Two Sides Of The Same Coin: Bootstrapping The Restricted Vs. Unrestricted Model

Panagiotis Mantalos
Lund University, Sweden

Recommended Citation

Two Sides Of The Same Coin:
Bootstrapping The Restricted Vs. Unrestricted Model

Panagiotis Mantalos
Department of Statistics
Lund University, Sweden

The properties of the bootstrap test for restrictions are studied in two versions: 1) bootstrapping under the null hypothesis, restricted, and 2) bootstrapping under the alternative hypothesis, unrestricted. This article demonstrates the equivalence of these two methods, and illustrates the small sample properties of the Wald test for testing Granger-Causality in a stable stationary VAR system by Monte Carlo methods. The analysis regarding the size of the test reveals that, as expected, both bootstrap tests have actual sizes that lie close to the nominal size. Regarding the power of the test, the Wald and bootstrap tests share the same power as the use of the Size-Power Curves on a correct size-adjusted basis.

Key words: Bootstrap, Granger-Causality, VAR system, Wald test

Introduction

When studying the small sample properties of a test procedure by comparing different tests, two aspects are of importance:

a) to find the test that has actual size closest to the nominal size, and given that (a) holds, and
b) to find the test that has the greatest power.

In most cases, however, the distributions of the test statistic used are known only asymptotically and, unfortunately, unless the sample size is very large, the tests may not have the correct size. Inferential comparisons and judgements based on them might be misleading. Gregory and Veall (1985) can be consulted for an illustrative example.
One of the ways to deal with this situation is to use the bootstrap. The use of this procedure is increasing with the advent of personal computers.

Panagiotis Mantalos, Department of Statistics Lund University, Box 743 SE-22007 Lund, Sweden. E-mail: Panagiotis.Mantalos@stat.lu.se

However, the issue of the bootstrap test, even it is applied, is not trivial. One of the problems is that one needs to decide how to resample the data, and whether to resample under the null hypothesis or under the alternative hypothesis.
By bootstrapping under the null hypothesis, an approximation is made of the distribution of the test statistic, thereby generating more robust critical values for our test statistic. Alternately, by bootstrapping under the alternative hypothesis, an approximation is made of the distribution of the parameter, and is subsequently used to make inferences.
In either case, it does not matter whether the nature of the theoretical distribution of the parameter estimator or the theoretical distribution of the test statistic is known. What matters is that the bootstrap technique approximates those distributions.
In this article, the bootstrap test procedure shows that

a) by bootstrapping under the null hypothesis (that is, bootstrapping the restricted model), and
b) by bootstrapping under the alternative hypothesis (that is, bootstrapping the unrestricted model)

will lead to the same results.

The properties of the two different methods will be illustrated and investigated using Monte Carlo methods. The Residual Bootstrap, (RB), will be used to study the properties of the test procedure when the errors are identically and independently distributed. To provide an example that is easy to be extended to a more general hypothesis, it is convenient to use the Wald test for restrictions for testing Granger-causality in a stable stationary VAR system.

The Model
Consider the general linear model

$$y = \mathbf{X}\beta + \delta \qquad (1)$$

where y is an $(n \times 1)$ vector, $\mathbf{X}$ is an $(n \times K)$ matrix and b is a $(K \times 1)$ vector. It is assumed that δ is an n-dimensional normal vector $\mathbb{N}(0, \Omega)$.

Consider testing q independent linear restrictions:

$$H_o : \mathrm{R}\beta = \mathrm{r} \quad \text{vs.} \quad H_1 : \mathrm{R}\beta \neq \mathrm{r}, \qquad (2)$$

where q and r are fixed (q x 1) vectors and R is a fixed $(q \times K)$ matrix with rank q. It is possible to base a test of H_0 on the Wald criterion

$$T_s = (R\hat{\beta} - r)' \left[Var(R\hat{\beta}) \right]^{-1} (R\hat{\beta} - r). \qquad (3)$$

Bootstrap critical values
The bootstrap technique improves the critical values, so that the true size of the test approaches its nominal value. The principle of bootstrap critical values is to draw a number of Bootstrap samples from the model under the null hypothesis, calculate the Bootstrap test statistic T_s^*, and compare it with the observed test statistic.

The bootstrap procedure for calculation of the critical values is given by the following steps:

a)   Estimate the test statistic as in (3)

b)   Use the adjusted OLS residuals $(\hat{\delta}_i - \overline{\delta})$ i = 1,...,T to draw i.i.d $\delta_1^*,...,\delta_T^*$ data. Define

$$y^* = X\hat{\beta}_0 + \delta^*. \qquad (4)$$

c)   Then, calculate the test statistic T_s^* as in (3), i.e., by applying the Wald test procedure to the (4) model. Repeat this step N_b times and take the $(1-\alpha)^{th}$ quintile of the bootstrap distribution of T_s^* to obtain the α - level Bootstrap critical values ($c_{t\alpha}^*$). Reject H_o if $T_s \geq c_{t\alpha}^*$.

Among articles that advocated this approach are Horowitz (1994) and Mantalos and Shukur (1998), whereas Davidson and MacKinnon (1999) and Mantalos (1998) advocated the estimate of the P-value. A bootstrap estimate of the P-value for testing is P*{ $T_s^* \geq T_s$ }.

Bootstrap-hypothesis testing
One of the important considerations for generating the y_t^* leading to the bootstrap critical values is whether to impose the null hypothesis on the model from which is generated the y_t^*. However, some authors, including Jeong and Chung (2001), argued for bootstrapping under the alternative hypothesis. 'Let the data speak' is their principle in apply the bootstrap. The bootstrap procedure to resample the data from the unrestricted model consists of the following steps:

a)   Estimate the test statistic as in (3)
b)   Use the adjusted OLS residuals $(\hat{\delta}_i - \overline{\delta})$ i = 1,...,T. to draw i.i.d $\delta_1^*,...,\delta_T^*$ data. Define $y^* = X\hat{\beta} + \delta^*$, noting that $\hat{\beta}$ is the unconstrained LS estimator of β. That is, the unrestricted model is used to simulate the y^*.
c)   Calculate

$$T_s^* = \frac{\left(R(\hat{\beta}^* - \hat{\beta}) \right)^2}{Var(R\hat{\beta}^*)}. \qquad (5)$$

By repeating this step N_b times the $(1-\alpha)^{th}$ quintile can be used of the bootstrap distribution of the (5) as the α - level Bootstrap critical values ($c_{t\alpha}^*$). Reject H_o if $T_s \geq c_{t\alpha}^*$. The bootstrap estimate of the P-value is P*$\{T_s^* \geq T_s\}$.

Since Efron's (1979) introduction of the bootstrap as a computer-based method for evaluating the accuracy of a statistic, there have been significant theoretical refinements of the technique. Horowitz (1994) and Hall and Horowitz (1996) discussed the method and showed that bootstrap tests are more reliable than asymptotic tests, and can be used to improve finite-sample performance. They provided a heuristic explanation of why the bootstrap provides asymptotic refinements to the critical values of test statistics. See Hall (1992) for a wider discussion on bootstrap refinements based on the Edgeworth expansion.

Davidson and MacKinnon (1999) provided an explanation of why the bootstrap provides asymptotic refinements to the p- values of a test. The same authors conclude that by using the bootstrap critical values or bootstrap test, the size distortion of a bootstrap test is at least of order $T^{-\frac{1}{2}}$ smaller than that of the corresponding asymptotic test.

Two sides of the same coin

Consider the general linear model

$$y = \mathbf{X}\beta + \delta \qquad (1)$$

and suppose that the interest is in testing the q independent linear restrictions

$$H_o: \mathbf{R}\beta = \mathbf{r} \quad \text{vs.} \quad H_1: \mathbf{R}\beta \neq \mathbf{r}. \qquad (2)$$

Let the LS unconstrained estimator of β be denoted by $\hat{\beta}$ and the equality-constrained estimator be denoted by $\hat{\beta}_0$. The bootstrap GDPs are:

a) Restricted: $y_R^* = \mathbf{X}\hat{\beta}_0 + \delta^*$. $\qquad (6)$

b) Unrestricted: $y_u^* = \mathbf{X}\hat{\beta} + \delta^*$. $\qquad (7)$

Let $\hat{\beta}^*$ be the LS estimator of the b coefficient in the model relating y_R^* to $\mathbf{X}$, and $\hat{\hat{\beta}}^*$ be the LS estimator of the b in the y_u^* on $\mathbf{X}$ model. Thus,

$$\hat{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y_R^* = \hat{\beta}_0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta^* \qquad (8)$$

and

$$\hat{\hat{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y_u^* = \hat{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta^*. \qquad (9)$$

From (8) and (9)

$$\hat{\beta}^* - \hat{\beta}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta^* \qquad (10)$$

and

$$\hat{\hat{\beta}}^* - \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\delta^*. \qquad (11)$$

Because the right-hand components of the (10) and (11) are equal,

$$\left(\hat{\beta}^* - \hat{\beta}_0\right) = \left(\hat{\hat{\beta}}^* - \hat{\beta}\right). \qquad (12)$$

It is not difficult to see from (12) that the same results from the both methods are expected: there are two sides to the same coin. These results will be illustrated by a Monte Carlo experiment.

Wald test for restrictions in a VAR model

Consider a data-generation process (DGP) that consists of the k-dimensional multiple time series generated by the VAR(p) process

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + \varepsilon_t, \qquad (13)$$

where $\varepsilon_t = \left(\varepsilon_{1t}, \ldots, \varepsilon_{kt}\right)'$ is a zero mean independent white noise process with non-singular covariance matrix Σ_ε and, for j = 1, ... ,

k, $\mathrm{E}|\varepsilon_{jt}|^{2+\tau} < \infty$ for some $\tau > 0$. The order p of the process is assumed to be known. Define

$$\mathbf{Y} := (y_1, \cdots, y_T) \qquad (\text{k x T}) \text{ matrix,}$$

$$\mathbf{B} := (v, A_1, \cdots, A_p) \quad (\text{k x (kp+ 1)}) \text{ matrix,}$$

$$\mathbf{Z}_t := \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} \qquad ((\text{kp +1}) \text{ x 1 }) \text{ matrix,}$$

$$\mathbf{Z} := (\mathbf{Z}_0, \cdots, \mathbf{Z}_{T-1}) \qquad ((\text{kp+1}) \text{ x T}) \text{ matrix}$$

and

$$\delta := (\varepsilon_1, \cdots, \varepsilon_T) \qquad (\text{k x T}) \text{ matrix.}$$

By using this notation, for t = 1, ..., T, the VAR (p) model including a constant term (v) can be written compactly as

$$\mathbf{Y} = \mathbf{BZ} + \delta. \tag{14}$$

Then, the LS estimator of the **B** is

$$\hat{\mathbf{B}} = \mathbf{YZ}'(\mathbf{ZZ}')^{-1}. \tag{15}$$

Let $\alpha_p = vec[A_1, \cdots, A_p]$ be the vector of the true parameters, and $\hat{\alpha}_p = vec[\hat{A}_1, \cdots, \hat{A}_p]$ be the vector the LS estimators of the parameters, where vec[.] denotes the vectorization operator that stacks the columns of the argument matrix. Then,

$$T^{1/2}(\hat{\alpha}_p - \alpha_p) \Rightarrow N(0, \Sigma_p), \tag{16}$$

where $\Rightarrow$ denotes weak convergence in distribution and the $[k^2(p) \text{ x } k^2(p)]$ covariance matrix Σ_p is non-singular.

Now, suppose that in testing q independent linear restrictions is of interest

$$H_o: \mathrm{R}\alpha_p = \mathrm{s} \quad \text{vs.} \quad H_1: \mathrm{R}\alpha_p \neq \mathrm{s}, \tag{17}$$

where q and s are fixed (q x 1) vectors and R is a fixed [q x $k^2 p$] matrix with rank q.

We can base a test of H_0 on the Wald criterion

$$T_{s\_wald} = (R\hat{\alpha}_p - s)'\big[Var(R\hat{\alpha}_p - s)\big]^{-1}(R\hat{\alpha}_p - s). \tag{18}$$

Let

$$\hat{\Sigma}_\delta = \frac{1}{T - kp - 1}\big(\mathbf{YY}' - \mathbf{YZ}'(\mathbf{ZZ}')^{-1}\mathbf{ZY}'\big) \tag{19}$$

be the estimate of the residual covariance matrix.

Then, the diagonal elements of $(ZZ')^{-1} \otimes \hat{\Sigma}_\delta$ form the variance vector of the LS estimated parameters. Substitute (19) into (18) in order to have

$$T_{s\_wald}$$
$$= (R\hat{\alpha}_p - s)'\Big[R\big((ZZ')^{-1} \otimes \hat{\Sigma}_\delta\big)R'\Big]^{-1}(R\hat{\alpha}_p - s). \tag{20}$$

The null hypothesis of no Granger-causality may be expressed in terms of the coefficients of VAR process as

$$H_o: \mathrm{R}\alpha_p = 0 \quad \text{vs.} \quad H_1: \mathrm{R}\alpha_p \neq 0. \tag{21}$$

Then, (20) can be written as

$$T_{s\_wald}$$
$$= (R\hat{\alpha}_p)'\Big[R\big((ZZ')^{-1} \otimes \hat{\Sigma}_\delta\big)R'\Big]^{-1}(R\hat{\alpha}_p) \tag{22}$$

and the bootstrap variations as

$$T^*_{s\_wald}$$
$$= (R\hat{\alpha}^*_p)'\Big[R\big((Z^*Z^{*'})^{-1} \otimes \hat{\Sigma}^*_\delta\big)R'\Big]^{-1}(R\hat{\alpha}^*_p) \tag{23}$$

for the restricted form and

$T^{*}_{s\_wald}$

$$= (R\hat{\alpha}^{*}_{p} - R\hat{\alpha}_{p})' \left[R\left(\left(Z^{*}Z^{*'}\right)^{-1} \otimes \hat{\Sigma}^{*}_{\delta} \right) R' \right]^{-1}$$

$$(R\hat{\alpha}^{*}_{p} - R\hat{\alpha}_{p})$$

(24)

for the unrestricted form.

Methodology

Monte Carlo experiment

This section illustrates various generalizations of the Granger-causality tests in VAR systems with stationary variables, using Monte Carlo methods. The estimated size is calculated by observing how many times the null is rejected in repeated samples under conditions where the null is true.

The following VAR(1) process is generated:

$$y_t = \begin{bmatrix} 0.5 & 0.3 \\ T^{-1/2}\gamma & 0.5 \end{bmatrix} y_{t-1} + \varepsilon_t,$$ (25)

where $\varepsilon_t \sim N\left(0, I_2\right)$, $y_t = \left(y_{1t}. y_{2t}\right)'$. If $\gamma = 0$, y_{1t} is Granger-noncausal for y_{2t} and if $\gamma \neq 0$, y_{1t} causes y_{2t}. Therefore, $\gamma = 0$ is used to study the size of the tests.

The order p of the process is assumed to be known. Because this assumption might be too optimistic, a VAR(2) is fitted:
$$y_t = v + A_1 y_{t-1} + A_2 y_{t-2} + \varepsilon_t.$$

For each time series, 20 pre-sample values were generated with zero initial conditions, taking net sample sizes of T = 25 and 50. The Bootstrap test statistic (T^{*}_{s}) is calculated. As for N_b, which is the size of the bootstrap sample used to estimate bootstrap critical values and the P-value, $N_b = 399$ is used. Note that there are no initial bootstrap observations in bootstrap procedure.

Next presented are the results of the Monte Carlo experiment concerning the sizes of the various versions of the tests statistics using the VAR(2) model. Graphical methods are used that were developed and illustrated by Davidson

and MacKinnon (1998) because they are easy to interpret. The P-value plot is used to study the size, and the Size-Power curves is used to study the power of the tests. The graphs, the P-value plots and Size-Power curves are based on the empirical distribution function, the EDF of the P-values, denoted as $\hat{F}\left(x_j\right)$.

For the P-value plots, if the distribution used to compute the p_s terms is correct, each of the p_s terms should be distributed uniformly on (0,1). Therefore the resulting graph should be close to the 45° line.

Furthermore, to judge the reasonableness of the results, a 95% confidence interval is used for the nominal size (π_0) as:

$\pi_0 \pm 2\sqrt{\dfrac{\pi_0(1 - \pi_0)}{N}}$, where N is the number

of Monte Carlo replications. Results that lie between these bounds will be considered satisfactory. For example, if the nominal size is 5%, define a result as reasonable if the estimated size lies between 3.6% and 6.4%. The P-value plots also make it possible and easy to distinguish between tests that systematically over-reject or under-reject, and those that reject the null hypothesis about the right proportion of the time.

Figure 1 shows the truncated P-value plots for the actual size of the bootstrap and the Wald tests, using 25 and 50 observations. Looking at these curves, it is not difficult to make the inference that both the bootstrap tests perform adequately, as they lie inside the confidence bounds. However, using the asymptotic critical values, the Wald test shows a tendency to over-reject the null hypothesis.

The superiority of the bootstrap test over the Wald test, concerning the size of the tests, is considerable, and more noticeable in small samples of size 25. The power of the Wald and bootstrap tests by using sample sizes of 25 and 50 observations was examined. The power function is estimated by calculating the rejection frequencies in 1000 replications using the value $\gamma = 2$.

The Size-Power Curves are used to compare the estimated power functions of the alternative test statistics. This proved to be quite

adequate, because those tests that gave reasonable results regarding size usually differed very little regarding power.

The same processes are followed for the size investigation to evaluate the EDFs denoted by $\hat{F}^{\oplus}\left(x_{j}\right)$, by using the same sequence of random numbers used to estimate the size of the tests. Size-Power Curves are used to plot the estimated power functions against the nominal size. The estimated power functions are plotted against the true size, that is, plotting $\hat{F}^{\oplus}\left(x_{j}\right)$ against $\hat{F}\left(x_{j}\right)$, which produces the Size-Power Curves on a correct size-adjusted basis.

Figure 2 shows the results of using the Size-Power Curves. The Wald test has higher power than the restricted and unrestricted bootstrap tests. A sample effect can also be seen. The larger the sample, the larger is the power of the tests. As the sample size increases, the power difference decreases.

However, the most interesting result is that both the restricted and unrestricted bootstrap tests share the same power. This result confirms the view that these two bootstrap methods are two sides of the same coin.

When using the Size-Power Curves on a correct size-adjusted basis, however, the situation is different concerning the power of the Wald and the bootstrap tests. Now the Wald, restricted and unrestricted bootstrap tests share the same power, as seen in Figure 3.

Conclusion

The purpose of this study was to provide advice on whether to resample under the null hypothesis or under the alternative hypothesis. In summary:

a)   the restricted bootstrap test was used, in which the distribution of the test statistic was approximated, generating more robust critical values for our test statistic, and

b)   the unrestricted bootstrap test, where the distribution of the parameter (coefficient) was approximated.

In both cases it does not matter whether or not the nature of the theoretical distribution of the parameter estimator or the theoretical distribution of the test statistic is known. What matters is that the bootstrap technique well approximates those distributions. Moreover, this article demonstrated the equivalence of these two methods.

The conclusion to this investigation for the Granger-causality test is that both bootstrap tests have an actual size that lies close to the nominal size. Given that the both unrestricted and restricted models have the same power, it makes sense to choose the bootstrap ahead of the classical tests, especially in small samples.

Figure 1. P-values Plots Estimated Size of the Wald and Bootstrap Tests.
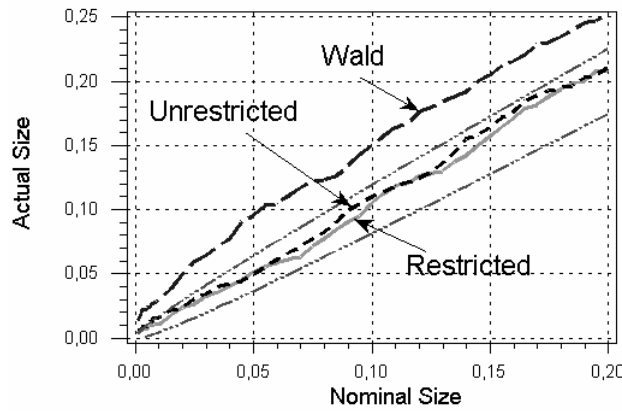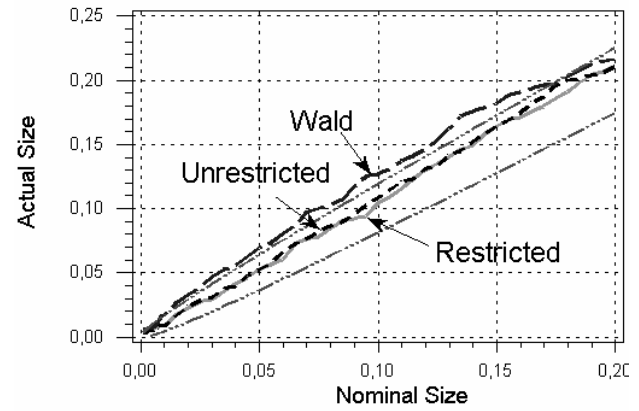
Figure 1a: 25 observations

Figure 1b: 50 observations

Dash 3Dot lines: 95% Confidence interval

Figure 2. Estimated Power of the Wald and Bootstrap Tests.
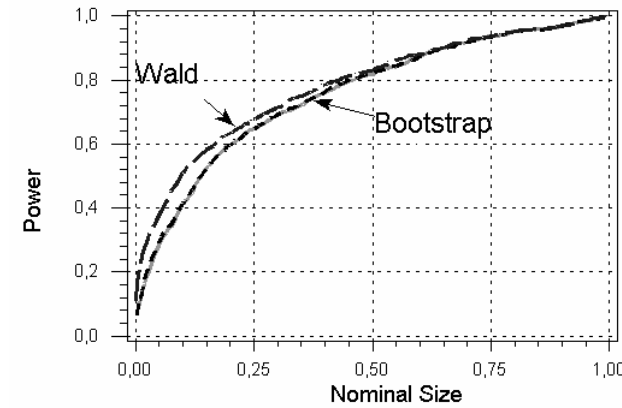
Figure 2a: 25 observations
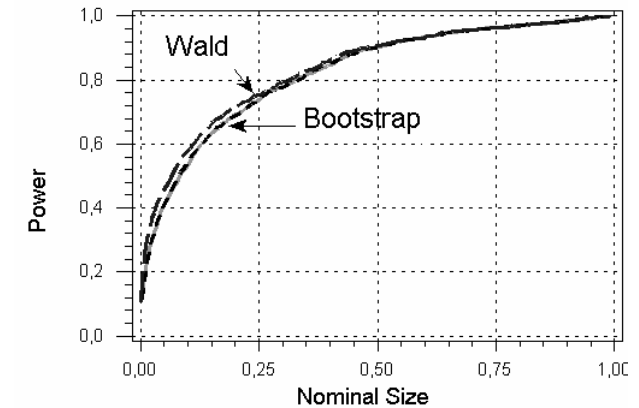
Figure 2b: 50 observations

Figure3. Size-adjusted Power of the Wald and Bootstrap Tests.
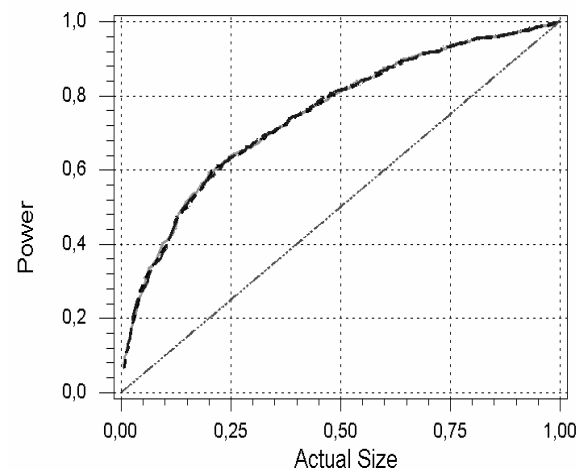
Figure 3a: 25 observations
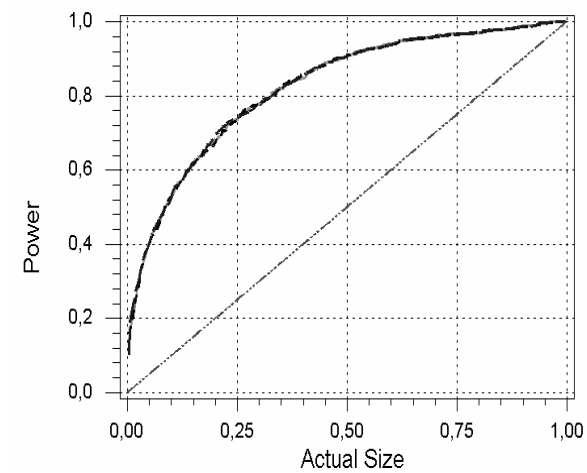
Figure 3b: 50 observations

References

Davidson, R., & MacKinnon, J. G. (1998). The size distortion of bootstrap tests. *Econometric Theory, 15*, 361-376.

Davidson, R., & MacKinnon, J. G. (1996). The Power of bootstrap tests. Discussion paper, *Queen's University, Kingston*, Ontario.

Davidson, R., & MacKinnon, J. G. (1998). Graphical methods for investigating the size and power of test statistics. *The Manchester School*, *66*, 1-26.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics 7*, 1-26.

Gregory, A. W., & Veall, M. R. (1985). Formulating Wald tests of nonlinear restrictions. *Econometrica 53*, 1465-1468.

Hall, P. (1992). *The bootstrap and Edgeworth expansion.* New York: Springer-Verlag.

Hall, P., & Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized - method - of - moments estimators. *Econometrica, 64*, 891–916.

Horowitz, J. L. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics, 61*, 395-411.

Jeong, J., & Chung, S. (2001). Bootstrap test for autocorrelation. *Computational Statistics and Data Analysis*, *38*(1) 49-69.

MacKinnon, J. G. (1992). Model specification tests and artificial regressions. *Journal of Economic Literature*, *30*, 102-146.

Maddala, G. S. (1992). *Introduction to Econometrics,* (2ed,). New York: Maxwell Macmillan.

Mantalos, P. (1998). A graphical investigation of the size and power of the Granger-causality tests in Integrated-Cointegrated VAR Systems. *Studies in Nonlinear Dynamics and Econometrics* 4.1, 17-33

Mantalos, P., & Shukur. G. (1998). Size and power of the error correction model (ECM) cointegration test - A Bootstrap approach. *Oxford Bulletin of Economics and Statistics*, *60*, 249-255.