Bootstrap test for a structural break under possible heteroscedasticity

Namba, Akio

(Citation)
Communications in Statistics - Simulation and Computation,46(5):4127-4139

(Issue Date)
2017

(Resource Type)
journal article

(Version)
Accepted Manuscript

Bootstrap Test for a Structural Break under Possible Heteroscedasticity

Akio Namba*

Graduate School of Economics
Kobe University
Rokko, Nada-ku, Kobe 657-8501
Japan

Abstract

In this paper we consider the Wald test statistic proposed by Watt (1979) for testing equality between the sets of regression coefficients in two linear regression models when the disturbance variances may possibly be unequal. This test can be also used as a test for a structural break. As shown by Ohtani and Toyoda (1985) and Honda and Ohtani (1986), the test based on the Wald test statistic suffers from severe size distortion in small sample when the disturbance variances of the two regression models are unequal. Our simulation results show that substantial improvements are made when the bootstrap methods are applied.

1 Introduction

The test proposed by Chow (1960) has been widely used to test equality between sets of coefficients in two linear regression models, or to test the existence of a structural break in a regression model. However, it is well known that the Chow test suffers from poor performance if the regression model is heteroscedastic, or the disturbance variances of the two linear regression models are unequal. [See Toyoda (1974) and Schmidt and Sickles (1977).]

To tackle this drawback of the Chow test, several authors proposed alternative testing procedures which are applicable when the disturbance terms are heteroscedastic. Some examples are

*The author is grateful to Kazuhiro Ohtani for his helpful comments and suggestions. He also would like to thank an anonymous referee who provided various comments which greatly improve the previous version of the paper. This work was supported by JSPS KAKENHI Grant Number 23243038, 26780136.

Watt (1979), Jayatissa (1977) and Rothenberg (1984). See also Thursby (1992) for comparisons of several testing methods and their performances. In particular, Watt (1979) proposed the test based on the Wald test statistic. According to Ohtani and Kobayashi (1986) and Thursby (1992), Watt's (1979) test is more powerful than Jayatissa's (1977) test.

Though the test statistic proposed by Watt (1979) is easy to compute, its exact distribution is very complex [See Kobayashi (1986) and Phillips (1986)]. Thus, Watt (1979) proposed to use critical values of a chi-squared distribution based on its asymptotic distribution. However, if we use the critical values of a chi-squared distribution, the test proposed by Watt (1979) suffers from size distortion when the sample size is small. See, e.g., Ohtani and Toyoda (1985) and Honda and Ohtani (1986). In order to avoid this size distortion, Ohtani and Kobayashi (1986) and Kobayashi (1986) proposed bounds test based on the Wald test statistic. Since their test is based on the upper and lower bounds of the Wald test statistic proposed by Watt (1979), there inherently exists an inconclusive region for the test statistic. Also, Weerahandi (1987) proposed a test which is exact under the normality assumption of disturbances. Though Weerahandi's (1987) test is exact, it requires a numerical integration when calculating the p-value of the test statistic and is not easy to implement.

The above procedures are required since the exact distribution of the Wald test statistic proposed by Watt (1979) is complex when the sample size is small. When the exact distribution of a statistic is complex or unknown, the bootstrap method proposed by Efron (1979) is sometimes useful. In particular, as shown by Beran (1987, 1988) [see also Hall (1992)], the procedure based on the bootstrap methods yields more accurate results than the conventional asymptotic procedure when the statistic is asymptotically pivotal, i.e., the asymptotic distribution of the statistic does not depend on unknown parameters. Since the Wald test statistic is asymptotically distributed as a chi-squared distribution with known degrees of freedom, it is asymptotically pivotal. Therefore, an improvement is expected if the bootstrap method is applied to the Wald test statistic proposed by Watt (1979). Also, Liu (1988) proposed the wild bootstrap method which is second-order correct under heteroscedasticity and Mammen (1993) proposed a procedure which satisfies Liu's (1988) condition.

Recent literature includes models which permit multiple structural breaks and unknown break points and the methods to investigate them. See, e.g., Andrews (1993,) Bai and Perron (1998), Perron (2006), Boldea et al. (2012), Hall et al. (2012), Perron and Yamamoto (2014) and references therein. However, to examine the validity of the proposed methods and for simplicity, we focus on the model with one possible structural break and a known break point. Thus, in this paper, we apply the bootstrap methods to the test statistic proposed by Watt (1979). We examine the size and the power of the bootstrap test by Monte Carlo simulations. The organization of the paper is as follows. In the next section, we introduce the model and the test statistic. Also the ways to apply the bootstrap methods to the test statistic are explained. It turns out that the bootstrap procedure gets simplified because of the structure of the test statistic. In section 3, we examine the performance of the bootstrap test by simulations. The simulation results show that the size distortion and of Watt's (1979) test is substantially improved by the bootstrap methods. It is also shown that the power of the test can be improved by the wild bootstrap methods. Finally, some concluding remarks are given in section 4.

2 Model, test statistic and the bootstrap methods

Consider two linear regression models

$$y_i = X_i \beta_i + \epsilon_i, \quad i = 1, 2, \tag{1}$$

where y_i is an $n_i \times 1$ vector of observations on a dependent variable, X_i is an $n_i \times k$ matrix of observations on nonstochastic explanatory variables, β_i is a $k \times 1$ vector of coefficients, and ϵ is an $n_i \times 1$ vector of error terms and $\epsilon_i \sim \mathrm{N}(0, \sigma_i^2 I_{n_i})$. Also, we assume that X_i is of full column rank.

The task considered in this paper is to test the null hypothesis $\mathrm{H}_0 : \beta_1 = \beta_2$ against the alternative $\mathrm{H}_1 : \beta_1 \neq \beta_2$. If i denotes the regime, accepting H_0 implies that there is no structural break.

If $\sigma_1^2 = \sigma_2^2$, i.e., the disturbance variances of the two regression models are equal, we can easily test the null hypothesis using the Chow test proposed by Chow (1960). However, as shown by Toyoda (1974) and Schmidt and Sickles (1977), the Chow test has a very poor performance when $\sigma_1^2 \neq \sigma_2^2$. Thus, Watt (1979) proposed the Wald test statistic which takes the heteroscedasticity into consideration:

$$W = (b_1 - b_2)' \left[s_1^2 (X_1' X_1)^{-1} + s_2^2 (X_2' X_2)^{-1} \right]^{-1} (b_1 - b_2), \tag{2}$$

where b_i and s_i^2 are the least squares estimator of β_i and σ_i^2. Though this Wald test statistic is asymptotically valid, as shown by Ohtani and Toyoda (1985) and Honda and Ohtani (1986), the test based on this statistic suffers from size distortion in small samples if the critical values of a chi-squared distribution are used. One way of coping with the size distortion is executing the test based on the the upper and lower bounds of the Wald test statistic as proposed by Ohtani and Kobayashi (1986) and Kobayashi (1986). However, this testing procedure inherently includes the inconclusive region. Also, Weerahandi (1987) proposed a test which is exact under normality of the disturbance. However, Weerahandi's (1987) test requires numerical integration when calculating the p-value of the test statistic and is not easy to implement.

Thus, in this paper, we consider a more direct method, i.e., the bootstrap method proposed by Efron (1979). As shown by Beran (1987, 1988), inferences based on asymptotic distributions can be improved by applying the bootstrap if the statistic considered is asymptotically pivotal, i.e., the asymptotic distribution of the statistic does not depend on unknown parameters. Since the asymptotic distribution of the Wald test statistic given in (2) is a chi-squared distribution with k degrees of freedom, it is asymptotically pivotal. Thus, by applying the bootstrap method to W, a reduction in the size distortion of the test is expected.

The application of the bootstrap method to W is summarized as follows:

1. Estimate β_i and σ_i^2 by the ordinary least squares (OLS) method and obtain b_i and s_i^2. Calculate the value of the Wald test statistic W given in (2).

2. Let $e_i = y_i - X_i b_i$ be the residual vector for $i = 1, 2$. Following Wu (1986), we first rescale the residual vector as $\tilde{e}_i = \sqrt{n_i/(n_i - k)}\, e_i$. Drawing a sample of size n_i from the elements of the rescaled residual with replacement and stacking them as an $n_i \times 1$ vector, we obtain a bootstrap sample vector e_i^* for $i = 1, 2$.

3. Regressing e_i^* on X_i, obtain bootstrap estimates b_i^* and s_i^{2*} for $i = 1, 2$. Using these estimates, calculate the bootstrap version of the Wald test statistic:

$$W^* = (b_1^* - b_2^*)' \left[s_1^{2*}(X_1'X_1)^{-1} + s_2^{2*}(X_2'X_2)^{-1}\right]^{-1} (b_1^* - b_2^*). \tag{3}$$

4. Repeating the steps 2 and 3 above B times, and calculating the ratio that W^* exceeds W, we obtain the p-value of the test based on the bootstrap method. Thus, if the obtained p-value is less than the assigned significance level, $\mathrm{H}_0 : \beta_1 = \beta_2$ is rejected.

We call the above procedure the ordinary bootstrap method. In order to cope with the heteroscedasticity, the bootstrap sample vector is e_i^* is constructed from the residuals obtained by regressing subsample y_i on X_i in step 2 above. Note that, in step 3 above, we simply regress e_i^* on X_i in order to obtain bootstrap estimates. In the ordinary bootstrap procedure for a regression model, we usually calculate a bootstrap sample of the dependent variable $y_i^* = X_i \bar{\beta}_i + e_i^*$

where $\bar{\beta}_i$ is any estimator of β_i, and obtain bootstrap estimates by regressing y_i^* on X_i. Since, when testing a null hypothesis, a bootstrap sample must be drawn from a model such that the null hypothesis is hold, we need to use an estimator which satisfies $\bar{\beta}_1 = \bar{\beta}_2 = \bar{\beta}$. However, if we regress y_i^* instead of e_i^* under the condition $\bar{\beta}_1 = \bar{\beta}_2 = \bar{\beta}$, we obtain

$$\begin{aligned} b_1^* - b_2^* &= (X_1'X_1)^{-1}X_1'y_1^* - (X_2'X_2)^{-1}X_2'y_2^* \\ &= (X_1'X_1)^{-1}X_1'e_1^* - (X_2'X_2)^{-1}X_2'e_2^*. \end{aligned} \tag{4}$$

This implies that the bootstrap version of the Wald test statistic W^* given in (3) is independent of the choice of $\bar{\beta}$ and that the value of W^* is unchanged whatever estimate $\bar{\beta}$ may be used. Thus, by using the zero vector as $\bar{\beta}$, we can simply regress e_i^* on X_i, and the bootstrap resampling gets simplified because of the structure of the Wald test statistic W.

We also consider to apply the wild bootstrap method proposed by Liu (1988) since the wild bootstrap is second-order correct under heteroscedasticity[1]. In the wild bootstrap method, step 2 in the ordinary bootstrap procedure explained above is modified as follows.

$2'$ Similar to the case of ordinary bootstrap, we first rescale the residuals as $\tilde{e}_{ij} = \sqrt{n_i/(n_i - k)}\, e_{ij}$, where $\tilde{e}_{ij}$ and e_{ij} are the jth elements of e_i and $\tilde{e}_i$. Drawing random samples z_j ($j = 1, 2, \ldots, n_i$) from a pick distribution defined below, and let $e_{ij}^* = z_j \tilde{e}_{ij}$ where e_{ij}^* is the jth element of e_i^*.

In the wild bootstrap methods, residuals can be calculated from either restricted or unrestricted estimator. In the above procedure, the residuals are calculated based on unrestricted estimator (i.e., $b_1 \neq b_2$). When we use the restricted estimator, e_i and n_i in step $2'$ are replaced by e and $n = n_1 + n_2$, where $e = y - Xb$, $b = (X'X)^{-1}X'y$, $y = [y_1', y_2']'$ and $X = [X_1', X_2']'$. Thus, the wild bootstrap resampling based on the unrestricted estimator can be considered as the method based on the subsamples, while the resampling based on the restricted estimator can be considered as the one based on the whole sample.

Liu (1988) showed that the wild bootstrap is second-order correct under heteroscedasticity if the random draw z_i from a pick distribution satisfies $\mathrm{E}[z_i] = 0$, $\mathrm{E}[z_i^2] = 1$ and $\mathrm{E}[z_i^3] = 1$. Thus, Mammen (1993) proposed to use the following distribution as a pick distribution in the wild bootstrap:

$$\mathrm{P}\left(z_i = \frac{1-\sqrt{5}}{2}\right) = \frac{\sqrt{5}+1}{2\sqrt{5}} \text{ and } \mathrm{P}\left(z_i = \frac{1+\sqrt{5}}{2}\right) = 1 - \frac{\sqrt{5}+1}{2\sqrt{5}}. \tag{5}$$

[1]The author is grateful to an anonymous referee who proposed to apply the wild bootstrap method.

It is easy to show that z_i satisfies the condition in Liu (1988). Also, Davidson and Flachaire (2008) proposed to use the following pick distribution:

$$\mathrm{P}(z_i = 1) = \mathrm{P}(z_i = -1) = \frac{1}{2}. \tag{6}$$

A random sample from this distribution does not satisfy the condition in Liu (1988) since $\mathrm{E}[z_i^3] = 0$. However, Flachaire's (2005) simulation results show that this pick distribution yields better performance than the pick distribution proposed by Mammen (1993).

3 Simulation results

In this section, we examine the performance of the bootstrap tests introduced in section 2 by Monte Carlo simulations. As for the pick distribution in the wild bootstrap, we use both of the distributions proposed by Mammen (1993) and Davidson and Flachaire (2008). Also, as stated in the previous section, the residuals based on both unrestricted estimator (i.e., b_1 and b_2) and restricted estimator (i.e., b) can be utilized in the wild bootstrap. To see the effect of the restriction, we executed simulations for both cases. The design of the simulation is as follows:

1. For simplicity, we assume $k = 2$ and $x_{ij} = [1, u_j]$, where x_{ij} is the jth element of X_i and u_j is a random sample drawn from $\mathrm{U}[0, 1]$. Thus, the regression model has an intercept and one explanatory variable.

2. Using the number of iteration of resampling in bootstrap $B = 1000$, and letting $\beta_{11} = \beta_{12} = \sigma_1 = 1$, and β_{21}, β_{22}, σ_2, n_1, n_2 = several values, where β_{ij} is the jth element of β_i, we iterated the procedure explained in the previous section $M = 100000$ times and test the null hypothesis at $\alpha = 0.10$ (10%), 0.05 (5%) and 0.01 (1%) significance levels. Calculating the ratio when the null hypothesis is rejected out of $M = 100000$ times, we obtain the empirical power of the test.

Through our simulations, we found that the size distortion of the Wald test which utilized critical values is severe when the differences between σ_1, n_1 and σ_2, n_2 are large. This coincides with the results in Ohtani and Toyoda (1985). Though we do not show all the results, the results shown here are typical ones obtained by our experiments.

Table 1a shows the empirical sizes (i.e., the empirical power for $\beta_{11} = \beta_{21}$ and $\beta_{12} = \beta_{22}$) of the ordinary bootstrap test and the Wald test based on asymptotic critical values (i.e., critical values from a chi-squared distribution) for the case of $n_1 = 10$ and $n_2 = 50$. Hereafter, we call the latter asymptotic test for simplicity. To evaluate the accuracy of the tests, we test the null hypothesis H_0' that the size of the test is equal to the assigned significance level α by means of the normal approximation of a binomial distribution. $*$, $\dagger$ and $\ddagger$ denote that the null hypothesis is rejected at the 10%, 5% and 1% significance levels, respectively.

We can also see from Table 1a that the size distortion of asymptotic test is very severe. H_0' is rejected at 0.01% significance level in all cases considered here. Note that severe size distortion exists even when the disturbance variance is not heteroscedastic (i.e., $\sigma_2 = 1.0$). We can also see that the bootstrap test yields more reliable results than the asymptotic test. There are cases where H_0' can not be rejected. Though H_0' is rejected in some cases, the bootstrap test yield empirical sizes much closer to the nominal significance level than the asymptotic test even in such cases.

Table 1b and Table 1c show the results for the wild bootstrap when the pick distributions proposed by Mammen (1993) and Davidson and Flachaire (2008) are used, respectively. From these tables, we can see that the choice of the pick distribution does not affect the sizes of the wild bootstrap tests considered in this paper. Also, we can see that the results based on the restricted estimator are preferable than the ones based on unrestricted estimator. This may be caused by the fact that the restricted estimator is calculated using the whole sample while the unrestricted estimator is calculated using the subsample. These tables indicates that the sizes of the wild bootstrap tests are not so accurate as the ordinary bootstrap test even when the restricted estimator is used in wild bootstrap tests.

Tables 2a–2c show the empirical sizes for $n_1 = 20$ and $n_2 = 50$. Comparing Tables 1a-1c and 2a–2c, we can see that the size distortions of both tests decrease as n_1 increases. Similar to the results in Tables 1a–1c, we can see that the size of the ordinary bootstrap test is most accurate among the proposed tests, and that the effect of the choice of the pick distribution in the wild bootstrap tests is small. Also, wild bootstrap tests based on the restricted estimator yield preferable sizes than the ones based on the unrestricted estimator. Though we do not show other results, we obtained the similar results for the other cases. Also, all our simulation results show that the size distortions of all the tests decrease as n_1 and n_2 increase and that the ordinary bootstrap test is most reliable when n_1 and n_2 is small.

Also, through our simulations, we found that the changes in β_{21} and β_{22} have similar effects on the empirical power of the tests. Thus, to investigate the power of the tests, we calculate the empirical power of the tests for various values of β_{21}, while fixing the other parameters

Table 1a: Empirical sizes of asymptotic and bootstrap tests
at 10%, 5% and 1% significance levels for $n_1 = 10$ and $n_2 = 50$.

	Asymptotic test			Ordinary bootstrap		
σ_2	10%	5%	1%	10%	5%	1%
0.1	$0.16074^{\ddagger}$	$0.10576^{\ddagger}$	$0.04583^{\ddagger}$	$0.09065^{\ddagger}$	$0.04187^{\ddagger}$	$0.00822^{\ddagger}$
0.2	$0.16051^{\ddagger}$	$0.10589^{\ddagger}$	$0.04587^{\ddagger}$	$0.09187^{\ddagger}$	$0.04321^{\ddagger}$	$0.00877^{\ddagger}$
0.3	$0.15943^{\ddagger}$	$0.10352^{\ddagger}$	$0.04440^{\ddagger}$	$0.09249^{\ddagger}$	$0.04492^{\ddagger}$	0.00954
0.4	$0.15838^{\ddagger}$	$0.10294^{\ddagger}$	$0.04351^{\ddagger}$	$0.09557^{\ddagger}$	$0.04694^{\ddagger}$	$0.01078^{\dagger}$
0.5	$0.15609^{\ddagger}$	$0.10058^{\ddagger}$	$0.04172^{\ddagger}$	$0.09679^{\ddagger}$	$0.04824^{\dagger}$	$0.01149^{\ddagger}$
0.6	$0.15226^{\ddagger}$	$0.09681^{\ddagger}$	$0.04000^{\ddagger}$	$0.09604^{\ddagger}$	$0.04863^{\dagger}$	$0.01170^{\ddagger}$
0.7	$0.14952^{\ddagger}$	$0.09592^{\ddagger}$	$0.03874^{\ddagger}$	$0.09790^{\dagger}$	0.05037	$0.01163^{\ddagger}$
0.8	$0.14838^{\ddagger}$	$0.09485^{\ddagger}$	$0.03663^{\ddagger}$	0.09956	0.05028	$0.01131^{\ddagger}$
0.9	$0.14940^{\ddagger}$	$0.09366^{\ddagger}$	$0.03599^{\ddagger}$	0.10059	$0.05187^{\ddagger}$	$0.01189^{\ddagger}$
1.0	$0.14785^{\ddagger}$	$0.09327^{\ddagger}$	$0.03566^{\ddagger}$	$0.10218^{\dagger}$	$0.05228^{\ddagger}$	$0.01205^{\ddagger}$
1.1	$0.14379^{\ddagger}$	$0.08848^{\ddagger}$	$0.03310^{\ddagger}$	0.09943	0.05037	$0.01165^{\ddagger}$
1.2	$0.14342^{\ddagger}$	$0.08834^{\ddagger}$	$0.03233^{\ddagger}$	0.10024	0.05097	$0.01187^{\ddagger}$
1.3	$0.14179^{\ddagger}$	$0.08737^{\ddagger}$	$0.03274^{\ddagger}$	0.10095	$0.05209^{\ddagger}$	$0.01215^{\ddagger}$
1.4	$0.14177^{\ddagger}$	$0.08682^{\ddagger}$	$0.03276^{\ddagger}$	0.10109	$0.05244^{\ddagger}$	$0.01261^{\ddagger}$
1.5	$0.14026^{\ddagger}$	$0.08662^{\ddagger}$	$0.03197^{\ddagger}$	0.10129	$0.05274^{\ddagger}$	$0.01194^{\ddagger}$
1.6	$0.13936^{\ddagger}$	$0.08524^{\ddagger}$	$0.03067^{\ddagger}$	0.10104	$0.05150^{\dagger}$	$0.01190^{\ddagger}$
1.7	$0.14051^{\ddagger}$	$0.08630^{\ddagger}$	$0.03035^{\ddagger}$	$0.10398^{\ddagger}$	$0.05353^{\ddagger}$	$0.01237^{\ddagger}$
1.8	$0.13818^{\ddagger}$	$0.08312^{\ddagger}$	$0.03006^{\ddagger}$	0.10178^{*}	$0.05182^{\ddagger}$	$0.01231^{\ddagger}$
1.9	$0.13878^{\ddagger}$	$0.08396^{\ddagger}$	$0.02969^{\ddagger}$	$0.10296^{\ddagger}$	$0.05284^{\ddagger}$	$0.01270^{\ddagger}$
2.0	$0.13607^{\ddagger}$	$0.08193^{\ddagger}$	$0.02927^{\ddagger}$	0.10119	$0.05181^{\ddagger}$	$0.01192^{\ddagger}$
2.1	$0.13601^{\ddagger}$	$0.08135^{\ddagger}$	$0.02893^{\ddagger}$	0.10068	$0.05186^{\ddagger}$	$0.01220^{\ddagger}$
2.2	$0.13689^{\ddagger}$	$0.08137^{\ddagger}$	$0.02886^{\ddagger}$	0.10116	$0.05225^{\ddagger}$	$0.01274^{\ddagger}$
2.3	$0.13521^{\ddagger}$	$0.08029^{\ddagger}$	$0.02708^{\ddagger}$	0.10043	$0.05198^{\ddagger}$	$0.01196^{\ddagger}$
2.4	$0.13476^{\ddagger}$	$0.08064^{\ddagger}$	$0.02739^{\ddagger}$	$0.10216^{\dagger}$	$0.05200^{\ddagger}$	$0.01215^{\ddagger}$
2.5	$0.13604^{\ddagger}$	$0.08130^{\ddagger}$	$0.02767^{\ddagger}$	$0.10257^{\ddagger}$	$0.05337^{\ddagger}$	$0.01251^{\ddagger}$
2.6	$0.13569^{\ddagger}$	$0.08111^{\ddagger}$	$0.02843^{\ddagger}$	$0.10300^{\ddagger}$	$0.05356^{\ddagger}$	$0.01315^{\ddagger}$
2.7	$0.13498^{\ddagger}$	$0.07870^{\ddagger}$	$0.02657^{\ddagger}$	$0.10197^{\dagger}$	0.05127^{*}	$0.01260^{\ddagger}$
2.8	$0.13288^{\ddagger}$	$0.07799^{\ddagger}$	$0.02689^{\ddagger}$	0.10031	$0.05194^{\ddagger}$	$0.01293^{\ddagger}$
2.9	$0.13428^{\ddagger}$	$0.07992^{\ddagger}$	$0.02702^{\ddagger}$	$0.10308^{\ddagger}$	$0.05287^{\ddagger}$	$0.01244^{\ddagger}$
3.0	$0.13229^{\ddagger}$	$0.07757^{\ddagger}$	$0.02581^{\ddagger}$	0.10134	0.05133^{*}	$0.01233^{\ddagger}$
3.1	$0.13222^{\ddagger}$	$0.07706^{\ddagger}$	$0.02589^{\ddagger}$	0.10124	$0.05155^{\dagger}$	$0.01236^{\ddagger}$
3.2	$0.13157^{\ddagger}$	$0.07689^{\ddagger}$	$0.02523^{\ddagger}$	$0.10197^{\dagger}$	$0.05149^{\dagger}$	$0.01265^{\ddagger}$
3.3	$0.13093^{\ddagger}$	$0.07709^{\ddagger}$	$0.02580^{\ddagger}$	0.10066	$0.05192^{\ddagger}$	$0.01267^{\ddagger}$
3.4	$0.13103^{\ddagger}$	$0.07594^{\ddagger}$	$0.02424^{\ddagger}$	0.10154	0.05121^{*}	$0.01199^{\ddagger}$
3.5	$0.13118^{\ddagger}$	$0.07655^{\ddagger}$	$0.02434^{\ddagger}$	0.10141	0.05128^{*}	$0.01225^{\ddagger}$
3.6	$0.13118^{\ddagger}$	$0.07683^{\ddagger}$	$0.02486^{\ddagger}$	0.10185^{*}	$0.05174^{\dagger}$	$0.01282^{\ddagger}$
3.7	$0.13123^{\ddagger}$	$0.07696^{\ddagger}$	$0.02435^{\ddagger}$	$0.10289^{\ddagger}$	$0.05192^{\ddagger}$	$0.01231^{\ddagger}$
3.8	$0.12994^{\ddagger}$	$0.07597^{\ddagger}$	$0.02442^{\ddagger}$	$0.10259^{\ddagger}$	$0.05232^{\ddagger}$	$0.01251^{\ddagger}$
3.9	$0.12821^{\ddagger}$	$0.07461^{\ddagger}$	$0.02280^{\ddagger}$	0.10071	0.05108	$0.01196^{\ddagger}$

Table 1b: Empirical sizes of Mammen's (1993) wild bootstrap tests at 10%, 5% and 1% significance levels for $n_1 = 10$ and $n_2 = 50$.

	Wild bootstrap (unrestricted)			Wild bootstrap (restricted)		
σ_2	10%	5%	1%	10%	5%	1%
0.1	$0.15130^{\ddagger}$	$0.09703^{\ddagger}$	$0.04069^{\ddagger}$	$0.11245^{\ddagger}$	$0.05548^{\ddagger}$	$0.01308^{\ddagger}$
0.2	$0.15217^{\ddagger}$	$0.09765^{\ddagger}$	$0.04168^{\ddagger}$	$0.11103^{\ddagger}$	$0.05543^{\ddagger}$	$0.01253^{\ddagger}$
0.3	$0.15302^{\ddagger}$	$0.09761^{\ddagger}$	$0.04095^{\ddagger}$	$0.11042^{\ddagger}$	$0.05507^{\ddagger}$	$0.01244^{\ddagger}$
0.4	$0.14962^{\ddagger}$	$0.09552^{\ddagger}$	$0.03953^{\ddagger}$	$0.10714^{\ddagger}$	$0.05286^{\ddagger}$	0.01028
0.5	$0.14897^{\ddagger}$	$0.09465^{\ddagger}$	$0.03812^{\ddagger}$	$0.10583^{\ddagger}$	$0.05161^{\dagger}$	0.00985
0.6	$0.14944^{\ddagger}$	$0.09302^{\ddagger}$	$0.03722^{\ddagger}$	$0.10531^{\ddagger}$	0.05098	$0.00887^{\ddagger}$
0.7	$0.14462^{\ddagger}$	$0.09042^{\ddagger}$	$0.03632^{\ddagger}$	$0.10213^{\dagger}$	0.04892	$0.00811^{\ddagger}$
0.8	$0.14362^{\ddagger}$	$0.08926^{\ddagger}$	$0.03497^{\ddagger}$	0.10110	0.04885^{*}	$0.00739^{\ddagger}$
0.9	$0.14277^{\ddagger}$	$0.08720^{\ddagger}$	$0.03320^{\ddagger}$	0.10078	$0.04653^{\ddagger}$	$0.00674^{\ddagger}$
1.0	$0.13985^{\ddagger}$	$0.08471^{\ddagger}$	$0.03121^{\ddagger}$	0.09850	$0.04526^{\ddagger}$	$0.00614^{\ddagger}$
1.1	$0.13727^{\ddagger}$	$0.08229^{\ddagger}$	$0.02944^{\ddagger}$	$0.09712^{\ddagger}$	$0.04527^{\ddagger}$	$0.00590^{\ddagger}$
1.2	$0.13445^{\ddagger}$	$0.08065^{\ddagger}$	$0.02844^{\ddagger}$	$0.09507^{\ddagger}$	$0.04326^{\ddagger}$	$0.00532^{\ddagger}$
1.3	$0.13256^{\ddagger}$	$0.07800^{\ddagger}$	$0.02702^{\ddagger}$	$0.09382^{\ddagger}$	$0.04212^{\ddagger}$	$0.00496^{\ddagger}$
1.4	$0.13121^{\ddagger}$	$0.07793^{\ddagger}$	$0.02613^{\ddagger}$	$0.09360^{\ddagger}$	$0.04170^{\ddagger}$	$0.00483^{\ddagger}$
1.5	$0.13039^{\ddagger}$	$0.07693^{\ddagger}$	$0.02526^{\ddagger}$	$0.09356^{\ddagger}$	$0.04146^{\ddagger}$	$0.00457^{\ddagger}$
1.6	$0.12794^{\ddagger}$	$0.07379^{\ddagger}$	$0.02402^{\ddagger}$	$0.09236^{\ddagger}$	$0.04030^{\ddagger}$	$0.00466^{\ddagger}$
1.7	$0.12576^{\ddagger}$	$0.07295^{\ddagger}$	$0.02321^{\ddagger}$	$0.09097^{\ddagger}$	$0.03891^{\ddagger}$	$0.00411^{\ddagger}$
1.8	$0.12400^{\ddagger}$	$0.07078^{\ddagger}$	$0.02156^{\ddagger}$	$0.08860^{\ddagger}$	$0.03845^{\ddagger}$	$0.00370^{\ddagger}$
1.9	$0.12154^{\ddagger}$	$0.06796^{\ddagger}$	$0.01969^{\ddagger}$	$0.08662^{\ddagger}$	$0.03603^{\ddagger}$	$0.00360^{\ddagger}$
2.0	$0.12283^{\ddagger}$	$0.06988^{\ddagger}$	$0.02076^{\ddagger}$	$0.08845^{\ddagger}$	$0.03785^{\ddagger}$	$0.00323^{\ddagger}$
2.1	$0.11864^{\ddagger}$	$0.06716^{\ddagger}$	$0.01961^{\ddagger}$	$0.08640^{\ddagger}$	$0.03706^{\ddagger}$	$0.00325^{\ddagger}$
2.2	$0.11907^{\ddagger}$	$0.06549^{\ddagger}$	$0.01831^{\ddagger}$	$0.08617^{\ddagger}$	$0.03535^{\ddagger}$	$0.00292^{\ddagger}$
2.3	$0.11925^{\ddagger}$	$0.06664^{\ddagger}$	$0.01879^{\ddagger}$	$0.08753^{\ddagger}$	$0.03631^{\ddagger}$	$0.00316^{\ddagger}$
2.4	$0.11811^{\ddagger}$	$0.06462^{\ddagger}$	$0.01785^{\ddagger}$	$0.08542^{\ddagger}$	$0.03512^{\ddagger}$	$0.00327^{\ddagger}$
2.5	$0.11354^{\ddagger}$	$0.06346^{\ddagger}$	$0.01775^{\ddagger}$	$0.08304^{\ddagger}$	$0.03504^{\ddagger}$	$0.00275^{\ddagger}$
2.6	$0.11495^{\ddagger}$	$0.06265^{\ddagger}$	$0.01753^{\ddagger}$	$0.08390^{\ddagger}$	$0.03515^{\ddagger}$	$0.00274^{\ddagger}$
2.7	$0.11637^{\ddagger}$	$0.06380^{\ddagger}$	$0.01755^{\ddagger}$	$0.08583^{\ddagger}$	$0.03518^{\ddagger}$	$0.00301^{\ddagger}$
2.8	$0.11397^{\ddagger}$	$0.06239^{\ddagger}$	$0.01641^{\ddagger}$	$0.08444^{\ddagger}$	$0.03402^{\ddagger}$	$0.00278^{\ddagger}$
2.9	$0.11344^{\ddagger}$	$0.06100^{\ddagger}$	$0.01677^{\ddagger}$	$0.08400^{\ddagger}$	$0.03409^{\ddagger}$	$0.00275^{\ddagger}$
3.0	$0.11307^{\ddagger}$	$0.06177^{\ddagger}$	$0.01632^{\ddagger}$	$0.08445^{\ddagger}$	$0.03432^{\ddagger}$	$0.00270^{\ddagger}$
3.1	$0.11105^{\ddagger}$	$0.06042^{\ddagger}$	$0.01660^{\ddagger}$	$0.08215^{\ddagger}$	$0.03360^{\ddagger}$	$0.00271^{\ddagger}$
3.2	$0.11197^{\ddagger}$	$0.05980^{\ddagger}$	$0.01588^{\ddagger}$	$0.08293^{\ddagger}$	$0.03350^{\ddagger}$	$0.00240^{\ddagger}$
3.3	$0.11049^{\ddagger}$	$0.05967^{\ddagger}$	$0.01531^{\ddagger}$	$0.08235^{\ddagger}$	$0.03247^{\ddagger}$	$0.00245^{\ddagger}$
3.4	$0.11072^{\ddagger}$	$0.05927^{\ddagger}$	$0.01568^{\ddagger}$	$0.08221^{\ddagger}$	$0.03352^{\ddagger}$	$0.00289^{\ddagger}$
3.5	$0.11087^{\ddagger}$	$0.05877^{\ddagger}$	$0.01491^{\ddagger}$	$0.08278^{\ddagger}$	$0.03294^{\ddagger}$	$0.00239^{\ddagger}$
3.6	$0.11114^{\ddagger}$	$0.05935^{\ddagger}$	$0.01510^{\ddagger}$	$0.08267^{\ddagger}$	$0.03378^{\ddagger}$	$0.00229^{\ddagger}$
3.7	$0.10934^{\ddagger}$	$0.05709^{\ddagger}$	$0.01464^{\ddagger}$	$0.08059^{\ddagger}$	$0.03122^{\ddagger}$	$0.00223^{\ddagger}$
3.8	$0.11162^{\ddagger}$	$0.06066^{\ddagger}$	$0.01594^{\ddagger}$	$0.08352^{\ddagger}$	$0.03384^{\ddagger}$	$0.00256^{\ddagger}$
3.9	$0.10902^{\ddagger}$	$0.05904^{\ddagger}$	$0.01531^{\ddagger}$	$0.08202^{\ddagger}$	$0.03336^{\ddagger}$	$0.00254^{\ddagger}$

Table 1c: Empirical sizes of Davidson and Flachaire's (2008) wild bootstrap tests at 10%, 5% and 1% significance levels for $n_1 = 10$ and $n_2 = 50$.

	Wild bootstrap (unrestricted)			Wild bootstrap (restricted)		
σ_2	10%	5%	1%	10%	5%	1%
0.1	$0.12597^{\ddagger}$	$0.07243^{\ddagger}$	$0.02563^{\ddagger}$	$0.11190^{\ddagger}$	$0.06058^{\ddagger}$	$0.01833^{\ddagger}$
0.2	$0.12709^{\ddagger}$	$0.07376^{\ddagger}$	$0.02688^{\ddagger}$	$0.11198^{\ddagger}$	$0.06081^{\ddagger}$	$0.01862^{\ddagger}$
0.3	$0.12831^{\ddagger}$	$0.07453^{\ddagger}$	$0.02770^{\ddagger}$	$0.11207^{\ddagger}$	$0.06209^{\ddagger}$	$0.01854^{\ddagger}$
0.4	$0.12599^{\ddagger}$	$0.07318^{\ddagger}$	$0.02696^{\ddagger}$	$0.11023^{\ddagger}$	$0.06019^{\ddagger}$	$0.01695^{\ddagger}$
0.5	$0.12711^{\ddagger}$	$0.07446^{\ddagger}$	$0.02663^{\ddagger}$	$0.10981^{\ddagger}$	$0.05923^{\ddagger}$	$0.01655^{\ddagger}$
0.6	$0.12837^{\ddagger}$	$0.07485^{\ddagger}$	$0.02601^{\ddagger}$	$0.10973^{\ddagger}$	$0.05877^{\ddagger}$	$0.01569^{\ddagger}$
0.7	$0.12618^{\ddagger}$	$0.07389^{\ddagger}$	$0.02628^{\ddagger}$	$0.10680^{\ddagger}$	$0.05697^{\ddagger}$	$0.01432^{\ddagger}$
0.8	$0.12619^{\ddagger}$	$0.07375^{\ddagger}$	$0.02635^{\ddagger}$	$0.10616^{\ddagger}$	$0.05584^{\ddagger}$	$0.01285^{\ddagger}$
0.9	$0.12631^{\ddagger}$	$0.07342^{\ddagger}$	$0.02477^{\ddagger}$	$0.10513^{\ddagger}$	$0.05339^{\ddagger}$	$0.01226^{\ddagger}$
1.0	$0.12448^{\ddagger}$	$0.07204^{\ddagger}$	$0.02437^{\ddagger}$	$0.10317^{\ddagger}$	$0.05330^{\ddagger}$	$0.01142^{\ddagger}$
1.1	$0.12487^{\ddagger}$	$0.07193^{\ddagger}$	$0.02398^{\ddagger}$	$0.10258^{\ddagger}$	$0.05210^{\ddagger}$	$0.01089^{\ddagger}$
1.2	$0.12470^{\ddagger}$	$0.07067^{\ddagger}$	$0.02268^{\ddagger}$	0.10181^{*}	0.05112	0.00987
1.3	$0.12277^{\ddagger}$	$0.06982^{\ddagger}$	$0.02211^{\ddagger}$	0.09888	0.04928	$0.00912^{\ddagger}$
1.4	$0.12203^{\ddagger}$	$0.06858^{\ddagger}$	$0.02121^{\ddagger}$	$0.09760^{\dagger}$	$0.04805^{\ddagger}$	$0.00844^{\ddagger}$
1.5	$0.12221^{\ddagger}$	$0.06920^{\ddagger}$	$0.02056^{\ddagger}$	0.09815^{*}	$0.04700^{\ddagger}$	$0.00754^{\ddagger}$
1.6	$0.12244^{\ddagger}$	$0.06921^{\ddagger}$	$0.02090^{\ddagger}$	$0.09676^{\ddagger}$	$0.04643^{\ddagger}$	$0.00772^{\ddagger}$
1.7	$0.12062^{\ddagger}$	$0.06714^{\ddagger}$	$0.01940^{\ddagger}$	$0.09538^{\ddagger}$	$0.04452^{\ddagger}$	$0.00710^{\ddagger}$
1.8	$0.11938^{\ddagger}$	$0.06693^{\ddagger}$	$0.01896^{\ddagger}$	$0.09452^{\ddagger}$	$0.04241^{\ddagger}$	$0.00657^{\ddagger}$
1.9	$0.11775^{\ddagger}$	$0.06499^{\ddagger}$	$0.01845^{\ddagger}$	$0.09167^{\ddagger}$	$0.04193^{\ddagger}$	$0.00572^{\ddagger}$
2.0	$0.11610^{\ddagger}$	$0.06273^{\ddagger}$	$0.01676^{\ddagger}$	$0.08912^{\ddagger}$	$0.03943^{\ddagger}$	$0.00518^{\ddagger}$
2.1	$0.11776^{\ddagger}$	$0.06455^{\ddagger}$	$0.01770^{\ddagger}$	$0.09141^{\ddagger}$	$0.04084^{\ddagger}$	$0.00540^{\ddagger}$
2.2	$0.11441^{\ddagger}$	$0.06255^{\ddagger}$	$0.01676^{\ddagger}$	$0.08870^{\ddagger}$	$0.04003^{\ddagger}$	$0.00487^{\ddagger}$
2.3	$0.11554^{\ddagger}$	$0.06142^{\ddagger}$	$0.01566^{\ddagger}$	$0.08892^{\ddagger}$	$0.03834^{\ddagger}$	$0.00469^{\ddagger}$
2.4	$0.11472^{\ddagger}$	$0.06287^{\ddagger}$	$0.01611^{\ddagger}$	$0.08951^{\ddagger}$	$0.03835^{\ddagger}$	$0.00480^{\ddagger}$
2.5	$0.11448^{\ddagger}$	$0.06104^{\ddagger}$	$0.01533^{\ddagger}$	$0.08777^{\ddagger}$	$0.03765^{\ddagger}$	$0.00480^{\ddagger}$
2.6	$0.11042^{\ddagger}$	$0.05987^{\ddagger}$	$0.01543^{\ddagger}$	$0.08482^{\ddagger}$	$0.03705^{\ddagger}$	$0.00401^{\ddagger}$
2.7	$0.11203^{\ddagger}$	$0.05984^{\ddagger}$	$0.01493^{\ddagger}$	$0.08592^{\ddagger}$	$0.03733^{\ddagger}$	$0.00414^{\ddagger}$
2.8	$0.11355^{\ddagger}$	$0.06088^{\ddagger}$	$0.01538^{\ddagger}$	$0.08777^{\ddagger}$	$0.03686^{\ddagger}$	$0.00409^{\ddagger}$
2.9	$0.11138^{\ddagger}$	$0.05926^{\ddagger}$	$0.01450^{\ddagger}$	$0.08619^{\ddagger}$	$0.03565^{\ddagger}$	$0.00378^{\ddagger}$
3.0	$0.11097^{\ddagger}$	$0.05811^{\ddagger}$	$0.01448^{\ddagger}$	$0.08457^{\ddagger}$	$0.03551^{\ddagger}$	$0.00369^{\ddagger}$
3.1	$0.11064^{\ddagger}$	$0.05905^{\ddagger}$	$0.01420^{\ddagger}$	$0.08577^{\ddagger}$	$0.03604^{\ddagger}$	$0.00368^{\ddagger}$
3.2	$0.10864^{\ddagger}$	$0.05767^{\ddagger}$	$0.01435^{\ddagger}$	$0.08330^{\ddagger}$	$0.03479^{\ddagger}$	$0.00356^{\ddagger}$
3.3	$0.11000^{\ddagger}$	$0.05709^{\ddagger}$	$0.01419^{\ddagger}$	$0.08408^{\ddagger}$	$0.03466^{\ddagger}$	$0.00353^{\ddagger}$
3.4	$0.10859^{\ddagger}$	$0.05639^{\ddagger}$	$0.01370^{\ddagger}$	$0.08323^{\ddagger}$	$0.03379^{\ddagger}$	$0.00324^{\ddagger}$
3.5	$0.10881^{\ddagger}$	$0.05698^{\ddagger}$	$0.01385^{\ddagger}$	$0.08321^{\ddagger}$	$0.03491^{\ddagger}$	$0.00377^{\ddagger}$
3.6	$0.10922^{\ddagger}$	$0.05653^{\ddagger}$	$0.01286^{\ddagger}$	$0.08353^{\ddagger}$	$0.03421^{\ddagger}$	$0.00295^{\ddagger}$
3.7	$0.10916^{\ddagger}$	$0.05703^{\ddagger}$	$0.01339^{\ddagger}$	$0.08361^{\ddagger}$	$0.03451^{\ddagger}$	$0.00315^{\ddagger}$
3.8	$0.10708^{\ddagger}$	$0.05470^{\ddagger}$	$0.01283^{\ddagger}$	$0.08170^{\ddagger}$	$0.03242^{\ddagger}$	$0.00310^{\ddagger}$
3.9	$0.10931^{\ddagger}$	$0.05849^{\ddagger}$	$0.01412^{\ddagger}$	$0.08426^{\ddagger}$	$0.03480^{\ddagger}$	$0.00313^{\ddagger}$

Table 2a: Empirical sizes of asymptotic and bootstrap tests at 10%, 5% and 1% significance levels for $n_1 = 20$ and $n_2 = 50$.

	Asymptotic test			Ordinary bootstrap		
σ_2	10%	5%	1%	10%	5%	1%
0.1	0.12725‡	0.07468‡	0.02378‡	0.09751‡	0.04806‡	0.00905‡
0.2	0.12700‡	0.07492‡	0.02350‡	0.09803†	0.04970	0.01018
0.3	0.12439‡	0.07231‡	0.02245‡	0.09817*	0.04866*	0.00995
0.4	0.12532‡	0.07121‡	0.02229‡	0.09896	0.04872*	0.01046
0.5	0.12401‡	0.07086‡	0.02140‡	0.10035	0.05040	0.01082‡
0.6	0.12128‡	0.06934‡	0.02101‡	0.09951	0.05023	0.01123‡
0.7	0.12178‡	0.06907‡	0.02037‡	0.10096	0.05156†	0.01116‡
0.8	0.11983‡	0.06693‡	0.01936‡	0.09945	0.04984	0.01105‡
0.9	0.11971‡	0.06708‡	0.01897‡	0.09993	0.05034	0.01112‡
1.0	0.12312‡	0.06852‡	0.01958‡	0.10304‡	0.05223‡	0.01104‡
1.1	0.11884‡	0.06651‡	0.01858‡	0.10002	0.05062	0.01094‡
1.2	0.11785‡	0.06542‡	0.01823‡	0.09986	0.05051	0.01098‡
1.3	0.12061‡	0.06727‡	0.01859‡	0.10292‡	0.05229‡	0.01113‡
1.4	0.11742‡	0.06548‡	0.01812‡	0.10026	0.05079	0.01109‡
1.5	0.12033‡	0.06693‡	0.01821‡	0.10246‡	0.05162†	0.01066†
1.6	0.11813‡	0.06657‡	0.01801‡	0.10109	0.05257‡	0.01096‡
1.7	0.11677‡	0.06473‡	0.01770‡	0.09973	0.05034	0.01080†
1.8	0.11877‡	0.06572‡	0.01833‡	0.10152	0.05111	0.01125‡
1.9	0.11964‡	0.06536‡	0.01772‡	0.10164*	0.05088	0.01077†
2.0	0.11845‡	0.06507‡	0.01777‡	0.10058	0.05055	0.01108‡
2.1	0.11820‡	0.06600‡	0.01780‡	0.10055	0.05132*	0.01114‡
2.2	0.11702‡	0.06475‡	0.01804‡	0.10056	0.05066	0.01140‡
2.3	0.11561‡	0.06405‡	0.01724‡	0.09905	0.04987	0.01061*
2.4	0.11817‡	0.06459‡	0.01720‡	0.10040	0.05081	0.01075†
2.5	0.11839‡	0.06535‡	0.01801‡	0.10148	0.05131*	0.01110‡
2.6	0.11704‡	0.06481‡	0.01704‡	0.10010	0.05066	0.01072†
2.7	0.11762‡	0.06461‡	0.01801‡	0.10036	0.05086	0.01107‡
2.8	0.11877‡	0.06452‡	0.01761‡	0.10204†	0.05043	0.01088‡
2.9	0.11740‡	0.06477‡	0.01696‡	0.10076	0.05073	0.01044
3.0	0.11671‡	0.06388‡	0.01681‡	0.09984	0.05026	0.01052*
3.1	0.11777‡	0.06466‡	0.01685‡	0.10105	0.05063	0.01098‡
3.2	0.11552‡	0.06361‡	0.01740‡	0.09926	0.04991	0.01082‡
3.3	0.11824‡	0.06597‡	0.01759‡	0.10267‡	0.05232‡	0.01156‡
3.4	0.11639‡	0.06466‡	0.01751‡	0.10092	0.05062	0.01143‡
3.5	0.11613‡	0.06367‡	0.01709‡	0.09974	0.05036	0.01098‡
3.6	0.11499‡	0.06374‡	0.01564‡	0.09931	0.04990	0.00999
3.7	0.11573‡	0.06343‡	0.01687‡	0.10061	0.04994	0.01094‡
3.8	0.11711‡	0.06446‡	0.01737‡	0.10144	0.05126*	0.01122‡
3.9	0.11700‡	0.06433‡	0.01745‡	0.10167*	0.05140†	0.01090‡

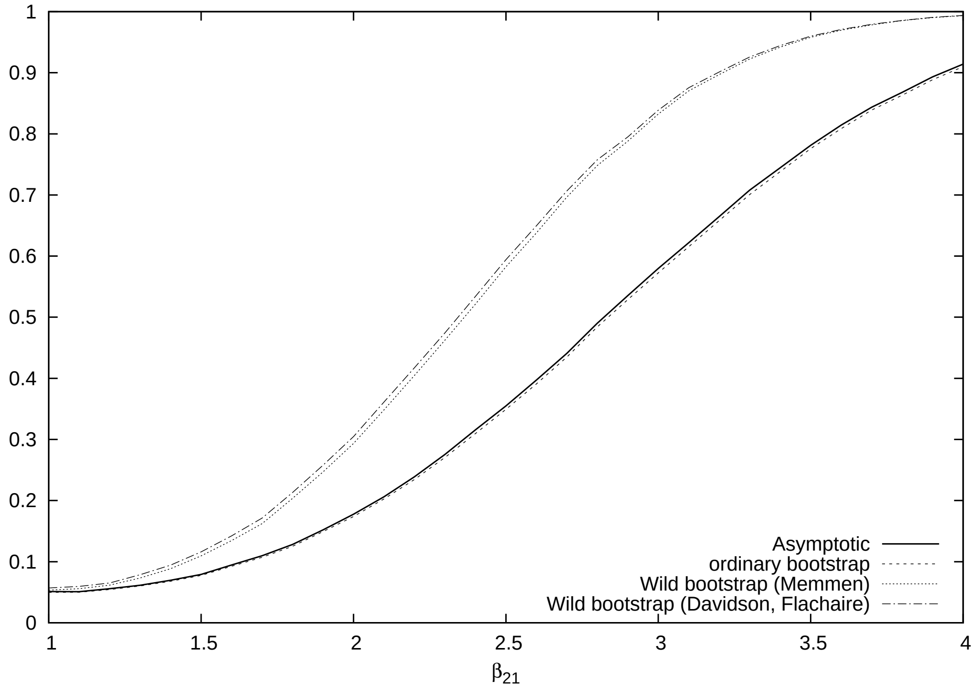
Table 2b: Empirical sizes of Mammen's (1993) wild bootstrap tests at 10%, 5% and 1% significance levels for $n_1 = 20$ and $n_2 = 50$.

	Wild bootstrap (unrestricted)			Wild bootstrap (restricted)		
σ_2	10%	5%	1%	10%	5%	1%
0.1	0.12517‡	0.07191‡	0.02347‡	0.10738‡	0.05498‡	0.01197‡
0.2	0.12595‡	0.07384‡	0.02370‡	0.10826‡	0.05688‡	0.01220‡
0.3	0.12507‡	0.07196‡	0.02364‡	0.10769‡	0.05540‡	0.01179‡
0.4	0.12414‡	0.07177‡	0.02276‡	0.10680‡	0.05503‡	0.01180‡
0.5	0.12195‡	0.06953‡	0.02214‡	0.10527‡	0.05394‡	0.01114‡
0.6	0.12114‡	0.06846‡	0.02086‡	0.10415‡	0.05304‡	0.01062†
0.7	0.11976‡	0.06873‡	0.02080‡	0.10505‡	0.05340‡	0.01074†
0.8	0.11822‡	0.06635‡	0.01926‡	0.10350‡	0.05176†	0.00994
0.9	0.11592‡	0.06393‡	0.01868‡	0.10089	0.05011	0.00971
1.0	0.11525‡	0.06320‡	0.01800‡	0.10031	0.05009	0.00932†
1.1	0.11485‡	0.06233‡	0.01736‡	0.10065	0.04930	0.00905‡
1.2	0.11435‡	0.06324‡	0.01684‡	0.10025	0.05009	0.00893‡
1.3	0.11482‡	0.06224‡	0.01685‡	0.10106	0.05010	0.00910‡
1.4	0.11246‡	0.06096‡	0.01604‡	0.09991	0.04892	0.00867‡
1.5	0.11099‡	0.06010‡	0.01603‡	0.09803†	0.04888	0.00884‡
1.6	0.10995‡	0.05911‡	0.01480‡	0.09796†	0.04787‡	0.00812‡
1.7	0.10877‡	0.05872‡	0.01499‡	0.09668‡	0.04762‡	0.00829‡
1.8	0.10880‡	0.05758‡	0.01464‡	0.09818*	0.04697‡	0.00817‡
1.9	0.10889‡	0.05723‡	0.01466‡	0.09728‡	0.04665‡	0.00830‡
2.0	0.10742‡	0.05692‡	0.01472‡	0.09557‡	0.04674‡	0.00822‡
2.1	0.10848‡	0.05729‡	0.01415‡	0.09699‡	0.04682‡	0.00816‡
2.2	0.10846‡	0.05742‡	0.01489‡	0.09777†	0.04720‡	0.00874‡
2.3	0.10703‡	0.05559‡	0.01416‡	0.09579‡	0.04566‡	0.00795‡
2.4	0.10725‡	0.05591‡	0.01408‡	0.09580‡	0.04619‡	0.00817‡
2.5	0.10658‡	0.05580‡	0.01346‡	0.09576‡	0.04622‡	0.00771‡
2.6	0.10737‡	0.05562‡	0.01334‡	0.09654‡	0.04610‡	0.00787‡
2.7	0.10696‡	0.05643‡	0.01343‡	0.09594‡	0.04633‡	0.00785‡
2.8	0.10697‡	0.05708‡	0.01397‡	0.09695‡	0.04816‡	0.00843‡
2.9	0.10676‡	0.05653‡	0.01348‡	0.09672‡	0.04685‡	0.00814‡
3.0	0.10687‡	0.05698‡	0.01430‡	0.09670‡	0.04770‡	0.00873‡
3.1	0.10657‡	0.05685‡	0.01360‡	0.09721‡	0.04761‡	0.00849‡
3.2	0.10554‡	0.05604‡	0.01365‡	0.09570‡	0.04691‡	0.00837‡
3.3	0.10594‡	0.05563‡	0.01389‡	0.09725‡	0.04611‡	0.00820‡
3.4	0.10540‡	0.05686‡	0.01398‡	0.09720‡	0.04799‡	0.00822‡
3.5	0.10857‡	0.05737‡	0.01430‡	0.09840*	0.04812‡	0.00847‡
3.6	0.10592‡	0.05695‡	0.01426‡	0.09651‡	0.04776‡	0.00860‡
3.7	0.10599‡	0.05639‡	0.01387‡	0.09676‡	0.04721‡	0.00816‡
3.8	0.10561‡	0.05543‡	0.01407‡	0.09602‡	0.04732‡	0.00788‡
3.9	0.10620‡	0.05746‡	0.01337‡	0.09695‡	0.04820‡	0.00858‡

Table 2c: Empirical sizes of Davidson and Flachaire's (2008) wild bootstrap tests at 10%, 5% and 1% significance levels for $n_1 = 20$ and $n_2 = 50$.

	Wild bootstrap (unrestricted)			Wild bootstrap (restricted)		
σ_2	10%	5%	1%	10%	5%	1%
0.1	$0.11369^{\ddagger}$	$0.06088^{\ddagger}$	$0.01641^{\ddagger}$	$0.10585^{\ddagger}$	$0.05537^{\ddagger}$	$0.01377^{\ddagger}$
0.2	$0.11517^{\ddagger}$	$0.06313^{\ddagger}$	$0.01715^{\ddagger}$	$0.10720^{\ddagger}$	$0.05712^{\ddagger}$	$0.01373^{\ddagger}$
0.3	$0.11459^{\ddagger}$	$0.06155^{\ddagger}$	$0.01736^{\ddagger}$	$0.10679^{\ddagger}$	$0.05565^{\ddagger}$	$0.01370^{\ddagger}$
0.4	$0.11505^{\ddagger}$	$0.06321^{\ddagger}$	$0.01705^{\ddagger}$	$0.10644^{\ddagger}$	$0.05584^{\ddagger}$	$0.01385^{\ddagger}$
0.5	$0.11362^{\ddagger}$	$0.06131^{\ddagger}$	$0.01666^{\ddagger}$	$0.10517^{\ddagger}$	$0.05426^{\ddagger}$	$0.01318^{\ddagger}$
0.6	$0.11340^{\ddagger}$	$0.06120^{\ddagger}$	$0.01617^{\ddagger}$	$0.10446^{\ddagger}$	$0.05385^{\ddagger}$	$0.01228^{\ddagger}$
0.7	$0.11380^{\ddagger}$	$0.06213^{\ddagger}$	$0.01657^{\ddagger}$	$0.10520^{\ddagger}$	$0.05436^{\ddagger}$	$0.01289^{\ddagger}$
0.8	$0.11280^{\ddagger}$	$0.06058^{\ddagger}$	$0.01567^{\ddagger}$	$0.10420^{\ddagger}$	$0.05335^{\ddagger}$	$0.01197^{\ddagger}$
0.9	$0.11073^{\ddagger}$	$0.05876^{\ddagger}$	$0.01537^{\ddagger}$	0.10182^{*}	0.05117^{*}	$0.01139^{\ddagger}$
1.0	$0.10993^{\ddagger}$	$0.05832^{\ddagger}$	$0.01462^{\ddagger}$	0.10087	0.05068	$0.01077^{\dagger}$
1.1	$0.11136^{\ddagger}$	$0.05825^{\ddagger}$	$0.01498^{\ddagger}$	0.10141	0.05000	$0.01114^{\ddagger}$
1.2	$0.11102^{\ddagger}$	$0.05951^{\ddagger}$	$0.01406^{\ddagger}$	0.10133	$0.05163^{\dagger}$	0.01054^{*}
1.3	$0.11188^{\ddagger}$	$0.05898^{\ddagger}$	$0.01477^{\ddagger}$	$0.10206^{\dagger}$	$0.05165^{\dagger}$	$0.01094^{\ddagger}$
1.4	$0.11041^{\ddagger}$	$0.05757^{\ddagger}$	$0.01393^{\ddagger}$	0.10095	0.04933	0.01020
1.5	$0.10891^{\ddagger}$	$0.05760^{\ddagger}$	$0.01480^{\ddagger}$	0.09910	0.04983	0.01027
1.6	$0.10844^{\ddagger}$	$0.05710^{\ddagger}$	$0.01310^{\ddagger}$	0.09883	0.04867^{*}	$0.00918^{\ddagger}$
1.7	$0.10714^{\ddagger}$	$0.05663^{\ddagger}$	$0.01373^{\ddagger}$	$0.09798^{\dagger}$	0.04890	0.00956
1.8	$0.10752^{\ddagger}$	$0.05543^{\ddagger}$	$0.01338^{\ddagger}$	0.09878	$0.04791^{\ddagger}$	0.00943^{*}
1.9	$0.10752^{\ddagger}$	$0.05572^{\ddagger}$	$0.01326^{\ddagger}$	0.09823^{*}	$0.04773^{\ddagger}$	$0.00932^{\dagger}$
2.0	$0.10583^{\ddagger}$	$0.05513^{\ddagger}$	$0.01306^{\ddagger}$	$0.09660^{\ddagger}$	$0.04728^{\ddagger}$	$0.00909^{\ddagger}$
2.1	$0.10722^{\ddagger}$	$0.05583^{\ddagger}$	$0.01252^{\ddagger}$	$0.09799^{\dagger}$	$0.04761^{\ddagger}$	$0.00904^{\ddagger}$
2.2	$0.10734^{\ddagger}$	$0.05545^{\ddagger}$	$0.01349^{\ddagger}$	$0.09806^{\dagger}$	$0.04776^{\ddagger}$	0.00963
2.3	$0.10543^{\ddagger}$	$0.05346^{\ddagger}$	$0.01256^{\ddagger}$	$0.09662^{\ddagger}$	$0.04599^{\ddagger}$	$0.00880^{\ddagger}$
2.4	$0.10509^{\ddagger}$	$0.05431^{\ddagger}$	$0.01256^{\ddagger}$	$0.09659^{\ddagger}$	$0.04711^{\ddagger}$	$0.00879^{\ddagger}$
2.5	$0.10481^{\ddagger}$	$0.05414^{\ddagger}$	$0.01210^{\ddagger}$	$0.09581^{\ddagger}$	$0.04675^{\ddagger}$	$0.00858^{\ddagger}$
2.6	$0.10588^{\ddagger}$	$0.05366^{\ddagger}$	$0.01199^{\ddagger}$	$0.09702^{\ddagger}$	$0.04644^{\ddagger}$	$0.00853^{\ddagger}$
2.7	$0.10520^{\ddagger}$	$0.05457^{\ddagger}$	$0.01210^{\ddagger}$	$0.09672^{\ddagger}$	$0.04693^{\ddagger}$	$0.00837^{\ddagger}$
2.8	$0.10558^{\ddagger}$	$0.05509^{\ddagger}$	$0.01256^{\ddagger}$	$0.09704^{\ddagger}$	$0.04814^{\ddagger}$	$0.00925^{\dagger}$
2.9	$0.10524^{\ddagger}$	$0.05433^{\ddagger}$	$0.01230^{\ddagger}$	$0.09693^{\ddagger}$	$0.04684^{\ddagger}$	$0.00855^{\ddagger}$
3.0	$0.10501^{\ddagger}$	$0.05464^{\ddagger}$	$0.01297^{\ddagger}$	$0.09733^{\ddagger}$	$0.04746^{\ddagger}$	$0.00915^{\ddagger}$
3.1	$0.10526^{\ddagger}$	$0.05486^{\ddagger}$	$0.01216^{\ddagger}$	$0.09671^{\ddagger}$	$0.04817^{\ddagger}$	$0.00884^{\ddagger}$
3.2	$0.10428^{\ddagger}$	$0.05364^{\ddagger}$	$0.01189^{\ddagger}$	$0.09545^{\ddagger}$	$0.04720^{\ddagger}$	$0.00857^{\ddagger}$
3.3	$0.10489^{\ddagger}$	$0.05294^{\ddagger}$	$0.01226^{\ddagger}$	$0.09666^{\ddagger}$	$0.04617^{\ddagger}$	$0.00823^{\ddagger}$
3.4	$0.10452^{\ddagger}$	$0.05454^{\ddagger}$	$0.01211^{\ddagger}$	$0.09657^{\ddagger}$	$0.04771^{\ddagger}$	$0.00887^{\ddagger}$
3.5	$0.10635^{\ddagger}$	$0.05464^{\ddagger}$	$0.01243^{\ddagger}$	0.09839^{*}	$0.04810^{\ddagger}$	$0.00871^{\ddagger}$
3.6	$0.10429^{\ddagger}$	$0.05483^{\ddagger}$	$0.01251^{\ddagger}$	$0.09634^{\ddagger}$	$0.04714^{\ddagger}$	$0.00899^{\ddagger}$
3.7	$0.10403^{\ddagger}$	$0.05415^{\ddagger}$	$0.01223^{\ddagger}$	$0.09642^{\ddagger}$	$0.04734^{\ddagger}$	$0.00867^{\ddagger}$
3.8	$0.10383^{\ddagger}$	$0.05296^{\ddagger}$	$0.01203^{\ddagger}$	$0.09618^{\ddagger}$	$0.04677^{\ddagger}$	$0.00853^{\ddagger}$
3.9	$0.10410^{\ddagger}$	$0.05448^{\ddagger}$	$0.01210^{\ddagger}$	$0.09633^{\ddagger}$	$0.04746^{\ddagger}$	$0.00877^{\ddagger}$

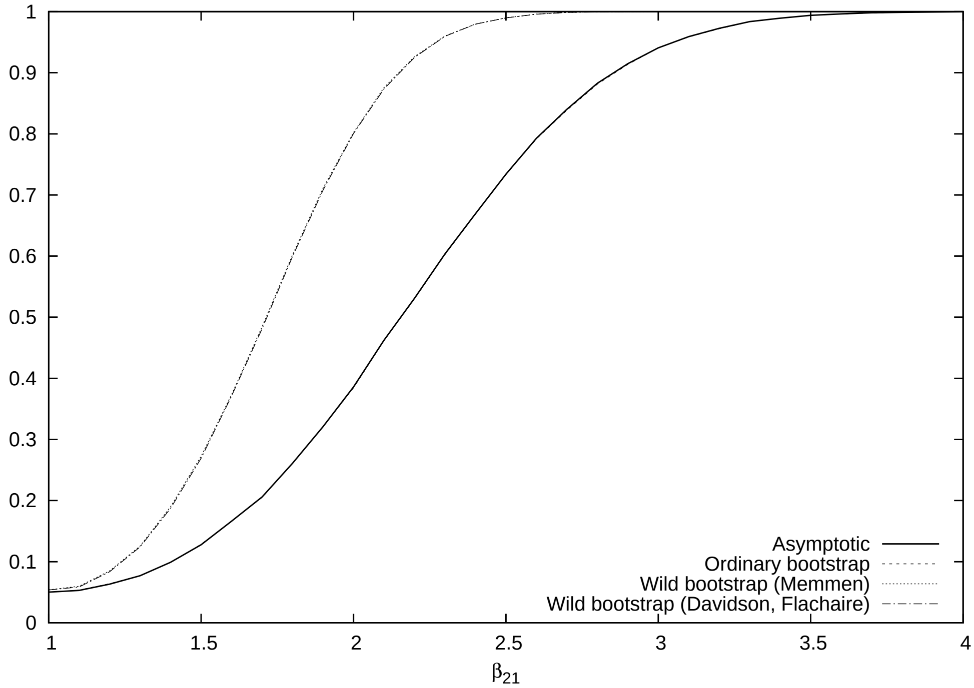
Figure 1: Power of the tests for $n_1 = n_2 = 15$ at the 5% significance level.



$\beta_{11} = \beta_{12} = \beta_{22} = 1.0$, $\sigma_1 = 1.0$ and $\sigma_2 = 2.0$. Since the size distortion of the asymptotic test is very severe even when $n_1 = n_2$, we first execute simulations to find the correct critical value of the asymptotic test, and calculated the size corrected power of the test. As for the bootstrap tests, size correction was not executed because their size distortions are not so severe. Also, the above tables indicate that the wild bootstrap tests based on the restricted estimator have better size performance than the ones based on the unrestricted estimator. Therefore, as for the wild bootstrap, we only show the powers obtained by the procedures based on the restricted estimator.

Figure 1 shows the power of the tests for $n_1 = n_2 = 15$ when the tests are conducted at the 5% significance level. From Figure 1, we observe that powers of the asymptotic test and ordinary bootstrap test are almost comparable. Also, wild bootstrap tests have much higher power than the asymptotic and ordinary bootstrap tests. The power of tests for $n_1 = n_2 = 50$ under the same parameter values are shown in Figure 2. From Figure 2, we can see that the powers of all tests get larger as n_1 and n_2 increase. Figure 2 indicates that the difference in the powers between asymptotic and ordinary bootstrap tests vanishes and these two tests have almost equivalent powers. Also, the difference between powers of wild bootstrap tests vanishes.

Figure 2: Power of the tests for $n_1 = n_2 = 50$ at the 5% significance level.



From the above results, we conclude that the ordinary bootstrap test has most accurate size than among the tests considered here, and the power of the ordinary bootstrap test is comparable to the asymptotic test. Also, though the sizes of the wild bootstrap tests are not so correct as the one obtained by the ordinary bootstrap test, the wild bootstrap tests based on the restricted estimator have higher powers than the asymptotic and ordinary bootstrap tests.

4 Concluding remarks

In this paper, we consider to apply the bootstrap methods to the Wald test statistic proposed by Watt (1979) for the equality of coefficients between two linear regressions under possible heteroscedasticity. As discussed by Ohtani and Kobayashi (1986), Thursby (1992) and others, tests based on the Wald test statistic proposed by Watt (1979) is more powerful than the one proposed by Jayatissa (1977). However, as stated by Thursby (1992), the drawback of the Wald test is either a requirement for numerical integration [Weerahandi (1987)] or the existence of the inconclusive region [Ohtani and Kobayashi (1986) and Kobayashi (1986)]. On the other hand, the bootstrap tests proposed in this paper is simple and easy to implement, and no inconclusive region exists. Our simulation results show that the ordinary bootstrap test is more reliable in size than the asymptotic test and that both tests are comparable in power. Also, though their sizes are not so correct as the ordinary bootstrap test, the wild bootstrap tests have much higher power than the asymptotic and ordinary bootstrap tests. Because of the way of the resampling, the ordinary bootstrap version of the test statistic can take $n_1^{n_1} \cdot n_2^{n_2}$ discrete values. However, the wild bootstrap version of the test statistic can take only $2^{n_1+n_2}$ discrete values. Since $2^{n_1+n_2}$ is much smaller than $n_1^{n_1} \cdot n_2^{n_2}$, the wild bootstrap method may not be able to mimic the tail probabilities of the test statistic so well as the ordinary bootstrap method. This may be the reason why the wild bootstrap test performs worse than the ordinary bootstrap test in terms of size. Further, as discussed in section 2, the bootstrap resampling is simplified because of the structure of the Wald test statistic. This suggests the usefulness of the bootstrap tests.

Also, in addition to the difficulty in implementation caused by the numerical integration, Weerahandi's (1987) test is valid only under the normality assumption of disturbances. Investigating the effect of the departure from the normality assumption is beyond the scope of this paper, however, the bootstrap test will be applicable under non-normality of disturbances because of the nature of the bootstrap methods.

In this paper, in order to examine the validity of the bootstrap methods for the model with structural breaks, we consider a simple model with a possible structural break and a known break point. Of course, it is possible to apply the methods considered in this paper to more general models. For example, as is discussed in introduction, some authors considered models with multiple breaks and unknown break points. Also, what happens if the true model is an AR model, or if the error terms are not independent? Further, what if the variance change occurs at the different point from the one where the change in regression coefficients occurs[2]? However, investigating such models are beyond the scope of this paper and a remaining problem for future research.

References

Andrews, D. W. K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Beran, R. (1987) Prepivoting to reduce level error of confidence sets. *Biometrika* 74, 457–468.

[2] Significance of the applications to this model and the AR models is suggested by an anonymous referee. The author is grateful for his insightful suggestions.

Beran, R. (1988) Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–697.

Boldea, O., A. R. Hall, & S. Han (2012) Asymptotic distribution theory for break point estimators in models estimated via 2SLS. *Econometric Reviews* 31, 1–33.

Chow, G. C. (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 591–605.

Davidson, R. & E. Flachaire (2008) The wild bootstrap, tamed at last. *Journal of Econometrics* 146, 162–169.

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

Flachaire, E. (2005) Bootstrapping heteroskedastic regression models: Wild bootstrap vs. pairs bootstrap. *Computational Statistics & Data Analysis* 49, 361–376.

Hall, A. R., S. Han, & O. Boldea (2012) Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of econometrics* 170, 281–302.

Hall, P. (1992) *The bootstrap and Edgeworth expansion*, Springer.

Honda, Y. & K. Ohtani (1986) Modified Wald tests in tests of equality between sets of coefficients in two linear regressions under heteroscedasticity. *The Manchester School* 54, 208–218.

Jayatissa, W. (1977) Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal. *Econometrica* 45, 1291–92.

Kobayashi, M. (1986) A bounds test of equality between sets of coefficients in two linear regressions when disturbance variances are unequal. *Journal of the American Statistical Association* 81, 510–513.

Liu, R. Y. (1988) Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics* 16, 1696–1708.

Mammen, E. (1993) Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics* 21, 255–285.

Ohtani, K. & M. Kobayashi (1986) A bounds test for equality between sets of coefficients in two linear regression models under heteroscedasticity. *Econometric Theory*, 220–231.

Ohtani, K. & T. Toyoda (1985) Small sample properties of tests of equality between sets of coefficients in two linear regressions under heteroscedasticity. *International Economic Review* 26, 37–44.

Perron, P. (2006) Dealing with structural breaks. *Palgrave handbook of econometrics* 1, 278–352.

Perron, P. & Y. Yamamoto (2014) A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. *Econometric Theory* 30, 491–507.

Phillips, P. C. (1986) The exact distribution of the Wald statistic. *Econometrica*, 881–895.

Rothenberg, T. J. (1984) Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica*, 827–842.

Schmidt, P. & R. Sickles (1977) Some further evidence on the use of the Chow test under heteroskedasticity. *Econometrica*, 1293–1298.

Thursby, J. G. (1992) A comparison of several exact and approximate tests for structural shift under heteroscedasticity. *Journal of Econometrics* 53, 363–386.

Toyoda, T. (1974) Use of the Chow test under heteroscedasticity. *Econometrica*, 601–608.

Watt, P. A. (1979) Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal: Some small sample properties. *The Manchester School* 47, 391–396.

Weerahandi, S. (1987) Testing regression equality with unequal variances. *Econometrica*, 1211–1215.

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