BOOTSTRAPPING FINANCIAL TIME SERIES

Esther Ruiz
Lorenzo Pascual
Universidad Carlos III de Madrid

Abstract. It is well known that time series of returns are characterized by volatility clustering and excess kurtosis. Therefore, when modelling the dynamic behavior of returns, inference and prediction methods, based on independent and/or Gaussian observations may be inadequate. As bootstrap methods are not, in general, based on any particular assumption on the distribution of the data, they are well suited for the analysis of returns. This paper reviews the application of bootstrap procedures for inference and prediction of financial time series. In relation to inference, bootstrap techniques have been applied to obtain the sample distribution of statistics for testing, for example, autoregressive dynamics in the conditional mean and variance, unit roots in the mean, fractional integration in volatility and the predictive ability of technical trading rules. On the other hand, bootstrap procedures have been used to estimate the distribution of returns which is of interest, for example, for Value at Risk (VaR) models or for prediction purposes. Although the application of bootstrap techniques to the empirical analysis of financial time series is very broad, there are few analytical results on the statistical properties of these techniques when applied to heteroscedastic time series. Furthermore, there are quite a few papers where the bootstrap procedures used are not adequate.

Keywords. Forecasting; GARCH models; Non Gaussian distributions; Prediction; Returns; Stochastic Volatility; Technical Trading Rules; Value at Risk (VaR); Variance ratio test.

1. Introduction

High frequency time series of returns are often characterized by having excess kurtosis and autocorrelated squared observations. These stylized facts can be explained by the presence of conditional heteroscedasticity, i.e. the volatility of returns evolves over time. Given that the marginal distribution of returns is usually non-Gaussian, the inference and prediction of models fitted to returns should not rely on methods based on Gaussianity assumptions. However, bootstrap methods can be adequate in this context; see Korajczyk (1985) for one of the earliest applications of bootstrap methods to analyze financial problems. Many of the earlier papers using bootstrap methods in finance, use procedures based on resampling directly from observed returns without taking into account that returns are sometimes correlated and often not independent. Given that the basic bootstrap techniques were originally developed for independent observations, the bootstrap inference has not the desired properties when applied to raw returns; see, for example, Bookstaber and McDonald (1987), Chatterjee and Pari (1990), Hsieh and Miller (1990) and Levich and Thomas (1993) for some applications where returns are directly bootstrapped.

The application of bootstrap methods in finance has been previously reviewed by Maddala and Li (1996) who pointed out these shortcomings in some of the applications. To take into account the dynamic dependence of returns and, in particular, the conditional heteroscedasticity, there are two possible bootstrap alternatives. First, it is possible to assume a particular model for the volatility and to resample from the returns standardized using the estimated conditional standard deviations. If the volatility is correctly specified, these standardized returns are asymptotically independent and, consequently, the bootstrap procedure has the usual asymptotic properties. Alternatively, the bootstrap procedure can be adapted to take into account that the observations are dependent without assuming a particular model as, for example, in the block bootstrap method.

The objective of this paper is to review the use of bootstrap methods in the analysis of financial time series. In general, these techniques can be used for two objectives. First of all, it is possible to estimate the distribution of an estimator or test statistic. Secondly, it is possible to estimate directly the probability distribution of returns. The paper is organized as follows. In Section 2, we briefly describe the main bootstrap procedures for time series. Section 3 reviews the application of bootstrap procedures for inference in financial models. The main application of bootstrap techniques in this context is to analyze the predictive ability of technical trading rules. In Section 4, we describe several studies that apply bootstrap methods to obtain the distribution function of returns that is fundamental in prediction and Value at Risk (VaR) models. Finally, Section 5 contains the conclusions.

2. Bootstrap techniques for time series

The bootstrap, introduced by Efron (1979), appeared originally as a procedure to measure the accuracy of an estimator. Its main attraction relies on the fact that it can approximate the sampling distribution of the estimator of interest even when this is very difficult or impossible to obtain analytically and only an asymptotic approximation is available. Even more, the bootstrap has the advantage that is very easy to apply independently of the complexity of the statistic of interest.

To illustrate the bootstrap methodology, let us consider one of the most common situations found in statistics. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a set of n independent and identically distributed (*iid*) observations with distribution function F, and let $\theta = s(F)$ be the unknown parameter to be estimated. Given that the empirical distribution function $\mathbf{F}_n$ is a good approximation of the true but unknown distribution F, a natural estimator for θ is $\hat{\theta} = s(\mathbf{F}_n)$. However, knowledge of the sampling distribution of the estimator or at least its mean and variance is only possible in very simple situations and, usually, the asymptotic distribution is used to approximate it. Furthermore, the standard errors are useful for summarizing the precision of estimates when the distribution is symmetric. However, when the estimator has a severely skewed finite sample distribution, bootstrap interval estimates summarize better the distribution. The bootstrap methodology allows an approximation of the distribution of $\hat{\theta}$ under very general conditions and it is based on obtaining a bootstrap replicate, $x_1^*, x_2^*, \ldots, x_n^*$, of the available data set $x_1, x_2, \ldots, x_n$, by drawing with replacement random samples from $\mathbf{F}_n$. Once B bootstrap replicates of the original data set, with the corresponding B bootstrap realizations of the parameter of interest $\hat{\theta}_i^*$, $i = 1, \ldots, B$, have been obtained, the resampling distribution of the bootstrap statistic θ^* is used to approximate the distribution of $\hat{\theta}$. Obviously, the bigger the value of B, the better is the Monte Carlo approximation of θ^*, with the only price of larger computational cost; see Efron and Tibshirany (1993) and Shao and Tu (1995).

With respect to the asymptotic validity of the bootstrap procedure, it is usual to prove that some distance, usually the Mallows distance, between the bootstrap distribution of θ^* and the sampling distribution of $\hat{\theta}$ goes to zero as the sample size increases to infinity. Under some circumstances, the bootstrap distribution enables us to make more accurate inferences than the asymptotic approximation.

The bootstrap method just described is the simplest version and is only valid in the case of *iid* observations. If the standard bootstrap is applied directly to dependent observations, the resampled data will not preserve the properties of the original data set, providing inconsistent statistical results. In particular, the standard bootstrap procedure is neither consistent nor asymptotically unbiased under heteroscedasticity; see Wu (1986) in the context of regression models. Recently, several parametric and nonparametric bootstrap methods have been developed for time series data. The parametric methods are based on assuming a specific model for the data. After estimating the model by a consistent method, the residuals are bootstrapped; see Freedman and Peters (1984) and Efron and Tibshirani (1986). If the serial dependence of the data is misspecified, the parametric bootstrap could be inconsistent. Consequently, alternative approaches that do not require fitting a parametric model have been developed to deal with dependent time series data. Kunsch (1989) proposed the moving block bootstrap method that divide the data into overlapping blocks of fixed length and resample with replacement from these blocks. The bootstrap replicates generated by the moving block method are not stationary even if the original series is stationary. For this reason, Politis and Romano (1994) suggest the stationary bootstrap method that resamples from blocks of data with random lengths. In the context of heteroscedastic time series, Wu (1986) proposed a weighted or wild bootstrap method that provides a consistent estimate of the variance of a test statistic in the presence of heteroscedasticity. The wild bootstrap is based on weighting each original observation with random draws with replacement from a standard normal distribution. Malliaropulos and Priestley (1999) propose a nonparametric implementation of this method that does not rely on the normal distribution. Hafner and Herwartz (2000) also use another alternative version of this procedure.

Li and Maddala (1996) and Berkowitz and Kilian (2000) review the most relevant developments in bootstrapping time series models, and show that the bootstrap algorithms that make use of some parametric assumptions about the model appropriate for the data, are preferable in many applications in time series econometrics.

With respect to testing a given null hypothesis, H_0, it is fundamental to bootstrap from the correct model. In the case of time series data, it is usually not recommended to bootstrap from the raw data but from the residuals from a given model. However, it is necessary to decide which are the residuals to be bootstrapped. Consider, for example, the following AR(1) model:

$$y_t = \phi y_{t-1} + u_t$$

and the null hypothesis H_0: $\phi = \phi_0$. In this case, we have mainly two alternative series of residuals from the following models:

$$a)\ y_t = \hat{\phi} y_{t-1} + \hat{u}_t$$

$$b)\ y_t = \phi_0 y_{t-1} + \tilde{u}_t$$

Denote by $\hat{u}_t^*$ the residuals resampled from $\hat{u}_t$ and by $\tilde{u}_t^*$, the residuals resampled from $\tilde{u}_t$. Then, it is possible to obtain bootstrap replicates of the variable y_t by one of the following schemes:

$$i)\ y_t^* = \hat{\phi} y_{t-1}^* + \hat{u}_t^*$$

$$ii)\ y_t^* = \phi_0 y_{t-1}^* + \hat{u}_t^*$$

$$iii)\ y_t^* = \phi_0 y_{t-1}^* + \tilde{u}_t^*$$

Although, the third scheme is the most appropriate for hypothesis testing, the other two alternatives have also been used in practice. For example, Hall and Wilson (1991) provide guidelines for hypothesis testing using the first alternative while Ferreti and Romo (1996) consider the second one to test for unit roots.

Bootstrap based methods can also be used to obtain prediction densities and intervals for future values of a given variable without making distributional assumptions on the innovations and, at the same time, allowing the introduction, into the estimated prediction densities, of the variability due to parameter estimation. The most influential bootstrap procedure to construct prediction intervals for future values of time series generated by linear AR(p) models, is due to Thombs and Schucany (1990). This method needs the backward representation of the autoregressive model to generate bootstrap series that mimic the structure of the original data, keeping fixed the last p observations in all bootstrap replicates. The use of the backward representation to generate bootstrap series makes the method computationally expensive and, what is more important, restricts its applicability exclusively to those models having a backward representation, excluding, for example, the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) class of models. Furthermore, the prediction in models with a moving average component is not possible with this methodology since, at least theoretically, the whole sample should be kept fixed when generating bootstrap replicates because of the infinite order of the corresponding autoregressive representation. Cao *et al.* (1997) present an alternative bootstrap method that does no require the backward representation. However, the corresponding prediction intervals do not incorporate the uncertainty due to parameter estimation as they are conditional on parameter estimates.

To overcome these drawbacks, Pascual *et al.* (1998) propose a new bootstrap strategy to obtain prediction densities for general ARIMA models. With this new methodology it is possible to incorporate the variability due to parameter estimation into the prediction densities without requiring the backward representation of the process. Therefore, the procedure is very flexible and easy to use, and what is more important, can be extended and adapted easily to processes without a backward representation and, in particular, to GARCH processes. Finally, Gospodinov (2002) analyses the prediction accuracy of another bootstrap procedure to compute the median unbiased forecast of near-integrated autoregressive processes. He illustrates the properties of this procedure analyzing one-month U.S. T-bill yields which are highly persistent although the presence of an exact unit root is inconsistent with the bond pricing theory. This procedure is also based on the use of the backward representation and could be modified along the lines of the procedure suggested by Pascual *et al.* (1998).

In a recent essay on bootstrap techniques, Horowitz (2001) points out that bootstrap methods for time series data are less well developed than methods for *iid* observations and that important research remains to be done. This fact is even more clear when looking at applications of bootstrap procedures to data generated by non-linear models and, in particular, by GARCH and Stochastic Volatility (SV) models.

3. Inference

In this section, we describe several applications of bootstrap procedures to analyze the dynamic properties of financial returns that appear after the review of Maddala and Li (1996). First, we consider tests related with the dynamic behavior of the conditional mean of returns. Then we review the papers where bootstrap procedures have been applied to test for dynamics in the conditional variance. One of the areas where bootstrap techniques have been widely applied is to test for the superiority of technical trading rules and we dedicate one separate subsection to inference on trading rules. Finally, we present other applications of bootstrap procedures to financial time series.

3.1. *Testing for dynamics in the conditional mean of returns*

In this subsection, we review the papers using bootstrap procedures to test for the dynamic components in the conditional mean of returns. Numerous studies have found that daily stock market returns exhibit positive low-serial correlation that is often attributed to non-synchronous trading effects. Consequently, there is great interest in testing for the presence of such autoregressive dynamics in returns. For example, Malliaropulos (1996) apply the variance ratio test, proposed by Cochrane (1988), to monthly observations of the FT-A All Share index and Pan *et al.* (1997) to currency futures prices. The latter authors use both the asymptotic standard errors and bootstrap p-values for the test and conclude that the results are similar. However, it is important to notice that both Malliaropulos (1996) and Pan *et al.* (1997) bootstrap directly from the raw returns that, as mentioned in the introduction, are not independent, although they are uncorrelated under the null hypothesis. Therefore, the bootstrap p-values may be inappropriate. To solve this problem, Malliaropulos and Priestley (1999) obtain the finite sample distribution of the variance ratio test using the weighted bootstrap method. They apply the variance ratio test to unexpected excess returns of several South Asian stock markets after accounting for time-varying risk and potential partial integration of the local stock market into the world stock market. It is concluded that, although excess returns exhibit mean reversion in a number of markets, the failure to reject the random walk hypothesis is related to mean-reversion of expected returns rather than to market inefficiency. Alternatively, Politis, *et al.* (1997) propose a subsampling method to test the null hypothesis of uncorrelated returns by means of the variance ratio test. This method has the advantage that it works for dependent and heteroscedastic returns.

To illustrate the effect of the presence of conditional heteroscedasticity on the bootstrap densities of the variance ratio test, 1000 series have been simulated by the following GARCH(1, 1) model

$$\begin{aligned} y_t &= \varepsilon_t \sigma_t, \qquad t = 1, \ldots, T \\ \sigma_t^2 &= 0.05 + 0.1 y_{t-1}^2 + 0.85 \sigma_{t-1}^2 \end{aligned} \tag{1}$$

where y_t represents the series of returns, i.e. $y_t = \log(p_t/p_{t-1})$, p_t is the stock price at time t, σ_t is the volatility and ε_t is a white noise that has been generated by both a Gaussian distribution and a standardized Student-t distribution with 5 degrees of freedom. Notice that the Student-t distribution has been proposed by many authors as the conditional distribution of returns; see, for example, Baillie and Bollerslev (1989). Table 1 reports the Monte Carlo results on the average p-values of the variance ratio statistic given by

$$VR(q) = \frac{\sum_{t=q+1}^{T} (p_t - p_{t-q})^2}{\sum_{t=2}^{T} (p_t - p_{t-1})^2} \tag{2}$$

for $T = 300$ and $T = 1000$ and $q = 2, 5, 10$ and 20 when series are generated by model (1) with ε_t being Gaussian. To make the comparisons simpler, the statistic

Table 1. Monte Carlo results on p values of $VR(q)$ statistic. GARCH (1, 1) returns with Gaussian errors

q	T	Average p values			
		Empirical	Asymptotic	Bootstrap 1	Bootstrap 2
2	300	0.4858	0.5183	0.5105	0.4965
	1000	0.4994	0.5250	0.5120	0.5035
5	300	0.4769	0.5246	0.5195	0.4872
	1000	0.4976	0.5158	0.5105	0.4935
10	300	0.4858	0.5568	0.5417	0.4958
	1000	0.5026	0.5299	0.5240	0.4966
20	300	0.4881	0.5941	0.5650	0.5026
	1000	0.5043	0.5468	0.5372	0.4987

has been standardized using its asymptotic standard deviation, as given by Lo and McKinlay (1989) so that all the statistics are asymptotically N(0,1). In this table, it is possible to observe that when the bootstrap is based on resampling from the raw returns (Bootstrap 1), the asymptotic and bootstrap p-values are similar, a result that was also reported by Pan *et al.* (1997). Furthermore, both the asymptotic and bootstrap p-values exceed the corresponding empirical p-values. However, when the bootstrap is based on the standardized returns (Bootstrap 2), the sampling and the bootstrap distributions of the VR(q) statistic are closer.

Figures 1 and 2 represent the empirical density of the VR(q) statistic for series generated by model (1) with ε_t having a Student-t distribution for $T = 300$ and 1000, respectively. These figures also represent the bootstrap densities obtained by resampling from the raw returns (bootstrap 1) and from the returns standardized using the estimated GARCH conditional standard deviations (bootstrap 2) for two particular series generated by the same model. It is clear that, in the latter case, the estimated sampling density is closer to the empirical density. Furthermore, notice that the performance of bootstrapping without taking into account the conditional heteroscedasticity deteriorates as the sample size increases. Consequently, the p-values based on bootstrapping directly from the raw returns may have important distortions. For example, for a series generated with $T = 1000$, the statistic VR(2) is 0.838. In this case, the asymptotic p-value and the bootstrap p-value obtained by resampling from the raw returns are nearly the same, 0.201 and 0.209, respectively. However, bootstrap p-values obtained from the heteroscedastic model is 0.249 that is closer to the empirical p-value of 0.271. Notice that, once more, it is possible to observe that the asymptotic and the bootstrap p-values based on raw returns are similar.

We now consider the empirical application of the VR statistic to test for the presence of autoregressive components in the exchange rate of the British Pound against the Dollar observed daily from the 1 January 1990 to 31 December 2001, with $T = 3040$. The series of returns, $y_t = 100 \log(p_t / p_{t-1})$, where p_t is the

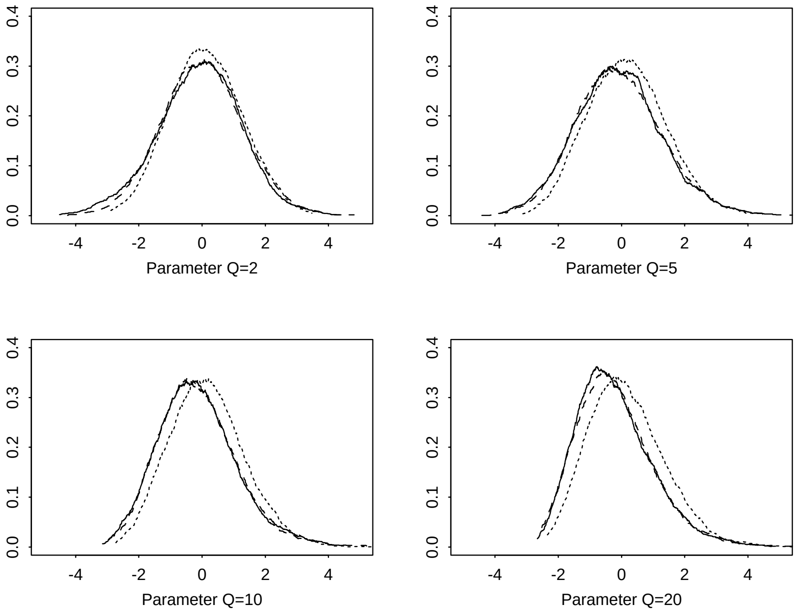

Figure 1. Empirical ( ) and bootstrap densities of variance ratio statistic for a series generated by a GARCH(1, 1) model with conditional Student t distribution with 5 degrees of freedom. $T = 300$.

exchange rate at time t, has been plotted in Figure 3. Table 2 reports the values of the VR(q) statistic for $q = 2, 5, 10$ and 20, together with the corresponding asymptotic p-values and the bootstrap p-values obtained from the raw returns. Notice that both p-values are very similar for all values of q considered. However, the returns are not independent. Fitting a GARCH(1, 1) model, the following estimates are obtained:

$$\hat{\sigma}_t^2 = \underset{(0.0001)}{0.0007} + \underset{(0.0047)}{0.0444}\; y_{t\;\;1}^2 + \underset{(0.0058)}{0.9443}\; \hat{\sigma}_{t\;\;1}^2 \qquad (3)$$

The p-values obtained by resampling from the corresponding standardized returns are also reported in Table 3. Observe that, these p-values are always greater than the corresponding asymptotic p-values. Furthermore, the results of the test can be reversed depending on which p-value is used. For example, when $q = 10$, the null of no autocorrelation is rejected using both the asymptotic and the bootstrap p-values based on raw returns. However, the null hypothesis is not rejected when bootstrapping from the standardized returns.

Summarizing, we have shown with both simulated and real data that bootstrapping raw returns in order to obtain p-values of the VR statistic when

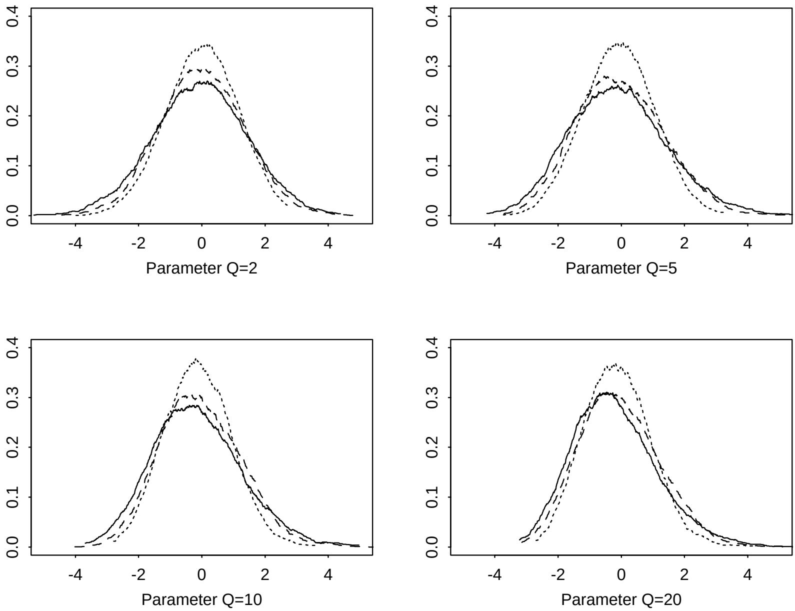

Figure 2. Empirical ( ) and bootstrap densities of variance ratio statistic for a series generated by a GARCH(1, 1) model with conditional Student t distribution with 5 degrees of freedom. $T = 1000$.

there is conditional heteroscedasticity, could seriously distort the results of the test. Furthermore, it could be expected that the corresponding bootstrap p-values are not very different from the asymptotic p-values. Using bootstrap procedures appropriate for the characteristics of returns, yields p-values remarkably close to the empirical p-values.

The variance ratio test is not the only statistic used in finance to test for autoregressive components in the conditional mean of returns. In a very interesting paper, Hafner and Herwartz (2000) consider two Wald tests based on Quasi-Maximum Likelihood (QML) estimation assuming a GARCH(1, 1) model for the conditional variance. As QML inference depends on the specification of the variance process, they also consider tests based on Ordinary Least Squares (OLS) estimation and a bootstrapped version of the OLS based statistics using the wild bootstrap. The asymptotic convergence of the distribution of the bootstrapped statistics to the asymptotic distribution of the original statistic is proven. By means of Monte Carlo experiments, they show that the wild bootstrap inference shows superior size properties relative to all the other tests considered. However, the power of the bootstrap tests is low in the cases were the volatility is highly persistent. Finally, they apply the alternative tests considered to

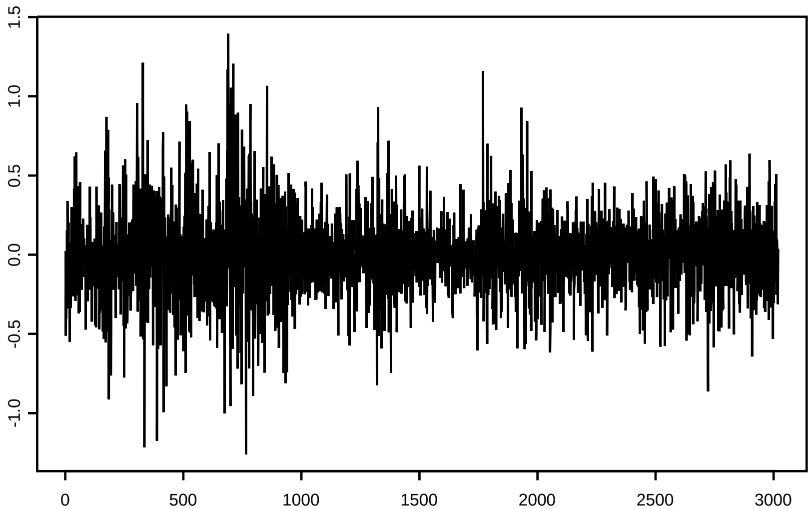

Figure 3. Daily returns of Pound Dollar exchange rate observed from 1 January 1990 to 31 December 2001.

Table 2. $VR(q)$ statistic and p values for British Pound Dollar exchange rate.

		p values		
q	Statistic	Asymptotic	Bootstrap 1	Bootstrap 2
2	3.0438	0.0012	0.0020	0.0070
5	2.5821	0.0049	0.009	0.0360
10	1.9790	0.0239	0.0270	0.0510
20	1.3291	0.0919	0.0960	0.1310

German stock returns, giving in many cases different decisions about acceptance or rejection of the null hypothesis of no autocorrelation.

White and Racine (2001) also test for predictable components in returns by applying bootstrap techniques for inference in artificial neural networks (ANN). They conclude that exchange rates do appear to contain information that is exploitable for enhanced point prediction, but the nature of the predictive relation evolves over time. However, they do not take into account the evolution of the conditional variance.

In relation to testing for the presence of unit roots in exchange rates, Kanas (1998) investigates whether the Dickey-Fuller (DF) test is affected by the presence

Table 3. Summary of recent bootstrap applications testing for dynamics of returns.

Author	Test for	Boot. procedure	Results
Conditional mean			
Malliaropulos (1996)	Autocor. of returns	Raw returns	Not take into account heteros.
Pan *et al*. (1997)	Autocor. of returns	Raw returns	Not take into account heteros.
Kanas (1998)	Unit root in prices		
Mallia. and Priest. (1999)	Autocor. of returns	Weighted boot.	Mean reversion is due to time varying expected returns and partial integration
Politis *et al*. (1999)	Autocor. of returns	Subsampling	Asymptotic properties
Gospodinov (2000)	Non linearities	Standard. returns	Finite sample properties
		Wild Bootstrap	Specifies a TAR model with GARCH errors
		Feasible GLS Boot	
Hafner and Herwa. (2000)	Autocor. of returns	Wild bootstrap	Boots. test have good size and power properties
White and Racine (2001)	Predictable regularities in exchange rates	Raw returns	Not take into account heteros.
Conditional variance			
Tauchen *et al*. (1996)	Persistence Asymmetry Relation vol. prices	Sampling from fitted conditional density	Dynamic impulse response analysis
Brock. and Chow. (1997)	Chaos	Raw returns	Not take into account heteros.
Bollers. and Mikk. (1999)	Fractional integration	Standard. returns	Best model: FIEGARCH
Blake and Kapet. (2000)	ARCH	Raw returns	Artificial neural network Under the null is appropriate
Eftekhari *et al*. (2000)	Measures of risk	Raw returns	Monthly data (homoscedastic)

(*continued*)

Table 3. *Continued.*

Author	Test for	Boot. procedure	Results
Technical trading rules			
Brock *et al*. (1992)	Performance of TTR	Standard. returns	TTR are profitable
Kho (1996)	Performance of TTR	Standard. returns	Different conclusions with standard and bootstrap tests
Mills (1997)	Performance of TTR	Standard. returns	Predictability dissapears after 1980
Besse. and Chan (1998)	Performance of TTR	Raw returns	Not take into account heteros.
Ito (1999)	Performance of TTR	Standard. returns	Importance of time varying expected returns
LeBaron (1999)	Performance of TTR	Raw returns	Effect of Federal Reserve Not take into account heteros.
White (1999)	Performance of TTR	Stationary boot.	Avoid data snooping
Sullivan *et al*. (1999)	Performance of TTR	Stationary boot.	Outperformance disappears out of sample
Chang and Osler (1999)	Performance of TTR	Raw returns	Not take into account heteros
Millet and Michel (2000)	Performance of TTR	Raw returns	Not take into account heteros
Taylor (2000)	Performance of TTR	Standard. returns	Test based on TTR have less power than standard uncorrelation tests
Other applications			
Ikenberry *et al*. (1995)	Event study		Long run returns are not zero
Kothari and Warner (1997)	Event study		Parametric long horizon tests can be missleading
Stanton (1997)	Term structure	Block boot.	Continuous time
Garrant *et al*. (2001)	Target zone nonlinearities	Block boot.	Nonlinearities in specific subsamples
Carriere (2000)	Forward rates	Block boot.	Uses splines
Groene. and Fraser (2001a)	Asset Pricing model	Block boot.	Monthly data (homosced.)
Groene. and Fraser (2001b)	Asset Pricing model	Block boot.	Monthly data (homosced.)

of structural breaks due to realignments in the central parities. Bootstrap simulations are used to generate critical values of the DF test in the presence of multiple dummy variables. He concludes that, once you take into account the realignments, there is no evidence of the presence of unit roots in exchange rates.

Although most of the previous authors conclude that stock returns are not predictable in the short run, there is an interest for long horizon regressions that usually take the following expression:

$$\sum_{i=1}^{k} y_{t+i} = \alpha_k + \beta_k x_k + u_{tk} \tag{4}$$

where x_t is some variable measuring fundamental values, usually dividend yield. Maddala and Li (1996) review extensively several papers using bootstrap techniques in this context. Ikenberry *et al*. (1995) also analyze the long-run behavior of returns by means of an event study analysis. They conclude that long-run abnormal returns are systematically nonzero. They defined the sample buy-and-hold abnormal return as the difference between the buy-and-hold return and the corresponding return on a portfolio of securities matched by book-to-market, size and event date. To assess the statistical significance, this difference is compared to a bootstrap distribution of buy-and-hold abnormal returns. However, Kothari and Warner (1997) point out several potential shortcoming of bootstrap techniques for long-horizon event studies.

In relation to testing for non-linearities in the conditional mean of a series in the presence of high persistence and conditional heteroscedasticity, Gospodinov (2000) proposes to use a Threshold Autoregressive of order one (TAR(1)) model with GARCH(1, 1) errors which is applied to the analysis of the term structure of interest rates. He uses bootstrap approximations to ensure the validity of the statistical inference. In particular, he proposes three alternative bootstrap procedures. The first one is based on bootstrapping the standardized residuals, the second is a wild bootstrap procedure and, finally, he considers a feasible GLS bootstrap. The size and power properties of these approximations are evaluated by simulation and the conclusion is that all of the bootstrap tests have excellent size properties.

Garrant *et al*. (2001) also test for the presence of target-zone nonlinearities in the Pound/Deutschmark exchange rate using the block bootstrap to compute the corresponding p-values.

3.2. *Testing for dynamics in the conditional variance of returns*

There are also hypothesis related to the dynamics of volatility that have been tested using bootstrap procedures. Lamoureux and Lastrapes (1990) were the first to use bootstrap procedures to test if the Integrated GARCH (IGARCH) models, often found in empirical applications, can be the result of structural changes in otherwise stationary GARCH models. However, Maddala and Li (1996) point out that they do not formulate correctly the null hypothesis to be tested and show how the test should be carried out properly.

Tauchen *et al.* (1996) investigate multi-step nonlinear dynamics of daily price and volume movements. Their objective is to examine the persistence properties of stochastic volatility, the asymmetric responses of conditional variances to positive and negative movements in prices and the nonlinear relation between volume and prices. They construct confidence bands for the corresponding impulse response functions by resampling from the fitted conditional densities. The bootstrap method they used is described in Gallant *et al.* (1993).

Later, Brockman and Chowdhury (1997) applied bootstrap techniques to distinguish whether the intra-day implied volatility of the S&P100 index call option is stochastic or has a chaotic deterministic behavior. However, they are bootstrapping from the raw returns series that are not independent. Therefore, the properties of the bootstrap procedure can be seriously affected.

Bollerslev and Mikkelsen (1999) analyze whether the long-run dependence in U.S. stock market volatility is best described by a slowly mean-reverting fractionally integrated process by inferring the degree of mean-reversion implicit in a panel data set of transaction prices on the S&P500 composite stock price index. They compare the observed prices with risk-neutralized prices bootstrapped from the residuals standardized with standard deviations estimated by different heteroscedastic models. They conclude that the Fractionally Integrated EGARCH (FIEGARCH) model of Bollerslev and Mikkelsen (1996) results in the lowest average absolute and relative pricing errors.

Also, in relation to testing the dynamics of volatility, Blake and Kapetanios (2000) propose a test for ARCH based on a neural network specification. As the test suffers from size distortions, they use bootstrap procedures to correct them.

Finally, Eftekhari *et al.* (2000) compare different measures of risk, namely the semi-variance, the lower partial moment, the Gini and the absolute deviation using both simulated and real series of monthly returns. They draw, with replacement, returns from each of the samples of real data and the alternative measures of risk are calculated for each of the bootstrapped samples.

3.3. *Technical trading rules*

One of the most popular methods to analyze the hypothesis that equity markets are efficient is based on technical analysis. Trading rules are used to classify each day t as either Buy, Sell or Neutral, using information available up to day t. Technical trading rules are rather important in practice given that they are almost universally used by practitioners; see the references in Chang and Osler (1999). A trading rule is said to uncover evidence of price predictability if expected returns depend on the Buy/Sell information. To assess this dependency, it is natural to test for the difference between the average returns for Buy and Sell days. The obvious test of the null hypothesis that there is no predictability is based on the following statistic

$$z = \frac{\bar{r}_I - \bar{r}_J}{\left(\frac{s_I^2}{n_I} + \frac{s_J^2}{n_J}\right)^{0.5}}$$

where $\bar{r}_I$, s_I^2 and n_I are, respectively, the sample mean, variance and number of returns for Buy days, and $\bar{r}_J$, s_J^2 and n_J are the corresponding measures for Sell days. The asymptotic distribution of the z statistic is standard normal when the returns process is a strictly stationary, martingale difference with finite second moments.

When several trading rules are considered, another interesting hypothesis is whether there exists a superior technical trading rule that significantly outperforms a benchmark of holding cash. The null hypothesis, in this case, is that the expected return of the best trading rule is no better than the expected return of the benchmark.

Due to the non-normality of returns, it is sensible to use bootstrap procedures to estimate the distribution of these statistics. In a seminal paper in this area, Brock *et al.* (1992) propose to combine technical analysis and bootstrap procedures. They proposed a bootstrap procedure to obtain a better approximation of these statistics and to decide if some specific statistical model can explain the observed trading rules results. A statistic z is calculated from a trading rule applied to the observed series. Then a particular statistical model is fitted to the observed returns and artificial price series are generated by sampling from the corresponding residuals together with the estimated parameters. The same statistic z is computed for each of the artificial price series, obtaining a sequence of bootstrap statistics, $z_1^*, z_2^*, ..., z_B^*$. The proportion of statistics z_i^* that are more extreme than z, is the p-value for the test of the null hypothesis that the particular model generates observed prices. They apply this bootstrap method to analyze the properties of the Dow Jones Index observed daily from 1897 to 1986, bootstrapping the p-values for the difference between Buy and Sell average returns by applying 26 technical trading rules, and conclude that they significantly outperform the benchmark. However, they explicitly mention that the asymptotic properties of the bootstrap procedure proposed are not known for some models of the GARCH family as, for example, EGARCH and GARCH-M. Furthermore, they suggest that the results of the test are not qualitatively altered whether the asymptotic or the bootstrapped standard errors are used. Finally, they note the dangers of data-snooping when testing the profitability of a large number of trading rules on the same sample of returns. Data-snooping occurs when a given data set is used more than once for inference or data selection. In this case, there is the possibility that positive results can be due simply to chance. As they are testing 26 trading rules one by one, there is a reasonable possibility that data-snooping could be occurring. Therefore, the evidence in favor of a superior performance of trading rules can be tempered. Finally, it should be mentioned that the combination of bootstrap methods with trading rules has been more fruitful as an instrument to check the adequacy of several commonly used models like Random Walks, GARCH and the Markov switching regression models. For this purpose, Brock *et al.* (1992) propose to bootstrap the residuals from a fitted model and the estimated parameters to obtain bootstrap replicates of the original data. They compute the trading rule profits for each bootstrap replicate and compare the corresponding bootstrap distribution with the trading rule profits derived from the actual data.

The application of the procedures proposed by Brock *et al.* (1992) is very extensive in the literature. For example, Mills (1997) applies their methodology to data on the London Stock Exchange FT30 index for the period 1935–1994. Although he found that trading rules outperform the benchmark when using data up to 1980, the predictive ability of the trading rules after this date disappears. Later, LeBaron (1999) tests whether the predictive ability of trading rules over future movements of foreign exchange rates changes after removing periods in which the Federal Reserve is active. Maillet and Michel (2000) apply the test proposed by LeBaron (1999) to twelve exchange rates. They also use bootstrap methods to estimate the distribution of both trading rule returns and raw returns to analyze whether filtering the raw exchange series with some trading rule significantly changes their characteristics. Finally, Taylor (2000) studies the predictability of several U.K. financial prices by fitting ARMA-ARCH models to the corresponding returns.

As noted by Brock *et al.* (1992), there is a danger of data-snooping when testing one by one the performance of a high number of trading rules. To avoid it, White (2000) applies the stationary bootstrap to test whether the performance of the best trading rule is no better than the benchmark. Later, Sullivan *et al.* (1999) apply White's (2000) bootstrap methodology to present a comprehensive test of performance across several technical rules. They show that, even after adjustments for data-snooping, some of the trading rules considered by Brock *et al.* (1992) outperform the benchmark. However, their results do not hold out-of-sample.

However, even after Maddala and Li (1996) highlighted the dangers of bootstrapping from raw returns, there are some authors who still do not take into account the presence of conditional heteroscedasticity when using bootstrap procedures to analyze the profitability of technical trading rules; see, for example, Bessembinder and Chan (1998) and Chang and Osler (1999).

Kho (1996) analyses the performance of trading rules on currency futures markets using an alternative procedure to the one proposed by Brock *et al.* (1992). He applies a bootstrap procedure based on observations standardized assuming a GARCH-M specification, to some versions of the conditional international Capital Asset Pricing Model (CAPM) for time-varying expected returns and risk. Subsequently, Ito (1999) evaluates the profitability of technical trading rules by using equilibrium asset pricing models. He found that using standard or bootstrap p-values, the conclusions can be reversed.

Finally, the Contrarian Hypothesis, also related to trading rules, states that stocks that consistently underperform (outperform) the market will outperform (underperform) over subsequent periods, those stocks that have previously outperformed (underperformed) the market. In two closely related papers, Mum *et al*. (1999, 2000), use exactly the same methodology to test this hypothesis for French and German stock markets in the first paper, and for US and Canadian stock markets in the second. The bootstrap procedure they use, however, is not appropriate, in the main because they are not resampling under the null hypothesis, but also because it is hard to believe that it is really a bootstrap procedure.

3.4. *Other tests*

There are other applications of bootstrap procedures to hypothesis testing related to financial data. For example, Stanton (1997) estimates non-parametrically the parameters of continuous time diffusion processes that are observed at discrete times using kernel estimators of the corresponding conditional expectations. He uses the block bootstrap to calculate confidence bands for the estimated densities.

Later, Carriere (2000) constructs confidence intervals for forward rates estimated with spline models that take into account the heteroscedasticity and correlation in the data. They resample from the residuals standardized to have constant variance and no autocorrelation.

Finally, in two very closely related papers, Groenewold and Fraser (2001a,b) analyze the sensitivity of tests of asset-pricing models to violations of the Gaussianity hypothesis. In the former paper, they use Australian data and in the latter, US and UK data, to compare the standard test with those based on GMM estimators and on bootstrap procedures. They conclude that standard methods are robust to Gaussianity. However, their results have two limitations. First, although they mention three alternative bootstrap procedures, the standard procedure based on resampling directly from the returns, a block bootstrap and a parametric bootstrap based on fitting a model for the conditional variance, the first is inappropriate and they do not implement the third. Therefore, only the block bootstrap may have the desired properties. The second limitation is concerned with the properties of the data they analyze, namely the observations are monthly and the presence of conditional heteroscedasticity is very weak. Therefore, it is not surprising that the results based on the bootstrap or on standard asymptotic distributions are similar.

Table 3 summarizes the main contributions described in this section.

4. Distribution of returns and volatilities

Bootstrap procedures can be used not only to estimate the sample distribution of a given statistic but also to obtain estimates of the density of the variable being analyzed. In this section, we review the papers that apply bootstrap procedures to obtain prediction densities of future returns and their volatilities and to estimate the VaR.

4.1. *Prediction*

Prediction is one of the main goals when a dynamic model is fitted to returns. In that sense, GARCH and SV models have the attraction that they can provide dynamic prediction intervals that are narrow in tranquil times and wide in volatile periods. Furthermore, there is an increasing interest in interval forecasts as measures of uncertainty; see, for example, Bollerslev (2001) and Engle (2001). On the other hand, the volatility of returns is a key factor in many models of option valuation and portfolio allocation problems. Therefore, accurate predictions of volatilities are critical for the implementation and evaluation of asset and derivative pricing theories, as well as trading and hedging strategies. Bootstrap-based methods lead to prediction intervals that incorporate the uncertainty due to parameter estimation without distributional assumptions on the sequence of innovations. As described in Section 2, these methods have proved to be very useful for obtaining prediction intervals for future values of series generated by linear ARIMA models. However, if the presence of conditional heteroscedasticity is not taken into account, the coverage properties of bootstrap intervals for high frequency returns can be distorted; see, for example, Kim (2001) in the context of VAR(1) models. Consequently, Miguel and Olave (1999a) extend the procedure of Cao *et al.* (1997) to stationary ARMA processes with GARCH(1, 1) innovations and prove the asymptotic validity of the corresponding bootstrap procedure to obtain prediction intervals for future returns. These prediction intervals are conditional on the parameter estimates and, consequently, do not incorporate the uncertainty due to parameter estimation. As volatility is specified as a function of past observations in GARCH models, future volatilities are known given the parameters and past observations. As a consequence, the bootstrap procedure proposed by Miguel and Olave (1999a) cannot be used to obtain prediction intervals for future volatilities. Miguel and Olave (1999b) carry out a Monte Carlo experiment to compare the performance of the conditional bootstrap intervals with the Cornish-Fisher approximation proposed by Baillie and Bollerslev (1992). They show that when the prediction horizon is longer than one period, the bootstrap prediction intervals have coverages closer to the nominal than the intervals based on Cornish-Fisher approximations. Gospodinov (2002) also proposes an alternative bootstrap procedure conditional on parameter estimates to forecast future returns modeled by a TAR(1) model with GARCH(1, 1) errors.

Pascual *et al.* (2000) generalize the bootstrap procedure of Pascual *et al.* (1998) to obtain prediction densities of both returns and volatilities of series generated by GARCH processes. The main advantage of their proposal is that the procedure incorporates the variability due to parameter estimation and, therefore, it is possible to obtain bootstrap prediction densities for the volatility process. The asymptotic properties of the procedure are derived and the finite sample properties are analyzed by means of Monte Carlo experiments which show that the properties of intervals for future returns are adequate. They also show that incorporating the uncertainty due to parameter estimation makes no difference when generating prediction intervals for returns if the error distribution is symmetric. However, when constructing prediction intervals for future volatilities, it is necessary to introduce this uncertainty to have coverage close to the nominal values. However, the length of intervals for future volatilities is well above the empirical values. Finally, they apply their bootstrap procedure to obtain prediction densities of future values and volatilities of the IBEX35 index of the Madrid Stock Exchange.

To illustrate the use of the bootstrap to obtain prediction densities for future returns and volatilities, the procedure proposed by Pascual *et al.* (2000) has been applied to the series of returns of the Pound–Dollar exchange rate described in Section 3. Although the series consists of $T = 3039$ observations, only 3019 have been used for estimation purposes, leaving the last 20 observations for out-of-sample forecast evaluation. Recall that the VR(q) test detects autoregressive components in the returns series for q = 2 and 5. Therefore, we fit an AR(1) model with GARCH(1, 1) errors. The estimated model is given by

$$\begin{aligned} y_t &= \underset{(0.0181)}{0.0652}\, y_{t-1} + a_t \\ a_t &= \varepsilon_t \sigma_t \\ \hat{\sigma}_t^2 &= \underset{(0.0001)}{0.0007} + \underset{(0.0047)}{0.0443}\, a_{t-1}^2 + \underset{(0.0058)}{0.9447}\, \hat{\sigma}_{t-1}^2 \end{aligned} \tag{5}$$

Figure 4 represents a kernel estimate of the density of the standardized residuals, $\hat{\varepsilon}_t = a_t/\hat{\sigma}_t$, together with the standard normal density. Notice that the density of $\hat{\varepsilon}_t$ has fat tails. In particular, the kurtosis is 4.6711. Therefore, the conditional

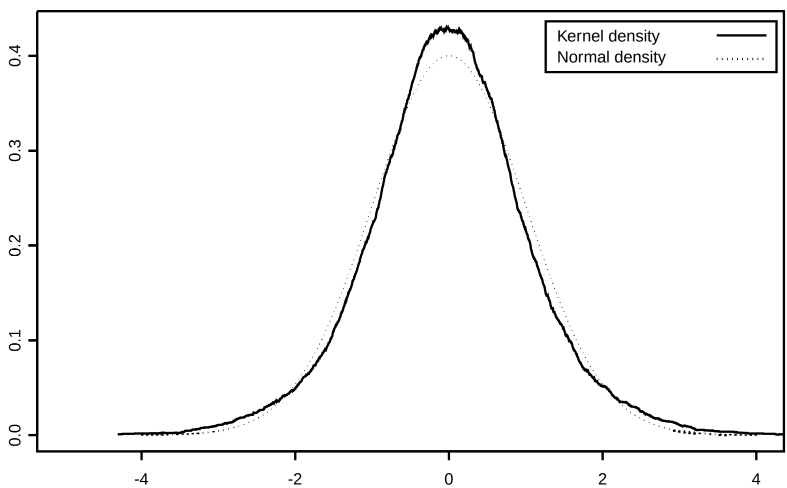

Figure 4. Kernel density of Pound Dollar exchange rate returns standardized with GARCH(1, 1) standard deviations.

Gaussianity of returns is rejected when a GARCH(1, 1) model is fitted. Figure 5 represents the bootstrap densities estimated for 1, 5, 10 and 20 steps-ahead predictions of returns. Using these bootstrap densities, it is possible to construct the corresponding prediction intervals for future returns. Figure 6 represents the 80% and 95% intervals for y_{T+k}, $k = 1, ..., 20$, together with the intervals obtained using the Box-Jenkins methodology. We also plot the point predictions that, in this case, are equal to zero and the actual values of y_{T+k}. Notice that approximately 4 of 20 observations are supposed to lie out the 80% prediction interval. However, the Box-Jenkins intervals are unnecessarily wide leaving only one outside. While the bootstrap intervals are thinner, they leave 4 observations outside. On the other hand, looking at the 95% intervals, they are supposed to leave one observation out.

With respect to the prediction of future volatilities, Figure 7 represents the bootstrap densities for different prediction horizons. The corresponding bootstrap prediction intervals for future volatilities have been plotted in Figure 8, together with the point predictions obtained from the estimated GARCH(1, 1) model in equation (5).

The extension of these bootstrap procedures to estimate prediction densities of returns and volatilities of series generated by SV models seems rather promising in the context of predicting future volatilities. Remember that while in GARCH models the volatility is known one-step-ahead, SV models introduce an

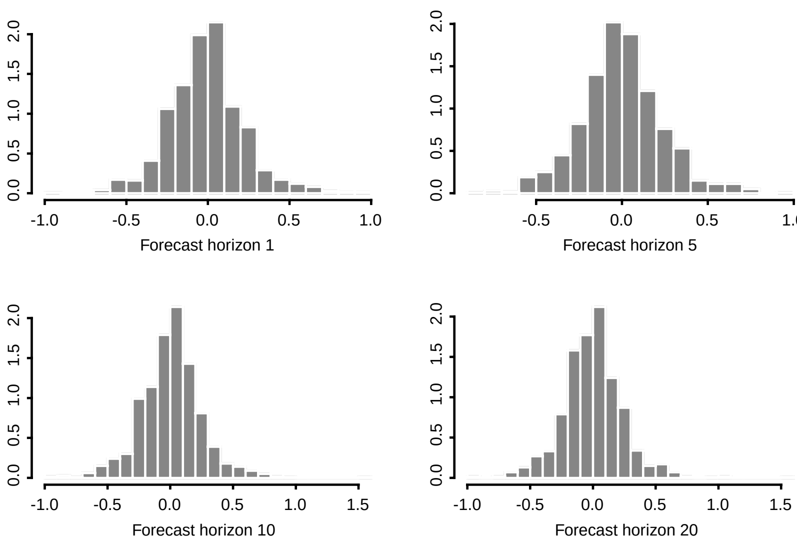

Figure 5. Bootstrap densities of 1, 5, 10 and 20 steps ahead forecasts of Pound Dollar exchange rate returns.

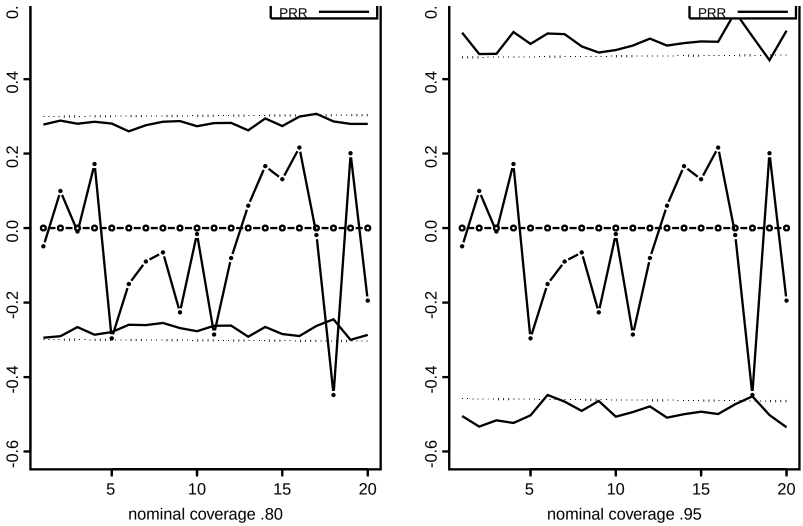

Figure 6. Box Jenkins and bootstrap 80% and 95% prediction intervals for Pound Dollar exchange rate returns together with point predictions (○) and actual values (□).

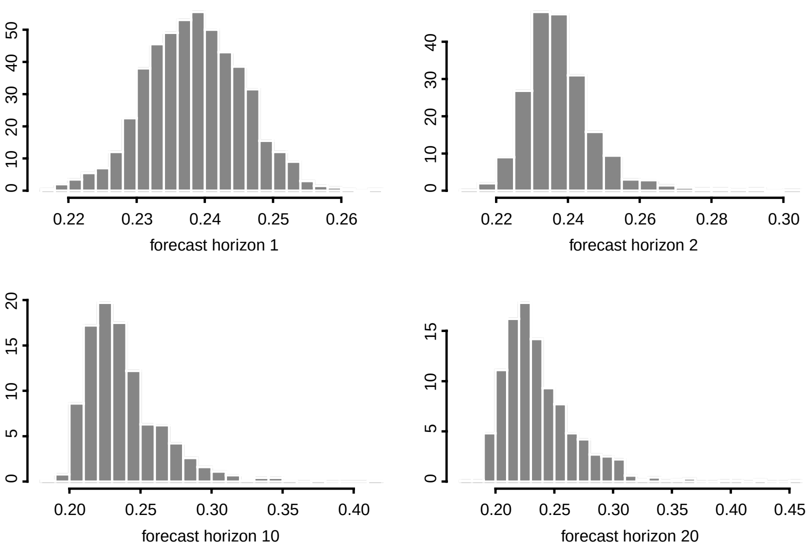

Figure 7. Bootstrap densities of 1, 2, 10 and 20 steps ahead Pound Dollar exchange rate volatilities.

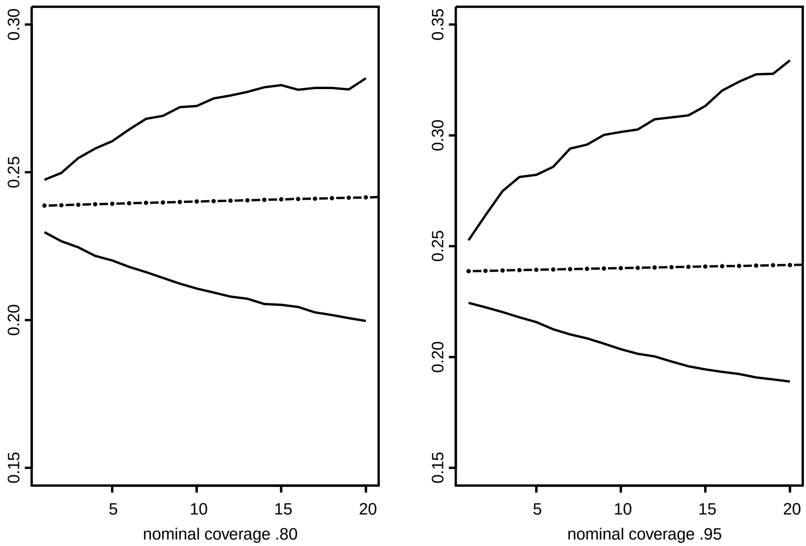

Figure 8. Bootstrap prediction intervals for future Pound Dollar exchange rates volatilities together with their GARCH point predictions (□).

unexpected component that could allow more realistic prediction intervals with better coverage.

4.2. *Value-at-Risk (VaR)*

Financial risk management is dedicated to providing density forecasts of portfolio values and to tracking certain aspects of the densities such as, for example, Value-at-Risk (VaR). The VaR can be defined as the expected loss of a portfolio after a given period of time (usually 10 days) corresponding to the $\alpha\%$ quantile (usually 1%).

The early VaR parametric models impose a known theoretical distribution to price changes. Usually it is assumed that the density function of risk factors influencing asset returns is a multivariate normal distribution. The most popular parametric methods are variance-covariance models and Monte Carlo simulation. However, excess kurtosis of these factors will cause losses greater than VaR to occur more frequently and be more extreme than those predicted by the Gaussian distribution. Consequently, many authors suggest using bootstrap techniques to avoid particular assumptions on the distribution of factors beyond stationarity of the distribution of returns. The procedure consists of generating scenarios by sampling observed returns associated with each risk factor included in the portfolio. The aggregate value of all linear and derivative positions produces a simulated portfolio value. Vlaar (2000) investigates the accuracy of various VaR models on Dutch interest rate-based portfolios and concludes that bootstrap techniques produce satisfactory results when long periods of data are available.

Early bootstrap procedures to compute the VaR of a portfolio assumed constant volatility of returns. However, the ability of bootstrap techniques to predict future losses can be undermined when the volatility evolves over time and, therefore the distribution of risk factors is not i.i.d.. In this case, the probability of having a large loss is not equal across different days. Barone-Adesi *et al.* (1999) propose a bootstrap procedure to obtain VaR estimates based on resampling from returns standardized using GARCH estimates of the volatility. The bootstrapping is done conditional on the parameter estimates and, therefore, is similar to the one proposed by Miguel and Olave (1999a) for obtaining prediction intervals. They illustrate the procedure with a very informative numerical example of a portfolio of three assets. Later, Barone-Adesi *et al.* (2001) compare this method with traditional bootstrapping estimates using three hypothetical portfolios on the S&P500 index and show that the advantages of the standardized bootstrap is magnified by the presence of options in the portfolio.

To illustrate the different alternatives to estimate the VaR, we perform the following experiment. We simulate 1000 series by the GARCH(1, 1) model in (1) and compute the empirical VaR for $\alpha = 0.01$, 0.05 and 0.1. Then for each simulated series, we estimate the VaR by each of the following procedures:

(i) Assuming that returns are $N(0, \hat{s}^2)$, where $\hat{s}^2$ is the sample variance.
(ii) Assuming that returns are a conditionally Gaussian GARCH(1, 1) process.
(iii) Resampling from the raw returns and estimating their density under conditional homoscedasticity.
(iv) Resampling from the returns standardized with the GARCH estimates of the conditional standard deviation and estimating the density conditional on parameter estimates, as proposed by Barone-Adessi *et al.* (1999).
(v) Resampling from the returns standardized with the GARCH estimates of the conditional standard deviation and estimating the density incorporating parameter uncertainty. Notice that, in this case, the procedure used to obtain the density is the one proposed by Pascual *et al.* (2000).

Tables 4 and 5 report the average VaR values across all the replicates when ε_t in (1) is a Student-t distribution with 5 degrees of freedom and a minus χ^2 distribution with 4 degrees of freedom, respectively. In these tables we do not report the average VaR values for the bootstrap procedure based on resampling the standardized returns conditional on parameter estimates because they are very similar to those obtained by incorporating the parameter uncertainty. Pascual *et al.* (2000) show that whether or not the parameter uncertainty is incorporated in intervals for returns does not have any significant effect. With respect to the Student-t distribution, Table 4 shows that, assuming marginal Gaussianity of returns, the VaR values obtained are well under the empirical

Table 4. Monte Carlo results on VaR values using conditionally Gaussian GARCH(1, 1) model and bootstrap methods. Student 5 distribution

Forecast horizon	Sample size		Average VaR values				
			Empirical	Normal	GARCH	Bootstrap 1	Bootstrap 2
1	*T*	Probability					
	300	10%	1.094	1.253	1.212	1.090	1.084
		5%	1.497	1.616	1.563	1.534	1.499
		1%	2.538	2.272	2.197	2.746	2.548
	1000	10%	1.081	1.264	1.202	1.073	1.079
		5%	1.478	1.630	1.549	1.504	1.477
		1%	2.504	2.291	2.178	2.747	2.505
5	*T*	Probability					
	300	10%	1.086	1.253	1.238	1.088	1.080
		5%	1.499	1.616	1.596	1.533	1.506
		1%	2.588	2.272	2.244	2.751	2.637
	1000	10%	1.075	1.264	1.224	1.079	1.073
		5%	1.485	1.630	1.578	1.513	1.483
		1%	2.565	2.291	2.218	2.755	2.592
10	*T*	Probability					
	300	10%	1.083	1.253	1.257	1.085	1.078
		5%	1.508	1.616	1.621	1.530	1.509
		1%	2.656	2.272	2.278	2.768	2.673
	1000	10%	1.074	1.264	1.242	1.075	1.072
		5%	1.495	1.630	1.601	1.513	1.496
		1%	2.630	2.291	2.251	2.769	2.654
20	*T*	Probability					
	300	10%	1.079	1.253	1.278	1.086	1.074
		5%	1.510	1.616	1.648	1.528	1.509
		1%	2.695	2.272	2.318	2.744	2.734
	1000	10%	1.074	1.264	1.264	1.078	1.069
		5%	1.503	1.630	1.629	1.518	1.497
		1%	2.682	2.291	2.291	2.767	2.692

values at the 0.05 and 0.1 probabilities, implying more expected losses than actual. However, at the most common 0.01 probability, the estimated VaR is larger than the empirical value. Therefore, the estimated loss is smaller than the actual. The same conclusions are reached for all horizons and the problem is not solved by increasing the sample size. Although the expected losses are slightly closer to the empirical values, the same results are observed when a conditionally Gaussian GARCH(1, 1) model is assumed. The estimated VaR values are clearly improved when they are computed using bootstrap

Table 5. Monte Carlo results on VaR values using conditionally Gaussian GARCH(1, 1) model and bootstrap methods. χ^2 distribution with 4 degrees of freedom

Forecast horizon	Sample size		Average VaR values				
			Empirical	Normal	GARCH	Bootstrap 1	Bootstrap 2
1	T	Probability					
	300	10%	1.284	1.254	1.220	1.271	1.280
		5%	1.869	1.616	1.572	1.881	1.863
		1%	3.178	2.272	2.211	3.323	3.140
	1000	10%	1.292	1.263	1.238	1.259	1.296
		5%	1.879	1.628	1.597	1.867	1.882
		1%	3.195	2.289	2.245	3.375	3.200
5	T	Probability					
	300	10%	1.278	1.254	1.242	1.272	1.272
		5%	1.873	1.616	1.601	1.887	1.860
		1%	3.247	2.273	2.251	3.321	3.204
	1000	10%	1.283	1.263	1.255	1.260	1.282
		5%	1.881	1.628	1.617	1.863	1.884
		1%	3.256	2.289	2.274	3.368	3.278
10	T	Probability					
	300	10%	1.276	1.254	1.257	1.275	1.265
		5%	1.881	1.616	1.620	1.884	1.862
		1%	3.332	2.273	2.278	3.305	3.260
	1000	10%	1.281	1.263	1.267	1.259	1.270
		5%	1.888	1.628	1.634	1.863	1.869
		1%	3.341	2.289	2.297	3.376	3.330
20	T	Probability					
	300	10%	1.252	1.254	1.274	1.275	1.253
		5%	1.869	1.616	1.642	1.886	1.855
		1%	3.365	2.273	2.309	3.285	3.296
	1000	10%	1.265	1.262	1.281	1.256	1.257
		5%	1.874	1.628	1.651	1.860	1.860
		1%	3.376	2.289	2.321	3.383	3.348

procedures. The expected losses when bootstrapping from the raw returns are generally bigger than the actual losses. However, when the bootstrap is done by resampling from the standardized returns, i.e. the presence of conditional heteroscedasticity is taken into account, the estimated VaR's are remarkably close to the actual values. The bootstrap procedure performs well in estimating the VaR.

With respect to the results for the asymmetric minus χ^2 distribution, Table 5 shows that the VaR values computed assuming a marginal Gaussian distribution of returns are systematically bigger than the actual values. Even larger or similar estimates are obtained when a conditionally Gaussian GARCH(1, 1) model is assumed. Therefore, the actual losses will be, on average, bigger than the losses predicted by both models. This problem is observed for all probabilities, forecast horizons and sample sizes considered. On the other hand, when the bootstrap procedures are applied, the estimated VaR's are closer to the empirical values. Once more, the estimated bootstrap VaR values are more accurate, specially for the shorter horizons, when resampling from the standardized returns.

Finally, we have obtained the VaR of the Pound–Dollar exchange rate using the normality assumption and by the procedure proposed by Pascual *et al*. (2000). For $\alpha = 0.05$ and three steps ahead, the expected loss assuming normality is -0.3865 while the bootstrap VaR is bigger at -0.3724. On the other hand, when $\alpha = 0.01$, under normality the VaR is -0.5435 and the bootstrap is smaller at -0.6108. Notice that the more important differences between both values of the VaR appear when looking at the tails of the distribution that are the focus of interest from the empirical point of view.

Table 6 summarizes the main contributions in this area.

Table 6. Summary of recent bootstrap applications to estimate the distribution of returns.

Author	Model	Boot. proced.	Results
Prediction of returns			
Mig. and Olav. (1999a)	ARMA GARCH(1, 1)	Stand. returns	Conditional on parameter estimates Asymptotic validity
Mig. and Olav. (1999b)	ARMA GARCH(1, 1)	Stand. returns	Conditional on parameter estimates Monte Carlo results
Pascual *et al.* (2000)	GARCH(1, 1)	Stand. returns	Parameter uncertainty. Asympt. and finite samp properties
Gospodinov (2002)	Highly persistent AR	Raw returns	Finite sample properties
		Backward repr.	Forecast interest rates
Prediction of volatilities			
Pascual *et al.* (2000)	GARCH(1, 1)	Stand. returns	Parameter uncertainty. Asympt. and finite samp properties
Value at Risk			
Baro. Ad. *et al.* (1999)	GARCH(1, 1)	Stand. returns	Conditional on parameter estimates
Vlaar (2001)	GARCH(1, 1)	Raw returns	Boot. satisfactory for long series
Baro. Ad. *et al.* (2001)	GARCH(1, 1)	Raw returns Stand. returns	Advantages of Stand. boot. when there are options in portfolio

5. Conclusions

In this paper, we reviewed the literature on the application of bootstrap procedures to the analysis of financial time series. We focused mainly on the papers that have appeared after the review of Maddala and Li (1996). High frequency financial returns are often characterized by a leptokurtic marginal distribution of unknown form. Consequently, bootstrap methods are especially well suited for their analysis. However, when applying these methods to the empirical analysis of financial returns, it should be kept in mind that they were originally designed for *i.i.d.* observations. Although financial returns are usually uncorrelated, they are not independent. Volatility clustering generates correlations between squared observations. Therefore, the bootstrap procedures should be adapted to take into account this dependence. There are two main alternatives. The first is to assume a parametric model for the dynamic evolution of the volatility and to bootstrap from the returns standardized with the estimated standard deviations. Alternatively, it is possible to adopt nonparametric bootstrap methods designed for dependent observations as, for example, the block bootstrap.

There are many empirical applications where bootstrap methods have been adopted to test a great variety of null hypothesis related with financial returns as, for example, the presence of predictable components in the conditional mean, the long-memory property of the conditional variance, or the predictive ability of trading rules. Bootstrap procedures have also been used to obtain the predictive densities of future returns and volatilities, which are fundamental, for example, for VaR models. However, there are very few analytical results on the finite sample and asymptotic properties of the bootstrap procedures when applied to heteroscedastic time series.

Although we focused on the application of bootstrap techniques to the analysis of univariate financial time series, there are also multivariate applications. For example, Engsted and Tanggaard (2001) use bootstrap procedures to compute the bias, standard errors and confidence intervals for the parameters of VAR models fitted to model the Danish stock and bond markets. Kim (2001) also uses bootstrap procedures in the context of VAR models applied to financial series.

Acknowledgements

We are very grateful to E. Ferreira and to the editors, M. McAleer and L. Oxley for useful comments. Financial support from project PB98 0026 from the Spanish Government is gratefully acknowledged by the first author.

References

Baillie, R. T. and Bollerslev, T. (1989) The message in daily exchange rates: a conditional variance tale. *Journal of Business and Economic Statistics*, 7, 297 305.

Baillie, R. T. and Bollerslev, T. (1992) Prediction in dynamic models with time dependent conditional variances. *Journal of Econometrics*, 52, 91 113.

Barone Adesi, G., Giannopoulos, K. and Vosper, L. (1999) VaR without correlations for non linear portfolios. *Journal of Futures Markets*, 19, 583 602.
Barone Adesi, G. and Giannopoulos, K. (2001) Non parametric VaR techniques. Myths and Realities. *Economic Notes*, 30, 167 181.
Berkowitz, J. and Kilian, L. (2000) Recent developments in bootstrapping time series. *Econometric Reviews*, 19, 1 48.
Bessembinder, H. and Chan, K. (1998) Market efficiency and the returns to technical analysis. *Financial Management*, 27, 5 17.
Blake, A. P. and Kapetanios, G. (2000) A radial basis function artificial neural network test for ARCH. *Economics Letters*, 69, 15 23.
Bollerslev, T. (2001) Financial econometrics: Past developments and future challenges. *Journal of Econometrics*, 100, 41 51.
Bollerslev, T. and Mikkelsen, H. O. (1996) Modelling and pricing long memory in stock market volatility. *Journal of Econometrics*, 92, 151 184.
Bollerslev, T. and Mikkelsen, H. O. (1999) Long Term equity anticipation securities and stock market volatility dynamics. *Journal of Econometrics*, 92, 75 99.
Bookstaber, R. M. and McDonald, J. B. (1987) A general distribution for describing security price returns. *Journal of Business*, 60, 401 24.
Brock, W., Lakonishok, J. and LeBaron, B. (1992) Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47, 1731 1764.
Brockman, P. and Chowdhury, M. (1997) Deterministic versus stochastic volatility: implications for option pricing models. *Applied Financial Economics*, 7, 499 505.
Cao, R., Febrero Bande, M., González Manteiga, W., Prada Sánchez, J. M. and García jurado, I. (1997) Saving computing time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Communications in Statistics, Simulation and Computation*, 26, 961 978.
Carriere, J. F. (2000) Non parametric confidence intervals of instantaneous forward rates. *Insurance: Mathematics and Economics*, 26, 193 202.
Chang, P. H. K. and Osler, C. L. (1999) Methodical madness: Technical analysis and the irrationality of exchange rate forecasts. *Economic Journal*, 109, 636 661.
Chatterjee, S. and Pari, R. A. (1990) Bootstrapping the number of factors in the arbitrage pricing theory. *J. Financ. Res.*, 13, 15 21.
Cochrane, J. (1988) How big is the random walk in GNP? *Journal of Political Economy*, 96, 893 920.
Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1 26.
Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, Chapman & Hall, New York.
Eftekhari, B., Pedersen, C. S. and Satchell, S. E. (2000) On the volatility of measures of financial risk: an investigation using returns from European markets. *European Journal of Finance*, 6, 18 38.
Engle, R. (2001) Financial econometrics. A new discipline with new methods. *Journal of Econometrics*, 100, 53 56.
Engsted, T. and Tanggaard, C. (2001) The Danish stock and bond markets: Comovement, return predictability and variance decomposition. *Journal of Empirical Finance*, 8, 243 271.
Ferreti, N. and Romo, J. (1996) Bootstrap tests for unit root AR(1) models. *Biometrika*.
Freedman, D. A. and Peters, S. F. (1984) Bootstrapping a regression equation: Some empirical results. *Journal of the American Statistical Association*, 79, 97 106.
Gallant, A. R., Rossi, P. E. and Tauchen, G. (1993) Nonlinear dynamic structures. *Econometrica*, 61, 871 907.
Garrant, A., Psaradakis, Z. and Sola, M. (2001) An empirical reassessment of target zone nonlinearities. *Journal of International Money and Finance*, 20, 533 548.
Gospodinov, N. (2000) Nonlinearities in short term interest rates, manuscript, Concordia University, Montreal.

Gospodinov, N. (2002) Median unbiased forecasts for highly persistent autoregressive processes. *Journal of Econometrics*, forthcoming.
Groenewold, N. and Fraser, P. (2001a) Test of asset pricing models: how important is the iid normal assumption? *Journal of Empirical Finance*, 8, 427 449.
Groenewold, N. and Fraser, P. (2001b) The sensitivity of tests of asset pricing models to the iid normal assumption: Contemporaneous evidence from the US and UK stock markets. *Journal of Business Finance & Accounting*, 28, 771 798.
Hafner, C. M. and Herwartz, H. (2000) Testing for linear autoregressive dynamics under heteroskedasticity. *Econometrics Journal*, 3, 177 197.
Hall, P. and Wilson, S. R. (1991) Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47, 757 762.
Horowitz, J. L. (2001) The bootstrap and hypothesis tests in Econometrics. *Journal of Econometrics*, 100, 37 40.
Hsie, D. A. and Miller, M. H. (1990) Margin regulation and stock market volatility. *Journal of Finance*, 45, 3 29.
Ikenberry, D., Lakonishok, J. and Vermaelen, T. (1995) Market underreaction to open share repurchases. *Journal of Financial Economics*, 39, 181 208.
Ito, A. (1999) Profits on technical trading rules and time varying expected returns: Evidence from Pacific Basin equity markets. *Pacific Basin Finance Journal*, 7, 283 330.
Kanas, A. (1998) Testing for a unit root in ERM exchange rates in the presence of structural breaks: evidence from the bootstrap. *Applied Economics Letters*, 5, 407 410.
Kho, B. (1996) Time varying risk premia, volatility and technical trading rule profits: Evidence from foreign currency futures markets. *Journal of Financial Economics*, 41, 249 290.
Kim, J. H. (2001) Bootstrap after bootstrap prediction intervals for autoregressive models. *Journal of Business & Economic Statistics*, 19, 117 128.
Korajczyk, R. A. (1985) The pricing of forward contracts for foreign exchange. *Journal of Political Economy*, 93, 346 368.
Kothari, S. P. and Warner, J. B. (1997) Measuring long horizon security price performance. *Journal of Financial Economics*, 43, 301 339.
Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217 1241.
Lamoureux, C. G. and Lastrapes, W. D. (1990) Persistence in variance, structural change and the GARCH models. *Journal of Business and Economic Statistics*, 8, 225 34.
LeBaron, B. (1999) Technical trading rule profitability and foreign exchange intervention. *Journal of International Economics*, 49, 125 143.
Levich, R. M. and Thomas, L. R. (1993) The significance of technical trading rule profits in the foreign exchange market: a bootstrap approach. *Journal of International Money and Finance*, 12, 451 474.
Li, H. and Maddala, G. S. (1996) Bootstrapping time series models. *Econometric Reviews*, 15, 115 158.
Lo, A. and McKinlay, A. C. (1989) The size and power of the variance ratio test in finite samples: A Monte Carlo investigation. *Journal of Econometrics*, 40, 203 238.
Maddala, G. S. and Li, H. (1996) Bootstrap based tests in financial models. In Maddala, G. S. and Rao, C. R. (eds). *Handbook of Statistics*, vol. 14, Elsevier, Amsterdam, 463 488.
Maillet, B. and Michel, T. (2000) Further insights on the puzzle of technical analysis profitability, *European Journal of Finance*, 6, 196 224.
Malliaropulos, D. (1996) Are long horizon stock returns predictable? A bootstrap analysis. *Journal of Business Finance and Accounting*, 23, 93 105.
Malliaropulos, D. and Priestley, R. (1999) Mean reversion in Southeast Asian stock markets. *Journal of Empirical Finance*, 6, 355 384.
Miguel, J. A. and Olave, P. (1999a) Bootstrapping forecast intervals in ARCH models. *TEST*, 8, 345 364.

Miguel, J. A. and Olave, P. (1999b) Forecast intervals in ARCH models: Bootstrap versus parametric methods. *Applied Economics Letters*, 6, 323 27.

Mills, T. (1997) Technical analysis and the London Stock Exchange: Testing trading rules using the FT30. *International Journal of Finance and Economics*, 2, 319 31.

Mun, J. C., Vasconcellos, G. M. and Kish, R. (1999) Test of the contrarian investment strategy. Evidence from the French and German Stock Markets. *International Review of Financial Analysis*, 83, 215 234.

Mun, J. C., Vasconcellos, G. M. and Kish, R. (2000) The contrarian/overreaction hypothesis. An analysis of the US and Canadian Stock Markets. *Global Finance Journal*, 11, 53 72.

Pan, M., Chan, K. C. and Fok, R. C. W. (1997) Do currency futures prices follow random walks? *Journal of Empirical Finance*, 4, 1 15.

Pascual, L., Romo, J. and Ruiz, E. (1998) Bootstrap predictive inference for ARIMA processes, Working paper 98 86, Universidad Carlos III de Madrid, Spain.

Pascual, L., Romo, J. and Ruiz, E. (2000) Forecasting returns and volatilities in GARCH processes using the bootstrap, Working paper 00 68(31), Universidad Carlos III de Madrid, Spain.

Politis, D. N. and Romano, J. P. (1994) The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1303 1313.

Politis, D. N., Romano, J. P., and Wolf, M. (1997) Subsampling for heteroskedastic time series. *Journal of Econometrics*, 81, 281 317.

Shao, J. and Tu, D. (1995) *The jackknife and bootstrap*, Springer, New York.

Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52, 1973 2002.

Sullivan, R., Timmermann, A. and White, H. (1999) Data snooping, technical trading rule performance and the bootstrap, *Journal of Finance*, 46, 1647 1691.

Tauchen, G., Zhang, H. and Liu, M. (1996) Volume, volatility and leverage: A dynamic analysis. *Journal of Econometrics*, 74, 177 208.

Taylor, S. J. (2000) Stock index and price dynamics in the UK and the US: new evidence from a trading rule and statistical analysis. *European Journal of Finance*, 6, 39 69.

Thombs, L. A. and Schucany, W. R. (1990) Bootstrap prediction intervals for autoregressions. *Journal of the American Statistical Association*, 85, 486 492.

Vlaar, P. J. G. (2000) Value at risk models for Dutch bond portfolios. *Journal of Banking & Finance*, 24, 1131 1154.

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

White, H. and Racine, J. (2001) Statistical Inference, the bootstrap and neural network modeling with application to foreign exchange rates. *IEEE Transactions on Neural Networks*, 12, 657 673.

Wu, C. F. J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14, 1261 1295.