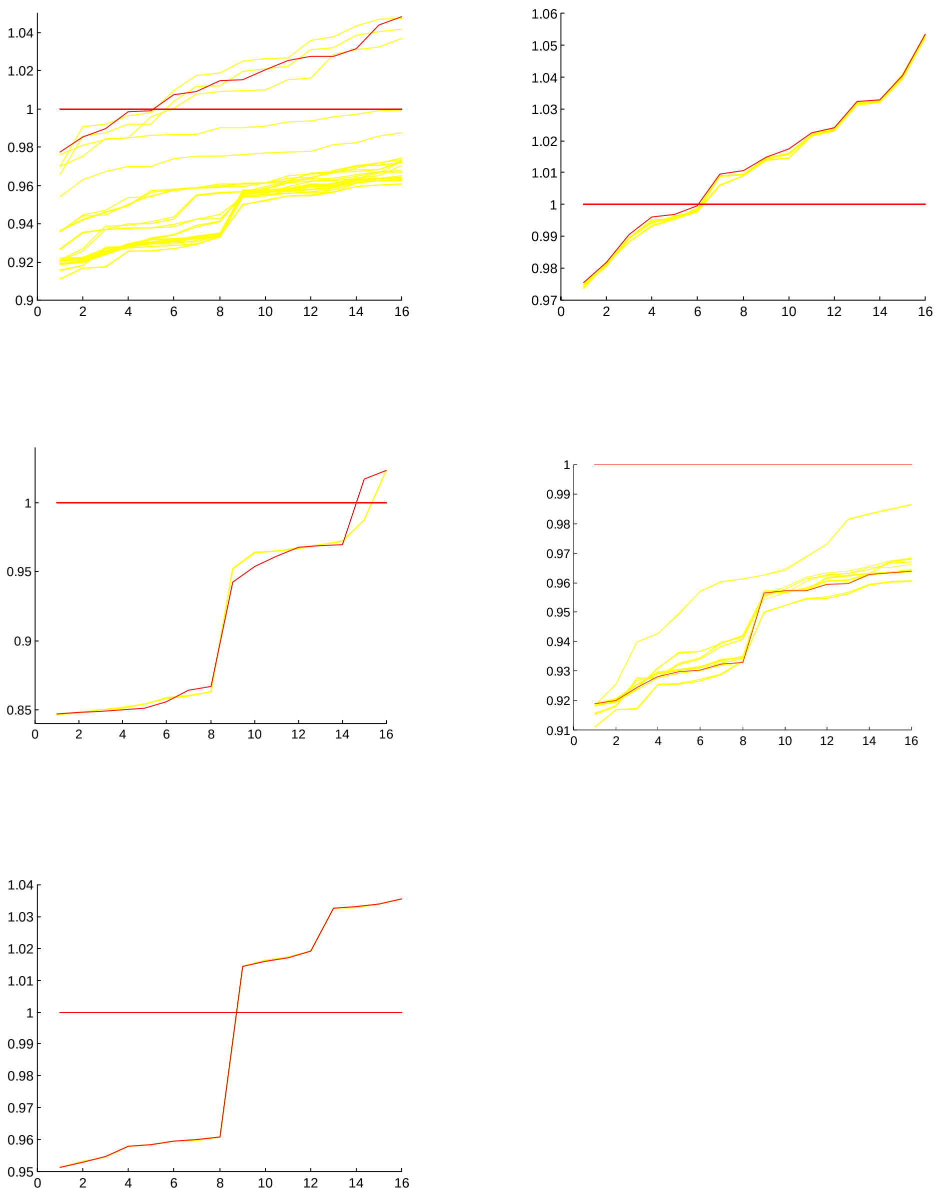


Figure 6: Filterings better than the best linear model: CAC, FTSE, DAX, PSI, AEX



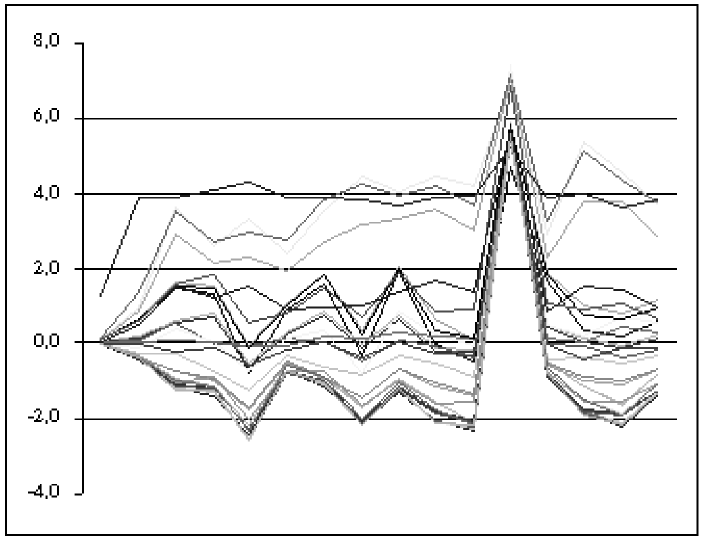


Figure 7: Diebold and Mariano test: CAC40

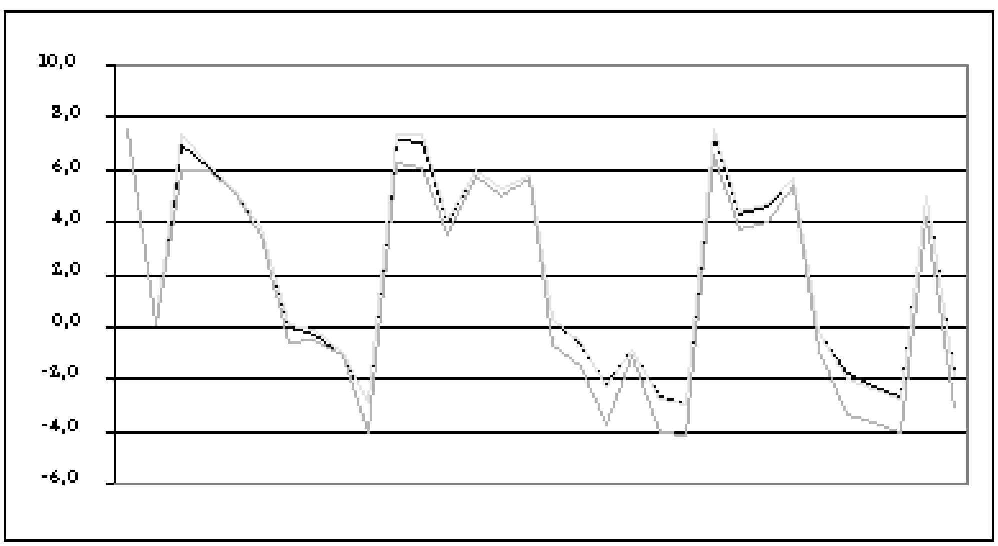


Figure 8: Diebold and Mariano test: DAX30