Bootstrapping the score vector to correct the score test

Dirk Hoorelbeke*
K.U.Leuven

June 2004

Abstract

In this paper a method is proposed to enhance the performance of the score test. The standard score test is corrected in two ways. A first step is to transform the score vector such that it is asymptotically (or exactly) pivotal. For this purpose one can use the inverse of a square root of a consistent variance estimate of the score vector, although other possibilities exist. By bootstrapping the transformed score vector, a second-order correct variance matrix estimate is obtained, to be used in the quadratic form score test statistic. In the second step, the bootstrap simulations are recycled to compute a second-order correct critical value. Monte Carlo simulations show that the corrected score test outperforms the standard score test in terms of error in rejection probability and power.

JEL classification: C12, C15
Keywords: parametric bootstrap, score test

*I am indebted to Geert Dhaene for helpful suggestions and comments. Address: K.U.Leuven, Department of Economics, Naamsestraat 69, 3000 Leuven, Belgium. Tel. +32 16 326652. Email: dirk.hoorelbeke@econ.kuleuven.ac.be

1 Introduction

The Lagrange multiplier test, or score test, suggested independently by Aitchison and Silvey (1958) and Rao (1948), tests for parametric restrictions. Although the score test is an intuitively appealing and often used procedure, the exact distribution of the score test statistic is generally unknown and is often approximated by its first-order asymptotic χ^2 distribution. In problems of econometric inference, however, first-order asymptotic theory may be a poor guide, and this is also true for the score test, as demonstrated in several Monte Carlo studies. See e.g. Breusch and Pagan (1979), Bera and Jarque (1981), Davidson and MacKinnon (1983, 1984, 1992), Chesher and Spady (1991), Horowitz (1994) and Aparicio and Villanua (2001), among others.

One can use the bootstrap distribution of the score test statistic to obtain a critical value which is more accurate then the asymptotic critical value (Hall 1992). However, the score test uses a quadratic form statistic. In the construction and implementation of such a quadratic form statistic two important aspects which determine the performance of the test (both under the null and the alternative), are (i) the weighting matrix (the estimate of the variance matrix of the score vector) and (ii) the critical value. Since the score test statistic is asymptotically pivotal (it is χ^2), the bootstrap critical value is second-order correct (whereas the asymptotic critical value is only first-order correct). Imagine now that the statistic being bootstrapped uses a poor weighting matrix (i.e. the variance matrix of the score vector is imprecisely estimated). Then the error in rejection probability (ERP)[1] of the test (using bootstrap critical values) will still be fine, but the power can be very low. Therefore it is important to obtain not only an accurate critical value, but also a good weighting matrix in the quadratic form test statistic. The importance of the weighting matrix for the score test is also discussed in Godfrey and Orme (2001). The problem of obtaining a good weighting matrix by using the bootstrap is also addressed in Dhaene and Hoorelbeke (2004).

This paper is about using the bootstrap to obtain both an accurate critical value and a good weighting matrix, using only a limited number of simulations. The method is related to Beran's (1988) technique of prepivoting. Prepivoting is transforming a (univariate) statistic by its bootstrap distribution function to obtain a new statistic whose ERP is smaller (using bootstrap critical values) than the ERP of the original statistic (using bootstrap critical values). Although Beran (1988) uses the bootstrap distribution function of the statistic to transform it, one can use any monotone mapping which makes the (possibly multivariate) statistic less dependent on the underlying probability distribution. I propose a multivariate rescaling of the score vector, such that it is asymptotically pivotal (or exactly pivotal if possible). One can always find such a transformation, given a consistent estimate of the (asymptotic) variance matrix of the score vector. Then a quadratic form statistic is constructed in this transformed score vector using

[1] The ERP of a test is the actual minus the nominal (i.e. chosen) probability of rejecting the null hypothesis when it is true.

the inverse of its bootstrap variance matrix as weighting matrix. This bootstrap variance matrix is more accurate then an estimate of the asymptotic covariance matrix. The simulations used to compute the bootstrap variance matrix are then recycled to compute a critical value which is more accurate then the asymptotic critical value. This avoids a nested bootstrap while it yields a second-order correct critical value.

Section 2 contains a short review on the score test and presents the bootstrap-based correction method. In Section 3 some simulation results on the information matrix test (White, 1982) in the probit model and the linear regression model are presented. The results indicate that the corrected score test performs very well, both in terms of ERP and (ERP-corrected) power. Section 4 concludes.

2 The score test and a bootstrap correction

Consider a parametric model defined by a density function $f(y, \theta)$, where θ is a $p \times 1$ vector of parameters. The hypothesis

$$H_0 : g(\theta) = 0,$$

where g is a known q-vector valued function, can be tested using the score test (or Lagrange multiplier test). The (suitably normalised) log-likelihood function from an i.i.d. sample of size n is

$$l = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \log f(y_i, \theta),$$

with score vector $s = \partial l / \partial \theta$. (The arguments of l and s are omitted since no confusion is possible.)

For a nice review on the score test see Godfrey (1990). The general form of the score statistic is

$$\omega = \hat{s}'\hat{V}^{-1}\hat{s},$$

where $\hat{}$ indicates evaluation at the restricted ML estimate $\hat{\theta}$ (i.e. $\hat{\theta}$ maximizes l subject to $g(\theta) = 0$) and $\hat{V}$ is a consistent estimator of the information matrix $\mathcal{I} = E[ss']$. Under H_0, ω is asymptotically $\chi^2(q)$ distributed. We assume that $\mathcal{I}$ is non-singular. This ensures that $\hat{V}$ will be non-singular for sufficiently large n.

Using asymptotic critical values the ERP of the score test is of order $O(n^{-1})$ (Horowitz, 2001). Since the score test statistic is asymptotically pivotal, bootstrap critical values are second-order correct (i.e. the ERP is then $O(n^{-2})$) (Horowitz, 2001). Different choices of $\hat{V}$ lead to different tests. As is evident from the literature on the information matrix test, this choice is not without importance. This is also noticed by Godfrey and Orme (2001).

Here a method is proposed to obtain both a second-order correct variance matrix estimate and a second-order correct critical value, using only one round of bootstrap simulations. To be more specific, assume there exists a matrix A such that the score vector premultiplied by A is asymptotically pivotal. An obvious choice for A is the inverse of a square root of a variance matrix estimate of the score vector, yielding a multivariate studentized score vector. This is not the only possible choice for A, though, as is shown in Section 3 for the information matrix test in the linear regression model.

Given a consistent estimate of the variance matrix of the score, this method is always applicable. Let $\hat{V}^{-1/2}$ denote the inverse of a square root of a consistent estimate of the asymptotic variance matrix of the score vector $\hat{s}$. For the remainder of this section, take $\hat{A} = \hat{V}^{-1/2}$. Then $\hat{A}\hat{s}$ is a multivariate studentized score vector. Since the multivariate studentized score vector is asymptotically pivotal, its bootstrap distribution is a second-order approximation to its exact finite sample distribution. This bootstrap distribution, however, can only be computed analytically in very simple cases. In general one has to resort to simulations.

The procedure is as follows. Let $\hat{d}_1 = \hat{s}$, $\hat{d}_2 = \hat{A}\hat{s}$, and let $B > q$ be the number of bootstrap replications. Then

1. compute $\hat{\theta}$ and $\hat{d}_2$;
2. for $b = 1, \ldots, B$:
   - generate a sample of size n from $f(\cdot, \hat{\theta})$;
   - for this sample compute $\hat{\theta}_b$ and $\hat{d}_{2b}$;
3. compute $\hat{V}_{2B} = (B-1)^{-1} \sum_{b=1}^{B} (\hat{d}_{2b} - \bar{d}_{2B})(\hat{d}_{2b} - \bar{d}_{2B})'$, where $\bar{d}_{2B} = B^{-1} \sum_{b=1}^{B} \hat{d}_{2b}$;
4. compute $\omega_2 = \hat{d}_2' \hat{V}_{2B}^{-1} \hat{d}_2$;
5. calculate the edf of $\omega_{2b} = \hat{d}_{2b}' \hat{V}_{2B}^{-1} \hat{d}_{2b}$ for $b = 1, \ldots, B$ and call this $\hat{F}_{2B}$.

Under the null hypothesis, ω_2 is asymptotically, as $n \to \infty$ and $B \to \infty$, χ^2_q distributed. If one uses simulations to approximate the bootstrap variance matrix of $\hat{d}_2$, then B is finite. The asymptotic distribution of ω_2 in this case is $T^2_{q,B-1}$ (Hotelling's T^2; see Dhaene and Hoorelbeke, 2004).

Since the studentized score vector $\hat{d}_2$ is asymptotically pivotal (by construction), the error made by the bootstrap distribution of $\hat{d}_2$ is $O(n^{-1})$ (Horowitz, 2001). The error made by the first-order asymptotic distribution [2] is $O(n^{-1/2})$. This also holds for the variance estimates of $\hat{d}_2$ derived from both distributions ($\hat{V}_{2B}$ and I respectively). If the studentized score vector is exactly pivotal, then the bootstrap variance matrix equals the exact finite sample variance matrix. So in this case, by choosing B sufficiently large, this matrix can be estimated to any desired accuracy.

Using the T^2 critical values for ω_2, one remains with first-order asymptotics. Hence, the ERP of the test based on ω_2 with T^2 critical values and the ERP of the test based on ω with χ^2 critical values are both $O(n^{-1})$. Given the correction in the weighting matrix, however, it is expected that in finite samples the ERP of ω_2 with T^2 critical values will be smaller than the ERP of ω with χ^2 critical values. The simulations in Section 3 indeed show that this is true, at least for the information matrix test in the probit model and in the linear model.

If one uses bootstrap critical values, the tests based on ω or ω_2 both have an ERP of $O(n^{-2})$, i.e. the bootstrap critical values are second-order correct (see e.g. Horowitz, 2001), since the score test statistic is asymp-

[2] The first-order asymptotic distribution of $\hat{d}_2$ under H_0 is multivariate normal with mean zero and variance matrix the identity matrix, if $\hat{V}^{-1/2}$ is used to transform the score vector. In other cases, $\hat{d}_2$ is multivariate normal with mean zero and some variance matrix, independent of the parameter vector.

totically pivotal. In some cases the score test statistic is exactly pivotal. Then bootstrap critical values are exact (i.e. the ERP goes to zero for B tending to infinity for fixed n). To avoid a nested bootstrap procedure for the bootstrap test based on ω_2, one can use the appropriate quantile of $\hat{F}_{2B}$ (as constructed above), which re-uses the simulations of the bootstrap variance calculation. If the test statistic is asymptotically (or exactly) pivotal, these critical values are also second-order (or exactly) correct. Thus, with only one round of simulations both a more accurate weighting matrix and a more accurate critical value (compared to first-order asymptotic theory) are obtained. This procedure, however, requires somewhat more simulations, due to the dependency introduced between $\hat{V}_{2B}$ and $\hat{d}_{2b}$, compared with the standard bootstrap-correction method (i.e. when only the critical value is corrected).

To explain the effect the weighting matrix can have on power, consider the following (somewhat artificial) example of the Jarque-Bera (1980) statistic. The Jarque-Bera statistic tests for skewness and non-normal kurtosis, using the following score vector:

$$\hat{s} = \begin{pmatrix} \hat{s}_1 \\ \hat{s}_2 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} \hat{\varepsilon}_t^3 \\ \hat{\varepsilon}_t^4 - 3 \end{pmatrix},$$

where $\hat{\varepsilon}_t = (y_t - \hat{\beta})/\hat{\sigma}$, $\hat{\beta}$ and $\hat{\sigma}$ are the estimated parameters of a normal model without covariates. The variance matrix of $\hat{s}$ is

$$V = \begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix}.$$

Naturally, the Jarque-Bera statistic equals $\hat{s}_1^2/6 + \hat{s}_2^2/24$. Suppose now instead of using V, the following variance estimate was used in the construction of the test statistic:

$$\hat{V} = \begin{pmatrix} 6 & 0 \\ 0 & 10^{10} \end{pmatrix}.$$

Then it follows that $\omega = (\hat{s}_1 \;\; \hat{s}_2)\hat{V}^{-1}(\hat{s}_1 \;\; \hat{s}_2)' \approx \frac{1}{6}\hat{s}_1^2$. As a consequence, the test based on this statistic would have no power against leptokurtic or platykurtic alternatives. This example is rather extreme, but it shows that also for power the weighting matrix matters.

The procedure set out above can be iterated to obtain further reductions in ERP. Having obtained $\hat{V}_{2B}$, take, for $j > 2$, $\hat{d}_j = \hat{V}_{j-1,B}^{-1/2}\hat{d}_{j-1}$ and $\omega_j = \hat{d}_j'\hat{V}_{jB}^{-1}\hat{d}_j$ where $\hat{V}_{jB}$ is the bootstrap variance matrix of $\hat{d}_j$. The error made by $\hat{V}_{jB}$ is $O(n^{-j/2})$. The test based on ω_j with a critical value from $\hat{F}_{jB}$ has an ERP which is $O(n^{-j})$.

3 Simulations

In this section some Monte Carlo results are reported on how the proposed method performs when correcting the information matrix test (White, 1982) in the probit model and the linear regression model. The information matrix test is a score test for parameter constancy (Chesher 1984). Let y_t have density $f(y, \theta)$, where θ is a $p \times 1$ vector of parameters, having density $h(\theta)$, expectation θ_0 and variance Σ. The information matrix test is a score test for the null hypothesis $H_0 : \Sigma = 0$, i.e. θ is degenerate with $P(\theta = \theta_0) = 1$. Chesher (1984) shows that the part of the score vector corresponding to $\Sigma^c = \text{vech}(\Sigma)$ (the vectorised lower triangular part of Σ) at $\Sigma = 0$ equals

$$s = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (F_{1t}F_{1t}' + F_{2t})^c,$$

where $F_{1t} = \partial \log f(y_t, \theta)/\partial\theta$ and $F_{2t} = \partial F_{1t}/\partial\theta'$. White (1982) and Chesher (1984) derive the asymptotic distribution of $\hat{s}$ (s evaluated at the restricted ML estimate $\hat{\theta}$). Under the null hypothesis, $\hat{s}$ is asymptotically multivariate normal with mean zero and variance matrix $V_\infty(\theta_0)$. The information matrix test statistic is then defined as

$$\omega = \hat{s}'\hat{V}^{-1}\hat{s},$$

where $\hat{V}$ is a consistent estimate of $V_\infty(\theta_0)$. The statistic has an asymptotic χ^2_q distribution, where $q = p(p+1)/2$. In the literature on the information matrix test a number of estimators of $V_\infty(\theta_0)$ have been proposed (White, 1982; Chesher, 1983; Lancaster, 1984; Orme, 1990; Davidson and MacKinnon, 1992). The ensuing tests behave quite differently in term of ERP and power, stressing the importance of the weighting matrix. In this paper the focus is on the original White (1982) statistic and the Chesher (1983) - Lancaster (1984) (CL) statistic. White's variance estimator just replaces all expectations in the formula of the asymptotic variance by sample averages. The variance estimator proposed by Chesher (1983) and Lancaster (1984) uses the information matrix equality in such a way that the computation of third derivatives of the log-likelihood can be avoided. The information matrix equality states that, if the model is true, the expectation of the hessian matrix of the log-likelihood equals minus the expectation of the outer product of the gradient vector of the log-likelihood. The CL statistic uses the latter expression of the information matrix in the variance estimator, hence the statistic is also called the outer-product-of-gradient (OPG) version. It also replaces expectations by sample averages.

The CL statistic is denoted $\omega_C = \hat{s}'\hat{V}_C^{-1}\hat{s}$. The bootstrap-corrected CL statistic is denoted $\omega_{BC} = \hat{d}_C'\hat{V}_{BC}^{-1}\hat{d}_C$, where $\hat{d}_C = \hat{V}_C^{-1/2}\hat{s}$ and $\hat{V}_{BC}$ is the bootstrap variance matrix of $\hat{d}_C$. The same notation applies to the White statistic: ω_W and ω_{BW}.

First, the more general case of asymptotic pivotalness is considered by looking at the probit model. Next, the IM test in the normal regression model is studied. In this model, exact pivotalness is obtained, and an alternative transformation of the score vector (different from the inverse square root of the variance matrix estimate) is proposed.

In the probit model

$$P(y_t = 1) = \Phi(x_t'\beta) = \Phi_t, \quad t = 1, \dots, n \quad ,$$

with $\Phi(\cdot)$ the standard normal cdf, y_t a binary variable, x_t a $k \times 1$ regressor and β a $k \times 1$ parameter. The score vector, evaluated at the ML estimate $\hat{\beta}$, is

$$\hat{s} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{(Y_t - \hat{\Phi}_t)\hat{\phi}_t'}{\hat{\Phi}_t(1 - \hat{\Phi}_t)} (x_t x_t')^c$$

where $A^c = K\text{vech}(A)$, $K = [0_{q\times 1}\ I_q]$, $q = k(k+1)/2 - 1$, $\hat{\phi}_t' = \phi'(x_t'\hat{\beta})$ is the derivative of the standard normal pdf and $\hat{\Phi}_t = \Phi(x_t'\hat{\beta})$. The first element of the score vector is disregarded since it is identically equal to zero if the first element of x_t is a constant (which is assumed here). In the probit model, the IM test is a consistent misspecification test.

In the simulation experiments x_t consists of a constant and $k-1$ independent standard normal variates, with $k = 2$ and $k = 3$. The sample size

n ranges over 50, 100, 200 and 400. The parameter β is equal to $(\frac{1}{2}, 1)'$ and $(\frac{1}{2}, 1, 1)'$. The number of simulations B to compute the bootstrap variance matrix and the critical value is set to 499 [3]. The same number is used for the standard parametric bootstrap method. The simulation study was carried out with 10000 Monte Carlo runs. The regressor matrix is held fixed across Monte Carlo replications, as in earlier Monte Carlo experiments on the information matrix test (Chesher and Spady (1991), Horowitz (1994) and Orme (1990)).

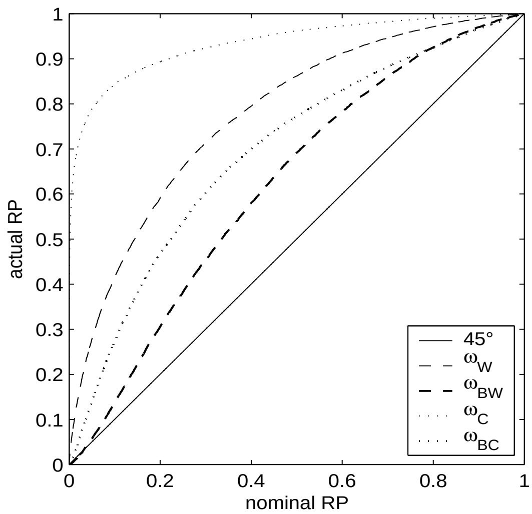
Three properties are investigated: the ERP of the tests using asymptotic critical values, the ERP of the tests using bootstrap critical values and the power against a heteroskedastic alternative.

First, the ERP of ω_W, ω_C and their bootstrap-corrected versions, using asymptotic critical values is studied. This means that the distributions of e.g. ω_C and ω_{BC} are approximated by the χ^2_q distribution and the $T^2_{q,B-1}$ distribution, respectively. If asymptotic critical values are employed, one can easily see the net effect of using a improved weighting matrix by comparing e.g.the ERPs of ω_C and ω_{BC}. The ERP is displayed using p-value plots (Davidson and MacKinnon, 1998). A p-value plot gives the (estimated) actual rejection probability (RP) of a test as a function of the nominal RP. On the 45° line actual and nominal RP agree, so ideally the p-value plot coincides with the 45° line.

The p-value plots for $k = 3$ and $n = 100$ are displayed in Figure 1. The

[3] Rather than 500, since with $B = 499$ the 5% bootstrap critical value is the 475th (largest) value of the ordered ω_{2b}.

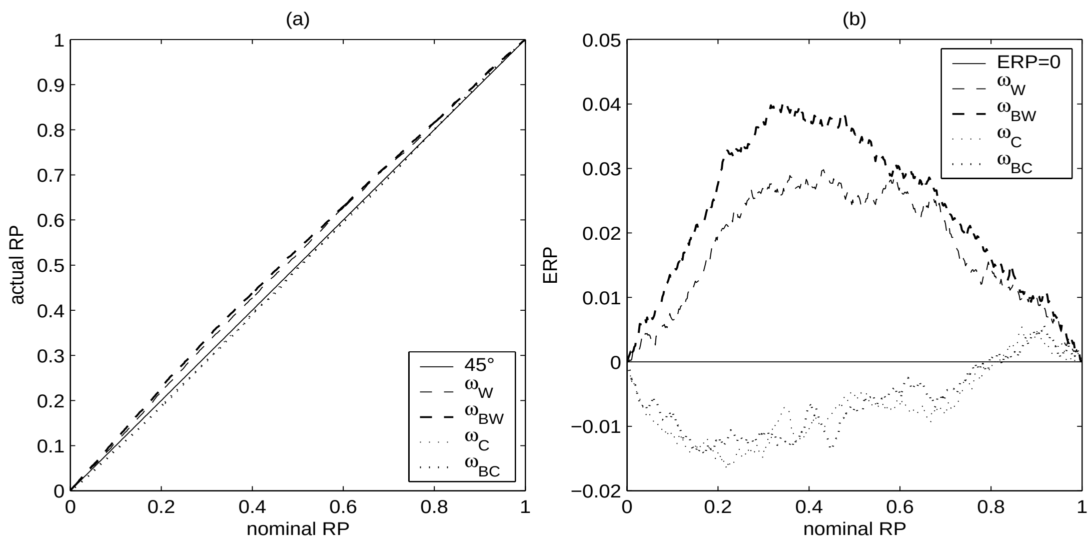
Figure 1: Probit model: p-value plots using asymptotic critical values for $k = 3$ and $n = 100$



discussion will focus only on this design point, but the findings hold more generally. The full set of results can be found in the Appendix. It is clear from the figure that ω_W and ω_C have an enormous ERP (as is also shown in many previous Monte Carlo studie), e.g. for a 5% level test the actual RP for ω_W is about 28% and for ω_C it is even about 79%. Also their bootstrap-corrected versions ω_{BW} and ω_{BC} have a non-zero ERP, but offer already a significant improvement upon ω_W and ω_C (e.g. at a 5% level, ω_{BW} has an actual RP of about 5.7%, and for ω_{BC} it is about 14%). The only difference between ω_W and ω_C, and ω_{BW} and ω_{BC}, is that the latter use an improved weighting matrix.

One can achieve a much better performance under the null hypothesis by using bootstrap critical values. The statistics are asymptotically pivotal, so

Figure 2: PROBIT MODEL: p-VALUE PLOTS USING BOOTSTRAP CRITICAL VALUES FOR $k = 3$ AND $n = 100$



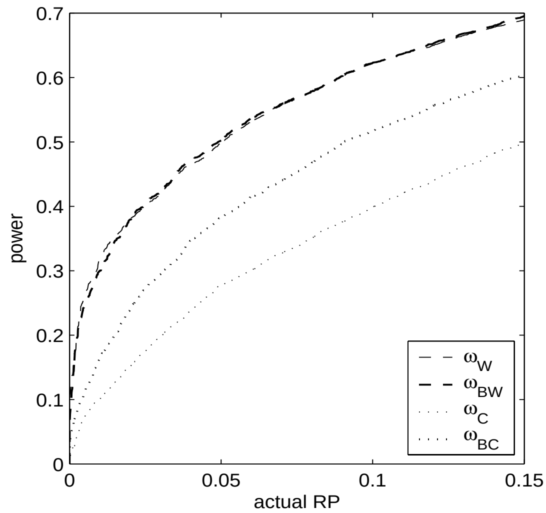
bootstrap critical values are second-order correct. The improvement upon the (first-order) asymptotic approximation is remarkable, as can be seen in Figure 2(a), where the p-value plots are given for ω_W and ω_C using bootstrap critical values (with 499 bootstrap simulations), and with critical values from the bootstrap recycling method (also with $B = 499$; see p. 6) for ω_{BW} and ω_{BC}.

The power is investigated against the following heteroskedastic alternative:

$$P(y_t = 1) = \Phi\left(\frac{x_t'\beta}{\eta(x_t)}\right),$$

where $\eta(x_t) = x_{2t}^2$. To save on CPUtime the number of simulations B to

Figure 3: Probit model: RP-power plots for $k = 3$ and $n = 100$



compute the bootstrap variance is decreased to 100, since now only the bootstrap variance matrix has to be computed, whereas for the experiments under the null also bootstrap p-values had to be computed.

Power is plotted as a function of actual RP under the pseudo-true null (called RP-power curves here), as in Horowitz (1994), Horowitz and Savin (2000) and Davidson and MacKinnon (1996). By using this method, the power of the tests is corrected for the (sometimes large) ERP under the null. Figure 3 plots the RP-power curves for $k = 3$ and $n = 100$. The power of the bootstrap-corrected statistic ω_{BC} is larger than that of ω_C, but the power of ω_{BW} is about the same as that of ω_W. Let us return to the example of the Jarque-Bera statistic given earlier for a possible explanation of this last observation. The version of the Jarque-Bera test using the very imprecisely estimated weighting matrix is not sensitive for departures from normal kurtosis, but it has power against skewed alternatives. So, if the alternative is only skewed (and thus has normal kurtosis), then there is less or no gain (with respect to power) in correcting the weighting matrix. I suspect that a similar story is true here for White's version of the information matrix test and the particular alternative.

Consider now the linear regression model

$$y_t = x_t'\beta + \sigma\epsilon_t, \quad t = 1, \ldots, n \quad ,$$

with x_t a $k \times 1$-vector of regressors (including 1), parameters β ($k \times 1$) and $\sigma > 0$ (thus $p = k + 1$ and $q = p(p+1)/2 - 1$), and an error term ϵ_t which is i.i.d. and standard normal. In the simulations the regression parameter β is set equal to a vector of ones, and also $\sigma = 1$, but this choice does not affect the results since the statistics are exactly pivotal. The score vector, evaluated at the ML estimate $\hat{\theta} = (\hat{\beta}', \hat{\sigma})$, equals

$$\hat{s} = \frac{1}{\sqrt{n}\hat{\sigma}^2} \sum_{t=1}^{n} \left(\begin{array}{l} (\hat{\epsilon}_t^2 - 1)(x_t x_t')^c \\ (\hat{\epsilon}_t^3 - 3\hat{\epsilon}_t) x_t \\ \hat{\epsilon}_t^4 - 5\hat{\epsilon}_t^2 + 2 \end{array} \right)$$

where $\hat{\epsilon}_t$ is ϵ_t evaluated at $\hat{\theta}$. As shown by Hall (1987), the information matrix test in the linear model is a combined test against heteroskedasticity (first subvector of $\hat{s}$), conditional skewness (second subvector of $\hat{s}$) and non-normal kurtosis (last element of $\hat{s}$).

In this model the information matrix test statistic is exactly pivotal, hence bootstrap critical values are exact. Also the bootstrap variance matrix of the studentized score vector is exact. In Section 2 it was already mentioned that $\hat{V}^{-1/2}$ is not the only possible matrix which makes the score vector asymptotically, or in this case, exactly pivotal. Given that the standardised residuals $\hat{\epsilon}_t$ are invariant with respect to θ, it suffices to multiply $\hat{s}$ by $\hat{\sigma}^2$ to make it exactly pivotal.

So, in the simulation study for the normal linear model, not only ω_W, ω_C and their bootstrap-corrected versions ω_{BW} and ω_{BC} are included, but also $\omega_{BT} = \hat{d}_T' \hat{V}_{BT}^{-1} \hat{d}_T$, where $\hat{d}_T = \hat{\sigma}^2 \hat{s}$ and $\hat{V}_{BT}$ is the bootstrap variance matrix of $\hat{d}_T$.

Figure 4 shows the p-value plots when asymptotic critical values are used. Again, the poor performance of ω_W and ω_C, and the improvement offered by ω_{BW} and ω_{BC} are obvious. Overall ω_{BT} is found to have the smallest ERP.

Figure 4: Linear model: p-value plots using asymptotic critical values for $k = 3$ and $n = 100$

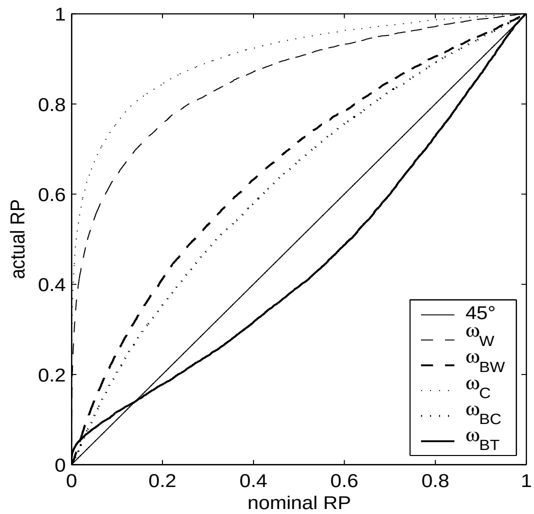
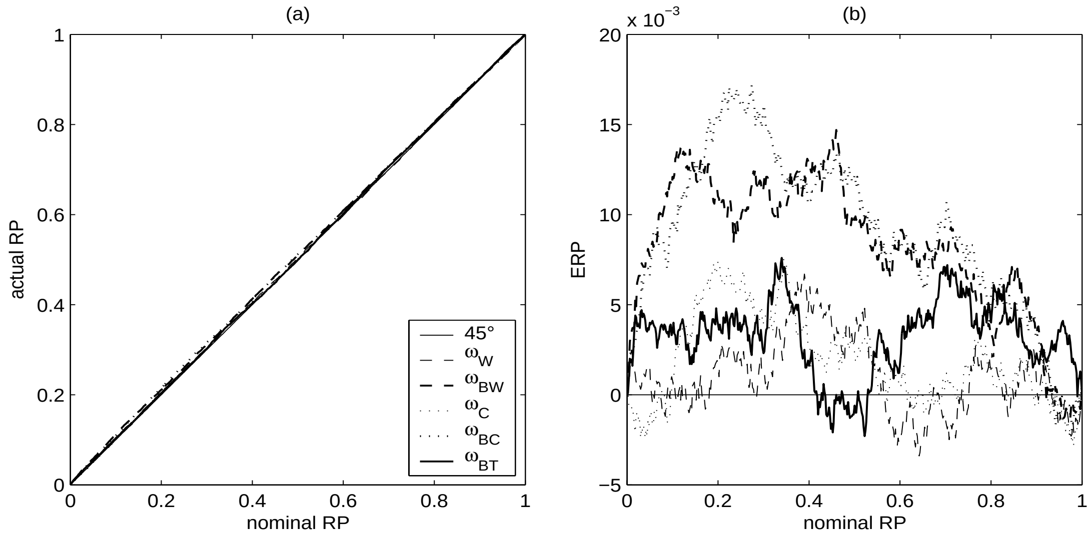


Figure 5: LINEAR MODEL: p-VALUE PLOTS USING BOOTSTRAP CRITICAL VALUES FOR $k = 3$ AND $n = 100$



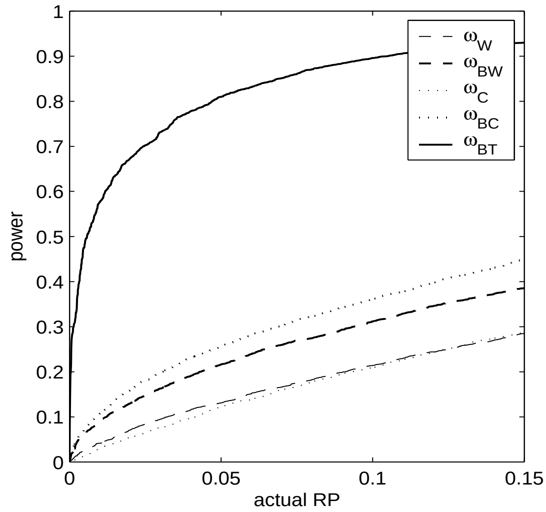
In the normal linear model the IM statistics are exactly pivotal, meaning that bootstrap critical values are exact. All tests have now an ERP which is zero for $B = \infty$, but the ERP is already very small for $B = 499$, as is evidenced by Figure 5(b).

The power of the test is studied against a heteroskedastic alternative with density $\phi((y_t - x_t'\beta)/(\sigma\eta(x_t))$, where $\eta(x_t) = \sqrt{|\,x_{2t}\,|}$ [4]. Figure 6 shows that ω_{BC} has more power than ω_C. Here, also the corrected White statistic ω_{BW} has a larger power than ω_W. The test based on ω_{BT}, however, has undeniably the largest power.

Thus, the bootstrap-corrected score test, which uses an improved weight-

[4]The tests in the normal linear model seemed to be somewhat more powerful against heteroskedasticity than in the probit model, therefore the alternative is weakened here.

Figure 6: LINEAR MODEL: RP-POWER PLOTS FOR $k = 3$ AND $n = 100$



ing matrix and more accurate critical values, outperforms the standard score test in terms of both ERP and power. The test based on ω_{BT}, which uses an almost trivial transformation of the score vector, is found to be the best in terms of ERP and power (at least in this Monte Carlo set-up), but such a statistic may not exist for all models and all score tests.

4 Conclusion

The usual bootstrap correction for score tests focusses on the critical value. In this paper it is argued that not only the critical value should be corrected, but also the weighting matrix used in the quadratic form statistic of the score test. By transforming the score vector to make it asymptotically (or exactly) pivotal, the bootstrap variance matrix is second-order correct (or exact). Then by re-using the (parametric) bootstrap simulations, a second-order correct (or exact) critical value is obtained for the corrected statistic. The Monte Carlo experiments show that the method indeed provides an improvement upon the standard score test.

References

[1] Aitchison, J. and Silvey, S.D. (1958) Maximum-likelihood estimation parameters subject to restraints. *Annals of Mathematical Statistics* 29, 813-828.

[2] Aparicio, T. and Villanua, I. (2001) The asymptotically efficient version of the information matrix test in binary choice models: A study of size and power. *Journal of Applied Statistics* 28, 167-182.

[3] Bera, A.K. and Jarque, M. (1981) Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters* 7, 313-318.

[4] Beran, R. (1988) Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687-697.

[5] Breusch, T.S. and Pagan, A.R. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47, 1287-1294.

[6] Chesher, A. (1983) The information matrix test: simplified calculation via a score test interpretation. *Economics Letters* 13, 45-48.

[7] Chesher, A. (1984) Testing for neglected heterogeneity. *Econometrica* 52, 865-872.

[8] Chesher, A. and Spady, R. (1991) Asymptotic expansions of the information matrix test. *Econometrica* 59, 787-815.

[9] Davidson, R. and MacKinnon, J.G. (1983), Small sample properties of alternative forms of the Lagrange multiplier test. *Economics Letters* 12, 269-275.

[10] Davidson, R. and MacKinnon, J.G. (1984) Convenient specification tests for logit and probit models. *Journal of Econometrics*, 25, 241-262.

[11] Davidson, R. and MacKinnon, J.G. (1992) A new form of the information matrix test. *Econometrica* 60, 145-157.

[12] Davidson, R. and MacKinnon, J.G. (1996) The power of bootstrap tests. Queen's Institute for Economic Research Working Paper, n° 937, forthcoming in Journal of Econometrics.

[13] Davidson, R. and MacKinnon, J.G. (1998) Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School* 66, 1-26.

[14] Dhaene, G. and Hoorelbeke, D. (2004) The information matrix test with bootstrap-based covariance matrix estimation. *Economics Letters* 82, 341-347.

[15] Godfrey, L.G. (1990) Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches. *Econometric Society Monographs* 16, Cambrigde University Press: Cambridge.

[16] Godfrey, L.G. and Orme, C.D. (2001) On improving the robustness and reliability of Rao's score test. *Journal of Statistical Planning and Inference* 97, 153-176.

[17] Hall, A. (1987) The information matrix test for the linear model. *The Review of Economic Studies* 54, 257-263.

[18] Hall, P. (1992) The bootstrap and Edgeworth expansion. *Springer Series in Statistics*, Springer-Verlag: New York.

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

[20] Horowitz, J.L. (2001) The Bootstrap in *Handbook of Econometrics*, Vol. 5 Chapter 52, edited by Heckman, J.J. and E.E. Leamer, Amsterdam: Elsevier.

[21] Horowitz, J.L. and Savin, N.E. (2000) Empirically relevant critical values for hypothesis tests: A bootstrap approach. *Journal of Econometrics* 95, 375-389.

[22] Jarque, C.M. and Bera, A.K. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6, 255-259.

[23] Lancaster, T. (1984) The covariance matrix of the information matrix test. *Econometrica* 52, 1051-1053.

[24] Orme, C. (1990) The small-sample performance of the information-matrix test. *Journal of Econometrics* 46, 309-331.

[25] Rao, C.R. (1948) Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of Cambridge Philosophical Society* 44, 50-57.

[26] White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-26.

Appendix

Figure 7: Probit model: p-value plots using asymptotic critical values for $k = 2$ and (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; and (d) $n = 400$

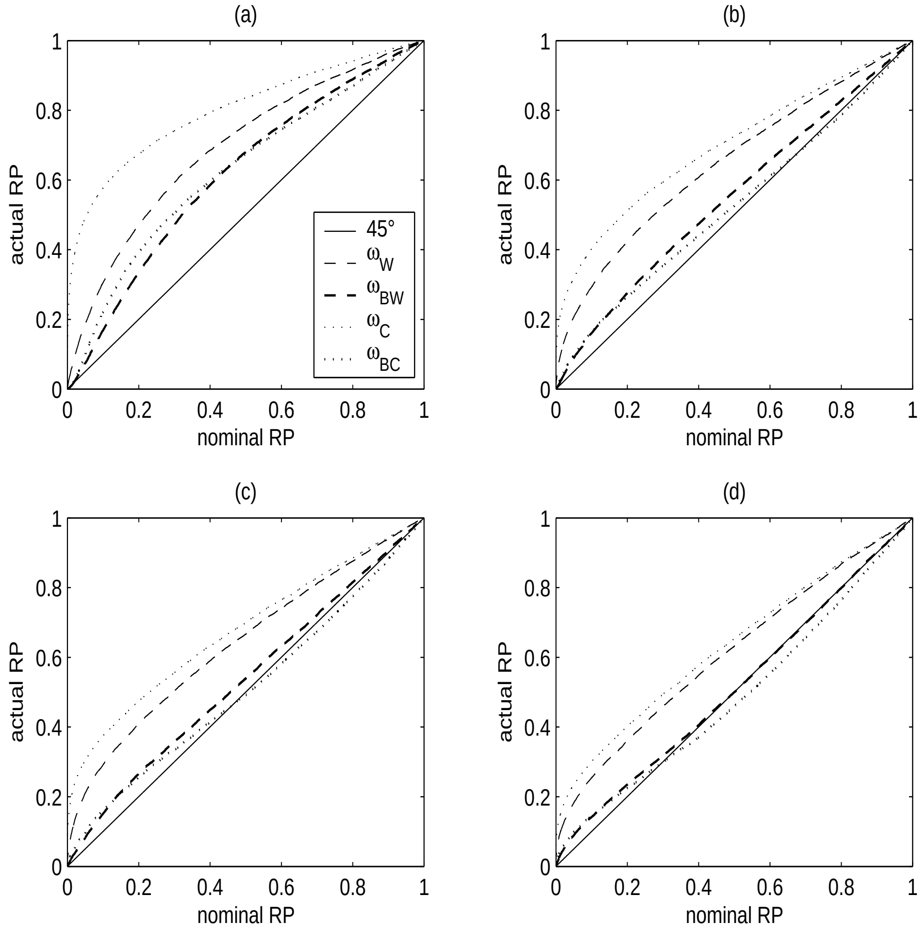


Figure 8: Probit model: p-value plots using asymptotic critical values for $k = 3$ and (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; and (d) $n = 400$

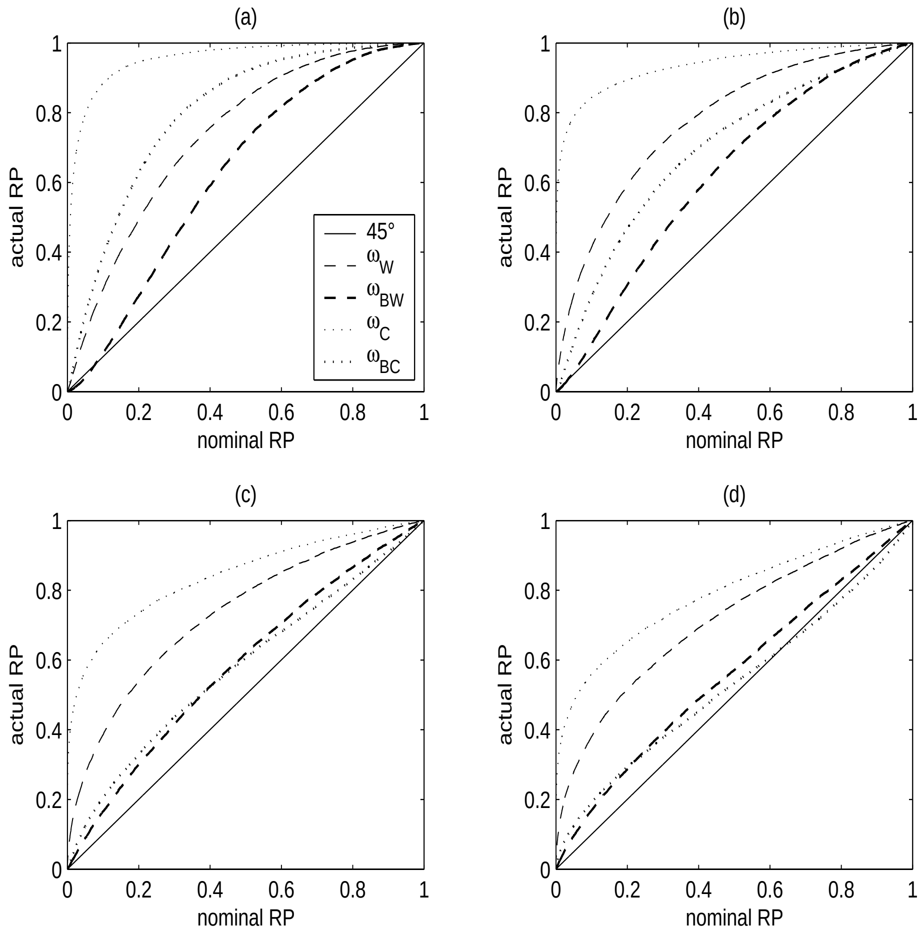


Figure 9: Probit model: ERP plots using bootstrap critical values for $k = 2$ and (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; and (d) $n = 400$

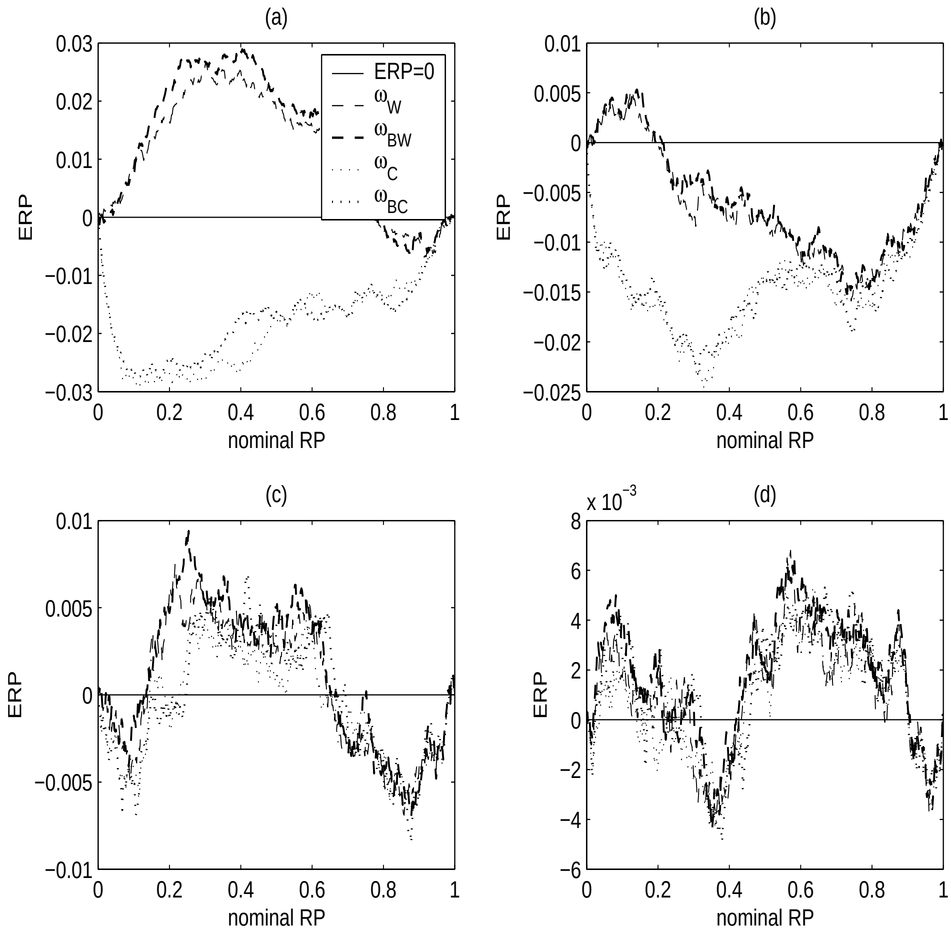


Figure 10: PROBIT MODEL: ERP PLOTS USING BOOTSTRAP CRITICAL VALUES FOR $k = 3$ AND (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; AND (d) $n = 400$

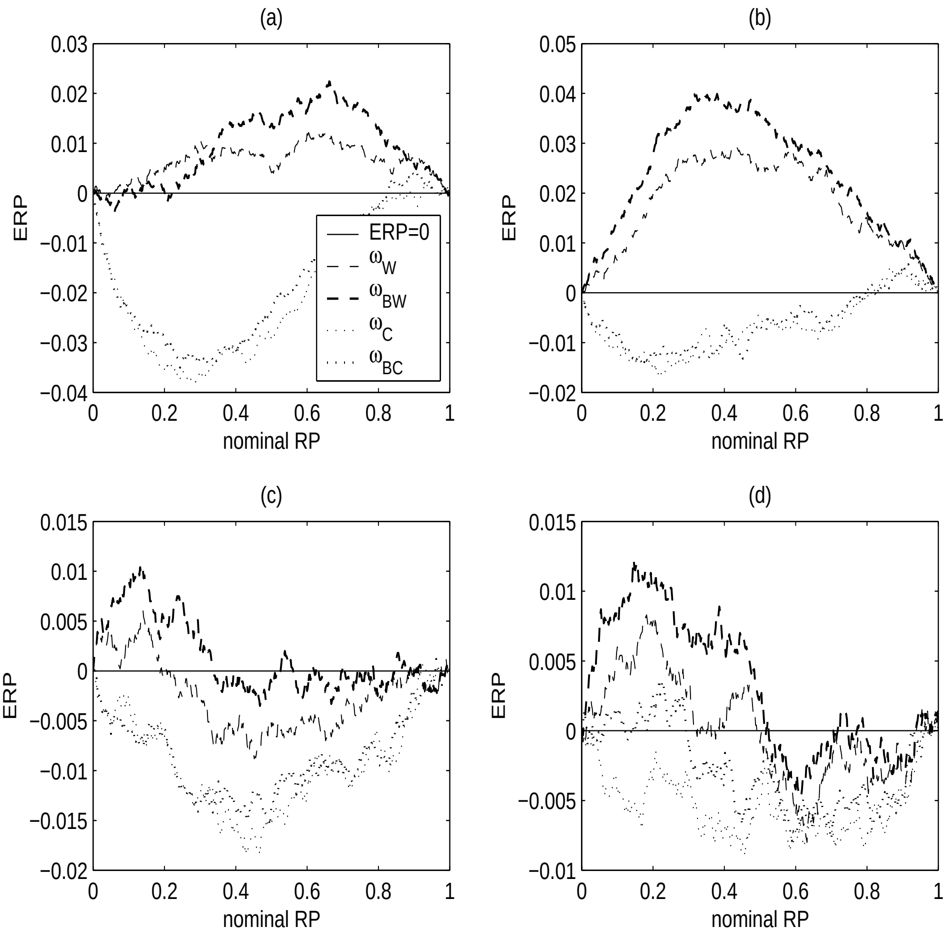


Figure 11: Probit model: RP-Power plots for $k = 2$ and (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; and (d) $n = 400$

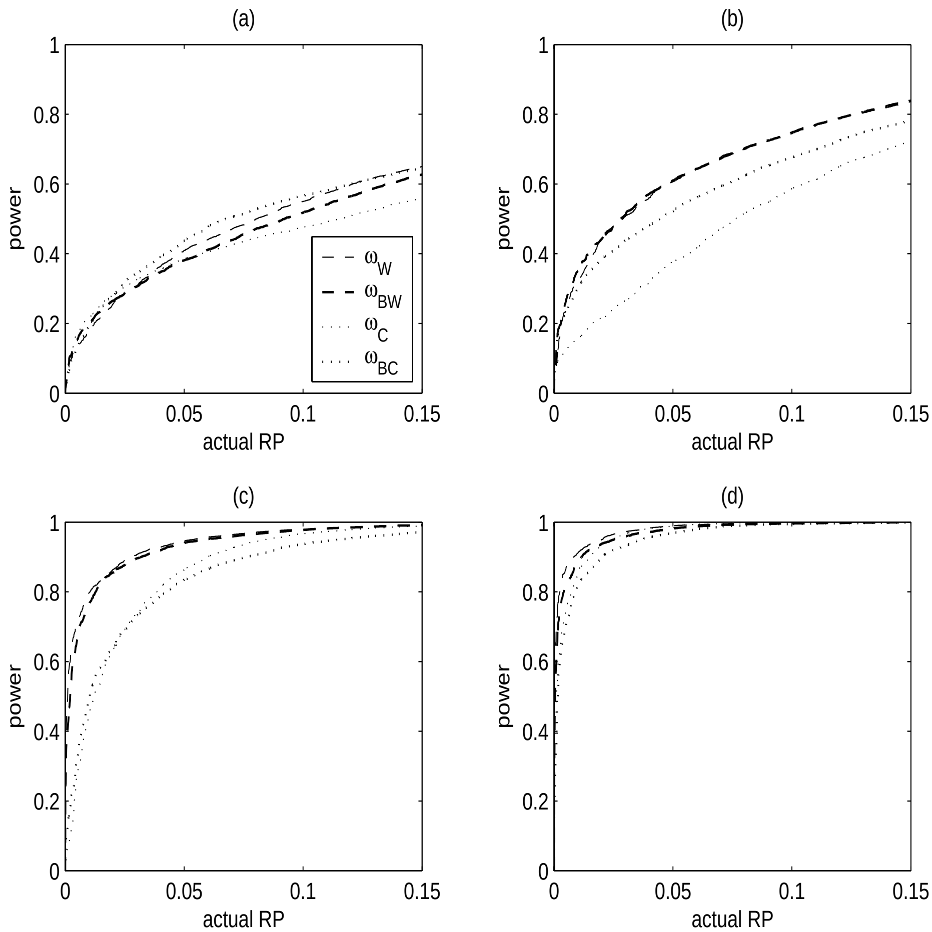


Figure 12: Probit model: RP-Power plots for $k = 3$ and (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; and (d) $n = 400$

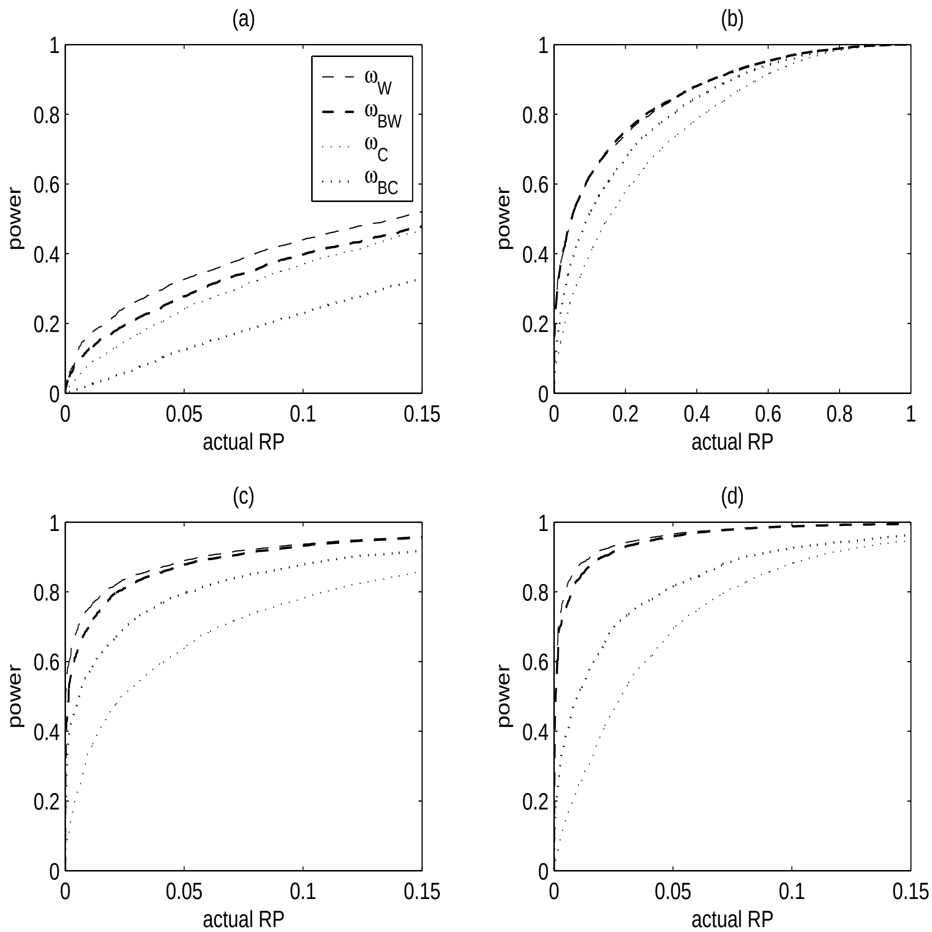


Figure 13: LINEAR MODEL: p-VALUE PLOTS USING ASYMPTOTIC CRITICAL VALUES FOR $k = 2$ AND (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; AND (d) $n = 400$

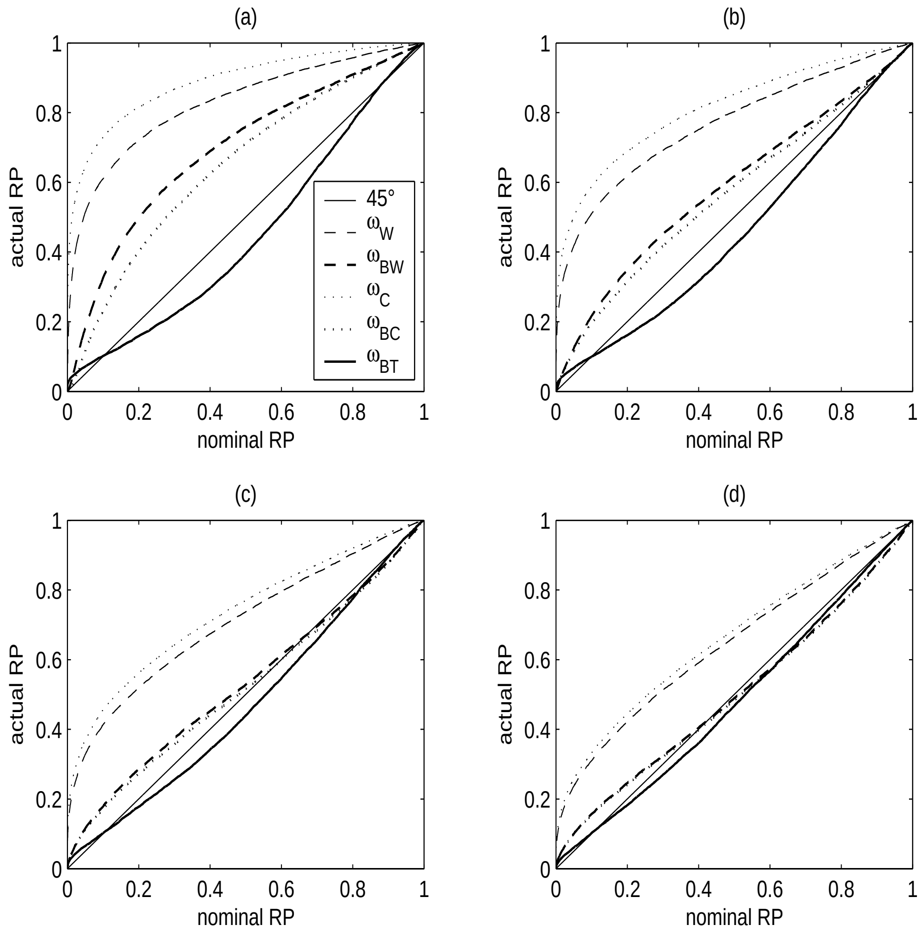


Figure 14: Linear model: p-value plots using asymptotic critical values for $k = 3$ and (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; and (d) $n = 400$

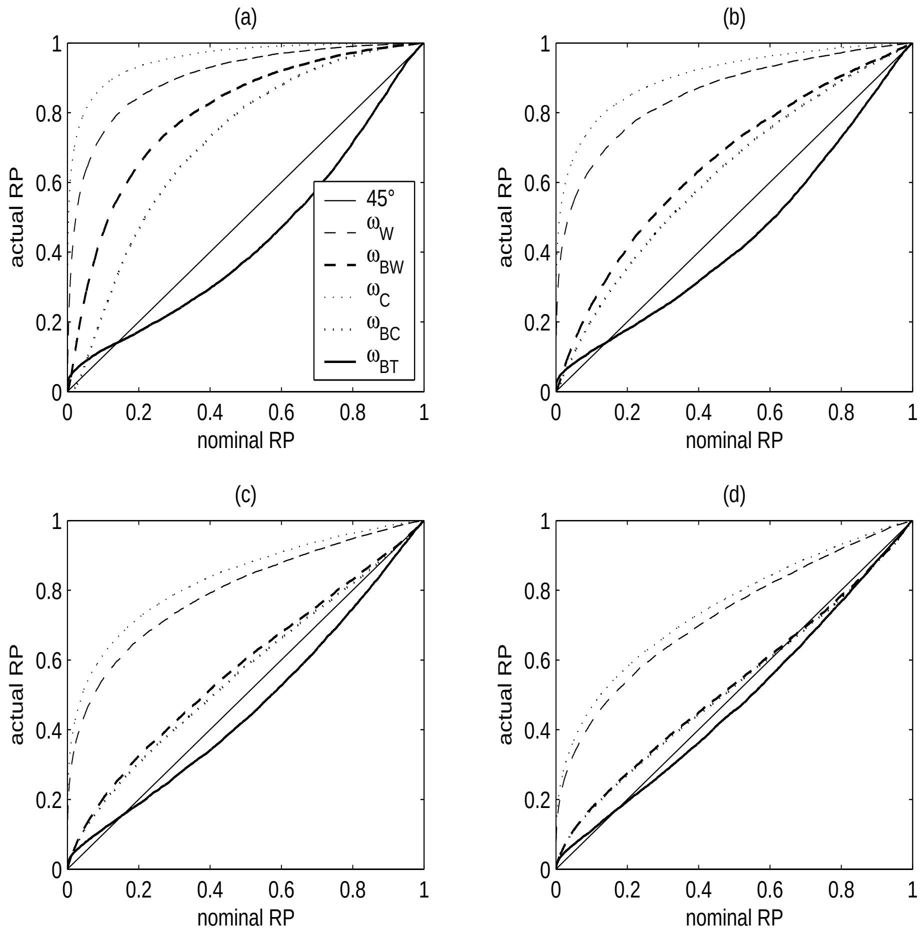


Figure 15: LINEAR MODEL: ERP PLOTS USING BOOTSTRAP CRITICAL VALUES FOR $k = 2$ AND (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; AND (d) $n = 400$

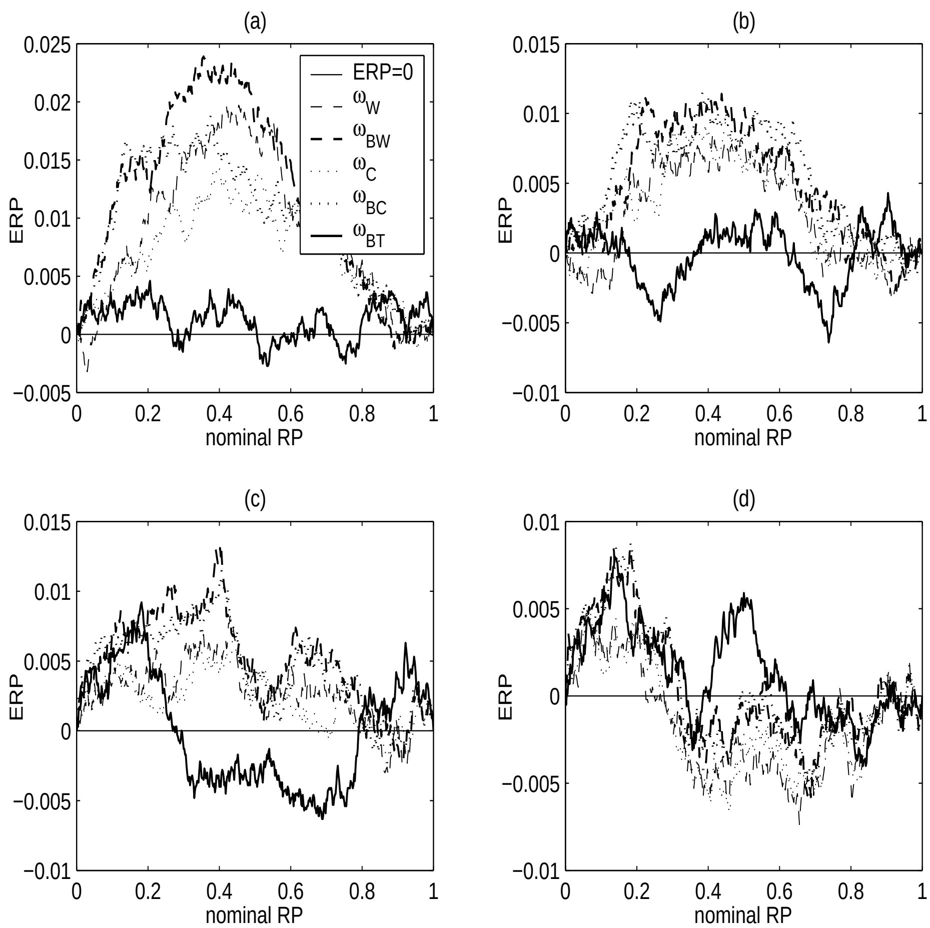


Figure 16: Linear model: ERP plots using bootstrap critical values for $k = 3$ and (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; and (d) $n = 400$

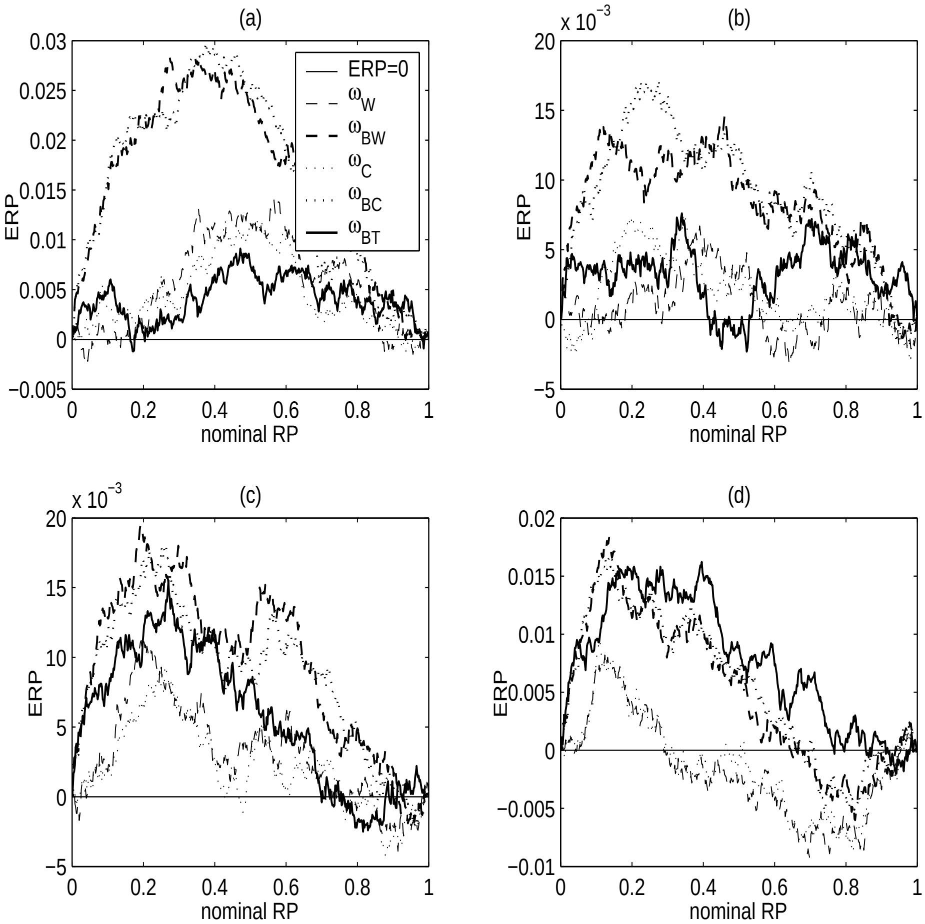


Figure 17: LINEAR MODEL: RP-POWER PLOTS FOR $k = 2$ AND (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; AND (d) $n = 400$

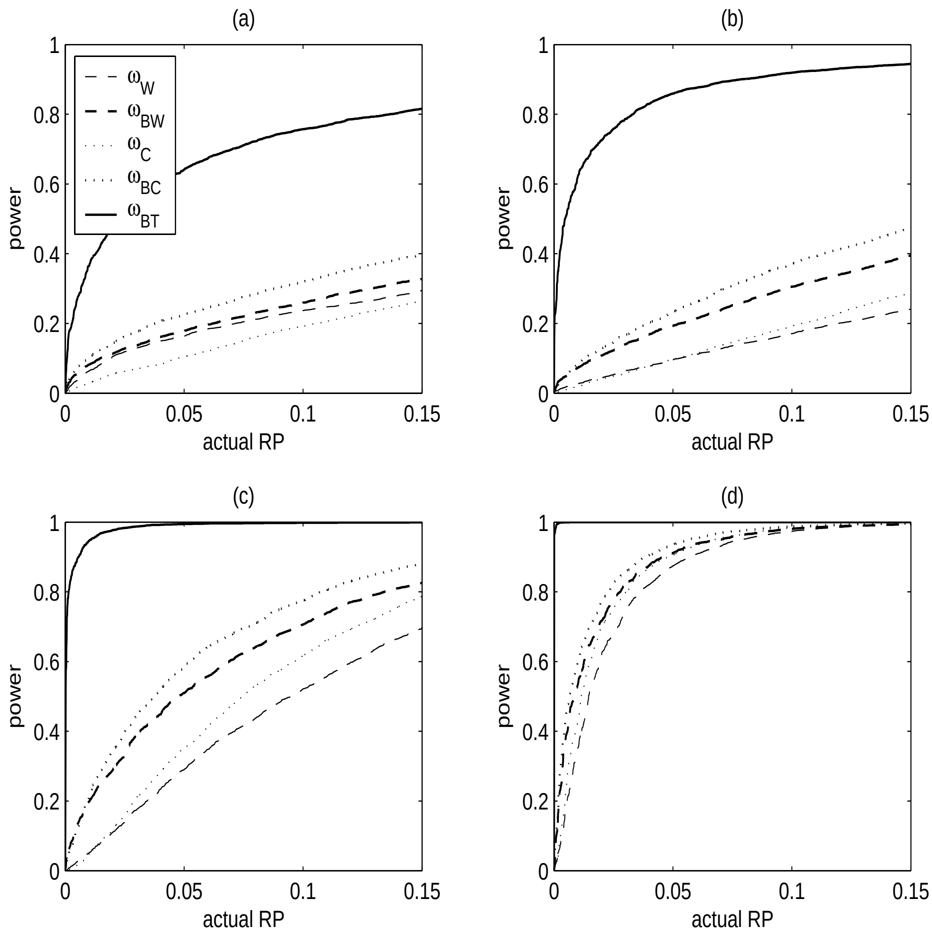


Figure 18: LINEAR MODEL: RP-POWER PLOTS FOR $k = 3$ AND (a) $n = 50$; (b) $n = 100$; (c) $n = 200$; AND (d) $n = 400$