



GREQAM

Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n°2006-41

Bootstrap Tests in Bivariate VAR Process with Single Structural Change: Power versus Corrected Size and Empirical Illustration

Jamel JOUINI

September 2006

DT-GREQAM

Bootstrap Tests in Bivariate VAR Process with Single Structural Change: Power versus Corrected Size and Empirical Illustration

Jamel JOUINI*

GREQAM, Université de la Méditerranée, Marseille, France

F.S.E.G.N., E.S.S.A.I.T. and L.E.G.I.

Université 7 Novembre de Carthage, Tunisie

September 14, 2006

Abstract: This paper evaluates the finite-sample performance of single structural change tests based on the asymptotic distribution and bootstrap procedures. In addition to the conventional case of stationary regressors, we consider nonstationary regressors and others characterized by the presence of a break in their structure. While our paper borrows the idea of assessing the performance of structural break tests from an other paper, ours is the first to examine and compare the power of such tests on the basis of corrected size by using graphical methods. We endeavour to see whether some conclusions, obtained for some tests, remain again valid for others based on an other type of processes. Some bootstrap procedures quasi-perfectly correct the size distortions of their asymptotic counterparts and have the same power performance as them on the corrected size basis; property often difficult to obtain. We finally propose a modelling strategy to study the relationship between U.S. interest rates. The results show that such relationship has been altered by a regime-shift located at the beginning of the 1980s.

Key-words: Break date, Bootstrap techniques, Graphical methods, Selection procedures.

JEL Classification: C20.

1. Introduction

In diverse fields of application, many time series might exhibit the phenomenon of structural change which is of capital importance in economics and econometrics. The problem consists in testing the null hypothesis of structural stability against the alternative of instability. Under the alternative hypothesis, we can have one break or more. The literature addressing the first issue is huge. Indeed, Chow (1960) was the first who considered a classic F test allowing to test for single structural

*Jamel JOUINI, Faculté des Sciences Economiques et de Gestion, Campus Universitaire Mrezgua, 8000 Nabeul, Tunisie. Tel.: +216 72 232 205; fax: +216 72 232 318; E-mail: Jamel.Jouini@fsegn.rnu.tn.

change when the break point is known a priori. Quandt (1960) extends the analysis of Chow (1960) to the case of unknown break date. Indeed, he computes a sequence of Chow statistics for all possible break dates contained in a restricted interval since we cannot consider break dates too close to the boundaries of the sample, as there aren't enough observations to identify all the subsample parameters. The estimated break point is the one that maximizes the Chow test. In the same context, some important tests are the "supremum" tests of Andrews (1993), and the related "exponential" and "average" tests of Andrews and Ploberger (1994). They derive the asymptotic distributions which are useful to provide approximations to the finite-sample distributions.¹ These asymptotic distributions have been derived under the assumption of stationary regressors.² This then precludes the presence of structural change in the marginal distribution of the regressors. In this context, Hansen (2000) remedies this deficiency and derives the asymptotic distributions of the test statistics. He finds that these distributions depend on structural change in the regressors.

These asymptotic distributions may be an unreliable guide to finite-sample behavior and as a result the true levels of tests based on asymptotic critical values can be very different from the nominal levels. An alternative approximation is the bootstrap distribution that gives evidence on the adequacy of tests and often provides a tractable way to reduce or eliminate finite-sample distortions of the sizes of statistical tests. In this context, Diebold and Chen (1996) consider a simulation study to obtain better size properties for the "supremum" tests of Andrews (1993) for single structural change using the nonparametric bootstrap procedure. Their study is a simple generalization of the study of Christiano (1992) who provides a bootstrap approximation to the distribution of the test statistic of Quandt (1960) to study the existence of structural change in the dynamics of U.S. aggregate output. In the same context and to obtain good size properties, Hansen (2000) introduces a fixed regressor bootstrap which leads to reasonable size properties in finite-samples. This bootstrap technique treats the regressors as if they are fixed (exogenous) even when they contain lagged dependent variables. It also allows for structural change in the regressors, including change in mean, change in trend function and change in exogenous stochastic trend function, and it finally allows for heteroskedasticity in the errors.

In this paper, through Monte Carlo analysis, we examine and compare the size and power³ properties of a "supremum" test of Andrews (1993), and the related "exponential" and "average" tests of Andrews and Ploberger (1994) to see whether the bootstrap procedures reduce the error in rejection probability⁴ (size distortion) while conserving their power compared to that of their asymptotic counterparts. Since it can be difficult to compare the power of alternative test statistics if they do not have the correct size, this comparative study is based on the graphical methods of

¹Note that Andrews *et al.* (1996) derive the finite-sample distribution for these tests, but the derivation only applies under a model with fixed regressors and normally distributed errors with known variance. In this paper, we explore the finite-sample performance under much more general conditions.

²Note that this cannot be a desirable hypothesis in practice.

³The size of a test is the probability that it rejects the null hypothesis when it is true, while the power of a test is the probability that it rejects the null hypothesis when it is false.

⁴The error in rejection probability of a test is the difference between the actual and nominal probability of rejecting the null hypothesis when it is true.

Davidson and MacKinnon (1998) whose interest is the principle of size correction that allows to show the true power of tests instead of the nominal power, and hence to examine accurately the properties of the tests. We consider, in addition to the conventional case of stationary regressors, nonstationary regressors and others characterized by the presence of regime-shift in their structure. While some single structural change tests enjoy certain optimality properties in many cases, little is known about their comparative power properties under other conditions. In this paper, we attempt to study this point based on a bivariate VAR(2) process. To illustrate empirically the usefulness of all the procedures, we propose a modelling strategy to study the relationship between U.S. interest rates, and to detect an eventual break date in this relationship.⁵

The remainder of the paper is organized as follows. Section 2 presents the basic model and the estimation method allowing to estimate the regression coefficients and the break date. Section 3 defines the "supremum" test of Andrews (1993), and the related "exponential" and "average" tests of Andrews and Ploberger (1994). In section 4, we introduce the bootstrap technique and the relating basic concepts being useful to carry out the simulation experiments. Section 5 defines the graphical methods of Davidson and MacKinnon (1998), namely the P value Plots, P value Discrepancy Plots and Size-Power Curves. Monte Carlo evidence on the size and power performance of both asymptotic and bootstrap approximations is given in section 6. The results indicate that the parametric and nonparametric bootstrap techniques outperform the asymptotic approximation since the bootstrap tests show more accurate size and the same power properties on the corrected size basis as their asymptotic counterparts. Thus, getting the size right and achieving high power are possible tasks. Note that this property is often difficult to obtain when evaluating the performance of tests in terms of size and power. Another feature of substantial importance is that the Hansen's (2000) bootstrap procedures are sometimes inadequate. This then calls these procedures into question. We also study the effect of change of the error distribution on the performance of the testing procedures. Section 7 reports an empirical application using U.S. interest rates to illustrate the usefulness of all the procedures in terms of selection of a break date in the data. The results show that the relationship between the interest rates has been altered by a break located in 1981:5. Some economic explanations are provided to highlight the underlying factors that fostered the change and altered the relationship between the considered series. Section 8 concludes the paper. The different graphs of the simulation study are reported in Appendices A and B, and the graphics of the interest rates are presented in Appendix C.

2. The model and estimation method

Let us consider the following linear regression model:

$$y_t = \begin{cases} x_t' \delta + u_t, & \text{if } 1 \leq t \leq k_0, \\ x_t' (\delta + \theta) + u_t, & \text{if } k_0 < t \leq T, \end{cases} \quad (2.1)$$

⁵It is well known that a proper treatment of parameter changes can be useful in uncovering the underlying factors that fostered the changes and altered a relationship between economic variables.

where y_t is the dependent variable, x_t is a m -vector of regressors, δ is the vector of regression coefficients, θ represents the magnitude of change, k_0 is the break date, and u_t is the disturbance that has a distribution $D(0, \sigma^2)$, with $\sigma^2 < \infty$. The break point k_0 is explicitly treated as unknown and is contained in the interval $[k_1, k_2]$ which is defined to restrict the break date to be bounded from the boundaries of the sample, as there are not enough observations to identify the parameters of the two subsamples. When testing for single structural change, we are not concerned with the distribution of x_t . Thus, the problem of structural change in the model (2.1) does not take into account the nature of the regressors; stationary or nonstationary. We are then interested in testing $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. In other words, we test the null hypothesis of stability against the alternative hypothesis of a single structural change.

Under the null hypothesis, the model (2.1) becomes

$$y_t = x_t' \delta + u_t, \quad t = 1, \dots, T, \quad (2.2)$$

which does not depend on the break date k_0 . The estimation method considered is that based on the least-squares principle (Bai, 1997). Indeed, under the null hypothesis the ordinary least-squares (OLS) estimates are $\hat{\delta}$, the residuals \hat{u}_t , and the estimator of the variance $\hat{\sigma}^2 = (T - m)^{-1} \sum_{t=1}^T \hat{u}_t^2$. Under the alternative hypothesis, the model can be written as

$$y_t = x_t' \delta + x_t' \theta I(t > k_0) + u_t, \quad (2.3)$$

where $I(\cdot)$ is an indicator function that takes the value 1 if its argument is true and 0 otherwise. For a given k , the model (2.3) can be estimated by OLS to get the estimates $(\hat{\delta}_k, \hat{\theta}_k)$, the residuals \hat{u}_{tk} , and the estimator of the variance $\hat{\sigma}_k^2 = (T - 2m)^{-1} \sum_{t=1}^T \hat{u}_{tk}^2$. The break date estimator \hat{k} is such that

$$\hat{k} = \arg \min_k \hat{\sigma}_k^2. \quad (2.4)$$

Finally, the OLS estimates of the regression coefficients and the residuals are respectively $\tilde{\delta} = \hat{\delta}_{\hat{k}}$, $\tilde{\theta} = \hat{\theta}_{\hat{k}}$, and $\tilde{u}_t = \hat{u}_{t\hat{k}}$.

3. The structural change tests

Chow (1960) was the first who considered a classic F test allowing to test for single structural change when the break date is known a priori. He uses the following Wald statistic:

$$F_k = \frac{(T - m) \hat{\sigma}^2 - (T - 2m) \hat{\sigma}_k^2}{\hat{\sigma}_k^2}. \quad (3.1)$$

This test statistic is equivalent to the likelihood ratio statistic when the errors u_t are *i.i.d.* $N(0, \sigma^2)$. In this paper, we are interested in testing the null hypothesis of stability when the true break date k_0 is unknown a priori.

Quandt (1960) extends the analysis of Chow (1960) and proposes a likelihood ratio test equivalent to

$$SupF_T = \sup_{k \in [k_1, k_2]} F_k. \quad (3.2)$$

Note that the break date corresponding to the maximal value of the sequence of the statistics F_k for all the possible break dates contained in the interval $[k_1, k_2]$ constitutes an estimator of the break point.⁶ The asymptotic distribution of the test statistic $SupF_T$ has been derived in Andrews (1993) for stationary regressors. Hansen (2000) derives the asymptotic distribution of the test statistic $SupF_T$ when the marginal distribution of the regressors is affected by a structural change. He finds the same distribution as Andrews (1993), but it depends on

$$\lambda^* = \pi_2^* (1 - \pi_1^*) / [\pi_1^* (1 - \pi_2^*)], \quad (3.3)$$

rather than

$$\lambda = \pi_2 (1 - \pi_1) / [\pi_1 (1 - \pi_2)], \quad (3.4)$$

as found by Andrews (1993). Note that $\pi_1^* = v(\pi_1)$, $\pi_2^* = v(\pi_2)$, and v is a function such that

$$v(r) = [r + \omega(r - \varkappa)I(r \geq \varkappa)] / [1 + \omega(1 - \varkappa)], \quad (3.5)$$

where ω is the magnitude of change in the regressors, $\varkappa = [\epsilon T]$ is the corresponding break date, and π_1 and π_2 are such that $\pi_1 = k_1/T$ and $\pi_2 = k_2/T$. Note that λ^* is a function only of π_1^* and π_2^* , the relative information cumulated in the regressors at the times π_1 and π_2 . From Table 1 of Andrews (1993), the asymptotic critical values for $SupF_T$ are increasing in λ . Thus, if $\lambda^* > \lambda$, then the statistic $SupF_T$ has high tendency to reject the null hypothesis if Andrews' critical values are used. On the other hand, if $\lambda^* < \lambda$, then this statistic tends to reject too infrequently, reducing power.⁷ In our simulation experiments presented in section 6 ($\pi_1 = 0.15$, $\pi_2 = 0.85$, $\omega = 3$ and $\epsilon = 0.5$), $\lambda^* = 49.61 > \lambda = 32.11$ and thus the statistic $SupF_T$ using the Andrews' critical values has severe size distortions as shown by the P value plots in Figure 8.

Andrews and Ploberger (1994) suggest two related tests, namely the "exponential" and "average" tests

$$ExpF_T = \ln \int_k \exp(F_k/2) dw(k), \quad (3.6)$$

$$AveF_T = \int_k F_k dw(k), \quad (3.7)$$

where w is a measure putting weight $1/(k_2 - k_1)$ on each integer k in the interval $[k_1, k_2]$.

⁶This estimator is the same as that obtained by minimizing the sum of squared residuals.

⁷Hansen (2000) plots λ^* as a function of \varkappa for four positive values of ω , $\pi_1 = 0.15$ and $\pi_2 = 0.85$. For each ω , λ^* is maximized at $\epsilon = \pi_1$ and minimized at $\epsilon = \pi_2$. As ω increases, λ^* can become arbitrarily large. The sampling implications is that the $SupF_T$ statistic can have arbitrarily large size distortion.

The asymptotic distributions of all the above-mentioned test statistics are derived under the regressor stationarity hypothesis.⁸ Consequently, the nonstationarity and structural change in the marginal distribution of the regressors affect the asymptotic distributions of these statistics, and then the critical values tabulated based on these distributions are not valid. This then justifies the recourse to the bootstrap as an approximation method of the distributions in such cases.

In this paper, we are interested in comparing the size and power of the asymptotic and bootstrapping versions of the $SupF_T$, $ExpF_T$ and $AveF_T$ tests. For the power, the attention is particularly paid to the location of the break and distance of the alternative from the null. Although, for example, the $ExpF_T$ and $AveF_T$ tests enjoy certain optimality properties in large samples, little is known about their comparative power properties under other conditions.

4. Bootstrap tests

Since the asymptotic tests may, in finite-samples, be biased in the sense that they have empirical sizes that differ from their nominal ones, an alternative approach is necessary. One such distribution-free method is the bootstrap method, introduced by Efron (1979). The bootstrap is based on the idea that the sample is a good representation of the underlying population distribution. As a result, the population distribution can be approximated by drawing samples repeatedly (so-called, resamples) from the original sample.⁹ It is well known that the bootstrap procedure can successfully approximate the finite-sample distribution especially when the statistic is asymptotically pivotal (Horowitz, 1997; and Hall, 1992).

A main feature of bootstrap tests is that, under certain conditions, their actual sizes will converge to the true ones faster than asymptotic tests and at times converge considerably faster. This paper proposes the following procedures for the structural change tests, to construct the bootstrapping distribution¹⁰ under the null hypothesis of stability.¹¹

4.1. Standard procedures

The standard bootstrap procedure is as follows:

1. Compute the asymptotic test statistic in the usual way.
2. Estimate the model under the null hypothesis by the least-squares principle and construct the following bootstrap DGP:

⁸The work of Ploberger and Krämer (1996) is a notable exception since the authors show that in the presence of general trends in the regressors, the asymptotic distribution of the CUSUM test is bounded by the asymptotic distribution in the case of stationary regressors.

⁹See Jeong and Maddala (1993) or Horowitz (1994) for a review of bootstrap methods in econometrics.

¹⁰As we mentioned above the nonstationarity and structural change in the marginal distribution of the regressors affect the asymptotic distributions of the test statistics in complicated ways. An alternative solution consists in using bootstrap methods to yield approximations to the distributions of the considered test statistics.

¹¹The idea of constructing the bootstrap DGP under the null hypothesis is proposed by Beran (1986), and Beran and Srivastava (1985).

$$y_t^* = x_t' \hat{\delta} + u_t^*, \quad (4.1)$$

where $\hat{\delta}$ is the restricted parameter estimate. This DGP is just an element of the model (2.2) using the parameters estimated under the null hypothesis.¹² For the parametric bootstrap, we have $u_t^* \sim \hat{D}(0, \hat{\sigma}^2)$ where $\hat{\sigma}^2 = (T - m)^{-1} \sum_{t=1}^T \hat{u}_t^2$. Note that without the normality assumption, it does not make sense to generate bootstrap errors from the normal distribution. Instead, we want to generate them nonparametrically by resampling with replacement from the residual vector $\check{u} = \{\check{u}_t\}_{t=1}^T$;

$$\check{u}_t = \sqrt{\frac{T}{T - m}} \left(\hat{u}_t - \frac{1}{T} \sum_{s=1}^T \hat{u}_s \right). \quad (4.2)$$

3. Based on these data, compute a bootstrap test statistic in exactly the same way as the real sample was used to compute the asymptotic statistic.

4.2. Hansen's (2000) procedures

The standard bootstrap procedures need the joint modelling of y_t and x_t which require the correct specification of the marginal distribution (including any structural change). To ignore the marginal distribution of the regressors when generating bootstrap data, Hansen (2000) proposes two alternative bootstrap procedures, called the homoskedastic bootstrap and the heteroskedastic bootstrap to yield approximations to the distributions of test statistics in the presence of nonstationarity and structural change in the regressors. These techniques treat the regressors x_t as if they are fixed (exogenous) even when they contain lagged dependent variables, and allow for structural change in the regressors and heteroskedastic errors. The Hansen's bootstrap techniques replicate the correct first-order asymptotic distribution, but it is easy to see that in general the bootstrap does not (in any way) replicate the finite sample distribution of the data or the test statistics. The only exception is when the regressors x_t are strictly exogenous and the errors *i.i.d.* normal, in which case the homoskedastic bootstrap described below will yield exact inference.¹³ Consequently, caution should be applied when interpreting the bootstrap tests.

For the homoskedastic bootstrap, the data y^* are generated by drawing observations from the $N(0, 1)$ distribution.¹⁴ To compute a realization of the bootstrap test statistic, we regress y^* on x_t to get $\hat{\sigma}^{*2}$, and we regress y^* on x_t and $x_t I(t > k)$ to get $\hat{\sigma}_k^{*2}$. The heteroskedastic bootstrap consists in generating bootstrap data y^* according to $e_t^* \tilde{u}_t$, where $e_t^* \sim i.i.d. N(0, 1)$ and \tilde{u}_t are the regression residuals defined in section 2. The bootstrap Wald statistic is then

¹²Note that all the parameter estimates are kept fixed during the bootstrap replications, i.e. they are treated as the true population parameters.

¹³This is a well known property of bootstrap inference with an exactly pivotal statistic. For more details, see Hall (1994).

¹⁴It may be possible to use alternative distributions such as the empirical distribution function (EDF) of the residuals.

$$F_k^* = \frac{(T - m) \hat{\sigma}^{*2} - (T - 2m) \hat{\sigma}_k^{*2}}{\hat{\sigma}_k^{*2}}.$$

The method applies as well for the other tests.

4.3. Computation of bootstrap P value

The bootstrap P value is computed as follows:

1. Draw B bootstrap samples from the above DGPs so as to obtain B bootstrap statistics.
2. The bootstrap P value, say \hat{p}^* , is then estimated by the proportion of bootstrap samples that yield a statistic greater than the asymptotic statistic.¹⁵

We reject the null hypothesis if the bootstrap P value \hat{p}^* is less than α , the significance level of the test.

4.4. The bootstrap when the null hypothesis is false

We know that to obtain an exact finite-sample distribution, it is preferred to construct the bootstrap DGP under the null hypothesis. When this hypothesis is false, a reasonable choice is the pseudo-true null, which is the DGP that satisfies the null hypothesis using pseudo-true values for the unknown parameters obtained under the incorrect assumption that the null hypothesis is true.¹⁶ Consequently, bootstrap sampling when the null hypothesis is false is equivalent to sampling from the null hypothesis model with pseudo-true parameter values; for more details, see Horowitz (1994, 1997).

5. Graphical methods

We use the graphical methods developed in Davidson and MacKinnon (1998), namely the P value plots, the P value discrepancy plots and the size-power curves to obtain evidence on the finite-sample properties of hypothesis testing procedures. Let us consider a Monte Carlo experiment in which N P values $\{\hat{p}_i\}_{i=1}^N$,¹⁷ are generated using a DGP that is a special case of the null hypothesis. All of the graphs we report are based on the EDF of the P values $\{\hat{p}_i\}_{i=1}^N$ which is then defined, at any point α_k in the $(0, 1)$ interval, by

$$\hat{F}(\alpha_k) \equiv \frac{1}{N} \sum_{i=1}^N I(\hat{p}_i \leq \alpha_k). \quad (5.1)$$

¹⁵In the case of fixed exogenous regressors and independent normal errors, the statistic \hat{p}^* can be interpreted as an exact Monte Carlo P value (Dwass, 1957). Dufour and Kiviet (1996) use this motivation in their analysis of the CUSUM test in regression models with fixed exogenous regressors and independent normal errors, and use this property to develop bounds for similar models with an added lagged dependent variable.

¹⁶The pseudo-true values for the unknown parameters are simply the OLS estimates of these parameters obtained under the incorrect assumption that the null hypothesis is true.

¹⁷Let \hat{p}_i denote the asymptotic P value which is computed using the methods of Hansen (1997).

Davidson and MacKinnon suggest the use of the following set of $\{\alpha_k\}_{k=1}^K$.¹⁸

$$\alpha_k = .001, .002, \dots, .010, .015, \dots, .990, .991, \dots, .999 \quad (K = 215). \quad (5.2)$$

5.1. P value plots and P value discrepancy plots

The P value plot that we will use is a plot of $\hat{F}(\alpha_k)$ against α_k . Each of the P values $\{\hat{p}_i\}_{i=1}^N$ should be distributed as $U(0, 1)$ and the resulting graph should be close to the 45 degree line. If the tests show size distortions, then the plots must be over or under the 45 degree line. The P value discrepancy plot is a graph of $(\hat{F}(\alpha_k) - \alpha_k)$ against α_k . The resulting graph must be close to the horizontal axis $y = 0$ if the tests don't show size distortions. These plots convey a lot more information than P value plots for test statistics that are well behaved but some of this information is spurious, simply reflecting experimental randomness.

5.2. Size-power curves

It can be difficult to compare the power of alternative test statistics if they don't have the correct size. The size-power curves are constructed with two EDFs \hat{F} and \hat{F}' . The first EDF is for an experiment in which the DGP satisfies the null hypothesis while for the second EDF, it does not. Note that it is preferable to use the same sequence of random numbers to reduce the experimental error. To generate the size-power curve on a correct size-adjusted basis, we plot the points $(\hat{F}(\alpha_k), \hat{F}'(\alpha_k))$ when α_k describes the $(0, 1)$ interval. Given several tests, if any test shows better power properties than the others, then the size-power curve of this test converges more rapidly to the horizontal axis $y = 1$ than those of the others.

Note that there is an infinite number of DGPs that satisfy the null hypothesis. When the test statistics are pivotal we can use any DGP to generate \hat{F} . However, when the test statistics are nonpivotal the choice of the DGP used to correct the size can matter greatly. Davidson and MacKinnon (1998) argue that a reasonable choice is the pseudo-true null, which is the DGP that satisfies the null hypothesis and is as close as possible, in the sense of the Kullback-Leibler information criterion, to the DGP used to generate \hat{F}' ; see also Horowitz (1994, 1997).

6. Monte Carlo design

Diebold and Chen (1996) compare the size performance of two approximations to the finite-sample distributions of the "supremum" test statistics developed in Andrews (1993), one based on the asymptotic distributions and the other based on the nonparametric bootstrap. Their results show that for an AR(1) process, the bootstrap technique improves on the asymptotic approximation and quasi-perfectly solves the inference problem. In this paper, through Monte Carlo analysis, we endeavour to see whether such conclusions remain again valid for the related "exponential" $ExpF_T$ and "average" $AveF_T$ tests of Andrews and Ploberger (1994) for a bivariate VAR(2) process and

¹⁸Let \hat{p}_i^* be the bootstrap P value. The EDF of the bootstrap P values is obtained by replacing \hat{p}_i by \hat{p}_i^* in (5.1).

by using, in addition, the Hansen's (2000) procedures. We also examine accurately the power properties of all the tests on the corrected size basis using the graphical methods outlined above. Thus, while our paper borrows the idea of evaluating the performance of single structural change tests from the paper of Diebold and Chen (1996), ours is the first to examine and compare the power of such tests on the basis of corrected size by using the size-power curves of Davidson and MacKinnon (1998).

We consider several ways of approximating the finite-sample distributions of the statistics. The first approximation is the Andrews (1993), and Andrews and Ploberger (1994) asymptotic distributions, the second approximation is the parametric bootstrap distribution (bootstrap 0), the third approximation is the nonparametric bootstrap distribution using the EDF of the residuals (4.2) (bootstrap 1),¹⁹ the fourth approximation is the homoskedastic bootstrap (bootstrap 2), and the last one is the heteroskedastic bootstrap (bootstrap 3).

The number of Monte Carlo replications is set at $N = 1000$ and we choose $B = 999$.²⁰ Note that when computing the probabilities of rejecting the null hypothesis when it is true, we don't require a large number of bootstrap simulations and the choice of a small number does not materially affect the simulation results because the experimental errors tend to cancel out across replications.²¹ However, when the null hypothesis is false, we must choose a large number so as to avoid the power loss which can be generated by the tests in many circumstances. For the computation of the test statistics, the possible break dates are contained in the interval $[0.15T, 0.85T]$. P value functions (PVFs) constructed using the asymptotic distributions are plotted. They show the frequencies of rejecting the null hypothesis at the 5% significance level when it is true or false as function of some characteristics of the DGP.

The Monte Carlo experiments were carried out using programs written in Gauss with Gauss pseudo-random number generators.

6.1. Inference under the null hypothesis

The theoretical results of Davidson and MacKinnon (1999) show that the bootstrap tests should perform better. But in many circumstances, in small samples, these tests can show size distortions. Here we show that, using the P value plots, some bootstrap procedures correct the size of the tests and solve the inference problem since the error in rejection probability committed by bootstrap tests is very minimal.

¹⁹We must confirm the fact that the nonparametric bootstrap gives results indistinguishable from those of the parametric bootstrap in some circumstances.

²⁰The number of bootstrap simulations is chosen to satisfy the condition that $\alpha(B + 1)$ is an integer, where α is the level of the bootstrap test (see, Davidson and MacKinnon (2000) for more details). This choice of B deletes all eventual bias of the bootstrap estimation of a P value. Globally, when the bootstrap tests are the sole concern, we recommend to consider large values of B .

²¹From Hall (1986), the error in rejection probability made by a test using bootstrap-based critical values is $O(T^{-(j+1)/2})$ (for some integer $j \geq 1$), regardless of the number of bootstrap samples used to estimate the bootstrap critical values. Consequently, the bootstrap methods provide an asymptotic refinement even with small numbers of bootstrap samples.

To evaluate the performance of all the procedures in finite samples, we consider the following bivariate VAR(2) process:

$$y_t = \delta_1 + \delta_2 y_{t-1} + \delta_3 y_{t-2} + \delta_4 x_{t-1} + \delta_5 x_{t-2} + u_t, \quad 1 \leq t \leq T, \quad (6.1)$$

where $T = 40$ and $u_t \sim i.i.d. N(0, 1)$. The regression coefficients are $\delta_1 = 0.80$, $\delta_2 = 0.90$, $\delta_3 = -0.20$, $\delta_4 = 1.00$, and $\delta_5 = -0.50$. This process satisfies the null hypothesis and is called the no-break model, so the P value plots should be ideally close to the 45 degree line. For a given significance level α , the asymptotic standard error of the empirical size $\hat{\alpha}$ is $(\alpha(1-\alpha)/N)^{1/2}$ which can be estimated by $(\hat{\alpha}(1-\hat{\alpha})/N)^{1/2}$. This last quantity decreases as N increases.

To justify the use of the bootstrap and to highlight the validity of all the approximations, we consider three different distributions for the regressors x_t . Let φ_t be the variable generated from the $N(1, 1)$ distribution;

1. Stationary (*i.i.d.*) regressors: $x_t = \varphi_t$.
2. Nonstationary (mean trend) regressors: $x_t = 2 + 0.05t + e_t$, where $e_t \sim i.i.d. N(0, 1)$.
3. Mean break in the regressors: $x_t = \varphi_t$ if $t \leq \varkappa$ and $x_t = \omega + \varphi_t$ if $t > \varkappa$.

The random regressors x_t are independently generated for each Monte Carlo replication. This means that the presented results are unconditional, rather than conditional on a particular set of regressors. For the first distribution, the asymptotic approximations are not supposed to present size distortions asymptotically since the asymptotic distributions of the test statistics have been derived under the hypothesis of stationary regressors. The other distributions then justify the recourse to the bootstrap to reduce the error in rejection probability committed by the asymptotic tests.

The bootstrap samples must be constructed recursively because of the presence of lagged dependent variables. This is necessary because y_t^* must depend on y_{t-1}^* and y_{t-2}^* , and not on y_{t-1} and y_{t-2} from the observed data. The recursive rule for generating a bootstrap sample is

$$y_t^* = \hat{\delta}_1 + \hat{\delta}_2 y_{t-1}^* + \hat{\delta}_3 y_{t-2}^* + \hat{\delta}_4 x_{t-1} + \hat{\delta}_5 x_{t-2} + u_t^*, \quad 1 \leq t \leq T, \quad (6.2)$$

where $y_0^* = y_0$ and $y_{-1}^* = y_{-1}$. $\hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3, \hat{\delta}_4$ and $\hat{\delta}_5$ are consistent least-squares estimates of $\delta_1, \delta_2, \delta_3, \delta_4$ and δ_5 . For the parametric bootstrap $u_t^* \sim i.i.d. N(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2$ is a consistent estimator of σ^2 whereas for the nonparametric one $u_t^* \sim \text{EDF}(\check{u}_t)$, where \check{u}_t is given by (4.2). As we see every bootstrap sample is conditional on the observed value of y_0 and y_{-1} . This initialization is certainly the most convenient and it may, in many circumstances, be the only method that is feasible.

Figures 1-2 (Appendix A) show the PVFs constructed using the asymptotic distributions and for the case where there is a mean break in the regressors. They show the percentages of rejecting the null hypothesis at the 5% level. In Figure 1, we present the PVFs as function of ω for $\varkappa = T/2 = 20$, and Figure 2 presents the ones as function of \varkappa for $\omega = 3$. We observe that the $AveF_T$ test is more accurate than the other tests especially the $ExpF_T$ test which presents quite severe over-rejection in

the overwhelming majority of cases. Another feature of substantial importance is that the $AveF_T$ test performs quite well in the first case. However, the other tests are slightly more adequate in the second case in comparison with the first case. For the second case and unlike the $AveF_T$ test, the size distortions of the $SupF_T$ and $ExpF_T$ tests are severer when the break date is far from the boundaries of the sample.

P value plots for the case where $\omega = 3$ and $\varkappa = T/2 = 20$ are presented in Figures 6-8 (Appendix B). From the ones of tests based on the conventional asymptotic approximation, we first observe that the error in rejection probability committed by all the tests is severe especially when there is mean trend and mean break in the regressor process since the P value plots are above the 45 degree line, i.e. the actual size is greater than the nominal rejection frequency.²² This is not surprising since the asymptotic distributions of the test statistics have been derived under the hypothesis of stationary regressors. In comparison with the other tests, the $ExpF_T$ test substantially over-reject more than the others, and the $AveF_T$ test is slightly more adequate than the $SupF_T$ test for the last regressor process.

Tests based on the Hansen's bootstrap procedures are poor since they don't allow to correct the size distortions of the asymptotic tests as they also over-reject relative to the nominal size. The heteroskedastic bootstrap $AveF_T$ test has higher tendency to over-reject than the one based on the conventional asymptotic approximation. This calls the usefulness and the validity of this bootstrap procedure into question. The parametric and nonparametric bootstrap procedures give similar results,²³ correct the size distortions of the asymptotic tests and allow to do correct inference since the P value plots are very close to the 45 degree line, i.e. these tests achieve correct size. These procedures are then more accurate than the Hansen's procedures which don't completely solve the inference problem.

We now consider another case in which we set $T = 50$, $\delta_2 = 0.50$, $\delta_3 = -0.75$, and the other characteristics of the DGP are kept fixed. The P value plots are presented in Figures 9-11. We first consider the tests based on the conventional asymptotic distribution. We observe that the $SupF_T$ test presents quite satisfactory results. The size distortions of the $ExpF_T$ test are quite severe especially when there is structural change in the regressor process. The $AveF_T$ test works well for the first two regressor processes since the P value plots are very close to the 45 degree line, i.e. the empirical size is very close to the correct rejection frequency. Thus, the error in rejection probability committed by this test is almost equal to zero ($|\hat{F}(\alpha_k) - \alpha_k| \simeq 0, \forall k$), which implies that the discrepancy P value plots are very close to the horizontal axis $y = 0$.

The bootstrapping versions of the $SupF_T$ test almost provide the same results as the conventional asymptotic distribution for the first two regressor processes. The performance of the heteroskedastic bootstrap is quite significantly poor when there is mean break in the regressors. The bootstrap procedures quasi-perfectly correct the size distortions of the asymptotic $ExpF_T$ test

²²This implies that $\hat{F}(\alpha_k) - \alpha_k > 0$ for all k , which means that the discrepancy P value plots are above the horizontal axis $y = 0$.

²³This confirms the fact that the nonparametric bootstrap gives results indistinguishable from those of the parametric bootstrap.

although there is slight under-rejection by the heteroskedastic bootstrap for the third regressor distribution. For the $AveF_T$ test, except for the heteroskedastic bootstrap which has tendency to over-reject, the other procedures are adequate. The parametric and nonparametric bootstrap procedures seem to be slightly more accurate than the Hansen's procedures.

For the two considered cases, the theoretical results of Davidson and MacKinnon (1999) showing that the bootstrap tests should perform better are well proved for the parametric and nonparametric procedures. But, those showing that in many circumstances and in small samples, the bootstrap tests can show size distortions are well proved for the Hansen's procedures in the overwhelming majority of cases.

From the obtained results of the two DGPs, we observe that there is difference in the results, and this may be explained by the fact that the performance of the tests may depend on how the data were actually generated. Thus, a test that works well against certain DGPs may not work well against others. This can be viewed in our Monte Carlo experiments for some testing procedures.

6.2. Inference when the null hypothesis is false

The anxiety for the size is understandable because a test with uncontrolled size is not useful regardless of the power. However, the fact that the bootstrap is able to eliminate the size distortions makes the test appropriate for a power study. In this case, we are interested in estimating the frequency of rejecting the null hypothesis when it is false. The main result in the analysis of power of the tests is found in Davidson and MacKinnon (2006). This result stipulates that the bootstrap tests with correct sizes can also often be shown to have the same power properties as their asymptotic counterparts. Consequently, if bootstrapping does result in a power loss when B is large, that loss arises simply because bootstrapping corrects the tendency of the asymptotic tests to over-reject.

To compare the performance of the different procedures we advocate the size-power curves constructed on the corrected size basis. We then require to carry out two experiences so as to generate \hat{F} and \hat{F}' using the same sequence of random numbers to reduce the experimental error. Note that for the bootstrap tests, we don't require to specify a priori the break date because the DGPs are constructed under the null hypothesis of no structural change.

Under the alternative hypothesis of structural change, the DGP is as follows:

$$y_t = \begin{cases} \delta_{11} + \delta_{21}y_{t-1} + \delta_{31}y_{t-2} + \delta_{41}x_{t-1} + \delta_{51}x_{t-2} + u_t, & 1 \leq t \leq T/2, \\ \delta_{12} + \delta_{22}y_{t-1} + \delta_{32}y_{t-2} + \delta_{42}x_{t-1} + \delta_{52}x_{t-2} + u_t, & T/2 < t \leq T, \end{cases} \quad (6.3)$$

where $T = 100$ and $u_t \sim i.i.d. N(0, 1)$. The regression coefficients are $\delta_{11} = 0.80$, $\delta_{21} = 0.50$, $\delta_{31} = -0.75$, $\delta_{41} = 1.00$, $\delta_{51} = -0.50$, $\delta_{12} = 0.30$, $\delta_{22} = 0.00$, $\delta_{32} = -0.25$, $\delta_{42} = 0.50$, and $\delta_{52} = 0.00$.²⁴ The parametric and nonparametric bootstrap DGPs are then constructed using the recursive rule given by (6.2) with the modification that in this case the regression coefficient estimates ($\hat{\delta}_1$, $\hat{\delta}_2$, $\hat{\delta}_3$, $\hat{\delta}_4$ and $\hat{\delta}_5$) are obtained under the incorrect assumption that the null hypothesis

²⁴The magnitude of change θ is then equal to 0.50.

is true (see, section 4.5).²⁵ We keep the same regressor processes considered in the study of the size performance.

Figures 3-5 (Appendix A) show the PVFs constructed using the asymptotic distributions for all the regressor processes, respectively. They show the rejection proportions of the null hypothesis at the 5% level as function of the sample size. We first observe that the $ExpF_T$ test is more powerful than the others whatever the value of the sample size, and the $SupF_T$ test is in turn more powerful than the $AveF_T$ test except for the case $T = 100$ and the stationary regressor process. This power discrepancy between all the tests tends to zero as the sample size increases. Another feature of substantial importance is that the probability of rejecting the null hypothesis by all the tests tends to unity as the sample size increases.²⁶ Thus, the tests are *consistent*.

The size-power curves are presented in Figures 12-14 for the case where $\omega = 3$ and $\varkappa = T/2 = 50$. For all the tests, the procedures show perfect power properties for all the regressor processes with very slight superiority for the stationary distribution since the size-power curves are too close to the horizontal axis $y = 1$. The bootstrap tests don't bring out any power loss since the intrinsic size-power curves are confused with those obtained for the asymptotic distribution. Thus, the theory that the test bootstrap power would be very similar to that of the asymptotic test on the basis of size correction is well proved.

The optimality properties of the $ExpF_T$ and $AveF_T$ tests obtained in large samples are maintained under the conditions imposed here. The presence of mean trend and mean break in the marginal distribution of the regressors has not strongly affected the results since almost the same power performance as for the stationary distribution is obtained.

The intuition often suggests that the bootstrap tests are less powerful than their asymptotic counterparts. This is because the formers have much less severe size distortions and then have low tendency to over-reject the null hypothesis unlike the asymptotic tests. The main result of this analysis is that in some cases, many bootstrap tests work very well since they have correct size and high power; property often difficult to obtain when evaluating the performance of the tests in terms of size and power.

We now consider another DGP for which we set $\delta_{21} = 0.90$, $\delta_{31} = -0.20$, $\delta_{22} = 0.40$, $\delta_{32} = 0.30$, and the other characteristics of the DGP are kept fixed. This is done to see whether the power performance of the tests depends on how the data are actually generated. The size-power curves are plotted in Figures 15-17. In comparison with the results of the previous case, the procedures show lower power for all the regressor processes. As for the first DGP, there is very slight superiority for the stationary distribution since the corresponding size-power curves are closer to the horizontal axis $y = 1$ than those of the other regressor processes. The size-power curves of the bootstrap procedures are confused with those obtained for the asymptotic distribution. This proves once again the theory that the power of the bootstrap tests would be very similar to the one of the

²⁵Note that $\hat{\delta}_1$, $\hat{\delta}_2$, $\hat{\delta}_3$, $\hat{\delta}_4$ and $\hat{\delta}_5$ cannot be consistent because when there exists a break date in the DGP while the estimation is done without breaks, the obtained coefficient regression estimates are not consistent because of the misspecification.

²⁶This is obvious since as the intuition suggests, a test statistic having tendency to over-reject (section 6.2) can have high power.

asymptotic tests on the corrected size basis. The results of this experiment show that change of some factors affects the power performance of the tests, but globally it remains satisfactory.

The samples used in the applications are often larger than those used in our Monte Carlo experiments. However, all the procedures are accurate with a sample size of 100 in the cases considered here. Thus, there is not much to gain by carrying out experiments with larger sample sizes. To that effect, we have carried out two other Monte Carlo experiments in which we set $|\theta| = 0.7$ for the first and $T = 150$ for the second, and we kept fixed, for the two experiments, the other characteristics of the first considered DGP.²⁷ The tests become more powerful for a larger magnitude of change. Thus, what is important to improve the power performance of the tests is the magnitude of jump and not the sample size.

6.3. Case of Student error terms

We now study the effect of change of the error distribution on the performance of the testing procedures. To that effect, we simulate leptokurtic data using the Student distribution with 5 degrees of freedom for the error terms. The same other characteristics of all the considered DGPs as well as the number of bootstrap and Monte Carlo replications are kept fixed. We claim that the number of bootstrap simulations, $B = 999$, is sufficient to capture the probability excess in the distribution tails.

For the study of the size performance, the P value plots are reported in Figures 18-20 for the first DGP (model (6.1)) and in Figures 21-23 for the second DGP (model (6.1) with $T = 50$, $\delta_2 = 0.50$ and $\delta_3 = -0.75$). We observe that there is a slight deterioration in the size performance of the asymptotic tests for the two DGPs with respect to the Gaussian case. The correction of the size distortions by the parametric and nonparametric bootstrap procedures is quasi-perfect for the two DGPs as in the Gaussian case. The Hansen's procedures show a slight degradation with respect to the Gaussian case for the $SupF_T$ and $ExpF_T$ tests only for the first DGP. For the study of the power performance, the size-power curves are reported in Figures 24-26 for the first DGP (model (6.3)) and in Figures 27-29 for the second DGP (model (6.3) with $\delta_{21} = 0.90$, $\delta_{31} = -0.20$, $\delta_{22} = 0.40$ and $\delta_{32} = 0.30$). The examination of these curves suggests that for the two DGPs, there is a power loss with respect to the case of normal errors for all the testing procedures and whatever the regressor distribution.

We can then conclude that as the change of the DGP characteristics, the change of the error distribution affects the size and power properties of the asymptotic and bootstrap testing procedures of single structural change.

7. Structural change analysis for U.S. interest rates

The importance of structural change in the U.S. inflation is considered by Ben Aïssa *et al.* (2004), Ben Aïssa and Jouini (2003), and Jouini and Boutahar (2003). Indeed, they use some selection procedures based on information criteria, sequence of tests for multiple breaks and a test similar

²⁷The results of these experiments are not reported here and are available upon request from the author.

to that based on Kolmogorov-Smirnov statistic applied to the evolutionary spectrum to determine the number of changes and their locations in the monthly U.S. inflation series. Both they find significant results since the breaks coincide with important facts and economic events such that the two Oil-Price Shocks and the major events in the International Monetary System.

The theory of the term structure of interest rates suggests that, if interest rates can be characterized as $I(1)$ processes, then they should be cointegrated. Stock and Watson (1988) carry out cointegration tests for three monthly U.S. interest rates covering the period 1960:1-1979:8,²⁸ and find evidence of two cointegrating vectors. In the same context, Hansen (1992) tests the hypothesis that relationships between U.S. interest rates are stable over the period 1960:1-1990:3. Indeed, he reports results of two fully modified regressions indicating that the change in the Federal Reserve's operating procedures in 1979 altered the relationship between some interest rates. He finds that this regime-shift only appears to have affected the relationship between the treasury-bill rates and the federal funds rate but not the relationship between the treasury-bill rates of different maturities.

7.1. The modelling strategy and the results

In the same context of studying the relationship between U.S. interest rates, we propose a different modelling strategy and apply the estimation and testing procedures discussed above using the monthly 90-day treasury-bill rate and the federal funds rate—plotted in Figure 30 (Appendix C)—covering the period 1957:1-2002:10 (yielding 550 observations) and obtained from the St. Louis Reserve Federal Bank database. Thus, we adopt a bivariate VAR(3) process:

$$y_t = \delta_1 + \delta_2 y_{t-1} + \delta_3 y_{t-2} + \delta_4 y_{t-3} + \delta_5 x_{t-1} + \delta_6 x_{t-2} + \delta_7 x_{t-3} + u_t, \quad (7.1)$$

where y_t is the 90-day treasury-bill rate and x_t is the federal funds rate.

The number of bootstrap replications is set at $B = 9999$, a large value that we would recommend using in practice to improve the inference accuracy. The results of the different testing procedures are as follows:

	Asymptotic P value	Bootstrap 0 P value	Bootstrap 1 P value	Bootstrap 2 P value	Bootstrap 3 P value
$SupF_T$	0.000	0.000	0.004	0.000	0.666
$ExpF_T$	0.000	0.001	0.008	0.001	0.703
$AveF_T$	0.165	0.169	0.156	0.132	0.376

The $AveF_T$ test fails to reject the null hypothesis of parameter stability for all the procedures even at the 10% significance level. For the other tests, only the heteroskedastic bootstrap does not lead to reject the null hypothesis. The fact that this procedure does not allow a rejection of the null hypothesis is not surprising because the reported simulation results show that this procedure has sometimes tendency to under-reject the null hypothesis. However, the rejection of the null by

²⁸They used data covering this period, presumably to exclude an eventual regime-shift in the term structure due to the change in the Federal Reserve's operating procedures in 1979.

the other procedures is obvious since the simulation results show that these procedures reject the null hypothesis with high proportions when there is a break in the data.

The relationship between the two series is then unstable and the break date thus selected by the $SupF_T$ test is located in 1981:5. Looking at the graphs of the series in Figure 30, we confirm the choice of this break since we observe that these series may be affected by structural breaks as there is a turbulence period at the beginning of the 1980s. The fact to find this regime-shift confirms the results of Hansen (1992) who found that the relationship between the 90-day treasury-bill rate and the federal funds rate has been affected by a regime-shift in 1980. Thus, we see that with two different modelling strategies and two different techniques we find the same regime-shift though the sample sizes are not the same. This gives us more information about the fact that the regime-shift is strongly evident and can be easily detected. The corresponding 95% confidence interval of the break date computed using the asymptotic distribution derived by Bai and Perron (1998) covers the quite large period stretching from 1979:3 to 1984:2, indicating that the break date is imprecisely estimated.

The break date 1981:5 may be explained by the following events and facts. In 1981, the United States of America have known the disinflation wave and which is known as "*disinflation of Volcker*" concerning a new control of the monetary policy. Indeed, before the year 1970, the repercussions of the real shocks on the economy were regarded as tiny. The considerable fluctuations of oil prices modified this point of view and led the Federal Reserve of the United States to reconsider the implementation of the monetary policy. In 1977, unemployment was running at 7%. Total labour costs to employers however, rose both in January 1978 and in 1979 because of increases in payroll taxes and the minimum wage, steps taken by administrative decision not market process but widely regarded as constituting almost as powerful an inflationary influence as were cash wage increases themselves. By 1979, spare capacity was becoming limited and production costs were rising at an annual rate of some 8%. During 1979-1980, moreover, the price of oil and oil-related products also rose sharply under the spur of supply restrictions and rising world demand. In some ways the restrictive policies initially adopted to counter the stagflation of 1979-1980, like some of the main causes at work, followed lines similar to those observed in 1974-1975. At a deeper level, however, the two sets of policies differed substantially. Couched primarily in terms of monetary aggregates, the post-1979 policies seemed to embody a willingness to permit, and even to bring about, greater rises in interest rates than did the policies five years before, and the Federal Reserve's discount rate at once edged up to 12%. After the election of Ronald Reagan in November 1980, the discount rates were raised to 14% in May 1981. The tax proposals of the president were totally different than the procedures taken in 1974-1975. President Reagan projected great tax cuts. The total effects of these last measures should lead to the restoration of the federal budget balance by 1984-1985. The year 1981 was marked by the rise of the tax deficit whereas the interest rates were on a rather high level of 9% and unemployment kept the margin of 8%. Actuated by the bad news coming from Mexico which had borrowed heavily and which stipulated that it doubts to refund its loans to the United States at the limits envisaged, all the American economy embarked on a productivity growth phase of the supported by a decrease of oil prices and the reduction of the inflation rate

of the with approximately 5%, but also by the deceleration of the raising of wages whereas the dollar, under the influence of the raised interest rates, attracted the flow of foreign capital, which appreciated approximately 30%.

8. Conclusion

The paper has presented a detailed evaluation of the performance of some tests for single structural change based on the graphical methods of Davidson and MacKinnon (1998) whose interest is the principle of size correction that allows to show the true power of tests instead of the nominal power. The bootstrap methods improve the performance of the tests. The exception is that the Hansen's heteroskedastic bootstrap has sometimes tendency to under-reject. Moreover, this procedure seems less precise than the asymptotic approximation for the $AveF_T$ test since the error in rejection probability committed by this bootstrapping version of the test is severer. The parametric and nonparametric bootstrap procedures are slightly more accurate than the hansen's procedures. We found that on the basis of the corrected size, the bootstrap procedures have the same power properties than their asymptotic counterparts. The empirical illustration highlights the practical importance of the simulation results and indicates that the relationship between the interest rates has been altered by a regime-shift located at the beginning of the 1980s.

Acknowledgements

The author is grateful to Emmanuel FLACHAIRE and the participants in the "*Journées Doctorales de l'ADRES*", Paris, February 2005 for useful comments and remarks.

Appendix A: P value Functions

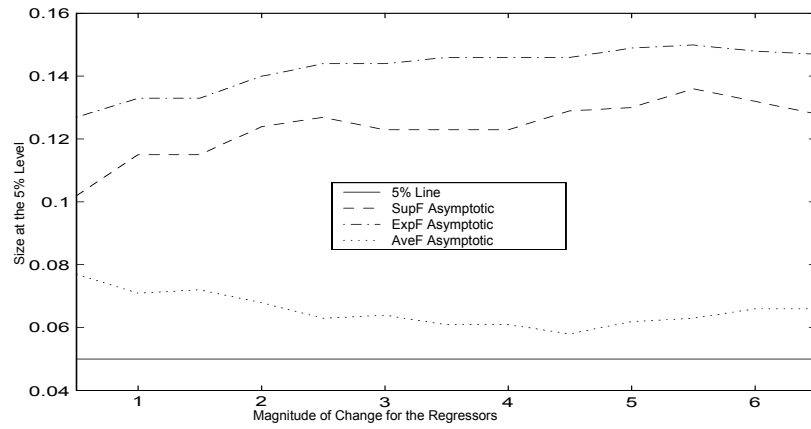


Figure 1. Size of the Asymptotic Tests as Function of the Jump Size of the Regressors

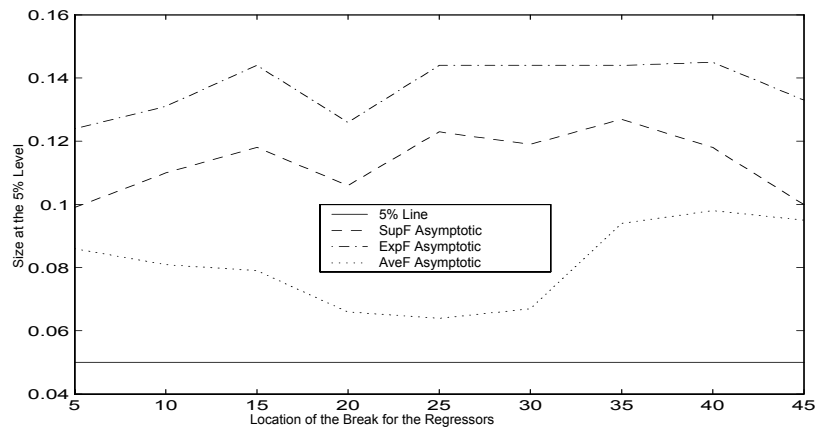


Figure 2. Size of the Asymptotic Tests as Function of the Location of the Break for the Regressors

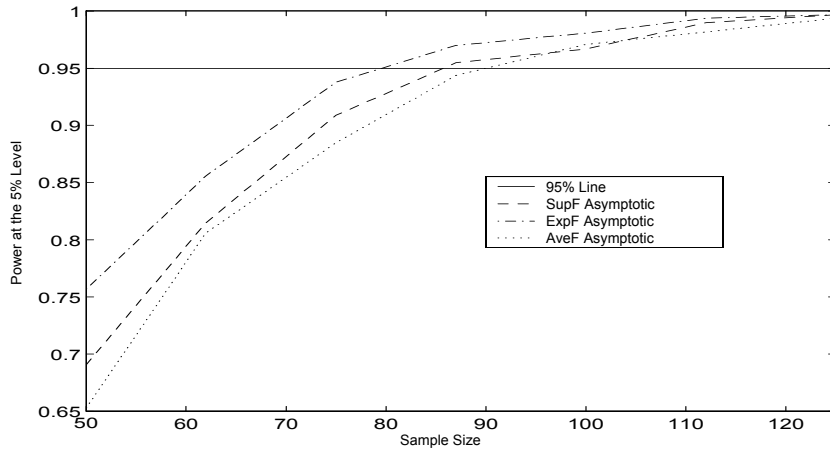


Figure 3. Power of the Asymptotic Tests as Function of the Sample Size for the Stationary Regressor Process

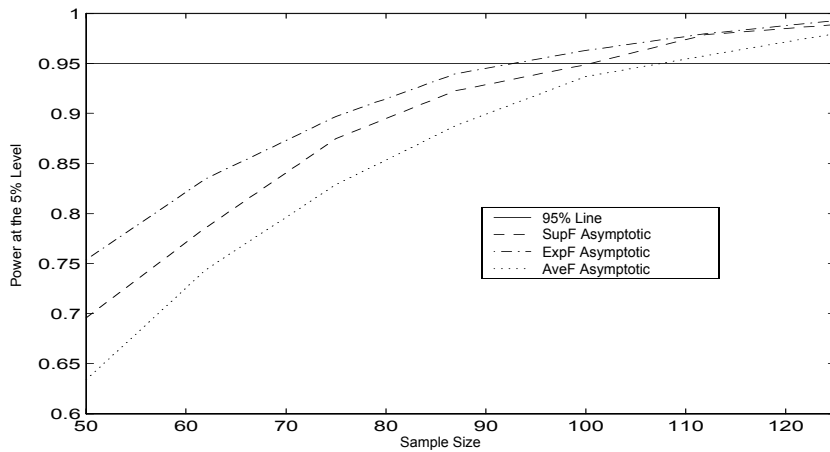


Figure 4. Power of the Asymptotic Tests as Function of the Sample Size for the Nonstationary Regressor Process

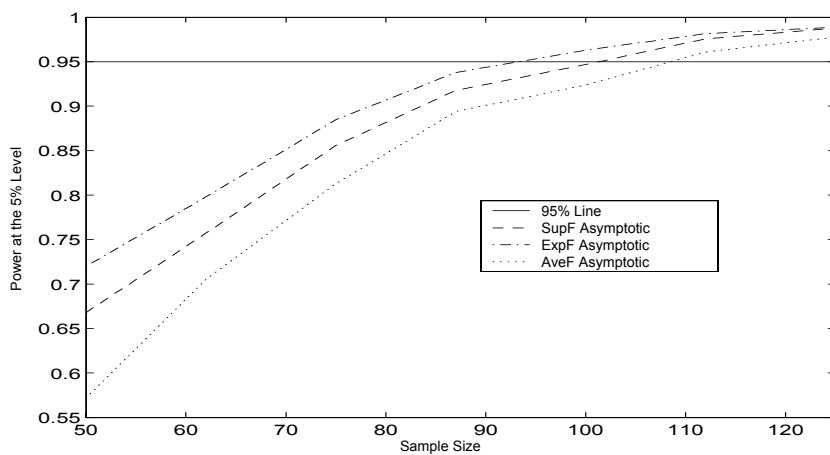


Figure 5. Power of the Asymptotic Tests as Function of the Sample Size for the Third Regressor Process

Appendix B: P value Plots and Size-Power Curves

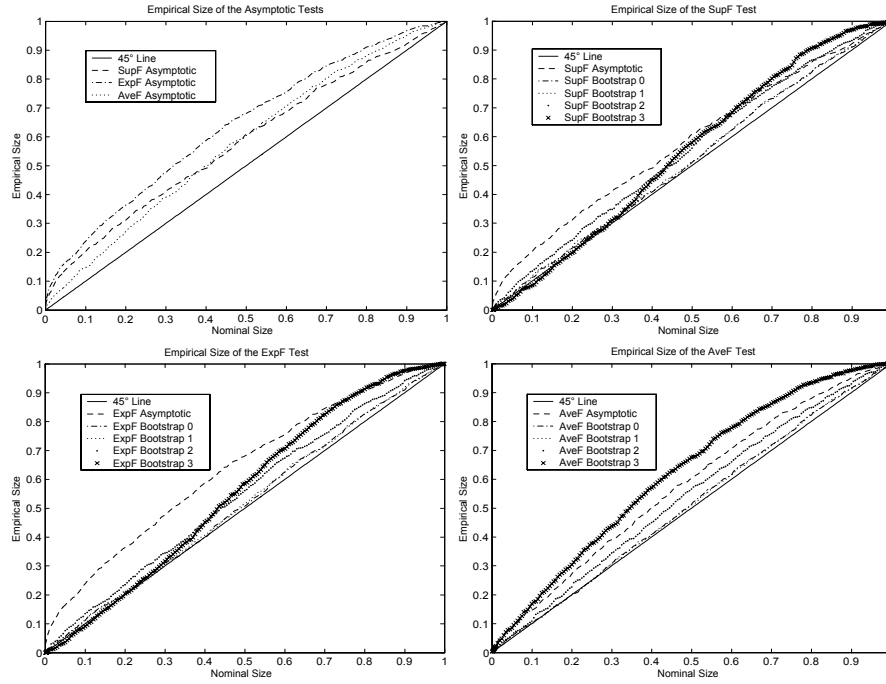


Figure 6. Empirical Size of the Tests for Stationary Regressor Process

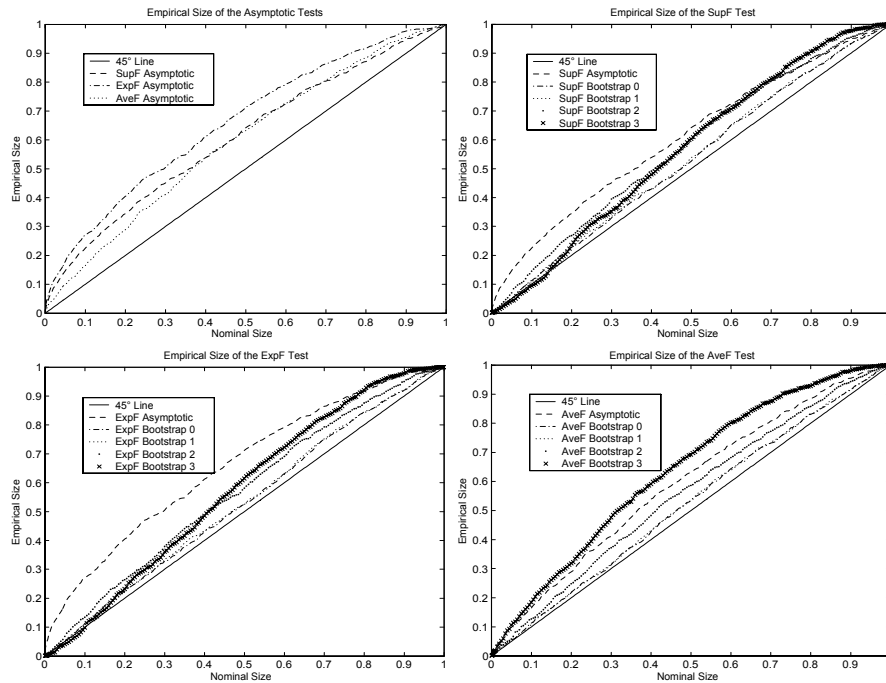


Figure 7. Empirical Size of the Tests for Nonstationary Regressor Process

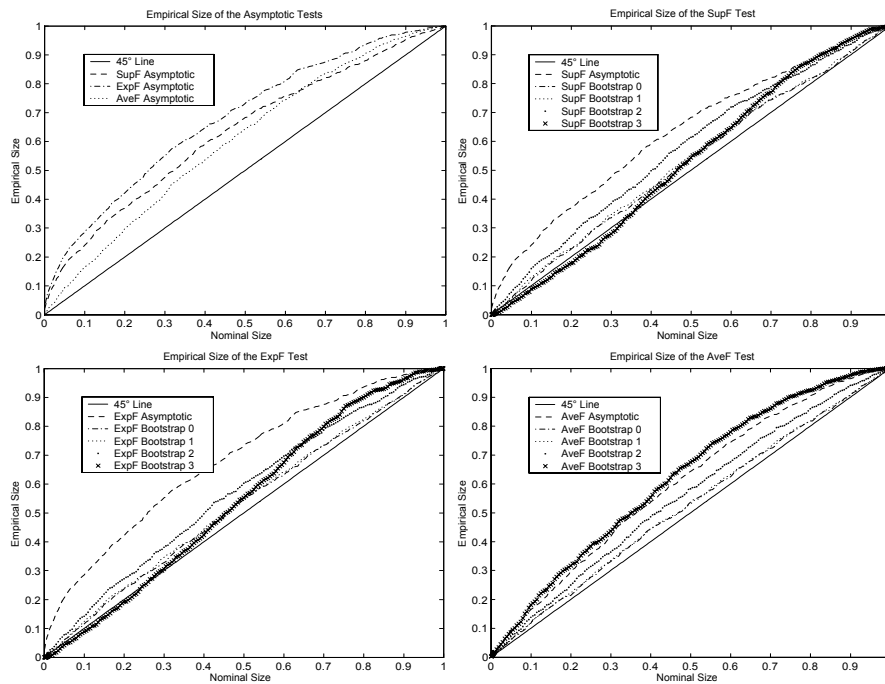


Figure 8. Empirical Size of the Tests when there is Mean Break in the Regressor Process

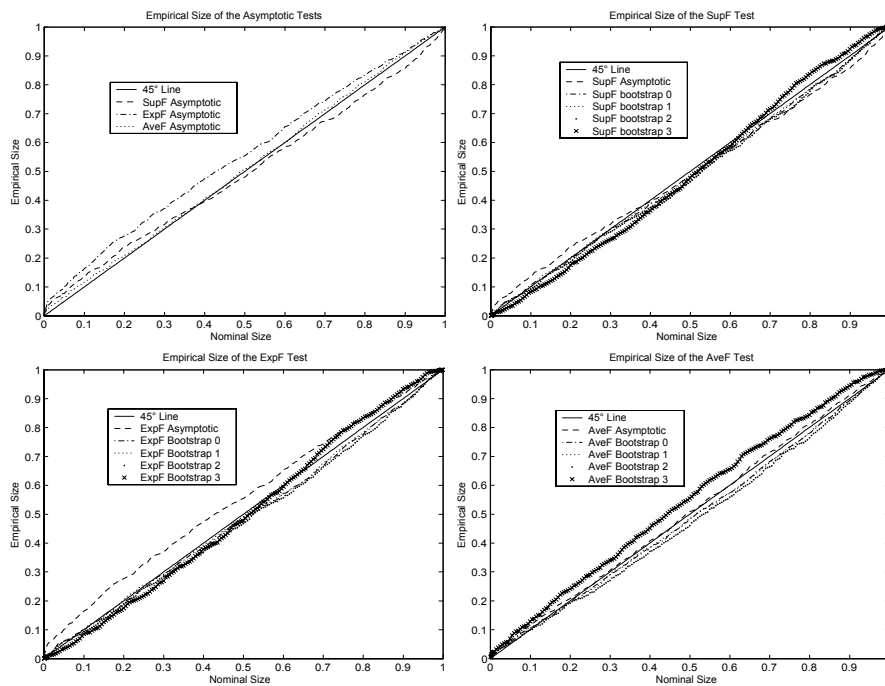


Figure 9. Empirical Size of the Tests for Stationary Regressor Process

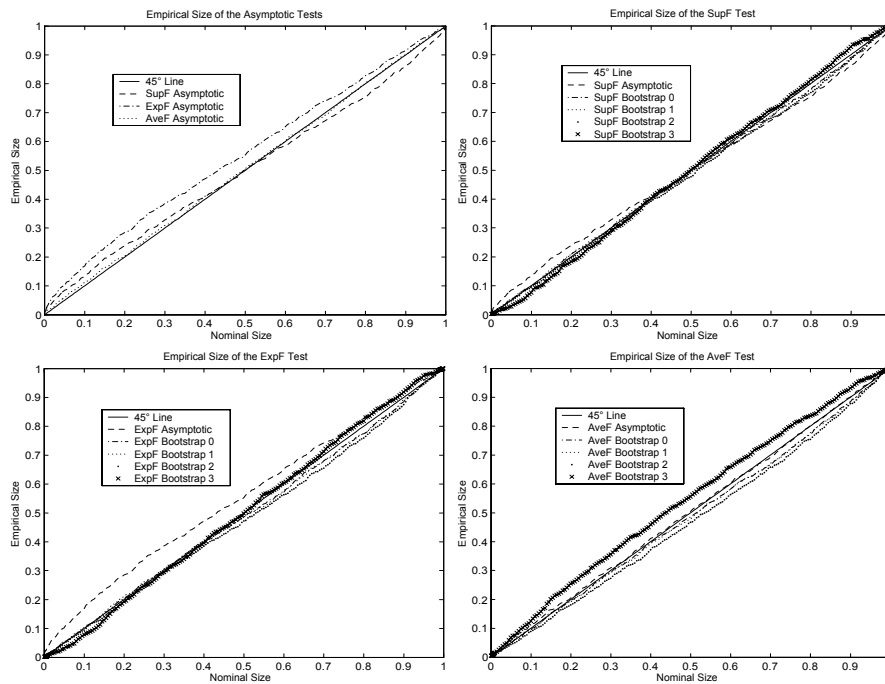


Figure 10. Empirical Size of the Tests for Nonstationary Regressor Process

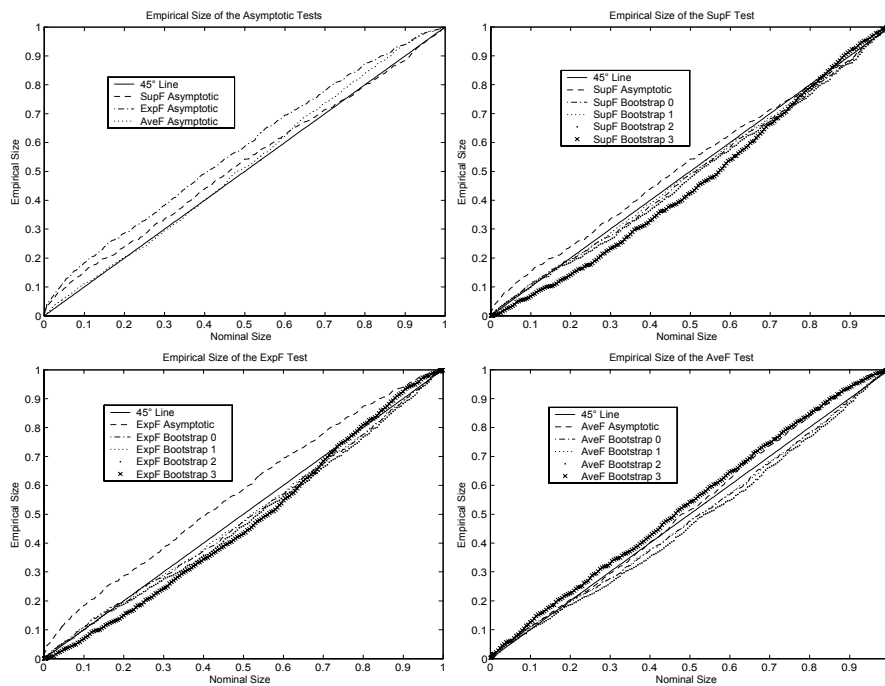


Figure 11. Empirical Size of the Tests when there is Mean Break in the Regressor Process

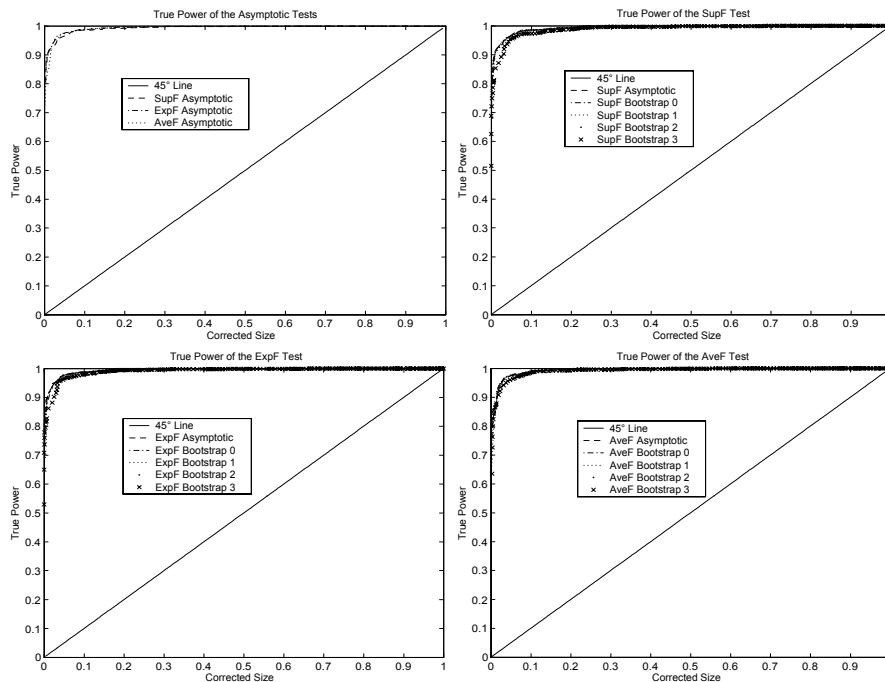


Figure 12. True Power of the Tests for Stationary Regressor Process

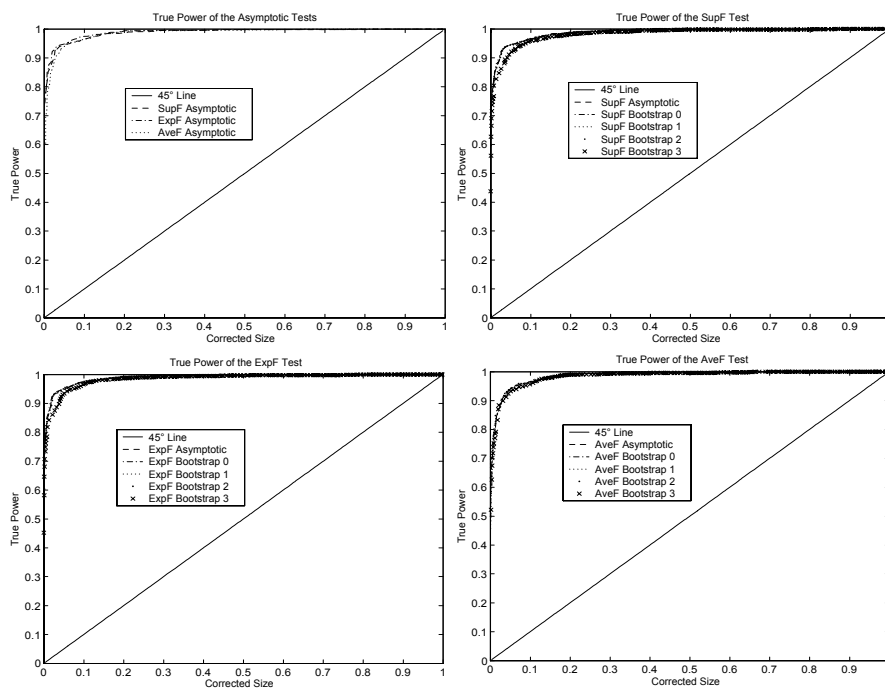


Figure 13. True Power of the Tests for Nonstationary Regressor Process

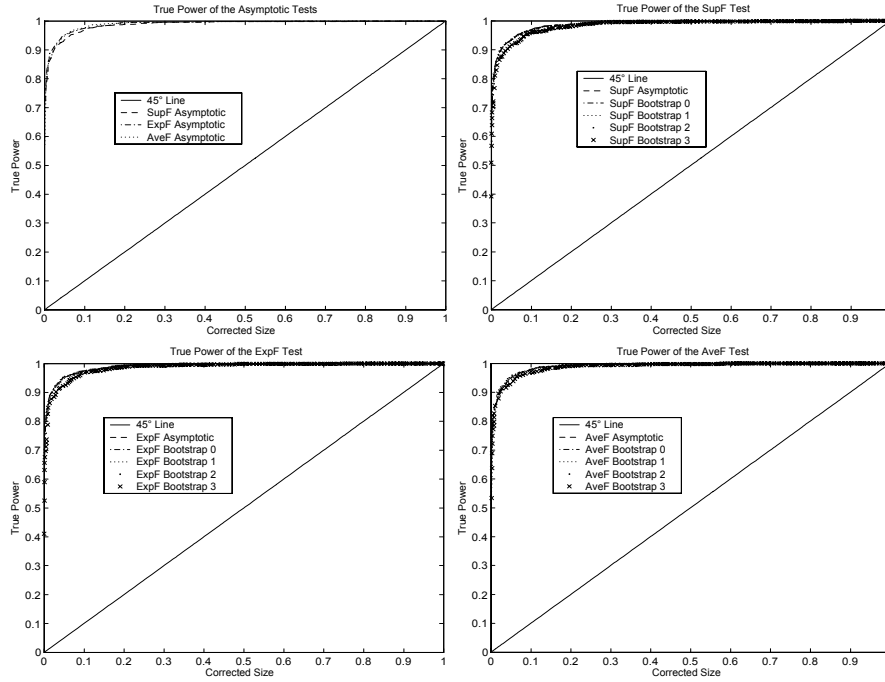


Figure 14. True Power of the Tests when there is Mean Break in the Regressor Process

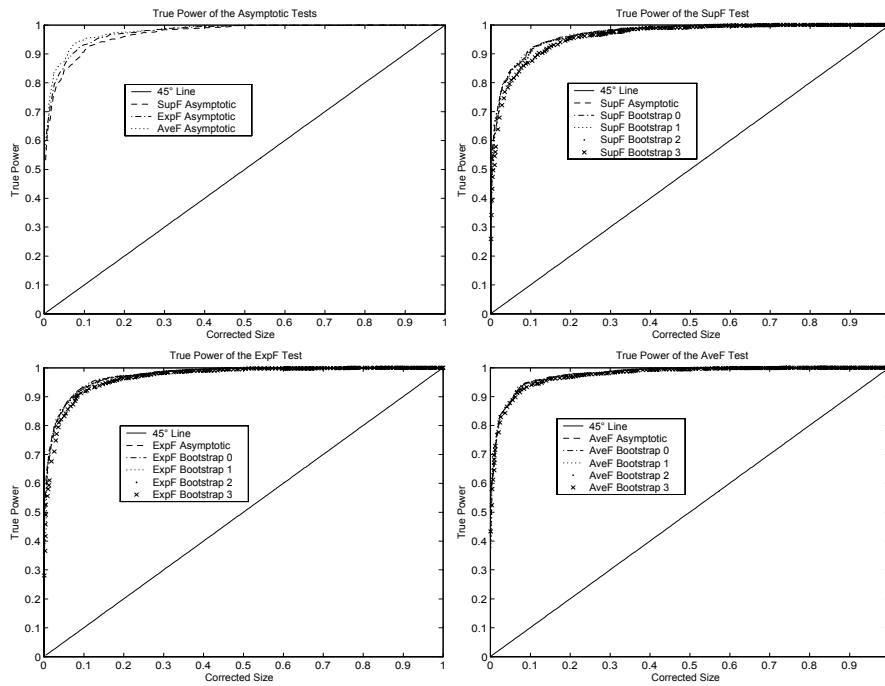


Figure 15. True Power of the Tests for Stationary Regressor Process

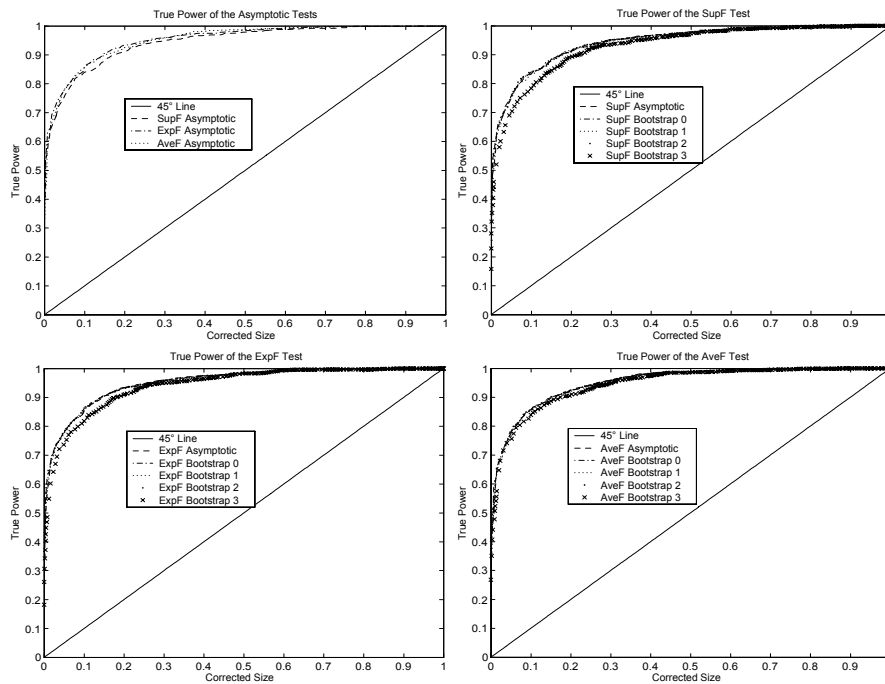


Figure 16. True Power of the Tests for Nonstationary Regressor Process

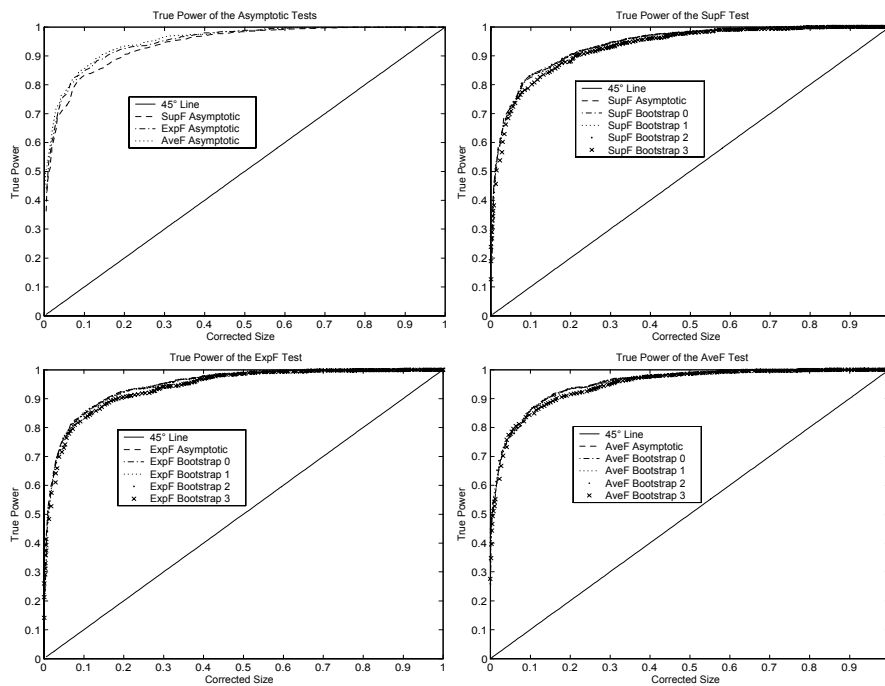


Figure 17. True Power of the Tests when there is Mean Break in the Regressor Process

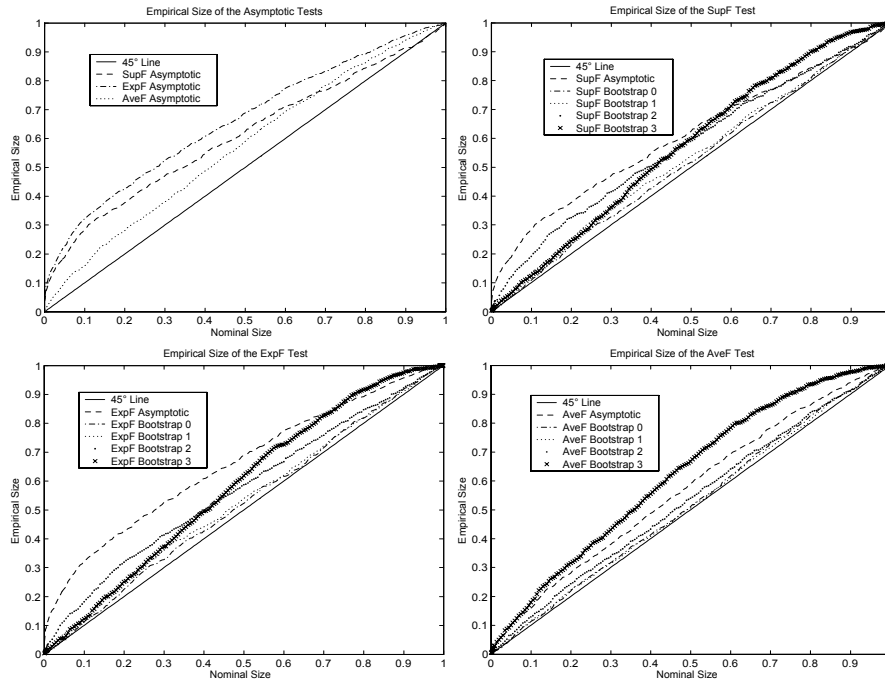


Figure 18. Empirical Size of the Tests for Stationary Regressor Process

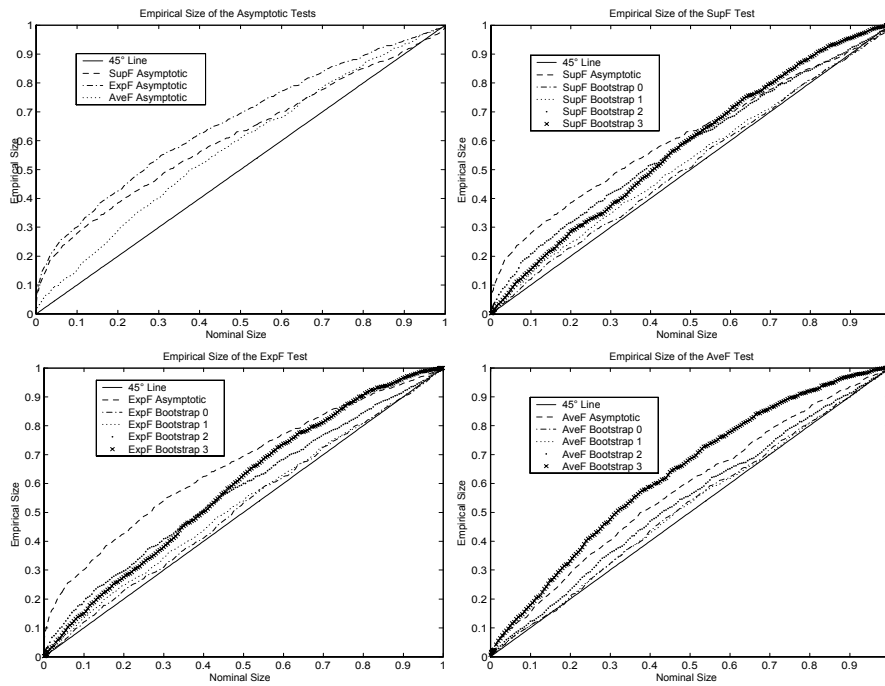


Figure 19. Empirical Size of the Tests for Nonstationary Regressor Process

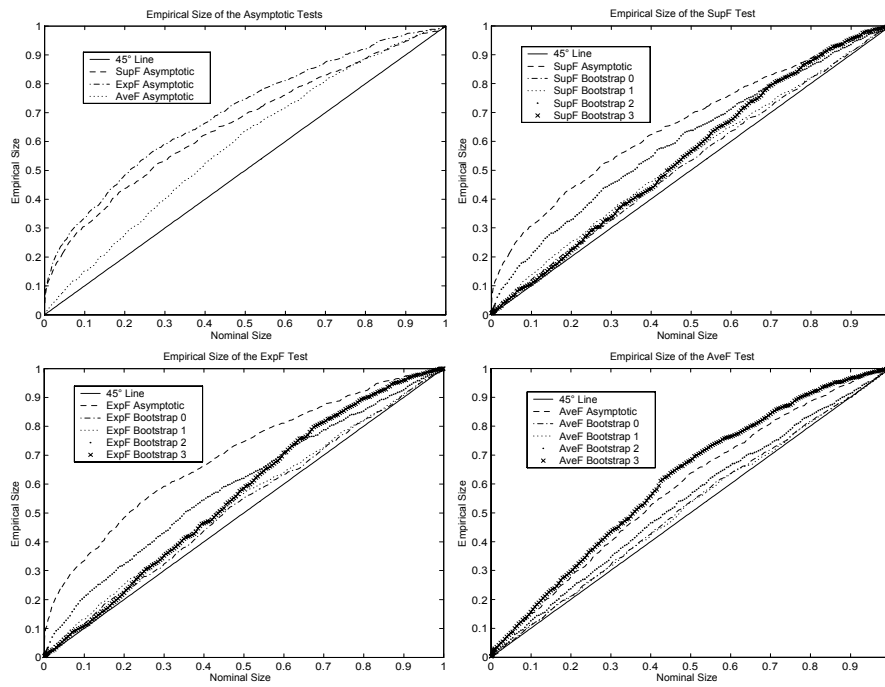


Figure 20. Empirical Size of the Tests when there is Mean Break in the Regressor Process

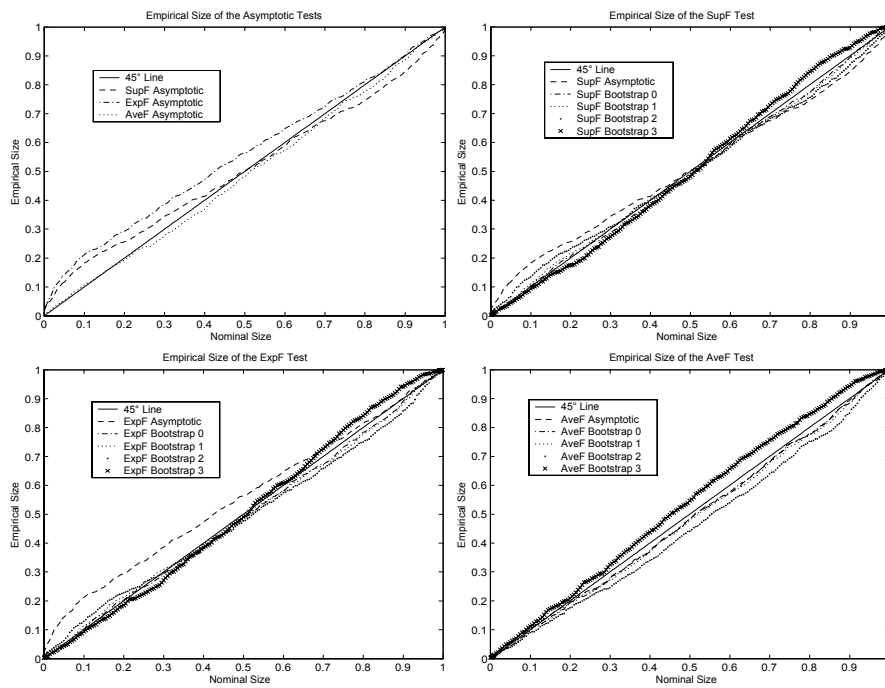


Figure 21. Empirical Size of the Tests for Stationary Regressor Process

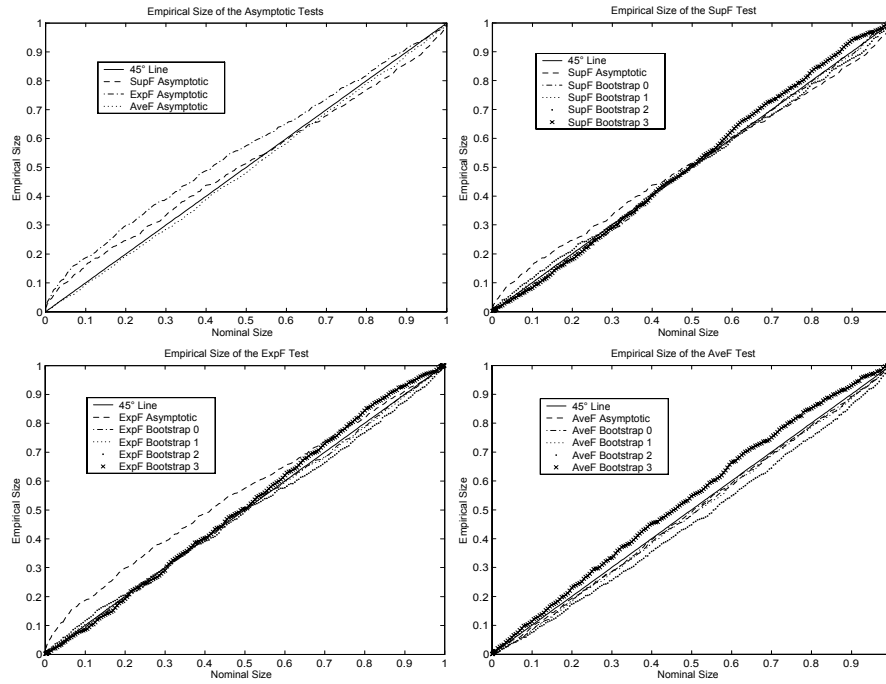


Figure 22. Empirical Size of the Tests for Nonstationary Regressor Process

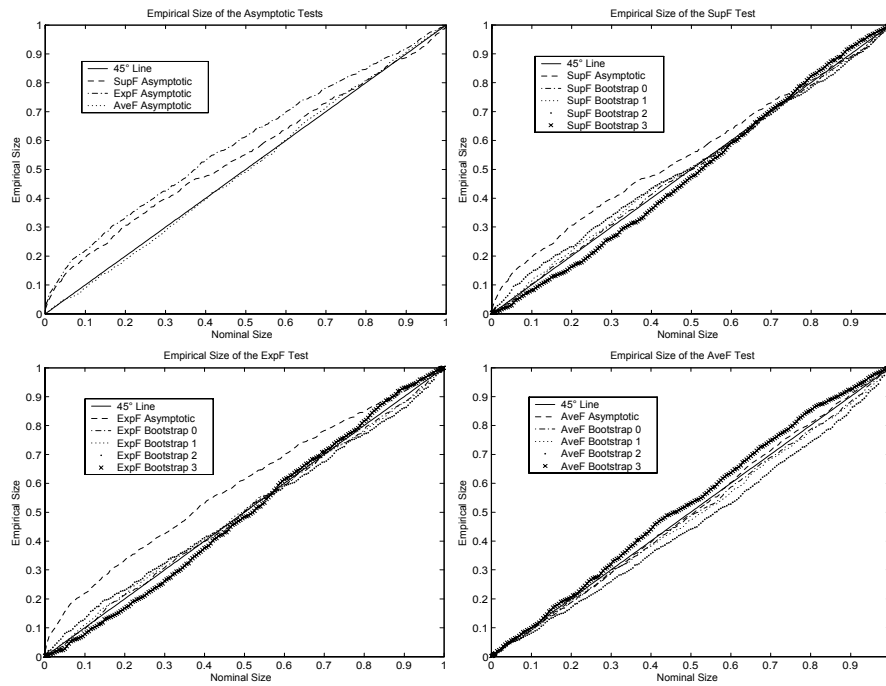


Figure 23. Empirical Size of the Tests when there is Mean Break in the Regressor Process

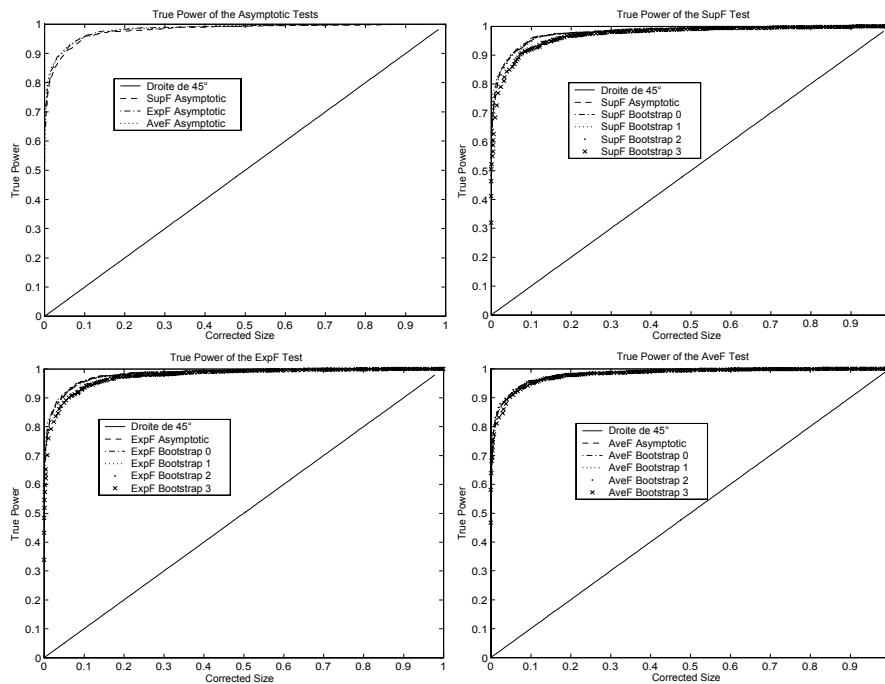


Figure 24. Empirical Size of the Tests for Stationary Regressor Process

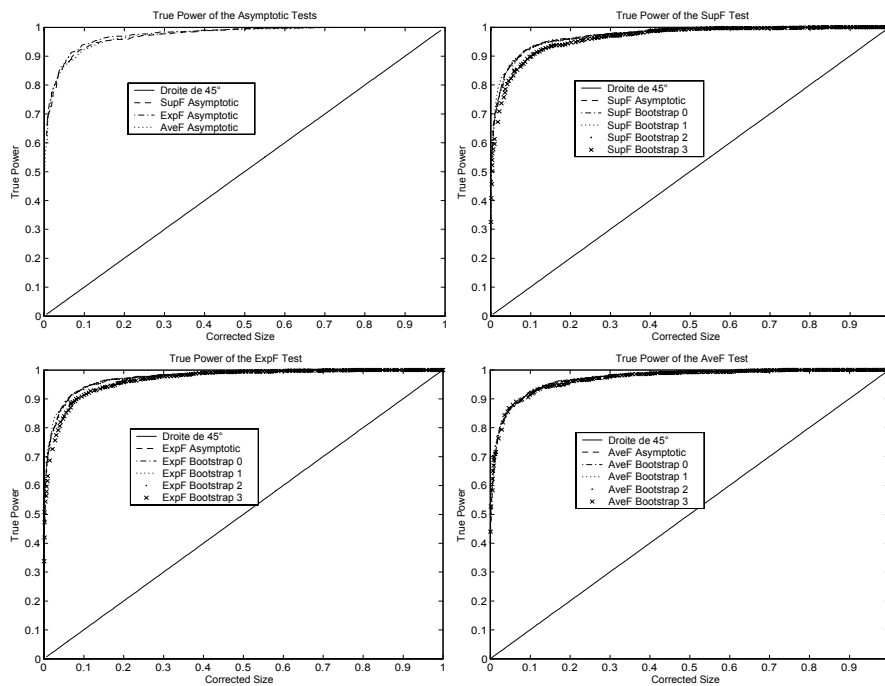


Figure 25. Empirical Size of the Tests for Nonstationary Regressor Process

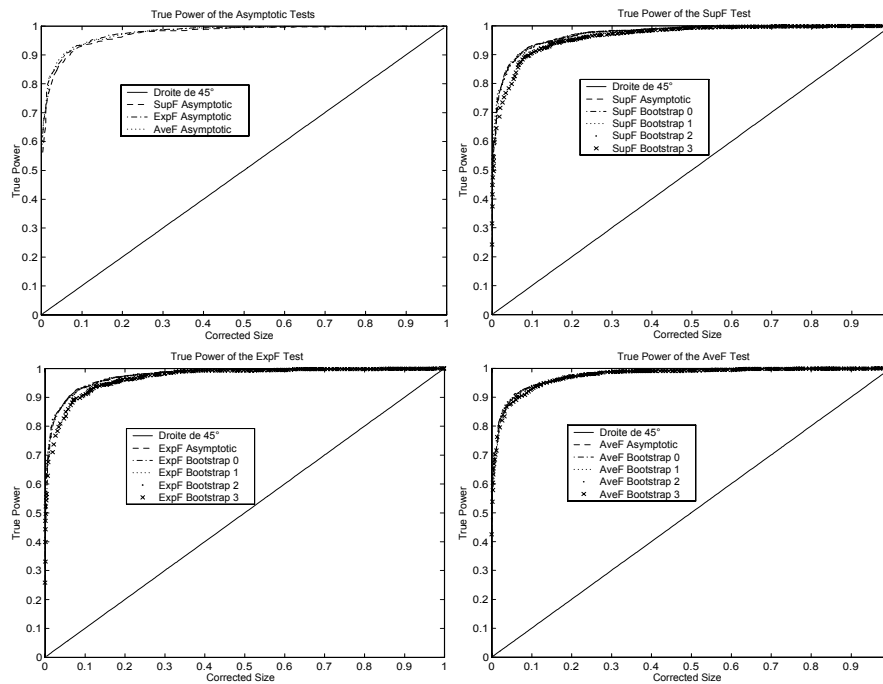


Figure 26. Empirical Size of the Tests when there is Mean Break in the Regressor Process

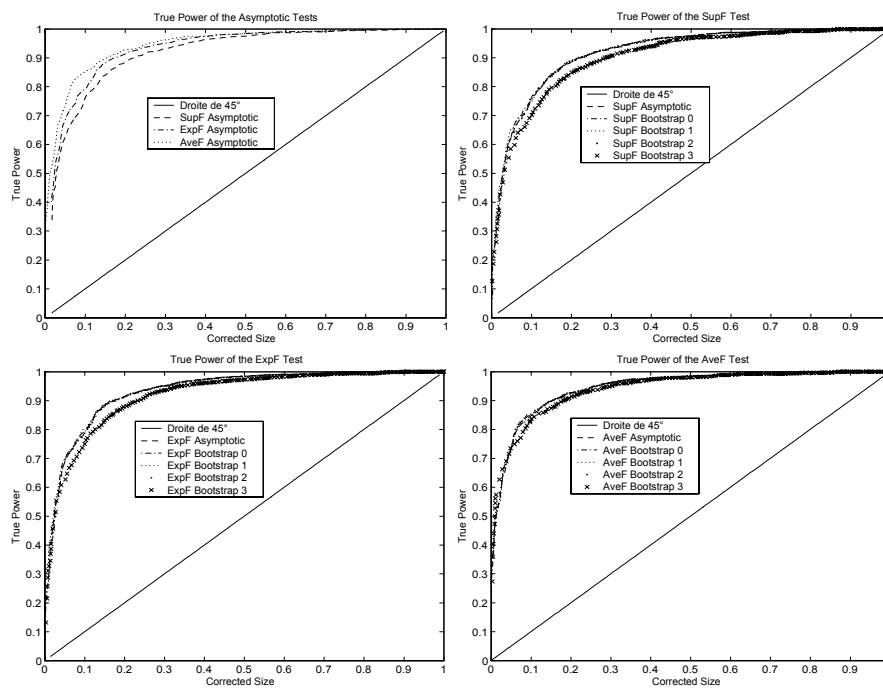


Figure 27. Empirical Size of the Tests for Stationary Regressor Process

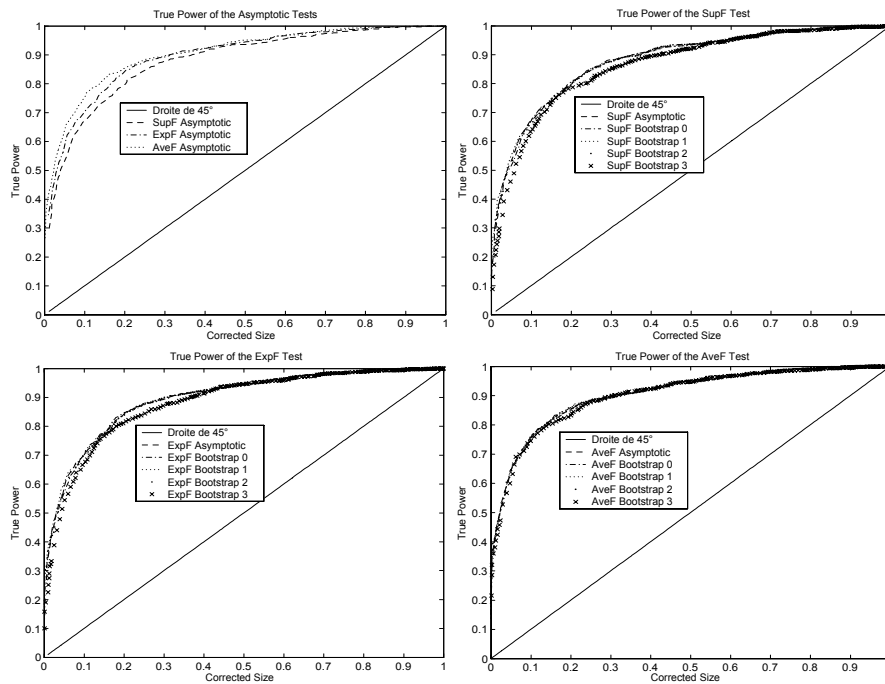


Figure 28. Empirical Size of the Tests for Nonstationary Regressor Process

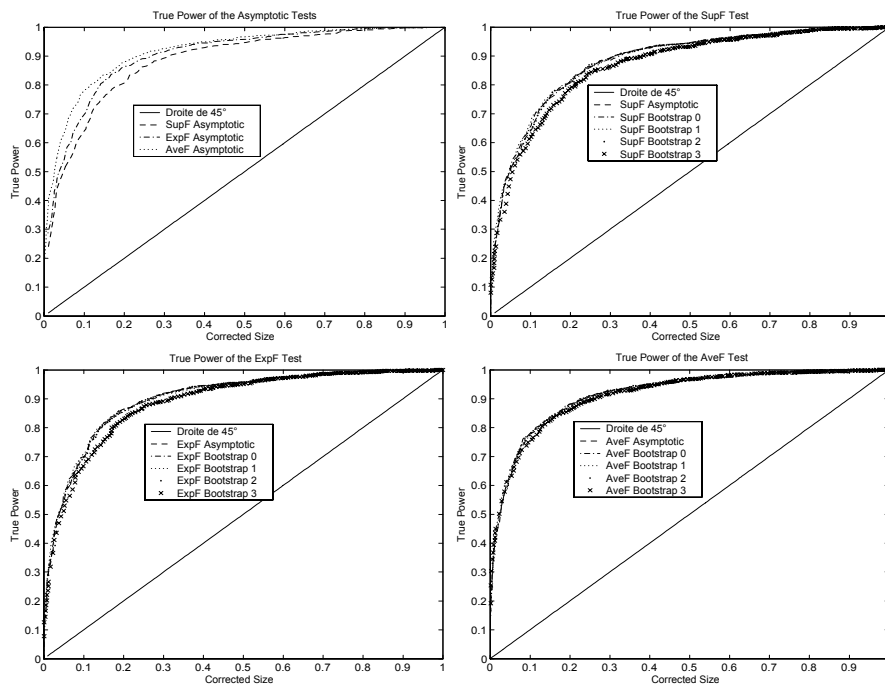


Figure 29. Empirical Size of the Tests when there is Mean Break in the Regressor Process

Appendix C: Graphics of the Interest Rates

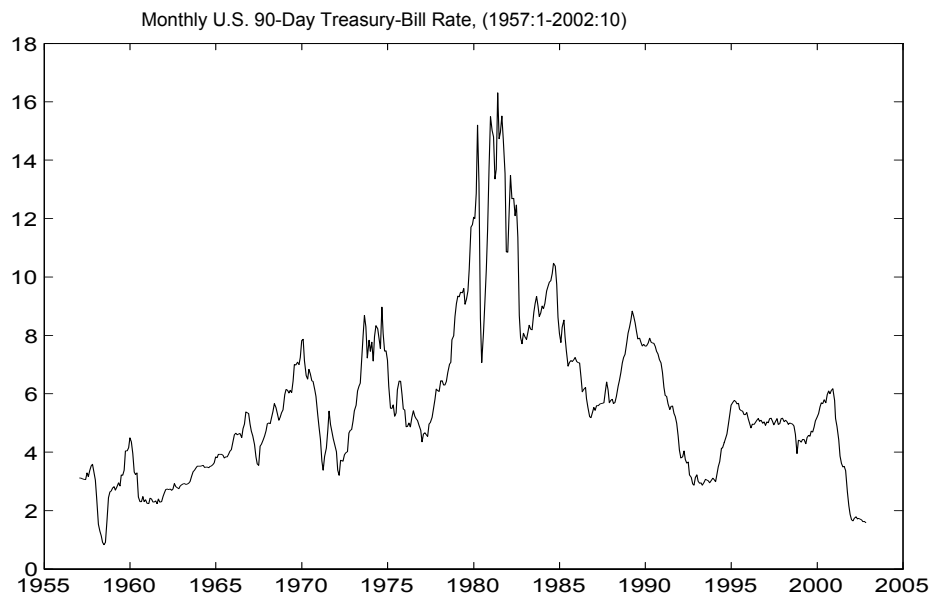
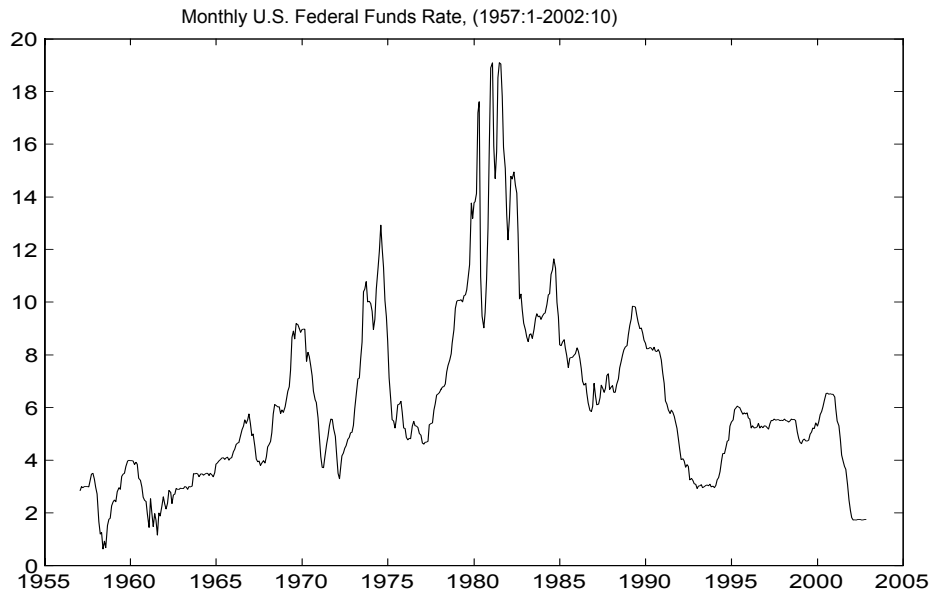


Figure 30. Graphs of the Series

References

- [1] Andrews DWK (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821-856.
- [2] Andrews DWK, Lee I, Ploberger W (1996) Optimal change point tests for normal linear regression. *Journal of Econometrics* **70**, 9-38.
- [3] Andrews DWK, Ploberger W (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62**, 1383-1414.
- [4] Bai J (1997) Estimation of a change point in multiple regression models. *Review of Economics and Statistics* **79**, 551-563.
- [5] Bai J, Perron P (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47-78.
- [6] Ben Aïssa MS, Boutahar M, Jouini J (2004) Bai and Perron's and spectral density methods for structural change detection in the US inflation process. *Applied Economics Letters* **11**, 109-115.
- [7] Ben Aïssa MS, Jouini J (2003) Structural breaks in the US inflation process. *Applied Economics Letters* **10**, 633-636.
- [8] Beran R (1986) Simulating power functions. *Annals of Statistics* **14**, 151-173.
- [9] Beran R, Srivastava (1985) Bootstrap tests and confidence regions for functions of a covariance matrix. *Annals of Statistics* **13**, 95-115.
- [10] Chow GC (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28**, 591-605.
- [11] Christiano LJ (1992) Searching for a break in GNP. *Journal of Business and Economic Statistics* **10**, 237-249.
- [12] Davidson R, MacKinnon J (1998) Graphical methods for investigating the size and power of test statistics. *The Manchester School* **66**, 1-26.
- [13] Davidson R, MacKinnon J (1999) The size distortion of bootstrap tests. *Econometric Theory* **15**, 361-376.
- [14] Davidson R, MacKinnon J (2000) Bootstrap tests: how many bootstraps? *Econometric Reviews* **19**, 55-68.
- [15] Davidson R, MacKinnon J (2006) The power of bootstrap and asymptotic tests. Forthcoming in *Journal of Econometrics*.
- [16] Diebold FX, Chen C (1996) Testing structural stability with endogenous break point: a size comparison of analytic and bootstrap procedures. *Journal of Econometrics* **70**, 221-241.
- [17] Dufour JM, Kiviet JF (1996) Exact tests for structural change in first-order dynamic models. *Journal of Econometrics* **70**, 39-68.

- [18] Dwass M (1957) Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics* **28**, 181-187.
- [19] Efron B (1979) Bootstrap methods; another look at the Jackknife. *Annals of Statistics* **7**, 1-26.
- [20] Hall P (1986) On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics* **14**, 1453-1462.
- [21] Hall P (1992) The bootstrap and edgeworth expansion. *Springer Series in Statistics*, Springer Verlag New York.
- [22] Hall P (1994) Methodology and theory for the bootstrap, In *Handbook of Econometrics*, Volume 4. Engle R. F. and McFadden D. L. (eds), Elsevier, New York.
- [23] Hansen BE (1992) Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics* **10**, 321-335.
- [24] Hansen BE (1997) Approximate asymptotic p-values for structural change tests. *Journal of Business and Economic Statistics* **15**, 60-67.
- [25] Hansen BE (2000) Testing for structural change in conditional models. *Journal of Econometrics* **97**, 93-115.
- [26] Horowitz JL (1994) Bootstrap-based critical values for the information matrix test. *Journal of Econometrics* **61**, 395-411.
- [27] Horowitz JL (1997) Bootstrap methods in econometrics: theory and numerical performance, In *Advances in Economics and Econometrics: Theory and Applications*, Volume 3, pp. 188-222. David M. Kreps and Kenneth F. Wallis (eds), Cambridge, Cambridge University Press.
- [28] Jeong J, Maddala GS (1993) A perspective on application of bootstrap methods in econometrics, In *Handbook of Statistics: Econometrics*, Volume 11, McGraw-Hill, pp. 573-610. Maddala G. S., Rao C. R. and Vinod H. D. (eds).
- [29] Jouini J, Boutahar M (2003) Structural breaks in the US inflation process: a further investigation. *Applied Economics Letters* **10**, 985-988.
- [30] Ploberger W, Krämer W (1996) A trend-resistant test for structural change based on OLS residuals. *Journal of Econometrics* **70**, 175-186.
- [31] Quandt RE (1960) Tests of the hypothesis that a linear regression obeys two separate regimes. *Journal of the American Statistical Association* **55**, 324-330.
- [32] Stock JH, Watson MW (1988) Testing for common trends. *Journal of the American Statistical Association* **83**, 1097-1107.